

OSL Platform: A Link to Open-access Scientific Information and Structured Data

Daniel Eisinger^a, George Tsatsaronis^a (✉), Alina Petrova^a, Efsthios Karanastasis^b, Vassiliki Andronikou^b, Efthymios Chondrogiannis^b

^a *Biotechnology Center, TU Dresden, Tatzberg 47-49, 01307 Dresden, Germany*
www.biotec.tu-dresden.de/

^b *National Technical University of Athens, 9 Iroon Polytechniou Str., 15773 Zografou, Greece*
http://www.ntua.gr/index_en.html

ARTICLE INFO:

RECEIVED: 20 Oct 2014
REVISED: 10 Nov 2014
ACCEPTED: 30 Nov 2014
ONLINE: 07 Jan 2015

KEYWORDS:

Open Science
Open Access to Scientific Data
Semantic Search
Life Sciences
Pilot Services
Evaluation of Science

ABSTRACT

The rapidly growing wealth of published scientific work, produced by researchers and scholars, has resulted in a pressing need for more effective processes towards reviewing scientific articles and research data, organizing data journals, as well as for improved tools and techniques for bibliographic analysis and management of scientometrics. The ongoing EU research project OpenScienceLink aims to address these needs, as well as offer a wide range of opportunities for better collaboration between researchers, by introducing a web-based Platform which offers efficient and intelligent applications and services for exploiting open access scientific information in the biomedical domain. The Platform is empowered by the semantic and social networking capabilities of three leading edge background infrastructures, which have been adapted and integrated for the scope of the project. In this paper, we present the five pilot services that are provided by the OpenScienceLink project. All five services are integrated into the web-based OpenScienceLink platform that is publicly accessible at <http://www.opensciencelink.org>.



Creative Commons BY-NC-SA 4.0

© 2015 Hosting by Procon Ltd. All rights reserved.

Introduction

During the last decade we have witnessed explosive growth in the amount of published scientific information in various forms. Apart from scientific/ research articles and monographs, this information increasingly includes research datasets that have been derived from experiments that

form the basis of research publications. Data journals are created specifically with the intent to allow authors to share this type of data in addition to the text that describes their results. These data journals may provide accurate links between datasets and their related scientific publications, which in turn can provide vital assistance to re-

searchers for the tasks of asserting research experiments and verifying research results.

This rapidly growing wealth of published scientific work has resulted in a pressing need for more effective processes towards reviewing scientific articles and research data, organizing data journals, as well as for improved tools and techniques for bibliographic analysis and management of scientometrics. Despite the proliferation of internet-based technologies and tools (including ones for data mining and semantic search) for organizing, searching and managing information, there are still significant obstacles towards the efficient and intelligent utilization of published scientific findings and data. In particular:

1. There are no universal well-structured repositories of scientific and research data for experimentation and benchmarking of pertinent research works in a given thematic area;
2. Current peer review processes remain fragmented, lengthy and, in several cases, weak and inefficient given the growing number of journals, magazines and conferences;
3. Conventional tools and metrics for assessing research work as well as individuals, institutions and organizations in terms of their scientific output and research contribution (e.g., h-index,¹ g-index,² impact factor³) are based on citation data that is usually not available without a license fee; and
4. Research articles are only poorly linked to data journals.

OpenScienceLink Platform Overview

OpenScienceLink is a project co-funded by the European Commission under the CIP-ICT-PSP 2012 work programme. The main motivation of OpenScienceLink is that the advent of tools, techniques and paradigms for open access to scientific information coupled with recent advances in the social semantic web can provide a remedy to the aforementioned problems, while also enabling a range of new business models for several stakeholders in the scientific publishing and academic value chains.

To this end, OpenScienceLink integrates, pilots and evaluates leading edge ICT infrastructures, applications and services for open access to scientific information that are intended to alleviate the lack of structured data journals and associated data models, the weaknesses of the review process, the poor linking of scientific information, as well as the limitations of current research evalua-

tion metrics and indicators. In particular, five pilot services are integrated and piloted.

Research dynamics-aware open access data journals development

This Pilot aims at covering the complete lifecycle of a data journal; from establishment and submission to evaluation and publishing. The first result of this effort is the *Biomedical Data Journal* (BMDJ), <http://biomed-data.eu>. The OSL platform provides the complete set of functionalities required for managing a journal:

- Via the “Author Activities” tab of the platform, authors can submit their datasets and manuscripts for any existing open calls;
- Via the “Reviewer Activities” tab of the platform, invited reviewers can view the review invitation with all required information, accept or reject the invitation and carry out the entire review process;
- For editors of the journal logged in to the platform, an additional “Editor Activities” tab is added. From here, editors can create calls, view the received submissions, oversee the reviewing process (cf. Pilot 2) and manage the different issues and volumes of the journal.

