



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

Project **CIP-Pilot 325101 / OpenScienceLink**  
Deliverable **D6.1.1**  
Distribution **Public**



OPENSOURCELINK  
ALERTTELEFERTIU

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

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

Authors: Matthias Zschunke, Julian Mendez, Stathis Katanastasis, Eythimios Chondroyannis, Vasiliki Andronikou, George Tsatsaronis, Daniel Eisinger, Alina Petrova, Giorgio Iervasi, Adomas Bunevicius, Iordanis Mourouzis

Status: Revision 1



D6.1.1 Report on Pilot Sites Preparation Activities, Detailed Planning of the Pilot  
Operations and Training

June 2014

### Project

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

### Deliverable

Deliverable type Report  
 Distribution level Public  
 Deliverable Number D6.1.1  
 Deliverable title Report on Pilot Sites Preparation Activities, Detailed Planning of the Pilot Operations and Training  
 Contractual date of delivery M12 (January 2014)  
 Actual date of delivery Revision 1: June 2014  
 Relevant Task(s) WP6/Task 6.1, 6.2, 6.3, 6.4  
 Partner Responsible TI  
 Other contributors NTUA, TUD, CNR, LUHS, NKUA  
 Number of pages 27  
 Author(s) Matthias Zschunke, Julian Mendez, Stathis Katanastasis, Eythimios Chondroyannis, Vasiliki Andronikou, George Tsatsaronis, Daniel Eisinger, Alina Petrova, Giorgio Iervasi, Adomas Bunevicius, Jordanis Mourouzis  
 Status & version Revision 1  
 Keywords Pilot Sites, Preparation Activities, Training

---

## Executive Summary

---

The purpose of the current deliverable is to describe the mobilization of resources (including human resources) in the project's pilot sites during the first year of the project. These are necessary for the final success of the project's pilot operations at the various sites of the consortium. The activities focus on the promotion of the project within the pilot sites, along with the specification of the resources (i.e., both infrastructure and personnel) needed for the successful conclusion of the pilot operations. In terms of infrastructures the respective work package 6 tasks ensure the required ICT applications and networking infrastructures required at the pilot sites. In terms of personnel, the preparation activities will ensure the involvement of the target numbers of researchers within the pilot sites, as well as the involvement of users from the external users/stakeholders group of the project. Finally, within this deliverable the strategy and the planning for the users training is described analytically.

# Contents

Executive Summary.....	3
Contents .....	4
List of Figures.....	6
List of Tables .....	7
1 Introduction.....	8
1.1 The OpenScienceLink project and platform landscape.....	8
1.2 Summary of the OpenScienceLink pilots.....	9
1.2.1 Pilot 1: Research dynamics-aware open access data journals development .....	9
1.2.2 Pilot 2: Novel open, semantically-assisted peer review process .....	10
1.2.3 Pilot 3: Data mining for biomedical and clinical research trends detection and analysis .....	10
1.2.4 Pilot 4: Data mining for proactive formulation of scientific collaborations.....	10
1.2.5 Pilot 5: Evaluation of research .....	11
1.3 Aspects of Launching, Disseminating, and Sustaining the OpenScienceLink Platform.....	11
1.3.1 Initial Launching and Dissemination .....	11
1.3.2 Users Training .....	12
1.3.3 Sustainability after launching .....	12
2 Initial Launching and Dissemination.....	13
2.1 Pilot Sites.....	13
2.1.1 Pilot Site Dresden.....	13
2.1.2 Pilot Site Kaunas .....	14
2.1.3 Pilot Site Athens.....	14
2.1.4 Pilot Site Pisa.....	15
2.2 Encourage existing user bases.....	16
2.2.1 GoPubMed users .....	16
2.2.2 Biotec, CRTD, BCube Dresden .....	16
2.2.3 PubMed authors.....	16
2.2.4 Customers and Partners of Consortium members .....	17
2.2.5 Mailing Lists.....	18
2.2.6 Other Channels .....	18
3 Users Training.....	19
3.1 Non-interactive Training .....	19
3.1.1 User Guide / Manual .....	19
3.1.2 Help Pages / In-place Help / FAQ.....	19
3.1.3 Tutorials (Written and Video).....	19
3.1.4 "Did you know, how to ..." Popup Screens.....	19
3.1.5 Summary and Schedule of Non-interactive Training Activities and Materials.....	20
3.2 Interactive Training .....	20
3.2.1 Workshop on Using Biomedical Search Technologies .....	20
3.2.2 Academic Lecture Tutorials .....	20
3.2.3 Summary and Schedule of Interactive Training Activities and Materials.....	21

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 Maintainance.....	24
4.2	Sustainability Methodology .....	24
4.2.1	3rd Party Funding.....	25
4.2.2	Customizing Services .....	25
4.2.3	Link with Open Research and Open Science Projects and Initiatives, and Existing Information providers .....	25
5	Summary and Conclusions.....	26
6	References .....	27

---

## List of Figures

---

Figure 1: The OpenScienceLink Eco-system. ....9  
Figure 2: Usage of GoPubMed to increase potential stakeholder's awareness of the OSL platform.... 17

---

## List of Tables

---

Table 1: Pools of users in Dresden .....	14
Table 2: Pools of users in Kaunas. ....	14
Table 3: Pools of users in Athens. ....	15
Table 4: Pools of users in Pisa.....	15
Table 5: Summary of non-interactive training schedule.....	20
Table 6: Summary of interactive training schedule. ....	21
Table 7: Summary of launching, training and dissemination processes. ....	26

# 1 Introduction

This section gives a short introduction to the OpenScienceLink project landscape and its main objective and pilots, which in turn define the nature of the pilot preparation activities that need to be conducted.

