

ERCIM



NEWS

Special theme:

**Data
Infrastructures
and
Management**

Also in this issue

Research and Innovation: Survival Skills in the Digital Era: Helping Teachers to Develop Digital Competences

D4SCIENCE: A Unique Infrastructure Delivering Virtual Research Environments as a Service

by Leonardo Candela (CNR-ISTI), Donatella Castelli (CNR-ISTI) and Pasquale Pagano (CNR-ISTI)

Nowadays, research challenges – often based on the collaborative analysis of a large amount of data – require suitable infrastructures and user-facing solutions promoting multidisciplinary collaboration and appropriate communication and sharing of data, processes, and outcomes. The D4Science infrastructure and its virtual research environments proved to be a viable and effective solution for many communities of practice and use cases.

Several initiatives are ongoing to develop infrastructures for supporting designated communities by providing them with data and computing capacity. D4Science [L1] is an infrastructure specifically conceived to deliver virtual research environments (VREs) via the as-a-Service paradigm [1,2]. Its development started 18 years ago as a testbed, and over time it progressed towards an operational infrastructure through the support of a series of EU Commission-funded projects.

D4Science-based VREs are web-based, community-oriented, collaborative, user-friendly, open-science-enabler working en-

vironments for scientists and practitioners willing to work together to perform a specific (research) task.

Each VRE manifests in a unifying web application hosted in a web gateway (and a set of application programming interfaces (APIs)). The application comprises several components made available by portlets organised in custom pages and menu items running in a plain web browser. Every component aims to provide VRE users with facilities implemented by relying on one or more services, possibly provisioned by diverse providers. Every VRE offers seamless access to the datasets and services of interest for the designated community while hiding the diversities originating from various resource providers.

Among the facilities each VRE offers, some basic ones allow VRE users to perform their tasks collaboratively, namely: (a) a workspace component to organise and share any digital artefact of interest, (b) a social networking component to communicate with co-workers by posts and replies, (c) a data analytics platform to share and execute analytics methods by relying on a distributed and scalable computing infrastructure, and (d) a catalogue component to document and publish any digital artefact worth sharing.

Along its lifetime, the D4Science infrastructure has supported the delivery of VREs for very diverse communities and usage scenarios. The creation of these VREs has been a continuous process. While some VREs were dismissed upon the completion of the activity that had motivated their deployment, others have been maintained to support continuously ongoing activities.

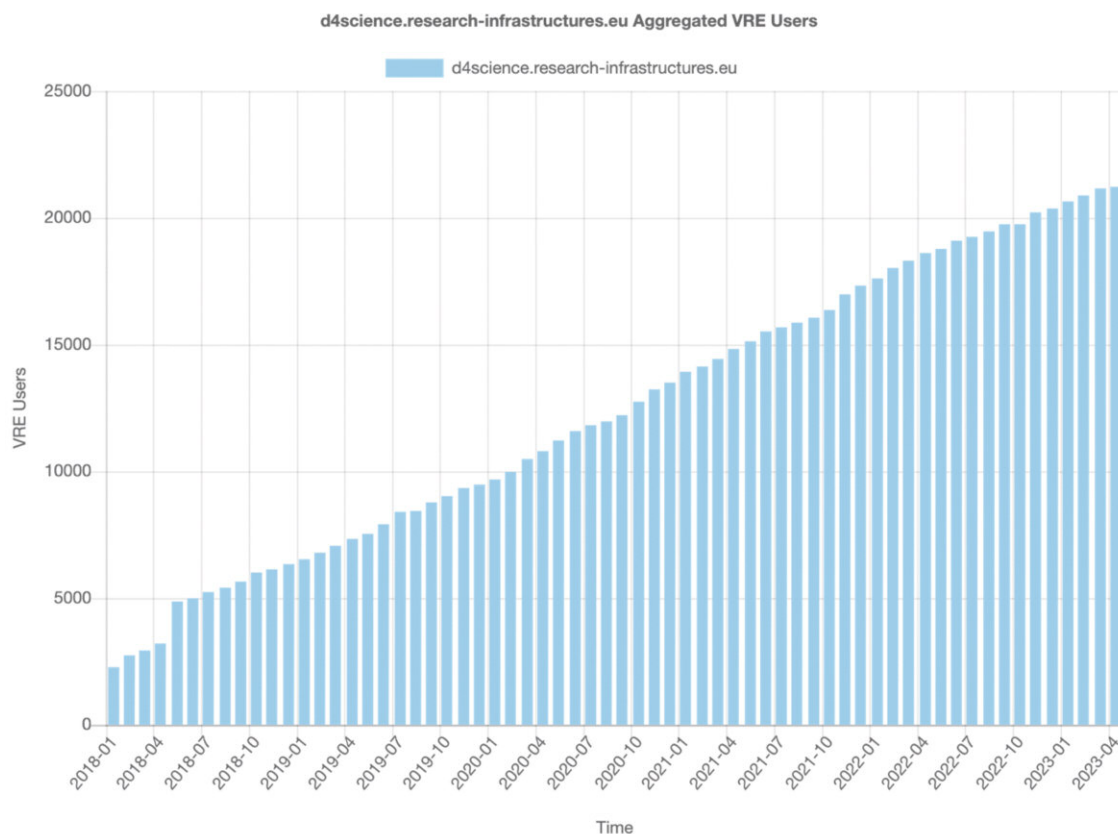


Figure 1: D4Science Users.

