

Un cruscotto di monitoraggio della ricerca basato su dati aperti per l'Università degli Studi di Milano

Dario Basset (ORCID 0009-0005-2320-0341), Stefano Bolelli Gallevi (0000-0002-9680-0461)

Direzione Performance Assicurazione Qualità Valutazione e Politiche di Open Science Università degli studi di Milano (<https://ror.org/00wjc7c48>)

Abstract. Progetto di un cruscotto di monitoraggio pubblicamente accessibile, che presenti in forma grafica una serie di dimensioni selezionate relative alle attività di ricerca. Nella prima fase del progetto, il focus sarà sulle collaborazioni dei ricercatori dell'Università viste attraverso le loro pubblicazioni. Le fonti di dati utilizzate nel cruscotto sono due:

AIR - <https://air.unimi.it/> - Archivio Istituzionale della Ricerca, che è l'archivio di dati sulla produzione scientifica dell'Università degli Studi di Milano e costituisce di fatto l'Anagrafe della ricerca dell'Ateneo.

OpenAlex - <https://openalex.org/> - che è il catalogo di pubblicazioni scientifiche aperto e gratuito a maggior copertura, in grado di mettere a disposizione i metadati di più di 250 milioni di pubblicazioni.

Il progetto, sviluppato con la collaborazione attiva del dipartimento di informatica dell'Università degli Studi di Milano, è stato avviato ad aprile 2024 e la fase 1 sarà completata a settembre 2024

Keywords. OpenAlex, Open Science, Pubblicazioni, Collaborazioni.

Come altre istituzioni soprattutto internazionali, l'Università degli Studi di Milano sta esplorando la possibilità di monitorare la propria attività di ricerca basandosi esclusivamente su dati aperti. Questo appare necessario a fronte di sistemi proprietari costosi e fortemente sbilanciati verso certe aree geografiche, linguistiche e verso specifiche tipologie di pubblicazione.

La Direzione Performance Assicurazione Qualità Valutazione e Politiche di Open Science dell'Università degli studi di Milano ha quindi fatto partire un progetto per la realizzazione di un cruscotto pubblicamente accessibile che presenti in forma grafica una serie di dimensioni relative alle attività di ricerca. In particolare nella prima release del cruscotto si presenteranno le informazioni sulle collaborazioni dei ricercatori e professori dell'Università attraverso le loro pubblicazioni.

Il cruscotto vuole superare la prassi di monitoraggio delle attività di ricerca basata su database proprietari e a pagamento e poco personalizzabili, i cosiddetti Scientific Knowledge Graph, quali Scopus (<https://www.scopus.com>) e Web of Science (<https://www.we>

bofsience.com). Infatti, oltre al pagamento di una licenza d'uso l'utilizzo di tali strumenti presenta l'ulteriore limite di non poter diffondere i dati alla base delle analisi effettuate e dei risultati ottenuti, rendendo così impossibile un efficace monitoraggio comparativo.

La scelta di utilizzare al posto di questi servizi OpenAlex offre notevoli vantaggi in termini di inclusività, accessibilità, disponibilità e riusabilità dei dati. La copertura è circa doppia rispetto a WOS e Scopus e comprende un maggior numero di opere in lingue diverse dall'inglese e del Sud globale. Anche per queste ragioni il ranking di Leiden lo ha utilizzato come fonte per l'edizione del 2024 (questa edizione ha ancora, parallelamente, la versione basata su dati proprietari, ma a partire dal 2025 ci sarà un passaggio alla sola versione basata su dati aperti).

OpenAlex offre un servizio veloce, moderno e ben documentato, ma soprattutto open, infatti OpenAlex è licenziato con licenza CC0, che consente il più ampio riutilizzo. Ispirato all'antica Biblioteca di Alessandria, OpenAlex è un indice di centinaia di milioni di entità interconnesse in un sistema di ricerca globale.

Con il cruscotto, quindi, si vuole ottenere un monitoraggio delle attività della ricerca della nostra istituzione totalmente trasparente e basato su strumenti del tutto aperti. In questa maniera, non ci saranno spese aggiuntive oltre lo sviluppo del cruscotto stesso, e i risultati del monitoraggio possono essere facilmente condivisi con tutta la comunità scientifica e chiunque sia interessato. Il monitoraggio incrocia OpenAlex (<https://openalex.org/>) e AIR (<https://air.unimi.it/>) come fonte dati e per la parte di rappresentazione grafica utilizza strumenti open source. La scelta è quindi aperta by default, come raccomandato dalla dichiarazione di Barcellona (<https://barcelona-declaration.org/>) di cui l'Università di Milano è uno dei primi firmatari in Italia

Il repository istituzionale AIR (<https://air.unimi.it/>) è la fonte dei dati anagrafici dei docenti e ricercatori in attività dell'Ateneo: per ciascun autore viene cercato in OpenAlex il corrispettivo utilizzando l'identificativo ORCID, e vengono poi scaricate tutte le pubblicazioni associate dell'ultimo quinquennio, dalle quali vengono poi ricavate le collaborazioni utilizzando i coautori e le loro affiliazioni; infine il risultato viene esposto online.

