Synergies between CLARIN-IT and OPERAS-IT within H2IOSC: Monitoring Communities and Orchestrating Digital Services

Pietro Sichera ILIESI, CNR, Italy pietro.sichera@cnr.it
Monica Monachini ILC, CNR, Italy monica.monachini@cnr.it
Valeria Quochi ILC, CNR, Italy valeria.quochi@cnr.it
Nicola Giampietro ILIESI, CNR, Italy nicola.giampietro@cnr.it
Vittoria Fabiani ILIESI, CNR, Italy vittoria.fabiani@cnr.it
Roberta Bianca Luzietti ILC, CNR, Italy robertabianca.luzietti@cnr.it
Roberta Ottaviani ILC, CNR, Italy roberta.ottaviani@cnr.it
Daniele Melaccio ILC, CNR, Italy daniele.melaccio@cnr.it
Laura Baggiani ILIESI, CNR, Italy laura.baggiani@cnr.it

Abstract

This paper presents two components developed within the Humanities and Cultural Heritage Italian Open Science Cloud, H2IOSC, the Observatory and the Marketplace. Together, they support the SSH federated national cluster of Research Infrastructures by combining evidence-based community monitoring with an environment for discovering and combining digital resources and services. The Observatory adopts a mixed methods approach, integrating quantitative and qualitative instruments to document researchers' practices, needs, and expectations across partially overlapping domains. The Marketplace provides a catalogue of services and a workflow environment based on controlled ingestion, semantic normalisation, and interoperability with external sources. We also describe how the two components interact, linking community analysis with service provision. This interaction shows how CLARIN-IT and OPERAS-IT work together within H2IOSC to monitor uptake and inform service development.

1 Introduction

The Open Science paradigm has created new challenges and opportunities for the Social Sciences and Humanities (SSH) and Cultural Heritage (CH) disciplines. While the availability of digital tools and datasets is constantly increasing, fragmentation still characterises the research landscape for many domains. Researchers often struggle to navigate various independent repositories, incompatible data formats, and isolated service platforms, which hinders the full realization of the FAIR principles (Findable, Accessible, Interoperable, Reusable) (Wilkinson et al., 2016).

In this context, the Italian nodes of four major European Research Infrastructures (namely, CLARIN-IT, DARIAH-IT, E-RIHS.it, and OPERAS-IT) jointly contribute to the creation of a federated cluster of Research Infrastructures (RIs), aiming to overcome these barriers by fostering an open, federated digital ecosystem for SSH research communities in Italy, while maintaining solid connections with corresponding European counterparts within the European Open Science Cloud (EOSC): the Humanities and Cultural Heritage Open Science Cloud (H2IOSC hereafter, Degl'Innocenti et al., 2023).

One output of this initiative is the development of two key components, the H2IOSC Marketplace and its Observatory. The Marketplace serves as the main access point for the cluster. Unlike traditional static catalogues, it acts as an active orchestration layer allowing researchers to discover services from different providers and combine them into coherent workflows. The Observatory complements this function by monitoring community needs, service uptake, and user experiences through survey data, focus groups, and usage indicators. This establishes a feedback loop: usage data from the Marketplace informs the Observatory, which in turn provides evidence that can inform the development of the service portfolio offered by the infrastructures. By integrating these components, the project supports more coordinated access to research tools and services across communities that have often operated separately. This paper describes the architectural design and the functional implementation of the Marketplace and the Observatory, illustrating how they support a dynamic and collaborative SSH research environment.

The remainder of this paper is organised as follows: Section 2 contextualizes the project within the European framework of Research Infrastructure clustering and introduces the H2IOSC initiative. Section 3 describes the two main instruments devised for community data collection: the survey, implementing a quantitative approach, and the focus groups, embodying a qualitative approach. Section 4 describes the

design and technical implementation of the H2IOSC Observatory. Section 5 presents the architecture of the Marketplace, highlighting its orchestration capabilities. Section 6 demonstrates the strategic integration between the Observatory and the Marketplace, illustrating the feedback loop behind the ecosystem. Finally, Section 7 draws conclusions and outlines future directions for the Marketplace and Observatory.

2 Context and Related Works: From Isolation to Federation

For at least the past two decades, the European Commission has heavily invested in Research Infrastructures (RIs) for the Humanities, Social Sciences, and Cultural Heritage. These investments have been driven by policy priorities increasingly aligned with Open Science and the FAIR principles, aiming to transform fragmented resources into a coherent scholarly ecosystem (European Commission: Directorate General for Research and Innovation, 2025). In the past, digital resources in the SSH domain were developed in isolation, leading to data silos. To address this, last-decade(s) EU strategies encouraged the construction and spread of distributed research infrastructures as a means for fostering both the data sharing and adherence to open science and FAIR principles also in the SSH disciplines, that were lagging behind life- and hard-sciences.

