Umanistica Digitale – ISSN: 2532-8816 – n. 20, 2025: Special Issue
R.B. Luzietti, A. Spadi, N. Giampietro, G. Mancuso, A. Caravale, A. D'Eredità, M. Caradonna, P. Moscati, V. Quochi, M. Monachini, E. Degli'Innocenti – *Digital Humanities and Heritage Science*
DOI: http://doi.org/10.6092/issn.2532-8816/21226

Digital Humanities and Heritage Science: moving from landscaping to a dynamic research observatory in an Open Science Cloud

Roberta Bianca Luzietti
Istituto di Linguistica Computazionale "Antonio Zampolli" (CNR-ILC), Italia
robertabianca.luzietti@cnr.it

Alessia Spadi
Istituto Opera del Vocabolario Italiano (CNR-OVI), Italia
alessia.spadi@cnr.it

Nicola Giampietro
Istituto per il Lessico Intellettuale Europeo e Storia, delle Idee (CNR-ILIESI), Italia
nicola.giampietro@cnr.it

Giacomo Mancuso
Istituto di Scienze del Patrimonio Culturale (CNR-ISPC)
giacomo.mancuso@cnr.it

Alessandra Caravale
Istituto di Scienze del Patrimonio Culturale (CNR-ISPC)
alessandra.caravale@cnr.it

Antonio D'Eredità
Istituto di Scienze del Patrimonio Culturale (CNR-ISPC)
antonio.deredita@cnr.it

Marta Caradonna
Istituto per il Lessico Intellettuale Europeo e Storia, delle Idee (CNR-ILIESI), Italia
marta.caradonna@cnr.it

Paola Moscati
Istituto di Scienze del Patrimonio Culturale (CNR-ISPC)
paola.moscati@cnr.it

Valeria Quochi
Istituto di Linguistica Computazionale "Antonio Zampolli" (CNR-ILC), Italia
valeria.quochi@cnr.it

Monica Monachini
Istituto di Linguistica Computazionale "Antonio Zampolli" (CNR-ILC), Italia
monica.monachini@cnr.it

Emiliano Degl'Innocenti
Istituto Opera del Vocabolario Italiano (CNR-OVI), Italia
emiliano.deglinnocenti@cnr.it

Abstract

This article presents a methodological framework for analyzing the digital infrastructure landscape in Italy in terms of availability of data resources, technologies, services, and training, as well as users' needs in the fields of Digital Humanities and Cultural Heritage Science. The initiative undertakes a comprehensive assessment of resources, tools, community requirements, standards, and best practices to facilitate the integration and FAIRification of digital assets of direct relevance to stakeholders while also identifying gaps and potential corrective actions. The collaborative involvement of four national research infrastructures (CLARIN, DARIAH, E-RIHS, and OPERAS) within the H2IOSC federation leads to the development of a broad and replicable set of instruments for analyzing the national landscape. This approach supports the identification of strategic priorities for improving scholarly access to and reuse of digital materials, and the enhancement of the exchange of best practices and training initiatives. The study also contributes to the design of an online observatory, meant to publish the findings of periodic landscape surveys over the coming decade and to serve as a community-building tool through interactive engagement with users and stakeholders.

Keywords: research Infrastructures, digital humanities, heritage science, infrastructural survey, #AIUCD 2024

Questo articolo presenta un quadro metodologico per l'analisi del panorama delle infrastrutture digitali in Italia, con riferimento all'offerta di risorse dati, tecnologie, servizi e attività formative, nonché ai bisogni degli utenti nei settori delle Digital Humanities e delle Scienze del Patrimonio Culturale. L'iniziativa prevede una ricognizione delle risorse, degli strumenti, delle esigenze delle comunità, degli standard e delle buone pratiche, con l'obiettivo di facilitare l'integrazione e la FAIRificazione degli asset digitali di diretto interesse per gli studiosi, individuando al contempo lacune per poter indirizzare efficacemente possibili interventi correttivi. La collaborazione tra i nodi nazionali di quattro infrastrutture di ricerca europee nell'ambito della federazione H2IOSC (CLARIN, DARIAH, E-RIHS e OPERAS) ha condotto allo sviluppo di un insieme replicabile di strumenti per l'analisi del panorama nazionale. Questo approccio consente di definire priorità strategiche per migliorare l'accesso e il riuso dei materiali digitali da parte della comunità scientifica, oltre a promuovere la condivisione di buone pratiche e iniziative formative. Lo studio contribuisce, infine, alla progettazione di un osservatorio online, concepito per pubblicare i risultati delle campagne di rilevazione che verranno periodicamente condotte nel prossimo decennio, e per fungere da strumento di community building attraverso un'interazione attiva con utenti e portatori di interesse.

Parole chiave: infrastrutture di ricerca, umanistica digitale, scienze del patrimonio culturale, ricognizione sulle comunità delle infrastrutture

Introduction

Research infrastructures (RIs) are becoming essential pillars supporting innovation, collaboration, and openness in research practices across many domains, including the Humanities and Cultural Heritage. Recognized as fundamental assets by the European Strategy Forum on Research Infrastructures (ESFRI)[1], infrastructures form the backbone of the European Research Area, enabling access to advanced technologies, data, and services across national and disciplinary boundaries, and thus helping better address societal challenges and

[1] https://www.esfri.eu (last accessed 19/01/2024).

consolidating the European Research Area (ERA) [10]. The European Open Science Cloud (EOSC)[2] consolidates this vision by offering a federated environment where researchers can store, share, and reuse data and services according to FAIR principles and Open Science practices. Within the EOSC and ESFRI umbrella, initiatives such as the SSHOC cluster have initiated the integration and consolidation of efforts within the Social Sciences and Humanities domain promoting the interconnection of existing and emerging infrastructures. At the Italian level, the Humanities and Heritage Italian Science Cloud[3] (H2IOSC) federation, clearly inspired by and connected to EOSC, takes a further step towards the operationalization of such clusters by federating the national operational nodes of four ESFRI research infrastructures (CLARIN[4], DARIAH[5], E-RIHS[6], and OPERAS[7]) with the ambitious goal of enabling access to advanced tools for conducting innovative research and encouraging multidisciplinarity.

In this context, landscaping activities are intended as the efforts to map existing resources, technologies, services, and community practices and needs, that serve as foundational instruments for guiding infrastructure development, fostering interoperability, and ensuring that research across domains remains inclusive, collaborative and sustainable. In the context of the H2IOSC landscaping activities represented a strategic value for infrastructure managers and providers as well as for users to collect information and suggestions on services and priorities in response to evolving scientific and scholarly practices, technological developments, and user expectations. To regularly continue to conduct landscape assessments will ensure that RIs remain up-to-date and capable of adjusting their trajectories in line with the communities they are serving.

