

# Wikidata e il controllo di autorità nel mondo della ricerca

Camillo Carlo Pellizzari di San Girolamo

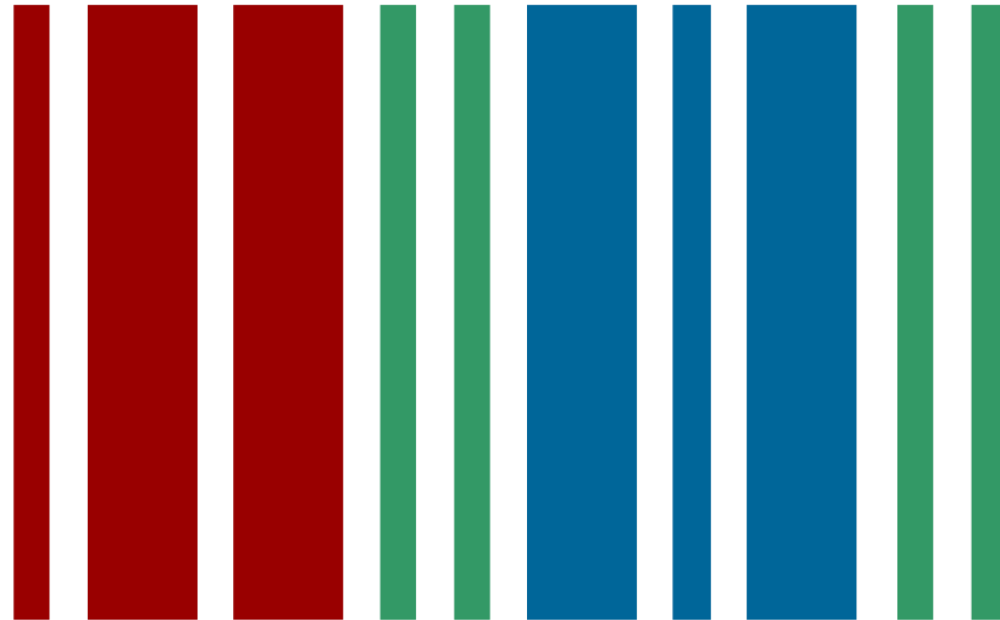
Scuola Normale Superiore; Wikimedia Italia

Alessandro Marchetti

Wikimedia Svizzera

*Università degli Studi di Firenze, 17 gennaio 2024*

# Una brevissima introduzione a Wikidata



**WIKIDATA**

# Wikidata



- **progetto Wikimedia** (fratello di Wikipedia), fondato il 29 ottobre 2012
- **base di conoscenza:** raccoglie dati strutturati in più lingue, disponibili in licenza CC0 (pubblico dominio), su (ad oggi) oltre 108 milioni di entità
- **piattaforma collaborativa:** oltre 11 mila utenti attivi
- **punto di snodo del web semantico:** le entità si connettono tramite identificatori ad oltre 8 mila database

# Un elemento di Wikidata



- ha un identificativo indipendente dalla lingua
- ha etichette, descrizioni e alias (forme alternative del nome) in potenzialmente ogni lingua
- contiene dati strutturati riguardo all'entità
- contiene identificatori di altri database che descrivono l'entità
- contiene i link alle pagine dei vari progetti Wikimedia (es. voci di Wikipedia) relative all'entità

# I dati di Wikidata



- possono essere riutilizzati senza alcun vincolo (grazie alla licenza CC0)
- sono scaricabili integralmente tramite i dump
- sono consultabili tramite API
- sono interrogabili tramite il query service in linguaggio SPARQL  
(<https://query.wikidata.org/>)

# Wikidata come luogo di incontro

**CATALOGAZIONE  
(METADATAZIONE)**

Cataloghi di biblioteche  
(*voci di autorità,  
pubblicazioni*)



**BIBLIOMETRIA**

Database citazionali e  
archivi della ricerca  
(*autori, citazioni,  
pubblicazioni*)

