



CIP Competitiveness and Innovation Framework Programme  
CIP-Pilot Actions, 2007-2013  
CIP-ICT-PSP-2012-6

Project **CIP-Pilot 325101 / OpenScienceLink**  
Deliverable **D6.1.2**  
Distribution **Public**



OPENSOURCELINK

<http://www.opensciencelink.eu/>

## **Report on Pilot Sites Preparation Activities, Detailed Planning of the Pilot Operations and Training**

Authors: Julian Mendez, Matthias Zschunke, Efsthios Karanastasis, Eyfhimios Chondroyannis, Vasiliki Andronikou, George Tsatsaronis, Michael Schroeder, Adomas Bunevicius, Alicja Juskiene, Iordanis Mourouzis, Costas Pantos, Christian Pilarsky, Yixin Zhang, Freyja Van den Boom, Griet Verhenneman, Alessandro Pingitore, Francesca Mastorci, Todor Tagarev

Status: Final (Version 1.0)

February 2015

### Project

Project ref.no.	CIP-Pilot 325101
Project acronym	OpenScienceLink
Project full title	Open Semantically-enabled, Social-aware Access to Scientific Data
Project site:	<a href="http://www.opensciencelink.eu/">http://www.opensciencelink.eu/</a>
Project start	February 2013
Project duration	3 years
EC Project Officer	David Guedj

### Deliverable

Deliverable type	Report
Distribution level	Public
Deliverable Number	D6.1.2
Deliverable title	Report on Pilot Sites Preparation Activities, Detailed Planning of the Pilot Operations and Training
Contractual date of delivery	M21 (October 2014)
Actual date of delivery	February 2015
Relevant Task(s)	WP6/Task 6.1, 6.2, 6.3, 6.4
Partner Responsible	TI
Other contributors	NTUA, TUD, CNR, LUHS, NKUA, KU LEUVEN, PROCON
Number of pages	28
Author(s)	Julian Mendez, Matthias Zschunke, Efsthios Karanastasis, Eyfhimios Chondroyannis, Vasiliki Andronikou, George Tsatsaronis, Michael Schroeder, Adomas Bunevicius, Alicja Juskiene, Iordanis Mourouzis, Costas Pantos, Christian Pilarsky, Yixin Zhang, Freyja Van den Boom, Griet Verhenneman, Alessandro Pingitore, Francesca Mastorci, Todor Tagarev
Status & version	Final (Version 1.0)
Keywords	Pilot Sites, Preparation Activities, Training

---

## Executive Summary

---

The current deliverable describes the mobilization of resources during the second year of the project for the training activities of the users and the promotion of the pilots at the pilot sites. The presentation is focused on the specific activities performed at the consortium pilot sites, and the resources (infrastructure and personnel) mobilized to accomplish the pilot operations. The presentation about the mobilized infrastructure focuses on the required information and communication technology applications and networking needed at the pilot sites. The presentation about the personnel focuses on getting researchers and, in general users, involved in the project and trained at the pilot operations.

# Contents

Executive Summary .....	3
Contents.....	4
List of Figures .....	6
List of Tables .....	7
1 Introduction.....	8
1.1 Summary of the OpenScienceLink pilots from the user's perspective .....	8
1.1.1 Pilot 1: Research dynamics-aware open access data journals development.....	8
1.1.2 Pilot 2: Novel open, semantically-assisted peer review process .....	10
1.1.3 Pilot 3: Data mining for biomedical and clinical research trends detection and analysis	13
1.1.4 Pilot 4: Identifying scientific collaborations.....	13
1.1.5 Pilot 5: Evaluation of research.....	14
1.2 Aspects of Launching, Disseminating, and Sustaining the OpenScienceLink Platform ..	15
1.2.1 Launching and Dissemination .....	15
1.2.2 Users Training.....	16
1.2.3 Sustainability after launching .....	16
2 Launching and Dissemination.....	17
2.1 Pilot Sites .....	17
2.2 Encouragement of existing user bases .....	17
2.2.1 GoPubMed users.....	17
2.2.2 Biotec, CRTD, BCube Dresden .....	17
2.2.3 PubMed authors .....	17
2.2.4 Mailing Lists .....	17
3 Users Training .....	18
3.1 Non-interactive Training .....	18
3.1.1 User Guide / Manual.....	18
3.1.2 Help Pages / In-place Help / FAQ .....	18
3.1.3 Tutorials .....	18
3.1.4 Call for Papers and Instructions for Datasets submission.....	18
3.1.5 "Did you know...?" .....	18
3.1.6 Summary and Updated Schedule of Non-interactive Training Activities and Materials	19
3.2 Interactive Training .....	19
3.2.1 Seminars/Workshops/Conference Presentations on Using Biomedical Search Technologies .....	19
3.2.2 Academic Lecture Tutorials.....	19
3.2.3 Summary and Schedule of Interactive Training Activities and Materials .....	20
4 Sustainability.....	22

4.1	Sustainable Platform Aspects .....	22
4.1.1	Sustaining Services .....	22
4.1.2	Sustaining Feature Evolution.....	23
4.1.3	Server Hosting and Platform / Service Maintenance.....	24
4.2	Sustainability Methodology .....	24
4.2.1	3rd Party Funding .....	24
4.2.2	Customizing Services .....	24
4.2.3	Link with Open Research and Open Science Projects and Initiatives, and Existing Information providers .....	24
5	Summary and Conclusions.....	25
6	References .....	26
7	APPENDIX I: Dissemination of the OSL platform and the BMDJ at the existing user pools	27
8	APPENDIX II: TUD Seminar on “Semantic Search in the Life Sciences in the Era of Open Access” .....	28

---

## List of Figures

---

Figure 1: Registering to the OpenScienceLink platform.....	8
Figure 2: Registering to the OpenScienceLink platform.....	9
Figure 3: Managing the issues in the OpenScienceLink platform with editor user privileges. ....	10
Figure 4: Finding potential reviewers. ....	11
Figure 5: Inviting a reviewer. ....	11
Figure 6: Filling the review form. ....	12
Figure 7: Visualization of trends in the biomedical domain.....	13
Figure 8: Visualization of collaborations. ....	14
Figure 9: Visualization of metrics for evaluating science and academic performance. ....	15
Figure 10: A novelty of the OSL platform: Semantic search in data repositories. ....	23

---

## List of Tables

---

Table 1: Summary of non-interactive training schedule.....	19
Table 2: Summary of interactive training schedule.....	21

# 1 Introduction

This section gives a short summary to the five OpenScienceLink pilots (OpenScienceLink Consortium, 2013) and their main operations. A description of what the users need in order to use the OpenScienceLink pilot is given, which is the basis for the executed and planned training activities at the project's pilot sites.

## 1.1 Summary of the OpenScienceLink pilots from the user's perspective

In the following, we present a summary of each of the five pilots from the user's perspective. The presentation is focused on the main points that the users need to know, and learn, in order to use the five OpenScienceLink pilots. As a general requirement, all users need to register to the OpenScienceLink platform (<http://opensciencelink.org>) in order to be able to use the pilots. This is a very easy process that takes only 1-2 minutes for a new user to perform, the only actual requirement being that the user needs to have a valid e-mail address. In **Figure 1** a screenshot from the registration process is illustrated. The user needs to provide some personal data (name, surname, affiliation) and mark some of her research interests, which are used as default queries for pilot 3 (trend detection), pilot 4 (research collaborations), and pilot 5 (evaluation of science). The declared research interests are also used by the platform in pilot 2 (peer-reviewing) in order to match the user as a candidate reviewer to a submitted data article for the Biomedical Data Journal.

The screenshot shows the OpenScienceLink registration interface. At the top, there's a navigation bar with links like Logout, John Doe (My Profile), Author Activities, Reviewer Activities, Editor Activities, Trends, Collaborations, Evaluation, and Search. The main content area is titled 'Personal information' and contains fields for E-mail Address, First Name, Surname, Password, and Confirm Password. Below these are checkboxes for 'You accepted the OpenScienceLink Terms and Conditions' and 'You agreed to the OpenScienceLink Privacy Notice'. The 'Professional information' section follows, with fields for Editor Privileges, Institution, Type of Institution, Country, and Job. The 'Research topics' section includes a list of topics with checkboxes for selection. The footer contains copyright information and a Creative Commons license logo.

