



CALL	<p>Ministry of University and Research (MUR), Directorate-General for Internationalisation and Communication, Notice D.D. 3264 dated 28/12/2021.</p> <p>Mission 4 "Education and Research" - Component 2 "From research to enterprise" - Investment Line 3.1 "Fund for the realisation of an integrated system of research and innovation infrastructure", Action 3.1.1 "Creation of new IRs or strengthening of existing ones that contribute to Horizon Europe's Scientific Excellence objectives and networking"</p>
PROJECT TITLE	Humanities and cultural Heritage Open Science Cloud
PROJECT ACRONYM	H2IOSC
PROJECT ID	IR0000029
PROJECT WEBSITE	https://www.h2iosc.cnr.it
CUP	B63C22000730005
WP	WP8 - Training, Capacity Building, Engagement
OU	OU 8 - CNR-ILC
DELIVERABLE NUMBER	8.1
DELIVERABLE TITLE	Training Strategy
EXPECTED DELIVERY DATE	31.08.2023
ACTUAL DELIVERY DATE	31.10.2023

AUTHORS	<p>Laura Benassi (CNR-INO), Federico Boschetti (CNR-ILC), Leonardo Canova (CNR-OVI), Antonina Chaban (CNR-INO), Emiliano Degl'Innocenti (CNR-OVI), Carmen Di Meo (CNR-OVI), Francesca Frontini (CNR-ILC), Monica Monachini (CNR-ILC), Roberta Ottaviani (CNR-ILC), Leonardo Pica Ciamarra (CNR-ISPF), Giulia Pedonese (CNR-ILC), Pietro Restaneo (CNR-ILIESI), Alessia Scognamiglio (CNR-ISPF), Alessia Spadi (CNR-OVI), Jana Striova (CNR-INO)</p>
<p>This project is funded by the European Union - NextGenerationEU. The views and opinions expressed are only those of the authors and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them</p>	

Abstract

In this document we describe the state of the art of the four participating infrastructures in terms of training and define the training strategy of the project H2IOSC in terms of overall vision, and along the following 3 lines:

- Building the H2IOSC the training infrastructure with a training portal and a depositing service for training materials (technical and functional requirements are described for both)
- Creating an offer of common H2IOSC training materials, aimed at facilitating the use of the H2IOSC marketplace and pilots
- Strengthening the disciplinary offer for the four participating infrastructures

This is a living document, evolving during the life of the project with the natural upgrade of the awareness of the needs expressed by the various groups interested (users already in the community and identified potential users).

Table of Contents

<u>LIST OF ACRONYMS.....</u>	<u>6</u>
<u>LIST OF FIGURES.....</u>	<u>8</u>
<u>1. CONTEXT AND STATE OF THE ART</u>	<u>9</u>
1.2 TRAINING, FAIR & OPEN SCIENCE.....	10
1.2.1 THE ESFRI PERSPECTIVE ON TRAINING	10
1.2.2 TRAINING IN EOSC AND SSHOC	12
1.2.3 FAIR TRAINING MATERIAL	13
1.3 CURRENT OFFER BY PARTICIPATING RIS: TRAINING AND OUTREACH ACTIVITIES.....	14
1.3.1 CLARIN ERIC & CLARIN-IT.....	14
1.3.2 DARIAH-ERIC & DARIAH-IT	19
1.3.3 E-RIHS	21
1.3.4 OPERAS.....	22
1.4 NEEDS OF THE H2IOSC COMMUNITIES - LINK TO WP2	23
1.5 TRAINING IN THE FUTURE H2IOSC MARKETPLACE - LINK WITH WP5	25
1.6 TRAINING FOR COMMUNITY PILOTS - LINKS WITH WP7	25
<u>2. H2IOSC TRAINING AND OUTREACH PLAN</u>	<u>27</u>
2.1 APPROACH AND LIMITATIONS.....	27
2.2 METHODOLOGY	29
2.3 TRAINING INFRASTRUCTURE	30
2.3.1 VIRTUAL TRAINING ENVIRONMENT	30
2.3.2 FAIR AND INTEROPERABLE TRAINING MATERIALS PUBLICATION PLATFORM	31
2.3.3 OTHER TRAINING EQUIPMENT	33
2.4 TRAINING AND UPSKILLING THE H2IOSC USERS.....	33
2.4.1 DEVELOPMENT OF JOINT MODULES AND INITIATIVES.....	33
2.4.2 TEACH CLARIN, TEACH WITH CLARIN	34
2.4.3 TEACH DARIAH, TEACH WITH DARIAH.....	37
2.4.4 TEACH E-RIHS, TEACH WITH E-RIHS.....	38
2.4.5 TEACH OPERAS, TEACH WITH OPERAS	40
2.5 INTERNAL TRAINING & UPSKILLING	41
2.5.1 COMMON TRAINING NEEDS.....	41
2.5.2 CLARIN	42
2.5.3 DARIAH	42
2.5.4 E-RIHS	42
2.5.5 OPERAS.....	43
<u>3. MONITORING TRAINING AND OUTREACH ACTIVITIES</u>	<u>44</u>

4. LONG TERM PERSPECTIVES AND CONCLUSION	44
REFERENCES	46
ANNEXES	48

LIST OF ACRONYMS

CLARIN-IT	Nodo italiano di CLARIN - Common Language Resources and Technology Infrastructure
DARIAH-IT	Nodo italiano dell'infrastruttura di ricerca europea DARIAH (Digital Research Infrastructure for the Arts and Humanities)
EOSC	European Open Science Cloud
ERA	European Research Area
E-RIHS.it	Nodo italiano dell'infrastruttura di ricerca europea E-RIHS sull'Heritage Science
ERIC	European Research Infrastructures Consortium
ESFRI	European Strategy Forum on Research Infrastructures
FAIR	Findable, Accessible, Interoperable and Reusable
H2IOSC	Humanities and Heritage Open Science Cloud
KPIs	Key Performance Indicators
OPERAS-IT	Nodo italiano dell'infrastruttura europea OPERAS - Open scholarly communication in the european research area for social sciences and humanities
RI	Research Infrastructure
SSHOC	Social Sciences & Humanities Open Cloud

LIST OF FIGURES

Figure 1	H2IOSC work packages overview	Section 1.6
Figure 2	H2IOSC training in a nutshell	Section 2.1
Figure 3	Integration of the FAIR and Interoperable Training Materials Publication Platform with the H2IOSC marketplace	Section 2.3.1
Figure 4	The UPSKILLS modules, to be further developed and adapted to H2IOSC	Section 2.4.2

1. CONTEXT AND STATE OF THE ART

The H2IOSC project aims at creating a national infrastructure for the Humanities and heritage in Italy, federating the existing national nodes of the four European infrastructures of the domain, CLARIN, DARIAH, E-RIHS and OPERAS.

Work Package 8 of the H2IOSC project is dedicated to “**Training, Capacity Building, Engagement**”, and aims to:

1. **empower user communities** with interdisciplinary as well as domain specific FAIR and Open Science skills and competences as well as training them on what disciplinary infrastructures can offer them at the national and international level. This will enable them **to better use the services of the RIs national nodes, of the ERICs and of the national H2IOSC marketplace**, as well as become capable of **correctly managing the lifecycle of their data**
2. **train a pool of new professionals** - data stewards, data curators, trainers – that can support and train future generations of scholars and professionals on **how to integrate RIs in their methods and practices**, thus *fostering a change of mentality* in research and possibly industry

Our target:

1. Users already identified:
 - **established researcher**, who need to correctly create a data management plan for their projects, and require the competence to manage their tools and data
 - **researchers** and **technologists**, developing datasets and tools and willing to share them under the aegis of Open Science
 - **lecturers** who can benefit from the integration of infrastructural resources in their curricula
 - **higher degree students** (e.g. Master and PhD students) and **early career researchers** who can become actors of change and fully integrate infrastructures in their research practices and projects, developing a data sharing mentality (for instance by adopting practices such as the DMP for their projects and dissertations)
 - **BA students**, who need to be exposed to the best practices of FAIR and Open Science early in their curriculum
 - **data stewards** and **data curators**, and their equivalents, employed and newly hired by research institutions, universities, research libraries, who need training

in a train the trainer perspective in order to be able to transmit their knowledge to end users and accompany them in integrating infrastructural services in their research project

- **professionals** of cultural industries

2. **New stakeholders:** they can be reached with our training materials, specifically created to reach a large number of users including those who have limited possibilities to attend face-to face events (this may be due to limited funding, teaching charges, but also disabilities), for more inclusive infrastructural services, and students close to graduation who can approach research directly with the tools provided by our courses (see also Section 4).

This document, developed in collaboration with all the H2IOSC participating institutions, aims at outlining a comprehensive training strategy, encompassing a technical training infrastructure, a roadmap for development of training modules and activities aimed at the community, as well as a plan of upskilling activities aimed at the project's personnel.

In the following sections we will briefly describe the most important initiatives at the international level (1.2), as well as the current state of the four infrastructures with respect to training (1.3), the needs of the project (1.4) as well as the interrelations between training and other parts of the project (1.5, 1.6). Section 2 will then outline the training strategy, aimed at building the training infrastructure (2.3), the development of training modules (2.4) and the internal training plan (2.5). Some perspectives will also be given as to the role of training in the future H2IOSC infrastructure in the long term, and beyond the end of the H2IOSC project.

1.2 TRAINING, FAIR & OPEN SCIENCE

The H2IOSC training strategy draws its inspiration and guiding principles from a number of important initiatives carried out within the framework of European Research Infrastructures, as well as FAIR and open Science practices. Thanks to the involvement of some of the partner infrastructures, the project will engage and participate in relevant initiatives at European and national level, such as for instance Trainers networks, EOSC and SSHOC task forces, RDA networks and WGs as well as with other disciplinary networks. We outline here the most relevant ones.

1.2.1 THE ESFRI PERSPECTIVE ON TRAINING

[ESFRI is the European Strategic Forum on Research Infrastructures](#) and monitors the activities of a number of RIs at the European Level, including the ones participating in H2IOSC.

[CLARIN](#) and [DARIAH](#) are currently ESFRI Landmarks (established infrastructures) while [E-RIHS](#) and [OPERAS](#) are on the [ESFRI roadmap](#) in 2021.

While ESFRI is defining a framework for transnational RIs, its guidelines and monitoring framework are relevant also for national infrastructures. Therefore the fact that ESFRI has identified a *fundamental role of education to achieve excellence* for RIs is also crucial for H2IOSC.