# Wikidata e bibliometria

## AUTOMATICO

grossi import:  
bot, OpenRefine

## SEMI-AUTOMATICO

QuickStatements,  
Mix'n'Match

## MANUALE

singoli utenti  
formazione

## ARTICOLI

- titolo
- rivista, volume, pagina
- stringhe autori
- DOI

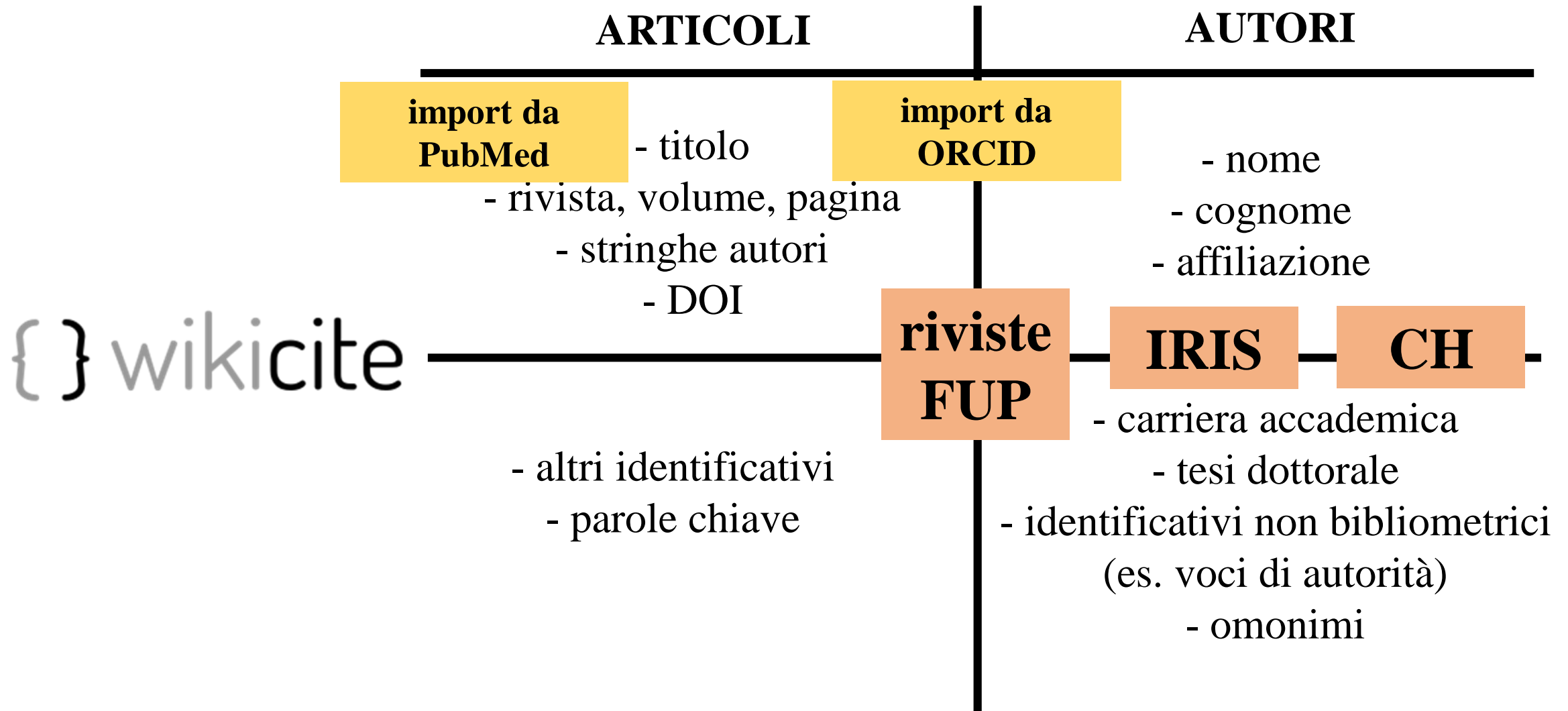
- altri identificativi
- parole chiave

## AUTORI

- nome
- cognome
- affiliazione

- carriera accademica
- tesi dottorale
- identificativi non bibliometrici  
(es. voci di autorità)
- omonimi

# Wikidata e bibliometria





# Wikidata:Gruppo Wikidata per Musei, Archivi e Biblioteche/IRIS

Abbreviazione: WD:IRIS

< Wikidata:Gruppo Wikidata per Musei, Archivi e Biblioteche  
Elemento di Wikidata non trovato.

Il progetto **IRIS** del Gruppo Wikidata per Musei, Archivi e Biblioteche mira a connettere a Wikidata gli autori le cui pubblicazioni sono presenti negli archivi open-access IRIS delle università italiane.

Per contatti, scrivere sulla mailing-list del Gruppo Wikidata per Musei, Archivi e Biblioteche.

Il progetto è il più attivo fra quelli in corso in Italia sul tema, si veda anche Wikidata:WikiCite/Italy.