**Figure 1: Registering to the OpenScienceLink platform.**

### 1.1.1 Pilot 1: Research dynamics-aware open access data journals development

The usage of Pilot 1 can be seen from the perspective of two user types. The first is the perspective of the researcher user, who can upload and submit her datasets to the Biomedical Data Journal. The submission process is very simple and fast. The user needs only to fill the metadata information and upload the actual



data set as an archive, along with a description of the dataset as a pdf file, following the template style offered by the BMDJ journal. The metadata information contains fields such as affiliation, dataset name, the current status of the submission, the issue to which the dataset was submitted, the date and time of the last update, a short description and selected keywords. For any user, it is also possible to perform some operations like to browse older submissions, and manage them, provided that the review process has not started. A screenshot taken from the platform for the process of submitting a dataset is shown in **Figure 2**.

OpenScienceLink PL...

https://www.gopubmed.org/web/oslplatform/107WEB01591q5dw5kprri3jib100f00101000200010.y

Google

Logout John Doe (My Profile) Author Activities Reviewer Activities Editor Activities Trends Collaborations Evaluation Search

### Upload Dataset

Before submission, please read our [Submission Guidance](#).

User: John Doe

Affiliation:

Dataset Name:

Description:

Call/Issue Name:

Main Topics:

Minor Topics:

Please list the PubMed IDs of the articles that the submitted dataset has been used in. Multiple IDs should be separated by commas, e.g.: 22345678, 2345786, 2223339

Papers that used the dataset:

Category:

☐ I accept the [Terms and Conditions](#), and have read the [Submission Guidance](#).

Please consult the [instructions to authors](#) on how your article should be structured and what type of information is required to describe your dataset. Your zip file can contain files of any type, e.g., .pdf, .txt, .xls, .xlsx, .doc, .docx,...

Article path:  No file selected.

Dataset path (zip):  No file selected.

or provide a URL:

OpenScienceLink is co-funded by the European Commission under the CIP-ICT-PSP 2012 work programme topic 2.2: "Open Data and Open Access to Scientific Information". Copyright © 2014-2015 OpenScienceLink. The entire website content is distributed under the Creative Commons 4.0 Attribution-NonCommercial-ShareAlike International license (CC BY-NC-SA 4.0).

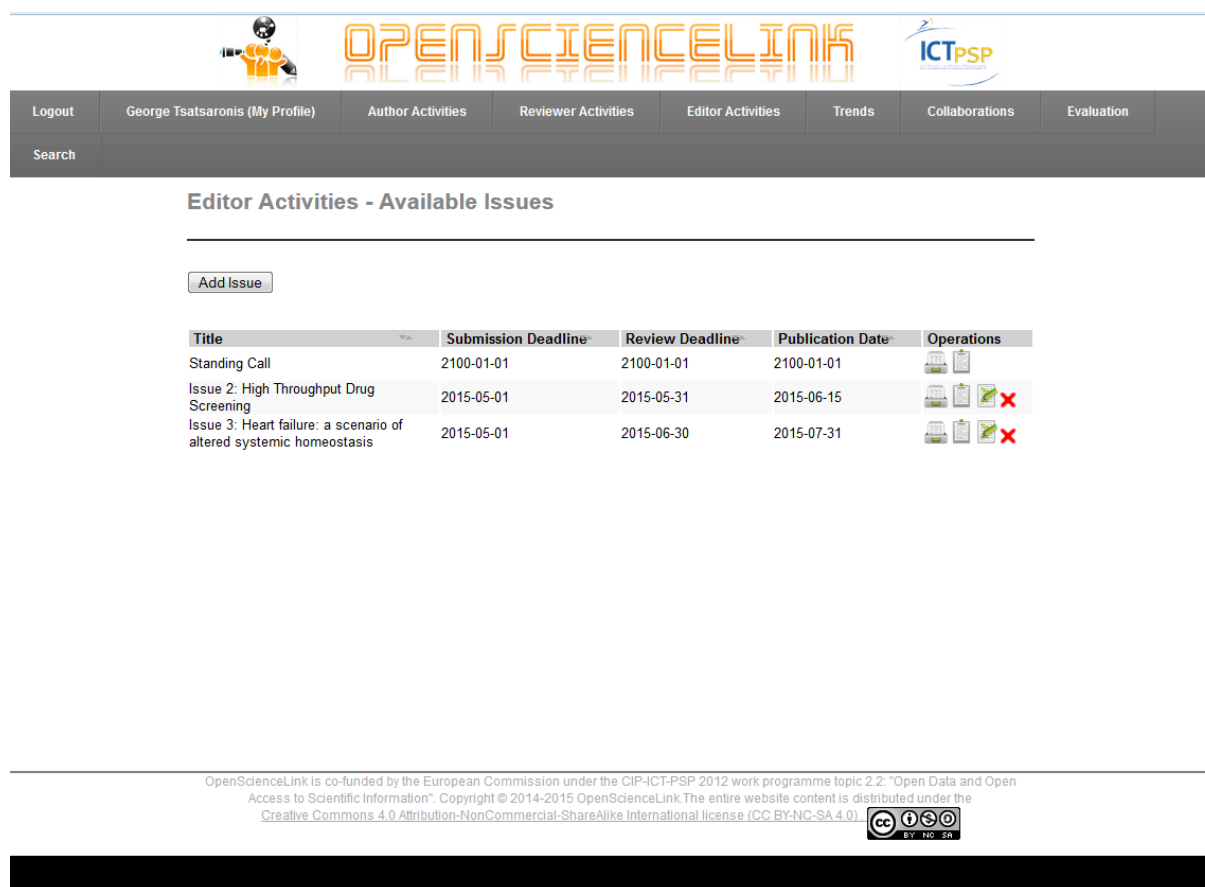
**Figure 2: Registering to the OpenScienceLink platform.**

The user can select to submit the dataset for any open call for a special issue, or for the 'Standing Call', if it is not addressed to any particular issue. In addition, the dataset can be 'research data' or a 'clinical trial'. The user has the option to either upload the dataset as a single ZIP file, or to provide the URI of the dataset. The acceptance of the Terms and Conditions<sup>1</sup> is mandatory to submit the dataset. The user must also read the Submission Guidance<sup>2</sup>. Finally, the user can always monitor the process of the submitted dataset. Its status should follow the following states: 'submitted', 'under review', 'reviewed', and (if accepted) 'published'.

<sup>1</sup><http://opensciencelink.eu/bmdj-copyright-and-licensing-policy/>

<sup>2</sup><http://biomed-data.eu/content/submission-guidance>

From the editor's perspective, the current issues, as well as calls for new special issues can be managed via the OpenScienceLink platform. The following screenshot illustrates the managing of the issues from the editor's perspective.



**Editor Activities - Available Issues**

[Add Issue](#)

Title	Submission Deadline	Review Deadline	Publication Date	Operations
Standing Call	2100-01-01	2100-01-01	2100-01-01	
Issue 2: High Throughput Drug Screening	2015-05-01	2015-05-31	2015-06-15	
Issue 3: Heart failure: a scenario of altered systemic homeostasis	2015-05-01	2015-06-30	2015-07-31	