In this contribution we present a coordinated and shared effort undertaken within the H2IOSC federation by all national RIs nodes, to define a replicable strategy to map and assess their operational landscape, particularly in the fields of Digital Humanities, Linguistics, and Heritage Science in Italy. Such investigation was driven by the need to identify gaps in the availability and use of data, tools, protocols and services; evaluate the degree of alignment with FAIR principles[8]; assess training needs in general and specific topics. As such, landscaping strategies and corresponding activities can be seen as complementary to the operation of RIs repositories and the H2IOSC Marketplace[9].

The primary objectives of this work are threefold: i) to strengthen the competitiveness of RIs as valuable means for content search and deposit of data, tools, and services; ii) to meet and better interpret the needs of different scholars and research communities; and iii) to encourage the integration of new resources and tools into the national cluster.

[2] https://eosc.eu/ (last accessed 19/01/2024).

[3] https://www.h2iosc.cnr.it (last accessed 19/04/2024)

[4] www.clarin.eu, www.clarin-it.it (last accessed 18/04/2024)

[5] www.dariah.eu, www.dariah.cnr.it (last accessed 18/04/2024)

[6] www.e-rihs.eu, www.e-rihs.it (last accessed 18/04/2024)

[7] https://operas-eu.org/ (last accessed 18/04/2024)

[8] https://www.go-fair.org/fair-principles/ (last accessed 18/04/2024)

[9] An advanced platform for the presentation, integration, and management of the H2IOSC service catalog, similar to the one developed within the SSHOC project https://marketplace.sshopencloud.eu (last accessed 28/01/2015)

The ultimate ambition of this initiative is to support RIs in increasing their reliability among scholars and capacity to align with the evolving demands of the digital humanities, linguistics, and heritage science research communities (i.e., students, researchers, and experts).

By adopting a mixed-methods approach, encompassing online questionnaire-based surveys, one-to-one interviews, focus groups, and data collection protocols, this work lays the groundwork for a permanent observatory, envisioned as a dynamic instrument for guiding infrastructure development and supporting long-term sustainability in alignment with Open Science principles. The involvement of users and academic communities is foundational since their direct engagement in surveying exercises, focus groups, and feedback can guide infrastructural development in real-world needs. In this context, landscape activities actively contribute to shaping and co-constructing the RIs.

Our efforts align with broader European efforts to map and monitor research infrastructures. Notably, the European Strategy Forum on Research Infrastructures (ESFRI) periodically publishes a Landscape Analysis[10], providing an overview of the European research infrastructure ecosystem by identifying key infrastructures operating transnational access in Europe across all fields of research, as well as major new or ongoing projects. Similarly, the European Open Science Cloud (EOSC) Observatory[11] serves as a sort of intelligence tool for monitoring policies, resources, and infrastructures related to EOSC, and offers a public interactive dashboard for data visualizations on the implementation and uptake of EOSC nodes at European and national levels. However, unlike most landscape analyses or observatories (where the sources of information, data collection protocols, or methodological criteria are often opaque) this study focuses on designing replicable instruments for collecting relevant data and information. Rather than concentrating on the analysis and interpretation of findings, the emphasis is placed on defining ways to capture community-informed evidence and to return it to the very communities that contributed it.

To present the working methodology and the results of the activities introduced, the article is structured as follows. The section "Methodology" describes the general approach adopted and theoretical framework. "Information gathering and classification" outlines the protocols used for data collection and the initial steps of data collection and analysis. The next two sections, "Questionnaire design" and "Focus groups design and organization" illustrate the design and implementation of the online questionnaire-based survey and of the focus group activities, respectively. "First results from the questionnaire and focus groups" presents and discusses preliminary findings from the analysis of questionnaire and the focus groups. Next, the section "Landscaping tools for digital cultural heritage", introduces a dedicated database and visualization service, particularly geared towards the Heritage Science domain. Finally, the contribution concludes with an overview of the design and objectives of a newly established H2IOSC Observatory, which collects all the main contributions of the work, and an outlook on future project developments.

[10] https://www.esfri.eu/landscape_analysis (last accessed 20/06/2025).

[11] https://eoscobservatory.eosc-portal.eu/home (last accessed 20/06/2025).

Methodology

This section is aimed at illustrating the general methodology adopted for landscaping the four RIs in Italy. The approach is conceived to be cyclically repeatable over time, serving both as a sort of health check of the four infrastructures and as a strategic compass to monitor developments, detect emerging needs, and inform future directions.

To address the complexity of the research landscape and the diversity of the communities involved, a Mixed Methods approach is adopted (Figure 1)[12]. This combines qualitative and quantitative strategies to:

- Provide a comprehensive overview of the current landscape of resources, tools and services, along with the needs expressed by the research communities
- Support related activities within the H2IOSC federation, particularly those focused on infrastructure construction and enhancement, service integration, user engagement, and training
- Allow for data-driven analysis and forecasting of user needs
- Elaborate a long-term strategy for the implementation and development of the H2IOSC Observatory.

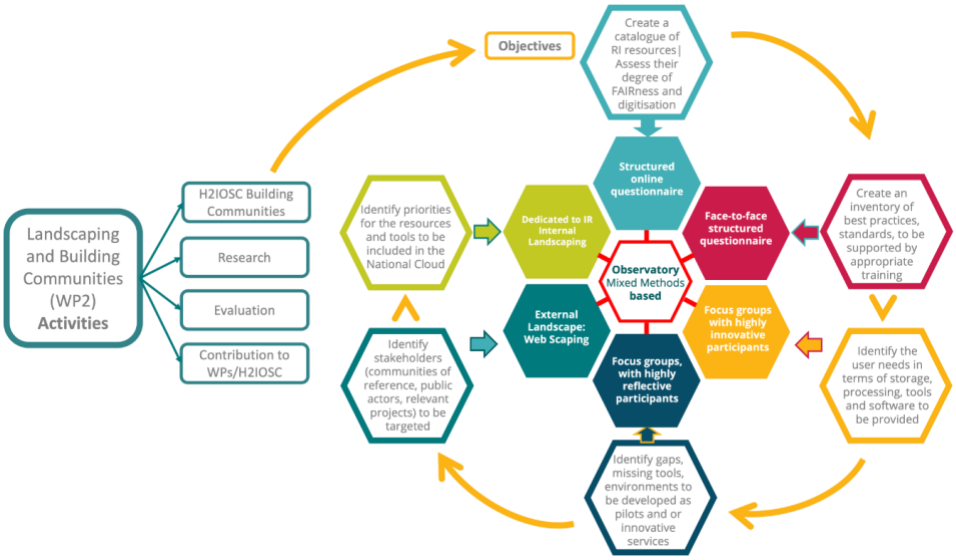

Figure 1 Schema illustrating the Mixed Methods approach employed in the H2IOSC surveying activity.

