



# D19.4

## INITIAL 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:** 31.10.2015/M6

**Actual Date of Submission:** 31.10.2015/M6



Project internal reviewer(s):

<b>Project internal reviewer(s):</b>	<b>Beneficiary/Institution</b>
Werner Kutsch	ICOS ERIC

Document history:

<b>Date</b>	<b>Version</b>
1.9.2015	Draft for comments
30.10.2015	Accepted by WK



## INITIAL DATA MANAGEMENT PLAN

### ABSTRACT

This document presents the overall ENVRIplus data management plan (DMP), developed by the project Data Management Team in their workshops. This is the initial 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<sup>PLUS</sup> 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):

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. Test observational data from new sensors – these are rarely needed for long term preservation, but important during project period;
  - b. Data collection of experiences (previous and project-time) from RI representatives;
2. Generate common solutions for shared information technology and data related challenges of the environmental RIs in data and service discovery and use, workflow documentation, data citations methodologies, service virtualization, and user characterization and interaction;
  - a. Software products for data management and virtualization;
  - 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, but 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, which in some cases are still in development. Similarly, the Beneficiaries might have their own overarching 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.

### **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 can be included in the TNA agreements.
- This DMP covers observational data, results from human subject research (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 from 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 approval for doing the study. The template and procedure above should be used as the basis of the ethical approval requests.

### **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.

### **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) 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).

### **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 PLANS

Note: This is done in the initial DMP in data collection level, i.e. we do not follow each separate data set, but instead work on identified major types of data produced in the project. The DMT together with WP leaders found 7 data collection types, which will be described in this section:

1. Project documentation
2. Questionnaire and interview data
3. Software products
4. Testing data
5. Citizen Scientist collected data
6. Data collected by Transnational Access participants

### **1. Project documentation**

#### **Data collected**

Deliverables, meeting notes, project personnel lists, email lists, email messages, other project documents (agreements, MoUs, etc.) collected by the Project Management Team (PMT). The work related to these datasets is done in the Work Packages 18 and 19.

#### **Collection method**

Deliverable collection is defined in the ENVRIplus deliverable guidelines. Meeting notes are collected by the meeting rapporteurs in the ENVRIplus ActiveCollab site. Project personnel lists, including email addresses, are collected on ActiveCollab site by the persons in question. Email messages of the project management team are collected on personal email archives.

#### **Metadata and documentation**

Documents will be included with necessary documentation to understand them later on. Email archives have their internal metadata schema attached to messages. The formal metadata will be included to the documents when they are inserted to the project long term storage by the PMT.

#### **Ethical and privacy issues**

Deliverables have no privacy issues as they are intended and clearly labeled to be open knowledge. Meeting notes are collected to avoid any personal information. Personnel lists, including email addresses, are collected only for project purposes, and their subjects are clearly informed on their collection – only name, institution, email address and project tasks are stored. The information is done using the standard Finnish Personal Data Act (523/1999) section 10 declaration sheet (**ANNEX I**). Email messages have potential privacy issues, although only directly project-related emails are stored, and the control of the store is by the recipient (not overall project management).

#### **IPR issues**

ENVRIplus default.





### **Storage and backup during project time**

Documents are stored in the ENVRIplus ActiveCollab site. The contract includes back up services. Additional backups are made by the project staff. Email archives are regularly backed up by the PMT on personal level.

### **Access management**

ENVRIplus ActiveCollab has an Access management mechanism, controlled by the Project Management Team. Only Beneficiaries have access to the collected email addresses or personnel list. Deliverables are open to all and are actively published in the project public website. Other project documents (i.e. agreements, MoUs) access is controlled by the Project Management Team and handled in the ActiveCollab site. Access is generally open for all Beneficiaries as defined by the Consortium Agreement.

### **Retention and preservation**

Deliverables will be retained. The email lists and personal information will be deleted latest 4 years after project ending to maintain necessary communications for reporting and financial auditing purposes. For email archives and other documents, the decision on retention will be done before project end by the PMT, after consulting the project Executive Board.

### **Long term preservation plan**

Retained documents will be stored with the minimum of 10 years. The long term preservation is done by using one of the following facilities (to be decided at Project finalization): OpenAire for documents suitable for that service, EUDAT B2SHARE for deliverables, etc. For documents not suitable for either, the Finnish national TTA ETSIN service is used.

### **Sharing policy**

Deliverables are openly shared, unless specifically determined to be confidential (in Description of Action, none are). Email addresses are only available to project participants. Stored documents (e.g. MoUs) are only shared to project beneficiaries. Email archives are not shared on default.