## 1.1 The OpenScienceLink project and platform landscape

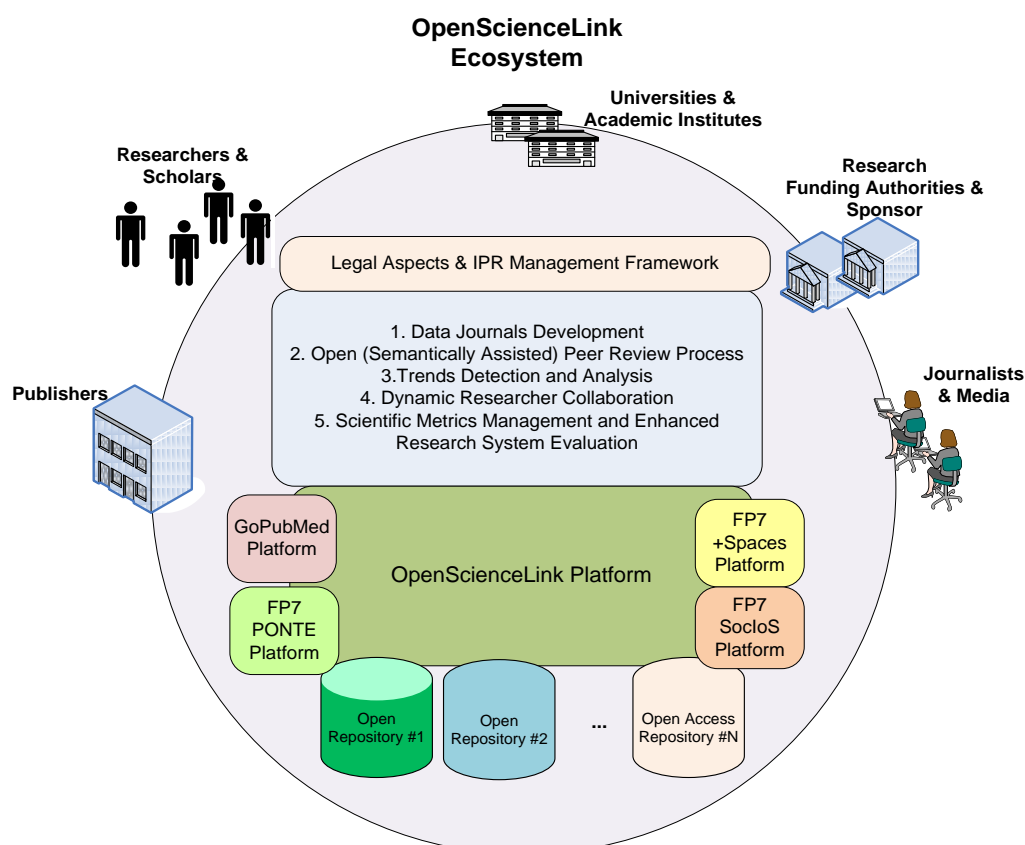
OpenScienceLink is a project co-funded by the European Commission under the CIP-ICT-PSP 2012 work programme topic 2.2 “Open Data and Open Access to Scientific Information”. The project aims at introducing and piloting a holistic approach to the publication, sharing, linking, review and evaluation of research results, based on the open access to scientific information. OpenScienceLink will pilot a range of novel services that could 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 evaluation metrics and indicators. **Five pilot services** will be integrated and piloted, in particular:

- (a) **data journals development** based on semantically-enabled research dynamics detection,
- (b) a novel open, **semantically-assisted peer review process**,
- (c) services for **detection and analysis of research trends**,
- (d) **services for dynamic researchers’ collaboration** based on non-declared, semantically-inferred relationships, and,
- (e) a set of **scientific field-aware, productivity- and impact-oriented enhanced research evaluation services**.

These services are developed and will be offered by the **OpenScienceLink platform**, which will be based on the semantic and social networking capabilities of background FP7 projects, as well as of the popular **GoPubMed** search engine (Schroeder & Doms, 2005). The OpenScienceLink services will be piloted with the aim to reach an active participation of over 1,200 researchers from the consortium organizations by the end of the project. OpenScienceLink has already established a group of external users / stakeholders that will contribute additional users/researchers in the scope of the validation process, while also engaging in the sustainable use of the services. OpenScienceLink also studies the business potential of open access paradigms, through investigating and pursuing multiple business models including author fees, hard copy sales, advertisements, sponsorship, as well as subscription based models.

As a summary, Figure 1 provides an overview of the OpenScienceLink eco-system, which is empowered by the OpenScienceLink ICT platform. A variety of elementary ICT services are implemented as part of the OpenScienceLink platform, in order to enable the deployment and integration of the five pilot services. More details with regards to the OpenScienceLink scope and objectives can be found within the Part B of Annex I of the project, i.e., the Description of Work of OpenScienceLink (OpenScienceLink Consortium, 2013), and more details with regards to the actual pilot and platform specification can be found in (Andronikou, et al., 2013) and (Tsatsaronis, et al., 2013) respectively.





## 1.2 Summary of the OpenScienceLink pilots

In the following, we present a summary of each of the five pilots.

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

Within Pilot 1 of the OpenScienceLink project, the Biomedical Data Journal (BMDJ) is being launched with the aim to allow the publication of datasets used in any aspect of biomedical, medical and clinical research. The datasets will be linked to the existing bibliography using annotation of their metadata with ontology concepts of the biomedical domain, and indexed accordingly to enable semantic search. The publication of large, complete and high-quality datasets is considered to be very important for all researchers and departments in order to prove their significant experience and expertise in a scientific field. Thus, publication in the BMDJ aims to become an important reference point for all researchers who aim to become established in a certain field and could be also used as an evaluation method for institutions and funding bodies.

In terms of technology, Pilot 1 requires advanced indexing and linking methodologies of the datasets and the scientific bibliography and access to a robust reviewing process that can be aided automatically by the platform. The role of the metadata submitted with the original dataset is, thus, crucial, and the users of the platform should be given the ability to provide such in an effortless manner, e.g., suggesting related headers and keywords, auto-filling and filtering abilities to select subject headings from existing and established domain ontologies.