[12]Among the various definitions of mixed methods, the one proposed by [8] best suits this case.

Since the design of an apparatus capable of accounting for the existing projects, resources, tools, communities, best practices, and standards, in relation to each RI community involved in the project, requires the elaboration of a composite strategy, four primary instruments have been put together to collect structured information:

1. a mapping matrix used to catalogue systematize data retrieved from the different sources on projects, datasets, tools, protocols and standards, across different disciplines;
2. an online questionnaire-based survey designed to engage the stakeholder communities at large, and used to collect data on the use, production, and awareness of digital resources, tools, and Open Science practices, as well as training needs and alignment with FAIR principles;
3. focus group meetings, intended to complement quantitative data, are conducted with selected representatives of the target communities, including students, researchers, senior scholars and experts pertaining to the different disciplinary areas represented by the RIs. These sessions gather new insights on expectations, perceived gaps, and specific needs of the target communities in the use of digital infrastructures. Focus groups also foster interest in the project's activities. Participants are selected to ensure disciplinary and career-stage diversity across the four RIs;
4. a database for storing and facilitating navigation of all the collected data. It facilitates longitudinal analysis and supports the development of dashboards and observatory functions. Strategic starting points for this activity are existing repositories and catalogues pertaining to two RIs involved in this project, which served as valuable initial sources of information (e.g. ILC4CLARIN[13]).

Quantitative data (obtained primarily from the questionnaires and integrated with data derived from interviews and focus groups) is analyzed using descriptive statistics to identify usage patterns, levels of awareness, and potential gaps. Qualitative data from focus groups is, on the other hand, thematically coded to extract recurrent concerns, needs, and suggestions. As a result, the triangulation of methods ensures robustness and consistency between findings and strategic recommendations.

The obtained insights are expected to address several key areas:

- **Prioritization**: identifying the most critical data resources, tools, and services that require urgent integration into the RIs and the H2IOSC Marketplace;
- **Enhancements, FAIRification, and servification**: determining which resources need improvement, FAIRification, or transformation into services;
- **Gaps identification**: detecting the absence of crucial resources, tools, and services within existing RIs, and addressing the specific needs of the Italian research communities;
- **Training needs**: identifying gaps in knowledge, skills, and competencies within the community to guide the development of targeted training programs, materials, and

[13] https://ilc4clarin.ilc.cnr.it/, https://dspace-clarin-it.ilc.cnr.it/

initiatives.

Information gathering and classification

The creation of an observatory aimed at mapping the panorama of the H2IOSC community in Italy is a challenging task. As mentioned above, our approach towards this goal relies on a Mixed Methods strategy. A notable aspect consists of the nature of the four participating RIs, which differs considerably since they deal with different issues in many aspects, including but not limited to different kinds data resources (e.g.: oral/written text corpora, digital scholarly editions, multimedia datasets, 2D/3D survey data such as photogrammetric surveys and laser scanning, GIS databases, etc.) and disciplines (e.g.: philology, lexicography, linguistics, cultural heritage studies, archaeology, etc.). In fact, while operating within the broader domain of the Humanities and Cultural Heritage, each RI focuses on only partially overlapping aspects and priorities. More in detail:

- CLARIN prioritizes high-quality language resources that are well-documented, relevant to a wide range of research areas, technically robust and interoperable. The reference community comprises researchers in linguistics, literature, history, sociology, and the humanities, but its utility reaches any discipline requiring high quality language technology. CLARIN's focus is on linguistic resources, including less-resourced languages and dialects.

- DARIAH focuses broadly on digital humanities research on disciplines such as philology, lexicography, archival science, history, Latin and Romance languages, and art history. It deals with resources including corpora, digital editions, archival records, and those related to the study of physical objects and artefacts (technical art history). Moreover, its landscape encompasses available records of written or photographic resources and/or other specific documentation datasets about artworks, catalogues, and other iconographic resources.

- E-RIHS landscaping activity includes existing digital resources such as texts, images, audio, video, and 3D models related to historical, cultural, and artistic heritage. Focusing on Cultural Heritage, E-RIHS is also interested in digital resources, repositories, services, tools, and best practices for data management and sharing to manage its resources.

- OPERAS focuses on Open Science and FAIR principles within the Humanities and Cultural Heritage research community in Italy. The mapping activity includes open publishing systems in Italy, focusing on journals, books, and other publication formats specific to Humanities and Cultural Heritage disciplines, as well as on existing tools, services, and training programs that can assist researchers in making their publications and research data more open and FAIR.

Therefore, the identification and classification of all the necessary information does not just require a shared strategy, but also an overall consideration of all the RIs' specific features and needs. The goal here consists of mapping tools, datasets and projects supported by the RIs,

emphasizing the importance of not just gathering data but also gaining an overall understanding of the activities taking place within the different discipline domains. The definition of common practices and shared parameters is imperative to map resources across the four infrastructures effectively. The methodology employed thus focuses on a detailed consideration of the current state of available resources and technological services within the linguistic, humanistic, and heritage sectors of the research institutions involved. The heterogeneous range of mapped resources poses a challenge in defining a shared set of work tools to classify, group and correctly describe different resource types, standards, reference ontologies, vocabularies, taxonomies, and specific domain features. Particular attention is paid to the FAIRness assessment process involving findability (online or in physical locations), accessibility, interoperability, and reusability in other contexts.

As the result of a collaborative effort among the four RIs, three data structures have been defined to map datasets, tools, and projects respectively. To describe each resource, the following pieces of information are thus recorded: i) general information about the resource; ii) data lifecycle management policies; iii) status (i.e., new, updated, etc.). More in detail, among the parameters of interest: acronym, name, description, classification, context of use, curator, data format (for input and output), standards, licenses, notes are also recorded. In this work, extensive documentation is also developed to support the data entry process, so that each contributor can proceed easily. The tables gather a compilation of information to support other activities, like FAIRness evaluation of the resources. By identifying and mapping existing resources, the landscape activity helps researchers within (and outside) the federation to discover relevant data, tools, and services that they may not have been aware of, leading to improved resource discovery and enhanced data accessibility, also in line with the Open Science principles. Overall, the insights gained from this effort support the creation of more effective research collaborations by putting together for the first time resources coming from different perspectives of the SSH and CH research domains.