This resulted in a proliferation of infrastructural initiatives both at EU and at national levels, that now run the risk of limiting sharing and hampering interoperability. In line with this trajectory, recent EU strategies encourage the aggregation of RIs into "clusters", aiming to optimize resources, avoid duplication of efforts, and foster even more interdisciplinary collaboration. A notable and successful example of this is the European Open Science Cloud (EOSC), which promotes the reuse and sharing not only of data but also of tools, methodologies, protocols, and workflows (Almeida et al., 2017). Following this supranational direction, several European countries have begun to operationalise this clustering model at the national level. Notable examples of integrated national nodes include, among others: France, with the *HumaNum* infrastructure[1] which aggregates services for digital humanities; Germany: with the *Text+* consortium, part of the NFDI initiative; Switzerland with *SSHOC-CH*[2], which brings together the Swiss national nodes of CLARIN, DARIAH and OPERAS.

The Italian strategic response to this trend is the H2IOSC initiative[3], which aims to answer the need for a coordinated federation of infrastructures (Degl'Innocenti et al., 2023). It establishes a cross-disciplinary infrastructure cluster designed to support end-user support, data stewardship, resource curation, and training activities for the national scholarly community. Within this federated model, CLARIN-IT plays a pivotal role focusing on linguistic data management, standardization, and service integration. Its contribution ensures that the Italian ecosystem remains fully aligned with the European CLARIN standards and related EOSC initiatives. By integrating CLARIN-IT's specific capabilities with the broader H2IOSC architecture, the project aims to implement a "system of systems" where specialized vertical nodes contribute to a horizontal, interoperable marketplace.

Similarly, OPERAS-IT contributes the perspective and practices of open scholarly communication, covering areas such as peer review, open access publishing, and dissemination of research outputs. Within H2IOSC, OPERAS-IT is responsible for the development of the Marketplace and has carried out pilot implementations, while also coordinating the qualitative community monitoring activities described in Section 3.2. The Observatory itself is coordinated by CLARIN-IT under WP2, ensuring that the analytical framework remains aligned with CLARIN's metadata standards and community practices.

This integration is not unidirectional. While H2IOSC benefits from CLARIN ERIC's established standards and discovery services, it also contributes back to the broader ecosystem, for instance by cataloguing internationally maintained CLARIN services alongside nationally developed tools in a shared Marketplace, and by formalising implicit semantic distinctions from CLARIN's metadata practices into a reusable domain ontology (Melaccio et al., 2025). The relationship with CLARIN ERIC is further discussed in Section 7.

[1] https://www.huma-num.fr/

[2] https://sshoc.ch/

[3] https://www.h2iosc.cnr.it/

3 Mixed-Methods Approach to Community Monitoring

This section outlines the mixed-methods strategy adopted to investigate the needs, practices, and expectations of the communities served by the four H2IOSC Research Infrastructures. It combines a quantitative survey, designed to collect comparable evidence at federation level, with a qualitative focus-group campaign aimed at validating and enriching survey findings and capturing emergent themes. Together, these instruments provide the empirical basis for the Observatory's monitoring activities and inform the evolution of the service portfolio and workflows exposed through the H2IOSC Marketplace.

3.1 The quantitative approach

The surveying activity was jointly designed by the four H2IOSC Research Infrastructures (CLARIN-IT, DARIAH-IT, E-RIHS.it, and OPERAS-IT) to collect comparable evidence on communities, practices, and needs across partially overlapping domains. A single questionnaire was intentionally adopted to (i) reduce respondent burden and community fatigue, and (ii) enable an integrated analysis of requirements and expectations at federation level.

The questionnaire was collaboratively drafted to address both shared and RI-specific information needs, covering: respondent profiles, awareness and use of RIs, training needs, research outputs, and the description of digital resources, tools, and services used or produced by participants. Before wide release, the instrument underwent a two-step validation process: an initial pilot with a small group of researchers and professors representative of the target domains, followed by a second review with members of key disciplinary associations. Feedback from these phases informed a revised version of the questionnaire, structured into two parts (profile/needs and resources/services), which was then disseminated broadly through mailing lists, social media, RI and project websites, conference presentations, and outreach events (Luzietti et al., 2024, 2025; Monachini, Quochi, Luzietti, Moscati, et al., 2025; Monachini, Quochi, Luzietti, Spadi, et al., 2025).

To encourage participation while respecting privacy, personal data collection was minimized and limited to a contact email address used exclusively for follow-up communication and engagement activities.

3.2 The qualitative approach

The questionnaire distribution phase was followed by a qualitative mapping exercise to:

- validate the data collected through the questionnaire;
- identify new concepts not included in the questionnaire;
- foster a sense of belonging.

Four focus groups were conducted, coordinated by the OPERAS-IT node. Twenty-four researchers from the four communities were involved, including five archaeologists, four philologists, eight linguists, and seven philosophers. The meetings were video-recorded and automatically transcribed using Microsoft Teams. The researchers corrected the transcripts and processed them using the MAXQDA26 tool to produce a textual corpus of 59,724 tokens. The corpus then underwent a preliminary analysis to assess lexical richness and vocabulary diversity in the language of the researchers' involved. A sample of unique words was selected from the corpus (The sample was stratified and weighted according to the following variables: age, biological characteristics, the scientific fields of the four IRs). There are 930 word types in the corpus with a Type-Token Ratio (TTR) of 0.0156[4].

