



## Learning Materials as FAIR Digital Objects: a Shared Methodology for Training in H2IOSC

---

Giulia Pedonese, Francesca Frontini, Roberta Ottaviani,  
Jana Striova, Alessia Spadi, Laura Benassi, Federico Boschetti,  
Alessia Scognamiglio, Pietro Restaneo, Antonina Chaban and  
Lucia Francalanci

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

April 27, 2024

# Materiali didattici come oggetti digitali FAIR: una metodologia condivisa per la formazione in H2IOSC

Giulia Pedonese<sup>1</sup>, Francesca Frontini<sup>2</sup>, Roberta Ottaviani<sup>3</sup>, Federico Boschetti<sup>4</sup>, Alessia Spadi<sup>5</sup>, Lucia Francalanci<sup>6</sup>, Alessia Scognamiglio<sup>7</sup>, Pietro Restaneo<sup>8</sup>, Antonina Chaban<sup>9</sup>, Jana Striova<sup>10</sup>, Laura Benassi<sup>11</sup>

<sup>1</sup> Istituto di Linguistica Computazionale “Antonio Zampolli” (CNR-ILC), Italia – [giulia.pedonese@cnr.it](mailto:giulia.pedonese@cnr.it)

<sup>2</sup> Istituto di Linguistica Computazionale “Antonio Zampolli” (CNR-ILC), Italia – [francesca.frontini@cnr.it](mailto:francesca.frontini@cnr.it)

<sup>3</sup> Istituto di Linguistica Computazionale “Antonio Zampolli” (CNR-ILC), Italia – [roberta.ottaviani@cnr.it](mailto:roberta.ottaviani@cnr.it)

<sup>4</sup> Istituto di Linguistica Computazionale “Antonio Zampolli” (CNR-ILC), Italia – [federico.boschetti@cnr.it](mailto:federico.boschetti@cnr.it)

<sup>5</sup> Istituto Opera del Vocabolario Italiano (CNR-OVI), Italia – [alessia.spadi@cnr.it](mailto:alessia.spadi@cnr.it)

<sup>6</sup> Istituto Opera del Vocabolario Italiano (CNR-OVI), Italia – [lucia.francalanci@cnr.it](mailto:lucia.francalanci@cnr.it)

<sup>7</sup> Istituto per la Storia del Pensiero Filosofico e Scientifico Moderno (CNR-ISPF), Italia – [alessia.scognamiglio@cnr.it](mailto:alessia.scognamiglio@cnr.it)

<sup>8</sup> Istituto per il Lessico Intellettuale Europeo e Storia, delle Idee (CNR-ILIESI), Italia – [pietro.restaneo@cnr.it](mailto:pietro.restaneo@cnr.it)

<sup>9</sup> Istituto Nazionale di Ottica (CNR-INO), Italia – [antonina.chaban@cnr.it](mailto:antonina.chaban@cnr.it)

<sup>10</sup> Istituto Nazionale di Ottica (CNR-INO), Italia – [jana.striova@cnr.it](mailto:jana.striova@cnr.it)

<sup>11</sup> Istituto Nazionale di Ottica (CNR-INO), Italia – [laura.benassi@cnr.it](mailto:laura.benassi@cnr.it)

## ABSTRACT

Il presente lavoro dettaglia la strategia per lo sviluppo di iniziative di formazione nell’ambito del progetto H2IOSC e mira a coinvolgere la comunità italiana di riferimento sulle modalità di design e di fruizione di moduli didattici che integrino l’uso delle Infrastrutture di Ricerca. In particolare, il contributo si sofferma sulla descrizione dei requisiti per l’implementazione dell’infrastruttura di *training* e sugli standard condivisi per la descrizione dei materiali didattici come oggetti digitali FAIR al fine di massimizzarne il riutilizzo in un’ottica *train the trainers*.