It is important to mention that criteria for recording information about resources and technologies established during this work are consistent in the design of the questionnaire, of the focus groups, and of the database, that will host all landscaping data to be made available to the permanent observatory. This will ensure that all the instruments are coherent and interoperable.

Questionnaire design

Among the core surveying activity there is a designed and comprehensive online questionnaire aimed at directly engaging with the target research communities to gather essential information on key aspects, including the usage and needs related to data resources and technologies, awareness and current utilization of existing research infrastructure services, desiderata for new services or offerings, and training needs as perceived by community members (students, researchers, and experts pertaining to the different disciplinary areas represented by the Ris) . The latter focused particularly on competencies and skills identified throughout the survey, which are crucial for the development of effective training programs, materials, and campaigns.[14]

[14] To do so, the group worked in strong collaboration with the project team dealing with training, also very involved with the target community.

The decision to create a single questionnaire applicable across all RI communities was initially made to prevent data fragmentation and mitigate the risk of "community overload" [7], given the interdisciplinary nature of many researchers' fields of specialization, which can often span multiple RIs. The elaboration process began with identifying the expected information to be collected [4], which encompassed several outcomes: the identification of stakeholders, user needs, training needs, priorities, gaps in available tools and services, existing resources both within and outside the RIs, the degree of FAIR compliance, the digitization of new resources, and best practices and standards.

A series of targeted questions were then formulated to acquire the necessary information for each of these objectives. For instance, to determine if a resource indicated by a respondent was FAIR-compliant, four additional questions were included to gather details on its location, cataloguing, persistent identifier (PID)[15], and licensing. Each question was accompanied by explanatory notes to provide further clarification and to minimize potential misunderstandings. Additionally, definitions and references for potentially ambiguous acronyms or technical terms, such as FAIR and PID, were included to help participants understand the questions and respond accurately. This supplementary information also aimed to spark curiosity and encourage engagement. Each question was carefully designed and assigned a specific response format, including open-ended, single-choice, and multiple-choice options. The questionnaire was developed and deployed using LimeSurvey[16] and was structured into the following sections:

1. Personal Information & Privacy Policy: was aimed to collect general information about the respondents age, career level, ERC research field, and institutional affiliation. Respondents were provided with a written informed consent form at the beginning of the questionnaire, before answering any questions, ensuring that they were fully aware of the purpose and scope of the data collection. All personal data collected follow the requirements of Article 13 of Regulation EU 2016/679 (General Data Protection Regulation) and responses have been analyzed and store in anonymized and aggregated form.

2. Data Resources, Software, Tools, and Technologies: served to gather information about existing or newly created data resources, tools, etc. The goals here were multiple, from acquiring information on the state-of-the-art material within the Italian panorama, users' knowledge, expectations, and needs, as well as identifying gaps that should be addressed by other teams within the federation (e.g., point 3 and 7 in this list).

3. Projects: focused on collecting information on projects that have developed or are currently developing data resources, digital tools, software, or other technologies relevant to the respondents' disciplines, such as linguistics, archaeology, and philology.

[15] https://www.clarin.eu/content/persistent-identifiers (last accessed 18/04/2024)

[16] https://www.limesurvey.org/ (last accessed 18/04/2024)

4. Training needs: aimed at collecting respondents' expectations and comments regarding their personal training needs, with a particular focus on competencies and skills they deemed essential for future professional development.

5. Prior knowledge of RIs: was included to assess respondents' awareness of and level of engagement with the partnering RIs, which was considered critical for designing outreach and training actions to increase participation and engagement.

6. Publications: was aimed to explore how respondents and their institutions publish scientific articles. Questions focused on publication practices, awareness of institutional policies, and the extent of commitment to open-access publishing.

7. Contacts: e-mail addresses were collected at the end of the survey solely for contact purposes and with the respondents' informed consent.

8. Feedback: to share impressions and suggestions to the questionnaire and provide insights on how they learned about the survey.

Regarding the deployment of the questionnaire, a first draft was initially distributed to a selected group of colleagues representing various subcommunities of interest for the investigation; these colleagues acted as a test group and provided valuable first-hand feedback. Following this preliminary phase, the revised first official version of the questionnaire was forwarded to a control group composed of members from key Linguistics, Digital Humanities, and Cultural Heritage associations, for which the exact numerosity was known. This step allowed for an estimation of the potential participation of the broader community in the subsequent stages of the investigation. As a result of incorporating all the important feedback acquired from the first communities involved, we produced a second version of the questionnaire[17] that was divided in two parts: the first comprising the set of questions regarding the profile of respondents, training needs, knowledge of the RIs and publications, while the second part[18] was dedicated to the description of resources, tools and services ever used and created by respondents. The final version of the questionnaire was widely disseminated across all target community members, associations, and mailing lists through various channels, including social networks, federation and RI websites, conference presentations, and other outreach events.

Focus groups design and organization

One of the main qualitative research instruments is focus groups, a widely used qualitative research technique, particularly in marketing studies, sociocultural analysis, and in contexts aiming to understand how individual opinions behave in a social setting. Through a discussion guided by a moderator, this method elicits shared perceptions, opinions, attitudes, and meanings among participants, enabling the collection of rich and in-depth data on topics that are often

[17] The questionnaire-based survey will be open until June 2025 at https://survey.cnr.it/index.php?r=survey/index&sid=825638&lang=it, results of the questionnaire will be made available in Zenodo after the end of the project.

[18] To complete this second part of the questionnaire, respondents were asked at the end of the first part to indicate their preference for providing information either via a second online questionnaire or through a guided interview.

difficult to access through other survey instruments. Among the advantages of this method there are its capacity to generate complex data, its efficiency compared to conducting numerous individual interviews, and its ability to reveal group dynamics. Focus groups are also effective tools for idea generation and collaborative knowledge construction.

In this work, we use this instrument to fulfil the following main objectives:

- facilitate the involvement of communities belonging to the four federated infrastructures;
- complement the questionary-based survey results with qualitative data;
- gather qualitative data on the digital humanities experiences of the different audiences.

Given these points, we decided to organize a series of focus groups in which participants are invited to deliberate on their expectations regarding the outcomes, future developments, and long-term sustainability of the RIs federation. Particularly, (i) about the nature and quantity of resources, services, and tools to which they expect to have access through the H2IOSC Cloud; and (ii) discussing how such offerings should be delivered to the community (which might also inform the refinement and dissemination of the H2IOSC Marketplace).

