

Il Glossario delle Infrastrutture di Ricerca (GIR)

Lucia Francalanci¹, Alessia Scognamiglio², Irene Falini³, Pietro Restaneo⁴, Giulia Pedonese⁵

¹ Istituto Opera del Vocabolario Italiano (CNR-OVI), Italia, lucia.francalanci@cnr.it

² Istituto per la Storia del Pensiero Filosofico e Scientifico Moderno (CNR-ISPF), Italia, alessia.scognamiglio@cnr.it

³ Istituto Opera del Vocabolario Italiano (CNR-OVI), Italia, irene.falini@cnr.it

⁴ Istituto per il Lessico Intellettuale Europeo e Storia delle Idee (CNR-ILIESI), Italia, pietro.restaneo@cnr.it

⁵ Istituto di Linguistica Computazionale "Antonio Zampolli" (CNR-ILC), Italia, giulia.pedonese@cnr.it

ABSTRACT (ITALIANO)

Il *Glossario delle Infrastrutture di Ricerca (GIR)* è un'iniziativa che si inserisce nell'ambito del progetto PNRR "H2IOSC - Humanities and Cultural Heritage Italian Open Science Cloud" del CNR, che mira a creare una federazione di Infrastrutture di Ricerca (IR) coinvolgendo i nodi italiani di quattro infrastrutture che fanno parte della *roadmap* dell'*European Strategy Forum on Research Infrastructure* nel settore dell'innovazione sociale e culturale: CLARIN-IT, DARIAH-IT, E-RHIS.it, OPERAS.it. Il GIR nasce dall'esigenza di avere una terminologia di riferimento all'interno del progetto H2IOSC. Nella redazione dei documenti legati al progetto è emersa la difficoltà legata alla mancanza di una terminologia univoca, condivisa e non ambigua, anche nell'ottica di una piena interoperabilità dei risultati della ricerca. Da qui la necessità di definire i termini che appartengono all'ambito della progettazione e implementazione delle IR, sciogliere i numerosi acronimi presenti, e documentare questo particolare linguaggio tecnico e specialistico. A partire da una rassegna critica sui più recenti studi teorici e applicati nel campo della terminologia della lingua italiana, del lessico delle IR, del PNRR e della progettazione europea, è stata individuata una metodologia per la costruzione del *Glossario* che prevede l'identificazione di un *corpus* di riferimento poi usato per la redazione delle voci. La struttura delle voci è stata progettata prendendo come riferimento alcuni vocabolari nati direttamente in rete e prevede, per ciascun lemma, l'indicazione della categoria grammaticale, la marca d'uso, le accezioni che il termine assume nel contesto delle IR, un campo "note" per eventuali integrazioni e le voci correlate, se presenti. Il GIR rappresenta un'iniziativa preziosa per promuovere la condivisione della conoscenza e la collaborazione tra le Infrastrutture di Ricerca in Italia.

Parole chiave: glossario; Infrastrutture di Ricerca; terminologia.

ABSTRACT (ENGLISH)

The *Glossary of Research Infrastructures (GIR)*. The proposal for a *Glossary of Research Infrastructures (GIR)* was conceived within the framework of the PNRR project 'H2IOSC - Humanities and Cultural Heritage Italian Open Science Cloud'. H2IOSC aims at creating a federation of Research Infrastructures (RI) among the Italian nodes of 4 infrastructures that are part of the roadmap of the *European Strategy Forum on Research Infrastructure*, in the field of social and cultural innovation: CLARIN-IT, DARIAH-IT, E-RHIS.it, OPERAS.it. The GIR has been developed to answer the need for a common reference terminology within the H2IOSC project. While drafting the H2IOSC documents, a difficulty emerged related to the lack of unambiguous shared terminology, that would contribute to achieving the full interoperability of the research results. Hence the need to define the terms belonging to the field of IR project management and implementation; to explain the numerous acronyms present; and to document such a technical and specialized language. Our paper will start from a critical review of the most recent theoretical and applied studies in the field of terminology of the Italian language, such as the RI lexicon and the PNRR and European planning. At first, we defined the methodology for the construction of the Glossary, which envisages the identification of a reference *corpus*. The *corpus* was then used to draft the entries. The structure of the entries was designed by taking as a reference some vocabularies born directly on the web. The grid provides, for each entry, an indication of the grammatical category, the brand of use, the meaning or meanings that the term takes on in the context of the RI, 'notes' field for possible additions and any related entries. The GIR is a valuable initiative to promote knowledge sharing and collaboration between Research Infrastructures in Italy.