### **Responsible person**

Deliverable, email list, personal information and general document management is by Magdalena Brus from DMT. Both Ari Asmi and Magdalena Brus are responsible for their own email archive. University of Helsinki will organize replacements for this staff if needed (personnel changes, long absences etc.).

### **Resources used**

ActiveCollab is funded by the ENVRIplus management budget. The work for data curation and storage is part of the Management PMs. Long term storage resources are institutional or handled by other EU funded activities (e.g. EUDAT2020).



## **2. Questionnaire/interview data**

### **Data collected**

Initially the questionnaires themselves, or audio recordings of the interview. For longer term storage, the reports collecting the findings of the questionnaires/interviews.

### **Collection method**

Individuals specifically selected for collecting such information collect the information. They are detailed in the Ethical Review of the questionnaire/interview done before the collection. The collection includes the initial information retrieval, analysis, consultation with the people interviewed (for correct interpretation) and synthesis. All participants are informed (via consent forms) of the data collection, processing and access before data collection.

### **Metadata and documentation**

The collected data sets will have machine created metadata. The end result (reports) will have the standard ENVRIplus documentation format (i.e. deliverable template) with included metadata.

### **Ethical and privacy issues**

All such activities go through ethical review, as they potentially include personal data and opinions. Procedures are summarized for project internal use in ANNEX II.

### **IPR issues**

ENVRIplus default.

### **Storage and backup during project time**

Raw documents are securely stored by the interviewing or collecting person. Responsibility of this storage and related backups is with this person. The interpreted reports are stored in the Project internal document management system, which have automatic back up systems.

### **Access management**

Access to the raw documents are only accessible to the person doing the interviews or data collection. He or she can provide access to the raw information to the participant during the after-interview discussion on conclusion for the final documentation. The final results in the form of the report are public, and meant for general use in the form of a deliverable. These documented final results will also be shared to collaborating H2020 projects, currently there is a data sharing MoU with EUDAT2020 project.

### **Retention and preservation**

The raw data (questionnaires and audio recordings) will be destroyed immediately after processing to the final report format. The resulting documents are stored as Deliverable reports (see earlier point).

### **Long term preservation plan**

For the deliverables, the long term preservation is explained in the previous section.

### **Sharing policy**

Deliverables are publicly available. None of the raw data will be shared in or out to the Project.



**Responsible person**

Each questionnaire/interview will have a responsible person, defined when doing the Ethical Review. The responsibility of final reports are with the Project Management Team.

**Resources used**

Local resources per partner with the raw results. Deliverables are using Project Management Team resources.



### **3. Software products**

#### **Data collected**

Developed software packages or modules. These are specifically those packages developed in the ENVRIplus project, i.e. these do not include existing software.

#### **Collection method**

Done during the project period, mostly in Theme II work packages. The data (software codes) are collected during work using suitable web-based versioning system.

#### **Metadata and documentation**

Documented by inline comments and documentation principles of each IT partner in question.

#### **Ethical and privacy issues**

None

#### **IPR issues**

All produced software will stay as the ownership of the original authors. However, all will need to be licensed using an open software license, as described in the Grant Agreement.

#### **Storage and backup during project time**

IT providers own service for version control will be used, with their internal backup systems.

#### **Access management**

Provided modules will have free access after project period. During the project work, the produced draft modules will be available to RI participants, and the author can also share it more widely. The access practicalities depend on the software platform, but these can be also shared via the project document management system.

#### **Retention and preservation**

Final versions of the program codes are stored and preserved. Draft versions can also be stored if the author or RI representative require this.

#### **Long term preservation plan**

The long term preservation of the software uses ENVRIplus standard methodologies (actual method will be chosen latest on project end).

#### **Sharing policy**

All final software products are to be shared publicly.

#### **Responsible person**

Individual researchers/programmers are responsible on version control, documentation and initial storage. WP leader is responsible to identify all software produced in his/her WP and to control that the final versions are made available.

#### **Resources used**

Partner individual version control systems. Long term storage: as ENVRIplus standard.



## **4. Testing data**

### **Data collected**

Data sets collected during testing of e.g. new equipment, or endurance tests. These are typically of only engineering interest, and do not have much additional value.

### **Collection method**

Depends on each test.

### **Metadata and documentation**

Only local metadata – the results of the instrument tests are reported in Deliverables, but actual testing data sets are not specifically documented.