Currently D4Science operates 187 VREs (with others to come) made available by 20 thematic gateways [L2]. These environments support the activities and tasks of diverse communities of practice centred around organisations like FAO, ESFRI Research Infrastructures, EU and national projects, and operating in different thematic areas from marine science to social science, humanities, agri-food, health, and geothermal science. The D4Science user base counts over 20,000 users from almost all over the world. Figure 1 displays the constantly growing trend of the user base in the past five years.

From the analysis of the requests and feedback received from this large user base it emerges that the following key features make the D4Science infrastructure a unique and effective solution for many use cases.

The VREs as-a-Service delivery approach is, to a large extent, the most relevant feature of the D4Science solution for most user communities. Most do not have the necessary resources and personnel to deploy, host and maintain such environments. Often they are also looking for solutions to help them to minimise the “time-to-market”, i.e. the time in which they can start using the VRE to support their specific activities. D4Science implements a VRE distribution model in which it hosts the whole application and provides it to users over the Internet as-a-Service [1]. The advantage of this design choice is that the actual management of the IT solution is in the hands of expert operators who manage it by providing reliable services, leveraging economies of scale, and using elastic approaches to scale. A new VRE is created by using a wizard to select the VRE's functional constituents among those available. The software components' deployment and configuration implementing the selected functionalities are completely automatic. It leads to a new and ready-to-use VRE made available through one of the gateways operated by D4Science.

The system of systems approach is paramount to promote the establishment of synergies with several service providers and to enlarge the capacity and service offering exploitable when creating and operating VREs. In fact, D4Science was designed to conceptually play the role of a central hub offering seamless access to its own resources (datasets, services, computing and storage capacity) as well as to services and computing capacity offered by other infrastructures and service providers. All the resources aggregated by the federated service providers are registered into a unifying information system, monitored, and exploited on demand to contribute to the creation and operation of the various VREs.

Catering for co-creation is paramount to guarantee community uptake and the incremental evolution of the VRE to meet the designated community's changing needs. Communities of practice have evolving needs and often refine their requirements when using the provided working environments and services. VREs cannot be static environments; they must evolve, making available new tools and datasets to meet emerging needs. D4Science supports integration patterns [2] to complement the offering and bring new resources into VREs by facilitating the incorporation of community-specific existing applications, analytics methods and workflows, datasets and other resources for discovery and access. This co-creation mechanism counts on the presence of a working ver-

sion of the VRE where community resources are “hot-plugged” without stopping or shutting down the environment.

Open Science is here to stay, yet it must be supported by an open access approach “as early as convenient”. It implies collaboration and sharing, reproducibility and transparency to as wide and great an extent as possible. Scientific communities willing to operate in line with this approach have found in the D4Science VREs concrete support for flexibly meeting these properties and implementing Open Science practices with the needed shades of openness. D4Science VREs are equipped with basic services supporting collaboration and cooperation among their users. They also continuously and transparently capture research activities, authors and contributors, as well as every by-product resulting from every phase of a typical research life cycle, thus offering a solid base for addressing Open Science practices like, for example, reproducibility, research assessment, communication and collaboration, and transparency.

At the current stage, we can state that this solution has many advantages, as demonstrated by its high uptake. The VREs as-a-Service represents, for many communities (especially communities of practice in long-tail science), the ideal solution for solving the need for their collaborative activities, especially when these are data-driven and computationally intensive and go beyond the boundaries of institutions and regions. Indeed, they largely reduce the time for a community of practice to become operational and the need for skilled personnel dedicated to technology development.

Links:

[L1] <https://www.d4science.org>

[L2] <https://services.d4science.org/thematic-gateways>

References:

[1] M. Assante et al., “Enacting open science by D4Science”, in *Future Gener Comput Syst*, vol. 101, pp. 555-563, 2019, doi:10.1016/j.future.2019.05.063.

[2] M. Assante et al., “Virtual research environments co-creation: The D4Science experience”, in *Concurrency Computat Pract Exper*, e6925, 2022, doi:10.1002/cpe.6925.

Please contact:

Leonardo Candela, CNR-ISTI, Italy
leonardo.candela@isti.cnr.it