We are inspired by ESFRI's strategic vision of a future oriented towards an interconnected and interoperable *Research Infrastructure ecosystem*, which, by providing high level services to European researcher communities, maintains an absolute edge of curiosity-driven, responsible and socially relevant research. ESFRI recognizes European Research Infrastructures as *important knowledge and innovation hubs*, and is boosting their role as drivers of economic growth, decisive elements for regional development, sources of social well-being, pushing for environmental transitions and for place-based innovation. At the same time, ***Research Infrastructures contribute to education by providing specialised training to students, researchers in academia and industry on cutting-edge scientific methods and technologies.***

RIs are also the forerunner of data exploitation, science digitalization, data engineering and data management, and thus, having an indispensable role for the development of the European Open Science Cloud. *The new ERA concept as an interconnected environment of which RIs are an important element is not restricted only to research, it extends also to other areas like education, innovation, health and public welfare, and others¹.*

Indeed, without proper investment in training and outreach users could be not adequately prepared to access the service that we aim to make available. ***With specific training programs and training materials we hope to increase skills and, with this new skills, redefine the profiles of researchers and create new professional profiles open to work with new possibilities offered by IR and compliance with FAIR principles.***

The [ESFRI Landmark Monitoring Framework](#) currently being applied to infrastructures such as CLARIN and DARIAH ERIC, is likely to be a source of inspiration also for national funding

¹ ESFRI Roadmap 2021 (<https://roadmap2021.esfri.eu/>).

agencies aiming at evaluating national infrastructures such as for H2IOSC²; it provides useful information and guidance on the monitoring of training activities as we will see below in Section 3.

1.2.2 TRAINING IN EOSC AND SSHOC

The [EOSC, the European Open Science Cloud](#), stems from a collaboration between the EU, the member states and the EOSC Association, to which several European RIs belong. It would be beyond the scope of the present document to go into the details of all the important EOSC initiatives related to training and upskilling of the various EOSC players. We will cite here but a few:

- The [Digital Skills for FAIR and Open Science](#) provides a taxonomy of skilled professionals to be trained as well as useful recommendations
- Various EOSC Task Forces are relevant for training: among them in particular the [Upskilling Countries](#) and [Data Stewardship](#) ones, in which members of our infrastructures participate, and that can provide useful references
- The [Skills4EOSC](#) project (*Skills for the European Open Science commons: creating a training ecosystem for Open and FAIR science*), currently ongoing, aiming at developing Data Stewardship training and a network of competence centres; at the national level, [ICDI](#) (Italian Computing and Data Infrastructure), the Italian mandated organisation to the EOSC, also maintains a [national competence centre](#)

While the aforementioned initiatives are aimed at all researchers, independently from the discipline, **our strategy aims to build on top of the already available Open Science and Data stewardship training materials and to target specifically the SSH domain with specific use cases for all of our communities.**

In this respect, the experience gathered in the **Social Sciences and Humanities Open Cloud (SSHOC)**³ project, and continued now in the **SSH Open Cluster**, is of high relevance for H2IOSC. The SSHOC project (2019-2022), was one of the [five EOSC thematic cluster projects](#) that aimed at strengthening the collaborations between ESFRI infrastructures of the same domain; SSHOC saw the participation of **all the 5 ERICs in SSH**, as well as of several other infrastructures including OPERAS and E-RIHS (the latter represented by CNR); currently the

² All details on the monitoring process can be found in the Public Guide https://www.esfri.eu/sites/default/files/ESFRI_LM-Monitoring_Guide_Public-2nd-batch.pdf.

³ Social Sciences & Humanities Open Cloud (SSHOC) is a project funded by the EU framework programme Horizon 2020 and unites 20 partner organisations and their 27 associates in developing the social sciences and humanities area of the European Open Science Cloud (EOSC).

collaboration continues with a memorandum of understanding and with the joint maintenance of the SSH Open Marketplace.

One of the most important outputs of the project we want to take into account for the development of H2IOSC training strategy is the **focus**, for SSHOC, in **providing training, advice, and educational resources** for producers, users, and curators of Social Sciences and Humanities data. ***The SSHOC perspective is centred in empowering users communities for achievement of a greater level of expertise in utilising services, tools, data and other resources in their research, according to the FAIR principles.*** Partners and associates worked together to develop a sustainable cross-disciplinary **training network** and **train-the-trainer toolkit** for research data management, also offering in person training events and webinars for a different type of audience.

In the same way, we want to create new skills and awareness about the tools that researchers can find in H2IOSC infrastructure to support their research activities. We are aware that infrastructures in our sector spend a large amount of resources in training and usually they have dedicated training spaces for disciplinary education and have training officers. But **training materials, like other types of data, run the risk of being “siloed”** and thus not fully visible to their potential audience. The experience of the SSHOC marketplace architecture and data model, where **training materials are considered digital objects** at the same level with datasets and other items, such as workflows and documentation, and interact with other types of objects, indicates the right pathway towards a greater visibility. This vision inspired us also to promote the creation of a **training infrastructure that enables not only the delivery of training, but also that could allow researchers and lecturers to share their own training materials.**

1.2.3 FAIR TRAINING MATERIAL

The FAIR principles (Barker et al., 2022; Engelhardt et al., 2022; Group, 2020; Martinez-Ortiz et al., 2022; Wilkinson et al., 2016) are applicable not only on research data, but also on other resources, like training materials, especially when these are produced during a research program. The demand for training and formal preparation for acquiring and up-skilling digital competences is growing: a major task in developing new training sessions is preparing the materials: this activity costs a lot in time and resources. One possible solution is to find and reuse existing materials, when they are useful for our education objectives. This requires that they have been *shared, correctly described, and made available for reuse* by their authors: suitable online materials that aren't subject to licensing and/or copyright restrictions. It can be

possible only if we all recognise the importance of developing sustainable training materials, shared with the community.

We will build on existing work for the publication of **FAIR training materials** and ***ensure that our modules are citable and made available*** in platforms such as DARIAH campus, DARIAH Teach, Teach with CLARIN, SSHOC Marketplace as well as, of course, in the future H2IOSC Marketplace (Garcia et al., 2020).

1.3 CURRENT OFFER BY PARTICIPATING RIS: TRAINING AND OUTREACH ACTIVITIES

One of the first objectives of H2IOSC will be to raise awareness and train the national communities to the services offered by the participating RIs, so as to encourage their use and reuse by different professions and skill levels, in order to increase competences, ethics and excellence in research. In the following section we describe the state of the art and existing initiatives for each of the four infrastructures involved (with an international and national perspective).

1.3.1 CLARIN ERIC & CLARIN-IT

The focus of CLARIN ERIC is on **language resources** (in all modalities, from all regions and with any topical orientation) and its training strategy serves researchers and educators from any discipline where language plays a major role, ranging from object of study, means of communication or expression, means to store and extract information, object of learning or teaching activities, to training source for data-driven analytics, and many others.

CLARIN ERIC's training activity has a specific disciplinary angle, and is tailored towards the **Language Resources and Technologies**, and with users from the Humanities and Social Sciences in mind. For each targeted group of users we will establish the recommended type of events, and specific learning outcomes will be identified, and associated with specific learning experiences. CLARIN is a distributed infrastructure, and the training activities take place not only at the central level, but also at the level of national consortia and centres⁴.

Central training activities

The central training activities are part of the CLARIN Knowledge infrastructure. They are coordinated by a Training officer, with the support of the CLARIN board of Directors.

[CLARIN's Training Hub](#) is a platform for developing and sharing training materials (both dedicated to central CLARIN tools and to individual services and resources). The CLARIN

⁴ See CLARIN value Proposition <https://office.clarin.eu/v/CE-2021-1830-ValueProposition.pdf>.

training strategy also encompasses best practices for developing and documenting training materials, a [Trainers' Network](#) and funding and support for the organisation of workshops and events. CLARIN supports and sends trainers to major international summer schools in the field of language resources and digital humanities⁵ and organises training events also within the framework of funded projects.

In [CLARIN Funding Hub](#) all the funding opportunities offered by CLARIN are shown and, between others, the **Teaching with CLARIN** Call forms part of a wider platform aimed at stimulating the uptake of CLARIN resources, tools and services into the curricula of SSH disciplines. The aims include the creation of a sustainable network of CLARIN trainers, the publication of training materials in open access, and to share best practices in teaching. Two funding instruments are already in place: the CLARIN [Training Network](#) and the [Training Suite](#).

CLARIN ERIC's action starts from an existing body of learning materials such as those surveyed within the [UPSKILLS project](#)⁶ (Gledić, Assimakopoulos, et al., 2021; Gledić, Đukanović, et al., 2021; Miličević Petrović et al., 2021), providing a list of already available resources relevant for teaching in the domain of languages and linguistics. As for CLARIN core training materials to be adapted and translated to meet the needs of the H2IOSC community (Section 2.3.1), are worth mentioning those developed within different projects, namely:

- the course [Introduction to Language Data: Standards and Repositories](#) developed for UPSKILLS and aimed at BA students of language related disciplines⁷ including modules on: the management of the LR's lifecycle; how to research data repositories; how to make resources FAIR; how to correctly cite data (Section 2.4.2)
- the [Training Materials](#) available in [CLARIN Learning Hub](#), collecting materials that integrate CLARIN resources, tools and services. Those include, but are not limited to, the resources winners of the *Teaching with CLARIN award*, where the Italian community already took a stand with 2022 award winner Dr. Rachele Sprugnoli, who presented the course on [Natural Language Processing Methods](#).

⁵ For a list see <https://www.clarin.eu/events>

⁶ "The UPSKILLS project is an Erasmus+ strategic partnership for higher education that has been seeking to identify and tackle the gaps and mismatches in skills for linguistics and language students through the development of a new curriculum component and supporting materials to be embedded in existing programmes of study", see <https://upskillsproject.eu/>.

⁷ This and many other courses are accessible from the project's Moodle platform after registering with a free account at <https://upskills.fil.bg.ac.rs/login/index.php>.

CLARIN ERIC's strategy has already been shaped by important events such as:

- [CLARIN@Universities](#) workshop (2019) on the experience with the integration of CLARIN content into university programmes
- [CLARIN in the Classroom](#) (2020), an initiative for university lecturers to present their experience with the use of CLARIN resources, tools and services in their courses and to suggest future steps to help such integration into university curricula
- [Teaching with CLARIN](#) (started 2021), an award given to lecturers who have successfully integrated CLARIN in their training materials

H2IOSC's approach will also benefit from the guidelines *Integrating research infrastructures into teaching: Recommendations and best practices* developed by CLARIN within the UPSKILLS project and providing an introduction to the CLARIN research infrastructure for lecturer, teachers, trainers and curriculum designers wishing to integrating linguistic resources, tools and research data repositories in their teaching materials (van der Lek et al., 2023). Another interesting aspect of UPSKILLS to be taken into account are the surveys conducted to analyse the necessities of teachers and students of language related courses (*Toward a new profile for twenty-first century language specialists: Industry, institutional and academic insights*, Bernardini & Miličević Petrović, 2021): the structure of those surveys could be tailored to the needs analysis of our community and their results could be taken into account when evaluating our own (Section 1.4).

