

Programma di formazione

Titolo

Integrare le citazioni aperte nell'European Open Science Cloud

Responsabile scientifico

Professor Silvio Peroni <silvio.peroni@unibo.it>, Dipartimento di Filologia Classica e Italianistica, Università di Bologna / Direttore di OpenCitations, che può essere contattato per ulteriori informazioni.

Obiettivi

OpenAIRE-Nexus è un progetto Horizon 2020 finanziato dalla Commissione Europea (<https://cordis.europa.eu/project/id/101017452>) che ha l'obiettivo di portare, all'interno della [European Open Science Cloud \(EOSC\)](#), quattordici nuovi servizi focalizzati sullo sviluppo e promozione della Scienza Aperta. Tra questi servizi rientra **OpenCitations** (<http://opencitations.net/>) [1], un'**infrastruttura aperta e completamente gratuita**, gestita dal Research Centre for Open Scholarly Metadata dell'Università di Bologna (<https://openscholarlymetadata.org>), che mette a disposizione una grossa mole di **metadati bibliografici e dati citazionali accademici**. Nel contesto del progetto OpenAIRE-Nexus, l'obiettivo è quello di rendere completamente compatibili i servizi offerti da OpenCitations con OpenAIRE, così da garantirne l'interoperabilità, semantica e tecnica, anche con tutti gli altri servizi Open Science a disposizione su EOSC.

La mole e la complessità di dati messi a disposizione da OpenCitations aprono numerosi problemi di scalabilità e ottimizzazione. È necessario quindi studiare e personalizzare le più recenti tecniche di gestione di Big Data per creare una base di conoscenza su cui costruire servizi scalabili, duraturi e facilmente personalizzabili. L'obiettivo dell'Assegno di Ricerca, infatti, è quello di studiare le limitazioni e i miglioramenti necessari all'attuale infrastruttura per essere integrata all'interno di OpenAIRE e di EOSC. Saranno affrontati, tra gli altri, temi di erogazione di servizi Web, gestione di basi di dati distribuite ed eterogenee, e processi di conversione e ingestione dati.

Il lavoro da svolgersi per questa posizione è di natura non commerciale.

Piano di attività

L'Assegno di Ricerca avrà durata di 12 mesi a partire da Maggio 2021, eventualmente rinnovabile per ulteriori 12 mesi. L'Assegnista di Ricerca lavorerà direttamente con il Professor Silvio Peroni e diventerà membro del Research Centre for Open Scholarly Metadata, presso il Dipartimento di Filologia Classica e Italianistica dell'Università di Bologna (Italia). Il Centro di Ricerca è un ambiente vivo e stimolante, ed è atteso che l'Assegnista di Ricerca fornisca contributi personali centrali alle attività di OpenCitations nel contesto del progetto OpenAIRE-Nexus. Il lavoro a distanza può essere possibile se strettamente necessario per la situazione relativa alla pandemia di COVID-19, ma altrimenti la presenza di persona nel Centro di Ricerca è preferibile.

Durante il primo mese, l'Assegnista di Ricerca verrà introdotto/a a tutti i sistemi hardware e software usati da OpenCitations per fornire i propri servizi, così da capire lo stato corrente dell'infrastruttura tecnologica. Nei restanti mesi, l'Assegnista di Ricerca assumerà la responsabilità di gestione e sviluppo di questi sistemi, proponendo soluzioni innovative basate su pratiche di ricerca aggiornate, ove necessario, in modo da permettere l'integrare dei servizi offerti all'interno di OpenAIRE e di EOSC. Questi sistemi includono:

- Un server fisico principale, che fornisce tutti i servizi di OpenCitations;
- Le basi di dati di tutti i dataset di OpenCitations, che includono l'OpenCitations Corpus [2], gli OpenCitations Indexes [3], il Open Biomedical Citations in Context Corpus [4], e il futuro OpenCitations Meta.

Mentre il professor Peroni dirigerà e supervisionerà il lavoro, l'Assegnista di Ricerca avrà la libertà decisionale e di responsabilità di gestire in modo autonomo e sistematico queste attività.

