ENVRI^{plus} DELIVERABLE

D17.1
Domain report on development needs and actions

WORK PACKAGE 17

LEADING BENEFICIARY: UNIVERSITY OF HELSINKI

Author(s):	**Beneficiary/Institution:**
Paolo Laj	UHEL, Finland
Andreas Petzold	FZJ, Germany
Hélène Leau	IFREMER, France
Helen Glaves	BGS, UK
Michèle Boichard	INRA France
Jacques Roy	CNRS, France

Accepted by: Sanna Sorvari Sundet (WP17 Leader)

Deliverable type: REPORT

Dissemination level: PUBLIC

Deliverable due date: 30.4.2019/M42

Actual Date of Submission: 15.6.2019/M48

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

Project internal reviewer(s):

Project internal reviewer(s):	**Beneficiary/Institution**
Technical Reviewer	Leysan Karimova
RI Reviewer	Sanna Sorvari Sundet

Document history:

Date	**Version**
01.06.2019	Draft for comments
16.06.2019	Corrected version
27.06.2019	Accepted by Sanna Sorvari Sundet

1. Rationale for D17.1

The objective of ENVRIplus is to provide common solutions to shared challenges for European Environmental and Earth System Research Infrastructures in their efforts to deliver new services for science and society. To reach this overall goal, ENVRIplus brings together Environmental and Earth System Research Infrastructures, projects and networks together with technical specialist partners to build common synergic solutions for pressing issues in Research Infrastructure construction and implementation. ENVRIPLUS is driven by 3 overarching goals: 1) favouring cross-fertilization between Research Infrastructures, 2) implementing innovative concepts and devices across Infrastructures, and 3) facilitating research and innovation in the field of environment to an increasing number of users outside the Research Infrastructures.

While finding common solutions is the main objective of ENVRIPLus, space was left in the project to maintain a level of exchange driven by thematic areas: terrestrial and ecosystem domain, atmospheric domain, Marine domain and Solid Earth Domain. The main focus of this activity in ENVRIPlus is connected to the implementation and further development of services at RI level while ensuring the highest possible level of standardisation at the domain level. It is clear that, from the view of a RI operator, interaction at domain level remains key for some of the ENVRIPlus discussion. The task 17.1 is therefore aimed at developing and operating smooth communication means in the 4 respective domains. Goals are sharing vision, strategies and ways toward interoperability, common operations and services taking account different development levels, needs and interests of participating research infrastructures, RI networks and RI related projects within the domains. From there the work has been designed addressing three levels of implementation and development: 1) the RI level accounting for the fact that participating RIs are of certainly different maturity and have individual constraints for their implementation, 2) the subdomain level (Atmosphere, Marine, Solid Earth, and Biodiversity / Terrestrial Ecosystems) which currently corresponds to the most-effective science-driven user level including for example institutional users such as Copernicus, and 3) the domain (or cluster) level which is most-effective for addressing standardisation with EOSC, e-infrastructure (dominantly EOSC) or with international organisations (GEO, WMO, RDA etc.).

Task 7.1 is therefore targeting level 2 and has been completed by establishing the four working groups, one for each domain (marine, atmospheric, biodiversity/ecosystem and solid Earth) to focus on domain specific development needs and issues which need to be overcome on the way to achieve the aim of higher level interoperability and joint strategy targets. Well advanced flagship RIs with a high level of maturity will lead the development and implementation of interoperable services and provide support and guidance to the less mature and upcoming RIs. The working groups ran throughout the project duration through regular physical meetings organized during each of the ENVRI meetings (twice a year).