[Indice](#) [mostra]

[Introduzione](#) [modifica]

*See also: Wikidata:Gruppo Wikidata per Musei, Archivi e Biblioteche/IRIS/Guida*



# Il progetto IRIS

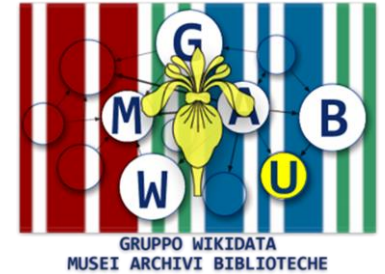
<https://www.wikidata.org/wiki/WD:IRIS>

# IRIS (Institutional Research Information System)

- software sviluppato dal CINECA per gli archivi istituzionali della produzione scientifica degli istituti di ricerca italiani
- usato da quasi tutte le università pubbliche italiane, nonché da alcune università private e da alcuni altri istituti di ricerca, per un totale di oltre 80 istanze

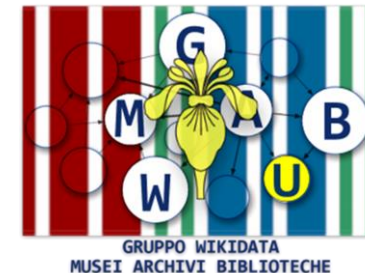


# Scopo



- il progetto IRIS mira, sul lungo periodo, a descrivere in modo esaustivo la produzione scientifica degli istituti di ricerca italiani, sulla base dei dati presenti negli archivi istituzionali che usano il software IRIS
- poiché ogni pubblicazione (scientifica o meno) presuppone l'esistenza di almeno un autore, la prima fase del progetto prevede la mappatura più completa possibile in Wikidata degli **autori** censiti negli archivi

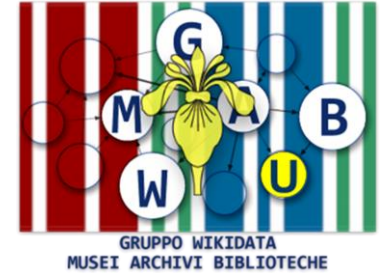
# Svolgimento



- per ciascun archivio che utilizza IRIS è stata creata una proprietà (attualmente sono 81)
- per ogni archivio, al momento della creazione della proprietà, è stata estratta la lista di tutti gli autori ed è stata importata nel tool Mix'n'match (<https://mix-n-match.toolforge.org/>)



# Svolgimento



- la riconciliazione degli identificatori assegnati agli autori avviene:
  - in modo automatico confrontando le pubblicazioni degli autori presenti negli archivi con quelle presenti in Wikidata (per autori di ambito medico-scientifico)
  - manualmente tramite Mix'n'match
- ad eccezione di due progetti conclusi (Università di Salerno e Scuola Normale Superiore) e di uno in corso (Università di Firenze), la riconciliazione finora è stata svolta da volontari