Keywords: glossary; Research Infrastructures; terminology.

1. IL PROGETTO E GLI OBIETTIVI

Il *Glossario delle Infrastrutture di Ricerca (GIR)* è un'iniziativa che si inserisce nell'ambito del progetto PNRR "H2IOSC - Humanities and Cultural Heritage Italian Open Science Cloud" (finanziato dall'Unione

europea NextGenerationEU - PNRR M4C2 - Codice progetto IR0000029 - CUP B63C22000730005)¹ del Consiglio Nazionale delle Ricerche, che mira a creare una federazione di Infrastrutture di Ricerca (IR), coinvolgendo i nodi italiani di quattro infrastrutture che fanno parte della *roadmap* dell'*European Strategy Forum on Research Infrastructure* (ESFRI) nel settore dell'innovazione sociale e culturale: CLARIN-IT, DARIAH-IT, E-RIHS.it., OPERAS.it.² Un'Infrastruttura di ricerca (IR) comprende un insieme di strutture, risorse e servizi correlati tra loro, utilizzati dalla comunità scientifica - e non solo - per condurre ricerche di alto livello nei rispettivi campi e promuovere l'innovazione nei territori.³ Il Forum Strategico Europeo per le Infrastrutture di Ricerca (ESFRI), organo consultivo del Consiglio dell'Unione Europea per le IR, sottolinea l'importanza strategica che queste rappresentano per lo Spazio Europeo della Ricerca (ERA) e per l'innovazione dei territori dove sono incardinate.

Il gruppo di lavoro è composto da ricercatrici e ricercatori, tecnologhe e tecnologi afferenti al Dipartimento di Scienze Umane e Sociali e Patrimonio Culturale (DSU) del Consiglio Nazionale delle Ricerche, coinvolti come parte attiva in H2IOSC.⁴

Il GIR nasce dall'esigenza di avere a disposizione una terminologia di riferimento all'interno del progetto H2IOSC. Nella redazione dei documenti è emersa, infatti, fin da subito una difficoltà profonda legata alla mancanza di una terminologia univoca, condivisa e non ambigua, soprattutto nell'ottica di una piena interoperabilità dei risultati della ricerca. Non tutti i ricercatori e i tecnologi coinvolti nel progetto hanno una buona padronanza della terminologia specialistica, e la corretta comprensione dei testi, insieme ad una efficace comunicazione tra le IR, è imprescindibile e può essere ottenuta soltanto ricorrendo a lemmi di uso più comune o esplicitando il significato del lessico tecnico-scientifico. Ad oggi mancano risorse lessicografiche a cui poter fare riferimento per la comprensione di questo linguaggio, e gli stessi vocabolari specialistici non sempre registrano le accezioni tecniche che alcuni lemmi possono assumere nell'ambito delle Infrastrutture di Ricerca.

Da qui la necessità e l'obiettivo di costruire un glossario per definire i termini che appartengono a questo specifico ambito e sciogliere i numerosi acronimi presenti, in modo da documentare un linguaggio tecnico e scientifico che risulta ampiamente condiviso all'interno delle IR, e che è presente anche in altri ambiti istituzionali, come ad esempio quello accademico.