Knowledge of and use of research infrastructures (RIs) were proposed as discussion prompts in the focus groups. Most participants stated that they were familiar with RIs and used them frequently. To specifically examine the researchers' opinions, the corpus was analyzed a second time. Stop words, filler words, and idiomatic repetitions were removed. One-on-one exchanges between researchers and

[4]Type-Token Ratio (TTR) is a well-known quantitative measure used in linguistic disciplines to assess lexical richness. It is calculated by dividing the number of unique words (or word"Types") by the total number of their occurrences (or "Tokens") and multiplying by 100, Types /Tokens ×100. In MAXQDA this measure excludes stop words (common and highly frequent words such as *the*, *of* and *that*) to avoid noise during topical analysis of the data.

participants that were unrelated to the research topic were also eliminated. As a result, the reference corpus was reduced to 11,177 tokens, with a TTR of 0.0042. The word cloud in Fig. 1 was constructed from this reference corpus.

Figure 1: Word Cloud referring to the four focus groups

This image illustrates how researchers at the four RIs can exploit the opportunity presented by tools such as the H2IOSC infrastructure.

The tokens in the word cloud, whose absolute and relative frequencies are partially shown in Fig. 2, were converted into codes using the MAXQDA26 software.

Words	Frequencies	%
infrastructure	234	2,09
research	220	1,97
tools	121	1,08
fair	60	0,54
digital	57	0,51
open access	41	0,37
linguistics	37	0,33
publication	18	0,16
computational	17	0,15
Clarin	**13**	**0,12**
interoperability	8	0,07

Figure 2: Table of absolute frequencies and percentages of the words included in the analysis

The resulting codes were visualized in a graph showing the relationships between them, creating a “code map.” (Fig. 3, Area 1).

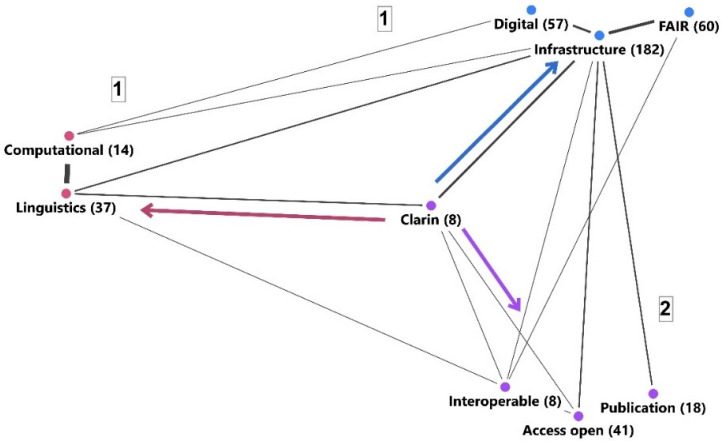

Figure 3: Concept map of stimuli on knowledge, use and requests for the CLARIN infrastructure

Using the "code map" tool, the relationships between all the codes and the "CLARIN" code were analyzed (Fig. 3). Regarding the linguistic domain, all eight researchers reported being aware of research infrastructures, and among these, CLARIN is the most widely used. Researchers associated the name CLARIN with the linguistic, computational, digital, infrastructure, and FAIR principles codes.

These code co-occurrences were then examined through MAXQDA's concept mapping functionality, which visualises proximity and frequency-based relationships between codes. The resulting map (Fig. 3, Area 2) revealed three convergent thematic clusters: a shared emphasis on resource availability and interoperability across RIs, a concern for the visibility and discoverability of research outputs, and a strong orientation towards Open Science compliance. These clusters were further validated by cross-referencing them with responses from the quantitative survey, confirming their representativeness across the broader community. The three expectations reported below correspond directly to these thematic clusters.

In addition, the expectations identified regarding the confederate infrastructure (Fig. 3, area 2) are:

- **Expansion of Digital Heritage:** H2IOSC should increase the availability of corpora and structured educational materials as resources that comply with FAIR principles. Linguists expect to have the resources of the four member research institutions immediately available in a single Marketplace.
- **Dissemination and Impact:** Leveraging the centralized infrastructure to enhance the visibility of linguistic research outputs, facilitating their discovery on a multidisciplinary scale. Users hope that H2IOSC will act as a catalyst for the outputs (publications) currently hosted in CLARIN.
- **Progress towards Open Science:** Integration into the H2IOSC Marketplace is seen as a critical opportunity to increase the CLARIN node's compliance with Open Science protocols.

The analysis of lexical relationships places CLARIN at the center (Fig. 3), highlighting that it is a digital linguistics infrastructure already aligned with open science and open access dissemination. The inclusion of linguists' needs in the H2IOSC Marketplace contributes to the development of a more inclusive research environment and may help CLARIN, as well as the other RIs, address interdisciplinary requirements more effectively. Our results suggest that infrastructure federation is not only a technical process, but also has important organizational and human dimensions. It can be informed by community feedback and, in turn, can inform methodological developments in related disciplines. In particular, the Observatory is the dedicated instrument through which this evidence-based community feedback, collected periodically through qualitative and quantitative methods, is gathered and returned to the communities.