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

Pilot 2 aims at offering a novel peer-review process methodology, which will help overcome a number of problems currently encountered by researchers, referees and editors. The peer-review system in its current form is considered as flawed by many, both because scientifically valid papers are frequently wrongly rejected, and scientifically flawed papers are accepted because errors, or even outright fraud, went undetected by the reviewers. Based on a number of new services that OpenScienceLink is introducing for this pilot, there exists the potential to substantially improve the speed and efficiency of the review process, which in turn will accelerate research itself.

Technology-wise the automatic identification of appropriate and available reviewers, which is a major problem for many editors, will be addressed with respective services from the OpenScienceLink platform. In addition, OpenScienceLink aims to aid the reviewers by providing the services that they can help them fill the review form efficiently and without consuming much time in the search process.

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

The aim of Pilot 3 is to discover and analyze research trends in open access biomedical and clinical research texts. The identification and analysis of such trends is essential for decision making, i.e., allocation of research funding (by private sponsors and governmental agencies), overall planning of research strategies, and evaluation of state of play of existing journals.

This pilot is primarily empowered by the advanced data mining capabilities that the OpenScienceLink platform offers, based on data from biomedical and clinical research fields. Research trend analysis could also be used to evaluate the degree of novelty of a publication or a research proposal. Furthermore, it is important to assess research trends in evaluating the impact of a research work in terms of citations. Thus, citation analysis should be weighed against research trends. Thus, the identification of emerging scientific fields related to each publication will be the basis for this analysis.

### 1.2.4 Pilot 4: Data mining for proactive formulation of scientific collaborations

The aim of Pilot 4 is to allow for researchers to investigate potential, non-declared relationships with other researchers, research groups and/or existing communities, as well as to build an open, dynamic scientific community for a given scientific research topic. This pilot focuses on automatically performing a series of intelligent matches among researchers based on their profile and research interests, as well as based on inferences through their published scientific work. The aim of these matches is to detect and propose collaborations among researchers who have not worked together and have no obvious, detectable connection, although they share the same or common research interests. Hence, the aforementioned matches are filtered by taking into consideration their past and existing collaborations as well as their deduced acquaintance in social networks. Similar matches are made between researchers and research groups as well as communities based on their scientific topics in order to propose their collaboration and/or participation in them. Publishers and editors may also use this pilot to obtain suggestions about researchers that are highly related to their journals' areas.

### 1.2.5 Pilot 5: Evaluation of research

The aim of Pilot 5 is to introduce, produce and track new objective metrics of research and scientific performance, beyond conventional metrics associated with conventional indices and impact factors. Funding agencies, institutions that employ scientists, and researchers, need to assess the quality and impact of scientific outputs. It is, thus, important that the scientific output is measured accurately and evaluated wisely. There is a constant need to improve the ways in which the output of scientific research is evaluated and all the parties involved are encouraging improved practices and methods in research assessment. Such steps are beginning to increase the momentum towards more sophisticated and meaningful approaches to research evaluation, especially the approaches that are based on publicly available data. Rating the quality and strength of scientific output is an important part of the evaluation process of researchers and institutions, taking decisions about funding allocation and research direction and even deciding about health practices and policies. The researchers may be evaluated based on a multitude of factors representative of their productivity, impact and domain, and important research work, in terms of potential. Based on the individual researchers' evaluation metrics, aggregated measures for research groups, universities, cities and countries may be designed, which could show individual assessments per research field. The complexity and availability of these measures will, naturally, be bounded to the amount of information that may be collected from publicly available sources for the respective entities.

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

The next sections explain the strategy the consortium follows to ensure a valuable dissemination and the sustainability of the work done for the OpenScienceLink platform. The major aspects of this strategy are described, including how the platform is launched, how the methodologies implemented for the OSL platform are disseminated across the corresponding potential stakeholders, how the users are trained to use the platform, and more precisely, how to use the platform for their corresponding information needs.

The main aspects of launching, disseminating and sustaining the OSL platform are:

- Initial Launching and dissemination,
- User Training, and
- Sustainability after launching

### 1.3.1 Initial Launching and Dissemination

The initial launching and dissemination of the platform is planned for the first 6 months of the second project year. Primarily, for the launching of the OSL platform a stable server environment must be ensured, that can offer accessibility and availability to the platform users. For this purpose, the OSL platform will be hosted in two sets of servers (primary and back-up) within TUD and TI respectively, under the reserved domain name: <http://opensciencelink.org>. The successful launching of the platform assumes that a respective implementation of the platform prototype is in place, following the specifications that are listed in the technical documents of work packages 3, 4 and 5.

Following the launching of the platform, the dissemination process commences. During this phase, the consortium will uptake the task of issuing press releases and using several types of media and fora to inform potential users, groups and other stakeholders through different channels: mailing lists, journal articles, conferences. The target is to convince the candidate audience about the good and high quality of the implemented services for the five pilots. This can be achieved via providing

analytical information about the features of the OSL platform, giving examples via case studies, and in parallel highlighting and motivating the need for sharing and publishing in an open access manner data sets, research work and results that can be made available via the platform.

### 1.3.2 Users Training

This second aspect of the launching, disseminating and sustaining the platform focuses on delivering a consistent set of materials and tools to candidate OSL users, so that they can train on the functionalities of the platform, and be able to get the maximum benefit from the offered services of the five pilots. In this regards, a set of actions has been planned, such as creation of online tutorials and videos, lectures, organisation of workshops, and demonstration of the platform in high impact venues, fora and conferences. The training process is split into two main categories: online and offline training. The former focuses in interactive training with the OSL platform users, e.g., tutorials in lectures and demonstration of the platform abilities, while the latter gives the ability to the users to train on their own pace via materials and manuals that can be used offline.

### 1.3.3 Sustainability after launching

This latter aspect of the launching and dissemination strategy focuses in two points: (1) the ensurance of the technical services availability in the after-project era, and, (2) the notion of keeping the underlying platform data and resources up-to date. Assuming a sufficient user pool by the end of the project, these two points require the creation of an efficient strategy that will be able to sustain the life of the platform after the end of the project. In this direction, the consortium will be working within the project's life time towards attracting the interest of stakeholders that are interested in any of the five pilots, with the aim to convince them to test them, and adopt them in their production processes.