2. METODOLOGIA

A partire da una rassegna critica riguardo i più recenti studi teorici e applicati nel campo della terminologia della lingua italiana,⁵ del lessico delle IR,⁶ del PNRR⁷ e della progettazione europea, è stata individuata una metodologia per la costruzione del *Glossario* che ha previsto prima di tutto l'identificazione di un *corpus* di riferimento da utilizzare per la redazione delle voci. Il *corpus* è composto da testi settoriali in lingua italiana e inglese e comprende piani nazionali, linee guida, capitoli e altra documentazione tecnica.⁸

L'interrogazione del *corpus* avviene tramite TIGRO (Gestore Ricerche del Tesoro Italiano delle Origini), un *software* per l'analisi lessicografica e linguistica di *corpora* testuali, che si basa su GATTO (Gestione degli Archivi Testuali del Tesoro delle Origini), il programma che gestisce la versione elettronica dei *corpora* dell'Opera del Vocabolario Italiano (OVI).⁹ Nello specifico TIGRO consente di fare ricerche all'interno dei testi, di visualizzare i contesti e di mostrare le statistiche di frequenza delle forme. La documentazione richiesta dal PNRR è perlopiù in inglese, ma poiché ad essa si devono necessariamente affiancare anche

¹ <https://www.h2iosc.cnr.it/> (cons. 08/04/2025). Cfr. Degl'Innocenti *et al.*, 2023.

² Common Language Resource and Technology Infrastructure (CLARIN), Digital Research Infrastructures for the Arts and Humanities (DARIAH), European Research Infrastructure for Heritage Science (E-RIHS) e Open Scholarly Communication in the European Research Area for Social Sciences and Humanities (OPERAS).

³ Cfr. Regolamento (CE) n. 723/2009 del Consiglio del 25 giugno 2009, art. 2, lett. a).

⁴ Al *Glossario delle Infrastrutture di Ricerca* collaborano, oltre a chi scrive, Antonina Chaban (Istituto Nazionale di Ottica, CNR-INO), Roberta Ottaviani (Istituto di Linguistica Computazionale "Antonio Zampolli", CNR-ILC), Salvatore Prinzi (Istituto per la Storia del Pensiero Filosofico e Scientifico Moderno, CNR-ISPF), Alessia Spadi (Istituto Opera del Vocabolario Italiano, CNR-OVI).

⁵ Si veda ad esempio Chiocchetti & Ralli, 2022; Grimaldi & Zanola, 2021.

⁶ Cfr. Liburdi, Marras & Russo, 2019.

⁷ Si veda in particolare Cozzio, 2022; Vellutino, 2022.

⁸ In particolare, sono inclusi tra i documenti ufficiali AgID, 2006; PNRR, 2021; PNIR, 2021-2027; PNSA, 2021-2027; PNR, 2021-2027.

⁹ <http://www.oivi.cnr.it/> (cons. 08/04/2025).

documenti in italiano, è emersa la necessità di progettare e costruire un *corpus* bilingue. Di conseguenza, una delle criticità che il gruppo di lavoro sta affrontando è la lemmatizzazione multilingue, motivo per cui il *corpus* non possiede ancora annotazioni morfosintattiche (*Part of Speech Tagging*).

Il lemmario è attualmente composto dai tecnicismi più frequenti nel *corpus*, dai termini rappresentativi di ciascuna unità di lavoro del progetto (*work package* o WP), da altri lemmi non frequenti nella documentazione scritta ma particolarmente caratteristici del linguaggio delle Infrastrutture di Ricerca e ricorrenti nella comunicazione orale legata a tale ambito.

Il gruppo di lavoro è consapevole che, ad accezione dei termini più attestati, la selezione delle voci può sembrare arbitraria; essa, tuttavia, risponde alle esigenze comunicative espresse dalle ricercatrici e dai ricercatori impegnati nei progetti infrastrutturali. Date queste premesse, il lemmario del GIR può considerarsi un contenitore sempre aperto ad accogliere voci che di volta in volta si attestano nell'uso.