4 The H2IOSC Observatory: Monitoring the Evolving Landscape of Digital Services

The H2IOSC Observatory is designed as a long-term instrument for monitoring the development and socio-economic impact of the four Research Infrastructures (RIs) of the H2IOSC federation. It provides

a continuous framework for analyzing the evolving landscape of digital resources, tools, and services in the domains of Linguistics, Digital Humanities, and Cultural Heritage, with particular attention to how these infrastructures evolve over time in relation to both community practices and infrastructural developments.

Originally designed as an internal instrument for monitoring communities and supporting evidence-based adjustments of services and priorities, the Observatory was redefined as a stable, publicly accessible online service. This evolution reflects its broader function as a shared environment for documenting researcher behaviors, practices, and needs across the H2IOSC communities. By displaying high-quality empirical data on researcher behaviors, practices, and needs, the Observatory establishes a data-driven feedback loop between community activities and the digital service environment of the H2IOSC ecosystem. The Observatory website is not intended as a space for active data collection, but rather as a curated and publicly accessible environment for the structured dissemination and interpretation of evidence produced within H2IOSC. This design choice is reflected in the organization of dedicated sections for different methodological instruments, such as surveys and focus groups, which are exposed through clearly separated thematic tabs (see Fig. 4). This structure supports transparency, reuse, and long-term comparability of results, while also facilitating navigation across heterogeneous forms of evidence.

Figure 4: Thematic sections of the H2IOSC Observatory dedicated to survey and focus-group evidence, exemplifying the platform's structured organisation of heterogeneous monitoring data.

4.1 Community-oriented monitoring

The Observatory operationalizes the mixed-methods approach described in Section 3, integrating quantitative and qualitative instruments such as surveys, focus groups, interviews, and landscape mapping activities (Tashakkori & Creswell, 2007). These instruments support the systematic investigation of the behaviors, needs, and practices of current and prospective RI users, providing an empirical basis for the analysis of both declared expectations and emerging research dynamics.

Alongside its analytical function, the Observatory also supports community engagement and participatory monitoring. By involving researchers, institutional stakeholders, and prospective users in the assessment and discussion of services, priorities, and gaps, it supports the co-definition of offerings, and

the development of infrastructures that respond to documented user needs and practices. In this sense, the Observatory bridges community-oriented analysis and infrastructural decision-making, helping to keep service development grounded in empirical evidence (Luzietti et al., 2024).

4.2 Platform implementation and content organisation

From an implementation perspective, the Observatory has been developed as a modular web platform, fully aligned with the approved UX/UI design and with the shared visual and functional ecosystem of H2IOSC infrastructures. The platform adopts a mobile-first approach and devotes specific attention to accessibility and usability requirements, in line with WCAG 2.1 principles[5], ensuring inclusive access and compliance with public-sector standards.

The organisation of contents reflects the core instruments of the Observatory and translates methodological choices into navigable sections. These include survey results and analytical summaries, qualitative outcomes of focus groups and interviews, and landscape mapping activities outlining available tools, datasets, and services. In addition, monitored projects are integrated through the Digital Platform for Cultural Heritage Research (DHeLO) platform via dedicated APIs, enabling the Observatory to connect empirical monitoring with a broader representation of infrastructural initiatives across the federation.

Selected usage indicators originating from the Marketplace are incorporated into the Observatory through dedicated APIs, complementing survey-based evidence with behavioural traces from real interactions with the service catalogue (see also Section 6).

4.3 Role within the H2IOSC ecosystem

The long-term sustainability of the Observatory is ensured through both organisational and technical measures. The Observatory Scientific Committee is responsible for guaranteeing methodological rigour, analytical depth, and the quality of published contributions, acting as a supervisory and theoretical reference point in alignment with Open Science principles (Rossi et al., 2022). At the same time, the platform includes a dedicated back-office environment based on a modular Content Management System and a Digital Asset Management component, enabling structured editorial updates and progressive enrichment of contents without continuous technical intervention.

Within the broader H2IOSC ecosystem, the Observatory operates as both a reflective research environment and a strategic infrastructural component. By consolidating heterogeneous evidence into a coherent monitoring framework, it supports adaptive governance of the federation and prepares the ground for the service orchestration mechanisms implemented in the H2IOSC Marketplace. In this perspective, the Observatory provides the interpretative and analytical layer of the ecosystem, while the Marketplace constitutes its operational interface, together enabling an evidence-based and continuously evolving research infrastructure.

5 The H2IOSC Marketplace as an Operational Infrastructure and Source of Evidence

The H2IOSC Marketplace is conceived as a platform for the consultation and management of resources and services exposed by the Italian nodes of the federated infrastructures, clearly separating functionalities aimed at end users from those devoted to governance and curation. It combines controlled metadata ingestion, semantic normalisation, indexing, and presentation, and is designed for interoperability with external sources and aggregators.