---

## 2 Initial Launching and Dissemination

---

### 2.1 Pilot Sites

In the following, a description of the first year's preparation activities for the launching of the pilots is given. The focus of the preparation was put to the linking of the project with libraries and organisations that are potentially interested to adopt part of / all of the OpenScienceLink platform and services. Furthermore, given the volume of the potential users' pool per pilot site, a discussion is made with regards to the utilized mediums, and the activities that will take place to mobilize as many new platform users as possible. The analysis is given on the basis of the four pilot sites, namely Dresden (Germany), Athens (Greece), Kaunas (Lithuania), and Pisa (Italy).

#### 2.1.1 Pilot Site Dresden

The pilot site in Dresden participates with four units. Primarily, the Bioinformatics Group of the Technische Universität Dresden (TUD), which are also the coordinators of the project. In addition, via TUD, the B CUBE and the Uniklinikum are active members of the consortium. Finally, Transinsight GmbH (TI), though technical partner within the OpenScienceLink project, can connect the OpenScienceLink platform and services to the open access industry in Germany.

With regards to the pilot preparation activities of the first year, Table 1 describes the main user pools with which contacts are already made, or there has been ongoing work to present to them the project's objectives and technologies. TUD and TI held two separate meetings with the SLUB, which is the university library of TU Dresden. An agenda of the main meeting is presented in APPENDIX I of the current document. The purpose of the meetings was to introduce to the representatives of SLUB the concepts and objectives of the OpenScienceLink project, highlight the key technologies and services that implement the five project pilots and identify SLUB systems and projects where the services can be integrated, or IT systems which can be linked to the OpenScienceLink platform as a whole. With regards to the second users pool, the Bioinformatics Group of BIOTEC, Uniklinikum and B CUBE have long standing collaborations with the Medical Faculty and the Faculty of Computer Science of TUD. Via these collaborations, the three entities in Dresden have started discussions with members of the two faculties, in order to investigate the processes and systems of the respective faculties that can benefit from the services of the OpenScienceLink project. These discussions are undergoing; however, as TUD is the largest technical university of Germany, this is a much promising user pool to attract and engage new users to the OpenScienceLink platform. In total, the two user pools add up to ~70,000 potential new platform users. With respect to the mediums that will be adopted for the attraction of these users, a series of lectures is foreseen to the two faculties for the TUD pool, and a demonstration will respective links is planned to be set up in the SLUB library systems; these are basically the planned activities for the second year of the project at the Dresden pilot site.

Pool of Users	Description	Estimated Number of End users (potential maximum)	Status
SLUB Dresden Library	The University Library of TUD.	~40,000	Linking was conducted. Further info in APPENDIX I of the current document.
Technische Universität Dresden	The largest technical university in Germany in terms of enrolled students.	~37,000	Ongoing process to link with the Medical Faculty and the Computer Science Faculty.

**Table 1: Pools of users in Dresden.**

### 2.1.2 Pilot Site Kaunas

The pilot site in Kaunas participates via the Behavioural Medicine Institute of the Lithuanian University of Health Sciences (LUHS). Two main user pools have been contacted within the first year of the project, shown in Table 2, and which may offer a potential of ~40,000 end users. The first pool is the University Library of LUHS with ~20,000 registered users. The second pool is one of the largest universities for social sciences in Lithuania with ~20,000 enrolled students and faculties in biology and psychology, which are absolutely related to the domains of the OpenScienceLink project. The process of linking with these two entities is ongoing, and the focus on the second year of the project will be given towards connecting part or all of the OpenScienceLink pilots with the IT systems of these two entities, to attract users.

Pool of Users	Description	Estimated Number of End users (potential maximum)	Status
Lithuanian University of Health Sciences Library	The University Library of LUHS.	~20,000	Ongoing process to link with the LUHS Library.
Vytautas Magnus University Library	One of the largest universities for social sciences in Lithuania in terms of undergraduate and graduate students. Contains faculties of biology and psychology.	~20,000	Ongoing process to link with the VMU Library.

**Table 2: Pools of users in Kaunas.**

### 2.1.3 Pilot Site Athens

The pilot site in Athens participates with two units, namely NKUA and NTUA. NKUA is participating via the Medical School of the National and Kapodistrian University of Athens, and NTUA via the Distributed Knowledge and Media Systems Group of the National Technical University of Athens. During the first year of the project, NKUA started the process of linking with the pool of users (~10,000) stemming from the Medical School of the National and Kapodistrian University of Athens amounting. The medium were presentations about the OpenScienceLink project (<http://opensciencelink.eu/results/presentations/>) within lectures of the faculty, and a separate meeting that took place with graduate and postgraduate researchers of the faculty to inform them about the initiative. NTUA started the process of linking with the users of the Central Library of the National Technical University of Athens (~9,500) by means of identifying parts of the services of the library that part or all of the pilot services of the OpenScienceLink project can be connected. These figures are summarized in Table 3, and in both cases the linking process will continue within the second year of the project.

Pool of Users	Description	Estimated Number of End users (potential maximum)	Status
Medical School, National and Kapodistrian University of Athens	The largest medical school in Greece in terms of enrolled students.	~10,000	Ongoing process to link with the Medical Faculty.
Central Library of the National Technical University of Athens	One of the largest libraries in Greece, of the largest technical universities of the country.	~9,500	Ongoing process to link with the library's systems.

**Table 3: Pools of users in Athens.**

### 2.1.4 Pilot Site Pisa

The pilot site in Pisa participates with one unit, namely the Institute of Clinical Physiology of the National Research Council (CNR). Within the first year of the project, CNR established two connections; primarily the connection with the National Council Research (CNR) Central Library (~10,000 end users), and with the Research Hospital Tuscany Region G Monasterio Foundation. In the former case, CNR started the communication with the management of the library, with the aim to present the OpenScienceLink project objectives and scope, and identify library services with which respective pilot services of the OpenScienceLink can be connected. In the latter case CNR members met the director of the CNR /Tuscany Region G Monasterio Foundation (Research Hospital) and obtained letter of interest and support. This can be found in APPENDIX II of the current document. In both cases, the linking process with these two pools of users will continue within the second year of the project.