The focus groups in H2IOSC are facilitated by a designated moderator, with the support of a 'critical friend', i.e., another researcher tasked with providing reflective feedback. This configuration enables the effective regulation of participation time and helps ensure that all cognitive objectives were successfully addressed (cf. [5] and [1]). A total of four focus groups have been planned, two of which were conducted between July and December 2024, with the remaining two held in March 2025, but not yet analyzed.

In this paper, we discuss the first two, for which preliminary analysis of the result is also available and presented in the dedicated section below.

Characteristics and theoretical assumptions of the two groups

In order to ensure effective moderation and high-quality discussion, each focus group is composed of a small group, homogeneous in terms of career level and attitude toward innovation, while also ensuring gender balance.

During the planning phase, it was decided that each group would consist of eight participants, and that the following variables would be monitored: gender, age, career level, orientation towards innovation and Research Infrastructure affiliation, estimated on the basis of the proximity of participants' disciplinary expertise to the macro research areas of the four H2IOSC infrastructures.

The first two groups, named FGSENIOR1 and FGJUNIOR1, are constituted as follows:

- both groups include four female and four male scholars;
- both groups ensure the presence of experts representing the four principal fields of study (i.e. Philosophy, Archaeology, Philology and Linguistics), with the aim of incorporating the contributions and perspectives of each of the four infrastructures;

- the first focus group, FGSENIOR1 (represented as a yellow hexagon in Figure 1 above), comprises individuals with a pronounced inclination towards exploring novel avenues of research, using innovative or experimental digital resources, and possessing substantial research experience. It would thus be formed of eight senior researchers or university professors (aged thirty-three to forty-one and over);
- the second focus group, FGJUNIOR1 (represented as a blue hexagon in Figure 1 above), has a strong focus on the present of learning and consists of early-stage researchers, i.e. ideally eight doctoral students and/or junior researchers (aged twenty-five to thirty-two).

From a theoretical standpoint, the research team designed the focus group activities around two distinct and complementary user profiles, each associated with a specific set of expectations regarding digital technologies and infrastructures (see Figure 1: yellow and blue hexagons).

The first group, FGSENIOR1, is expected to demonstrate a strong enthusiasm for innovation and experimentation with digital tools. Composed of senior researchers and university professors (primarily from Generation X[19]) this group is expected to show a proactive attitude towards the potential of digital research infrastructures. Their extensive research experience, coupled with a generational trajectory in which digital technologies became increasingly central over time, leads us to expect a high degree of engagement and forward-looking reflections on the role of digital infrastructures in scholarly practice.

Conversely, the second group, FGJUNIOR1, comprises doctoral students and early-career researchers (mainly from Generation Z[20]) who, being 'digitally native', are expected to approach digital tools as part of their everyday academic routine. As a result, they might take the presence of such tools for granted and be more critical when infrastructures do not meet their expectations. These assumptions are grounded in the observation that generational differences and varying degrees of research maturity could shape divergent expectations and perceptions regarding digital research infrastructures.

Characteristics of the conducted two groups

The first two focus groups were convened between July and December 2024. Compared to the original plans, the actual implementation involved smaller groups, with a total of 12 participants recruited. This reduction was due to the unavailability of some invited participants. The July session included five senior researchers from Italian universities and public research institutions (PRIs), comprising two archaeologists, one linguist, one philologist, and one philosopher. The September session brought together seven early-career scholars—specifically, two linguists, two philosophers, two philologists, and one archaeologist. The average age of the twelve respondents was 35.4 years, with the youngest participant being 25 and the oldest 61.

[19] Sociological study on young Italians, useful for understanding the context of the Italian Gen X [12]. It also contains data and interpretations relating to Gen X (now adults), with reflections on their transition to adulthood [11].

[20] Demographic and sociological analysis of Gen Z in Italy, with a focus on work, political participation and sustainability; see [13].

The rationale for incorporating focus group into the landscaping activities is predicated on the premise that, in comparison with alternative methodological constructs that have been adopted to date, focus groups constitute a more flexible instrument regarding adaptability and reusability for subsequent follow-ups.

In accordance with the definition provided by [1], the focus group conducted have facilitated the acquisition of an in-depth understanding by the landscaping team of the preferences of the stakeholders regarding resources, tools, and services related to all the communities involved in the project, for both the Marketplace implementation and the empowerment of the national RIs nodes. Preliminary findings that emerged from these first groups are discussed in the following section.

First results

Questionnaire

The community analysis conducted on the two versions of the questionnaire provided valuable insights into the participation, awareness, and use of Research Infrastructures (RIs), as well as the respondents' training needs and publication habits.

- Participation: overall, the first version of the questionnaire recorded 396 views with a completion rate of 22% (86 responses), while the second version had a higher relative participation, with 65% (97 responses) of participants completing the first part and 56% (29 responses) completing the second. The responses primarily came from researchers affiliated with the National Research Council (CNR), with 106 respondents, and universities, with 70 respondents.

- Career and disciplinary field: the distribution of respondents by career level shows a higher presence of senior researchers, with 99 participants, followed by 44 mid-career researchers, 17 early-stage researchers, and 14 students. Most respondents belong to diverse disciplinary fields, with a predominance in the social sciences and humanities, particularly in fields such as SH4_9 (theoretical and computational linguistics), SH5_8 (cultural studies, cultural identities and memories, cultural heritage), and SH6_3 (archaeology of early literate societies and early civilizations).

- Use and creation of resources and tools: relational databases are the most frequently used, with 66 responses, followed by linguistic databases and written language corpora, both with 54 responses. GIS resources and geographic datasets are used by 49 participants, while the creation of new resources is most common in relational databases, with 56 responses, followed by written language corpora with 36 responses, and linguistic databases with 28. Less frequently used technologies include artificial intelligence applications and 3D/BIM modeling. Seventy-nine percent of respondents reported being aware of research infrastructures, with CLARIN, DARIAH, and ERIHS being the most well-known, with 84, 100, and 104 responses, respectively. However, 20% of participants are unfamiliar with the available resources, while 73% of those who do not currently use them express an interest in exploring them further.

Regarding the sharing of research data, 109 respondents expressed their willingness to deposit their materials, while 22 have already done so. The reasons cited by those who have not yet deposited data include the perception of outdated platforms, complexities in managing data ownership, and the need for more information on sharing procedures.

- Awareness and involvement with H2IOSC and Research Infrastructures: awareness of the H2IOSC federation stands at 53%, with 51 respondents aware of the initiative and 37 actively involved. However, 13 involved individuals are not sufficiently familiar with it, highlighting possible gaps in information dissemination. The identified training needs show a strong interest in access to archives and repositories, indicated by 112 responses, data analysis tools, with 104 preferences, and data management and preservation, with 80 responses. The preferred training methods are self-paced distance learning, chosen by 109 participants, followed by live online sessions, with 87 preferences. In-person sessions were selected by 36 participants, while the hybrid mode received 40 preferences. Regarding the sharing of training materials, 89 respondents expressed their willingness to share, while only 6 were opposed.