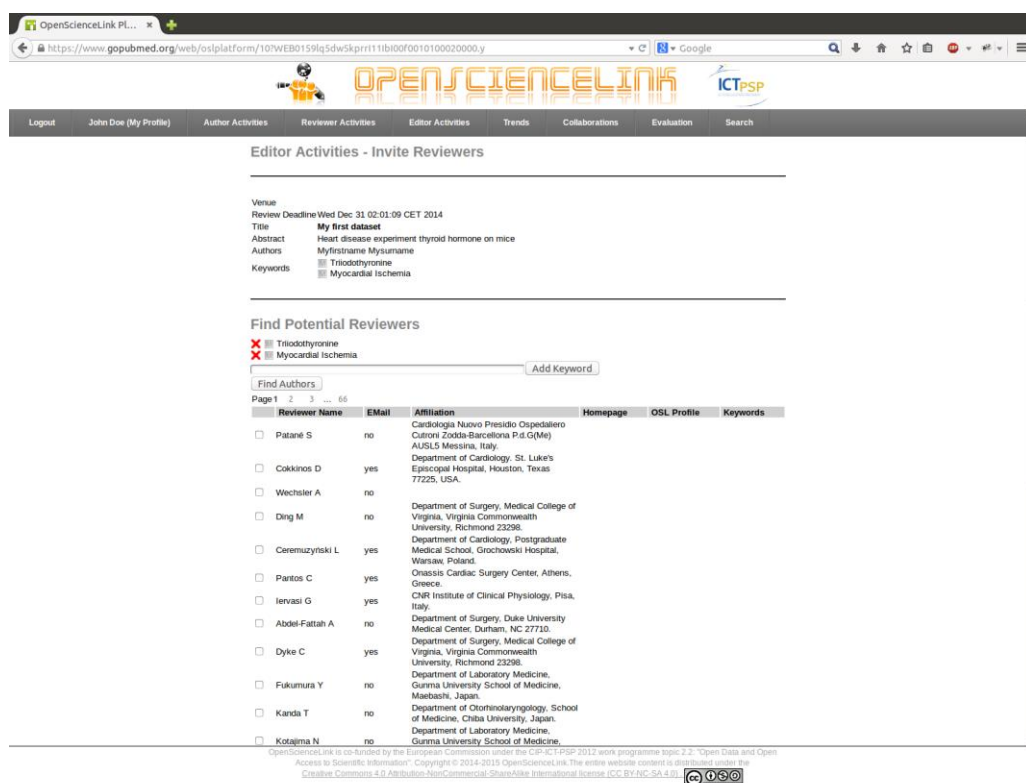
OpenScienceLink is co-funded by the European Commission under the CIP-ICT-PSP 2012 work programme topic 2.2: "Open Data and Open Access to Scientific Information". Copyright © 2014-2015 OpenScienceLink. The entire website content is distributed under the Creative Commons 4.0 Attribution-NonCommercial-ShareAlike International license (CC BY-NC-SA 4.0).

**Figure 3: Managing the issues in the OpenScienceLink platform with editor user privileges.**

### 1.1.2 Pilot 2: Novel open, semantically-assisted peer review process

Pilot 2 aims at offering a transparent peer-review process of the datasets and their accompanying descriptions submitted at the Biomedical Data Journal via the OpenScienceLink platform. As part of the process the platform has distinguished roles for *authors*, *reviewers*, and *editors*. After an author submits a dataset, the editors are informed via e-mail for the submission and they can see it in the list of submitted datasets for a specific issue. An editor who finds the submission can verify that the dataset corresponds to the journal's scope, and that it meets some scientific standards. The editor can create a 'review call', which starts the review process, or can alternatively write directly an e-mail to the authors, noting any problems with the submission. The editor can see the list of review calls (or review invitations), and can select the newly created review call. Following a link, the author can select reviewers that are automatically recommended or write the final review, based on the reviewers' comments.

The selection of reviewers uses the topics provided in 'main topics' to retrieve potential reviewers. A list of potential reviewers is shown after running a search using the main topics or other provided topics. **Figure 4** shows an example list for the topics '*Triiodothyronine*' and '*Myocardial Ischemia*'. This list of candidates contains the reviewer name, affiliation, and whether the e-mail address is available or not.



OpenScienceLink Platform

https://www.gopubmed.org/web/osplatform/107WEB0159q5dWskpr111b00f0010100020000.y

Logout John Doe (My Profile) Author Activities Reviewer Activities Editor Activities Trends Collaborations Evaluation Search

### Editor Activities - Invite Reviewers

Venue: Review Deadline Wed Dec 31 02:01:09 CET 2014

Title: My first dataset

Abstract: Heart disease experiment thyroid hormone on mice

Authors: Myfirstname Mysurname

Keywords: ☒ Triiodothyronine ☒ Myocardial Ischemia

### Find Potential Reviewers

☒ Triiodothyronine ☒ Myocardial Ischemia

Find Authors

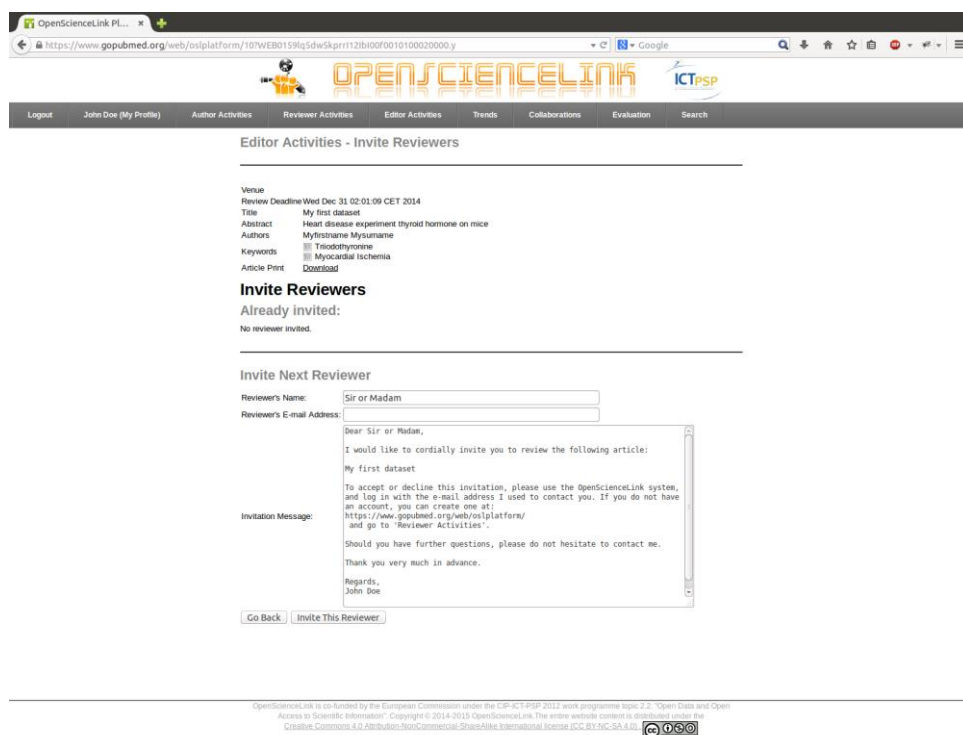
Page 1 2 3 ... 65

Reviewer Name	Email	Affiliation	Homepage	OSL Profile	Keywords
<input type="checkbox"/> Patané S	no	Cardiologia Nuovo Presidio Ospedaliero Cattone Zucchi-Bacchioni P.d.G(Me) AUSL D. Messina, Italy			
<input type="checkbox"/> Cokkino D	yes	Department of Cardiology, St. Luke's Episcopal Hospital, Houston, Texas 77225, USA.			
<input type="checkbox"/> Wechsler A	no				
<input type="checkbox"/> Ding M	no	Department of Surgery, Medical College of Virginia, Virginia Commonwealth University, Richmond 23298			
<input type="checkbox"/> Ceremuzynski L	yes	Department of Cardiology, Postgraduate Medical School, Grochowski Hospital, Warsaw, Poland.			
<input type="checkbox"/> Pantos C	yes	Onassis Cardiac Surgery Center, Athens, Greece.			
<input type="checkbox"/> Iervasi G	yes	CNR Institute of Clinical Physiology, Pisa, Italy.			
<input type="checkbox"/> Abdel-Fattah A	no	Department of Surgery, Duke University Medical Center, Durham, NC 27710.			
<input type="checkbox"/> Dyke C	yes	Department of Surgery, Medical College of Virginia, Virginia Commonwealth University, Richmond 23298			
<input type="checkbox"/> Fukumura Y	no	Department of Laboratory Medicine, Gunma University School of Medicine, Maebashi, Japan.			
<input type="checkbox"/> Kanda T	no	Department of Otorhinolaryngology, School of Medicine, Chiba University, Japan.			
<input type="checkbox"/> Kotajima N	no	Department of Laboratory Medicine, Gunma University School of Medicine.			

OpenScienceLink is co-funded by the European Commission under the CP-ICT-PSP 2012 work programme topic 2.2: "Open Data and Open Access to Scientific Information". Copyright © 2014-2015 OpenScienceLink. The entire website content is distributed under the Creative Commons 4.0 Attribution-NonCommercial-ShareAlike International license (CC BY-NC-SA 4.0).

**Figure 4: Finding potential reviewers.**

Once a group of potential reviewers is selected, the editor can use an automatically generated form to write an e-mail message to each invited reviewer, as shown in **Figure 5**. The platform provides a default message which can be freely changed by the editor. This message is the invitation sent by the editor.