### **Ethical and privacy issues**

None

### **IPR issues**

ENVRIplus default.

### **Storage and backup during project time**

During testing with the RI individual storage options, or ENVRIplus default methods.

### **Access management**

Only to WP in question.

### **Retention and preservation**

The WP leader will make the final decision on storing such test data. If the data is determined to be stored, it will then be included in the documentation process and will be stored according to ENVRIplus standard methodology.

### **Long term preservation plan**

Usually not retained. If retained as ENVRIplus standard.

### **Sharing policy**

Only within the WP. Retained data is openly available, and will be shared according to ENVRIplus standard methods.

### **Responsible person**

Persons doing the testing, WP leader in question (for retention decision)

### **Resources used**

Only local resources. If data is to be retained, then ENVRIplus standard.



## **5. Citizen Scientist collected data**

### **Data collected**

Data collected by Citizen Scientists during the (potential) Citizen Science pilot

### **Collection method**

Unknown yet – depends on the method chosen

### **Metadata and documentation**

This will be decided when the CS WP decides on the potential test case.

### **Ethical and privacy issues**

Most likely these will be issues. These will be investigated later in the project.

### **IPR issues**

ENVRiplus default.

### **Storage and backup during project time**

To be decided

### **Access management**

To be decided

### **Retention and preservation**

To be decided

### **Long term preservation plan**

If retained as ENVRiplus standard.

### **Sharing policy**

To be decided

### **Responsible person**

WP leader of Citizen Science.

### **Resources used**

To be decided



## **6. Data collected by Transnational Access participants**

### **Data collected**

Data sets collected Transdisciplinary Transnational Access Pilot (TTAP). The content will depend on selected TTAP projects.

### **Collection method**

Depends on each TTAP project.

### **Metadata and documentation**

We will require clear documentation and (discipline specific) typical and complete metadata records.

### **Ethical and privacy issues**

Depends on TTAP, but unlikely.

### **IPR issues**

The data ownership must remain with the TTAP participant, but we require these data to be openly available (as in the ENVRIplus Grant Agreement).

### **Storage and backup during project time**

Up to the TTAP participant. However, the Project will provide support for this and will require TTAP proposals to explain how this is done.

### **Access management**

To be decided, but should follow ENVRIplus normal procedures.

### **Retention and preservation**

To be decided, in collaboration with the TTAP participant.

### **Long term preservation plan**

Up to the TTAP participant. However, the Project will provide support for this and will require TTAP proposals to explain how this is done.

### **Sharing policy**

Data must be made open and accessible. All produced data sets must follow the ENVRIplus general directions, although exceptions can be made if necessary, on decision of ENVRIplus Executive Board.

### **Responsible person**

TTAP participant and the TTAP WP leader.

### **Resources used**

TTAP participants' own data services, General ENVRIplus resources.



## 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.





## APPENDICES

### ANNEX I DESCRIPTION OF PROJECT INTERNAL PERSONNEL INFORMATION STORAGE

#### DESCRIPTION OF THE FILE

Personal Data Act (523/1999) section 10

Date 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"

OFFICE OF THE DATA PROTECTION OMBUDSMAN      MODEL FORM      25 November 2004  
Tel: +358 10 36 66700  
Tel: +358 10 36 16670 (information service 9 a.m. to 3 p.m.) Fax: +358 10 36 66735 [www.tietosuojafi.fi](http://www.tietosuojafi.fi)

[Unofficial translation]



<p>5. Content of the register</p> <p>(For instance: name, address and telephone number of the data subject)</p>	<p>Name, professional title, address, telephone number, email address, institution they work in, (potentially) associated Research Infrastructure, involvement on specific tasks in the project.</p>
<p>6. Regular sources of information *</p> <p>(Which data is received, by whom and on what ground. For instance: consent or provision of a law)</p>	<p>The information is collected from the project participants, specifically from the Primary Investigators of project Beneficiaries.</p>
<p>7. Regular destinations of disclosed data and whether the data is transferred to countries outside the the European Union or the European Economic Area</p>	<p>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.</p>
<p>8. The principles how the data file/register is secured.</p>	<p>A. Manual register (place of storage and the methods of protection)</p> <p>Copy of the register is stored in encrypted drive of the personal work computer of M. Brus.</p> <p>B. Data register/ADP register (principles for the right to use a register, monitoring of the use and actual protection of hardware)</p> <p>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.</p>

\* 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.

