

Common Operations of Environmental Research Infrastructures: the ENVRI project

Yannick Legré⁽¹⁾, Jacco Koning⁽²⁾, Alex Hardisty⁽³⁾, Roberto Cossu⁽⁴⁾, Sanna Sorvari⁽⁵⁾ and Wouter Los⁽²⁾ on behalf of the ENVRI consortium

(1) *yannick.legre@idgrilles.fr, Institut des Grilles et du Cloud - CNRS*

(2) *University van Amsterdam, Netherlands (the)*

(3) *Cardiff University, United Kingdom*

(4) *European Space Agency*

(5) *Finnish Meteorological Institute, Finland*

Overview: (police Arial Narrow 12 points Gras)

Frontier environmental research increasingly depends on a wide range of data and advanced capabilities to process and analyse them. The ENVRI project, "Common Operations of Environmental Research infrastructures" is a collaboration in the ESFRI Environment Cluster, with support from ICT experts, to develop common e-science components and services for their facilities. The results will speed up the construction of these infrastructures and will allow scientists to use the data and software from each facility to enable multi-disciplinary science.

The target is on developing common capabilities including software and services of the environmental and e-infrastructure communities. While the ENVRI infrastructures are very diverse, they face common challenges including data capture from distributed sensors, metadata standardisation, management of high volume data, workflow execution and data visualisation. The common standards, deployable services and tools developed will be adopted by each infrastructure as it progresses through its construction phase.

Two use cases, led by the most mature infrastructures, will focus the development work on separate requirements and solutions for data pre-processing of primary data and post-processing toward publishing.

The project will be based on a common reference model created by capturing the semantic resources of each ESFRI-ENV infrastructure. This model and the development driven by the testbed deployments result in ready-to-use systems which can be integrated into the environmental research infrastructures.

The project puts emphasis on synergy between advanced developments, not only among the infrastructure facilities, but also with ICT providers and related e-science initiatives. These links will facilitate system deployment and the training of future researchers, and ensure that the inter-disciplinary capabilities established here remain sustainable beyond the lifetime of the project.

Enjeux scientifiques, besoin en calcul, stockage et visualisation : (police Arial Narrow 12 points Gras)

The ESFRI-ENV projects are not starting from nothing. They all build upon many years (even decades) of existing practice. All of them are currently working with some form of data infrastructures and contributing data and knowledge to international observation systems and for broad user communities. The ENVRI project is aiming to minimise the heterogeneity among environmental data infrastructures to offer common best practices and to provide technical tools to help them to better fulfil their scientific remit. But environmental research is inherently multidisciplinary and scientific data are of many different kinds.

Scientific questions drive a wide variety of experimental and data-driven methodologies. Data are generated by a huge ensemble of diverse instrumentation and observations.

Here follows some examples of different data kinds and providing facilities.

- **EISCAT:** Reflected (interference) radar signals from solar troposphere disturbances
- **EPOS:** Seismic data; satellite data on volcanic activities; ocean circulation data
- **EMSO:** Marine sensor data about salinity, temperature, Ph, fluxes, etc. Circulation data.
- **ICOS:** CO₂ and other gas (CH₄, NO_x, etc) concentrations in the air, measured at various places and heights, including fluxes through time
- **LifeWatch:** Species level data (species names, distributions, locality data on individual organisms, attribute data as parasite (host), ecology, life history, etc). Mainly provided by GBIF. Ecological data (habitats, species composition per area, canopy density, trunk and other biomass, phenology (= changes through time), etc). Mainly provided by monitoring in Long-Term Ecological Research (LTER) sites. Genetic data (DNA sequences, genomes) from EBI-EMBL. Sensor data (Species presence or abundance). Remote sensing data (canopy cover, etc)
- **SIOS:** All kinds of above mentioned environmental data, but restricted to the island and surrounding seas of Svalbard.

Today, it is well understood that precise, long-term observations are essential to quantify the patterns and trends of on-going environmental changes, and that continuously evolving models are needed to integrate our fundamental knowledge of processes with the geospatial and temporal information delivered by various monitoring activities. These increasing requirements for observation and modelling mean that the volume and transfer rates of environmental data are massively increasing.

This makes it critically important that the environmental science community should put a strong emphasis on analysing the best practices and adopting common solutions on the management of heterogeneous data and data flows, both among ESFRI ENV projects and beyond. Frontier research and societal challenges call for the effective exploitation of the full range of global environmental data and the wide distribution of such data to multiple user communities. Such a research landscape presents important challenges, for example in data management, data access and data preservation. Current developments in ICT capabilities, especially in eFP7-infrastructures, not only provide a potential means of responding to these challenges, but also offer an excellent opportunity for the development of the environmental infrastructures

Développements, utilisation des infrastructures : (police Arial Narrow 12 points Gras)

The ENVRI European environment cluster framework for IT infrastructure is aiming to minimise divergence from what is already developed, while at the same time seeking to bring about long-term convergence and interoperability. It is not the plan to deploy a new single infrastructure for multi-disciplinary collaboration.

ENVRI contributes to the construction of the ESFRI-ENV research infrastructures with a flexible framework, facilitating their current and future interoperability. The ENVRI project delivers results for the ESFRI-ENV infrastructures as users. As such, all the ESFRI-ENV infrastructures are members of a Stakeholders Advisory Board in which the overall discussions on future strategies and project development will take place.

