ENVRI, Integrated infrastructures, Environmental research in harmony

Yannick Legre

HAL Id: in2p3-00798132
http://hal.in2p3.fr/in2p3-00798132

Submitted on 8 Mar 2013

Integrated
infrastructures

Dr Sanna Sorvari, a key collaborator in the ENVRI project, explains how the venture will contribute to the construction of environmental research infrastructures by sharing experiences and technologies and solving crucial common challenges and e-science issues in an integrated manner

Firstly, could you describe what led to the creation of the Common Operations of Environmental Research infrastructures (ENVRI) project and outline its central aims and objectives?

Environmental research is inherently multidisciplinary and the scientific data produced is multifarious in nature. Scientific questions drive a wide variety of experimental and data-driven methodologies. Data is generated by a huge ensemble of diverse instrumentation and observations. Today, it is well understood that precise, long-term observations are essential to quantify the patterns and trends of ongoing environmental changes. 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 made us realise that it is critically important that the environmental science community places a strong emphasis on analysing the best practices and adopting common solutions on the management of heterogeneous data and data flows, both among European Strategy Forum on Research Infrastructures (ESFRI) ENV projects and beyond.

Can you outline the requirements needed to develop more advanced analytical and modelling techniques and software?

With several different infrastructures and project participants from different backgrounds, reference models ensure a common vision and understanding, as well as meaningful cooperation and sharing of resources. They provide a basis for easier communications between partners and for designing ICT architectures that support interoperability between infrastructure components and between infrastructures themselves.

Since the project's inception, what common needs and issues have required attention within the environmental research infrastructure projects?

Most attention has been addressed to issues and challenges, such as data streaming and storage management; data discovery and access to distributed data archives; linked computational, network and storage infrastructure; data curation; data integration, harmonisation and publication; data mining and visualisation; and scientific workflows.

What activities are currently underway in pursuit of developing a European environment cluster framework for an IT infrastructure that enhances interoperability and collaboration?

The technical work within ENVRI is now focusing on: analysing and categorising ENVRI projects; the development of a common reference model; file discovery and interoperability (OpenSearch); data discovery and interoperability (Linked Data); and data integration and publication (visualisation).

ENVRI is not attempting to build a new e-infrastructure system, but rather to build on existing information and tools. Could you highlight the main areas of development?

Data management, data discovery, cross-infrastructure data integration and service federation are all within the scope of ENVRI. All four domains are of interest to

the reference model. New tools are being developed specifically for the data discovery and data integration domains. These will assist ESFRI infrastructures that adopt them to better enable the integration and harmonisation of resources across the cluster's infrastructures, with emphasis on digital repositories and interoperating infrastructures; linked computational, network and storage infrastructure; semantic interoperability of data and resources; and data integration, processing and publication facilities.

Can you describe the main features of the common reference model, designed to capture the semantic resources of each ESFRI-ENV infrastructure?

The reference model is based on the design experiences of the state-of-the-art environmental research infrastructures, with a view to informing future implementation. It tackles multiple challenging issues, such as integrated data access and discovery, real-time data handling and service federation, which are prioritised for many existing initiatives. It uses a standard language - Open Distributed Processing (ODP) – to describe the model, which provides a uniform platform for communication and discussion, and ensures unifying understanding.

What ultimate impact do you hope this work will have on the environmental research landscape?

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 in data management, data access and data preservation. The central goal of the ENVRI project is to provide software tools to harmonise, integrate and analyse the heterogeneous data across domains and research infrastructures by promoting accessibility and preserving specificity. By using various data from different research infrastructures, new scientific breakthroughs can be achieved and environmental challenges tackled.

Environmental research
in harmony

As a pan-European collaboration, the **ENVRI** project aims to provide infrastructures to facilitate the sharing and analysis of multidisciplinary environmental science data in order to harmonise and promote greater access to environmental research

THE EARTH IS highly complex, with an atmosphere, oceans, and geological and biological systems that all interlink in a complex web of feedback mechanisms. For example, the climate system is directly influenced by oceanic and atmospheric processes. Adverse effects on this system can affect biodiversity and influence the development of the cryosphere and lithosphere.