5.1 Three-layer architecture

From an architectural standpoint, the Marketplace adopts a three-component structure: Back-End, Back-Office, and Front-End. The Back-End constitutes the application core and concentrates the exposure of APIs; the Back-Office is the operational environment for authorized users (administrators, curators, editors), responsible for curatorial management and configuration; the Front-End is the public interface for consultation, search, and catalog navigation. This separation is not an implementation detail, but an infrastructural constraint: it allows governance functions to be isolated from user-facing access paths and,

[5] https://www.w3.org/TR/WCAG21/

above all, enables programmatic access to Marketplace data through APIs without depending on the UI. The technical design includes infrastructural components dedicated to persistence and search, employing technologies such as SOLR, Redis, and MongoDB, as well as integration components to interact with external systems.

5.2 Identity management

Identity and authentication management is based on Keycloak. This aspect is relevant for the relationship with the Observatory for two reasons: on the one hand, it enables the separation of public access for consultation from the use of services that require credentials; on the other hand, it creates the technical conditions to distinguish types of interaction and use (anonymous vs. authenticated) in compliance with access policies.

5.3 A data model as the basis for observability

The descriptive uniformity of the catalogue is ensured by a data model organised around the *Item* entity, intended as the main unit of cataloguing for heterogeneous types (for example, tools, datasets, publications, projects, and services). Items are described through a structured set of fields, properties, and relations; the mandatory status and formats of fields are defined to ensure consistency and interoperability. This choice has a direct consequence for the relationship with the Observatory: it makes both the composition of the catalogue and its temporal evolution observable and comparable in a systematic way, because variability in descriptions is constrained by schemas and controlled vocabularies. In particular, lexical and semantic normalisation through managed dictionaries and vocabularies enables reliable metrics (e.g., by Item type, disciplinary area, or access mode) that the Observatory can reuse directly.

5.4 Feeding the Marketplace

The feeding of the Marketplace is designed to combine automated import and controlled management, with the aim of maintaining alignment and traceability with respect to the original sources. For the acquisition and integration of metadata and Items, mechanisms based on the OAI-PMH protocol and on REST APIs are envisaged, with a dedicated endpoint `/api/items/harvest` and `Bearer`-type authentication, and with a data schema aligned with the OAI-PMH model. Beyond enabling ingestion, this approach defines a technical contract that makes it possible to distinguish what is curated locally from what is synchronized with external providers.

On the Back-Office side, harvesting mechanisms are operationalized through a harvesting console that allows users to create, modify, and schedule jobs, monitor their executions, and download outcome reports. The orientation is explicitly curatorial: the Marketplace does not merely "collect", but provides tools to govern the ingestion process, including filters, sorting, and monitoring of the status of jobs and processed items.

In parallel with harvesting, the design envisages providing APIs to export Items towards other marketplaces or aggregators, through token-authorized access and an OpenAPI description. Providing is therefore part of the Marketplace's federated positioning: not only ingestion, but also the ability to "expose" contents in exchange formats and to support scenarios in which third-party systems select and reuse the catalog contents.

5.5 Descriptive workflows and executable workflows

A qualifying element of the H2IOSC Marketplace is workflow management as a orchestration of services. The implementation includes descriptive workflows, which represent and document how different services can be combined, and executable workflows (Sichera et al., 2024), in which the orchestration is designed and guided at run-time, subject to compatibility constraints between output/input pairs and adherence to specifications such as OpenAPI 3. In this setting, the Marketplace does not merely describe resources, but encodes operational relationships among services that may also come from different sources.

In the technical design, the construction and management of executable workflows and integration is associated with the use of WSO2 API Manager (for API management and exposure) and WSO2 Micro

Integrator (for orchestration), coupled with the management of manifest files for service description.

At the time of writing, the harvesting pipeline for CLARIN-IT services has been fully developed and validated. The Marketplace catalog already includes, or will include at launch, a range of services provided by CLARIN-IT within H2IOSC — such as an eScriptorium instance for handwritten text recognition, a deposit service, and a training platform — alongside internationally maintained CLARIN services such as UDPipe and resources such as ParlaMint. This combination of national and international assets illustrates the federated nature of the catalog and its alignment with the broader CLARIN ecosystem. An example of cross-infrastructure orchestration is provided in Section 6.4.

6 The Interaction between Marketplace and Observatory as an Infrastructural Mechanism

The relationship between the Marketplace and the Observatory, in the H2IOSC design, is explicitly conceived as a continuous synergy between two complementary nodes of the same federated research infrastructure. The Marketplace aggregates and exposes digital assets and services from the participating infrastructures, while the Observatory analyses the use and evolution of such resources over time, with the aim of supporting adaptive management of the ecosystem and aligning the service offer with emerging needs of the community. This interaction is not incidental, but intentionally designed as an infrastructural mechanism in which observation, interpretation, and service provisioning are functionally decoupled yet operationally interconnected.