Pool of Users	Description	Estimated Number of End users (potential maximum)	Status
National Council Research (CNR) Central Library	Founded in 1927, it is the oldest and principal multidisciplinary technical and scientific Italian library	~ 10,000	Linking with the management of the library started.
Tuscany Region G Monasterio Foundation (Research Hospital)	Research hospital specialized in the treatment of cardiopulmonary diseases.	~ 3,000	Linking with the management of the hospital started.

**Table 4: Pools of users in Pisa.**

## 2.2 Encourage existing user bases

Within the consortium of OpenScienceLink there are technologies from partners that have been deployed in the past, e.g., GoPubMed, and which already contain active users which are registered and are using them frequently. On this basis, in the following we present an overview of these user pools, which will also be mobilized towards attracting new users in the OSL platform. In addition, we also present the means via which additional external users will be invited to use and adopt the platform in their work.

### 2.2.1 GoPubMed users

GoPubMed is the first semantic search engine in the life-sciences. It enables semantic search functionality on the PubMed content using the MeSH (medical subject headings) as an ontology. GoPubMed is available online for the biomedical research community since 2003. Users are given the opportunity to create a profile in order to benefit from additional features. Meanwhile, GoPubMed has reached a substantial user base in the biomedical research community. The whole GoPubMed user base consists of potential stakeholders interested in OpenScienceLink functionalities. The active GoPubMed community consists of currently (December 2013) approx. 3,500 registered and active users with academic and commercial background, as well as approx. 1,000 anonymous users per day.

### 2.2.2 Biotec, CRTD, BCube Dresden

In the Dresden pilot site, there are three research entities which are active in several different aspects of research in life sciences. Biotec, which is also the host of the Bioinformatics group that coordinates 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. This pool of approximately 450 active researchers in life sciences which are located in this triple complex in Dresden is easily accessible, and most of them are researchers that are already aware of the OpenScienceLink project and its aims, and are willing to test the pilots of the platform. This user pool will constitute the main feedback vehicle to the deployment of the first prototype of the OpenScienceLink platform.

### 2.2.3 PubMed authors

PubMed provides an index of biomedical publications. The index consists of author affiliation information as well. This information is valuable with regards to potential stakeholders from the pool of scientific authors in very specific life sciences field.

The research community published 3.4 million articles in the fields covered by the BMDJ Editorial Board member expertise in the past 5 years (2010 - 2014)<sup>1</sup>. This pool consists of a correspondingly high number of active authors.

The OSL consortium does not intend to invite the whole set of these authors with news and irrelevant information, but rather aims at inviting specific authors from this pool to participate in pilot 2 as candidate reviewers, and encourage them to also use additional features of the OSL platform via the other 4 pilots.

<sup>1</sup> Determined by the GoPubMed Query: ("Cell Biology"[mesh] OR Anatomy[mesh] OR "Cardiovascular Diseases"[mesh] OR "Biomedical Research"[mesh] OR "Nervous System"[mesh] OR Pathology[mesh] OR Pharmacology[mesh] OR "Molecular Biology"[mesh] OR Biotechnology[mesh] OR Biophysics[mesh] OR Neurosurgery[mesh] OR "Chemicals and Drugs"[mesh] OR "Clinical Trials as Topic"[mesh]) last5years[time].



## 2.2.4 Customers and Partners of Consortium members

TI is a software vendor specialized in semantic web technologies especially for the life science industry. In its 9 years history it attracted a substantial customer record of big players in the biomedical industry. Amongst others Unilever, Bayer, BASF Janssen by Johnson&Johnson are long lasting customers of Transinsight. It can be assumed that through the track record of the company a first trust in the project is granted. Through the companies network of these important big customers and through our partners we'll at first disseminate the project in general. The hope is to tease relevant people (ideally key players in the industry) and make them ambassadors for the idea of OpenScienceLink.

A second ideal possibility to disseminate OSL is the landing page of GoPubMed.org, the semantic search engine for the life sciences. The search engine is now online for almost 11 years and visited by thousands of users daily. Placing advertisement in a prominent position (e.g. below the search line) will surely attract a huge number of users. The following draft shows how it is prepared to be launched as soon as the platform is ready for public release.

The screenshot displays the GoPubMed website. On the left is a 'help index' menu with categories like 'In a Nutshell', 'Semantic search', 'Concept Categories', 'Examples of semantic search', 'Statistics', and 'Usability features'. The main content area features the GoPubMed logo with the tagline 'searching is now sorted!', a search input field, a 'find' button, and a 'Login' link. Below the search area is an illustration of three people holding a globe, with the text 'Publish in OpenScienceLink ...' and a 'Donate' button. On the right side, there are three vertical advertisements: 'AdipoGen' (Connecting Immunology to Metabolism), 'BioA2Q' (A challenge on large-scale biomedical semantic indexing and question answering), and 'publish in Open Science Link'.

**Figure 2: Usage of GoPubMed to increase potential stakeholder's awareness of the OSL platform.**

In addition, TI is often presenting technologies at large conferences and exhibitions. The most prominent example is CeBIT, the world wide biggest trade show in computation. Also the annual conference of the EAGE (European Association of Geoscientists and Engineers) can be named for the field of geo-research. With around 5000 participants and visitors this conference and exhibition is one of the most important fairs world wide and an ideal platform to disseminate the idea of OSL. Also smaller events will allow to spread the platform in the scientific community via TI's business.