2. Main findings of D17.1

The main outcome of work in 17.1 is not necessarily directly reported in D17.1 as the domain meetings also were the place to discuss a number of issues for preparing new projects, reports presented in other WPs or decisions to be held at the BEERI meetings. The domain level discussions in WP17 can be summarized into 3 main issues, as follow:

Subdomain level approach towards FAIRness in RI data

Enhancing the harmonisation of RI data FAIRness within each single RI is key to future use of services and it was clear from the beginning that work towards FAIRness should partly be approached at the domain level, reinforcing the functional collaboration among the RIs, and capitalising on RI specificities and strength points. This was concretized in the follow-up project ENVRI-FAIR where domain approaches were identified as the most efficient ways to respond to the following objectives: 1) To analyse and define priorities to implement the FAIR principles into the data life cycle of the Research Infrastructures; 2) to realise case-studies showing the achieved and integrative FAIRness on interoperability and reusability applied to 'essential variables' of relevance to the domain; and 3) to share strategies towards developing proper Data Management Plans (DMPs) and their integration in the respective Information Systems in order to then guarantee the development of modelling techniques and tools through the provision of distributed e-Services associated. In ENVRI-FAIR, following discussions in WP17 implementation of tools and services is organised by the four subdomains of the ENV domain. This approach provides the appropriate granularity of the implementation process to secure the best possible provision of thematic data services and tools at RI level, at domain level and at ENV domain level under the EOSC catalogue of services.

The main observation targets of the RIs are based on the essential variables (Climate ECV and Ocean EOV) which are covered in time and space when combining the observations of these RIs. Further integration of the RIs at domain level will be developed along the lines of data workflow and management needs and by the joint services discussed within ENVRI-FAIR. In that sense, the ENVRI-FAIR Domain Work Packages will ensure maintaining the domain-level discussions from the ENVRI cluster, however with a focus on data related topics. Further integration efforts towards, e.g., joint governance structures, and site collocations etc. requires a separate platform currently in the process of being implemented.

Developing the domain landscape

One key outcome from the ENVRIplus work at subdomain level is the report on the Landscape of European ENVRIs in 3 out of the 4 subdomains (Solid Earth domain being already integrated). White papers from the landscape analysis are presented as part of WP12 activities but discussed at domain level. The current report states that the landscape appears well structured, with a clear task distribution among the RIs, which are complementing each other.

Other activities undertaken at domain level

→ *Atmospheric subdomain*

- In the Atmospheric Subdomain major work was undertaken in terms of technical innovation and training across RIs and domains, and in the field of data for science. Goals of the innovation and training efforts were the joint development, application and integration of technology between RIs as a preparation step towards the advanced availability of technical solutions for common challenges like operation of instruments under extreme conditions, for future joint services and harmonised data workflow management concepts with particular focus on data QA/QC, for data analysis concepts for long time series, and as a joint approach towards the support to space missions and assimilation systems based on RI data.
- The joint development of data management and analysis concepts was promoted by two training and knowledge-building activities, which were a workshop and conference on time series analysis in Tromso, Norway in January 2019, and a

workshop on harmonizing quality control documentation and flagging for observation data, held during the 7th ENVRI week in Riga on 6 November 2018.

- The joint development of technology across RI boundaries was tackled by two use cases in the Atmospheric Subdomain, namely the case of drones as new common observation platforms, which resulted in an Unmanned Vehicle Workshop in Italy and broad field demonstration, and the coordinated measurement of the aerosol light extinction coefficient as the main parameter used by space borne and ground-based platforms for characterizing the atmospheric load of aerosol particles, resulting in a coordinated field deployment of instruments across the RIs. The cross-domain operation of marine-atmosphere common platforms will be demonstrated as a case study on methane from seafloor to atmosphere during the upcoming Black Sea cruise. In order to enhance the observation capabilities of the Atmospheric Subdomain particularly at remote sites, energy production units used by various RIs were evaluated during intense field tests under extreme conditions at high altitude to come up with specifications matching the needs of a broad range or RI users.
- Linking ENV RIs to the European earth observation programme Copernicus was tackled during an intense workshop with RIs and Copernicus representatives, held during the 3rd ENVRI week. The development of a common strategy to produce RI standardized observations of atmospheric variables that meet the specific requirements to support the current and future Copernicus Space missions is ongoing and will provide a major outcome of the Atmospheric subdomain efforts.
- The development and implementation of software and services towards interoperable data management was in the focus of Data for Science activities. For the Atmospheric Subdomain, a coherent strategy towards DOI was developed, which reflects the different needs of the RIs with respect to e.g. granularity of information for supporting correct, quantitative credit to data providers. Key action was also taken towards (meta)data interoperability among atmospheric RI's which included an agreement to make metadata and data available among atmospheric RIs using standardized machine-to-machine interfaces, and metadata transmission (location, time, availability) to other portals for data discovery. Joint services like air-mass flagging and typing have been developed and partially implemented.