OpenScienceLink Platform

https://www.gopubmed.org/web/osplatform/107WEB0159q5dWskpr111b00f0010100020000.y

Logout John Doe (My Profile) Author Activities Reviewer Activities Editor Activities Trends Collaborations Evaluation Search

### Editor Activities - Invite Reviewers

Venue: Review Deadline Wed Dec 31 02:01:09 CET 2014

Title: My first dataset

Abstract: Heart disease experiment thyroid hormone on mice

Authors: Myfirstname Mysurname

Keywords: ☒ Triiodothyronine ☒ Myocardial Ischemia

Article Print: Download

### Invite Reviewers

Already invited:

No reviewer invited.

### Invite Next Reviewer

Reviewer's Name:

Reviewer's E-mail Address:

Invitation Message:

Dear Sir or Madam,

I would like to cordially invite you to review the following article:

My first dataset

To accept or decline this invitation, please use the OpenScienceLink system, and log in with the e-mail address I used to contact you. If you do not have an account, you can create one at: <https://www.gopubmed.org/web/osplatform/> and go to "Reviewer Activities".

Should you have further questions, please do not hesitate to contact me.

Thank you very much in advance.

Regards,  
John Doe

OpenScienceLink is co-funded by the European Commission under the CP-ICT-PSP 2012 work programme topic 2.2: "Open Data and Open Access to Scientific Information". Copyright © 2014-2015 OpenScienceLink. The entire website content is distributed under the Creative Commons 4.0 Attribution-NonCommercial-ShareAlike International license (CC BY-NC-SA 4.0).

**Figure 5: Inviting a reviewer.**

The receiver of the review invitation needs to have or create an OpenScienceLink user account to accept or reject the invitation. To make the decision, the potential reviewer can see the dataset and the related article (if provided). Once the invitation is accepted, the reviewer can fill out a predefined form with the main points to consider in the review. Most of these points are shown in **Figure 6**. The content of the form is verbalized, and can be sent and read as a text.

OpenScienceLink PL...

https://www.gopubmed.org/web/oslplatform/4?WEB0z3se4p7b846k1819100f0010100003

Google

Logout John Doe (My Profile) Author Activities Reviewer Activities Editor Activities Trends Collaborations Evaluation Search

### Submit Review

User: **John Doe**  
 Name: Examining Heart Attacks in Young Women  
 Description: Examining Heart Attacks in Young Women  
 Article path: Examining Heart Attacks in Young Women  
 Dataset path: Examining Heart Attacks in Young Women  
 URL: Examining Heart Attacks in Young Women

Recommendation

Is the description of the methodology used detailed enough?

Have the authors previous publications based on the same methodologies described in the submitted work?

Are there missing data in the submitted dataset?

The number of subjects in the submitted dataset is

The number of variables measured per subject in the submitted dataset is considered

Do the authors provide insights on how these datasets could be potentially used for computational modeling?

Language

List of references

Are there ethical issues concerning the submitted work?

Strengths (Optional)

OpenScienceLink is co-funded by the European Commission under the CIP-ICT-PSP 2012 work programme topic 2.2: "Open Data and Open Access to Scientific Information". Copyright © 2014-2015 OpenScienceLink. The entire website content is distributed under the Creative Commons 4.0 Attribution-NonCommercial-ShareAlike International license (CC BY-NC-SA 4.0)

CC BY-NC-SA 4.0

**Figure 6: Filling the review form.**

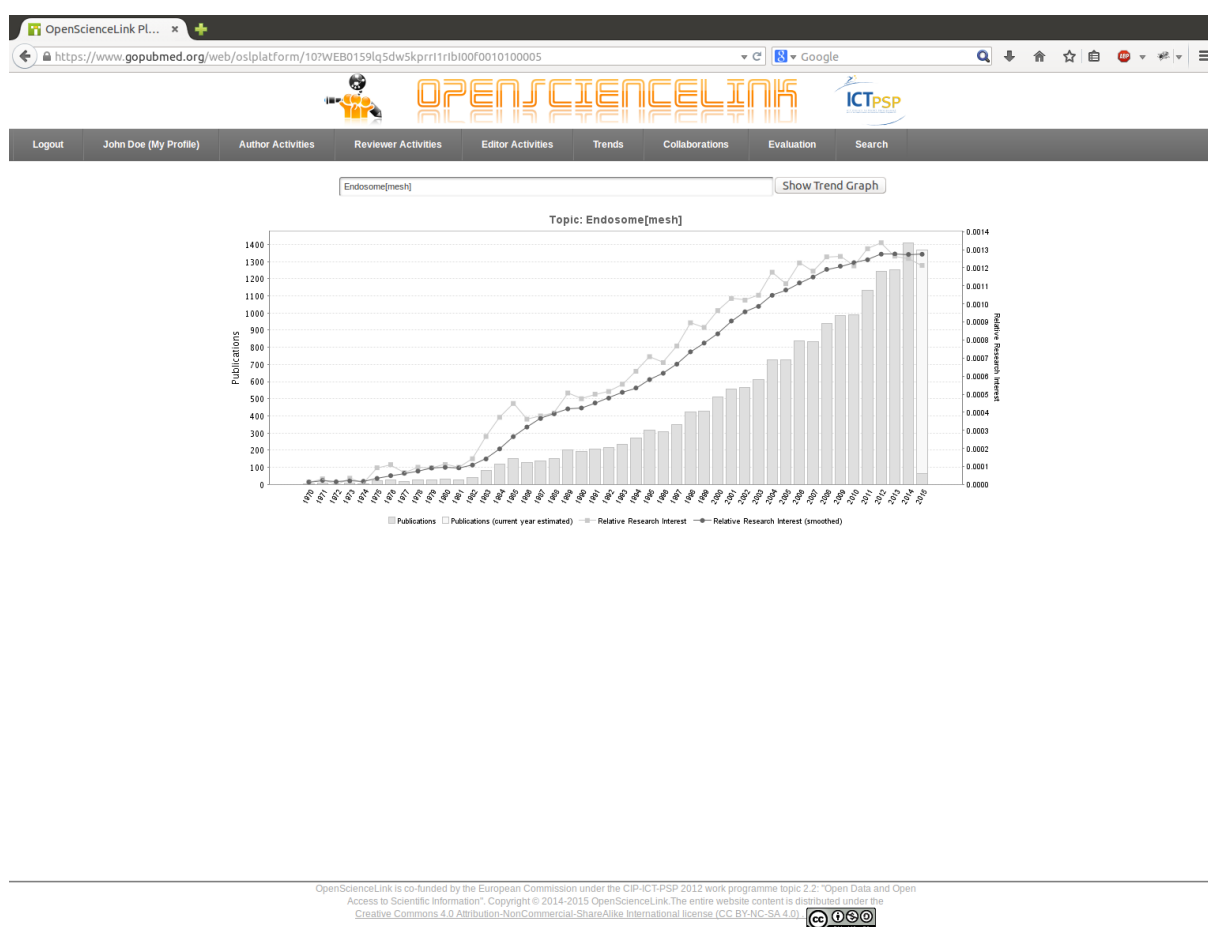
The completed reviews can be browsed by the editor who makes the final decision for the submitted paper. The final decision with the individual reviews is communicated to the author. If positive, the dataset is published at the respective planned issue of the Biomedical Data Journal.

The publication process from the platform perspective is simple. The editor requests a digital object identifier (DOI) from the publisher, Procon Ltd. The article and the dataset have two very similar, but distinct DOIs. One links to the article, technically stored by Procon Ltd, and the other one to the dataset, technically stored by Transinsight GmbH. The editor must carefully give these two identifiers to the published dataset, and this last step finishes the process. Thereafter, the dataset is available for online preview and becomes also searchable over the OSL platform. Further alterations, like changing the issue of a submission, or fixing an orthographic error in the title, are handled individually by Transinsight GmbH and Procon Ltd, of course in coordination with the editorial board.

### 1.1.3 Pilot 3: Data mining for biomedical and clinical research trends detection and analysis

The aim of Pilot 3 is to discover and analyse research trends in open access biomedical and clinical research texts. The identification and analysis of such trends is essential in order to allocate research funding, to plan research strategies, to evaluate the state of play of existing journals, and to strategically decide the expansion of journals into hot, trendy topics.