Finally, an important **joint initiative between CLAIRN and DARIAH ERIC** is the [Digital Humanities Course Registry \(DHCR\)](#), a platform aimed at students wanting to enrol in Digital Humanities programmes at university level. The DHCR lists programmes (BA and MA) and courses in the broader domain of DH, among which a number of Italian courses. The relevance of this initiative as a discovery tool to disseminate H2IOSC and its training content in the academic community, cannot be underestimated. H2IOSC could improve the coverage of the Italian section of DHCR, ensuring that more courses and programmes at the national level are recorded, in particular from the domain of Cultural Heritage which is currently underrepresented. This would in turn provide a **ready to use directory of potential stakeholders for our training strategy**.

National consortia and User Involvement Committee

Of course a large number of training happen at the level of individual consortia, and are aimed towards the national communities. Training activities are monitored as part of the national User Involvement activities and User Involvement Representatives from each country exchange,

share and align their best practices within the framework of the activities of the [User Involvement Committee](#). **The participation in this network will provide the H2IOSC consortium with a large body of useful information as to the best practices and success stories from other national consortia.**

K-centres activities

[CLARIN's Knowledge Centres \(K-centres\)](#) are a cornerstone of the [CLARIN Knowledge Infrastructure \(KI\)](#), one of the main components ensuring a continuous transfer of knowledge between all players involved in the construction, operation and use of the infrastructure. The mission of the CLARIN KI is to ensure that the available knowledge and expertise does not exist as a fragmented collection of unconnected bits and pieces, but is made accessible in an organised way to both the CLARIN community and the social sciences and humanities research community more widely.

K-centre's mission is to share their knowledge and expertise on one or more aspects of the domain covered by the CLARIN infrastructure and can be mostly found in CLARIN countries, but also exist elsewhere, and they all have a virtual presence. Alongside the important service of helpdesk, **training and education are crucial** missions of K-centres.

More specifically, nodes focus on:

- fostering the creation of data management plans for specific types of Language Resources, so as to facilitate a new generation of FAIR-by-design resources
- facilitating the use of CLARIN services and resources, not only those offered by CLARIN-IT but also those offered via the CLARIN technical infrastructure (Virtual Language Observatory, Language Resources Switchboard, etc., see Section 2.4.2)
- promoting access and use of the CLARIN knowledge infrastructure (consulting services offered by CLARIN k-centres, transnational access, support for networking, starting grants and other initiatives) for young researchers and PhD students
- ensuring the support for Training and dissemination activities

CLARIN in Italy

The [CLARIN-IT national consortium](#) has always been very committed over the years to user involvement initiatives including training, which we plan to enhance with H2IOSC. CLARIN-IT is involved in a number of national training activities both directly and via the [DiPText](#)

[Knowledge Centre](#). Among all the interventions⁸, it is worth mentioning the **Summer Schools and permanent training initiatives**, such as:

- the [Digital Tools for Humanists Summer School](#) organised with the Laboratory for Digital Culture at the University of Pisa
- the Summer School [Digital Humanities and Digital Communication: AI and \(New\) Literacies](#) organised within the Doctoral Programme in Human Sciences of the University of Modena and Reggio Emilia
- the [Multi3Generation 2022 Training School](#) organised by the COST Action with the support of the DiPText-KC
- the [Digital Text Processing Training School](#) held at the Ca' Foscari University of Venice
- the [Permanent Seminar Series "A Bridge Between Two Worlds"](#) by the CNR-ILC and the VeDPH through the DiPText-KC
- the [Seminars in Digital and Public Humanities](#) held at the VeDPH and supported by CLARIN-IT
- Curricular internship at the Ca' Foscari University of Venice and at the University of Pisa

Furthermore, CLARIN-IT's educational initiatives also involve **curricular internships specifically aimed at language students**. The primary objective of these internships was to enhance skills related to translation assisted by Artificial Intelligence (AI). The participating students undertook the responsibility of correcting the bilingual (Italian and English) subtitles of webinars organised by CLARIN-IT on Digital Humanities (DH) topics. Through this hands-on experience, students were able to gain practical insights into the evolving field of AI in translation while contributing to the dissemination of knowledge on DH topics. The multimedia products resulting from this collaboration are freely available online⁹.

Within the framework of CLARIN ERIC national node, CLARIN-IT reported an increasing number of training activities, encompassing workshops, lectures, webinars and Summer Schools **for a total of 9 in 2020, 19 in 2021** (year when there was an increase in online initiatives due to the pandemic) **and 16 in 2022**, reaching a total of around **1200 participants in the time span of three years**.

As for the current year, the census of initiatives is still in progress, but it can be estimated that the results will exceed those of 2022 with a growing trend. Compared to previous years, a

⁸ A complete list of user involvement initiatives is available here: <https://www.clarin-it.it/en/content/user-involvement-initiatives>.

⁹ https://www.youtube.com/playlist?list=PLzkDEscu2r2qt_PCxxL23eSauQa22lyWG

more consistent effort has been made to monitor the initiatives carried out by the individual members of the national consortium, with periodic questionnaires on the measurement of user involvement interventions, including training initiatives.

1.3.2 DARIAH-ERIC & DARIAH-IT

DARIAH-ERIC's general objective is that of enhancing and supporting digitally-enabled research across the humanities and arts' field, anchoring them at the centre of a technologically evolving knowledge society. DARIAH's mission is to empower data communities of researchers across EU Member Countries and Associated Partners with digital methods, in order to create, share and connect knowledge and to generate impact about socio-cultural research themes.

Description of general training activities

DARIAH empowers research communities with digital methods to create, connect and share knowledge about culture and society. Training and education is central to this mission¹⁰; research in a fast-changing environment requires sustained access to new perspectives and methods. The digital is not a goal in itself, but a means to enable and enhance the highest quality arts and humanities research.

In order to develop, curate and maintain resources and services, DARIAH relies on four **Virtual Competency Centres (VCC)** and [Working Groups](#). The VCCs coordinate the Working Groups and define the general parameters under which they work. The [Virtual Competency Center Research and Education Liaison \(VCC2\)](#) aims to promote and support the use of research data and ICT methods and technologies, including the DARIAH infrastructure. It serves as the primary contact with the research and teaching communities for Arts and Humanities (A+H), providing the interface between the DARIAH research infrastructure and researchers undertaking research at different levels of depth across A+H. It seeks to understand A+H research activities and processes, and to understand and promote the use and application of ICT-enabled methods and tools, with particular emphasis on interdisciplinary understanding and exchange.

VCC Research and Education encourages, supports and enables researchers to use DARIAH tools and services to assist in asking new questions and addressing old problems in new ways,

¹⁰ For further details see the [Training and Education section of the DARIAH portal](#).

through the use of research data and the application of ICT methods. In this way, it aims to address a range of people and interests, from established researchers to postgraduates and students, as well as different disciplines and sectors within and outside of higher education.

In addition to essential support services for the design and promotion of the DARIAH infrastructure, and among other key concepts and components, this VCC contributes a knowledge base that captures and connects A+H methods, tools and projects, and refers to curricula in digital humanities.

DARIAH-ERIC Training Platforms:

- [DARIAH Teach](#) offers free community-driven and multilingual learning and teaching materials
- [DARIAH Campus](#) is both a discovery framework and a hosting platform for DARIAH and DARIAH-affiliated offerings in training and education. It brings together external resources, such as those from the following 3 projects, as well as hosting original training resources that can be reused by academics in the community to create bespoke training. DARIAH-Campus also offers *pathfinders*, a curated discussion of resources related to a specific area, and the ‘captured events’ tool, sustaining training events for future use. DARIAH-Campus records are currently also automatically injected in the SSHOC Marketplace.
- [PARTHENOS Training](#) offers lectures and exercises covering a wide variety of topics related to digital humanities, digital heritage and research infrastructures. PARTHENOS Training is particularly suitable for instructors looking to reuse materials or sharpen their own skills, but is also designed to be used by self-learners
- The aforementioned CLARIN-DARIAH [DH Course Registry](#), compiles courses at all levels, based in European higher education institutions.

DARIAH-IT training activities

DARIAH-IT is a partner of the [Master Course Infotext in Text Computing and EElectronic Publishing at the University of Siena](#). DARIAH-IT held classes for the Master on the topic of digital research infrastructures and the projects that DARIAH-IT supported throughout the year.

DARIAH-IT participated in the workshop “Filologia digitale e testi italiani antichi - Verso un sistema integrato di ricerca”, held at Scuola Normale Superiore of Pisa, Palazzo della Carovana, Sala Stemmi, 27-28 Giugno 2022. The event was organised by the Scuola Normale

Superiore in collaboration with DARIAH-IT and the Istituto Opera del Vocabolario Italiano of the National Research Council (OVI-CNR)

1.3.3 E-RIHS

E-RIHS is the European infrastructure delivering cross-disciplinary innovation in heritage science. It aims to extend the scientific impact and boundaries of the heritage science field and to foster knowledge, growth and expertise of the whole heritage science community.

The E-RIHS training offer, in its current form, is organised under the label of [HS Academy](#).

HS Academy initiative delivers heritage science training activities, based on the experience of previous projects [IPERION HS](#) and [IPERION CH](#), integrating the most recent scientific developments and taking into consideration actual needs of the growing international heritage science community¹¹.

HS Academy offers online and in-presence training events, provides open access materials on various topics including but not limited to: art conservation, art history, architectural history and criticism, theory, restoration, digitalization, chemistry, physics, informatics and other studies applied to cultural heritage.

Currently, HS Academy offers in-presence, hybrid (held simultaneously online and in-presence) and online training activities. Streaming of training activities, when recorded upon the consent of interested parties, is made available for all HS Academy learners as free recorded online training videos. As for now, all the activities are conducted in the English language only and are open to an international audience.

The current HS Academy training offer includes series of activities described as follows:

- [The IPERION HS Doctoral Summer School](#), held every two years, provides primarily theoretical and technical lectures on a specific topic of interest for the heritage science community: e.g. collections care, digitalization of built heritage etc. The admission to the Doctoral Summer Schools is upon selection process, the course fees are covered
- [The IPERION HS Training Camp](#), held every two years, provides primarily hand-on practical training on a specific topic of interest for the heritage science community: e.g. paleontological and archaeological studies, built heritage etc. The admission to the

¹¹ See also the IPERION HS training plan (Mazzeo et al 2021) and D8.3 Deliverable on existing centralised procedures (Benassi et al 2023).