OFFICE OF THE DATA PROTECTION OMBUDSMAN      MODEL FORM      25 November 2004  
 Tel: +358 10 36 66700  
 Tel: +358 10 36 16670 (information service 9 a.m. to 3 p.m.) Fax: +358 10 36 66735 [www.tietosuojaja.fi](http://www.tietosuojaja.fi)

[Unofficial translation]



## ANNEX II

### PROCEDURE FOR QUESTIONNAIRES AND INTERVIEWS

1. Report immediately on plans to generate such actions. The Project Management Team must be informed about this issues early enough. This reporting should include the following information:
  - a. What information is needed
  - b. Who are the targets of the request (as detailed as possible)
  - c. When this information is needed
  - d. Could this be done collectively (i.e. combining many questionnaires or interviews to one)
2. The PMT will try to combine these to smaller amount of total interviews/questionnaires
3. Each interview/questionnaire must have a partner and a person responsible for the action.
4. This person must then fill out the Ethical Considerations report (see ANNEX III for an example), contact the Qualified Ethical Board of their institution (or country).
  - a. We specifically require that this report must follow the Questionnaire/Interview requirements of the Data management Plan.
  - b. If necessary additions and changes must be done based on the interaction with the Qualified Ethical Board.
5. The Qualified Ethical Board must then provide acceptance letter in writing, which is then stored in the Document Management system
6. The actual action can only be done after submitting such acceptance letter (and typically the consent form from the interviewed person).
7. The reports based on the interviews and questionnaires must be also shared to the interviewed person.



## 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.



# ENVRIplus Task 5.1 Ethics Considerations

---

*Malcolm Atkinson and Cristina Alexandru*  
*14 August 2015*

## Context

The European H2020 project ENVRIplus, [www.envriplus.eu](http://www.envriplus.eu), will facilitate interoperability between or improve the sustainability and cost-effectiveness of the participating Research Infrastructures (RI), which are all involved in environmental research. The majority are endorsed by the *European Strategy Forum on Research Infrastructures* (ESFRI) <https://ec.europa.eu/research/infrastructures>. ENVRIplus is more fully described in the following abstract:

ENVRIPLUS is a cluster of research infrastructures (RIs) for 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-fertilisation 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 multi-disciplinary 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 socio-economic 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 multi-disciplinary 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.

Theme 2 of ENVRIplus is focused on identifying, prototyping and developing key information and communications technology (ICT) elements, particularly those concerning the full research-data lifecycle, to improve the capabilities, interoperability and sustainability of the cluster of RIs.

Work Package 5 builds on the work of the antecedent project ENVRI, [envri.eu](http://envri.eu), to establish guidelines for the design of common services, systems and practices.

The Task 5.1 is part of Work Package 5. Its function is to gather and collate requirements, predominantly ICT related, and to report them as a coherent input to subsequent tasks. Task 5.1 is led by Malcolm Atkinson, assisted by Rosa Filgueira and Cristina Alexandru, in the School of Informatics, University of

Edinburgh (UoE). It is undertaken by the staff of a number of consortium partners in several EU countries, and Norway and Switzerland.

## Purpose

The purpose of Task 5.1 is to gather and integrate an initial overall requirements analysis and to develop an initial survey of technology and engineering options. It will refresh the requirements gathering performed and formalised by its predecessor, ENVRI <http://envri.eu>, where the RIs were previously partners in ENVRI and take an initial look, where they are new in ENVRIplus.

The primary focus of ENVRIplus is to help with data for research, but this needs to be set in the context of the factors governing the data life cycle in each RI, the related computational and data transmission requirements, the communities associated with each RI, their existing technologies and commitments, and their research goals. All of those involved directly in the requirements gathering will be explicitly identified by name and institution on the reports produced as they take responsibility and credit for the quality of the information gathered. *No other required information concerns individuals and if any third-party individual is mentioned for any reason, they should be pseudonymised by the procedure that follows.*

Though information about individuals is not required, it may on a few occasions be helpful to discuss a third-party individual not directly involved in the requirements gathering process, e.g. to clarify a role, researcher behaviour or skill set. On all such occasions their identity should be systematically hidden by using a consistent naming scheme as follows:

- A prefix corresponding to the context of the conversation, e.g., ACTRIS
- A suffix as an integer randomly chosen from 0..99.
- To enable a consistent mapping, the ICT go between should maintain a carefully concealed map from true identities to the above token so that a pseudonymised individual has a consistent pseudonym in the written reports.
- The map should not be supplied to anyone else inside or outside the ENVRIplus project.
- The map should be deleted when Task 5.1 is complete or earlier, when conversation about requirements with that RI is complete.