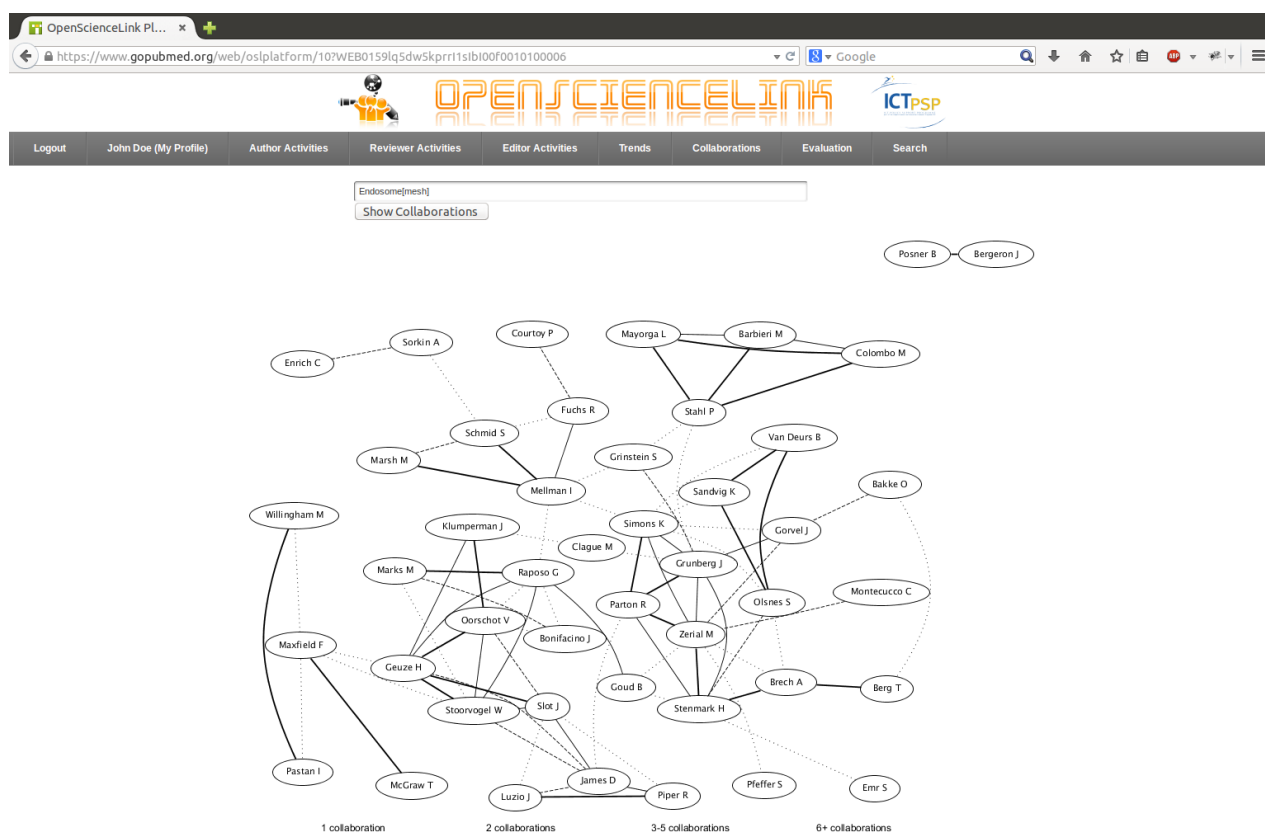
The trends are shown with bars and lines, as it is illustrated in **Figure 7**. The bars represent the number of publications related to the given topic. The lines are relative frequencies normalized by the overall number of publications per year. There are two line variants; the grey line shows the relative frequencies per year and the black line shows a five year sliding average of the grey line.



**Figure 7: Visualization of trends in the biomedical domain.**

### 1.1.4 Pilot 4: Identifying scientific collaborations

Scientific collaborations are crucial for the development of science. The main purpose of Pilot 4 is to present a graph of potential collaborations, as the one illustrated in **Figure 8**. In the text box, the user can enter any particular topic or the name of an author. After giving this information, the platform computes a graph containing the network. The vertices (or nodes) in the graph contain the authors. The edges represent collaboration between two authors, in at least one published research work. The thickness of an edge represents the number of collaborations between the authors it connects.



**Figure 8: Visualization of collaborations.**

### 1.1.5 Pilot 5: Evaluation of research

The aim of Pilot 5 is to introduce novel metrics for the evaluation of research and of academic performance, beyond conventional metrics associated with conventional indices and impact factors, and based exclusively on openly accessible data and information. Via the OSL platform the metrics are currently based on the counts of published works, per author, journal, city and country. The counts of countries, cities, journals, and authors are presented in different tables as shown in **Figure 9**. Their content is dynamically generated based on the text entered in the text box. Each row in each table contains a horizontal bar to show the percentage with respect to the top.

A novelty of the OSL platform, is the introduction of the OpenScore. This is currently integrated only for authors, but in future releases it will also be available for journals, cities and countries. The OpenScore is produced by a supervised machine learning model that uses a set of 12 publicly available features. The model was trained on a development set of selected authors (10,000 authors). The selection was made using stratified sampling to include authors which cover a large spectrum of h-indexes. The h-indexes were computed for these authors for the purposes of training the model, from the ISI Web of Science citation services. Therefore, the model is trained to predict the impact of the research works of the authors, trained on h-indexes. The features that were used per author are: (1) number of papers, (2) number of distinct journals the author published at, (3) number of distinct MeSH terms annotating the author's work, (4) number of cumulative MeSH terms annotating the author's work, (5) number of MeSH trees covering the areas of the author's published work, (6) number of years since first paper was published, (7) number of years since last paper was published, (8) number of years the author had at least one paper published

(active years), (9) distinct number of co-authors, (10) cumulative number of co-authors, (11) number of author's affiliations since the beginning of his/her career, and, (12) the author's PageRank score, computed on the co-authorship graph that is created from all PubMed indexed articles.

