

## Proposta di Programma di Formazione

**Titolo:** Visualizzazione ed analisi di dati citazionali globali e in linea presenti nell'Open Biomedical Citations in Context Corpus

**Tutor:** Silvio Peroni <[silvio.peroni@unibo.it](mailto:silvio.peroni@unibo.it)>, che può essere contattato per ulteriori informazioni

### Obiettivi

OpenCitations (<http://opencitations.net>) è un'infrastruttura che promuove e mette a disposizione strumenti per la diffusione di dati bibliografici e citazionali aperti nel contesto accademico, messi a disposizione nel Pubblico Dominio (<https://creativecommons.org/publicdomain/zero/1.0/deed.it>). Utilizzando le tecnologie del web semantico - che permettono la creazione e la pubblicazione di dati FAIR (findable, accessible, interoperable, reusable) sul web in modo che siano facilmente processabili anche da agenti software - l'infrastruttura nasce per offrire tanto alle istituzioni che ai singoli la possibilità di analizzare e riutilizzare le citazioni scientifiche nelle raccolte bibliografiche e nelle attività di ricerca.

L'*Open Biomedical Citations in Context Corpus Project*, finanziato dal Wellcome Trust (Londra, Regno Unito), vuole rendere i dati citazionali in OpenCitations ancora più utili per la comunità accademica estendendo significativamente le tipologie di dati citazionali attualmente gestiti, in modo da fornire dati relativi ai riferimenti citazionali testuali in linea e al relativo contesto semantico, così da poter distinguere i riferimenti che sono citati solo una volta da quelli citati più volte, vedere quali riferimenti sono citati insieme (ad esempio, nella stessa frase), determinare in quale sezione dell'articolo i riferimenti sono citati (ad esempio, nell'introduzione, nei metodi), e, potenzialmente, estrarre la funzione della citazione – ovvero la ragione per cui un autore cita un altro lavoro. A questo proposito, l'*Open Biomedical Citations in Context Corpus Project* ha finanziato la borsa di ricerca in oggetto. Il principale obiettivo del progetto da svolgere da parte del/della borsista è quello di sviluppare tutto il software necessario per visualizzare ed analizzare dati citazionali e il loro relativo contesto.

La borsa di ricerca associata al progetto *non* è di natura commerciale.

### Piano di attività

Si prevede uno svolgimento di 4 mesi per il programma complessivo. Seppur il *Open Biomedical Citations in Context Corpus Project* è una collaborazione con l'École de Bibliothéconomie et des Sciences de l'Information (Université de Montréal, Canada), l'Oxford e-Research Centre (University of Oxford, Inghilterra), il Centre for Science and Technology Studies (Leiden University, Olanda), ed è formalmente supportato da Europe PubMed Central (EMBL-EBI, Inghilterra), il/la borsista lavorerà con il Dr. Silvio Peroni presso il Research Centre for Open Scholarly Metadata del Dipartimento di Filologia Classica e Italianistica (Università di Bologna, Italia). Il centro è un ambiente vivo e stimolante, in cui il/la borsista dovrà fornire il suo significativo contributo personale al progetto. In una prima fase di circa un mese, il/la borsista dovrà impraticarsi delle tecnologie utilizzate per l'interrogazione dei dati citazionali presenti in OpenCitations e dovrà studiare software esistenti per lo sviluppo di interfacce Web che ben si sposano con le tecnologie del web semantico. Nei restanti mesi il borsista sarà responsabile:

- dello sviluppo di interfacce utente per interrogare e navigare i nuovi dati citazionali globali e testuali in linea;
- dell'analisi di questi dati citazionali mediante le interfacce realizzare, per proporre ulteriori estensioni che mostrino viste significative sui dati così da permetterne una descrizione ad alto livello e da evidenziarne eventuali caratteristiche distintive.