I dati in OpenAlex

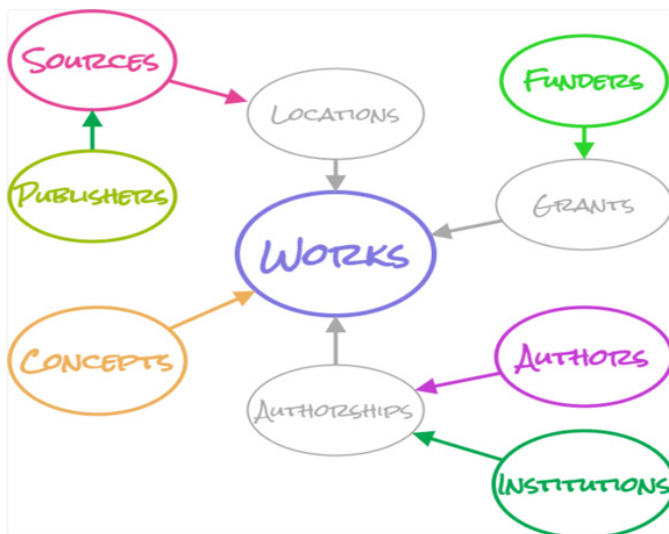


Fig. 1
<https://docs.openalex.org/api-entities/entities-overview>

La figura mostra i dati disponibili in OpenAlex.

Le entità in gioco sono di seguito descritte:

- Authors sono le persone che creano pubblicazioni (works)
- Works articoli, libri, dataset, tesi
- Funders Organizzazioni che prestano fondi alla ricerca
- Publishers Compagnie ed organizzazioni che distribuiscono le pubblicazioni
- Concepts Gli argomenti delle pubblicazioni
- Institutions Le istituzioni, cioè Università, enti di ricerca ed altre istituzioni
- Sources I giornali ed altri database che ospitano le pubblicazioni

Funzionalità della prima release

Le ricerche che possono essere effettuate sono:

- Collaborazioni (numero di autori, numero di paper) dell'Università degli Studi di Milano con altre istituzioni in tutto il mondo, raggruppate per Paesi
- Collaborazioni (numero di autori, numero di papers) di ricercatori dell'Università di Milano con altre istituzioni di tutto il mondo, raggruppate per Paesi. Scelta interattiva del ricercatore.
- Collaborazioni (numero di autori, numero di papers) di gruppi di istituzioni di tutto il mondo con l'Università degli Studi di Milano. Scelta interattiva delle istituzioni.

Il cruscotto è interattivo e permette di eseguire ricerche scegliendo lo specifico elemento (il ricercatore, il paese) restituendo grafici e dati tabellari di rendicontazione riguardanti le pubblicazioni scientifiche.

La visualizzazione dei risultati si compone di diverse viste. Nella prima vista vengono illustrate geograficamente le collaborazioni degli autori con altri paesi nel mondo.