Doctoral Summer Schools is upon selection process, the course fees are covered, with a small contribution to be paid by participants

- [IPHS webinar programme](#), a series of webinars held monthly with invited talks on different Topics related to TNA services. The participation online is free upon registration
- The lecture series “[Current Topics in Heritage Science](#)” of the IPERION HS Academy, held monthly with invited talks structured as lectures providing fundamental knowledge on a specific topic of interest for the heritage science community
- [User meeting events](#), exchange of experience of using TNA services or with other projects

1.3.4 OPERAS

The main aim of [OPERAS](#) is to foster an **Open scholarly communication system** in the social sciences and humanities (SSH) in the European Research Area, by federating and coordinating resources in order to assess and address the needs of European Researchers in the field of SSH regarding various topics such as Open Access, Open scholarly publications, Open Science, etc.

OPERAS as a network is involved in two projects aimed at promoting knowledge and competencies on FAIR and Open Science, and Open scholarly publishing:

- [Skills4EOSC](#) ‘Skills for the European Open Science commons: creating a training ecosystem for Open and FAIR science’ is funded by the European Commission Horizon Europe programme (GA 101058527). Coordinated by Consortium GARR and supported by 44 partners in 18 European countries. Skills4EOSC will set up a pan-European network of competence centres to speed up the training of European researchers and harmonise the training of new professional figures for scientific data management. Skills4EOSC launched its first [Fellowship programme](#) call in August, aimed at providing data professionals with relevant skills in Open and FAIR data practices
- [TRIPLE](#) “Transforming Research through Innovative Practices for Linked Interdisciplinary Exploration”¹². At the heart of the project is the development of the [GoTriple platform](#), an innovative multilingual and multicultural discovery solution. It will

¹² [TRIPLE](#) has received funding from the European Union’s Horizon 2020 Research and Innovation action funding scheme INFRAEOSC-02-2019 “Prototyping new innovative services” (grant agreement #863420). The content of this website reflects only TRIPLE’s view and the Commission is not responsible for any use that may be made of the information it contains.

be one of the dedicated services of [OPERAS](#). Within the context of TRIPLE, OPERAS and its partners developed two main outputs dedicated to training:

- [TRIPLE Open Science Training Series](#), a series of 12 open and reusable training events specifically designed to upskill researchers in FAIR and Open Science. The training originally ran from March 2021 to June 2022 and focused on Open Science and EOSC related topics as well as on GoTriple services
- [TRIPLE Training Toolkit](#), designated to facilitate trainers in organising FAIR-by-design training events and sharing the related training materials as Open Educational Resources (OERs).

1.4 NEEDS OF THE H2IOSC COMMUNITIES - LINK TO WP2

To assess the training needs of our communities, WP8 can draw from a significant body of previous work carried out in international projects involving our research infrastructures. In addition, the results of H2IOSC WP2 will be crucial for the adaptation to our needs. One of the objectives of WP2 is, in fact, to produce web-based questionnaires to assess **scientific communities' educational needs** and the **visibility and the level of awareness of the four RIs services and tools** within the Italian community. Such a nationwide and cross-infrastructure survey has never been conducted before, especially at this scale and this analysis. Such work will provide more details about the rising needs of communities.

In line with the project timeline, the WP2 working group has:

- **Developed a questionnaire** to identify the educational needs of the H2IOSC communities
- **Tested** the questionnaire within a restricted and selected group of users (representing the different disciplines that coexist within each research CNR Institute)
- **Refined** the questionnaire, based on the outputs-comments-suggestions that emerged from the first respondents
- **Validated** the questionnaire with the involvement of a selected part of the target community - consisting of mailing-list associations whose members are considered representative for each IR

The questionnaire is articulated in different sections, aimed to explore:

- **resources developed/used** - the type of resources created/used, description, compliance with FAIR requirements, purpose of creation/use

- **digital technologies developed/used** - type of technologies/software created/used, description, compliance with FAIR requirements, formats, usage licences, purpose of creation/use
- **knowledge and use of IRs and the services they offer**
- **training needs** - interests, preferences about course fruition, topics to be explored in depth
- **needs for new resources/technologies/software to be developed**
- **publications policy** - concerning the different institution affiliation

Future developments:

- dissemination of the survey to the entire target community
- definition of the data analysis criteria for questionnaire results
- operative planning of interviews

WP8 will monitor the results of the WP2 survey, and its analysis, and ***will develop the training catalogue (in continuous evolution) with educational offerings based on currently identified and periodically reviewed needs.*** Our aims are:

- filling gaps in training, expressly indicated by stakeholders, with structured and diversify training programs, addressed to specific targets (priority in the development of our training strategy); answer to the personal and professional goals of different groups of stakeholders - learner centred perspective
- filling needs that are still unexpressed, but nonetheless identified by our team (we take them into account during the design of our training offer)
- build an integrated and interconnected training environment based on already existing training activities and material, enhancing and creating synergies between what has been done individually by the 4 IRs
- upgrade technology resources for improve the training opportunities and enrich the quality and methods of our training strategy
- monitor the efficiency and effectiveness of our learning system and the quality of courses and materials
- analyse feedback from our users through post-event questionnaires
- lifelong learning to be always in line with technological progresses, legislative evolution, national and international guidelines related to the world of research that bring changes and innovation.

1.5 TRAINING IN THE FUTURE H2IOSC MARKETPLACE - LINK WITH WP5

H2IOSC Marketplace will be an online tool to increase visibility and valorisation of the data sources, services, and resources provided by the IRs and the research community at large, and hosted in the H2IOSC National Cloud under full compliance with the FAIR principles. The Marketplace structure has been identified in the European research space as an essential step in supporting the transition from the ordinary research landscape to cloud-based infrastructures.

The H2IOSC Marketplace will be a universal entry point for various user types and different usage models of the national research communities involved, to find and employ services and resources. In this sense it will allow for knowledge as well as data sharing, advertising tools and methodologies, and fostering collaboration and innovation. ***It will allow participants to share and make findable research datasets and tools as well as training materials,*** through dedicated hubs, contextualised catalogues, instruments for presentation, visualisation, and aggregation.

The training infrastructure will thus be a part of the H2IOSC Marketplace and should enable users to find and share training materials, to enrol in courses both online and face to face, and to engage in self learning activities.

1.6 TRAINING FOR COMMUNITY PILOTS - LINKS WITH WP7

WP7 is dedicated to define and implement a series of innovation oriented services and proof-of-concept sets of resources for direct use by researchers in the form of Pilot applications hosted on the H2IOSC platform.

Pilots in the form of platforms or hubs will coalesce custom-designed domain-specific services, workflows, and interfaces, operating on appropriate data subsets for relatively rapid prototyping. Pilots will be designed so as to be both scalable and extendable, by adding new resources and data, in order to gain insight and adjust aims as the implementation proceeds. User groups from the national research community and contributions from the European counterpart of the national RIs, will be involved/included in the design process from the early phases.

The Pilots will allow information and tools to be ***accessible to scholars***, institutions, and other members and stakeholders of the scientific community in the Open Science and Open Data

spirit and according to the FAIR principles, also **enabling the use of innovative research results and practices in** their own sites of interest and **fields of enquiry, communication, and learning.**

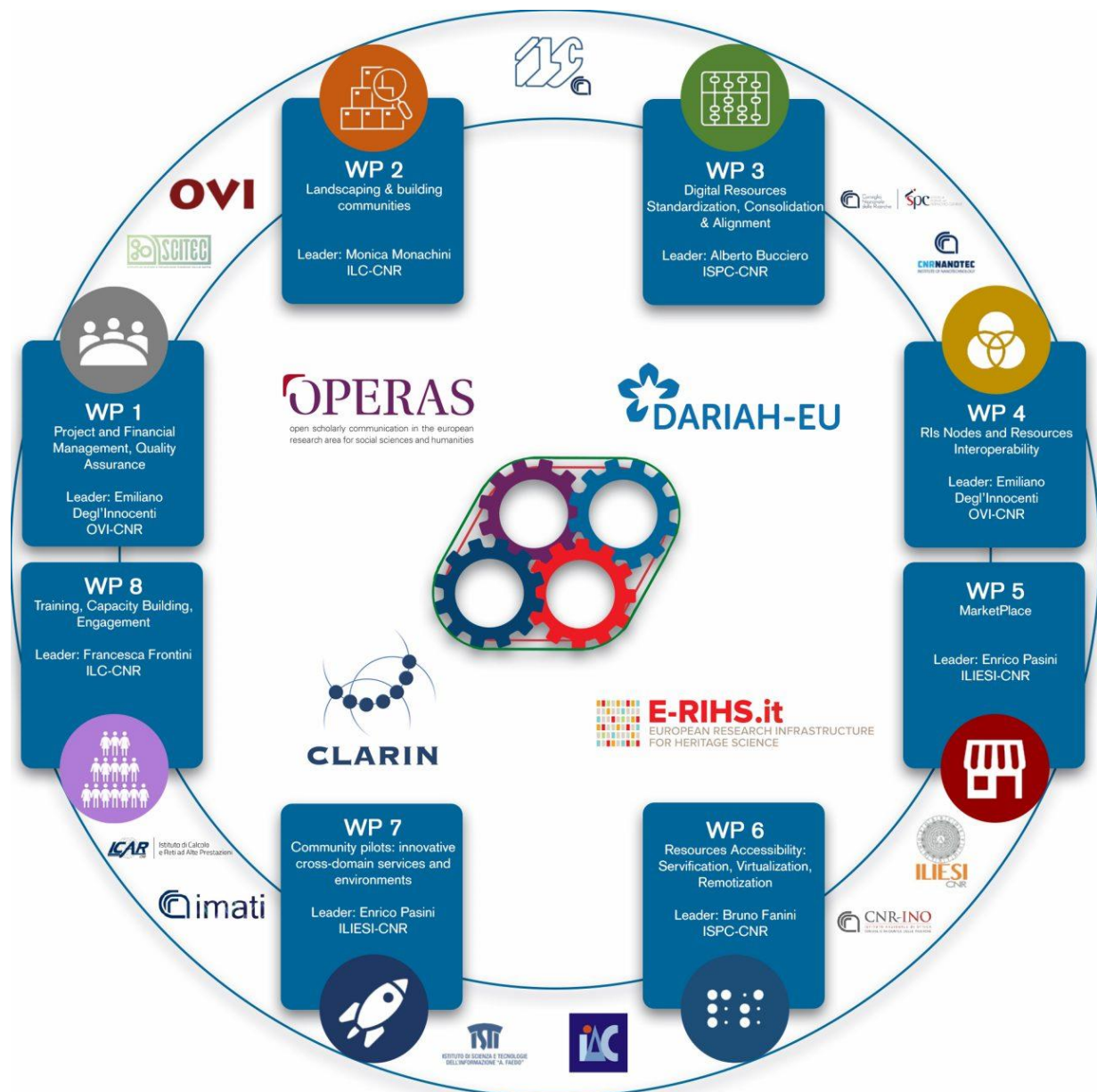


Figure 1 - H2IOSC work packages overview.