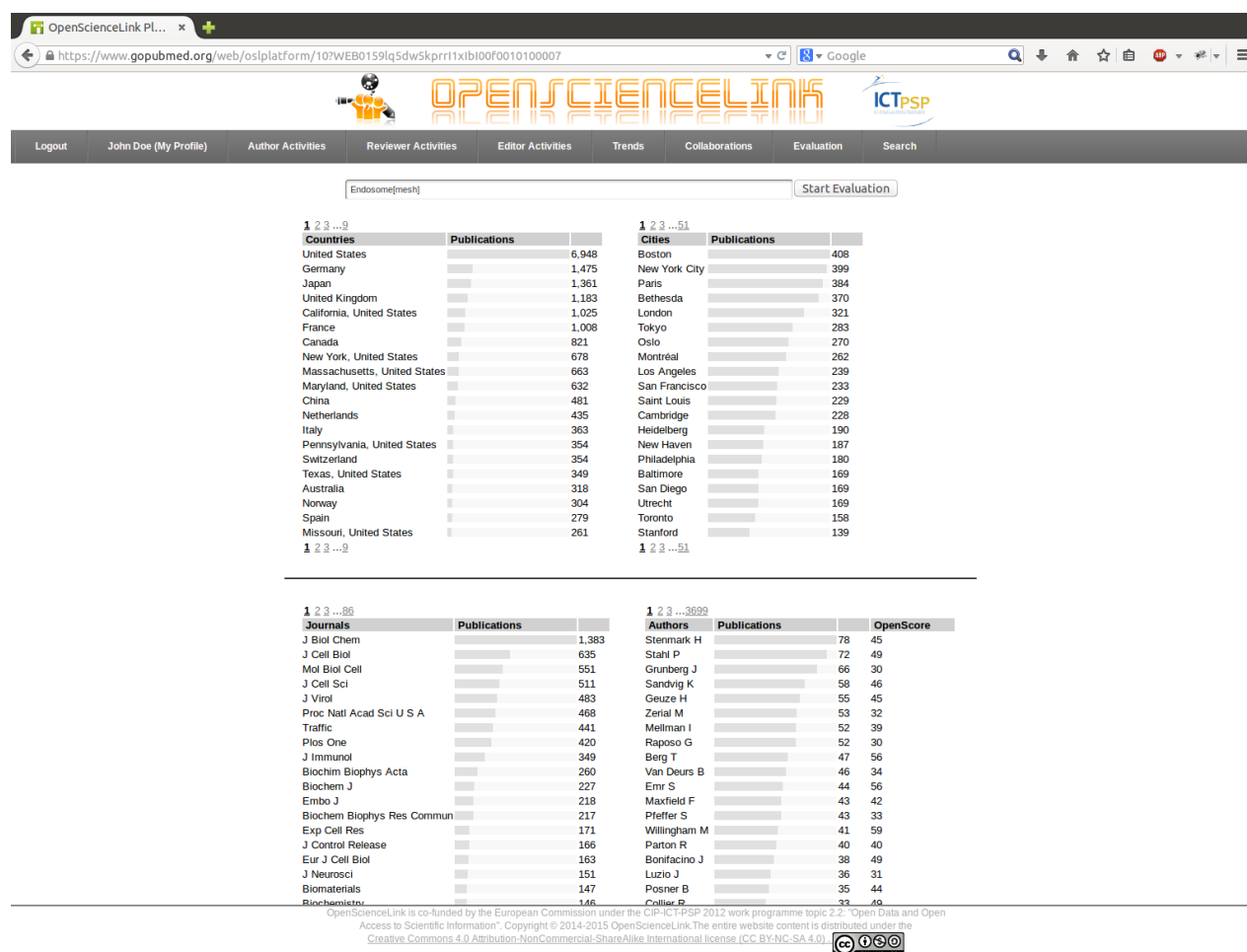


Figure 9: Visualization of metrics for evaluating science and academic performance.

## 1.2 Aspects of Launching, Disseminating, and Sustaining the OpenScienceLink Platform

In the following the main directions of dissemination, training and sustainability are summarized for the OpenScienceLink platform and the Biomedical Data Journal (BMDJ: ISSN 2367-5322). Analytical descriptions of the dissemination activities can be found in D9.2.2 (Tsatsaronis, et al., 2015) and of the sustainability plans in D9.4.2 (Andronikou, et al., 2014).

### 1.2.1 Launching and Dissemination

The platform has been launched, and runs on a stable server environment. It can be accessed in the following address: <http://opensciencelink.org>. An analytical description of the releases has been reported in D7.1.2 (Bunevicius, et al., 2015). The following major steps are directed to attract users. This dissemination can be achieved through different channels like mailing lists, journal articles, tutorials, seminars, workshops and conferences.



### **1.2.2 Users Training**

The user training can be conducted with the help of online tutorials, videos, lectures and seminars delivered by the consortium members. The organisation of workshops, and live demonstrations of the platform, can contribute importantly in bringing the interest and simplifying the teaching of new users and their familiarization with the five pilots.

### **1.2.3 Sustainability after launching**

A critical aspect for long-term success is the sustainability of the platform and the Biomedical Data Journal in the after project era. Two main points have been considered: (1) to ensure the technical services availability, (2) to keep the underlying platform data and resources up-to-date, and, (3) to sustain the interest of researchers and other stakeholders to the Biomedical Data Journal.



## 2 Launching and Dissemination

### 2.1 Pilot Sites

In this section we discuss the involvement of the pilot sites. This is based on the four pilot sites, namely Dresden (Germany), Athens (Greece), Kaunas (Lithuania), and Pisa (Italy). Dresden participates with four units: (1) the Bioinformatics Group of the Technische Universität Dresden (TUD), which is also the project coordinator, (2) the B CUBE (Center for Molecular Bioengineering), (3) the Universitätsklinikum Carl Gustav Carus Dresden, and, (4) Transinsight GmbH (TI), which provides technical assessment to develop and connect the OpenScienceLink platform with the journal and the users. The pilot site in Kaunas participates with the Behavioural Medicine Institute of the Lithuanian University of Health Sciences (LUHS). The pilot site in Athens participates with two units: (1) the Medical School of the National and Kapodistrian University of Athens (NKUA), and, (2) the Distributed Knowledge and Media Systems Group of the National Technical University of Athens (NTUA). The pilot site in Pisa participates with the Institute of Clinical Physiology of the National Research Council (CNR). The CNR established two connections; primarily the connection with the CNR Central Library, and with the Research Hospital Tuscany Region G Monasterio Foundation (Research Hospital). The volume and capacity of the aforementioned pilot sites has been reported analytically in D6.1.1 (Zschunke, et al., 2014).

### 2.2 Encouragement of existing user bases

In this section we discuss the involvement of the pilot sites towards encouraging existing user bases. In all cases, a text was disseminated to these user pools, via e-mail, in order to invite them to use the OpenScienceLink platform and the submit datasets at the Biomedical Data Journal. The text that was used as a basis in all cases can be found in Appendix I. The text was adapted per case, to become more personalized to the user community that addressed.

#### 2.2.1 GoPubMed users

GoPubMed is the first semantic search in the life science domain. It performs semantic search on PubMed content using the Medical Subject Headings (MeSH). GoPubMed is available for the biomedical research community since 2003. All of the GoPubMed users have been invited to test the pilots of the OpenScienceLink platform, and explore the new journal as a means of publishing their datasets. The number of the users invited was approximately 4,300.

#### 2.2.2 Biotec, CRTD, BCube Dresden

Three research entities are active in the pilot site in Dresden. Biotec, which is also the host of the Bioinformatics group that co-ordinates the OpenScienceLink project, has approximately 300 active researchers, while BCube, which is also one of the OpenScienceLink project partners, has approximately 100 researchers. CRTD, co-located in the same premises with the other two entities, has approximately 150 research employees. All of them have been invited to use the OpenScienceLink platform and explore the Biomedical Data Journal as a means to publish their data sets.

#### 2.2.3 PubMed authors

PubMed, which provides an index of biomedical publications, contains also information of author affiliation. This information gives pointers to acknowledged research groups for specific life science fields. All of the PubMed authors are used as a reviewer pool in the OpenScienceLink platform.

#### 2.2.4 Mailing Lists

In addition to the aforementioned invitations to external user pools, the OpenScienceLink consortium has utilized a number of publicly available mailing lists in order to advertise the platform and the journal. The lists that were used are: DBWORLD, BioNLP, PASCAL Network, elsnet, UIB, and KDNNet.

## 3 Users Training

### 3.1 Non-interactive Training

In the following we analyse the different non-interactive training activities conducted within year 2 and we array the plan for year 3 of the project.

#### 3.1.1 User Guide / Manual

The user guide or manual is intended to explain how the platform works. It contains different use cases to exemplify the use of the platform, and the five pilots. The user guide will be available for download from the platform on April 2015 (M27), as a pdf document. It will also be uploaded at the project's web site. In parallel, a document with user guidelines on the legal aspects of the OpenScienceLink platform will be prepared and uploaded to the platform as a future guide for all platform users. The document will be prepared by the consortium's legal experts, namely KU Leuven.

#### 3.1.2 Help Pages / In-place Help / FAQ

Help pages and Frequently Asked Questions (FAQ) are fast ways to train users, especially if they show concrete examples. Help and FAQ pages will be deployed in the platform during April 2015 (M27).

#### 3.1.3 Tutorials

Tutorials can give the user a complete understanding of the available services. Each tutorial can be HTML, PDF, or a video, although it always focuses on explaining each functionality step by step. A series of pdf tutorials (one for each pilot), and a video tutorial introducing the OpenScienceLink platform and the Biomedical Data Journal will be released during May 2015 (M28).

#### 3.1.4 Call for Papers and Instructions for Datasets submission

During November 2014, the Biomedical Data Journal was preparing its official launching. For this purpose, a series of documents were created to: (1) publicize the launching of the journal, and, (2) provide concrete guidelines on how the authors can submit their papers at the Biomedical Data Journal. The former point comprised the announcement of the call for papers for the upcoming special issues, namely:

- (1) High Throughput Drug Screening. Special Issue Editor: Yixin Zhang
- (2) Heart Failure and Stress Response. Special Issue Editors: Alessandro Pingitore and Giorgio Iervasi
- (3) Biomarkers in Brain Disorders: Towards a Global Database. Special Issue Editors: Adomas Bunevicius and Arimantas Tamasauskas

For the guidelines to authors, specific instructions were created, with respective style templates and guidance on how to fill them, for the five types of papers that the BMDJ accepts as submissions: (a) editorial, (b) policy, (c) data, (d) data-based modeling and models paper, and, (e) commentary. These instructions, with the respective template styles and guidance for authors can be found at:

<http://www.biomed-data.eu/content/submission-guidance>

#### 3.1.5 "Did you know...?"

A "Did you know...?" panel showing tips on startup can arouse interest of the users in specific functionalities. This feature will be deployed in the platform during May 2015 (M28).

