

Supporting digitally enhanced scientific workflows in a clustered Infrastructure: H2IOSC

Emiliano Degl'Innocenti, Francesco Coradeschi, Alessia Spadi, Federica Spinelli, Irene Falini, Michela Perino,
Lucia Francalanci, Francesco Pinna
Istituto Opera del Vocabolario Italiano del Consiglio Nazionale delle Ricerche (OVI-CNR), Italy

Keywords: research infrastructures; semantic framework; scientific pilots; FAIR principles

H2IOSC (Humanities and cultural Heritage Italian Open Science Cloud) is a project funded by the National Recovery and Resilience Plan (PNRR) in Italy, aiming to create a federated cluster comprising the Italian branches of four European research infrastructures (RIs) - CLARIN, DARIAH, E-RIHS, OPERAS - operating in the "Social and Cultural Innovation" sector of ESFRI (European Strategy Forum for Research Infrastructures).

H2IOSC facilitates collaboration among researchers from various disciplines in the social sciences and humanities fields. Its primary objective is to conduct data-driven research activities, by supporting several scientific workflows. Around the workflows selected by each participating infrastructure H2IOSC will build as many Scientific Pilots supported by digital data, tools and services, in the form of Virtual Research Environments, by using 8 high-performance computing data centers built by the cluster. Additionally, it aims to enhance the adaptation, implementation, and efficiency of existing facilities, based on the needs identified and expressed by the respective communities. H2IOSC will establish an onboarding process for each federated infrastructure within its framework, that will ensure the designated maturity threshold. Moreover, H2IOSC promotes training and dissemination activities to foster proficiency in the FAIR (Findable, Accessible, Interoperable, Reusable) and Open Science domain. Ultimately, H2IOSC seeks to drive the digital transformation of cultural and creative industries.



Fig. 1: H2IOSC project overview

The role of DARIAH.it in the project is to build the distributed infrastructure and federate its 8 nodes, elaborate the semantic framework to represent the knowledge gathered in H2IOSC in an interconnected way and develop the Scientific Pilots supporting the digital philology workflow. To do so, DARIAH.it will undertake a set of preparatory activities, including collecting and evaluating existing tools, datasets, and services relevant for the above task. A crucial part of this process is the semantization of selected data and metadata to promote interoperability of diverse resources; this involves the definition of standardized vocabularies and ontologies alongside the deployment of tools and workflows which actually implement the semantic transformation. The Pilots, presented as platforms or hubs, integrate domain-specific services, workflows, and interfaces. They operate

on meticulously parsed data-subsets to facilitate the development of prototypes. Connected to this research and development activities, DARIAH.it will promote training, outreach, and dissemination efforts, as the funding of a Doctoral Scholarship in Digital Philology.

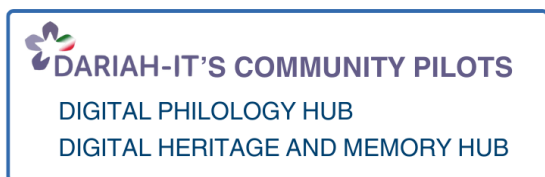


Fig 2: DARIAH.it pilots in the H2IOSC project

DARIAH.it assumes the responsibility of designing, developing, and populating AEON, a service-provision oriented infrastructure interacting with the H2IOSC Marketplace and aligned with SSHOC and EOSC platforms. Ultimately, by participating in the project, DARIAH.it endeavors to provide an indispensable service that ensures the development and sustainability of humanities and human sciences research, as well as cultural heritage preservation, by facilitating the transition to digital approaches in research activities, overcoming the obsolescence of current sector-specific research product systems, and maximizing the potential impact of research communities by developing innovative approaches.

Moreover, DARIAH.it establishes a foundation for interactions with the Italian Cultural and Creative industry and GLAM sectors, fostering collaborative research and development endeavors to enable the adoption of cutting-edge technologies. Simultaneously, it strengthens training programs and enriches content through valuable knowledge transfer activities.

References

1. European Strategy Forum on Research Infrastructures (ESFRI, <https://www.esfri.eu/>, date of access: 01.02.2024).
2. COUNCIL RECOMMENDATION on a Pact for Research and Innovation in Europe. 13701/21 EL/DOS/en ECOMP.3.B.
3. MINISTERO DELL'ISTRUZIONE, DELL'UNIVERSITA' E DELLA RICERCA, DECRETO 9 maggio 2019 Ammissione del progetto «DARIAH-IT - Developing nAtional and Regional Infrastructural nodes of dAriaH in ITaly» al finanziamento previsto dal decreto direttoriale 28 febbraio 2018. (Decreto n. 900/2019). (19A04611) (GU Serie Generale n.165 del 16-07-2019).
4. Degl'Innocenti, Emiliano, Monica Monachini, Alberto Bucciero, Enrico Pasini, Bruno Fanini, e Francesca Frontini. «H2IOSC: Humanities and Heritage Open Science Cloud». *La memoria digitale: forme del testo e organizzazione della conoscenza. Atti del XII Convegno Annuale AIUCD*, a cura di Emmanuela Carbé, Gabriele Lo Piccolo, Alessia Valenti, e Francesco Stella, 63–64, 2023. <https://iris.unive.it/retrieve/0f226d38-e332-418b-9b14-d5558d1a0d9d/AIUCD2023.pdf>.
5. Wilkinson, Mark D. et al., «The FAIR Guiding Principles for scientific data management and stewardship». *Sci Data* 3, 160018 (2016), <https://doi.org/10.1038/sdata.2016.18>.