### 2.2.5 Mailing Lists

Another important medium to disseminate the OSL platform and attract new users is via the communication of press releases to related communities reachable with publicly available mailing lists. Example of such lists are: DBWORLD, BioNLP, Researchers at the PASCAL Network, elsnets, corpora at the UIB, KNet, and many more that have audience with interests in applied bioinformatics, or other life sciences. The consortium has compiled a list of 19 such communities, which will be used to disseminate the news of the launching of the platform. In addition, LinkedIn groups, such as Text Mining, Semantic Web Research, Bioinformatics, SYNC3 and other groups with members that their interests lie in applications of life sciences, will be used to communicate the press releases of the platform as well.

### 2.2.6 Other Channels

Besides the aforementioned mobilization, several other channels will be used to communicate the launching of the platform, and advertise the project and its results. More precisely, press releases to the national research points (e.g., ministries and secretaries) will be issued, conference presentations and demonstrations in high impact venues will be planned, where also material that will be prepared, such as brochures about the project, will be distributed. Specialized workshops where audience with a niche in life sciences and applications is present will be also considered.

---

## 3 Users Training

---

In the following we analyze the training activities, materials and schedule that is planned in order to get users accustomed to the OpenScienceLink pilots and platform services. The presentation is split into two parts, which separate the training schedule in materials and activities according to the criterion of interaction with the audience, and more precisely, into: (a) non-interactive training, and, (b) interactive training.

### 3.1 Non-interactive Training

In the following, we present the activities, materials and schedule planned for the non-interactive training of users in the OpenScienceLink pilots and platform services. The final table in Section 3.1.5 summarizes the planned schedule.

#### 3.1.1 User Guide / Manual

A common strategy to explain the platform functionality to the potential users is a user guide. The user guide will present the use cases that were described reported in the deliverables of WP2-WP5 and WP7 in the platform context and show the users how to achieve the corresponding use cases goals by using the OpenScienceLinkk platform prototype. The user guide is planned to be published through the platform website in the form of HTML pages in M21 (October 2014), and will give a detailed guide through all of the OSL platform pilots.

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

A strategy where help is given when needed is appreciated by users not completely familiar with a computer program, software or web site. For each pilot there will be a side bar with examples and methods explained within the deployed platform. The help will also be linked to the corresponding page in the users manual. Many questions arise multiple times by different users, therefore we will collect the users questions and answers on an FAQ page hosted on the projects website. The Help Pages are planned to be in place by M22 (November 2014).

#### 3.1.3 Tutorials (Written and Video)

Tutorials will be made to show users step-by-step use case scenarios as defined in the deliverables of WP2-WP5 and WP7. The tutorials will be part of the platforms user guide, and there will be at least one tutorial per pilot. In many cases, potential first-visit users are discouraged by plenty of documentation to be read, so we will also put the tutorials as videos. The videos will be hosted at youtube.com to make them accessible to every stakeholder or interested party. The tutorials are expected to be prepared by M23 (December 2014).

#### 3.1.4 "Did you know, how to ..." Popup Screens

A possibility to show users specifically new features and functionalities is to show a "Did you know"-popup info box at the opening of the platform website. In this box users are presented with information and links regarding new functionalities and tutorials, but also to interesting key functionalities that existed already. The info boxes are planned to be integrated into the platform in M23 (December 2014).

### 3.1.5 Summary and Schedule of Non-interactive Training Activities and Materials

Type	Planned Delivery Date
User Guide / Manual	M21 (October 2014)
Help Pages / In-place Help	M22 (November 2014)
FAQ	M22 (November 2014)
Tutorials / Video-Tutorials	M23 (December 2014)
"Did you know, how to ..." -Popup screens	M23 (December 2014)

**Table 5: Summary of non-interactive training schedule.**

## 3.2 Interactive Training

In the following, we present the activities, materials and schedule planned for the interactive training of users in the OpenScienceLink pilots and platform services. The final table in Section 3.2.3 summarizes the planned schedule.

### 3.2.1 Workshop on Using Biomedical Search Technologies

One very important training vehicle to introduce candidate users to the concepts of the OpenScienceLink pilots and platform services is via the organisation of a workshop on the usage of biomedical search technologies for research purposes. TUD, in collaboration with TI and NTUA, are planning to organise such an international workshop in Dresden, most probably as a satellite event in a major conference or symposium about life sciences. The aim of the workshop will be to invite speakers from Europe that work for major biomedical search engines and related technologies, and to present these technologies and their usage in research. A major pillar of this workshop will be the presentation of the GoPubMed and OpenScienceLink platforms, where the aim will be to motivate the audience to use them, via the demonstration of interesting case studies, and the capabilities of the platforms. From this regards, this workshop with invited speakers will have a tutorial flavour with the aim to explain to the biomedical researchers how they can get the best out of these tools. As the optimal organisation will have to coincide in time and place with a major conference or symposium, the exact time frame cannot be decided at this stage. However, candidate venues are: ISMB/ECCB 2015, and the Biosaxony Symposium in 2015, hence first half of 2015.

### 3.2.2 Academic Lecture Tutorials

Besides providing an easy to use platform per sé it is important to tease potential users by educational activities. Academic lecture-tutorials presented in short demonstration videos are also good ways of transporting the ideas behind OSL. The aim must be to: (a) attract the interest of scientists in the platform by communicating the clear benefits for their scientific work, and, (b) provide a very easy start with the platform.

It is also imaginable to cross-link OSL with the BioASQ project (<http://bioasq.org/>). The yearly meeting of BioASQ is held at the CLEF conference, also an ideal platform to bring the idea to researchers to scientists in the field of text mining and semantics. A small "right-to-the-point-lecture" should interest this community too. Transinsight provides its channels to relevant players to disseminate the (video) lectures.

In addition, all of the academic partners of the consortium (TUD, NTUA, CNR, NKUA, LUHS, and KU Leuven) will embed demonstrations of the OpenScienceLink platform in lectures of relevant courses, when applicable. These activities are planned for both the fall and spring semester of the academic year 2014-2015.