Care must be taken not to provide other information that would allow their identity to be inferred.

## Analysis of Ethical Issues

There are no long-term recorded data that cannot be made public, as all of the recorded information will be about RIs as a whole, collated into a published report. The only identified individuals are those who are authoring the parts of the required material. They are only identified by their name and the institution that employs them during the requirements gathering process.

None of the individuals collecting requirements in Task 5.1 are chosen (by the process described below) on any other basis than their role in their organisation, their RI and their role in the ENVRIplus project. They are all adults capable of deciding whether to proceed with the requirements gathering discussions and that agreement will be made explicit and recorded at the start of such discussions using a consent form provided in the Appendix.

There is no sensitive information involved. The only personal information is the names and institutions of those collecting the requirements and developing the written reports, which will be made public within the reports. Any other personal information will be kept confidential.

The written record of each discussion will not be used until those who have been involved in the discussions have each agreed that it is correct and fit for its planned use.

The uses of the resulting requirements analyses will be subsequent tasks in the ENVRIplus project, the public reports developed by the project and the partners in the project within the constraints of the Consortium Agreement. These uses may be extended, e.g. to collaborate with other H2020 projects pursuing similar issues, only after an MoU between ENVRIplus and such projects has been formally agreed.

## Procedures and Methods

### Identification of participants

For the purposes of Task 5.1 the participants undertaking the requirements gathering process fill four roles:

1. **Task Leaders:** who focus on a particular technical issue from the ICT viewpoint and shape the questions about requirements for their topic. They volunteered from the community of people assigned person months in Theme 2. In many cases, they are going to lead subsequent tasks using the information. Their selection was reviewed and confirmed by the weekly Theme 2 conference call.
2. **Research Infrastructure Representatives:** who gather and provide information about their RI. The leaders or formal contacts in ENVRIplus for each RI were contacted requesting that they nominate a representative. Replies contained their nominees. The University of Edinburgh group oversaw this process and the weekly Theme 2 conference call validated the selection.
3. **Go Betweens:** ICT experts who will communicate with a small number (typically 2 or 3) Research Infrastructure Representatives on behalf of the topic leaders, and also gather general information. They will ensure that the procedures, e.g. pseudonymisation, are properly conducted, and will help in the production of the required report. These were volunteers, from the group assigned person months in Task 5.1. The Theme 2 weekly meeting considered their suitability.

4. **Additional sub-topic experts:** In the event of the requirements gathering needing a particular expertise, more detail than the above group can handle, or more effort, those involved in a particular requirement gathering discussion may identify delegates to handle a sub-topic. These delegates from partner institutions may be put forward by a topic leader or by an RI representative. Their appointment will be reviewed and ratified by the Theme 2 weekly call, and recorded by the University of Edinburgh team. They will then conduct their requirements gathering under exactly the same controls and conditions, e.g. consent form, as govern the other three groups of contributing requirements gatherers.

All four of the above roles in requirement gathering may contribute to the written record of requirements. The Go Betweens will take responsibility for ensuring the discussions, development of the text and agreement that the text is ready for wider use, is compliant with these ethical procedures. This means that all of those contributing must use the agreed procedures including pseudonymisation of third party identities and exclusion of confidential personal data other than the names and institutions of those directly compiling the material.

#### Information consent

There is no personal or sensitive information; however, at the start of each requirements gathering discussion, which will typically involve several sessions by tele-conference services or face-to-face meetings, the Go Between or Topic Leaders, or their delegates will present a consent form, which the Research Infrastructure Representative and they will sign to indicate that they are prepared to proceed. A copy of this form can be found in the appendix. If additional parties are involved at any stage of the discussion they should also sign the consent form.

In addition, there are two further safeguards:

1. **Release formality:** the outcome of a requirements gathering discussion will be a written record produced principally by the Go Between but may include contributions from Topic Leaders and sub-topic experts. It will not be passed into further use until all those who have contributed agree that it is correct and acceptable. This acceptance will be recorded as comments on a version held in the consortium's repository [envriplus.manageprojects.com](https://envriplus.manageprojects.com) in the <https://envriplus.manageprojects.com/projects/requirements/notebooks/470> Notebook.
2. **Acoustic recordings:** If those involved in a discussion all agree that a recording of their conversation would be helpful (and sign the section of the consent form indicating this) then a recording of the discussion may be made. It should be on a locally controlled medium at one of their sites, normally the Go Between's. It should be kept while the written report of the discussion is being prepared, and should only be accessed by those involved in the discussion. Once the participants in the discussion have agreed to the release of the written version, it should be deleted. Care



should be taken not to record inappropriate material and to protect the recordings against misuse.

