

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



OPENSOURCELINK

<http://opensciencelink.eu>

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

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

Status: Final (Version 1.1)

March 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	<a href="http://opensciencelink.eu">http://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	D6.1.1
Deliverable title	Report on Pilot Sites Preparation Activities, Detailed Planning of the Pilot Operations and Training
Contractual date of delivery	M12
Actual date of delivery	M14
Relevant Task(s)	WP6/Task 6.1
Partner Responsible	TI
Other contributors	NTUA, TUD, CNR, LUHS, NKUA
Number of pages	18
Author(s)	Matthias Zschunke, Julian Mendez, Stathis Katanastasis, Eythimios Chondroyannis, Vasiliki Andronikou, George Tsatsaronis, Daniel Eisinger, Alina Petrova, Giorgio Iervasi, Adomas Bunevicius, Iordanis Mourouzis
Status & version	Final (Version 1.1)
Keywords	Pilot Sites, Preparation Activities

---

## 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.

---

# Contents

---

Executive Summary .....	3
Contents .....	4
List of Figures .....	5
List of Tables .....	6
1 Introduction.....	7
1.1 The OpenScienceLink project and platform landscape.....	7
1.2 Summary of the OpenScienceLink Pilots .....	8
1.2.1 Pilot 1: Research dynamics-aware open access data journals development.....	8
1.2.2 Pilot 2: Novel open, semantically-assisted peer review process.....	9
1.2.3 Pilot 3: Data mining for biomedical and clinical research trends detection and analysis	9
1.2.4 Pilot 4: Data mining for proactive formulation of scientific collaborations .....	9
1.2.5 Pilot 5: Evaluation of research.....	9
2 Preparation Activities and Description of Existing Services and Resources .....	11
2.1 Pilot Site Dresden .....	11
2.2 Pilot Site Kaunas.....	12
2.3 Pilot Site Athens .....	12
2.4 Pilot Site Pisa .....	13
3 Conclusions .....	14
4 References .....	15
5 APPENDIX I - Agenda of the Dresden Pilot Site Meeting with the SLUB Library.....	16
6 APPENDIX II - Letter of Interest and Support from CNR /Tuscany Region G Monasterio Foundation (Research Hospital).....	18

---

## List of Figures

---

Figure 1: The OpenScienceLink Eco-system.....8

---

## List of Tables

---

Table 1: Pools of users in Dresden.....	11
Table 2: Pools of users in Kaunas.....	12
Table 3: Pools of users in Athens.....	13
Table 4: Pools of users in Pisa.....	13

---

# 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.

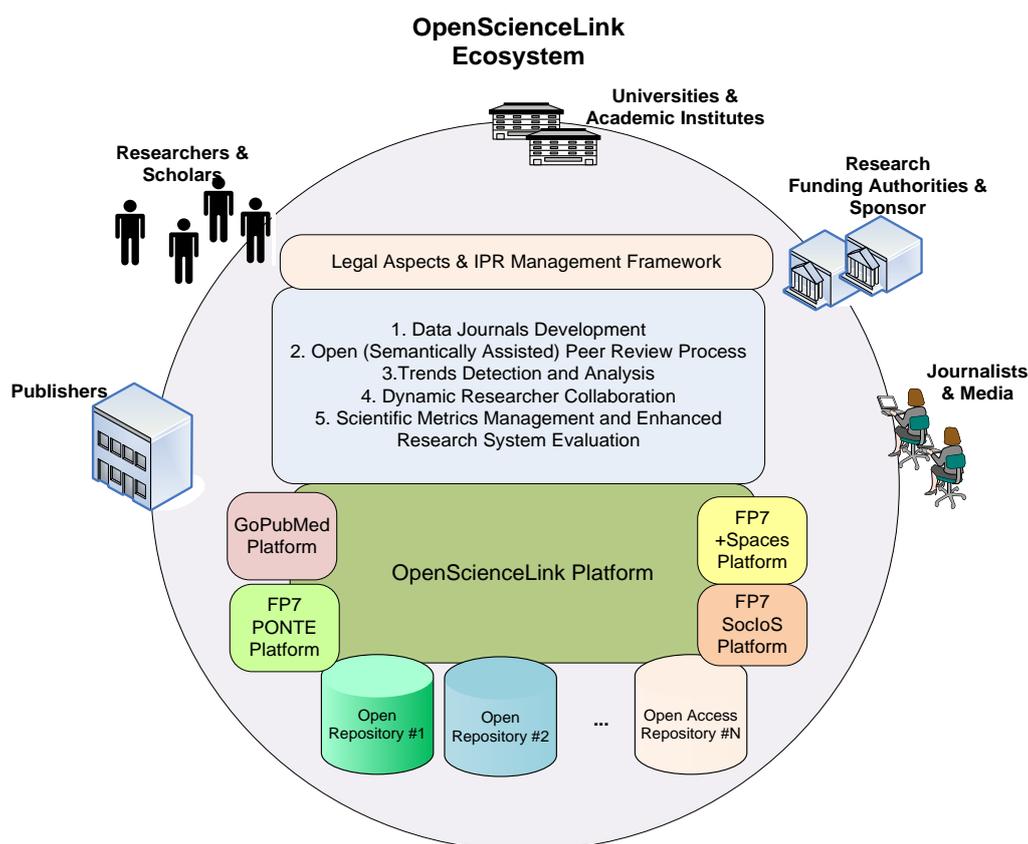
## 1.1 The OpenScienceLink project and platform landscape

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.