## PAROLE CHIAVE

formazione; training; infrastrutture di ricerca; H2IOSC; principi FAIR

## 1. FORMAZIONE NELLE INFRASTRUTTURE DI RICERCA

Il progetto Humanities and cultural Heritage Italian Open Science Cloud (H2IOSC) [1] intende creare un cluster federato dei servizi e delle risorse sviluppate dai nodi nazionali di quattro Infrastrutture di Ricerca (IR) per la Scienza Aperta [2] che fanno parte della *roadmap* European Strategy Forum on Research Infrastructure (ESFRI)<sup>1</sup> nel settore dell’innovazione sociale e culturale: Digital Research Infrastructures for the Arts and Humanities (DARIAH); European Research Infrastructure for Heritage Science (E-RIHS), Common Language Resource and Technology Infrastructure (CLARIN) e Open Scholarly Communication in the European Research Area for Social Sciences and Humanities (OPERAS)<sup>2</sup>. Il presente contributo ha l’obiettivo di aprire la strategia di formazione sviluppata nell’ambito del progetto H2IOSC alla comunità scientifica in modo da favorire l’interscambio di obiettivi, metodologie e strumenti didattici volti ad integrare l’uso delle IR all’interno dei programmi di istruzione superiore, universitaria e professionale.

Per il funzionamento delle IR intese come «strutture, risorse e servizi correlati che sono utilizzati dalla comunità scientifica per condurre ricerche di alto livello nei rispettivi campi del sapere» [3] la formazione è un aspetto cruciale in quanto costituisce un ponte fra le comunità scientifiche e ciò che le IR offrono, risultando strettamente connessa alle attività di disseminazione che promuovono l’uso di prodotti e servizi da parte di ricercatori, studenti e cittadini. Più in generale, le IR contribuiscono all’istruzione fornendo una formazione specializzata a studenti, ricercatori e professionisti

<sup>1</sup> <https://roadmap2021.esfri.eu/landscape-analysis/section-1/social-cultural-innovation/>

<sup>2</sup> H2IOSC rientra nell’ambito del PNRR Missione 4, “Istruzione e Ricerca” - Componente 2, “Dalla ricerca all’impresa” - Linea di investimento 3.1, “Fondo per la realizzazione di un sistema integrato di infrastrutture di ricerca e innovazione”, finanziato dall’Unione europea – programma NextGenerationEU, di cui al Decreto Direttoriale MUR n. 0003264 del 28/12/2021 codice progetto IR0000029, CUP B63C22000730005. Sito ufficiale: <https://www.h2iosc.cnr.it/>

del mondo accademico e industriale su metodi e tecnologie scientifiche all'avanguardia. Per questo le iniziative di formazione hanno da sempre un posto di primo piano attraverso l'organizzazione di eventi come seminari, *workshop* e *summer school*, l'erogazione di borse di studio e di mobilità e la creazione di archivi digitali attraverso cui gli utenti possono accedere al materiale didattico, indicizzato secondo standard condivisi in quanto oggetto digitale in aderenza ai principi FAIR [4]. A questo proposito si ricordano non solo le iniziative di infrastrutture disciplinari come CLARIN (con il suo Learning Hub) e DARIAH (con la piattaforma DARIAH-Campus<sup>3</sup>), che forniscono l'accesso a risorse formative quali registrazioni di eventi, slide deck e tutorial indicizzati per tipologia di argomento o evento, ma anche le risorse dedicate al *training* da grandi progetti cluster europei fra i quali si segnala il Training Discovery Toolkit<sup>4</sup> sviluppato nell'ambito del Social Sciences and Humanities Open Cloud<sup>5</sup> (SSHOC). Inoltre, per potenziare le competenze trasversali degli utenti in termini di principi FAIR, Scienza Aperta e gestione dei dati della ricerca, il progetto Skills for the European Open Science commons: creating a training ecosystem for Open and FAIR science (Skills4EOSC) finanziato dal programma Horizon Europe della Commissione Europea e coordinato dal Consorzio GARR, ricopre un ruolo d'avanguardia nello sviluppo di metodologie, attività e risorse condivise per unificare la formazione dei professionisti della ricerca che costituiranno il futuro bacino d'utenza dello European Open Science Cloud<sup>6</sup> (EOSC). Il progetto H2IOSC si inserisce pienamente in questo contesto dedicando un intero *work package* (WP) alla formazione e basandosi sulle iniziative preesistenti per la Scienza Aperta e la gestione di dati FAIR per concentrarsi sull'ambito delle Social Sciences and Humanities (SSH) con casi d'uso specifici per le comunità afferenti.