→ *Marine Subdomain*

Marine Subdomain representatives regularly met at ENVRIPLus domain level meetings. This has resulted in some important ENVRIplus deliverables.

- The marine domain is not very much integrated and shows a large diversity, but this integration work is in progress. Harmonisation and consistency between the various marine RIs strategies is important to be developed and ENVRIPLUS has been a place for developing the European Ocean Observing System (EOOS) initiative. EOOS is a community-driven coordinating framework for Europe's ocean observing capacity that helps linking the disparate components of the ocean observing system, including the different RIs contributing, and promote shared strategies, infrastructure development, data standardization, open access, and capacity building. In Task 17.1, the marine-borne in-situ ocean observations had the capacity to foster a better integration and a stronger engagement across the different ocean observing sectors.
- One of the outcomes of marine domain activity was to contribute to organizing specific tasks in the project such as 1) Task 2.3: observation continuum: enhancing use

of RIs for satellite validation from and to assimilation and services and 2) Task 4.2: Marine – atmosphere common operation of platforms: case study on methane from seafloor to atmosphere

- Discussions for defining the RI strategies in the preparation of several key EU projects were held at domain level. This is the case for BG-09-2016: An integrated Arctic observation system which is now EU INTAROS using services from several marine RIs. An identical process was put in place for BG-12-2016: Towards an integrated Mediterranean Sea Observing System and for BG-13-2016: Support to the BLUEMED Initiative: Coordination of marine and maritime research and innovation activities in the Mediterranean.
- The thematic discussions in 17.1 also dealt with addressing specific topics such as the continuum coast / offshore in RIs, biogeochemical parameters, and also included technical discussions on technology, metrology (such as CO2 ocean / atmosphere measurements, and operations (such as the WP4 measurement campaigns).
- 17.1 was also the place for preparing the ENVRI FAIR for marine domain RIs. The improvement of the FAIRness of the Marine Research Infrastructures (RIs) is critical to developing the integrated services systems required by a broad variety of research, regulatory and operational communities. Marine domain RIs were built by different specific communities, with heterogeneous maturity and capacity of data systems, as evidenced by the FAIR maturity matrix organized within 17.1. It is expected that ENVRI-FAIR will be a significant step toward European coordinated and homogeneous ocean data management.

→ *Solid Earth Domain*

The key aspect at domain level has been the work jointly between the EPOS team and the Theme 2 team of ENVRIPlus. This has resulted in some important ENVRIplus deliverables.

- The final deliverable on curation (lead, EPOS) derived from work within EPOS and commented upon by the ENVRIPlus WP8 and WP5 teams especially. It is based on the methods of DCC (Digital Curation Centre) but takes account of the EPOS team experience in curating datasets and also software, workflows, computing environments (to allow later reproducibility of results) information on ancillary entities/objects such as organisations, persons which assist in contextualisation.
- The final deliverable on cataloguing (lead IFREMER) which includes the EPOS catalog in CERIF format and a first matching and mapping between CERIF and the CKAN-based metadata format used in EUDAT. The key finding is that the rich metadata of CERIF is needed for autonomic (as opposed to human-driven) systems. Further work is planned in the last months of ENVRIPlus leading to ENVRIFAIR where the target catalog is that of eInfraCentral at the EOSC Hub.
- The final deliverable on curation (lead EAA) where EPOS contributed in two ways; first the use case from Verce where workflows using Jupyter Notebooks have associated with them provenance information collected in a slightly specialised version of the W3C PROV group of standards. In parallel it was demonstrated that CERIF can provide provenance information natively due to its linking relations with role and date/time