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

Type	Planned Delivery Date
Workshop on the usage of biomedical search technologies	First half of 2015
Academic lecture tutorials	Academic year 2014-2015, at the sites of the 6 academic partners of the consortium

**Table 6: Summary of interactive training schedule.**

---

## 4 Sustainability

---

In this section, the notion of sustainability of the platform in the era after the end of the project's lifetime is analyzed. The sustainable platform aspects are discussed, and an initial planning is made with regards to the efforts that are being scheduled to constitute the platform sustainable after the project finishes.

### 4.1 Sustainable Platform Aspects

Generating a sustainable user base is key for the OpenScienceLink project. At the same time the most difficult challenge. Pitching against e.g., ResearchGate with a funding of at least 40 million dollars will be the toughest task in the project. The only way of being able to sustain is to provide a best of breed system that is easier to use and provides better features than competitors. The biggest advantage we see here is our semantic search technology in paper and dataset repositories, which allows to intelligently link research papers, a feature that is not observed in other platforms. This again is what scientists are truly interested in: a machine (OSL + Semantics) that helps to identify collaborators and monitor research in their field on a content level rather than on a meta-data level.

#### 4.1.1 Sustaining Services

A very important aspect of the sustainability of the platform is the sustainability of the services that are part of the platform's prototype. More precisely, this challenge is how to maintain/keep the services running and how to continue expanding the services after the project ends. Though the steps for future improvements are analytically described in deliverables D5.2.1, D5.2.2 and D7.1.1-D7.1.3, the key is how to break down the advantages of the platform in an effort to maximize the chances for adoption of parts, or whole of the OSL platform services. From this regard, the following components have been identified as the major pillars that can be used to define the sustainability strategy for the platform services, analyzing them independently, and as a whole for future exploitation.

**Data Journal:** OpenScienceLink develops and manages a data journal (BMDJ) which encapsulates the OpenScienceLink services. This data journal facilitates the presentation, validation, use, and re-use of datasets, with focus on publishing biomedical datasets that can serve as a source for simulation and computational modelling of diseases and biological processes. Computational modelling can bridge the gap between experiments and patients by integrating data obtained from experimental cell and animal based models to patients. Datasets availability is critical for training, optimisation and validation of 'integrative' mathematical models based on experimental (cell and animal datasets) and clinical observations (human datasets).

**Data journal management services:** A series of tools and services allowing for the management of data journals which will publish openly accessible data about experiments, clinical observations and/or trials and making it available for aggregation, creation, testing and validation of models. These services will be enhanced with intelligent mechanisms enabling the identification of hot topics to be covered by upcoming data journal issues, as well as the linking of the published data sets with highly related openly accessible scientific papers. *Relevant/involved Stakeholders: Publishers, Editors, Researchers*

**Scientific work peer review services:** These services cover and enhance the review process of scientific work. Focusing mainly on, but not limited to, the reviewers tasks, they provide reviewers with a set of tools which facilitate the scientific work's review process, reduce the temporal requirements and increase the confidence for the review outcome by offering them instant access to highly relevant, intelligently processed and filtered scientific work. They also assist editors in selecting the most appropriate reviewers for each paper based on the semantic analysis of the latter. A crowd-sourcing approach is followed for the evaluation of the reviews which allows for

researchers to comment upon the reviewers work and the scientific work under review.  
*Relevant/involved Stakeholders: Researchers, Publishers, Editors*

**Research Trends Detection and Analysis services:** These services semantically process openly accessible scientific work in order to detect and analyse potential research trends in the domain, including hot/novel scientific topics, research areas with the highest recent accessibility and published work, research trends associated with specific research, etc. These trend mining services will allow users to either view a trend analysis for a specific topic of their interest or be presented with a ranked list of research-wise “hot” scientific

**Dynamic researchers’ collaboration building services:** These services will present researchers with automatically inferred suggestions 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 (aka Scientometrics) services:** These services deliver and visualise evaluations of the openly published scientific work through not only conventional metrics (such as Impact Factor) but also novel ones. The latter take into consideration several aspects of a scientific work, including author, number of research work views, downloads and citations (as well as their source), level of dissemination, etc. *Relevant/involved Stakeholders: Researchers, Publishers, Editors, Research Groups and Laboratories, Universities, Research Departments in Biomedicine and Pharma Companies, Research Sponsors, Research Funding Authorities, Media.*

**Semantic search services:** These mechanisms will offer unique capabilities associated with semantic search over a wide range of openly accessible scientific repositories in the biomedical domain. *Relevant/involved Stakeholders: Researchers, Publishers, Editors, Research Groups and Laboratories, Universities, Research Departments in Biomedicine and Pharma Companies, Research Sponsors, Research Funding Authorities, Media*

**OpenScienceLink platform:** The OpenScienceLink platform itself will comprise an exploitable product of the project. This platform will encapsulate all the aforementioned mechanisms, models and services and will be customisable in terms of covered domain and set-up ecosystems.

#### 4.1.2 Sustaining Feature Evolution

From the point of view of features evolution, in the following we briefly described how we can open up the access to the five pilots, and make them attractive to stakeholders that would be willing to adopt the respective services commercially in their existing or future products.

##### Pilot 1

The key to the feature evolution in pilot 1 is to allow more journals to be created and managed by the services of this pilot. Cross-discipline journals would also show signs of advantage compared to other data and paper management suites. Integration of data sets from other resources, and examining the OpenScienceLink Platform as a Meta search engine for datasets is also a key view that gives the consortium a strategic advantage in the absence of any related engine or technology.

##### Pilot 2

In Pilot 2, the key is to allow editors from other journals (other than BMDJ), and also to non-dataset journals to use the pilot, suggest reviewers and to setup and follow the review processes of the platform..

### Pilot 3

In Pilot 3, additional data may be considered in order to detect trends, such as financial and funding project's data, as well as clinical trial information.

### Pilot 4

Suggestion of potential collaborations may be expanded with additional social networks. This would allow the OSL users to link their profiles with existing and popular social networks. From another perspective, other bibliographical resources can also be included, such as DBLP and citeseer, from which the co-authorship networks could be expanded.