## 2. LA STRATEGIA DI TRAINING H2IOSC

Il WP dedicato alla formazione è rivolto alle comunità di utenti italiani con lo scopo di dotarli di capacità e competenze interdisciplinari specifiche del settore SSH e formarli sulle risorse che le infrastrutture disciplinari possono offrire a livello nazionale e internazionale. Parte di questo obiettivo è quello di formare nuove figure professionali in grado di formare a loro volta le future generazioni su come integrare le IR nei metodi e nelle pratiche di ricerca delle rispettive discipline, con una prospettiva *train the trainers*, da cui l'esigenza di disporre di materiali didattici modulari e facilmente riutilizzabili da poter essere integrati nei propri corsi di formazione a seconda delle necessità. Le unità operative coinvolte nel WP hanno sviluppato una strategia condivisa per il coordinamento delle attività di *training*, sia a livello di singole IR che di progetto, che comprende innanzitutto l'individuazione dei bisogni formativi delle diverse tipologie di utenti, l'implementazione di un'infrastruttura per il *training* che metta a disposizione una piattaforma per l'erogazione diretta di moduli formativi (sia in modalità sincrona che asincrona) e una per il deposito dei materiali didattici e infine lo sviluppo di una metodologia comune per il design e l'adattamento di tali materiali.

Le tipologie di utenti oggetto della formazione sono state identificate in: ricercatori, esperti e all'inizio della loro carriera, con la necessità di integrare le proprie competenze allo scopo di gestire correttamente i dati della ricerca; tecnologi che hanno bisogno di sviluppare dataset e strumenti da aprire alla comunità secondo i dettami della Scienza Aperta; docenti e formatori di ambito universitario e professionale che possano integrare l'uso delle IR nei loro curricula di insegnamento; studenti di corsi di laurea e di dottorato, che devono essere esposti il prima possibile alle buone pratiche della comunità scientifica; professionisti a supporto della ricerca come *data stewards* e *data curators* per cui è necessario trasmettere le proprie competenze a diversi tipi di utenti; e infine i professionisti delle istituzioni culturali (archivi, biblioteche, musei), che possono beneficiare dei servizi offerti dalle IR. Per sondare le necessità formative dei vari gruppi, il WP dedicato al training lavora in stretto contatto con il WP dedicato all'analisi delle comunità di riferimento, che ha sviluppato un questionario e lo ha somministrato in via preliminare a una selezione di utenti. Inoltre, sono state implementate misure per il coordinamento con le varie associazioni, con cui a breve saranno presi i primi contatti.

L'infrastruttura di *training* diventerà una componente fondamentale del Marketplace H2IOSC e sarà composta da due piattaforme separate: un Virtual Training Environment, cioè un *learning management system* per l'erogazione di corsi in modalità sincrona e asincrona che permetterà ad H2IOSC e alle IR coinvolte di offrire formazione diretta sia al proprio

---

<sup>3</sup> I materiali di *training* sviluppati da DARIAH-EU sono disponibili su <https://campus.dariah.eu/>. All'interno del framework, le collezioni esterne sono cercabili tramite lo strumento DARIAH Pathfinders <https://campus.dariah.eu/source/dariah-pathfinders/page/1>

<sup>4</sup> [https://training-toolkit.sshopencloud.eu/entities?search=&f%5B0%5D=content\\_type%3Asource](https://training-toolkit.sshopencloud.eu/entities?search=&f%5B0%5D=content_type%3Asource)

<sup>5</sup> <https://sshopencloud.eu/>

<sup>6</sup> <https://eosc-portal.eu/>

personale che alle comunità di utenti; e la piattaforma FAIR and Interoperable Training Materials Publication Platform allo scopo di conservare, condividere, citare e riutilizzare i materiali formativi come oggetti digitali FAIR dotati di licenza e di un sistema di versionamento, in modo da creare una comunità di formatori che condividano e riutilizzino i rispettivi materiali all'interno dei programmi accademici e di formazione. Questi requisiti tecnici e funzionali sono stati tradotti in documenti di specifica destinati a bandi pubblici per l'implementazione di entrambe le piattaforme, le cui procedure di aggiudicazione si sono appena concluse.