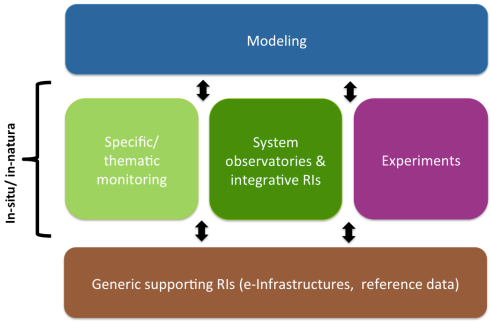
start and end attributes. Some initial work on mapping between PROV (specifically PROV-O) ad CERIF was presented; unfortunately, the work on the PROV side was being done by a colleague in Australia who had to stop the work due to serious illness. IT is expected that the work will be picked up again in ENVRIFAIR.

- EPOS contributed extensively to the work on the Reference Model in particular proposing and having accepted major revisions in the areas of curation, cataloguing and provenance. Detailed work was done relating the Reference Model to the Requirements specification from T5.1and then proposing improvements to the RM.
- In addition EPOS team members have contributed in other areas: EPOS has contributed to the strategic management of ENVRI through attendance and active participation in BEERi. EPOS has maintained a member on the editorial board of the newsletter and contributed extensively. Furthermore, EPOS team members have made presentations and posters at EGU, AGU and other for a as well as generating scholarly publications.

→ *Ecosystem Biodiversity domain*

- **Mapping the landscape**

The understanding of ecosystem functioning requires a combination of approaches and an integration of scales. Meetings within ENVRIplus explored various ways to visualise a landscape map of biodiversity & ecosystem research infrastructures taking into account these aspects. This was relevant for both internal domain considerations (understanding each one niche and cooperative linkages, and identifying any gaps), and for promoting external awareness (for ESFRI, policy and funding bodies, and the public in general). The following diagram was initially proposed for each RI to insert its position according to its methodological approaches:



This map was then used to work on complementarities and specificities of each RI. The domain discussions facilitated mutual understanding between RIs at different stages of their development.

A first step is to recognize specific leading expertise to different RIs depending on the methodological approaches:

- Observation: leading expertise with eLTER
- Experimentation and instrumentation: leading expertise with AnaEE
- Modeling: leading expertise with Lifewatch
- Social interactions: leading expertise with eLTER
- Samples repositories and characterization: DISSCO (terrestrial animals, plants), EMBRC (marine, animals, plants), MIRRI (micro-organisms), AnaEE (soils)

- **Site documentation**

Setting up joint site documentation for *in situ* research sites was needed to identify overlaps and provide a basis for across RIs cooperation as well as for the use of sites by other infrastructures and for integrating strategies for observational and experimental sites. It was recommended to build this documentation on the existing DEIMS system (http://data.lter-europe.net/deims/), which already contains a big part of European sites and can serve as a global site documentation facility.

Further integration between RIs of the EcoBio domain was particularly developed in deliverable D12.3 of ENVRIplus, which developed the concept of co-location and proposed to identify master sites.

Following this first phase of mapping, EcoBio domain meetings focused on specific issues aimed at facilitating synergies and interoperability between RIs.

- **Data and metadata standard**

ENVRIplus and its theme 2 provided a strong support to the organization of data in the Eco-Bio domain, as it did for the other domains. Significant progress was made towards the definition of metadata description for experimental and observational studies in ecosystem science.

Although the RIs can be grouped in 4 clusters (atmospheric, biodiversity/ ecosystems, marine and solid earth), biodiversity is present in all these domains. Essential Biodiversity Variables (EBVs) were promoted by the EcoBio domain as a common language to connect data on biodiversity across dimensions. Several partners of ENVRI were also partners of the Globus project. Whereas the population genetics theory provides a solid theoretical basis to the EBV categories 'Genetic composition' and 'Community composition', a unified theory is still missing for the 'Ecosystem structure' and 'Ecosystem Function' categories, which need further discussion within the scientific community in order to identify the best 'indicator' variables. Data standards and data management procedures are now being worked out in ENVRI FAIR as for the other domains.