6.1 *Ex ante* and ongoing evidence of use

The Observatory inherits and systematises evidence built with replicable mixed-methods instruments (questionnaires, interviews, focus groups, and semi-automated collection protocols), producing a knowledge base on the needs and practices of the reference communities. The opening and use of the Marketplace add a further dimension: the possibility of correlating what users declare or discuss with what they actually explore and use in the catalogue, making expectations and observed behaviour comparable. In this perspective, the Marketplace becomes a generator of traces and metrics that transform observation from a periodic snapshot into a continuous process. This ongoing monitoring is supported by usage statistics and interaction indicators, which provide continuous evidence of service uptake and navigation dynamics within the platform (see Fig. 5). For instance, the Observatory dashboard currently displays the distribution of catalogue Items by type, the temporal evolution of published resources, registered user counts over time, and saved search activity, all drawn from the Marketplace via dedicated APIs. These indicators, though preliminary, already enable the identification of trends in community engagement and catalogue growth. In this sense, the Marketplace enables a form of infrastructural observability, whereby user interactions, navigation paths, and service combinations can be analysed as structured signals rather than isolated events.

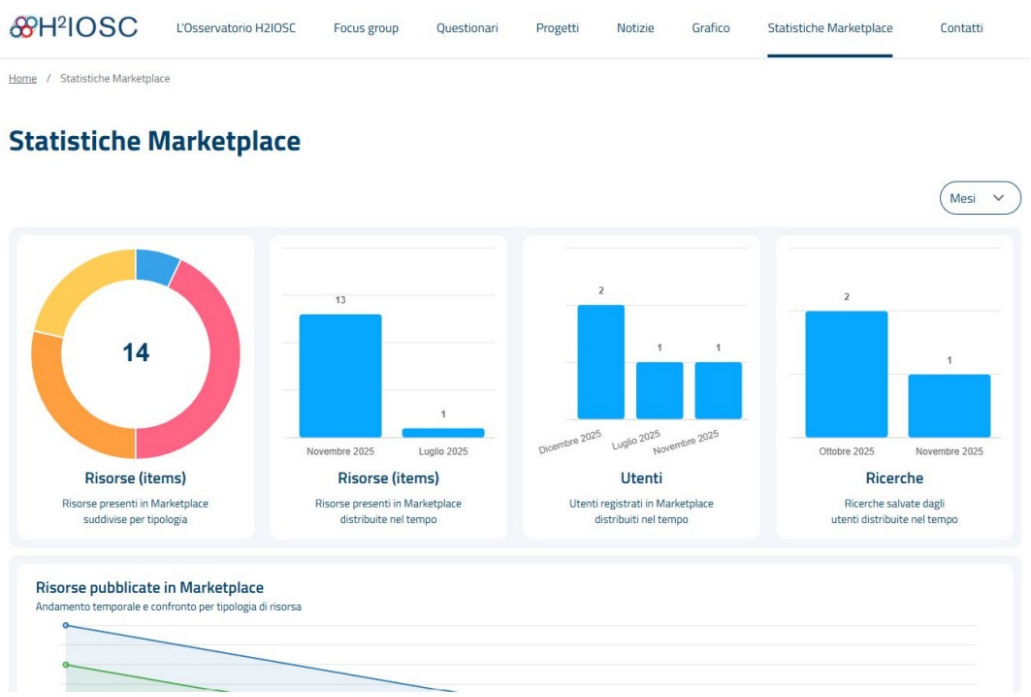

Figure 5: Usage statistics from the H2IOSC Marketplace as presented in the Observatory dashboard for monitoring catalogue growth and user engagement.

6.2 Service provision and analysis

The Observatory can be considered a modular web application developed "in synergy" with the Marketplace, with a bidirectional flow of information: the Marketplace aggregates and exposes resources, the Observatory analyses their use and evolution and returns insights that guide subsequent developments.

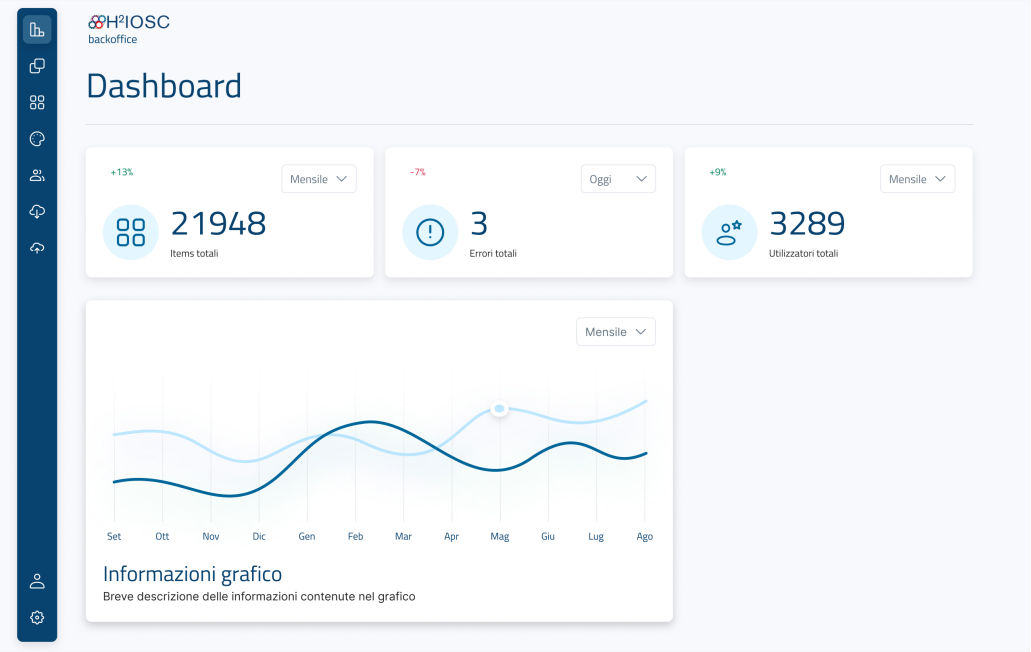

Figure 6: Analytical dashboard based on Marketplace data, used to observe user behaviour, resource uptake, and patterns of interaction across the platform.