2. H2IOSC TRAINING AND OUTREACH PLAN

2.1 APPROACH AND LIMITATIONS

Our aim is to define a **comprehensive shared strategy for the training and engagement activities, at the level of single infrastructures and for the whole project H2IOSC**. Based on the aspect explored above and starting from the stakeholders training needs, H2IOSC strategy will aim at the following **objectives**:

1. the creation of an **infrastructure for sharing and delivery of educational materials**, that we will describe below, and that will focus in particular on the FAIRification of training materials and on self-training and distance learning
2. the creation of learning materials and content for the upskilling of the H2IOSC user communities, aimed at
 - providing disciplinary training on FAIR and Open Science, built also on existing training materials and taking into account the specific needs of our communities
 - providing training on data management, data lifecycle and guidelines for the creation of Data Management Plans, including also data preparation, metadata and standards, data citation, legal and ethical issues, personal data protection, GDPR compliance, specific focus on Italian laws and their evolution
 - adapting and creating training materials for the existing services of CLARIN, DARIAH, OPERAS, E-RIHS to allow national users to take advantage of the full potential of the RIs
 - providing tutorials for the depositing services, aimed at increasing the use of our national nodes and stressing the importance of depositing in trusted repositories
 - providing tutorials on how to integrate own tools and resources as services for CLARIN, DARIAH, OPERAS, E-RIHS
 - developing new training materials to use within the innovative services of the Marketplace, with a special focus on the Community Pilots
3. the **upskilling of the H2IOSC workforce** itself for the duration of the project, in particular for what concerns Open Science and a wider knowledge of the services of each RI by all partners.

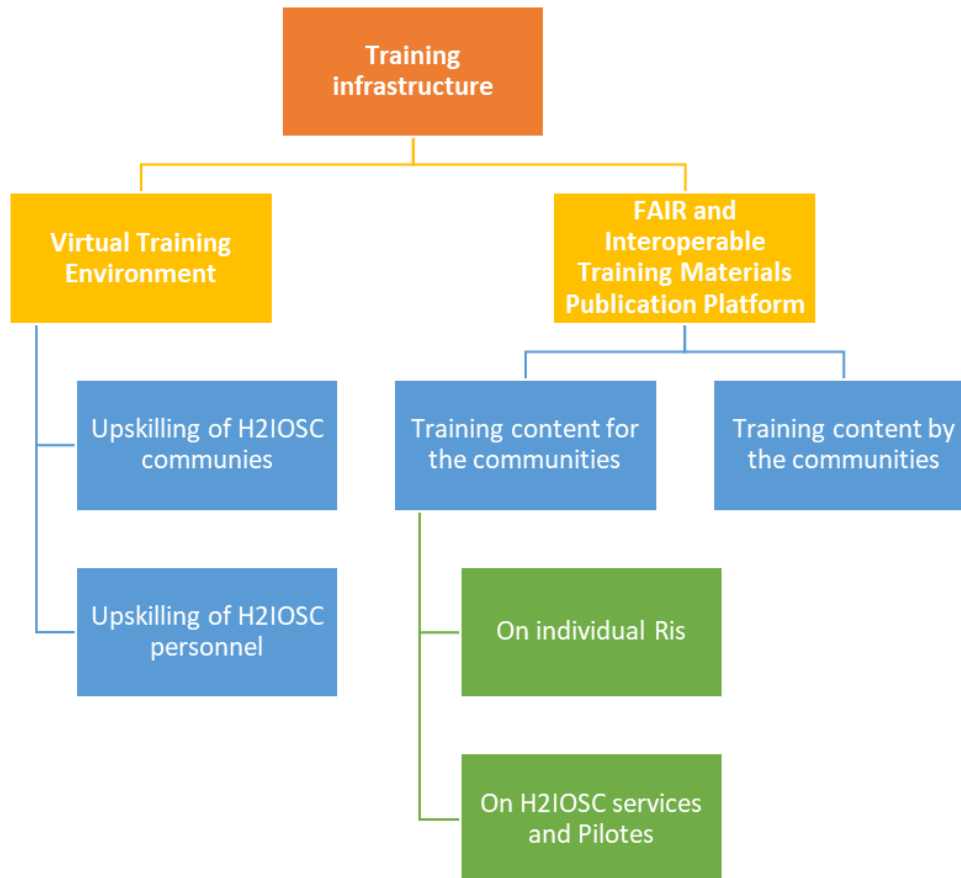


Figure 2 - H2IOSC training strategy in a nutshell.

A major limitation, imposed by our funding rules, is that the H2IOSC budget can only cover training expenses for personnel specifically hired on the project. For this reason, all training and outreach activities aimed at external personnel, including the members of our consortia and our users, should be funded by independent means. Thus the **WP8 effort will be aimed at developing training materials that can be published, reused by instructors from our infrastructures and by others as well as used for self learning and distance learning.**

While outreach activities are not strictly speaking funded by the H2IOSC project, we believe that it is the task of WP8, in collaboration with WP1 and WP2, to monitor major events and promote the organisations of workshops, training events, conferences, aimed at disseminating the outcomes of the H2IOSC project and the services offered by our partners. A monitoring framework will be outlined in this first document (Section 3).

In the following sections we will outline the methodology, sketch out our approach for the 3 aforementioned objectives.

2.2 METHODOLOGY

The implementation of this strategy will follow the path indicated in previous paragraphs, first building on the existing resources, and then extending to new areas. The first 12 months after the publication of this document will be devoted to the **development of the training infrastructure** (Section 2.3) as well as to the **adaptation and creation of materials related to each RI involved**. Subsequently, **new H2IOSC-related materials will be developed** (see Section 1.6 on Community Pilots).

The adjustment of existing **guidelines** (Section 1.3.1) and the creation of **new best practices** will be pursued to ensure the application of FAIR principles and Open Science methods, the use of clear licences to enable their easy reuse and a system of standard metadata that will allow training materials to be linked to tools, resources and publications also with a view to their accessibility to users with different needs.

In order to provide advanced training, the needs analysis will be targeted to the Language Technologies, Humanities and Cultural Heritage and specifically to the figures identified in the WP2 (Section 1). Surveys will be conducted through questionnaires to be administered before and after the delivery of the training (Section 1.4) so as to enable targeted adaptation of materials and the monitoring of the progress (Section 3.1).

Depending on the various teaching requirements and the needs of the target community, teaching materials will be delivered in the form of lectures, individual **webinars, workshops, training events, online courses, seminars, tutorials, podcasts and also with possible use of innovative microlearning methodologies** and, in relation to the target audience specified above, will be oriented towards:

- the improvement of the users' ability to understand the tools that the H2IOSC infrastructure provides
- raising their awareness about the connection between IR and research activities
- making them conscious about the common needs of the funded programs (for example a well done Data Management Plan) and about the new skills essential in the world of Digital Humanities

These materials will be made available on learning management systems that will provide learners with an intuitive interface, foster engagement and allow progressive evaluation as well as feedback. **For each training material, the prerequisites, modular objectives and possible insights will be clearly detailed so as to target a specific level of users.** A

particular focus will be given on the connections between courses, with the definition of suggested “paths” across modules.

Considering the possible call for proposal of new training materials funded by each RI, **specific workflows will be developed and aimed at participants in order to accompany and support them in the creation of data management plans for their project**, so as to facilitate a new generation of FAIR by Design resources in parallel with the adaptation to FAIR standards of already existing ones. Besides the adjustment and creation of new training modules for the benefit of students and researchers, the same effort will be directed to creating training opportunities for trainers, data curators and data stewards in order to facilitate the use of existing services and resources.

These materials will be published according to the FAIR principles for training materials and will become visible on various platforms (such as the SSHOC training platform and the national marketplace). These will also include **specific training modules for legal and ethical issues, standards, data management plan and servification** in compliance with each RI infrastructural requirements.

2.3 TRAINING INFRASTRUCTURE

A core part of WP8 activities and budget are devoted to the creation of **a training infrastructure that will become a core component of the H2IOSC Marketplace**. A number of discussions with the partner in this work package, as well as a thorough analysis of the state of the art and of existing platforms offered by various infrastructure, has led us to identify the need for **two separate platforms**, one that will support training activities directly organised by H2IOSC and its partner infrastructures (both directed to its personnel and to the communities of users), and one aimed at creating a trusted repository of FAIR training materials, open to all users for deposit and reuse. This latter platform, inspired by and improving on the example of DARIAH Campus, should promote the idea of **Training Materials as digital objects that can be cited, licensed, versioned and reused in accordance with the FAIR principles**. Currently the technical and functional requirements for both platforms have been defined, and we will briefly outline them in the following paragraphs.

These technical and functional requirements have been transformed into specification documents aimed at public calls for the development of both platforms.

2.3.1 VIRTUAL TRAINING ENVIRONMENT

From the technological point of view this platform will be a **Learning Management System (LMS)**, that is to say a hub for building and running courses; core functionalities will be:

- a simple interface allowing for the publication of courses, with search functionality and calendar of courses
- templates for learning design
- access Policy and Users Management Strategy
- management of enrolment, and issuing of certificates
- functionalities for testing and monitoring
- possibility to create virtual communities and facilitating exchange by users, facilitating feedback by teachers
- integration with other applications (in particular video conferencing apps, notebooks for coding, audio video recording, compatibility with various formats and multimedia content)
- possibility to run synchronous and asynchronous training sessions
- compliance with security and privacy requirements
- automation of tasks

Moreover, the platform should be integrated with the H2IOSC marketplace and community in two ways:

- adopt the H2IOSC visual identity
- allow for single sign on so as to allow users to access with their institutional identifiers

We consider that an adaptation of the Moodle platform would be the most appropriate solution, also taking into account needs for scalability and sustainability. The management of the platform and of the courses offered will be restricted to the H2IOSC consortium members, with the possibility to give access to external trainers for individual courses.