- **Standard protocols for experimentation and measurements**

In order to benefit from the complementarity between the RIs of the domain, standard protocols for biogeochemical processes were discussed. Initial efforts were conducted under the I3 ExpEER and was further develop by the Climani network. The output from the Climani network was presented at an ENVRIweek and provided a basis for further collection of protocols within the domain as well as across domains.

- **Scientific methodological crucial issues**

Although not all the exclusivity of the Ecosystem Biodiversity domain, several issues hindering our understanding and prediction capacities have been discussed, in particular:

Tackling complexity: unravel the relations between the levels of biological organisation and functions; understand (temporal, spatial, functional) scaling effects; define sources of uncertainty in model outcomes (decision support); explicit interaction of biodiversity & ecosystems with other systems (climate, oceans, humans).

Modeling: high priority for an RI service framework allowing users to link existing and new models together, implying interoperability of workflows. Develop modeling capacities on demand – on the fly, cross cutting infrastructure domains. Define milestones for realizing priorities for data analytics and dealing with large and complex data volumes. Liaise with ISBE.

Virtual environments and web services: users need RI services to work together and benefit from finding, processing and modeling data in virtual environments of their choice. Such services to reach out to workflows, tools and methods. There is a big gap between what was done so far and what is needed. Virtual environments have to be much larger with capabilities, being robust and user-friendly.

- **Outcome of the community building discussions**

As a result of sustained exchanges between RIs, mutual understanding has increased and several RIs have elaborated projects submitted to the last INFRAIA (eLTER) or INFRADEV (AnaEE, Lifewatch) calls, mentioning ENVRI cluster as the best place to trigger collaborations between RIs. In addition, eLTER submitted its Preparatory Phase project.

3. Conclusions

The subdomains are vital to ENVRI. They provide (a) requirements; (b) ideas; (c) skilled research scientists with understanding of both data and methods; (d) skilled IT professionals with not only general IT expertise but also expertise in applying IT to the domain.

Because the individual subdomains have their own ways of doing science and of using IT, Task 17.1 was key to maintain the necessary level of discussion between RIs within single domains. The two key outcomes were 1) the success of the ENVRI FAIR project and in particular the organisation of 4 WPs in the project, dedicated to develop FAIRness in domain specific RIs and 2) the landscape reports of European ENVRIs, delivered in WP12 but discussed in WP17. By today, the current report states that the landscape appears well structured, with a clear task distribution among the RIs, which are complementing each other. In the course of ENVRI-FAIR these papers will continuously developed to keep pace with the ongoing integration activities within ENVRI-FAIR.

Task 17.1 was also fundamental to prepare for the cross-domain discussion in BEERI. There are cross-domain issues (links with Copernicus for instance) where strategies are most efficiently implemented at domain level first before being addressed as cross-domain activities. This was the basis for the work proposed in the NEURONE project, which the ENVRI

domain response to call LC-SPACE-05-2019 of the European commission. The overarching objective of the proposed NEURONE project is to build upon EEA efforts to compile a comprehensive overview of the current state of play on in situ contributions to Copernicus, with a view on identifying and analysing the data-, governance-, technological- and sustainability- related challenges and gaps of the in situ infrastructures of Copernicus interest. This information will be utilized for constructing a roadmap for the evolution of the "Copernicus in situ Interface Layer", a system to provide homogeneous and interoperable in situ observations for the validation and quality enhancement of Copernicus data and information in a cost-effective way.

All RIs agree that maintaining subdomain level discussions is key to future developments of RIs, and that it would definitely be an added value to the ENVRI community. The subdomains provide a solid basis for sustainability since they are (progressively) becoming ERICs with state funding and a well-defined governance. Subdomain discussions in the future will likely concern RI sustainability. ENVRI-FAIR provides a mechanism for the ENVRI community to be sustained for 4 more years with a focus on using EOSC as a hub. However, longer-term a mechanism needs to be found to encourage and fund the interoperability between the ENVRI domains each of which is sustained for its own purposed by mechanisms such as ERICs.