La struttura delle voci è stata progettata prendendo come riferimento alcuni vocabolari nati direttamente in rete, come il *Tesoro della Lingua Italiana delle Origini* (TLIO), il dizionario storico dell'italiano antico elaborato dall'OVI.¹⁰ La scheda lessicografica prevede, per ciascun lemma, l'indicazione della categoria grammaticale, la marca d'uso,¹¹ le accezioni che il termine assume nel contesto delle IR, un campo "note" per eventuali integrazioni, che possono essere di varia natura, e le voci correlate, se presenti. Le voci risultano così corredate da un set completo di metadati descrittivi. Le Fig. 1 e 2 illustrano alcuni esempi di schede del GIR:

FAIR

CATEGORIA GRAMMATICALE: acronimo usato come aggettivo invariabile.

MARCA D'USO: istruzione; Pubblica Amministrazione; ricerca.

DEFINIZIONE

Insieme dei requisiti che i dati e i risultati della ricerca devono rispettare per aderire a un modello di scienza aperta, ovvero essere *findable* 'rintracciabili' (devono essere reperibili, ossia descritti con metadati appropriati e identificativi univoci e persistenti), *accessible* 'accessibili' (devono essere resi accessibili ad altri e persistenti nel tempo), *interoperable* 'interoperabili' (devono poter essere integrati e utilizzati con altri dati o strumenti e in diversi contesti) e *reusable* 'riutilizzabili' (devono essere riutilizzabili e dunque ben descritti e documentati).

NOTE

L'acronimo inglese FAIR (*Findable, Accessible, Interoperable, Reusable*) viene coniato nel 2014, in occasione di un workshop Lorentz a Leida, nei Paesi Bassi. Negli anni successivi, un gruppo di esperti lavora a una lista di principi guida utili a garantire un uso ottimale dei dati della ricerca e nel 2016 vengono pubblicati ufficialmente sulla rivista «Nature» i 15 Principi FAIR (*FAIR Guiding Principles for scientific data management and stewardship*).

VOCI CORRELATE: *fairificazione, fairificare.*

Figura 1. Scheda della voce FAIR

¹⁰ <http://tlio.oivi.cnr.it/TLIO/> (cons. 08/04/2025).

¹¹ Posto che tutti i vocaboli del GIR sono legati a un uso marcatamente tecnico-specialistico, la marca d'uso indica gli ambiti specifici in cui ciascun termine risulta attestato, indipendentemente dall'accezione che assume nel contesto delle IR. La difficoltà nel rintracciare tutti i contesti d'uso dei tecnicismi sommata all'eterogeneità dei settori scientifici dei redattori rende inevitabile il rischio che l'indicazione della marca d'uso possa essere talvolta lacunosa.

INFRASTRUTTURA DI RICERCA (IR)

CATEGORIA GRAMMATICALE: locuzione sostantivale femminile.

MARCA D'USO: istruzione; Pubblica Amministrazione; ricerca.

DEFINIZIONE

Organizzazione che comprende strutture, risorse e servizi collegati che possono essere usati dalla comunità scientifica per condurre ricerche di alta qualità e promuovere l'innovazione nei rispettivi settori. Le Infrastrutture di Ricerca includono impianti, laboratori di ricerca, banche dati, strumenti e tecnologie scientifiche, risorse informatiche e reti di comunicazione. Possono essere utilizzate, oltre che per la ricerca, anche per l'istruzione e per i servizi pubblici e possono trovarsi in un'unica sede, essere distribuite in più sedi o essere virtuali.

NOTE

Si fa qui riferimento alla definizione proposta dalla Commissione europea (CE) nel Regolamento (CE) N. 723/2009 del Consiglio del 25 giugno 2009, art. 2, e a quella del Regolamento (UE) n. 1291/2013 del Parlamento Europeo e del Consiglio dell'11 dicembre 2013, art. 2, che istituisce il programma quadro di ricerca e innovazione (2014-2020).

