Interoperable ENVRI data management services: RM guided design methodology, service portfolio and demonstrators

Zhiming Zhao on behalf of the technical team of the ENVRIplus "*data for science*" theme
University of Amsterdam, the Netherlands
 Z.Zhao@uva.nl

In the ENVRIᴾᴸᵁˢ project, the "data for science" theme (theme2) has established a reference model guided ICT approach for handling the lifecycle of managing scientific data in research infrastructures. This approach inspired interoperability and reusable solutions that can benefit research infrastructures (RI) and facilitate interdisciplinary research.

The technical team of theme2 starts from a set of common problems identified from the ENVRI RI cluster, including 1) how to identify and cite data from different sites or infrastructures; 2) how to control the quality of nearly real-time data from sensors and annotate them; 3) how to catalog the data and to allow users to search and access data from different sites or infrastructures; 4) how to support scientists to perform experiments using data, software tools and resources from different remote infrastructures; 5) how to effectively manage the infrastructure resources in scientific experiments and allow scientists to achieve their goals more quickly; and 6) how to effectively record the events and results generated during experiments so that scientists can reproduce them independently.

The team uses the ENVRI reference model to establish a common understanding of data solutions and components within an RI. By applying the *agile software engineering approach,* the technical team closely collaborates with the RI developers through a set of use cases in the cycle of software requirement collection, technology review, system design, service prototype, and case validation. The agility of development activities has been greatly enhanced via more than 10 technical workshops in the context of research infrastructure site visits organized by theme2.

The "Data for Science" theme has a duration of four years. During the past project phases, the theme2 team has created
1) A collection of the knowledge for interoperable RI data management, including requirement analysis (D5.1), reference model guided RI design (D5.2, D5.3, D5.4 and D5.5), data management service design (D6.1, D7.1, D7.3, D8.1, D8.3 and D8.5), data management service implementation and demonstration (D7.2, D7.4, D8.2, D8.4, and D8.6), use case management and demonstration (D9.1, D9.2), and community exploitation (D6.2 and D9.3). By the end of the project, two more deliverables D6.3 and D9.4 on community exploitation will be delivered.
2) The portfolio of the software solutions, tools and services developed by theme2 have been made available to the RI communities. It grouped development results into four parts: methodology, interoperable data management services, reusable solutions from RI community, and e-Infrastructure test bed.
3) A number of use case demonstrators have also been tested on the e-Infrastructures of EGI and EUDAT, including user-defined data processing (EISCAT-3D), observation data subscription (EURO-Argo), eddy covariance fluxes (ICOS), nearly real-time data quality control, sensor registry, particle event information analysis, and mosquito disease distribution study (LifeWatch).

The interaction between the Theme 2 team and RI developer communities resulted in not only a much better appreciation of the challenge faced by the RIs but also recommendations to the RIs from the Theme2 team on solutions to assist scientific research using advanced ICT.

In the last phase of the project, the theme2 team will focus on knowledge transfer and community support. The summer school organized by theme2 and LifeWatch Eric was a typical example.

Furthermore, the team is aligning the results of ENVRI[PLUS] with the development of European Open Science Cloud to allow greater access by ENVRI researchers to powerful computing facilities and access by a much wider community to the assets of ENVRI.