# Campagna social

WD:IRIS TOSCANA

P9761 P9913  
P9795 P9914  
P9796 P9915



UNIVERSITÀ STATALI

WD:IRIS UMBRIA

P10070  
P10071



UNIVERSITÀ STATALI

WD:IRIS MARCHE

P10054 P10147  
P10105 P10148



UNIVERSITÀ STATALI

WD:IRIS SARDEGNA

P10258  
P10259



UNIVERSITÀ STATALI

WD:IRIS LIGURIA

P10359



UNIVERSITÀ STATALI

WD:IRIS FRIULI-VENEZIA GIULIA

P10011 P10506  
P10505



UNIVERSITÀ STATALI

WD:IRIS PIEMONTE

P10104 P10638  
P10277 P10639



UNIVERSITÀ STATALI

WD:IRIS EMILIA-ROMAGNA

P10058 P10762  
P10761 P10763



UNIVERSITÀ STATALI

WD:IRIS ABRUZZO

P10009 P10828  
P10827 P10829



UNIVERSITÀ STATALI

WD:IRIS MOLISE

P10866



UNIVERSITÀ STATALI

WD:IRIS CAMPANIA

P9843 P10948  
P10056 P10949  
P10947 P10950 P10935



UNIVERSITÀ STATALI E PRIVATE

WD:IRIS CALABRIA

P11059 P11061  
P11060



UNIVERSITÀ STATALI

WD:IRIS BASILICATA

P11115



UNIVERSITÀ STATALI

WD:IRIS SICILIA

P11207 P11209  
P11208 P11210



UNIVERSITÀ STATALI E PRIVATE

WD:IRIS PUGLIA

P10102 P11304  
P11303 P11306 P11305



UNIVERSITÀ STATALI E PRIVATE

WD:IRIS TRENINO-ALTO ADIGE

P11506 P11509  
P11507 P11508



UNIVERSITÀ STATALI E ALTRO

WD:IRIS VENETO

P10055 P11511  
P11510 P11597



UNIVERSITÀ STATALI

WD:IRIS LAZIO

P9956 P11673 P11655 P10938  
P10276 P11674 P11653  
P11672



UNIVERSITÀ STATALI E PRIVATE

WD:IRIS LOMBARDIA

P9844 P11717 P11723 P10936  
P10010 P11718 P11724 P10937  
P10057 P11719 P11725 P11654  
P10103 P11720



UNIVERSITÀ STATALI E PRIVATE

WD:IRIS 81 PROPRIETÀ



66 STATALI  
15 PRIVATE

19 REGIONI





# Campagna social

WD:IRIS ●●●  
COMPLETATA...


 Università  
per Stranieri  
di Perugia

P10071  
~ 100



WD:IRIS ●●●  
COMPLETATA...


 HIPPOCRATICA CIVITAS  
STUDIUM SALERNITANUM

P10056  
~ 5100



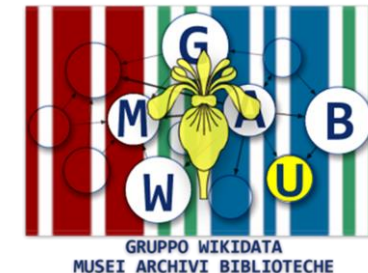
WD:IRIS ●●●  
COMPLETATA...

SCUOLA  
NORMALE  
SUPERIORE



P9761  
~ 1800



# Per maggiori informazioni

- <https://www.wikidata.org/wiki/WD:IRIS> contiene ulteriori dati sul progetto, altre presentazioni, video e bibliografia

Cfr. Stefano Bargioni, Carlo Bianchini, Camillo Carlo Pellizzari di San Girolamo, «IRIS e Wikidata: un progetto per una migliore valorizzazione e fruizione degli archivi della ricerca scientifica italiana», in Giovanni Bergamin, Tiziana Possemato (edd.), *Guardando oltre i confini. Partire dalla tradizione per costruire il futuro delle biblioteche. Studi e testimonianze per i 70 anni di Mauro Guerrini* (2023), 31-45.

- <https://www.wikidata.org/wiki/WD:GWMAB> contiene tutti i dati sui progetti del Gruppo Wikidata per Musei, Archivi e Biblioteche



# Wikidata:WikiCite/Researchers in Switzerland

< Wikidata:WikiCite  
Elemento di Wikidata non trovato.

This is a side project funded by WMCH as part and is part of [Wikidata:WikiCite/Switzerland](#)

This initiative can be viewed as the second attempt at achieving comprehensive coverage of researchers within a country, following the organized example of IRIS proect in Italy.

## Indice [\[nascondi\]](#)

- 1 History
- 2 Main goal
- 3 Steps
- 4 Other related edits
  - 4.1 Items of institutions or departments created during the project
  - 4.2 Items of institutions improved during the project
  - 4.3 Identical or similar names refined or created on the way
  - 4.4 Mistakes found in external archives
- 5 Future work



# Il progetto ricercatori in Svizzera

([https://www.wikidata.org/wiki/Wikidata:WikiCite/Researchers in Switzerland](https://www.wikidata.org/wiki/Wikidata:WikiCite/Researchers_in_Switzerland))

# Italia vs Svizzera



	Italia	Svizzera
<b>Promotore</b>	GWMAB, comunità	Wikimedia Svizzera
<b>Punto di partenza</b>	ID e Mix'n'match	siti ufficiali
<b>Spin-off di...</b>	sforzo di controllo di autorità e prosopografia	Wiki Science Competition
<b>Status degli archivi istituzionali</b>	frammentato ma omogeneo (IRIS)	molto frammentato
<b>Metriche aggiuntive</b>		monitoraggio degli omonimi
<b>Obiettivo qualitativo</b>	accuratezza, completezza	funzionalità
<b>Comunità</b>	uso di canali sociali, mailing list ecc.	festival della qualità con premi
<b>Social media</b>	campagna dedicata	social media in attesa
<b>Partner istituzionali</b>	contatti con alcune università	primi contatti
<b>Didattica, formazione</b>	sì	non ancora