Attempting to make sense of this complexity, environmental science encompasses a number of different disciplines that enable us to study the Earth. However, these disciplines are often studied independently and it is difficult to link data and to view the Earth systems as a whole. In recent years, the environmental science community has placed increased emphasis on combining data from all aspects of study in order to gain better understanding of the Earth as an intricate, yet single system.

Environmental data is collected by a wide range of instruments and observations in many different environments worldwide. International Research Infrastructures are being set up to collect and handle specific data forms. For example, European Incoherent SCATter (EISCAT) focuses on monitoring reflected (interference) radar signals from solar troposphere disturbances, while the European Multidisciplinary Seafloor Observatory (EMSO) collects and analyses

© EISCAT_3D

marine sensor data about salinity, temperature, Ph, fluxes, etc., along with circulation data.

Although these infrastructures focus on specific areas of study, it is necessary to combine data in order to study the Earth's systems and produce accurate models or simulations of the processes underway. The environment is constantly evolving and in order to quantify the patterns and trends that lead to this evolution, models are required that combine our knowledge of the processes occurring with geospatial and temporal information from environmental monitoring apparatus. In order to create an integrated environmental e-infrastructure that enables such models to be produced, the incorporation of sufficient data from instruments, sensors or observations, along with reliable databases, advanced analytical software and sufficient computational capacity is required.

A COMMON GOAL

The project Common Operations of Environmental Research Infrastructures (ENVRI) is a collaboration conducted within the European Strategy Forum on Research Infrastructures (ESFRI) Environmental Cluster*. This 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. The ethos of the ENVRI project is not to build an entirely new e-infrastructure system. Rather, it plans to use the existing portfolio of tools, services and data products already available within the infrastructure communities, as well as those offered and innovated by ICT specialist partners.

Data comes from a wide variety of sources or sensors in potentially remote locations. In addition, metadata must specify how, when and where data was collected and additionally low-level data must be managed and processed giving higher-level data products that can be more easily utilised by scientists. "Such a research landscape presents important challenges, for example in data management, data access and data preservation," reveals Dr Sanna Sorvari, Research Manager at the Finnish Meteorological Institute who

is involved in the ENVRI collaboration. "Current developments in ICT capabilities, especially in e-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."

ENVRI is not starting entirely from scratch. Over many years, data infrastructures have been built up, so ENVRI looks to minimise heterogeneity among these infrastructures in order that scientists follow similar approaches and use the same technical tools to produce reliable and comparative outcomes. In the long term, the aim is to see convergence and interoperability of the environmental e-infrastructures.

REFERENCE MODEL

The primary objective of ENVRI at present is to agree on a reference model for joint operations. This will enable greater understanding and cooperation between users: "Fundamentally, the model serves to provide a universal reference framework for discussing many common technical challenges facing the ESFRI-ENV infrastructures," Sorvari explains. It achieves this by virtue of the components that are modelled and how they are named and characterised. "By drawing analogies between the reference components of the model and the actual elements of the infrastructures (or their designs) as they exist now, we can identify gaps and points of overlap, as well as being able to identify and justify possible common services which could be provided by other projects and infrastructures," Sorvari states.

The reference model is to be based on the Reference Model for Open Distributed Processing (RM-ODP), which has

gained increased use within environmental e-infrastructures. Its architecture is based on the premise that a distributed system can be considered from different independent viewpoints that represent different abstractions of the original system, in this case: enterprise, information, computation, engineering and technology.