Figura 2. Scheda della voce INFRASTRUTTURA DI RICERCA (IR)

3. LAVORI FUTURI

I lavori futuri prevedono l'ampliamento della bibliografia e la pubblicazione online sia del *Glossario delle Infrastrutture di Ricerca* sia del *corpus* di riferimento. Il passaggio successivo consisterà nella mappatura dei metadati attraverso SKOS (*Simple Knowledge Organisation System*), un linguaggio formale creato per rappresentare glossari, classificazioni, tassonomie e qualsiasi tipo di vocabolario strutturato. SKOS fornisce uno standard per la descrizione delle relazioni fra i sistemi di organizzazione della conoscenza nel quadro del web semantico. L'adozione di tale linguaggio aumenta a nostro avviso il valore del *Glossario*, trasformandolo in una risorsa di conoscenza interoperabile, accessibile e riutilizzabile.

Auspichiamo che il GIR e la banca dati di riferimento per la sua redazione possano rappresentare uno strumento utile per la comunità scientifica che ruota attorno alle IR, oltre che fungere da base per vari studi di tipo linguistico.

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 "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

Agenzia per l'Italia Digitale (AgID). 2006. Linee Guida recanti regole tecniche per l'apertura dei dati e il riutilizzo dell'informazione del settore pubblico. Art. 12 D.Lgs. n. 36/2006.

https://www.agid.gov.it/sites/default/files/repository_files/ig-open-data_v.1.0_1.pdf.

Chiocchetti, E., & Ralli, N. (a cura di). 2022. Risorse e strumenti per l'elaborazione e la diffusione della terminologia in Italia. Bolzano: Eurac Research.

Cozzio, M., (a cura di). 2022. Conoscere il PNRR. 150 parole chiave per capire regole, strumenti e funzionamento del Piano Nazionale di Ripresa e Resilienza. Santarcangelo di Romagna (RN): Maggioli Editore.

Degl'Innocenti, E., Monachini, M., Bucciero, A., Pasini, E., Fanini, B. & Frontini, F. (2023). H2IOSC: Humanities and Heritage Open Science Cloud. In Carbé, E., Lo Piccolo, G., Valenti, A., Stella, F. (a cura di), *La memoria digitale: forme del testo e organizzazione della conoscenza*. Atti del XII Convegno Annuale AIUCD (pp. 63-64). <https://iris.unive.it/retrieve/0f226d38-e332-418b-9b14-d5558d1a0d9d/AIUCD2023.pdf>

Grimaldi, C., & Zanola, M. T. (a cura di). 2021. Terminologie e vocabolari. Lessici specialistici e tesauri, glossari e dizionari. Firenze: Florence University Press.

- Liburdi, A., Marras, C., & Russo, A., (2019). Infrastrutture, terminologie e policy per la ricerca umanistica: note per un confronto interdisciplinare. Conferenza GARR 2018 - Data (R)evolution - Selected Papers, 82-86. DOI: 10.26314/GARR-Conf18-proceedings.
- Piano Nazionale di Ripresa e Resilienza (PNRR) 2021.
<https://www.governo.it/sites/governo.it/files/PNRR.pdf>
- Piano Nazionale Infrastrutture di Ricerca (PNIR) 2021-2027.
<https://www.mur.gov.it/sites/default/files/2021-10/Decreto%20Ministeriale%20n.1082%20del%2010-09-2021%20-%20PNIR%202021%20-%202027.pdf>.
- Piano Nazionale per la Scienza Aperta (PNSA) 2021-2027.
https://www.mur.gov.it/sites/default/files/2022-06/Piano_Nazionale_per_la_Scienza_Aperta.pdf
- Programma nazionale per la ricerca (PNR) 2021-2027. <https://www.mur.gov.it/sites/default/files/2021-01/Pnr2021-27.pdf>.
- Vellutino, D. (2022). Neologismi istituzionali per la ripresa e resilienza dell'Italia. Analisi tipologico-strutturale dei nuovi termini del PNRR. AIDAinformazioni, 3-4, luglio-dicembre, 113-14.