Digital Arts and Humanities in the European Open Science Cloud

Davor Davidović, Eva Cetnić, Karolj Skala
Rudjer Boskovic Institute, Bijenicka cesta 54, 10000 Zagreb, Croatia
{ddavid, ecetinic, skala}@irb.hr, +385 1 4561091

With the growth of digital arts and humanities related research content, emerges the need for data storage resources, as well as the need to easily access, search and analyse this type of data. While other research areas are already making considerable use of e-Infrastructure services, arts and humanities disciplines are still underexploiting these facilities. Researchers in this area often experience difficulties with using advanced digital research infrastructures and services because they either don't have access to infrastructure services or cannot use them in an efficient way due to technical, functional or legal constraints. Often the reasons for underutilizing e-infrastructure services is the lack of awareness of their existence, as well as the lack of technical knowledge that is required to install, setup and run those services.

To move beyond the state-of-the-art and alleviate these problems, it is important to foster the collaboration between digital arts and humanities research groups and e-Infrastructure providers. As the leading European social and technical infrastructure for arts and humanities, DARIAH (Digital Research Infrastructure for the Arts and Humanities) is currently working on the integration with existing and emerging e-Infrastructure initiatives through the framework of European Horizon 2020 projects EGI-Engage[1] and INDIGO-DataCloud[2]. The aim of these two projects is to accelerate and support the use of cloud infrastructures among a broad variety of research communities, including the digital arts and humanities community, represented by DARIAH.

The mission of the EGI-Engage project is to promote the implementation of the Open Science Commons[3] by extending the capacities of European foundation of federated services, data, knowledge and expertise among researchers from various disciplines with the aim to foster their collaboration and productivity. EGI-Engage plans to expand the facilities offered to scientists (e.g. improved cloud or data services) and the scope of its user base by engaging with large Research Infrastructures (RIs), the long tail of science and industry/SMEs. The main engagement mechanism of the EGI-Engage project is a network of eight Competence Centres (CC), where each centre is operating using joint forces of Research Infrastructures, National Grid Initiatives (NGIs), user communities, technology and service providers to collect requirements, integrate community-specific applications, promote interoperability across e-Infrastructures and evolve services through a user-centric development model. The DARIAH CC[4] is one of eight competence centres that form the core structure of the EGI-Engage project. The DARIAH CC is a virtual research group, consisting of computer scientists, service providers and researchers coming from the arts and humanities. The main

1 EGI web page: www.egi.eu, EGI-Engage wiki page: wiki.egi.eu/wiki/EGI-Engage:Main_Page
2 INDIGO-Datacloud web page: www.indigo-datacloud.eu
3 Open Science Commons: www.opensciencecommons.org
4 EGI DARIAH CC: wiki.egi.eu/wiki/Competence_centre_DARIAH

goal of the DARIAH CC is to widen the usage of the advanced research infrastructures and technologies, such as cloud-oriented services, computational and storage resources in the domain of the arts and humanities research. By extending generic EGI e-Infrastructure services according to the unique needs of arts and humanities users, the DARIAH CC aims to provide customised e-Infrastructure services and implement specific use-cases from the arts and humanities that depend on those services.

Based on the presumption of the underutilization of e-Infrastructure services among arts and humanities, the DARIAH CC developed an engagement strategy to achieve a far-reaching impact by incorporating both generic solutions and services, as well as provide concrete solutions and applications for specific research domains or groups. One of the main generic solution developed within the DARIAH CC is the DARIAH Science Portal[5] used for accessing EGI FedCloud compute and storage resources. The portal is based on the generic-purpose WS-PGRADE technology and its main purpose is to serve as a key access point for domain specific applications and their execution in a cloud-based environment. To address the quest for targeting specific research groups, the DARIAH Science Portal hosts several domain-specific applications such as multi-source distributed search and information retrieval (based on CDSTAR) and Bavarian dialects repository (based on gLibrary). The idea of the portal approach is to hide the technical complexity of the underlying cloud system by providing an easy-to-use, web-based graphical interface, enabling users to setup, manage and execute their research cases in the Cloud. Since DARIAH is a member of EduGain federation, the DARIAH science portal is currently in the process of becoming a registered EduGain service provider. Once registered, DARIAH members will be able to access the portal by using their own DARIAH credentials and exploit offered services, applications and the FedCloud infrastructure.

The DARIAH CC is in tight collaboration with the INDIGO-DataCloud project which represents another initiative for fostering the usage of cloud-based solutions among researchers from different domains. INDIGO-DataCloud focuses on filling the gaps of current cloud technologies regarding redundancy, scalability, resource utilization, streaming, multimedia, big data, performance, etc. Different case studies have been proposed and analysed in detail by the different research communities on the basis of the interest on INDIGO-DataCloud solutions. One of the research communities involved in the development of INDIGO solutions is the DARIAH community. The mission of INDIGO-DataCloud is to develop an open source platform that meets the data and computation needs of various scientific communities. Considering the heterogeneous properties of arts and humanities data, as well as the insufficient experience with use of e-Infrastructure resources among researchers in this area, a challenging solution addressed within INDIGO-DataCloud project is the development of a user-friendly data repository platform for preserving and managing arts and humanities research data in the cloud.

5 DARIAH Science Portal: dariah-gateway.lpds.sztaki.hu

The main goal of this platform is to simplify the process of creating and managing repositories of digital assets from various arts and humanities disciplines. To be able to meet this requirement, one of the key features of the platform is user-friendliness. The platform will enable users, depending on the user role, the possibility to easily create data collections, store and retrieve data responsively and reliably, as well as include a data discovery mechanism and metadata management. Each DARIAH community member (i.e. a person authorized by the DARIAH IdP) will be able to access this platform and create her/his own customized repository of data collections, upload data and related metadata, annotate data and validate other data annotations, define access rights to end-users and invite other users to access data collections. The platform comprises multiple cloud-based services including federated authentication and authorization, distributed data storing and management and cloud resources allocation. To provide multi-cloud support and easiness of the deployment, all the services are packed and run inside containers, such as Docker, and hosted in the cloud.

The EGI-Engage and INDIGO-Datacloud projects represent two interweaving initiatives which aim to foster the usage of cloud technologies among scientists, as well as promote open access and collaborative and sustainable research. The outstanding results of those two projects serve as a foundation for building the European Open Science Cloud (EOSC) which aims to strengthen Europe's position in data-driven innovation by federating existing research infrastructures currently scattered across disciplines. In our digital era, innovative digital services have become a backbone for excellent research in all scientific disciplines. Arts and humanities disciplines are not an exception, although the full exploitation of digitized and new-born digital arts and humanities data, as well as domain-specific digital tools and methods, is yet to come.