- Publication practices: the most widely used system is OJS, with 87 responses, followed by WordPress, used by 24 participants, and Drupal, selected by 5 participants. The most common type of open-access publications includes scientific journals, with 141 responses, followed by conference proceedings with 75 and datasets with 42. Creative Commons licenses are the most frequently adopted, with CC BY used by 38 respondents and CC BY-NC by 29, while 44% of respondents do not use any specific license. Finally, the follow-up of the questionnaires recorded a high level of interest, with 59 respondents willing to be contacted again and 52 ready to participate in further questionnaires. Additionally, 14 participants have expressed their availability for in-depth interviews, highlighting a significant willingness to collaborate and further explore the topics addressed.

Focus groups

The first two focus groups—held in July and September 2024—already provide rich qualitative insights into participants' perceptions, experiences, and expectations concerning digital research infrastructures. A detailed analysis of the results lies beyond the scope of this article, as a comprehensive view will only be possible once the remaining two focus groups are analyzed. In what follows; therefore, we present a selection of highlights and preliminary findings from the first two focus group sessions described above, focusing on the applications, perceptions and expectations that coalesce or differentiate the two groups, and on those insights that methods such as the questionnaires cannot capture.

Both groups of respondents, FGSENIOR1 and FGJUNIOR1, share an overall familiarity with research infrastructures, with some participants already using their offerings. Among those who had only been exposed to the concept of Research Infrastructure, most expressed a great sense of curiosity regarding their potential for future exploitation.

Another point of convergence between the two groups is their familiarity with the principles of Open Science and the FAIR guidelines. However, their perspectives diverge with regard to the practical implementation of these principles in research methodologies. Senior researchers (FGSENIOR1), drawing on extensive field experience, emphasize that applying these principles

poses significant challenges, as certain research and study practices are constrained by specific agreements between the researcher and the commissioning body. In archaeology, for instance, such agreements often exist between the host state of the excavation and the institution to which the researchers are affiliated. Senior researchers highlight that these agreements are frequently based on property rights and intellectual property rights over the research, which stands in contrast to the principle of immediate and open dissemination promoted by the FAIR principles.

Doctoral students and early-career researchers (FGJUNIOR1) exhibit greater enthusiasm towards Open Science and the FAIR Principles, accompanied by a sound theoretical understanding. However, their attitude appears to be more ideologically driven, likely due to their relatively limited empirical experience. The younger cohort focused on the dialectical tension between the FAIR Principles and copyright law. While they expressed strong support for the principles of Open Science, they also noted a perceived conflict between these principles and the protection of intellectual property rights.

FGSENIOR1 participants all reported having published in Open Access and expressed strong interest in becoming more FAIR-compliant, despite highlighting the challenges. FGJUNIOR1 participants expressed that they have generally tried to be FAIR in their research and seemed quite confident that "*(...) science is not science if it is not open (...)*". However, as above, they also point out encountering difficulties regarding copyright.

Despite their varying levels of experience and different backgrounds, all participants have high expectations of research infrastructures in the SSH. Almost all (only a few have some doubts) expect to find within them all the resources, tools and services necessary for their research activities.

Divergences are instead observed mostly in relation to the participants' disciplinary field. Above all, expectations differ considerably across disciplines. For instance, while linguists desire a platform offering all the resources, tools and services essential for the management of linguistic corpora, capable of facilitating the reuse of existing resources in innovative ways, archaeologists emphasized the importance of a platform capable of serving as a repository for both publications and "field" materials.

One of the most important outcomes of FGSENIOR1, elicited by the spontaneous discussion, encouraged by the open format, on participants' personal experiences in using, developing, and creating digital resources, is the valuable observation that in some disciplines, such as computational linguistics, the boundary between "using" and "creating" resources is often perceived as fluid. In contrast, in other contexts/disciplines, this boundary appears more clear-cut. For example, linguists employing corpora to train large language models (LLMs) consider themselves not just "users" of existing resources but also creators of new ones, since trained models become new digital and reusable resources in their own right. This close relationship between resource use and production is therefore particularly important in fields like computational linguistics, where researchers are both users and producers. Their work exemplifies how using a resource often leads to the creation of something new, thus contributing to the broader digital research ecosystem.

FGJUNIOR1, composed of PhD students and early-career researchers, brought to light a different set of concerns. Particularly important for these participants was dissatisfaction with existing training opportunities, with institutional courses often described as too theoretical and lacking practical training. Many noted that, to acquire those practical skills, they had to seek out training independently. When asked about their familiarity with infrastructures, a distinction

emerged between research fellows, who generally knew and used them, and doctoral students, who were mostly aware but lacked hands-on experience. Notably, several expressed curiosity toward the H2IOSC project, particularly among those who had only limited, indirect exposure to it.

Another especially striking reflection from FGJUNIOR1 concerned the question of access. When asked to express their opinion on whether access to Research Infrastructure offerings should be free or paid, participants were divided between those favoring institutional payment and those advocating for completely free access. The motivation for free access presented by one participant was very interesting, relating the concept of citizenship to infrastructure access. If the infrastructure is national and publicly funded, the participant argued, then access should be free of charge, like other civic (digital) services. In this sense, the infrastructure is seen as a virtual home: "*(...) I like the idea of being present as a digital user in there, with my things and then making a little bit of my home in there, and I imagine it to be very clear and defined in the definition of the services that it offers of the working objects, so clearly the resources, the digital resources, the resources, the tools, the resources. And above all, also that there is a whole series of areas related to the sharing of ideas and therefore a space for sharing with other people, therefore contact with others*".

As a final remark, somewhat unexpectedly, this first analysis shows that the views and conceptions of the digital future of the Humanities and Cultural Heritage disciplines expressed within the first two focus groups are less divergent than expected, given the generational and professional differences represented. Although the differences certainly shape specific needs and expectations, all participants demonstrate a motivated, constructive and inquisitive position towards the evolving digital research ecosystem.