The integration of the two platforms aims to enable monitoring of resource ingestion, user behavior, and service uptake. This defines a minimum set of requirements for interoperability between the two

platforms: availability of structured data on the catalog (composition and changes), availability of usage metrics (with adequate granularity and access rules), and semantic consistency to enable meaningful aggregations. The "Item-centric" model and the controlled vocabularies of the Marketplace thus become a technical prerequisite for observational analysis, because they stabilize the dimensions along which to measure and compare. This separation of concerns is technically supported by stable data models, controlled vocabularies, and API-based exchanges, which allow analytical components to evolve independently from service provisioning layers.

6.3 Catalogue and process data

Observational interaction concerns not only which Items exist in the catalog, but also how they reach it and how they are maintained. Harvesting and providing mechanisms, as well as the management of ingestion jobs and operational reports, produce process data that can be interpreted as indicators of the maturity and reliability of the catalog feeding chains. In particular, the ability to distinguish between functionalities implemented and verifiable in staging and functionalities implemented but not verifiable end-to-end due to external prerequisites helps to avoid improper readings of the metrics, and suggests that the Observatory should incorporate metadata about the "state" of the platform and of the flows, in addition to usage metrics. Such information is particularly relevant for governance purposes, as it enables infrastructure managers to distinguish between structural limitations, integration bottlenecks, and actual patterns of community uptake. As a result, observability is not retrofitted onto the system, but emerges from design choices that constrain variability and make catalog evolution analytically tractable.

6.4 From observational feedback to action

The operational value of the Observatory–Marketplace cycle lies in transforming insights into verifiable actions. A first trajectory concerns the evolution of the catalog: evidence on emerging needs and domains can guide scouting and integration activities through harvesting from external sources and through curatorial processes in the Back-Office. The Marketplace design explicitly requires full interoperability with European catalogs (including EOSC and SSHOC), which, within the observational cycle, can also be read as a strategy to reduce the time between identifying a need and making relevant resources available, by exploiting federation and exchange channels.

A concrete example of how community evidence informs service design is provided by the workflow orchestration capabilities of the Marketplace. Focus group findings (Section 3.2) highlighted that researchers across disciplines regard interoperability and the ability to combine services from different infrastructures as a key expectation. This evidence directly informed the development of executable cross-infrastructure workflows. For instance, a workflow has been designed that takes a IIIF manifest served by E-RIHS.it as input, sends it to the CLARIN-IT eScriptorium instance for handwritten text recognition, and automatically imports the resulting transcription into a TEI Publisher instance for publication and annotation. This pipeline illustrates how the Marketplace translates declared community needs into operational service compositions that span multiple federated infrastructures.

A second trajectory concerns the quality of the presented information: the Back-Office includes a Content Management System that enables the management of editorial pages and the creation of content (simple pages, collections, blog posts), complementing the technical description of Items with layers of contextualisation and usage-oriented guidance. In a data-driven cycle, the Observatory can help identify areas of the website and catalogue that require greater clarity or informational completeness, intervening not only in the addition of new Items but also in their presentation and in the construction of editorial paths consistent with observed usage practices. In this way, observational evidence is translated into concrete curatorial, interoperability, and editorial decisions, closing the loop between monitoring, interpretation, and infrastructural action. Taken together, the Observatory–Marketplace interaction exemplifies how analytical monitoring and service orchestration can be combined into a single adaptive mechanism within federated research infrastructures.

7 Conclusions and Future Work

In the near future, the Observatory will be further developed by cyclically running the defined monitoring strategies as well as extending current data collection methods with additional monitoring strategies and impact assessment procedures. The aim is to improve the identification of changes in community needs and service usage patterns, in order to inform decisions about the extension of available tools and possible technical developments.

In this context, requests for additional data from the Marketplace, such as performance indicators or engagement metrics, may offer useful insights for refining both the Observatory's analytical methods and its overall development.

At the same time, the Marketplace may expand its functionality and interoperability on the basis of feedback emerging from the Observatory's activities. This may include the introduction of new modules for orchestrating complex workflows and procedures for integrating third-party services and external infrastructures. The effectiveness of these enhancements will depend on continued coordination between the two components: the Observatory will provide ongoing monitoring and identify areas for improvement; the Marketplace will supply the data and resources required to support incremental, modular updates and alignment with user needs.

In this context, the partnership between infrastructures will continue to develop through iterative exchanges between the Observatory and the Marketplace. This interaction supports an integrated Open Science ecosystem, where coordination between research infrastructures contributes to the development of tools, broadens user engagement and increases the use of available services, thereby improving the uptake of open science digital approaches in humanities, social and cultural heritage sciences.

