

OpenCitations: enabling the FAIR use of open citation data

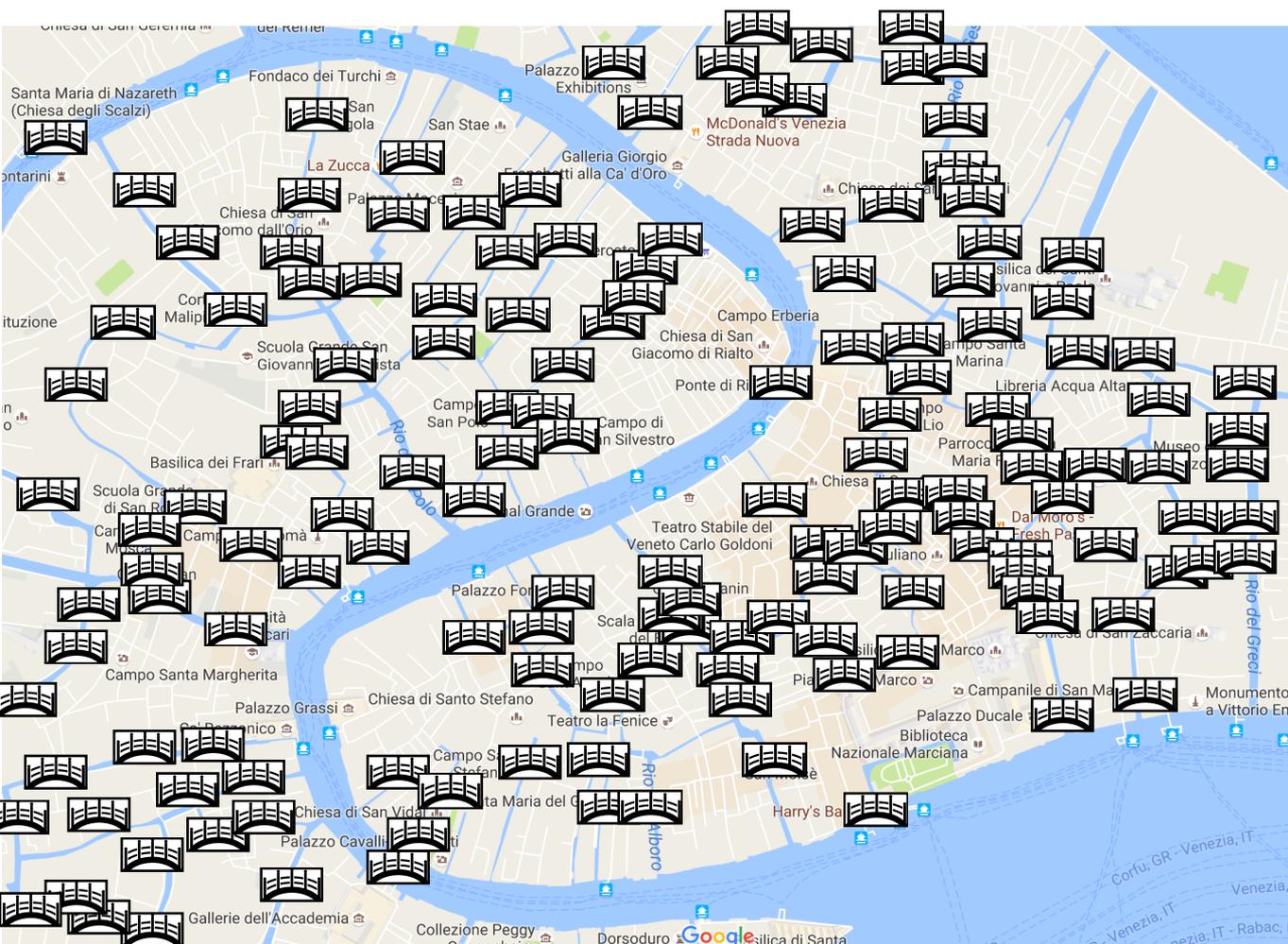
<https://w3id.org/people/essepuntato/papers/oc-garr2017.html>