### 3.1.6 Summary and Updated Schedule of Non-interactive Training Activities and Materials

Table 1 summarizes the activities mentioned above, and gives a plan for the third year of the project.

Type	Planned Delivery Date
<b>YEAR 2</b>	
<b>Instructions for Authors, Styles and Author Support</b>	<b>M22 (November 2014). Material released under:</b> <a href="http://www.biomed-data.eu/content/submission-guidance">http://www.biomed-data.eu/content/submission-guidance</a> <b>Special issues call for papers with description of focus and instructions announced at:</b> <a href="http://www.biomed-data.eu/active-calls-for-papers">http://www.biomed-data.eu/active-calls-for-papers</a>
<b>YEAR 3</b>	
<b>User Guide / Manual</b>	<b>M27 (April 2015)</b>
<b>Help Pages</b>	<b>M27 (April 2015)</b>
<b>FAQ</b>	<b>M27 (April 2015)</b>
<b>Tutorials / Video-Tutorials</b>	<b>M28 (May 2015)</b>
<b>"Did you know...?"</b>	<b>M28 (May 2015)</b>
<b>User Guidelines on Legal Aspects of the OSL Platform</b>	<b>M28 (May 2015)</b>

Table 1: Summary of non-interactive training schedule.

## 3.2 Interactive Training

In the following we analyse the different interactive training activities conducted within year 2 and we array the plan for year 3 of the project.

### 3.2.1 Seminars/Workshops/Conference Presentations on Using Biomedical Search Technologies

All the pilot sites consortium partners are engaged into activities that explain how the semantic-enabled technologies work in the biomedical domain, and how users can benefit from the usage of the OpenScienceLink platform and its five pilots. A series of seminars, workshops and conference presentations have been conducted by the partners within year 2, and another series is planned for the third year of the project. Their description can be found in **Table 2**.

#### 3.2.2 Academic Lecture Tutorials

Since many participating partners are universities, this creates a good opportunity to reach current and future researchers. Short videos or presentations in academic lecture-tutorials could arouse the interest of scientists and provide an easy introduction to the platform. For this purpose, all partners who are engaged into the delivery of academic lectures, have incorporated examples of the usage of the OpenScienceLink platform into their lectures. Their description can be found in **Table 2**.

### 3.2.3 Summary and Schedule of Interactive Training Activities and Materials

Table 2 summarizes the activities mentioned above, and gives a plan for the third year of the project.

Type	Planned Delivery Date
<b>YEAR 2</b>	
Conference panel: “Open Europe: Open Data for Open Society”, by KU Leuven.	M15 (February 2014)
Conference presentation at the 13 <sup>th</sup> Annual Conference of the Behavioral Medicine Institute of LUHS, by LUHS.	M17 (April 2014)
Invited lecture at the Institute of Experimental Endocrinology, Charite University in Berlin, by NKUA.	M19 (June 2014)
Workshop for scholars on how to use the pilot services of the OSL platform, by LUHS.	M22 (November 2014)
Conference presentation: “Behavioral Neurosurgery: From Patient to Data”, by LUHS.	M23 (December 2014)
<b>YEAR 3</b>	
Training material and slides for the OSL platform included in academic lectures: “Internet Programming”, by NTUA.	2014-2015 academic fall semester of NTUA
Seminar: “Semantic Search in the Life Sciences in the Era of Open Access”, by TUD.	M25 (February 2015) see Appendix II.
Conference Presentation: “Molecular Diagnostics and New Therapeutic Approaches for Pancreatic Cancer”, Ulm University, by Uniklinikum TUD.	M25 (February 2015)
Working groups seminar of the Hellenic Cardiological Society, by NKUA.	M25 (February 2015)
Workshop session: “Share-PSI 2.0 workshop session on Open Data to identify best practices”, by KU Leuven.	M26 (March 2015)
Planned workshop for scholars on how to use the OSL platform, by LUHS.	M27 (April 2015)
Conference presentation at the 14 <sup>th</sup> Annual Conference of the Behavioral Medicine Institute of LUHS, by LUHS.	M27 (April 2015)
Seminar on how to use the OpenScienceLink platform, at the National Kapodistrian University of	M29 (June 2015)

<b>Athens, by NKUA.</b>	
<b>Workshop with focus on: “Translating thyroid hormone action in non-thyroid care setting: current options and future perspectives”, supported from the European thyroid Association, by NKUA.</b>	<b>M29 (June 2015)</b>
<b>Academic lectures at universities and research institutes in Bulgaria, by PROCON.</b>	<b>M26-M29 (March-June 2015)</b>

Table 2: Summary of interactive training schedule.

## 4 Sustainability

### 4.1 Sustainable Platform Aspects

Sustainability is a key aspect to be considered for the success of the OpenScienceLink project. OpenScienceLink offers SoA services for the five pilots and can compete at high level with respective existing tools providing novel semantic-enabled technologies for papers and dataset repositories. In the following, the aspects of sustainability for the major project outcomes and novel services are summarized. An analytical description on these aspects can be found at (Andronikou, et al., 2014). These are the main services that need to be sustained and frequently updated in the meta-project era.

#### 4.1.1 Sustaining Services

**Biomedical Data Journal:** OpenScienceLink develops and administrates the Biomedical Data Journal (BMDJ). This data journal uses the OpenScienceLink platform to present, validate, use, and reuse datasets in the biomedical domain. Availability of datasets is critical for analysing models based on experimental (cell and animal datasets) and clinical observations (human datasets).

**Data journal management services:** A collection of tools and services to manage a data journal, and to publish articles and datasets publicly available. *Relevant/involved Stakeholders: Publishers, Editors, Researchers*

**Scientific work peer review services:** These services cover the review process of scientific work, and assist editors in selecting the most appropriate reviewers for each paper. *Relevant/involved Stakeholders: Researchers, Publishers, Editors*

**Research Trends Detection and Analysis services:** The analysis of potential research trends in the domain can help finding hot/novel scientific topics, research areas with the highest recent accessibility and published work, and research trends associated with specific research.

**Dynamic researchers' collaboration building services:** These services give the user a global view of potential research collaborations with other researchers, research groups and organisations based on the web "trace" of the scientists' work, including published papers, articles, data, research interests and involvement in research groups and communities. *Relevant/involved Stakeholders: Researchers, Publishers, Editors, Research Groups and Laboratories, Universities, Research Departments in Biomedicine and Pharmaceutical Companies*

**Research evaluation (also called Scientometrics) services:** These services use the openly published scientific work to get not only conventional metrics (such as the volume of publications) but also novel ones, like the OpenScore. The OpenScore takes into consideration several aspects of a scientific work, including co-authorships, number of research works, scientific linking and status via the collaborations graph (PageRank score), and number of years of scientific activity. *Relevant/involved Stakeholders: Researchers, Publishers, Editors, Research Groups and Laboratories, Universities, Research Departments in Biomedicine and Pharmaceutical Companies, Research Sponsors, Research Funding Authorities, Media.*

**Semantic search services:** The semantic search uses a wide range of openly accessible scientific repositories in the biomedical domain. The result of an example search ('Endosome') is shown in **Figure 10**. *Relevant/involved Stakeholders: Researchers, Publishers, Editors, Research Groups and Laboratories, Universities, Research Departments in Biomedicine and Pharmaceutical Companies, Research Sponsors, Research Funding Authorities, Media*

**OpenScienceLink platform:** The OpenScienceLink platform encapsulates all of the aforementioned services and components.

The screenshot displays the OpenScienceLink (OSL) platform interface. At the top, there is a navigation bar with links for 'Logout', 'John Doe (My Profile)', 'Author Activities', 'Reviewer Activities', 'Editor Activities', 'Trends', 'Collaborations', 'Evaluation', and 'Search'. Below this, a search bar contains the query 'Endosome'. The main content area shows '10,000 documents found' and lists several articles related to endosome maturation, the endosomal network, and endosomal acidification. A sidebar on the left provides 'refine search' options, including 'Favorite concepts' and 'Highly related concepts'. The bottom of the page features a Creative Commons license logo and a footer with project information.