CHALLENGES

The research Infrastructures involved in the ENVRI project are in different stages of their development which brings certain challenges. Therefore a collaborative workspace system has been developed to support the exchange of information, results and deliverables and to facilitate easy and timely inter-project communication. Although much of the communication is carried out via email, teleconferences and intranet, there are also regular ENVRI meetings and a policy body of ESFRI Environmental Research Infrastructures – the Stakeholders Advisory Board (SAB) – has been established which ensures that an active dialogue with ESFRI is maintained. The SAB consists of coordinators of current ENVRI projects along with those from other ESFRI environmental projects, thus ensuring those involved are aware of the activities of ENVRI within other environmental ESFRI activities.

Currently, facilities tend to address their e-science requirements individually, yet with the successful implementation of the ENVRI project they would work collectively and greatly advance current ESFRI environmental infrastructures. In order to facilitate this, the main challenges faced by ENVRI in the near future include: federation over existing (national and international) infrastructures/services, (near) real-time data handling and integrated data discovery across various centres/catalogues. Other challenges include quality control of data, and

metadata definition and assignment.

The ENVRI project is still in its initial stages, but its aim of enabling the convergence and interoperability of European environmental Research Infrastructures is regarded as highly achievable by Sorvari and her colleagues: "All of the ESFRI ENV infrastructures and their various user communities will benefit from the common solutions and better interoperability brought by the ENVRI project," enthuses Sorvari. The future infrastructure established by the ENVRI project is likely to facilitate environmental scientists' investigations, enabling them to respond to scientific questions that cannot currently be addressed. This will largely be due to the methods of project coordination and good communication channels that have already been established. In the US, similar approaches are being adopted, which could potentially lead in the distant future to global cooperation and sharing of data and approaches; a future in which, through homogenised scientific data, we may finally be able to view the Earth as a single homogeneous system.

*ENVRI PROJECT PARTNERS OF ENVIRONMENTAL ESFRI RIS:

EISCAT-3D

Integrated Carbon Observation System (ICOS)

European Plate Observing System (EPOS)

European Multidisciplinary Seafloor Observatory (EMSO)

EURO-ARGO

LifeWatch

INTELLIGENCE

ENVRI

COMMON OPERATIONS OF ENVIRONMENTAL RESEARCH INFRASTRUCTURE

OBJECTIVES

ENVRI 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 multidisciplinary science.

PARTNERS

Istituto Nazionale di Geofisica e Vulcanologia, Italy • **The University of Edinburgh**, UK • **Helsingin Yliopisto**, Finland • **Umweltbundesamt Gmbh**, Austria • **Cardiff University**, Wales • **Csc-Tieteen Tietotekniikan Keskus Oy**, Finland • **Commissariat a l'Energie Atomique et aux Energies Alternatives**, France • **European Space Agency**, France • **Institut Francais de Recherche pour l'exploitation de la Mer**, France • **Koninklijk Nederlands Meteorologisch Instituut-Knmi**, The Netherlands • **Centre National de la Recherche Scientifique**, France • **Stichting European Grid Initiative**, The Netherlands • **Universitaet Bremen**, Sweden • **Eiscat Scientific Association**, Sweden • **Consiglio Nazionale Delle Ricerche**, Italy

FUNDING

EU Seventh Framework Programme (FP7) – contract no. 283465

CONTACT

Yannick Legre
Research Manager

CNRS
Institut des Grilles et du Cloud (IdGC)
c/o Laboratoire de Physique Corpusculaire
Campus des Cezeaux
24 avenue des Landais
BP 80026
63171 Aubiere CEDEX
France

T +33 473 405 373
E yannick.legre@idgrilles.fr

www.envri.eu

DR SANNA SORVARI is a Research Manager at the Finnish Meteorological Institute. Sorvari has over 15 years' experience in climate research. Currently working with ERA initiatives, such as environmental research infrastructures (ICOS, ACTRIS, ENVRI, COOPEUS) and Joint Programming Initiatives (JPI Climate). He is a member of many committees and belongs to the secretary of Climate Panel to support Finnish Government.