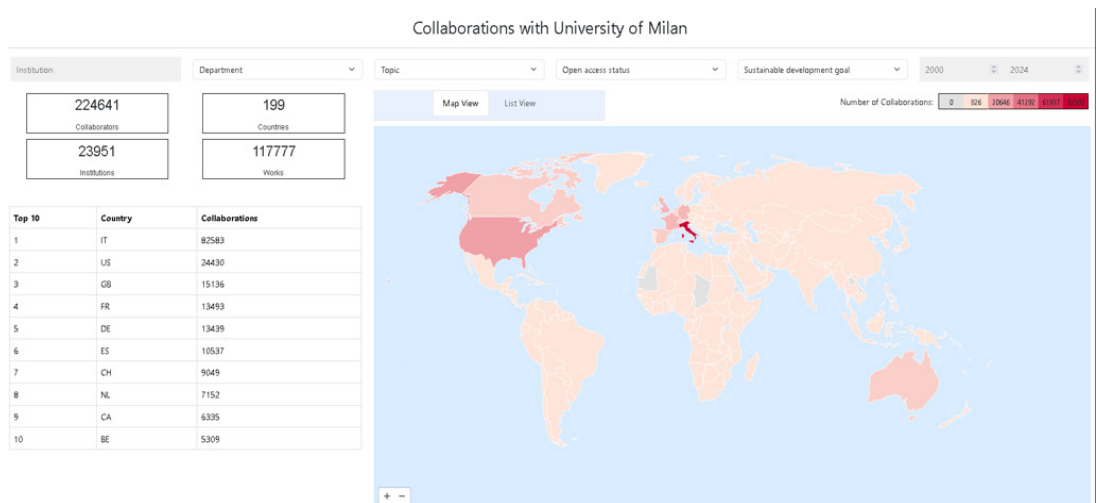
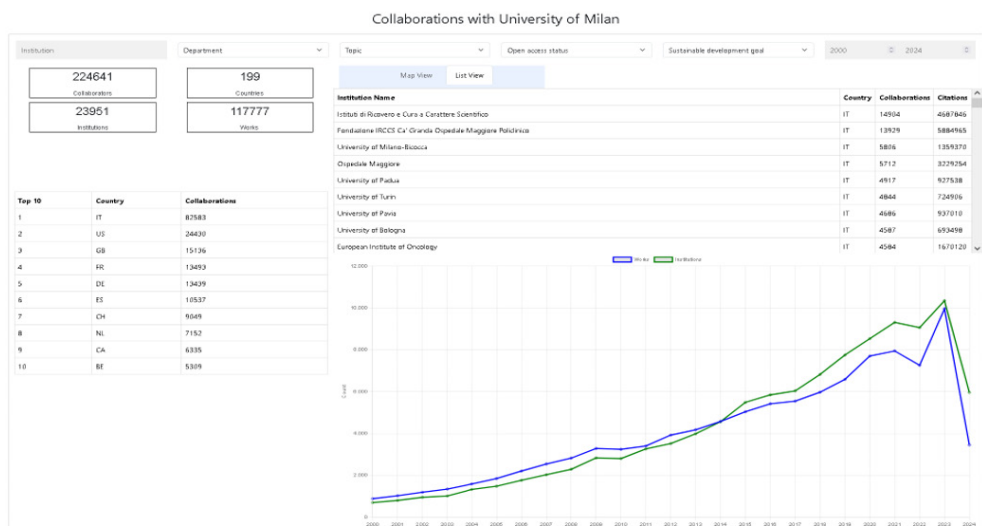


Fig. 2

Vengono mostrati il numero di collaboratori, di istituzioni e di pubblicazioni effettuate in maniera collaborativa dall'Ateneo o dal dipartimento specifico. Esistono inoltre alcuni filtri che hanno effetti selettivi sulla visualizzazione in maniera immediata.



Le stesse informazioni possono essere ricavate dalla vista tabellare come da esempio. Tale vista mostra anche un grafico a linee che descrive l'andamento nel tempo. Anche in questo caso i filtri agiscono in maniera selettiva sui dati, mostrando le viste relative alle scelte effettuate.

Tecnologie utilizzate

Dal punto di vista tecnologico, si è optato per una scelta completamente open-source, utilizzando librerie e strumenti di sviluppo facilmente disponibili su internet. Il progetto è stato realizzato costruendo un applicativo che usa le api di OpenAlex per ricavare i dati relativi ai ricercatori come censiti in AIR, con i quali alimenta un database da cui un altro applicativo in javascript costruisce l'interfaccia pubblica di esposizione grafica dei dati.

L'applicazione si compone delle seguenti parti:

- Il database utilizzato è postgres 15.
- Il programma che recupera i dati è stato scritto con il linguaggio python versione 3.9 (<https://www.python.org>). Il programma si compone di una serie di moduli che recuperano i dati dapprima dalle tabelle AIR e successivamente da OpenAlex.
- L'interfaccia utente è il cliente ed è stata scritta utilizzando gli strumenti javascript react (<https://react.dev>) e nodeJs (<https://nodejs.org>). In particolare, react realizza l'interfaccia utente attraverso una serie di componenti sul browser, mentre nodeJs agisce da server, ricevendo le chiamate dal client e accedendo al database postgres.

Autori:

Dario Basset presso la Direzione Performance, Assicurazione Qualità, Valutazione e Politiche di Open Science dell'Università degli studi di Milano. Si occupa degli strumenti dell'Ateneo per i dati della ricerca e più in generale dell'open science. È coinvolto nelle attività di formazione che riguardano i dati della ricerca e gli strumenti relativi, e si occupa da più di un anno della qualità dei dati del repository Dataverse.

Stefano Bolelli Gallevi lavora presso la Direzione Performance, Assicurazione Qualità, Valutazione e Politiche di Open Science dell'Università degli studi di Milano. Si occupa del repository istituzionale AIR e collabora alle altre attività di Open Science della Direzione.