ENVRIplus Science Demonstrators

Yin Chen holds a PhD from the University of Edinburgh School of Informatics. Currently, Yin works for the EGI foundation that leads ENVRIplus WP9. Yin also contributed to the previous ENVRI project (2011-2014), where she led the development of the ENVRI Reference Model.

The recent ENVRIplus deliverable D9.2 documents seven Science Demonstrators that showcase the prototype implementations of the Theme 2 service solutions, and confirms the added value that these services potentially bring to the ENVRIplus community and research infrastructures (RIs) in general.

The first Science Demonstrator addresses a requirement of the EISCAT RI community, to allow individual scientists to process their experimental data using their own algorithms. The challenge is common to many ENVRIplus 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).

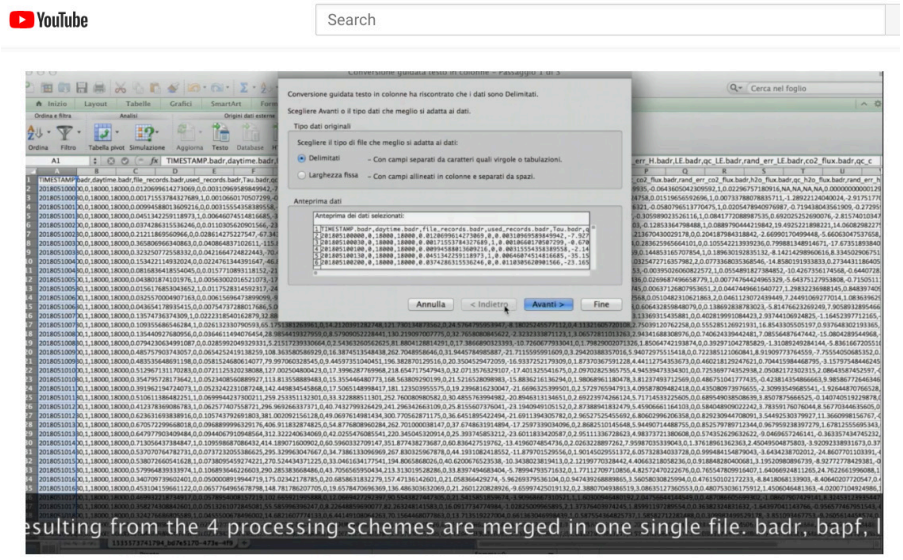
This demo showcases a model making use of the D4Science gCube platform developed by WP7, which enables scientific researchers to re-process data by implementing and adapting algorithms and parameters from other sources. Watch the demo: https://youtu.be/YEEMUvnSHUM

The second Demonstrator showcases a novel implementation of a computationally efficient tool for processing of Eddy Covariance (EC) data that allows users to calculate EC fluxes using the EddyPro® software. These calculations normally involve a complex set of data processing steps that require a considerable amount of computational resources, which can be a constraint for RIs (e.g. ICOS) that aim to simultaneously process raw data sampled at multiple sites in Near Real Time (NRT) mode (i.e. provide each day fluxes estimates relative to the previous day). To reduce the computational runtime required, four 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, achieving a final computational runtime for Near Real Time processing of around 4 minutes. Watch the demo: https://youtu.be/hod2WksKzV8

The third Science Demonstrator addresses a common problem for ENVRI RIs that deliver data services, in particular, the preparation of data prior to transmission are often not 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 supports the submission and publishing of 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 quality control procedures through an Apache Storm NRT QC Topology, which publishes the quality controlled and labelled data

via a messaging output queue. Watch the demo: https://youtu.be/p3UQZkR-RWlw

Science Demonstrator 4 describes the EuroArgo Data Subscription Service (DSS)



ENVRIplus IC 13 ICOS Eddy Covariance Raw Data Processing

that allows researchers to subscribe to customized views of Argo data. Subscribers can select specific regions and timespans, and choose the frequency of these tailored updates which are delivered directly to their 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 WP7 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. Watch the demo: https://www.youtube.com/watch?v=PKU_JcmSskw

Science Demonstrator 5 illustrates a "sensor registry" that aims to support 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 deployed on platforms by multiple research infrastructures. The sensor registry uses data technologies and standards from the OGC Sensor Web Enablement family including SensorML, Observations and Measurements (O&M), and Sensor Observation Service (SOS). 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. Watch the demo: https://youtu.be/4QxTZ2iiznk

Science Demonstrator 6 describes a service prototype that supports aerosol scientists studying new atmospheric particle formation events by moving the data analysis from their local computing environments to interoperable infrastructures. This harmonization of the data analysis and, more importantly the syntax and semantics of the derived data, also illustrates the concept of implementing the global agenda of FAIR data and promoting the notion of "FAIR by Design" by weaving data FAIRness into the fabric of infrastructures. It builds on the principle of not leaving it to researchers to make their data FAIR, but instead guaranteeing it through design of well-engineered infrastructures. Watch the demo: https://youtu.be/ra9W7b5DbgI

Finally, Science Demonstrator 7 illustrates how a LifeWatch researcher can easily upload and integrate an analysis algorithm in D4Science, and share it with other researchers in a VRE. Watch the demo at: https://youtu.be/IBJkSys5tVo