Quanto allo sviluppo di materiali di *training*, si è deciso di dedicare una prima fase all'adattamento di moduli didattici preesistenti, originati dalle iniziative di formazione già in atto nelle IR coinvolte, alle necessità degli utenti di H2IOSC. In un primo momento, l'erogazione di questi moduli sarà dedicata alla formazione del personale interno delle IR coinvolte in una prospettiva di potenziamento delle competenze, ma sono in programma proposte di disseminazione come, ad esempio, la partecipazione del nodo italiano di CLARIN alla *summer school* Digital Tools for Humanists<sup>7</sup> organizzata dall'Università di Pisa sotto la direzione del Laboratorio di Cultura Digitale. Successivamente, è previsto il potenziamento dell'offerta formativa di ciascuna IR con la creazione di nuovi corsi specifici per la propria comunità di riferimento e infine verranno messe in atto iniziative congiunte per consolidare le competenze trasversali del bacino di utenza H2IOSC e formazione specifica per i membri del consorzio affinché possano a loro volta diventare formatori esperti nelle buone pratiche della Scienza Aperta e dei principi FAIR. A coordinamento delle operazioni, è stata sviluppata una metodologia di adattamento e design dei materiali didattici come oggetti digitali FAIR che verrà dettagliata nella prossima sezione.

### **3. MATERIALI DIDATTICI FAIR: UNA METODOLOGIA CONDIVISA**

Consapevoli che la produzione ed erogazione diretta di moduli didattici potrà soddisfare solo in parte le esigenze della comunità, H2IOSC vede nella costruzione di un ecosistema di FAIRificazione e condivisione di materiali didattici per il potenziamento e la trasmissione di competenze digitali una missione importante, anche al fine valorizzare le molte delle iniziative già messe in campo da AIUCD. Questa sezione illustra in particolare i principi alla base dello sviluppo della Publication Platform, dove andranno depositati dapprima tutti i materiali sviluppati dalle realtà coinvolte in H2IOSC e dai partner ed in seguito anche materiali della comunità scientifica, al fine di massimizzarne il riutilizzo. Secondo la strategia dettagliata nella sezione precedente, lo sviluppo dei moduli didattici prenderà le mosse da una prima fase di raccolta e sistematizzazione di materiali preesistenti (*slide deck*, infografiche, materiale audiovisivo, collezioni digitali ecc.). Ciò comporta l'ideazione di un flusso di lavoro capace sia di allineare i materiali con l'applicazione dei principi FAIR che di integrare moduli già concepiti come FAIR, ma spesso descritti secondo standard differenti, in formati diversi e con licenze tra loro incompatibili. Per questo, nella definizione di una metodologia condivisa sono state definite pratiche comuni per la descrizione dei materiali con metadati standard, l'applicazione di licenze adatte al riuso e l'attribuzione di identificativi univoci e persistenti (PID) che ne permettano la corretta citazione e attribuzione ai rispettivi autori.

L'unità minima modulare di riferimento è stata individuata nella singola lezione descritta in accordo con la definizione di *learning object* come un pacchetto di una lezione, un'attività e una valutazione con un singolo obiettivo di apprendimento e un risultato di apprendimento concreto adottata dal progetto Skills4EOSC [5]. In questo modo la lezione, accompagnata da opportuni metadati, può essere utilizzata come base per la creazione di contenuti didattici più complessi. I metadati per descrivere ogni singolo *learning object* sono stati individuati nel Minimal Metadata Set for Learning Resources proposto dalla Research Data Alliance [6] che stabilisce 14 campi suddivisi in 3 categorie di informazioni (*descriptive, access, educational*) ed è attualmente in fase di valutazione presso progetti come OpenPlato<sup>8</sup>, SSHOC Training Discovery Toolkit e NI4OS Training Platform<sup>9</sup>. Si tratta di uno schema flessibile, aperto ad eventuali integrazioni a seconda delle necessità connesse, ad esempio, alle diverse prospettive di insegnamento formali, professionali e informali, ed è volto a massimizzare la cercabilità dei dati senza appesantire il sistema dei descrittori, garantendo conformità e riusabilità del materiale esistente. In una prospettiva FAIR dal punto di vista dei formatori, i metadati più importanti sono quelli relativi alle parole chiave, che possono connettersi a vocabolari condivisi specifici per il settore disciplinare e quindi rendere il materiale facilmente trovabile; alle licenze utilizzate, che sono alla base delle possibilità di accesso e riutilizzo; e infine i