2.3.2 FAIR AND INTEROPERABLE TRAINING MATERIALS PUBLICATION PLATFORM

The second platform will be open to the broader community of the Italian Humanities and Cultural Heritage. From the technical point of view it will be a repository, allowing for the deposit of training resources by trainers, teachers, lecturers. The training materials deposited will be first and foremost those produced by the H2IOSC consortium, but will not be limited to them.

The focus and principal use of this platform will be **to preserve, share, cite and reuse training materials as FAIR digital objects** in such a way as to create a community of trainers sharing and reusing each other's materials. For this reason, guidelines for deposit will also be developed so as to facilitate reuse and integration in academic programmes and training.

The technical and functional requirements for this platform are:

- compliance with the semantic framework of H2IOSC and with the set of metadata defined to describe training materials
- possibility to harvest metadata via standard protocols (OAI-PMH) so as to make materials visible in the H2IOSC marketplace, but also in other meta catalogues
- a guided depositing procedure facilitating the correct documentation of courses
- single sign on so as to allow users to deposit with their institutional identifiers
- possibility to deposit multimodal materials and compatibility with different formats
- a curation system allowing for administrators to check and validate deposits
- possibility to gather materials in channels and collections
- attribution of persistent identifiers (DOIs) to each deposited material
- versioning of materials
- clear indication of authorship and suggested citation format
- support and clear indication of licence and conditions for reuse

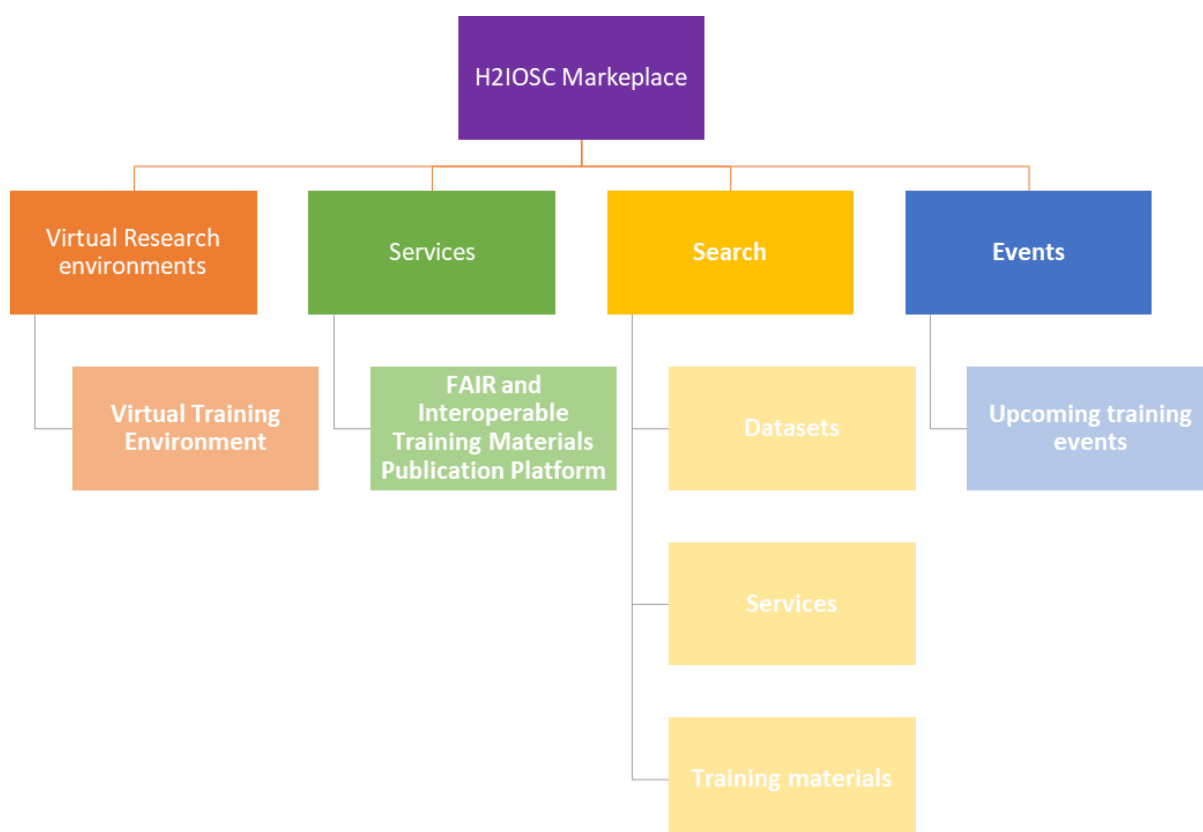


Figure 3 - integration of the FAIR and Interoperable Training Materials Publication Platform with the H2IOSC marketplace.

The **integration of the FAIR and Interoperable Training Materials Publication Platform** with the H2IOSC marketplace will be twofold:

- the platform will be listed as one of the H2IOSC services, alongside other depositing services such as for instance data repositories
- the metadata from the platform will be made visible and searchable via the H2IOSC search engine, so as to allow users to **discover Training Materials alongside other digital objects**

2.3.3 OTHER TRAINING EQUIPMENT

In addition to these platforms, which will be implemented by CNR-ILC and hosted at the CLARIN Pisa Data centre, all units participating in WP8 have identified needs in terms of software and instrumentation that will enable them to strengthen their training offer in terms of production and adaptation of existing modules.

Namely, we refer to software for audio and video monitoring, screencasting, creating multimodal online courses and related instrumentation such as microphones, cameras and tripods, video conferencing equipment, workstations for audio and video editing and for creating online courses.

2.4 TRAINING AND UPSKILLING THE H2IOSC USERS

This part mostly concerns the development of training modules, aimed at the services of our infrastructures and of the H2IOSC marketplace, as well as at SSH specific Open Science Skills. We will present below the envisaged joint training activities as well as those of the four participating RIs.

2.4.1 DEVELOPMENT OF JOINT MODULES AND INITIATIVES

The RIs involved will produce both joint and specific training activities tailored to their own communities. In general, the joint commitment of the RIs will be aimed at improving the coverage of the DHCR in Italy and consolidating the presence of H2IOSC output within major scientific and training events in our community.

More specifically, among the cross-domain efforts, the following modules will be implemented:

- Module: ***Guidelines Integrating RIs in teaching and learning***, inspired by work carried out in UPSKILLS, focusing on:
 - higher education
 - research (PhD training)
 - high school

- Module: ***How to Use of the H2IOSC marketplace***
 - Including how to use the training platforms
 - How to deposit and FAIRify training materials
- Module: ***How to write a Data Management Plan***
 - building and improving on existing material
 - integrating the H2IOSC offer of services for FAIRification

2.4.2 TEACH CLARIN, TEACH WITH CLARIN

As the Infrastructure for language technologies, CLARIN has a broad user base, covering a wide range of disciplines, going from linguists, digital humanists, language resources and technologies experts, researchers in computational linguistics and natural language processing, historians and oral historians, (oral) history, social scientists, cognitive scientists and beyond.

One of the objectives of training aimed at these scientific communities is to be able to **use the CLARIN infrastructure** and in particular **the core services** such as:

- search with the [Virtual Language Observatory](#) (VLO)
- deposit with the CLARIN centres (with particular reference to the [ILC4CLARIN B-centre](#))
- create [virtual collections](#) and appropriately cite Language Resources
- use the [Language Resources Switchboard](#)'s services
- use the CLARIN Knowledge Infrastructure and K-centres, in particular the [CLARIN Knowledge Centre for Digital and Public Textual Scholarship](#) (DiPText-KC)
- use the services of the [ILC4CLARIN repositories](#)

In order to do this, we plan to adapt and translate the core CLARIN courses already developed under previous initiatives, such as the courses shared by the CLARIN ERIC consortium and available in the [Learning Hub](#) and the courses developed by CLARIN for the UPSKILLS project (see section 1.3.1), and notably the [Introduction to Language Data: Standards and Repositories course](#), which is intended as teaching material for lecturers in language and linguistics BA courses to export and adjust depending on their purposes.

The UPSKILLS course is composed of five modules plus a glossary (see figure 3), which cover the basic aspects of Language Resources data management, outlining services provided by CLARIN for each aspect. The overall workload associated with the first 5 units of this block

amounts to 6 ECTS. All UPSKILLS modules are licensed under Creative Commons Lincense (CC-BY) and allow reuse.

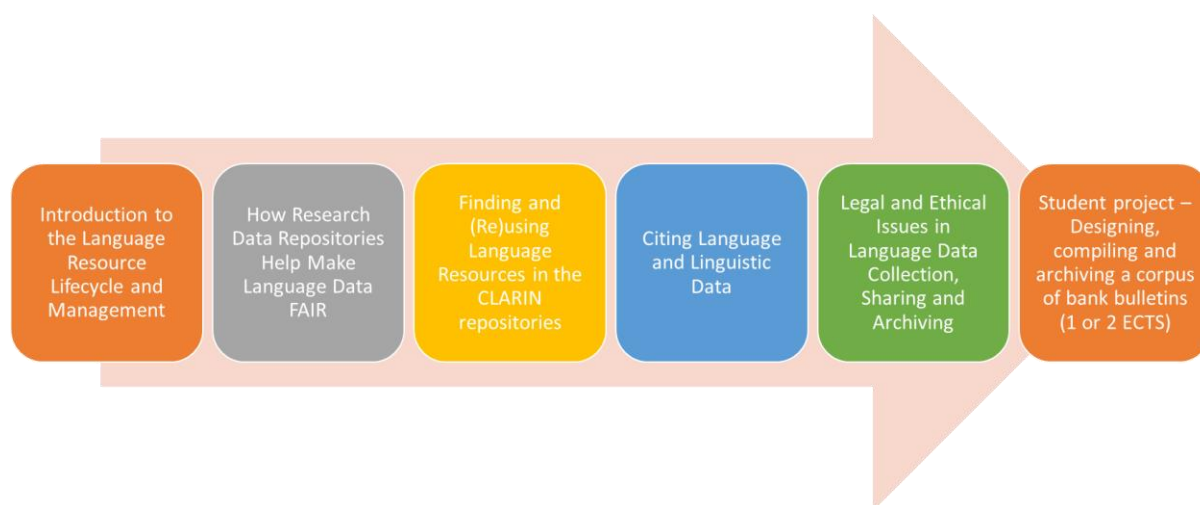


Figure 4 - The UPSKILLS modules, to be further developed and adapted in H2IOSC.

The course follows a modular approach and is already intended for resume and integration by lecturers in higher education. Originally designed for BA students in applied linguistics, the modules will be adapted and translated to meet the needs of the Italian community, broadening the scope both in terms of users, level of complexity and of disciplines.