**Figure 1: The OpenScienceLink Eco-system.**

## 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.

## 2 Preparation Activities and Description of Existing Services and Resources

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 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 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.**

With regards to existing infrastructure, the first official release of the OpenScienceLink platform will be hosted at the Bionformatics Group's biocluster, which is a very strong computational cluster that can accommodate for scalability and high performance, comprising 330 CPUs and 2TB of main memory. Depending on the user volume in the following years, there exists the ability to transfer the OpenScienceLink platform in one of the largest computational clusters in Europe, which belongs to TUD, namely the ZIH cluster ([http://tu-dresden.de/die\\_tu\\_dresden/zentrale\\_einrichtungen/zih/hpc](http://tu-dresden.de/die_tu_dresden/zentrale_einrichtungen/zih/hpc)) which will soon reach 100,000 CPUs.

## 2.2 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.**

## 2.3 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.**

## 2.4 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.**

---

## 3 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.

---

## 4 References

---

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

Schroeder, M., & Doms, A. (2005). GoPubMed: exploring PubMed with the Gene Ontology. *Nucleic Acids Research*, W783--W786.

Andronikou, V., Karanastasis, E., Tsatsaronis, G., Zschunke, M., Tagarev, T., Tagareva, P., et al. (2013). *Detailed specification of OpenScienceLink Pilot Scenarios and Services*. OpenScienceLink Consortium.

Tsatsaronis, G., Eisinger, D., Mendez, J., Petrova, A., Karanastasis, E., Chondrogiannis, E., et al. (2013). *OpenScienceLink Platform Architecture and Background Platform Specifications*. OpenScienceLink Consortium.

## 5 APPENDIX I - Agenda of the Dresden Pilot Site Meeting with the SLUB Library

In the following, the agenda for the meeting that took place in Dresden, Germany, on January 24, 2014, with the aim to introduce the OpenScienceLink project to the SLUB library of Dresden and identify related services of the OSL platform that the library is willing to adopt to their systems, is reported.



**CIP-ICT PSP-2012-6**

**Pilot Type B, Open Data and open access to scientific information**



**OpenScienceLink: Open Semantically-enabled, Social-aware  
Access to Scientific Data**

**Grant Agreement No: 325101**

**1<sup>st</sup> Dresden Pilot Site Meeting - Dresden**

**Dresden, January 24, 2014**

<b>Document title:</b>	<b>OpenScienceLink 1<sup>st</sup> Dresden Pilot Site Meeting Agenda</b>
<b>Document version:</b>	<b>Final</b>
<b>Workpackage - Task</b>	<b>WP1- Management</b>
<b>Circulation:</b>	<b>Internal</b>
<b>Document Description</b>	<b>This document provides the agenda and necessary information for the 1<sup>st</sup> Dresden pilot site meeting of the OpenScienceLink project, held in Dresden, January 24, 2014.</b>

## Introduction

The 1<sup>st</sup> Dresden Pilot Site Meeting of the OpenScienceLink project takes place in Dresden, Germany, at the premises of BIOTEC. The main objective of the meeting is to exchange views with SLUB (Library of TU Dresden) on the major services and processes of the OSL platform, and investigate ways of promoting the platform to the library users. In addition, discussions on services that may be re-used by the library infrastructure will take place.

### Agenda (Friday January 24, 2014)

<b>1<sup>st</sup> Dresden Pilot Site Meeting Agenda</b>		
<b>10:00-10:15</b>	<b>Welcome and Intro</b>	
10:15-10:35	Presentation of TI and GoPubMed	TI
10:35-10:55	Presentation of the Bioinformatics Group (BIOTEC)	TUD
10:55-11:15	Presentation of SLUB	SLUB
11:15-11:45	Presentation of OSL major services	TUD, TI
<b>11:45-12:45</b>	<b>Lunch Break</b>	
12:45-13:15	Real Time Demonstrations of Tools and Services (Semantic Annotation, Trends, Sentiment Analysis)	TUD, TI
13:15 - 14:00	Demonstration of SLUB Data Management Platform	SLUB
<b>14:00-14:15</b>	<b>Coffee Break</b>	
14:15-15:00	Discussion: (a) Next Steps: Integration of OSL/GoPubMed services to SLUB platforms/projects. (b) Using Full Text Licenses for Text Mining (c) Attracting users to the OSL platform. Conclusions and Closing of the Meeting Planning of Next Meeting	All

### Participants List

<b>TUD</b>	Daniel Eisinger, George Tsatsaronis, Michael Schroeder
<b>TI</b>	Matthias Zschunke, Michael Alvers
<b>SLUB (TU Dresden Library)</b>	Jens Mittelbach, Ralf Talkenberger, Felix Lohmeier, Robert Glaß

## 6 APPENDIX II - Letter of Interest and Support from CNR /Tuscany Region G Monasterio Foundation (Research Hospital)

In the following, the original letter of interest and support is appended, as obtained by CNR.



Fondazione CNR/Regione Toscana per la  
Ricerca Medica e di Sanità Pubblica  
(L. R.T. n. 85/2009)

Fondazione Toscana Gabriele Monasterio

Prot.: 0002371 - 2014 - P

Data: 18/02/2014



Gent.mo  
Dr. Giorgio Iervasi  
Area della Ricerca CNR  
Via Moruzzi, 1  
56100 PISA

Facendo seguito alle precedenti comunicazioni, sono a confermare formalmente l'interesse di questa Fondazione a collaborare ai test pilota della nuova piattaforma informatica del progetto europeo OpenScience Link di cui la S.V. fa parte quale Partner e Task Leader.

Cordialmente

IL DIRETTORE GENERALE  
(Dr. Luciano Ciucci)

Direzione: Via Trieste 41, Pisa - Tel. 050-3153711 Fax 050-503596  
cod. fisc. 93062260505 - p. iva 01851550507