Requisiti

Tutti/e i/le candidati/e devono avere eccellenti abilità come programmatori/trici, e devono essere in grado di parlare, scrivere, e presentare verbalmente a conferenze in un buon inglese. Esperienze dimostrabili nell'usare e gestire server Linux (in particolare Debian e Raspbian), server Web che implementano FastGCI (in particolare lighttpd), basi di dati basate su grafi (in particolare Blazegraph), programmazione in Python e utilizzo dei più comuni Python Web

Application Framework (in particolare web.py e Flask), e sistemi di versionamento basati su Git (in particolare GitHub) sono fortemente desiderabili. In più, è altresì fortemente desiderabile che il/la candidato/a abbia una forte e dimostrabile attitudine verso la scienza aperta e la capacità di lavorare in gruppo. Conoscenze dimostrabili nelle tecnologie del Web Semantico, Linked Data e tecnologie Web in generale sono elementi favorevoli per la candidatura.

I requisiti minimi formali per la posizione sono il possesso di una Magistrale in Informatica, Ingegneria Informatica, Ingegneria delle Telecomunicazioni, Informatica Umanistica, o equivalente. Il candidato deve avere un'esperienza adeguata e dimostrabile come programmatore, comprovata dai documenti da allegare in fase di domanda. La candidatura (in Italiano o in Inglese) deve almeno includere un Curriculum Vitae completo di informazioni riguardanti attività scientifico-professionali e relative alla produttività scientifica. Eventuali lettere di raccomandazioni sono opzionali, ma fortemente consigliate.

L'Università di Bologna è un'istituzione che da pari opportunità di impiego, e la selezione per questa posizione verrà fatta esclusivamente sul merito.

Riferimenti

1. Peroni, S., & Shotton, D. (2020). OpenCitations, an infrastructure organization for open scholarship. *Quantitative Science Studies*, 1(1), 428–444. https://doi.org/10.1162/qss_a_00023
2. Peroni, S., Shotton, D., & Vitali, F. (2017). One Year of the OpenCitations Corpus. In C. d'Amato, M. Fernandez, V. Tamma, F. Lecue, P. Cudré-Mauroux, J. Sequeda, C. Lange, & J. Heflin (Eds.), *The Semantic Web – ISWC 2017 (Lecture Notes in Computer Science Vol. 10588, pp. 184–192)*. Springer. https://doi.org/10.1007/978-3-319-68204-4_19
3. Heibi, I., Peroni, S., & Shotton, D. (2019). Software review: COCI, the OpenCitations Index of Crossref open DOI-to-DOI citations. *Scientometrics*, 121(2), 1213–1228. <https://doi.org/10.1007/s11192-019-03217-6>
4. Daquino, M., Peroni, S., Shotton, D., Colavizza, G., Ghavimi, B., Lauscher, A., Mayr, P., Romanello, M., & Zumstein, P. (2020). The OpenCitations Data Model. In J. Z. Pan, V. Tamma, C. d'Amato, K. Janowicz, B. Fu, A. Polleres, O. Seneviratne, & L. Kagal (Eds.), *The Semantic Web – ISWC 2020 (Lecture Notes in Computer Science Vol. 12507, pp. 447–463)*. https://doi.org/10.1007/978-3-030-62466-8_28

Research programme

Title

Integrating open citations into the European Open Science Cloud

Supervisor

Professor Silvio Peroni <silvio.peroni@unibo.it>, Dipartimento di Filologia Classica e Italianistica, Università di Bologna / Director of OpenCitations, from whom further information may be obtained.

Goals

OpenAIRE-Nexus is an H2020 project funded by the European Commission (<https://cordis.europa.eu/project/id/101017452>) which aims at bringing, in the [European Open Science Cloud \(EOSC\)](#), fourteen new services focused on the development and promotion of Open Science. OpenCitations (<http://opencitations.net/>) [1] is one of these fourteen services. OpenCitations is a **fully free and open infrastructure**, managed by the Research Centre for Open Scholarly Metadata of the University of Bologna (<https://openscholarlymetadata.org>), that provides access to global **scholarly bibliographic and citation data**. In the context of the OpenAIRE-Nexus project, the goal is to make all services offered by OpenCitations compatible with OpenAIRE, so as to guarantee the semantic and technical interoperability also with all the other Open Science services available in the EOSC.

The amount and complexity of the data made available by OpenCitations opens up to several issues related to the scalability and optimization of its infrastructure. Therefore, it is crucial to study and customize recent approaches for the management of Big Data to create a knowledge base to build scalable, long-lasting and easily-customizable services. Indeed, the goal of the Research Fellowship is to study the current limitations and possible improvements to introduce in the OpenCitations infrastructure to integrate it with OpenAIRE and the EOSC. The Research Fellow is expected to address issues related with the provision of Web services, the management of distributed and heterogeneous databases, and data conversion and ingestion processes.

