

# Science Demonstrators

By Science Demonstrator we mean *"a showcase of a service solution illustrated through a prototype implementation, which serves as proof or evidence that the Theme2 services can bring added value for supporting ENVRlplus community to deliver scientific research"*. Nine well-developed use cases have been selected as the final Science Demonstrators.

NEWS: A questionnaire-based tool for user evaluation of the Science Demonstrators is now available at <https://survey2.icos-cp.eu/ENVRlplus-evaluator> - please participate and use it to give the developers valuable feedback!

The tool is the final deliverable (D9.4) of ENVRlplus Work Package 9. It was primarily developed to provide a concrete example of how the dialogue between the developers of these services and their user communities can be facilitated and streamlined by allowing the collection of evaluation information and other feedbacks with the help of a structured on-line questionnaire. The evaluation criteria covered by the survey include Accessibility or Ease of use, Maturity, Sustainability, Visibility, Impact, Portability, Scope, Potential, Domain coverage, and Explicit ENVRlplus contribution.

The complete survey consists of 70 questions arranged in 15 groups. Depending on which service is evaluated, the respondent is presented with ca 35 questions, which on average should take 15-20 minutes to complete. The publicly available survey response summary is available at <https://survey2.icos-cp.eu/ENVRlplus-evaluator-summary>; this displays basic statistics for a selected subset of the survey questions.

## Science Demonstrator 1: Support EISCAT\_3D Users to Reprocess Data Using User's Algorithm (Use Case IC\_3)

SD1 addresses a requirement of the EISCAT RI community, namely to allow individual scientists to process their experimental data using their own algorithms. The challenge is common to many ENVRlplus RIs, where data is often processed using standard models and methods. As researchers want to use different analysis models, easily modify parameters or algorithms, and collaborate with each other, they need a Virtual Research Environment (VRE). This demo showcases a model making use of the D4Science gCube platform developed by T7.1, which enables scientific researchers to re-process data by implementing and adapting algorithms and parameters from other sources.

## Science Demonstrator 2: The Eddy Covariance Fluxes of GHGs (Use Case IC\_13)

SD2 showcases a novel implementation of a computationally efficient tool for processing of Eddy Covariance (EC) data which offers to users the possibility to calculate EC fluxes through the EddyPro® software (LI-COR Biosciences, 2017; Fratini and Mauder, 2014) according to 4 processing schemes resulting from a different combination of existing methods. To reduce the computational runtime required, the 4 processing schemes were implemented and executed in parallel mode. The whole service setup including a metadata management algorithm, was implemented and tested in the D4Science gCube Virtual Research Environment provided by Task 7.1, and the final computational runtime for Near Real Time (NRT) processing (i.e. flux estimates based on raw data collected the previous day) is of about 4 minutes, similar to those required for a standard run involving only a single processing scheme.

## Science Demonstrator 3: SOS & SSN Ontology Based Data Acquisition & Near Real Time Quality Control (Use Case IC\_14)

SD3 addresses a common problem for ENVRlplus RIs (specifically observatories that build on environmental sensor networks) that data acquisition service, in particular, the preparation of data transfer prior to data transmission are often not yet sufficiently standardized. This hinders the operation of efficient cross-RI data processing routines, e.g., for data quality checking. The demonstrator showcases a service prototype that allows submitting and publishing raw observational (non-geophysical) environmental time series data in common standard formats (T-SOS XML and SSNO JSON). A messaging API (EGI ARGO) is used to perform Near Real Time (NRT) quality control procedures by an Apache Storm NRT QC Topology, which publishes the quality controlled and labelled data via a messaging output queue.

## Science Demonstrator 4: EurpArgo Data Subscription Service (Use Case TC\_2)

SD4 describes the EuroArgo Data Subscription Service (DSS) that allows researchers to subscribe to customized views of Argo data, selecting specific regions and time-spans, and choose the frequency of updates. Tailored updates are then provided on schedule to researchers' private storage. The demo showcases an integration solution that combines the EuroArgo community data portal with e-Infrastructure services (EUDAT B2SAFE, EGI FedCloud, etc.), and uses the DRIP service developed by T7.2 for optimised service deployment. The pilot activity was initiated by the marine research community, however, the possibility to receive regular transmissions of data, especially in near-real time, directly from the organisation responsible for data collection and (pre-)processing, is very important to many large initiatives. ENVRlplus RIs can benefit from the subscription services, e.g., to create more elaborated data products by requesting data from other sources, and can optimise their internal workflows by signing up for automatic updates.