---

<sup>7</sup> <https://digitaltools.labcd.unipi.it/>

<sup>8</sup> <https://openplato.eu/blocks/catalog/list.php>

<sup>9</sup> <https://training.ni4os.eu/>

campi connessi a caratteristiche didattiche specifiche come *target group*, *learning outcome(s)* ed *expertise level*, che permettono di valutare la coerenza della lezione con le proprie finalità didattiche ed eventualmente pianificarne l'adattamento.

In questa fase del lavoro si è deciso inoltre di optare per la conversione del materiale preesistente nei formati più aperti disponibili, in modo da garantire la massima libertà e flessibilità di riutilizzo. Il formato non proprietario più versatile è sicuramente il plaintext, che verrà preferito laddove possibile, ad esempio rispetto a formati PDF (.ppt) o Moodle (.mbz). Per il design di nuovo materiale, che si tratti di creazione ex novo o di una rielaborazione di moduli preesistenti, inclusi quelli convertiti in plaintext, si prefigura l'utilizzo di Markdown come linguaggio open-source estensibile, intuitivo e facilmente convertibile in HTML, JSON, XML, YAML e molti altri formati. Lo stesso principio sarà alla base della scelta delle licenze<sup>10</sup>, che saranno più aperte possibile in considerazione delle regolamentazioni vigenti per la tipologia di dati in oggetto e della compatibilità delle licenze applicate al materiale progressivo. Un aspetto importante per facilitare il lavoro dei formatori e promuovere l'integrazione e il riutilizzo di *learning objects* è costituito dalle modalità di citazione dei materiali riutilizzati. A questo scopo, H2IOSC ha in progetto di adottare pratiche formalizzate e condivise per permettere la citazione e il riutilizzo dei propri materiali didattici, sia originali che derivati, considerati in tutto e per tutto come prodotti della ricerca. Fondamentale in questo senso sarà anche l'interscambio con le comunità scientifiche ed in particolare con AIUCD, da sempre sensibile a queste tematiche, per la definizione di modalità di citazione, riuso e deposito chiare e di facile implementazione.

#### 4. PROSPETTIVE FUTURE E CONCLUSIONI

Il passo successivo comporterà l'applicazione di tale strategia ai corsi di formazione specifici per ogni IR coinvolta a partire dal materiale già presente, fra cui: il corso "Introduction to Language Data: Standards and Repositories" sviluppato da CLARIN nell'ambito del progetto UPSKILLS<sup>11</sup>; il corso "La lessicografia web based all'Opera del Vocabolario Italiano: metodi e strumenti per la ricerca filologica e linguistica", elaborato dal gruppo di lavoro di DARIAH-IT con tutorial sugli strumenti descritti, e i webinar<sup>12</sup> della serie didattica online "Current Topics in Heritage Science" che verranno adattati dal gruppo di lavoro di E-RIHS. In seguito, saranno sviluppati dei progetti pilota su servizi innovativi nell'ambito delle diverse IR: questi prototipi, sotto forma di piattaforme o *hub*, riuniranno servizi, flussi di lavoro e interfacce specifici per il dominio disciplinare di afferenza e saranno progettati in modo da essere scalabili ed estendibili con l'aggiunta di nuove risorse e dati, al fine di acquisire conoscenze e adeguare gli obiettivi man mano che l'implementazione procede. Alcuni tra i progetti pilota previsti sono, ad esempio i tutorial che saranno sviluppati da CLARIN per facilitare l'uso dei Linguistic Linked Open Data, per la cura e il deposito di dati di storia orale e per la curatela di dati neurolinguistici e psicolinguistici. Infine, vi saranno corsi dedicati all'uso dei servizi di H2IOSC come, ad esempio, il futuro Marketplace, e ad aspetti gestionali delle IR come l'Infrastructure Management per le Digital Humanities, con particolare focus sulla creazione di *data management plan* e aspetti legali ed etici connessi alla protezione dei dati digitali per le tipologie di dati rilevanti per H2IOSC.