The UPSKILLS modules were licensed in such a way as to allow for reuse, but we envisage a collaboration with CLARIN ERIC, in the common interest of producing more general modules beyond the scope of the now concluded UPSKILLS project, to be published on the CLARIN Learning Hub and then adapted for H2IOSC.

Based on the same approach, new modules will be developed for CLARIN-IT's Community Pilots (Section 1.6), namely:

- **A CLARIN for Oral History tutorial**, aimed at researchers who need to curate and deposit legacy archives in the platform that will be developed
- **A CLARIN for Linguistic Linked Open Data and Terminologies Tutorial**, aimed at facilitating the use of the LLOD and terminological platform, and touching upon the following points:
 - Creation, conversion and deposit of (multilingual) terminological resources in Skos

- Creation, conversion and deposit of LLOD datasets to be hosted on a dedicated endpoint
- Reuse of terminological and LLOD resources also using the dedicated API and query functionality
- **A Neurolinguistic and Psycholinguistic Data Curation Tutorial**, aimed at researchers who need to curate and deposit in ILC4CLARIN

All of the training materials will be aimed at producers to safely and FAIRly deposit their resources and they will be available in two formats for two different ways of fruition:

- as self learning modules directly implemented in our **Virtual Training Environment**
- as reusable training modules for trainers and lecturers to be shared via the **FAIR and Interoperable Training Materials Publication Platform**

Moreover, inspired by the [CLARIN café series](#), as well as by the [Open Science cafés](#) organised by the [Competence Centre ICDI](#), we will host a number of **training and outreach webinars**, dedicated to showcasing CLARIN-IT's resources and tools and its contribution to H2IOSC. These events will have a particular focus on the **training of trainers and lecturers**, and will provide **best practices and guidelines** on how to reuse the training materials.

As previously mentioned, CLARIN-IT content is already present in the curricula of summer schools and courses in Italy, thanks also to members of our consortium. We plan to expand this offer, thanks also to the availability of the new training materials, in particular to:

- **BA and Master programmes** in Italian institutions
- **PhD training** programmes
- **Summer Schools** organised in Italy (extending the coverage beyond what is currently offered by CLARIN-IT)
- Training programmes for **Data Stewards** (which are currently being developed at various universities¹³)
- **Tutorials at relevant conferences** in the domain, such as in particular [AIUC](#) and [CLIC](#)

A particular attention will be given to contents focussing on **new and disruptive technologies linked to AI and large language models**, which are likely to change the job market in the coming years. It is crucial that students in the humanities are trained to understand and use

¹³ Possible collaborations with the [University of Bologna Data Stewardship programme](#) have already been discussed within the framework of joint EOSC task forces collaborations.

such technologies, also in view of **filling the gap in skills that are expected in the job market** for the coming years. This aspect has proved crucial in already established projects such as UPSKILLS (Section 1.3.1), which could be taken as a model in terms of structure and technology while remodelling existing materials for a wider range of specialists than the students in the BA degree programs for which they were designed.

In particular, among new professions that are likely to be in demand and that would benefit from training by H2IOSC and CLARIN-IT we can list the following:

- **Language data analyst**
- **Language data manager**
- **Language data scientist**
- **Language project manager**

We will also liaise with the **CLARIN Training Officer**, increase the participation of Italian researchers to the CLARIN [Trainers network](#) and work towards the training and upskilling of PhD students also with the contribution of K-centres mobility grants.

2.4.3 TEACH DARIAH, TEACH WITH DARIAH

DARIAH-IT aims to give a consistent contribution in supporting the data communities and the pool of researchers - at the different stages of their own career path - collaborating with the H2IOSC network to acquire and up-skill their digital competences, when regarding research work in the Arts and Humanities field. Special focus will be given to strategies of data integration in the GLAMs' sector and disciplines. In line with the ongoing development of Research Pilot in WP7, the two foreseen training lines will be: a) digital philology; b) heritage science data integration with historical, archival and librarianship data.

With the aim to improve and train the people in the RIs network who will have access to the testing phases and the use of our Digital Hubs, training materials and courses will be held about:

- data ingestion, with reference to LoD data published from the Italian Ministry of Culture and Cultural Institutes and stakeholders at a national level;
- data integration, and how to merge and align structured semantic data to the main reference ontologies (i.e., the CIDOC-crm model, listed as the SSHOCro¹⁴) for the HS sector, such as the: Europeana Data Model (EDM), Cultural-ON, ArCo - Architettura

¹⁴ Bekiari, Chrysoula, Kritsotaki, Athina, & Tsouloucha, Eleni. (2020). SSHOC D4.18 SSHOC Reference Ontology (beta version) (v1.0). Zenodo. <https://doi.org/10.5281/zenodo.3744861>.

della Conoscenza, Ontologie del Sistema Archivistico Nazionale (SAN), EAC CPF (Encoded Archival Context Corporate Bodies, Persons and Families), Ontology of Archival Description (OAD), FRBR-aligned Bibliographic Ontology, MADS (Metadata Authority Description Schema), MODS (Metadata Object Description Schema), DBPedia ontology, Friend Of A Friend (foaf), Geonames Ontology;

- creation of vocabularies, taxonomy, thesauri for the HS and CH objects description;
- data semantization, annotation, and metadata of images datasets IIIF-compliant (<https://iiif.io/api/>);
- fostering semantic interoperability of cataloguing standards and research data;
- opening and sharing of cultural research facilities;

Training Strategy:

- upload of custom tools manual (to be also accessed offline);
- online guides to the available H2IOSC toolboxes supporting Memory, Cultural, Research institutions and citizen scientists with an interest in historical sources from archives, libraries, and museum collections in the creation of a FAIR digital ecosystem

2.4.4 TEACH E-RIHS, TEACH WITH E-RIHS

Within E-RIHS, the overall approach and the main objectives of HS Academy are related to informing potential users of IPERION HS and E-RIHS TNA services about the available scientific instruments, on one side, and to providing the most advanced knowledge on specific subjects of interest to the Heritage Science community, on the other side (Strlic et al., 2020). To address the needs of a wider audience, represented by the interdisciplinary H2IOSC community, HS Academy will customise its training offer and develop a new training strategy concerning the topics of interest, format of training activities, post-event analysis etc. The approach will be based on FAIR principles and will be adapted to the whole H2IOSC strategy in order to:

- address the specific learning needs of H2IOSC users audience
- adapt to the standardised format of H2IOSC training platform and training activities
- organise, promote and analyse KPIs of training activities in a way to comply with the standards of H2IOSC communities

Based on the aspects mentioned above, E-RIHS is expected to develop and deliver the customised training strategy at H2IOSC, in accordance with these 5 cross-cutting and interconnected strategic objectives:

- informing the communities about Italian E-RIHS node, services offered and other aspects
- informing the communities about services at other E-RIHS national nodes at the european level
- providing training about advanced digital tools available at E-RIHS, identified as of interest to the H2IOSC communities
- fostering interdisciplinary approach and continuous learning among researchers and their scientific domains around E-RIHS and all H2IOSC communities
- fostering contact and collaboration of emerging professionals with E-RIHS service providers, institutions and research groups of excellence

Types of E-RIHS training activities through the H2IOSC platform, can include a selection of already existing and new custom designed training modules. In particular, the HS Academy training strategy for H2IOSC can be based on:

1. an upgrade and adaptation of already existing E-RIHS Academy training activities, according to the identified needs of H2IOSC communities and other factors;
2. development of new activities that address the training needs of H2IOSC community and other factors.

With regards to **point 1**, *the existing training materials should be transformed into the digital resources* with the following characteristics:

- training materials digitised and equipped with metadata, following standardised H2IOSC approach
- training materials made Findable, Accessible, Interoperable, Reusable (FAIR), in compliance with the international guidelines on this matter.

With regards to **point 2**, *the new training materials for H2IOSC can be developed around the crucial topics of interest and training needs of the H2IOSC communities*. For instance, they can contain tutorials concerning E-RIHS Catalogue of Services and training modules regarding the services provided by E-RIHS Italian national node.

The specific content of proposed training activities will be discussed and proposed internally among E-RIHS members of H2IOSC. It can be finalised after consulting the survey and analysis of training needs performed in activity 2.4 and described in D2.1. The first report on the H2IOSC Landscapes [M12] is expected to contain the first version of the Landscape, with a first set of resources to be piloted (with common parts and RI specific sections).

2.4.5 TEACH OPERAS, TEACH WITH OPERAS

As the aim of OPERAS is the creation of a multi-layered Open scholarly communication system, its training strategy will reflect the different dimensions needed to support its goal, as represented by the work of the diverse [Special Interest Groups](#).

As its main training activity, OPERAS will invite tenders for the creation of training materials, conforming to the FAIR and Open principles, aimed at:

- introducing the logic and benefits of Open Science and Open Access
- promoting the creation of a common conceptual and terminological framework on Open Science in the European scholarly community
- presenting publishing opportunities in Open Access
- supporting researchers embracing Open Science and Open Access by exploring EU policies, funders' policies and publishing best practices
- introducing the core principles and methodologies of FAIR data management, as well as providing researchers with the competence to create FAIR-by-design research outputs
- offering specific competencies on citizen science
- offering practical knowledge on the use of OPERAS services

The training material will converge in the common H2IOSC training infrastructure.

At the same time, an internal survey within the different OPERAS nodes will be conducted to collect all the training materials produced outside of the common activities. Collected materials will be conformed to the FAIR standards, if needed, and included in the common training infrastructure.

Finally, OPERAS training strategy foresees the creation of a multimedia classroom, connected to similar structures in the network, for the in-presence training on Open Science and Open scholarly publications. Through the classroom OPERAS will offer courses that will also be preparatory and synergistic to the work of the other infrastructures, and that have the potential to reach a wider audience.

2.5 INTERNAL TRAINING & UPSKILLING

Internal training is the best way to build efficient and productive teams that positively reflects the importance given by partners to training for improving professional knowledge and skills of the group. **Identification of gaps and selection of appropriate training pathways** to fill them is essential. The following sections will detail both common initiatives and those targeting specific communities in each RI.

2.5.1 COMMON TRAINING NEEDS

A dedicated budget for the **training and upskilling of H2IOSC personnel** is present in all work packages and for all Participating Units. The dedicated **WP8 budget** will be used in particular for three purposes:

- **Transversal courses** that can benefit all participating units and infrastructures
- **Cross-IR training**, aimed at increasing the knowledge of services offered by national nodes of the four infrastructures across the consortium members
- **IR specific training**, with a particular focus on **Open Science** and **FAIRification**, in a trainers perspective

Based on the common training needs, we envisage the following modules **could benefit a number of H2IOSC dedicated personnel from all units**:

- **training on research infrastructure management** for DH and CH (specifically oriented to the creation of **data management plans**)
- training on **FAIR principles** and **Open Science**
- training on **Infrastructure Management**
- **ethical and legal issues**, with a specific focus on **data protection and data subjects protection** for the data types that are relevant to H2IOSC

In addition to this, each infrastructure could contribute with training modules dedicated to its services (both at a national and international level) either by offering them in kind to the consortium, or by inviting experts from their respective networks.