**Figure 10: A novelty of the OSL platform: Semantic search in data repositories.**

### 4.1.2 Sustaining Feature Evolution

In this section we discuss how the different pilots can be split in order to make them attractive for stakeholders, and make the provided services commercially viable.

#### Pilot 1

The inclusion of functionality to create and manage multiple journals could be a valuable feature. Cross-discipline journals can be easily managed under the same infrastructure.

#### Pilot 2

A key feature is to allow editors from other different journals use the pilot. The editors can suggest reviewers and follow the review process of the platform. This can be easily supported and expanded under the same underlying infrastructure.

#### Pilot 3

Additional data can be considered and processed. This extra information can be used to detect trends. The inclusion of additional repositories to extract information from trends needs to be considered, e.g., the inclusion of the domains of funded projects from the EU or National Ministries of research from the partners' countries.

#### Pilot 4

Potential collaborators can be extended by using popular social networks and profiles of OpenScienceLink users. Other bibliographical resources, like DBLP and citeseer, can also contribute to expand the network.



## **Pilot 5**

The OpenScore can be enriched with the inclusion of financial and budget data of universities and research centers. The OpenScore will also be expanded within the third year of the project, in order to be measured for the rest entities, i.e., cities, countries, journals, and academic departments/research centers.

### **4.1.3 Server Hosting and Platform / Service Maintenance**

The OpenScienceLink platform is hosted on the server of TUD. Development and administration are carried out jointly by TUD and TI. The implementation of the platform is flexible enough to easily add additional information sources.

## **4.2 Sustainability Methodology**

Our sustainability methodology is underpinned by the following three strategies.

### **4.2.1 3rd Party Funding**

Journals with large impact factor are owned by Nature Group, Elsevier, Springer and other publishing groups. We aim at presenting the added value that the OpenScienceLink platform can offer to these major publishers, in order to attract their interest in using the OSL platform for managing some of their journals. As result of our meetings, we expect useful feedback from their side. Meetings with additional stakeholders from the industry of academic networking platforms, e.g., ResearchGate and LinkedIn, as well as with stakeholders from the industry of managing huge scientific repositories, e.g., US National Library of Medicine, will be pursued in the third project year.

### **4.2.2 Customizing Services**

The modular approach of the OpenScienceLink platform allows configuring the system to present it as different products. In particular, this can be directed to relevant stakeholders, and charged as "pay-per-click" or "pay-per-service-call". Thus, each platform service can be considered as a single component. Strategies on utilizing the service-based view for raising funding will be considered within the third year of the project.

### **4.2.3 Link with Open Research and Open Science Projects and Initiatives, and Existing Information providers**

An important aspect for the sustainability is linking OpenScienceLink to other related projects, such as OpenAire. The OpenScienceLink consortium is already communicating with several of these big initiatives, in order to share experiences and join forces in common activities of shared interest, e.g., open access policies.



---

## 5 Summary and Conclusions

---

In this deliverable we described the efforts towards mobilizing resources in the project's pilot sites during the second year of the project, and provided a plan for the training and sustainability activities for the third year of the project. We mention the four pilot sites of the project (Dresden, Kaunas, Athens, and Pisa) and how their participants have contributed and will contribute to the project with respect to the pilot operations and training, until the end of the project, and which efforts and actions will be taken to ensure the sustainability of the platform and the journal in the meta-project era.

---

## 6 References

---

- Andronikou, V., Tagarev, T., Tagareva, P., Pantos, C., Mourouzis, I., Iervasi, G., et al. (2014). *D9.4.2: Sustainability Plans*. OpenScienceLink Consortium.
- Bunevicius, A., Juskiene, A., Birbilaite, I., Tsatsaronis, G., Schroeder, M., Zschunke, M., et al. (2015). *D7.1.2: Report on the Operations of the OpenScienceLink Pilot Services and Services/Processes Improvements*. OpenScienceLink Consortium.
- OpenScienceLink Consortium. (2013). *OpenScienceLink: OpenSemantically-enabled, Social-aware Access to Scientific Data*. EC.
- Tsatsaronis, G., Schroeder, M., Zhang, Y., Pliarsky, C., Pantos, C., Mourouzis, I., et al. (2015). *D9.2.2: Dissemination Report*. OpenScienceLink Consortium.
- Zschunke, M., Mendez, J., Karanastasis, E., Chondroyannis, E., Andronikou, V., Tsatsaronis, G., et al. (2014). *D6.1.1, REVISION 1, Report on Pilot Sites Operations and Training*. OpenScienceLink Consortium.

## 7 APPENDIX I: Dissemination of the OSL platform and the BMDJ at the existing user pools

In the following, the text that acted as a basis for the dissemination of the OpenScienceLink platform and the Biomedical Data Journal is shown. In all cases, the text was customized to address the specific user communities that were invited.

Dear colleagues,

It is with great pleasure that I would like to announce you the launching of the OpenScienceLink platform (<http://opensciencelink.org/>) and of the Biomedical Data Journal (<http://www.biomed-data.eu/>), which are both the results of the European co-funded project: OpenScienceLink (<http://opensciencelink.eu/>), coordinated by the Bioinformatics Group of BIOTEC, in TU Dresden.

Via the OpenScienceLink platform, you can:

- search efficiently in the biomedical literature and data sets,
- mine and visualize trends in the biomedical domain,
- find collaborators for your work,
- browse an evaluation of science based on metrics that use openly accessible data. For example, you can see your "OpenScore" as an author, which is a novel metric developed within the project and that correlates well with the h-index, but is trained on publicly accessible information about you, your publications and your collaborations.

The platform also supports the publication of research data sets and models generated from any discipline across the life sciences. The submitted data sets undergo a transparent peer-reviewing process, and are published in the Biomedical Data Journal (BMDJ).

With an editorial board of 20 renowned scientists in clinical research, cell biology, functional anatomy, pharmacology, cancer research, neurosurgery and bioinformatics, the BMDJ leads the way for high-impact, transparent, open access publication of research models, results and data sets. The journal will be indexed by all major engines, such as PubMed, Thomson Reuters, and DBLP.

Hence, I would like to invite you to try out the platform, and submit your data sets for publication at the BMDJ. The journal invites data sets and models submissions independently of whether they are new, or have been used before in your published research, as the aim of the project is to provide open access to scientific information, data sets and results for every researcher, academic and citizen interested to follow and contribute to the advancement of research.

If you have any questions, or feedback, about the platform or the journal, please do not hesitate to contact us.

The OpenScienceLink Project Consortium

## 8 APPENDIX II: TUD Seminar on “Semantic Search in the Life Sciences in the Era of Open Access”

In the following, the invitation to the seminar delivered by Dr. George Tsatsaronis follows. The seminar took place at the premises of CRTD in Dresden. Dr. Tsatsaronis demonstrated the pilots of the OpenScienceLink platform and the GoPubMed platform at the researchers of CRTD, BIOTEC, BCube and Uniklinikum.

Dear all,

I would like to invite you to the extraordinary seminar which will take place this Friday at 4pm:

Speaker:	George Tsatsaronis (Bioinformatics Group BIOTEC; Transinsight GmbH)
Title:	“Semantic Search in the Life Sciences in the Era of Open Access“
Date:	Friday, 06 February 2015
Time:	4:00 pm
Place:	CRTD, auditorium left

Abstract:

Researchers in the life sciences are struggling to keep up-to-date with the avalanche of the publicly available information. Scientific literature and biomedical resources such as domain ontologies and databases are growing extremely fast in numbers and volume. The indexed scientific literature in PubMed contains more than 24,000,000 articles. In parallel, the latest release of the Unified Medical Language System (UMLS), a major knowledge source in the life sciences, integrates information from 173 distinct biomedical resources and contains more than 3,000,000 concepts and 11,000,000 associated concept names. Given the plethora of the publicly available and highly heterogeneous resources (literature, databases, ontologies, microarray experiments, assays, biological networks) the modern biomedical knowledge worker needs efficient tools and engines to harness this information. Semantic search and related semantic-enabled technologies are the key to draw semantic bridges across the resources and integrate the information, providing powerful tools to the researchers. The presentation will focus on two SoA semantic search engines in the life sciences, GoPubMed and OpenScienceLink, explaining how the underlying technologies can cope with the heterogeneity and volume of the data, and provide novel tools and services to scientists working in the life sciences.

Enjoy the talk! Hope to see all of you at 4 pm this Friday!