Novel, open, semantically-assisted peer review process

The first purpose of this Pilot is to cover the reviewing needs of BMDJ, but it is also intended to be used by other traditional, as well as data journals. Editors can get suggestions for reviewers and invite them directly from the aforementioned “Editor Activities” tab. After receiving the invitation, reviewers can go through the reviewing process via the platform (tab “Reviewer Activities”) with the help of automatically retrieved supporting information such as relevant research articles and datasets. Additionally, the platform will enable registered users to have post-review discussions about the published content in the near future.

Data mining for biomedical and clinical research trends detection and analysis

This Pilot intends to help users discover and analyse research trends in support of decisions concerning the allocation of research funding (both by private sponsors and governmental agencies) and the overall planning of research strategies. This functionality can be accessed from the “Trends” tab, and trends can be retrieved for free text queries, MeSH terms or combinations of the two.

Data Mining for Proactive Formulation of Scientific Collaborations

This Pilot has the goal of supporting the networking and collaboration of researchers and scholars working on scientific fields with collaboration potential. It is available at the “Collaborations” tab of the OpenScienceLink platform and enables the user to enter very specialized queries corresponding to their research area of interest. After submitting the query, co-authorship information is combined with social networking data to generate a collaboration network of important researchers in this area which is displayed to the user.

Scientific field-aware, productivity- and impact-oriented enhanced research evaluation services

The main objective of this Pilot is to introduce, produce and track new metrics of research and scientific performance with the goal of helping research sponsors, funding authorities and governmental agencies shape their research strategies and evaluate researchers fairly. We have developed a new metric that is based on public data, and more precisely on features that can be measured via the processing of openly accessible bibliographic data; the metric is therefore called “OpenScore.” The metric is used to assign scores to individual researchers. Future expansions will offer the computation of the metric for research groups, institutes, cities, countries and journals. From the “Evaluation” tab of the platform, the user will be able to retrieve the OpenScore for all these different scientific entities.

In addition to the five pilot functionalities described above, the OpenScienceLink platform offers faceted search functionality for all PubMed abstracts and all datasets published via BMDJ. Dataset metadata from additional databases will be added constantly, with direct access to the respective datasets if possible. The first iteration of the platform contains metadata for all functional genomics datasets indexed in the RapidExpress database. The “Search” tab offers a semantic search engine based on the existing GoPubMed⁴ engine for the biomedical domain that integrates all types of documents and datasets, available on the platform, into one result set. However, the

user can also narrow their search to specific resources they are interested in, e.g., ArrayExpress datasets.

Conclusion

As the role of the Open Data contribution towards *Open Science* is becoming more and more important, the OpenScienceLink platform supports the concept that the publication of datasets is part of a bigger and worthier idea than just a peer-reviewed article. From this perspective, in this paper we described the functionality of the OpenScienceLink platform for each of the five pilots implemented within the OpenScienceLink project, one of which is focused on the sharing and publication of datasets. The platform is available at <http://www.opensciencelink.org> to all interested users and aims at serving the wider concept of Open Science, adding transparency to the review process of manuscripts, bringing together researchers with shared interests, identifying trends, and reporting on the performance of the research entities via OpenScore, computed on the basis of publicly available bibliographic data.

Acknowledgement

The preparation of this manuscript was supported by the OpenScienceLink project (Grant agreement #318652).

References

- ¹ Hirsch JE. An index to quantify an individual’s scientific research output. *Proc Natl Acad Sci USA.* 2005 Nov 15;102(46):16569-72. DOI: 10.1073/pnas.0507655102.
- ² Egghe L. Theory and practice of the g-index. *Scientometrics.* 2006 Oct;69(1):131-52. DOI: 10.1007/s11192-006-0144-7.
- ³ Garfield E. The history and meaning of the journal impact factor. *JAMA.* 2006 Jan 4;295(1):90-93. DOI: 10.1001/jama.295.1.90.
- ⁴ Doms A, Schroeder M. GoPubMed: exploring PubMed with the gene ontology. *Nucleic Acids Res.* 2005 Jul 1;33(Web Server issue):W783-6. DOI: 10.1093/nar/gki470.