A dedicated training event covering these common courses is envisaged in the second year of the project, and a dedicated budget is set aside for its organisation.

2.5.2 CLARIN

Beyond the common training plan, **H2IOSC personnel** dedicated to the development of CLARIN related tools and activities will benefit from **specific training on the following topics**:

- basic **programming in python and creation of Notebooks**, with a specific focus on their use in teaching and learning
- development, fine tuning, testing and use of **large language models (LLMs), including the management of the related resources** (such as prompt datasets)
- dedicated training on **legal and ethical** issues related to the management of speech data, as well as **psycholinguistic and neurolinguistic data** collected from humans

Modules on the aforementioned topics could be delivered from experts within the CLAIRN network to H2IOSC personnel.

2.5.3 DARIAH

Training activities for DARIAH concern:

- Training on **specific services** offered by DARIAH ERIC
- Specific training on **digital lexicography**
- Web based **lexicography of ancient Italian**
- Training on **use of digital resources for lexicographic and philological disciplines**

These activities are implemented with the other IR for the common points, aiming for a **synergic collaboration that will result in more extensive and in-depth training**.

2.5.4 E-RIHS

Internal training and upskilling of H2IOSC staff, with regards to E-RIHS, should cover important **methodological and technical aspects of heritage science analytical data treatment, storage and use**. The topics can be developed on the basis of human capital enhancement needs for heritage science, as identified and reported by [SHINE-StrengtHening the Italian Nodes of E-RIHS](#) project:

- **digitization of cultural heritage** in its different facets, methodologies and procedures
- **principles and operation of advanced software digitalization tools** (GIS, 3d modelling, virtual world design, etc)

- **tools of intelligent data processing** (Semantics, AI, IoT, etc.)
- **data storage management**
- technical aspects of **software and hardware tools** related to E-RIHS services

In addition, the internal staff working on the creation of training activities for the interconnected H2IOSC communities, could benefit from specific training on how to produce digital resources for training, making them FAIR and equipped with metadata. This aspect is essential for valorization of the available materials and successful delivery of external training activities through the H2IOSC platform.

2.5.5 OPERAS

Beyond the common training plan, OPERAS staff involved in training activities should benefit from specific training on:

- how to design and produce FAIR resources
- how to make a resource FAIR-compliant (data FAIRification)
- principles and methodologies of Open Sciences
- principles, methodologies and good practices in Open scholarly publications
- e-learning methodologies and technologies
- use of the multimedia classroom

3. MONITORING TRAINING AND OUTREACH ACTIVITIES

Outreach activities will be carried out in close cooperation with WP1.3 (Training and engagement Plan Developed) and following the methodology and monitoring structure delivered by D1.3. As indicated in Section 1.2.1 of this document, this methodology is inspired by the guidelines provided by the [ESFRI landmark monitoring framework](#).

Both jointly and per infrastructure we will consider in particular the following results as **parameters to monitor the progress** of the initiatives:

- number of **training activities** carried out
- number of **modules** produced and published
- number of **outreach events**
- total number of **users** targeted with breakdown per activity
- total number of **post-event questionnaires** received
- number of **accesses** to online training materials
- **presence** at Summer Schools and relevant events for our community
- number of **courses** entered in the DHCR

4. LONG TERM PERSPECTIVES AND CONCLUSION

The sustainability plan for the training infrastructure and related training activities will be aligned with the long term strategies for the maintenance of the project's knowledge, outputs and services **for at least 10 years after the end of the funding period** developed in WP1, specifically in D1.4 (H2IOSC Sustainability Financial Plan).

Starting from this premise, the working group of WP8 decided to design, from this first edition of Training Strategy, **a long term plan** for training, capacity building and engagement framework. We define **a list of strategic objectives** not only for the duration of the H2IOSC project, but also and especially those related to the second phase, in which **our effort will be aimed to give rise to further development of the H2IOSC infrastructure**.

In addition to the target audience outlined in WP2 and prioritised in the first 24 months of the project (Section 1), we plan to **extend the training activities to a more general public**, including:

- **professionals** interested in upgrading their skills within the SSH domain

- **teachers outside universities**, especially high school professors
- **cultural industries**
- **industries and professionals** involved in the **development of new technologies**, especially those related to **language and AI**
- **Cultural Heritage specialists**
- **Galleries, Libraries, Archives and Museums** curators
- **citizens** invested in their personal education

This program, therefore, will increase the consciousness, the knowledge and the abilities of figures already belonging to the world of research and will also lead to the creation of new professional figures, needed in a national and international landscape increasingly characterised by IR.

REFERENCES

Benassi, Laura, Strlič, Matija, Chaban, Antonina, Bertasa, Moira, Di Gianvincenzo, Fabiana, Rijavec, Tjaša, Novotný, Jakub, & Striova, Jana. (2023). IPERION HS D8.3 Report on existing RI centralised procedures (1.2). Zenodo. <https://doi.org/10.5281/zenodo.7745729>

Bernardini, Silvia, and Maja Miličević Petrović. 'Toward a New Profile for Twenty-First Century Language Specialists: Industry, Institutional and Academic Insights', 25 June 2021. <https://zenodo.org/record/5030873>.

Directorate-General for Research and Innovation (European Commission), EOSC Executive Board, Natalia Manola, Emma Lazzeri, Michelle Barker, Iryna Kuchma, Vinciane Gaillard, and Lennart Stoy. *Digital Skills for FAIR and Open Science: Report from the EOSC Executive Board Skills and Training Working Group*. LU: Publications Office of the European Union, 2021. <https://data.europa.eu/doi/10.2777/59065>.

Engelhardt, Claudia, Raisa Barthauer, Katarzyna Biernacka, Aoife Coffey, Ronald Cornet, Alina Danciu, Yuri Demchenko, et al. 'How to Be FAIR with Your Data', 2022. <https://doi.org/10.17875/gup2022-1915>.

Garcia, Leyla, Bérénice Batut, Melissa L. Burke, Mateusz Kuzak, Fotis Psomopoulos, Ricardo Arcila, Teresa K. Attwood, et al. 'Ten Simple Rules for Making Training Materials FAIR'. *PLOS Computational Biology* 16, no. 5 (21 May 2020): e1007854. <https://doi.org/10.1371/journal.pcbi.1007854>.

Gledić, Jelena, Stavros Assimakopoulos, Iva Buchberger, Jelena Budimirović, Maja Đukanović, Tihana Kraš, Martina Podboj, Nađa Soldatić, and Michela Vella. 'UPSKILLS Guidelines for Learning Content Creation', 20 September 2021. <https://zenodo.org/record/8302296>.

Gledić, Jelena, Maja Đukanović, Maja Miličević Petrović, Iulianna van der Lek, and Stavros Assimakopoulos. 'Survey of Curricula: Linguistics and Language-Related Degrees in Europe', 25 June 2021. <https://zenodo.org/record/5030861>.

Group, FAIR Data Maturity Model Working. 'FAIR Data Maturity Model. Specification and Guidelines', 25 June 2020. <https://doi.org/10.15497/rda00050>.

Lek, Iulianna van der, Darja Fišer, Tanja Samardzic, Marko Simonovic, Stavros Assimakopoulos, Silvia Bernardini, Maja Milicevic Petrovic, and Genoveva Puskas. 'Integrating Research Infrastructures into Teaching: Recommendations and Best Practices', 31 August 2023. <https://zenodo.org/record/8114407>.

Martinez-Ortiz, Carlos, Daniel S. Katz, Anna-Lena Lamprecht, Michelle Barker, Axel Loewe, Anne Fouilloux, Jane Wyngaard, et al. 'FAIR4RS: Adoption Support', 24 February 2022. <https://doi.org/10.5281/zenodo.6258366>.

Mazzeo, Rocco, Campbell, Jill, Cassar, JoAnn, Furche, Andreas, Gibson, Adam, Kilian, Ralf, Leisner, Johanna, Martinez, Daniel Garcia, Muscat Azzopardi, Elizabeth, Sier, Mark, Strlič, Matija, Striova, Jana, & Benassi, Laura. (2021). D7.1 IPERION HS Training Plan (1.0). Zenodo. <https://doi.org/10.5281/zenodo.5788561>

Miličević Petrović, Maja, Jelena Gledić, and Maja Đukanović. 'Overview of Learning Materials Aimed to Educate Students of Languages and Linguistics for the 21st Century Job Market', 22 February 2021. <https://zenodo.org/record/6186227>.

Strlic, Matija, Buchczyk, Magdalena, Heritage, Alison, Liang, Haida, Ropret, Polonca, & Vopálenský, Michal. (2020). E-RIHS PP D7.2 E-RIHS Training strategy (1.0). Zenodo. <https://doi.org/10.5281/zenodo.3949504>

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 'The FAIR Guiding Principles for Scientific Data Management and Stewardship'. *Scientific Data* 3, no. 1 (15 March 2016): 160018. <https://doi.org/10.1038/sdata.2016.18>.

ANNEXES

Annex 1 - Overview of training initiatives

The following overview summarises **training initiatives implemented by each RI involved in H2IOSC** as mentioned in Section 1.3. The initiatives at ERIC level highlighted in blue.

CLARIN	DARIAH	E-RIHS	OPERAS
CLARIN @Universities Workshop	DARIAH Teach	IPERION HS Doctoral Summer School	Participation in Skills4EOSC
CLARIN in the Classroom	DARIAH Campus	IPERION HS Training Camp	Participation in the TRIPLE project
Teaching with CLARIN award	PARTHENOS Training	IPHS webinar series	TRIPLE Open Science Training Series
Introduction to Language Data: Standard and Repositories (UPSKILLS course)	Lecturing for the Master Course in Informatica del Testo ed Edizione Elettronica dell'Università di Siena	Monthly lecture series Current Topics in Heritage Science of the IPERION HS Academy	TRIPLE Training Toolkit
Digital Humanities Course Registry		User Meeting events	
Digital tools for Humanists (Summer School)			
Digital Humanities and Digital Communication: AI and (new) literacies (Summer School)			
Digital Text Processing Training School			

Permanent Seminar Series “A bridge between two worlds”			
--	--	--	--