## Science Demonstrator 5: Sensor Registry (Use Case TC\_4)

SD5 showcases a "sensor registry" that aims at supporting the management of sensors deployed for in-situ measurements. Common sensors or families of sensors are used across different research infrastructures, for example, oxygen optodes that are equipped on platforms in multiple research infrastructures. The goal of this work is to define common methods to access the sensor metadata in such cases. The sensor registry applies the design principle of data catalogue developed in WP8, and uses data technologies and standards from the OGC Sensor Web Enablement family including SensorML, Observations and Measurements (O&M), and Sensor Observation Service (SOS). It brings together a marine domain implementation of these standards (the Marine SWE profile) developed by several European projects demonstrating the viability for future sensor and observation activities. The service can be integrated to various types of platforms, deep-sea observatories (e.g., EMSO), marine gliders (e.g., EuroGOOS) as well as solid earth (e.g., EPOS) or atmosphere observations (e.g., ICOS). It can also be used to track usage of specific sensor models (e.g., CO2) across the RI's observation networks.

## Science Demonstrator 6: New Particle Formation Event Analysis on Interoperable Infrastructure (Use Case TC\_17)

SD6 describes a service prototype that supports aerosol scientists in studying new atmospheric particle formation events by moving data analysis from local computing environments to interoperable infrastructures, thus harmonizing data analysis itself and more importantly the syntax and semantics of data derived from analysis. As researchers interpret primary data and thus gain information and transfer information into knowledge, we are studying and advancing in particular some technical aspects of a knowledge infrastructure i.e., a robust network of scientists, artefacts such as virtual research environments and research data, and institutions such as research infrastructures and e-Infrastructures that acquire, maintain and share scientific knowledge about the natural world. The science demonstrator showcases a possible architecture of a socio-technical infrastructure that "transforms data into knowledge." The proposed approach highlights a range of novel possibilities, in particular enabling researchers to focus on data analysis and interpretation while leaving data access and transformation from and to systems to interoperable infrastructure. It significantly contributes to implementing the global agenda of FAIR data by promoting the notion of "FAIR by Design", weaving data FAIRness into the fabric of infrastructures. It builds on the principle not to leave making data FAIR to researchers but to guarantee it by design of well-engineered infrastructures. The demonstrator is first and foremost of primary interest to a specific scientific community, namely the various aerosol research groups that study new particle formation events.

## Science Demonstrator 7: gCube-based VRE for Mosquito Diseases Study (Use Case SC\_2)

SD7 illustrates how a LifeWatch researcher can easily upload and integrate an analysis algorithm in D4Science, and share it with other researchers in a VRE. The use case proposed an integration solution that links the D4Science/gCube VRE to the LifeWatch RI and to the EGI e-Infrastructure. This integration, for example, enables individual researchers to repeat and reuse algorithms at will, run trend analysis, and add new parameters and custom data. The VRE provides provenance registration that improves reproducibility and also allows retention of computation results in the user's workspace. This facilitates editing and adaptation of algorithms, features that are not provided by the existing LifeWatch ICT.

#### **Science Demonstrator 8: Dynamic Ecological Information Management System - Site and Dataset Registry (DEIMS\_SDR)**

SD8 addresses a common problem identifying in-situ observation sites across RIs. It can be challenging to identify synergies and potential collaboration between different in-situ environmental research infrastructures for cross-RI activities due to a lack of a common site registry. DEIMS-SDR ([deims.org](http://deims.org)) not only acts as a RI-independent site registry, but also issues unique, persistent and resolvable identifiers that can be used to easily detect co-location and overlaps in different RIs. Site information can be extracted using standardised OGC services like WFS or CSW and therefore addresses the requirements for curation and cataloguing defined in use case IC8. DEIMS-SDR is also used in task 12.3.3 for the supply of site information with the aim to come up with a proof of concept for a federated site concept.

#### **Science Demonstrator 9: PROV-Template Registry and Expansion Services (Use Case IC\_10)**

SD9 shows a pilot implementation about a ENVRplus PROV-Template registry and extension service. PROV-Template is a proposed standard for converting existing process output such as log files into representations following the PROV Data Model (PROV-DM) specification for describing provenance of electronic resources in machine readable, structured form. Besides the potential advantage that existing process implementations can be enabled to generate PROV-DM conforming data without the need to change their underlying codebase, the general notion of using templates for describing provenance traces for recurring workflows can be used to foster interoperability and best practices for provenance data generation across individual communities. Following this motivation, the ENVRplus PROV-Template registry and expansion service prototype has been designed as public platform for describing, storing and sharing PROV-Templates across members of different RI, including a dedicated Web API for instantiating stored templates with individual data.