Le potenzialità di una metodologia condivisa per il design e l'adattamento di materiali didattici come prodotti della ricerca sono molteplici, primi fra tutti il superamento di gap metodologici e terminologici esistenti tra ambiti disciplinari diversi e l'agevolazione della formazione professionale nell'ambito di progetti nazionali e internazionali. Per questo, in accordo con i principi della Scienza Aperta, il presente contributo apre alla comunità la metodologia sviluppata in seno al progetto H2IOSC per l'adattamento e il design di materiali didattici come *learning objects* modulari e riutilizzabili in modo da massimizzare l'impatto delle iniziative di formazione che integrano l'uso delle Infrastrutture di Ricerca.

#### 5. RINGRAZIAMENTI

Progetto H2IOSC - Humanities and cultural Heritage Italian Open Science Cloud finanziato dall'Unione europea NextGenerationEU - Piano Nazionale di Ripresa e Resilienza (PNRR) - Missione 4 "Istruzione e Ricerca" Componente 2

---

<sup>10</sup> <https://creativecommons.it/chapterIT/index.php/license-your-work/>

<sup>11</sup> Il progetto UPSKILLS è stato un partenariato strategico Erasmus+ con lo scopo di identificare e affrontare le lacune nelle competenze degli studenti di linguistica attraverso l'integrazione dei curricula esistenti con materiali didattici di supporto: <https://upskillsproject.eu/>. Il materiale del corso è disponibile qui: [https://upskillsproject.eu/project/standards\\_repositories/](https://upskillsproject.eu/project/standards_repositories/)

<sup>12</sup> Si tratta di materiale originato da incontri mensili di formazione online su aspetti fondamentali dell'Heritage Science: <https://www.iperionhs.eu/lectures-series/>

“Dalla ricerca all’impresa” Linea di Investimento 3.1 “Fondo per la realizzazione di un sistema integrato di infrastrutture di ricerca e innovazione” Azione 3.1.1 “Creazione di nuove IR o potenziamento di quelle esistenti che concorrono agli obiettivi di Eccellenza Scientifica di Horizon Europe e costituzione di reti” - Codice progetto IR0000029 - CUP B63C22000730005. Soggetto attuatore CNR.

## BIBLIOGRAFIA

- [1] Degl’Innocenti, Emiliano, Monica Monachini, Alberto Bucciero, Enrico Pasini, Bruno Fanini, e Francesca Frontini. «H2IOSC: Humanities and Heritage Open Science Cloud». In *La memoria digitale: forme del testo e organizzazione della conoscenza. Atti del XII Convegno Annuale AIUCD*, a cura di Emmanuela Carbé, Gabriele Lo Piccolo, Alessia Valenti, e Francesco Stella, 63–64, 2023. <https://iris.unive.it/retrieve/0f226d38-e332-418b-9b14-d5558d1a0d9d/AIUCD2023.pdf>.
- [2] «UNESCO Recommendation on Open Science - UNESCO Digital Library». Consultato 1 dicembre 2023. <https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en>.
- [3] Frontini, Francesca, e Monica Monachini. «Infrastrutture digitali per le scienze umane e sociali». In *Digital Humanities. Metodi, strumenti, saperi*, 197–213. Roma: Carocci, 2023. <https://www.carocci.it/prodotto/digital-humanities>.
- [4] GO FAIR. «FAIR Principles». Consultato 12 gennaio 2024. <https://www.go-fair.org/fair-principles/>.
- [5] Filiposka, Sonja. «D2.2 Methodology for FAIR-by-Design Training Materials», 31 agosto 2023. <https://doi.org/10.5281/ZENODO.8305540>.
- [6] RDA. «Recommendations for a Minimal Metadata Set to Aid Harmonised Discovery of Learning Resources», 1 aprile 2022. <https://www.rd-alliance.org/group/education-and-training-handling-research-data-ig/outcomes/recommendations-minimal-metadata-set>.