In addition, the work carried out within H2IOSC has also contributed back to the CLARIN ecosystem. The ontology-based mediation approach developed for aligning CLARIN-IT metadata with the H2IOSC Marketplace enables interoperability between heterogeneous schemas while preserving semantic consistency. This approach represents a concrete methodological contribution that can be reused by other CLARIN national nodes facing similar integration challenges, and may support the broader harmonisation of services and metadata within CLARIN ERIC and EOSC-related environments (Melaccio et al., 2025).

Finally, H2IOSC is conceived in continuity with the EOSC vision, as a federation of Italian nodes of European ESFRI RIs already involved in EOSC. CNR is a member of the EOSC Association, and several H2IOSC team members have participated in EOSC-related and SSHOC initiatives. H2IOSC draws on these experiences. In particular, the Marketplace and the Observatory are informed by the EOSC and SSH Open Marketplace, while adapting their functions to national and domain-specific requirements. The Observatory is intended to complement the EOSC Observatory by providing more detailed, community-driven information on research practices and service usage at the community level. These developments point to a layered model of integration, in which national infrastructures contribute specialized components that can be connected and may enrich their European counterparts. In this perspective, CLARIN-IT within H2IOSC remains aligned with the mission of CLARIN ERIC, while OPERAS-IT continues its participation and strengthens its role within OPERAS, ensuring that both infrastructures contribute to and benefit from broader open science initiatives at national and European levels.

Acknowledgments

This work is supported by the H2IOSC Project - Humanities and Cultural Heritage Italian Open Science Cloud funded by the European Union – NextGenerationEU – NRRP M4C2 - Project code IR0000029 - https://www.h2iosc.cnr.it/

References

Almeida, A. V. d., Borges, M. M., & Roque, L. (2017). The european open science cloud: A new challenge for europe. *Proceedings of the 5th International Conference on Technological Ecosystems for Enhancing Multiculturality*. https://doi.org/10.1145/3144826.3145382

Degl'Innocenti, E., Monachini, M., Bucciero, A., Pasini, E., Fanini, B., & Frontini, F. (2023). H2iosc: Humanities and heritage open science cloud. *La Memoria Digitale: Forme Del Testo e Organizzazione Della Conoscenza. Atti Del XII Convegno Annuale AIUCD*, 63–64.

European Commission: Directorate General for Research and Innovation. (2025). The european strategy on research and technology infrastructures [Accessed: 2025-10-22].

Luzietti, R. B., Spadi, A., Giampietro, N., Giacomo, M., Caravale, A., D'Eredità, A., Caradonna, M., Moscati, P., Quochi, V., Monachini, M., & Degl'Innocenti, E. (2024). Digital humanities and heritage science: Moving from landscaping to a dynamic research observatory in an open science cloud. In A. Di Silvestro & D. Spampinato (Eds.), *Me.te. digitali. mediterraneo in rete tra testi e contesti, proceedings del xiii convegno annuale aiucd*. AIUCD Associazione per l'Informatica Umanistica e la Cultura Digitale.

Luzietti, R. B., Spadi, A., Giampietro, N., Mancuso, G., Caravale, A., D'Eredità, A., Caradonna, M., Moscati, P., Quochi, V., Monachini, M., & et al. (2025). Digital humanities and heritage science: Moving from landscaping to a dynamic research observatory in an open science cloud. *Umanistica Digitale*, *9*(20), 419–439. https://doi.org/10.6092/issn.2532-8816/21226

Melaccio, D., Boschetti, F., & Monachini, M. (2025). Interfacing CLARIN with H2IOSC: Metadata Interoperability through Ontology-based Mediation. *Proceedings of the CLARIN Annual Conference 2025*. https://doi.org/10.5281/zenodo.17357825

Monachini, M., Quochi, V., Luzietti, R. B., Moscati, P., Caravale, A., Chirivì, A., Degl'Innocenti, E., Spadi, A., & Caradonna, M. (2025, January). *D2.1 first report on the H2IOSC landscapes* (tech. rep.). Zenodo. https://doi.org/10.5281/zenodo.14680021

Monachini, M., Quochi, V., Luzietti, R. B., Spadi, A., Mancuso, G., D'Eredità, A., Caravale, A., Giampietro, N., Caradonna, M., & Melaccio, D. (2025, October). *D2.2 updated report on the H2IOSC landscapes* (tech. rep.). Zenodo. https://doi.org/10.5281/zenodo.17737101

Rossi, G., Caso, R., Castelli, D., & Giglia, E. (2022). National plan for open science.

Sichera, P., Marras, C., & Pasini, E. (2024, November). Orchestrazione api per workflow applicativi nell'ambito delle digital humanities. https://doi.org/10.5281/zenodo.14187534

Tashakkori, A., & Creswell, J. W. (2007). The new era of mixed methods. *Journal of mixed methods research*, *1*(1), 3–7.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The fair guiding principles for scientific data management and stewardship. *Scientific Data*, *3*(1). https://doi.org/10.1038/sdata.2016.18