### **Data protection**

There is no retained personal data. Only the names and institutions of those engaged in the requirements gathering discussion. The data will be retained in the consortium's repository, which will be subject to a data policy agreed by the consortium. Local copies used while the requirements gathering is underway will be deleted on or before the completion of Task 5.1.

### **Sensitive data**

There is no sensitive data.

### **Data collection procedures**

These are as set out above, and as detailed further in <https://envriplus.manageprojects.com/projects/requirements/notebooks/431/pages/25>.

### **Commitment to ethical standards**

The requirements will be gathered by members of the ENVRIplus project who are committed to and bound by the ethical standards and guidelines of Horizon2020. Throughout Task 5.1 the University of Edinburgh team will oversee this adherence to ethical standards.

## Appendix: Notes to be used with the consent form

### ENVRiplus Task 5.1- INFORMATION SHEET

#### Background

My name is *<please complete with your name>* and I am a team member on the *<please complete with the name of your institution>* team of the ENVRiplus project. ENVRiplus is a Horizon 2020 project aiming to develop software and services which can improve the interdisciplinarity and interoperability of Environmental Research infrastructures (RIs) across Europe. We are responsible for Task 5.1, the purpose of which is to gather and integrate requirements from different RIs, and review the technology and engineering tools that they use.

My role is that of a Go-between, which means that I communicate with RIs to gather their general and specific requirements on behalf of our topic leaders. I have been assigned to gather requirements from your RI.

#### What I would like from you

I would be very grateful if you would agree to have a few discussions with me by means of your preferred voice call tool (e.g. Skype) or face-to-face meetings. During these discussions, I will be interested in your responses to different sections of the list of questions attached to this document. The list contains generic questions, and specific questions which were proposed by our topic leader. For our initial discussion, I will only ask you the generic questions. For subsequent ones, I will invite you to propose the next topic to discuss. Should you ever feel that you cannot reply to a question, we could always postpone it to a later discussion. Alternatively, it would be of great help if you could indicate to me a person from your RI who could reply to it, or to the questions on a certain topic. I could then contact that person.

The purpose of our discussions will be to incrementally add information to a compiled list of requirements. After each discussion, I will write a report summarising it, which I will share with you for comments and amendments. Once you agree with the report, I will also forward it to the appropriate topic leaders, and I hope that you would agree to have a discussion with them, or with any delegates that they assign, in case they need any clarification (also by means of your preferred voice call tool). Also, whenever our discussion on a topic becomes too technical, I would be very grateful if you would agree to have a discussion directly with that topic's leader, or with his/her delegates. I will also share with you the final list of requirements for comments and amendments. You are free to propose the date and duration of each discussion, and to end the discussion at any time.

With your permission, I would like to audio record our discussions, which would help me with the production of reports. If you agree, each audio recording would only be held on a locally controlled medium within my institution, only I would have access to it, and I would delete it once we agree on the final version of the report. However, if you are not happy with being audio recorded, I would only take notes.

Your name and institution will be recorded as co-authors of the discussion on intermediary reports, and within our project's final deliverables. Please rest assured that any other personal information that you may reveal during our discussions will be kept confidential.

I would be happy to answer any questions regarding this information sheet. Thank you for the time taken in reading it.

## Appendix: The Consent form for requirements gathering CONSENT FORM

**Title of Project:** ENVRplus Task 5.1

**Date:**

1. I confirm that I have read and understand the information sheet dated  for the above project and have had the opportunity to ask questions	<input type="checkbox"/>
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.	<input type="checkbox"/>
3. I agree to take part in the discussion, help draft the internal reports, and I understand that these reports will not be used until I agree with their contents.	<input type="checkbox"/>
4. I agree for the discussions to be audio recorded, and for each recording to be kept until the associated report is completed.	<input type="checkbox"/>
5. I agree to my name appearing in the internal reports and project deliverables.	<input type="checkbox"/>

Please include your details below, in the format that you would like them to appear in the internal reports and project deliverables.

<b>Researcher representing the RI</b>	
Title:	Name:
Affiliation:	

<b>Researcher representing ICT</b>	
Title:	Name:
Affiliation:	

Signature of the Researcher representing the RI	Signature of the Researcher representing ICT