Silvio Peroni, David Shotton

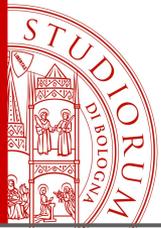
*Conferenza GARR 2017
Venezia, Italia, 16 Novembre 2017*

Venezia: un'analogia

<https://w3id.org/oc/paper/the-venice-analogy.html>



- Isola = pubblicazione
- Ponte = citazione
- Situazione attuale:
 - OK spostarsi verso un'isola vicina
 - biglietto molto caro per spostarsi liberamente in tutta la città
 - “macchine” non sono ammesse



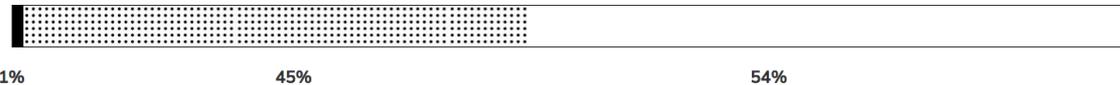
Initiative for Open Citations

<https://i4oc.org>

Initiative for Open Citations

The Initiative for Open Citations **I4OC** is a collaboration between scholarly publishers, researchers, and other interested parties to promote the unrestricted availability of scholarly citation data.

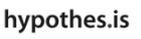
How many citations are open today?

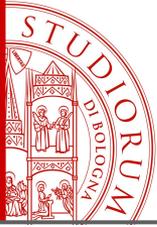


Fondatori



Stakeholder

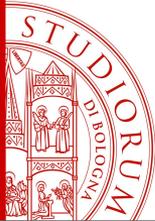




OpenCitations Corpus

<http://opencitations.net>

- **OpenCitations** ha come scopo la creazione di una base di dati aperta di citazioni tra lavori accademici – l'**OpenCitations** Corpus (OCC) – messi a disposizioni come materiale di pubblico dominio
- Attualmente il processo di generazione dei dati relativi alle citazioni usa gli articoli disponibili nel sotto-insieme Open Access di PubMedCentral, grazie all'uso delle Europe PubMedCentral API, e li arricchisce con altri dati ottenuti interrogando le Crossref API e le ORCID API
- Al 10 Novembre 2017, l'**OpenCitations** Corpus contiene:
 - 273,242 risorse bibliografiche citanti
 - 6,094,196 risorse bibliografiche citate
 - 11,544,405 di citazioni (articolo A cita articolo B)
- Attuale generazione dati: > 500,000 citazioni al mese
- Nuova infrastruttura (fondi Sloan): > 500,000 citazioni **al giorno**
- Altre sorgenti verranno considerate: arXiv, Crossref+I4OC, etc.



FAIR

- **Findable**

- Risorse univocamente identificate da un URL `w3id.org` persistente
- SPARQL endpoint per interrogare la base di dati
- Dump dell'intero [OpenCitations](#) Corpus su Figshare

- **Accessible**

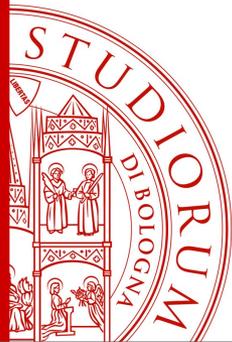
- HTTP per accedere ai dati delle risorse
- Informazioni in formati comprensibili sia da umani (HTML) sia da macchine (RDF)

- **Interoperable**

- Modello descrizione dei dati ben documentato (<https://doi.org/10.6084/m9.figshare.3443876>)
- SPAR Ontologies per descrivere dominio editoriale
- Vocabolari standard per descrizione di collezione di dati (PROV-O, DCAT, VOID)

- **Reusable**

- Dati rilasciati in CC0
- Provenance (sorgenti primarie, agenti responsabili della curatela, etc.)
- Tracciamento evoluzione dei dati nel tempo
- Codice script disponibile su GitHub rilasciato con licenze open source



Grazie per l'attenzione

Silvio Peroni, David Shotton

*Conferenza GARR 2017
Venezia, Italia, 16 Novembre 2017*