Landscaping tools for digital cultural heritage

The decision to develop a database to collect and analyze landscaping data comes from the acknowledgement of several needs within the H2IOSC research community, emerged during the surveying activity, such as i) the necessity of creating a database for existing digital products, tools/software and research projects utilized in heritage science and digital cultural heritage; ii) the requirement to classify and index those products; iii) the need for analytical tools to help RIs adjust their strategic plans over time. Its primary objective is thus to gain a deeper understanding of the evolving landscape of digital cultural heritage/heritage science and to effectively address its requirements in terms of digital tools ([2]: 516-518). Due to its purpose, it was named Digital Heritage Landscaping platfOrm (DHeLO - https://dhelo.cnr.it). One of the first challenges was to structure a data model able to fully represent the complexity and heterogeneity of the available data, while also allowing for sufficient segmentation for analytical purposes. Therefore, so far, six distinct entities have been identified (people, institutions, products, product types, tools, and research projects), all interconnected with each other with multiple relations. This conceptual scheme has proven to be an efficient mapping solution for digital products developed for Social Sciences and Humanities, providing an in-depth insight into the current progress of the disciplines ([6]: 525-528).

At present, although the data gathering process is still ongoing, enough information has been collected to enable a preliminary analytical approach. The initial phase of data collection involved scraping outputs and research projects from ISPC and the E-RHIS infrastructure, alongside integrating findings from the landscaping questionnaire conducted during the early stages of WP2. Subsequently, additional data were sourced from key data-sharing platforms relevant to the CH/HS disciplines, such as Zenodo (www.zenodo.org), AriadnePlus (https://ariadne-

infrastructure.eu), Iperion HS (https://www.iperionhs.eu), AdS (https://archaeologydataservice.ac.uk), tDAR (https://www.tdar.org), and OpenContext (https://opencontext.org). This was complemented by a parsing phase grounded in sector-specific literature, with a focus on publications from the past five years to ensure an up-to-date representation of the state of the art. In the field of Digital Archaeology (DA), peer-reviewed journals such as *Archeologia e Calcolatori*, *Virtual Archaeology Review*, *Open Archaeology*, and *Internet Archaeology* were analyzed[21]. For Digital Cultural Heritage (CH), key journals including the *Journal of Cultural Heritage* and *Digital Applications in Archaeology and Cultural Heritage* were indexed[22]. The Heritage Science (HS) domain was explored through publications in *Heritage Science* and *Heritage*[23]. Additionally, proceedings from discipline-specific series, such as *Computer Applications in Archaeology* and *MetroArchaeo*[24], published in the last five years, were considered. The data entry phase also included conducting individual interviews with selected professionals, such as university professors and leaders of research teams, who have extensive experience in managing projects related to open data and heritage science. These interviews were structured around the questionnaire previously described but evolved naturally, facilitating the collection of insights and perspectives. The discussions particularly focused on the role of research infrastructures in advancing discipline-specific research within the field.

Throughout this process, in compliance with the requirements of the H2IOSC federation, priority was assigned to Italian research outputs, while also considering those produced abroad, provided they were carried out by Italian researchers or research groups. Currently, the DHeLO web app hosts 265 records in the project table, encompassing 1,121 individual product types, along with metadata from 121 research projects and 114 tools. DHeLO also enables seamless interconnection and interoperability with other platforms, facilitated by the release of an API in JSON format (https://dhelo.cnr.it/api); it also supports statistical and spatial analysis using third-party software. Even at this early stage, the ongoing analysis of the collected data is progressively unveiling common practices, shared standards, and similar workflows across various research projects and institutions. One interesting and preliminary insight is the apparent lack of a direct correlation between the production of datasets and research projects. Specifically, almost 40% of the 265 catalogued products are not associated with a research project. This finding is particularly interesting as it suggests that the creation of datasets is increasingly becoming an integral part of the disciplinary workflow and notably this process appears to occur independently of the large, interdisciplinary research projects traditionally required for their

[21] Archeologia e Calcolatori: https://www.archcalc.cnr.it; Virtual Archaeology Review: https://polipapers.upv.es/index.php/var; Open Archaeology: https://www.degruyter.com/journal/key/opar/html?srsltid=AfmBOorEYnCFCBhOGM11ZFKbjKyxzMpmgDd0qMwn4HWLbaFI57FL2xl4#specialIssues; Internet Archaeology: https://intarch.ac.uk. Last access for all resources on the 09th January 2025.

[22] Journal of Cultural Heritage: https://www.sciencedirect.com/journal/journal-of-cultural-heritage; Digital Application in Archaeology and Cultural Heritage: https://www.sciencedirect.com/journal/digital-applications-in-archaeology-and-cultural-heritage. Last access for all resources on the 09th January 2025.

[23] Heritage: https://www.mdpi.com/journal/heritage. Last accessed on the 9th January 2025.

[24] Computer Application in Archaeology: https://caa-international.org; MetroArchaeo: https://www.metroarcheo.com Last accessed on the 9th January 2025.

production. This preliminary example shows how this holistic approach contributes to a more comprehensive understanding of the digital cultural heritage environment, fostering better cooperation and informed decision-making.

In parallel, the need also arose to develop a tool to gather the rich sector-specific literature systematically, the lifeblood of any research field that plays a crucial role in the knowledge production process and the information dissemination system. In this perspective, it was decided to implement a large work of collection and systematization of these resources, named Bibliography of Digital Archaeology (BiDiAr), starting with the references published at the end of each article in the journal *Archeologia e Calcolatori*, a peer-reviewed open access publication that has been a key scientific benchmark in the wide field of digital archaeology for over thirty years. The references were collected within Zotero, a well-renowned open-source software commonly used by the scientific community. Through Zotero API, BiDiAr relates to DHeLO, providing bibliographic references to the entry of the database. The cataloguing work has started with the volumes published in the last five years, with the idea to cover, over time, all the issues published since 1990. The volume of data collected (to date, more than 9000 entries) will provide an impressive bibliographic corpus related to the discipline, which will then be analysed using various keys and tools. Analyses will highlight how ICT (Information and Communications Technologies) tools and techniques for studying the past have evolved over time, identify the currently most advanced research areas and the schools and teams most active in this research field, which appears to be constantly evolving and updating given the pervasive nature of technology in all its aspects. The integration of both these landscaping tools will be one of the cornerstones of H2IOSC's permanent observatory to monitor the status of all these research communities.