### Pilot 5

For the evaluation of research entities, the inclusion of financial and budget data for universities and research centers can become an additional feature to the OpenScore. In addition, a methodology to include evaluations/ranking data from countries' local secretaries and authorities, or public stakeholder, would give an additional set of features that can be exploited, certainly for the cities and countries evaluation.

Finally, as the usage of mobile information technology increases, there is a demand for Mobile App for pilots 3,4,5. In this direction, such Apps can be definitely part of the future expansions and additions of the OSL platform.

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

The OpenScienceLink platform is hosted on the servers of the TUD. Administrative and developmental tasks are organized by TUD and TI as well. The platform implementation is based on Transinsight's highly scalable, semantic search technology ESI (Enterprise Semantic Intelligence).

The implementation allows for direct horizontal scaling, to allow a near-linear scaling of the number of users as well as the number of semantically indexed material. The platform is prepared to be enriched with additional information sources, like external dataset repositories.

The platform's implementation is flexible in a way that additional information sources can easily be implemented and added to the overall semantic index of the OSL platform. Also additional ontologies can easily be integrated to further extend the platform to other domains, like social science, computer science, physics, geo science, etc.

This large spectrum of additional knowledge domains will increase the number of potential stakeholder willing to finance the OSL platform's operations, e.g., funders and publishers.

The methodology adopted to sustain the operation of the OSL platform are explained in the next sections.

## 4.2 Sustainability Methodology

Our sustainability methodology aims at analyzing the cost/benefit of the developed technologies, as well as conducting a comparative value analysis, an utility analysis and a risk assessment. These quantifiers and qualifiers are discussed and expanded in the related Sustainability Plans deliverables (D9.4.1-D9.4.3). In the following we are only focusing in arraying the three main vehicles for constituting the project and the platform sustainable in the after-project era. The successful utilization of those vehicles is of course heavily dependent on the aforementioned aspects, and on how these qualifiers can be embedded into presentation and demonstration portfolios that may convince the interested stakeholders to adopt the OpenScienceLink technologies.



### 4.2.1 3rd Party Funding

As the publishing industry is at a large extent the referee of the published scientific material, i.e., especially the large impact factor journals owned by Nature Group, Elsevier, Springer and other publishing groups, we focus on the publishers as a major stakeholder role which can prove a key to the sustainability of the project. We aim at meeting with these important stakeholders, and present them the added value that the OpenScienceLink platform services can incorporate into their processes. More precisely, the consortium is planning to meet with Springer, Nature Group and NLM within the second year of the project, in order to identify which of the five pilots are more important and interesting for them, and under which conditions they would be willing to adopt any of those in their processes. Their feedback will also give the consortium the necessary input to decide on where to turn our focus in the third and final year of the project. We hope that via these meetings, a mature ground will be occupied towards opening future collaborations between the OSL consortium and publishers, with the aim to raise third party funding in the after-project era.

### 4.2.2 Customizing Services

The architecture and implementation of the OpenScienceLink platform has the advantage of modularity at the services level, since it is based on a Service-oriented design. This advantage allows the consortium to be able to present the platform services as individual technologies that can be promoted towards interested stakeholders, and further customized for their needs. TI has a large experience with customers that are interested in service-based approaches. Service oriented products, customizations and "pay-per-click" or "pay-per-service-call" schemes are all part of the current TI's business models, and based on this experience, the consortium can create interesting, attractive and economic offers towards stakeholders that would be willing to adopt small components of the OSL platform functionality. As an example, the services for the trends in the drug design and the biomedical research can be customized for the pharma industry, while in parallel the services for the evaluation of researchers and universities can be customized for public bodies and agencies that are assigned the task of accessing the research at the national or international level. Hence, each OSL platform service will be considered as a single component that may be attractive for relevant stakeholders.

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

Another very important aspect for the sustainability of the OpenScienceLink project is the linking with existing related initiatives and projects, such as OpenAire. Towards the implementation of the Open Science idea, OpenScienceLink can benefit by sharing content and exchanging technologies with related projects. This will also raise the visibility of the project, and bring the consortium in touch with the major efforts towards this vision. Among others, the OpenScienceLink project may adopt similar sustainability strategies that are adopted by the related projects, digest this past experience and embed it in the future plans, by making corrections and adaptations that are customized to the specific OpenScienceLink technologies and related stakeholders. A series of such meetings with related initiatives and projects has already been conducted at the first half of the second year of the project and further meetings will be planned for the third year.

## 5 Summary and Conclusions

In this deliverable the efforts towards mobilizing resources in the project's pilot sites during the first year of the project was described. In the four pilot sites of the project (Dresden, Kaunas, Athens, and Pisa) the linking with large pools of potentially interested pilot users started. The user pools are mainly libraries and research or academic organizations, and in all of the cases offer services which may benefit from the pilot operations of the OpenScienceLink project. The total number of the end users covered by these pools amounts to ~150,000 end users, a large percentage of which is related to activities covered by the OpenScienceLink pilots. The potential is great, and in the second year of the project the consortium will focus into attracting as many of these users as possible to the pilot site operations of the OpenScienceLink project. In this document we presented in addition the planning for the training of the users, as well as the main vehicles for the sustainability of the project. A short summary of the presentation is given in the following table, that covers the remaining half of the project.

Activity	Period
Official release of the platform	June 2014
Inviting users from existing user bases and disseminating the platform and the project	June 2014 - January 2016
Training Material (Tutorials, User Manuals, Help Pages, Workshops, Academic Lectures, Conference Presentations)	October 2014 - June 2015
Meetings with stakeholders (slub, Springer Open, Elsevier, Nature Group, NLM)	September 2014 - February 2015

**Table 7: Summary of launching, training and dissemination processes.**

---

## 6 References

---

Andronikou, V., Tagarev, T., Pantos, C., Mourouzis, I., Iervasi, G., Hugelier, S., et al. (2014). *D9.4.1 Sustainability Plans*. OpenScienceLink Consortium.

OpenScienceLink Consortium. (2013). *OpenScienceLink: OpenSemantically-enabled, Social-aware Access to Scientific Data*. EC.