Apart from these strategy considerations, the overall strategy is focussing on:

- Contributions to the architecture of decentralised infrastructures in the ENVRI cluster
- Standards, harmonisation and interoperability
- Metadata frameworks and other ICT related issues
- Support for access and deposit policies
- Ready-to-uptake and deploy services, systems, standards etc. by the cooperating infrastructures
- Support for a consistent European research infrastructure ecosystem
- Strong interaction between ESFRI infrastructures and e-infrastructures

Outils, le cas échéant complémentarité des ressources, difficultés rencontrées : (police Arial Narrow 12 Gras)

Scientific communities have already developed and are rapidly developing their own IT tools; often in global collaborations. Other ESFRI projects are building on existing tools and infrastructure from many European countries, as well as non-European countries. They all made considerable progress and it is now the right time to agree on a common reference model for joint operations. The Reference Model for Open Distributed Processing or RM-ODP1 (ISO/IEC 10746-1:1998) is a framework that is being increasingly widely used in the environmental e-infrastructure community. For example: the European INSPIRE Directive (2007/2/EC) on an environmental data sharing infrastructure, the ISO Technical Committee 211 on standards for geographic information and geomatics, the Open Geospatial Consortium (OGC) standards body, and large infrastructure projects like ORCHESTRA, SANY-IP, GENESIS, GENESI-DR, GENESI-DEC and LifeWatch all base their work on it. RM-ODP factors an architecture into five mutually consistent viewpoints: *Enterprise, Information, Computation, Engineering and Technology*.

Résultats scientifiques : (police Arial Narrow 12 points Gras)

No result yet. The project just started last november.

Perspectives : (police Arial Narrow 12 points Gras)

Environmental science is placing an increasing emphasis on the understanding of the Earth as a single, highly complex, coupled system. It is well accepted, for example, that the feedbacks involving oceanic and atmospheric processes can have major consequences for the long-term development of the climate system, which in turn affects biodiversity and can control the development of the cryosphere and lithosphere. Our future researchers must be empowered to address scientific questions that nowadays cannot, or not sufficiently be addressed. The overarching impact of ENVRI is that the collaborating environmental infrastructures may provide these much needed new opportunities.

Improved access and service for users; strengthening the European Research Area and international cooperation

Despite the diversity of environmental science, many of the ENV-ESFRI projects share the same significant challenges. These include the collection of data from multiple distributed sensors (potentially in very remote locations), the management of large low-level data sets, the requirement for metadata fully specifying how, when and where the data were collected, and the post-processing of those low-level data into higher-level data products which need to be presented to scientific users in a concise and intuitive form. Through the constituent projects of the European E-infrastructure Forum, the European Commission is already putting in place an underlying framework to support the requirements of the European Research Area in areas such as networking, high-performance computing, grids/workflow, and (in the near future) data storage and management. The EEF cooperation constitutes a major resource of expertise and capabilities on which European projects

can draw, and the engagement of the EEF with the ENVRI project is essential in order not to avoid duplication of the intensive efforts which have already been made to make these underpinning technologies available.

Interoperable research infrastructures

The development of deployable components for shared e-infrastructure operations can be implemented by any new research infrastructure in the area of environmental science. The deployable components are providing a selection of standardised solutions to key problems common to all environmental facilities, resulting in substantial savings of time and effort to new ESFRI environmental research infrastructures seeking to support their project through the exploitation of the leading capabilities in European e-science. Infrastructures adopting these shared components will be uniquely well-placed to inter-operate with the data from other environmental facilities and potentially also with those in quite different fields, making it much easier to carry out interdisciplinary studies and to apply the “whole system” approach which is an emerging paradigm in the geosciences.

Cost efficiency

The ethos of the ENVRI project is not to build an entirely new e-infrastructure system from first principles. It is planned to build from the existing portfolio of tools, services and data products already available within the infrastructure communities, as well as those offered and innovated by our ICT specialist partners. Defining the common requirements, and developing new common capabilities by incorporating existing capabilities, provides the best chance to achieve a sustainable system for future use, since it will mean that the various infrastructures will each have some familiarity with aspects of the final deployable systems. To this extent, the ENVRI project provides a good demonstration of harmonisation, cost efficiency and the promotion of existing synergies between projects.

Reference Model to support innovation in interdisciplinary research

The development of common elements for a reference model, which all of the facilities can adopt to support their data collection, storage and processing needs, is contributing to a basic level of compliance with the other infrastructures in the cluster. An important result is the combination of the semantic resources of each of the various infrastructures into a single conceptual model recognisable to any scientist working within the environmental domain.

Scope for wider use and technology transfer

The construction of a fully inter-operable e-infrastructure for environmental science is a journey which each infrastructure has to pursue at its own pace. The ENVRI contributions do not provide a “one size fits all” system for each environmental infrastructure. The ENVRI developments have an impact on balancing contributions from across Europe. The ENVRI consortium includes entities from various European countries, and there is clear scope for this number to increase as users from other European nations engage with the ENVRI infrastructures, while many of the ICT specialists are multi-national collaborations combining contributions from several European nations.

The ENVRI project is relevant to a number of other national and international e-infrastructure initiatives. The related US communities are also actively discussing its exploitation of “cyberinfrastructure” and various contacts are fostered for international cooperation.

Références : (police Arial Narrow 12 points Gras)

- <http://envri.eu>