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

Activity plan

The Research Fellowship will have a duration of 12 months commencing in May 2021, that can be renewed for additional 12 months. The Research Fellow will work directly with Professor Silvio Peroni and will become a member of the Research Centre for Open Scholarly Metadata at the Department of Classical Philology and Italian Studies, University of Bologna, Italy. This is a multicultural, lively and stimulating environment, and the Research Fellow will be expected to provide a key personal contribution to the OpenCitations activities in the context of the project OpenAIRE-Nexus. Distance working will be permitted if necessary because of the COVID-19 pandemic, but otherwise attendance in person in the Research Centre will be preferable.

During the first month, the Research Fellow will be introduced to all the hardware and software systems currently used by OpenCitations to provide its services, so as to become familiar with the current status of the technical infrastructure. Subsequently, the Research Fellow will assume responsibility for managing and developing these systems, with the introduction of innovative solutions coming from current research practices when necessary, so as to integrate OpenCitations services with OpenAIRE and the EOSC. They comprise:

- The main physical server providing the OpenCitations services;
- The databases storing the data of all the OpenCitations datasets, including the OpenCitations Corpus [2], the OpenCitations Indexes [3], the Open Biomedical Citations in Context Corpus [4], and the future OpenCitations Meta dataset.

While Professor Peroni will direct and supervise his/her work, the Research Fellow will be given, and will be expected to exercise, a considerable degree of personal freedom and responsibility in undertaking these tasks.

Requirements

Applicants are expected to have excellent computer programming skills, and the ability to speak, write and make verbal conference presentations in good English. Demonstrable expertise in using and managing Linux servers (especially Debian and Raspbian), Web servers implementing FastGCI (especially lighttpd), graph databases (especially Blazegraph), Python programming and Python Web Application Frameworks (especially web.py and Flask), Git-based version-control systems (especially GitHub), plus a strong and demonstrable commitment to open science and

team-working abilities are highly desirable. Demonstrable knowledge in Semantic Web technologies, Linked Data and Web technologies would also be very beneficial.

The minimal formal requirement for this position is a Master degree in Computer Science, Computer Science and Engineering, Telecommunications Engineering, Digital Humanities, or equivalent. It is expected that the successful applicant will have had substantial and demonstrable experience in computer programming, described in the documents to submit for the application. The Research Fellowship application (written either in Italian or in English) should minimally comprise a full Curriculum Vitae containing information about prior scientific and professional jobs and about the scientific production of the candidate. Letters of recommendation are optional but highly desirable.

The University of Bologna is an equal-opportunities employer, and selection for this post will be made solely on merit.

References

1. Peroni, S., & Shotton, D. (2020). OpenCitations, an infrastructure organization for open scholarship. *Quantitative Science Studies*, 1(1), 428–444. https://doi.org/10.1162/qss_a_00023
2. Peroni, S., Shotton, D., & Vitali, F. (2017). One Year of the OpenCitations Corpus. In C. d'Amato, M. Fernandez, V. Tamma, F. Lecue, P. Cudré-Mauroux, J. Sequeda, C. Lange, & J. Heflin (Eds.), *The Semantic Web – ISWC 2017 (Lecture Notes in Computer Science Vol. 10588*, pp. 184–192). Springer. https://doi.org/10.1007/978-3-319-68204-4_19
3. Heibi, I., Peroni, S., & Shotton, D. (2019). Software review: COCI, the OpenCitations Index of Crossref open DOI-to-DOI citations. *Scientometrics*, 121(2), 1213–1228. <https://doi.org/10.1007/s11192-019-03217-6>
4. Daquino, M., Peroni, S., Shotton, D., Colavizza, G., Ghavimi, B., Lauscher, A., Mayr, P., Romanello, M., & Zumstein, P. (2020). The OpenCitations Data Model. In J. Z. Pan, V. Tamma, C. d'Amato, K. Janowicz, B. Fu, A. Polleres, O. Seneviratne, & L. Kagal (Eds.), *The Semantic Web – ISWC 2020 (Lecture Notes in Computer Science Vol. 12507*, pp. 447–463). https://doi.org/10.1007/978-3-030-62466-8_28