The H2IOSC Observatory

Within the H2IOSC federation, an Observatory is currently under development, designed to provide a comprehensive and continuous assessment of outcomes resulting from landscaping activities and to facilitate a synergies and integration between research community practices and the digital environment provided by the H2IOSC Marketplace. By acting as a dynamic connector between these two domains, the Observatory is meant to ensure a continuous, bidirectional flow of information and insights. The Observatory is equipped to gather data through both quantitative and qualitative methods presented here, such as surveys, focus groups, and interviews to produce insights that will inform the advances in the needs and activities of the H2IOSC cluster community. These insights contribute to inform Ris on the latest developments, gaps in services, and help prioritizing new tools for FAIRification, and shaping future strategies for the integration of resources into the H2IOSC Marketplace. One of the key features of the Observatory will be its direct connection to the H2IOSC Marketplace, ensuring that data collected from user interactions are fully integrated with the federation's broader goals. The Observatory will generate knowledge outputs on a regular basis, including reports on innovative research practices and technical guidelines that can benefit a wide range of stakeholders, from research communities to policy makers. To ensure its sustainability over time, the Observatory will adopt an iterative approach to regularly map and survey the research communities for updating its outputs and potentially expanding to emerging areas of interest. By providing an evidence-based understanding of how researchers interact with digital tools and infrastructures,

the Observatory helps to optimize the H2IOSC Marketplace offerings and ensure that H2IOSC continues to meet the demands of its target communities.

Conclusion

This contribution presented the landscaping goals and explained the methodology implemented for the H2IOSC federated RIs for mapping the panorama of resources, tools, services, and needs characterizing the digital humanities, linguistic and cultural heritage sciences. The information collected throughout this investigation are meant to help:

- defining the H2IOSC federation target research communities;
- laying the foundation for the creation of the H2IOSC Marketplace;
- assigning the priorities for the integration of resources, tools and services in the H2IOSC Marketplace;
- providing information for the FAIRification of the identified resources to other teams involved in the construction, maintenance or enhancement of the federated RIs;
- identifying the training needs to be addressed;
- developing a permanent Observatory.

To do so, we employed a Mixed Methods approach to identify the main research communities' representatives, sources of information across online repositories, publications, academic events, etc. To encourage the direct participation of the RI's target research communities, we implemented a questionnaire and focus groups. Concurrently with the collection and analysis of the first results, we developed a database to easily reconcile, visualize, and interpret the acquired data that will be made findable and accessible via the permanent H2IOSC Observatory.

The present results reveal a positive trend in awareness and participation in research infrastructures and the mixed-methods approach allowed to gather many information and feedback from the research communities. Looking ahead, continuous refinement of community-building strategies and ongoing collaboration with research infrastructures will be essential to ensure that the H2IOSC Observatory and Marketplace will evolve into valuable, sustainable resource hubs. By aligning infrastructure offerings with the interests and needs of the research community, H2IOSC aims to contribute significantly to Italy's Open Science efforts, fostering greater cross-disciplinary collaboration and innovation.

Over the next decade, the surveying activities initiated through the initial H2IOSC project will play a pivotal role in shaping the strategic development of the H2IOSC Marketplace and RIs repositories, fostering a more interconnected and responsive research communities across the Humanities, Cultural Heritage, and Social Sciences domains. However, this long-term vision is accompanied by several challenges, including the continuous evolution of user needs, the sustainability of efforts, and the cultivation of active and diverse community engagement. Additionally, ensuring compliance with FAIR principles across disciplines and integrating emerging technologies, such as AI and big data analytics, will require ongoing innovation and

adaptability. To address these challenges a strategy will collect regular feedback mechanisms with user communities, implement modular and scalable infrastructure updates, and develop targeted training programs to enhance familiarity with best practices and new tools. From the efforts and first results of such collaboration, the H2IOSC federation will continue to monitor and sustain the needs of its research communities, ensuring impactful and sustainable RI cluster environment for years to come.

Acknowledgements

This work is supported by the H2IOSC Project - Humanities and cultural Heritage Italian Open Science Cloud funded by the European Union NextGenerationEU - National Recovery and Resilience Plan (NRRP) - Mission 4 "Education and Research" Component 2 "From research to business" Investment 3.1 "Fund for the realization of an integrated system of research and innovation infrastructures" Action 3.1.1 "Creation of new research infrastructures strengthening of existing ones and their networking for Scientific Excellence under Horizon Europe" - Project code IR0000029 - CUP B63C22000730005. Implementing Entity CNR.

For the redaction of this paper all authors contributed reporting their work activities within the H2IOSC project and together the redaction of the introduction, methodology and conclusion section. More specifically, Roberta Bianca Luzietti, Valeria Quochi and Monica Monachini worked on the coordination of surveying activities and development and dissemination of the questionnaire. Alessia Spadi worked on the information gathering and classification task. Nicola Giampietro and Marta Caradonna conceived and conducted the focus groups meetings. Finally, Giacomo Mancuso, Antonio D'Eredità, Alessandra Caravale and Paola Moscati worked on the development of the Landscaping tools for Digital Cultural Heritage.

References

[1] Bezzi, Claudio. *Fare ricerca con i gruppi. Guida all'utilizzo di focus group, brainstorming, Delphi e altre tecniche*. Milano: Franco Angeli, 2020.

[2] Caravale, Alessandra, Paola Moscati, and Irene Rossi. "Landscaping and Integrating Digital Archaeology and Digital Epigraphy Resources: New Challenges and Future Opportunities. Introduction to the Special Section." *Archeologia e Calcolatori* 35, no. 1 (2024).

[3] Corbetta Piergiorgio. *La ricerca sociale: metodologia e tecniche: I paradigmi di riferimento*. Vol. I, Bologna: Il Mulino, 2014.

[4] Dewaele, Jean-Marc. "Online questionnaires." *The Palgrave handbook of applied linguistics research methodology*, 2018: 269-286.

[5] Frisina, Annalisa. *Focus Group, Una Guida Pratica*. Bologna: Il Mulino, 2010.

[6] Mancuso, Giacomo, and Antonio D'Eredità. "DHELO and BIDIAR: New Digital Resources for the H2IOSC Project." *Archeologia e Calcolatori* 35, no. 1 (2024).

[7] Nardi, Peter M. Doing survey research: A guide to quantitative methods. Routledge, 2018.

[8] Tashakkori, Abbas, and John W. Creswell. "The New Era of Mixed Methods." *Journal of Mixed Methods Research* 1, 2007: 3–7.

[9] European Strategy Forum on Research Infrastructures. Roadmap 2021. *Strategy Forum on Research Infrastructures*. ESFRi: 2021

[10] ESFRI White Paper. *Making Science Happen*. ESFRI, 2020.

[11] Buzzi, Carlo, Alessandro Cavalli, and Antonio De Lillo. *Rapporto giovani*. Vol. 568. il Mulino, 2007.

[12] Cavalli, Alessandro. *La gioventù tra disincanto e nuove speranze*. Bologna: Il Mulino, 2002.

[13] Rosina, Alessandro. *Il futuro non invecchia*. Vita e pensiero, Milano, 2018.