# Tre grandi sfide per Wikidata ...

1. **distinguere gli autori omonimi** adeguatamente
2. **arricchire gli elementi** di Wikidata con informazioni e identificatori (di biblioteche, altri IRIS, database accademici ecc.)
3. **individuare dati imprecisi** presenti negli archivi (autori inesistenti, duplicati, conflati, nominati in modo impreciso) per non importarli

# ... tre opportunità per gli archivi

1. **ottenere un'identificazione precisa** dei propri autori anche al di fuori del singolo archivio
2. **avere più informazioni** sugli autori, tratte da fonti diverse; essere pienamente inseriti nel web semantico, facendo perno su Wikidata
3. **trovare e correggere i propri dati imprecisi** (autori inesistenti, duplicati, conflati, nominati in modo impreciso)

# Cosa manca quindi?

1. trovare (e formare) **più persone** che connettano gli identificatori degli autori presenti negli archivi a Wikidata e ne sistemino gli omonimi
2. e, ancora, trovare (e formare) **più persone** che migliorino gli elementi in Wikidata degli autori descritti negli archivi aggiungendovi informazioni e altri identificatori

# Cosa manca quindi?

3. creare un meccanismo efficiente per permettere agli utenti di Wikidata di **segnalare** a ogni singolo archivio e agli authority file nazionali **i dati imprecisi** da correggere
4. sviluppare **gruppi di lavoro simili in altri paesi** per condividere le conoscenze e per alleggerire i rispettivi lavori

# Wikidata:Gruppo Wikidata per Musei, Archivi e Biblioteche/Riviste FUP

---

[< Wikidata:Gruppo Wikidata per Musei, Archivi e Biblioteche](#)  
Elemento di Wikidata non trovato.

Il progetto (novembre-dicembre 2022), finanziato  da [Wikimedia Italia \(Q15136611\)](#), prevede l'importazione su Wikidata da parte di [Alexmar983](#) dei metadati degli articoli disponibili in open access delle riviste [Caryologia \(Q3661018\)](#) e [Substantia \(Q111198802\)](#), pubblicate dalla [Firenze University Press \(Q3745980\)](#).

Gli articoli importati si caratterizzano per l'uso di [autore \(P50\)](#) per indicare gli autori (che sono quindi rappresentati da singoli elementi, disambiguati dai principali omonimi) e, nella gran parte dei casi, per l'uso di [argomento principale \(P921\)](#) per indicare gli argomenti (che sono anch'essi rappresentati da elementi, integrati nell'ontologia di Wikidata e connessi ove possibile a thesaurus generalisti e disciplinari). L'importazione ha inoltre permesso di [segnalare alcune imprecisioni](#) nei dati bibliografici presenti nei siti delle riviste.

## Il progetto riviste FUP

[https://www.wikidata.org/wiki/Wikidata:Gruppo Wikidata per Musei, Archivi e Biblioteche/Riviste FUP](https://www.wikidata.org/wiki/Wikidata:Gruppo_Wikidata_per_Musei,_Archivi_e_Biblioteche/Riviste_FUP)

# Contenuto

- tra novembre e dicembre 2022 Alessandro Marchetti ha importato in Wikidata i metadati bibliografici degli articoli di due riviste pubblicate in open access dalla Firenze University Press (FUP), *Caryologia* e *Substantia*
  - (nota: l'import ha interessato solo i numeri delle riviste pubblicati dalla FUP)
- sono ora presenti in Wikidata, per queste due riviste, gli elementi di 601 articoli ([esempio](#)), collegati agli elementi di 956 autori unici



# Specificità metodologiche

- importazione degli **autori in quanto elementi** e non in quanto semplici stringhe di testo (fondamentale a fini bibliometrici)
- sistemazione manuale dei **principali omonimi degli autori** degli articoli, risolvendo conflazioni e duplicazioni esistenti
- aggiunte manuale di **identificativi esterni agli autori**, per renderne più chiara l'identificazione (e ridurre la possibilità di confusioni future)
- aggiunta manuale sistematica degli **argomenti degli articoli** (le 'parole chiave', inserite però come elementi e non come semplici stringhe)

# Grazie per l'attenzione!

[camillo.pellizzaridisangirolamo@sns.it](mailto:camillo.pellizzaridisangirolamo@sns.it)

[alessandro.marchetti@wikipedia.ch](mailto:alessandro.marchetti@wikipedia.ch)