### Requisiti

Il/la borsista deve avere ottime competenze di ricerca, di creazione e gestione di dati e modelli, di programmazione, e di comunicazione. Inoltre, deve essere in grado di scrivere e presentare oralmente i lavori

svolti in Inglese. Passate esperienze in Bibliometrics, Scientometrics, Semantic Publishing, Python, Web Interface Design, tecnologie Web, Semantic Web e Linked Data sono valori aggiunti, così come la dedizione alle tematiche di Open Science e l'abilità di lavorare in gruppo. Il requisito minimo per applicare per la posizione è avere una laurea magistrale in LM-84, con adeguato curriculum scientifico-professionale. Sono altresì ben valutate e auspicabili eventuali esperienze di ricerca conformi ai temi del progetto e con un possibile percorso di dottorato di ricerca.

## Research programme

**Title:** Visualization and analysis of global and in-text citation data in the Open Biomedical Citations in Context Corpus

**Academic supervisor:** Silvio Peroni <[silvio.peroni@unibo.it](mailto:silvio.peroni@unibo.it)>, from whom further information may be obtained

### Goals

OpenCitations (<http://opencitations.net>) is an independent infrastructure organization for open scholarship dedicated to the publication of open bibliographic and citation data made available under a Creative Commons public domain dedication (CC0, <https://creativecommons.org/publicdomain/zero/1.0/>). Using Semantic Web technologies – which enable the creation and publication of FAIR (findable, accessible, interoperable and re-usable) data on the Web to make them easily processable by software agents – the infrastructure allows single users and large institution to analyze and reuse scientific citations in bibliographic collections and research activities.

The *Open Biomedical Citations in Context Corpus Project*, funded by the Wellcome Trust (London, United Kingdom), wants to make citation data in OpenCitations more useful to the academic community by significantly expanding the kinds of citation data currently handled, so as to provide data for each individual in-text reference and its semantic context, making it possible to distinguish references that are cited only once from those that are cited multiple times, to see which references are cited together (e.g. in the same sentence), to determine in which section of the article references are cited (e.g. Introduction, Methods), and, potentially, to retrieve the function of the citation – i.e. the reason why an author cites another work. To this end, the *Open Biomedical Citations in Context Corpus Project* has funded the salary of a short-term research fellow, who will primarily be involved in the development of all the necessary software for visualizing and analyzing citation data and their semantic context.

The project related to this position is non-commercial in nature.

### Activity plan

The short-term research fellowship position has a duration of 4 months. While the *Open Biomedical Citations in Context Corpus Project* is a collaboration with the École de Bibliothéconomie et des Sciences de l'Information (Université de Montréal, Canada), the Oxford e-Research Centre (University of Oxford, United Kingdom), Centre for Science and Technology Studies (Leiden University, The Netherlands), and supported by Europe PubMed Central (EMBL-EBI, United Kingdom), the short-term research fellow will work with Dr Silvio Peroni in the Research Centre for Open Scholarly Metadata at the Department of Computer Classical Philology and Italian Studies (University of Bologna, Italy). This is a lively and stimulating environment, and the short-term research fellow will be expected to provide a key personal contribution to the project. During the first month, the short-term research fellow will practice the technologies used for querying citation data available in OpenCitations and will study existing software for the development of Web-based interfaces that can be used with Semantic Web technologies. In the remaining months, the short-term research fellow will be responsible for:

- the development of appropriate user interfaces for querying and browsing these new global and in-text citation data;
- the analysis of these citation data by means of the interfaces developed, so as to propose additional extensions that show meaningful view on such data with the goal of describing them from an high-level perspective and to identify distinctive patterns.

### Requirements

Applicants are expected to have excellent research skills, computer programming skills, and the ability to

communicate, undertake academic writing and make verbal conference presentations in good English. Expertise in Bibliometrics, Scientometrics, Python, Web Interface Design and Information Visualization, and in Semantic Web technologies, Linked Data and Web technologies would be highly beneficial, plus a strong and demonstrable commitment to open science and team-working abilities. The minimal formal requirement for this position is a Masters degree in one of the following “classi di laurea”: LM 84, with appropriate experience in the topics of the project. In addition, it is expected that the successful applicant will have had research experience leading to a doctoral degree.