



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

Project **CIP-Pilot 325101 / OpenScienceLink**
Deliverable **D3.1**
Distribution **Public**



<http://opensciencelink.eu>

Detailed Specification of OpenScienceLink Pilot Scenarios and Services

Authors: Vassiliki Andronikou, Efstathios Karanastasis, George Tsatsaronis, Matthias Zschunke, Todor Tagarev, Petya Tagareva, Vladimir Savov, Greta Keremedchieva, Konstantinos Pantos, Giorgio Iervasi, Adomas Bunevicius

Status: Final (Version 1.14)

September 2013

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://opensciencelink.eu>
 Project start February 2013
 Project duration 3 years
 EC Project Officer Kirsti Ala-Mutka

Deliverable

Deliverable type Report
 Distribution level Public
 Deliverable Number D3.1
 Deliverable title Detailed Specification of OpenScienceLink Pilot Scenarios and Services
 Contractual date of delivery M6 (July 2013)
 Actual date of delivery 17 September 2013
 Relevant Task(s) WP3/Task3.1-3.4
 Partner Responsible NTUA
 Other contributors TUD, LUHS, NKUA, CNR, Procon, TI
 Number of pages 48
 Author(s) Vassiliki Andronikou, Efstathios Karanastasis, George Tsatsaronis, Matthias Zschunke, Todor Tagarev, Petya Tagareva, Vladimir Savov, Greta Keremedchieva, Konstantinos Pantos, Giorgio Iervasi, Adomas Bunevicius, Efthymios Chondrogiannis, Iordanis Mourouzis
 Internal Reviewers
 Status & version Final
 Keywords OpenScienceLink, scenario, functionality, service, model, semantics, social media

Executive Summary

The aim of D3.1 (henceforth "Deliverable") is to provide a detailed specification of the pilot scenarios and services (henceforth "Scenarios" and "Services") of each OpenScienceLink pilot (henceforth "Pilot") during the first iteration of the OpenScienceLink project. These scenarios and services are based on the requirements analysis which was held in WP2 and has been documented in D2.1. The outcome of the Deliverable should serve as input for the development of the architecture and the determination of the specifications of the OpenScienceLink platform (henceforth "Platform") architecture which comprise the main focus of WP4. Throughout the Deliverable our work is presented in a per Pilot basis and is organized in four sections (henceforth "Sections") as follows.

In Section 2, the detailed Pilot Scenarios descriptions are provided, based on the Pilot requirements (henceforth "Requirements") identified in D2.1, according to the input of the Platform's end users. Each Pilot Scenario is presented in discrete steps, in order to be easily followable by the reader. At the end of each step, the relevant OpenScienceLink platform functionalities (henceforth "Functionalities") are elicited. Towards the end of Section 2, the common Functionalities among Pilots are identified. Additionally, all Functionalities are summarized and numbered for easier reference.

In Section 3, the association of the Requirements stemming from D2.1 with each of the elicited Functionalities is presented. This also serves as a verification that all of the user requirements have been taken into account and are fully covered by the identified Functionalities.

In Section 4, the Functionalities are analyzed into OpenScienceLink Services, which are described briefly. The suitable user interfaces are identified and presented alongside, wherever applicable. Relevant models, that could be of value within the scope of each Pilot, are also presented, wherever applicable. At the end of Section 4, the common Services among the Pilots are presented, and all the Services are summarized and numbered.

Finally, Section 5 concludes the Deliverable.

Contents

Executive Summary	iii
Contents	iv
List of Tables	vi
1 Introduction.....	7
2 OpenScienceLink Pilot Scenarios and Functionalities	8
2.1 Research Dynamics-aware Open Access Data Journals Development	8
2.1.1 New journal issue suggestion	8
2.1.2 Dataset submission	8
2.1.3 Dataset peer review	9
2.1.4 Publishing.....	9
2.1.5 Assessment and evaluation.....	9
2.2 Novel open, semantically-assisted peer review process	10
2.2.1 Paper-based reviewers suggestion and selection	10
2.2.2 Review support.....	10
2.2.3 Review form submission.....	11
2.2.4 Open-identity post-review discussion	11
2.3 Data Mining for Biomedical and Clinical Research Trends Detection and Analysis	11
2.3.1 Trend Mining	11
2.4 Data Mining for Proactive Formulation of Scientific Collaborations.....	12
2.4.1 Dynamic and automated identification of researcher’s relationships	12
2.4.2 Dynamic and automated identification of groups of relevant researchers (around specific areas or interests)	12
2.5 Scientific field-aware, Productivity- and Impact-oriented Enhanced Research Evaluation Services	12
2.5.1 Research work evaluation.....	13
2.5.2 Researcher evaluation	13
2.5.3 Research group or community evaluation	13
2.5.4 Journal evaluation.....	13
2.5.5 Institution evaluation.....	14
2.5.6 Country evaluation	14
2.6 Summary of scenarios and functionalities	14
3 OpenScienceLink Functionalities and User Requirements	17
3.1 Research Dynamics-aware Open Access Data Journals Development	17
3.2 Novel open, semantically-assisted peer review process	19
3.3 Data Mining for Biomedical and Clinical Research Trends Detection and Analysis	21
3.4 Data Mining for Proactive Formulation of Scientific Collaborations.....	21
3.5 Scientific field-aware, Productivity- and Impact-oriented Enhanced Research Evaluation Services	25
4 OpenScienceLink Pilot Services and relevant Models.....	31
4.1 Research Dynamics-aware Open Access Data Journals Development	31

4.1.1	Services	31
4.1.2	Models and indicators.....	32
4.2	Novel open, semantically-assisted peer review process	33
4.2.1	Services	33
4.3	Data Mining for Biomedical and Clinical Research Trends Detection and Analysis	35
4.3.1	Services	35
4.4	Data Mining for Proactive Formulation of Scientific Collaborations.....	35
4.4.1	Services	35
4.5	Scientific field-aware, Productivity- and Impact-oriented Enhanced Research Evaluation	36
4.5.1	Services	37
4.5.2	Models and indicators.....	39
4.6	Common Services	43
4.6.1	Services	43
4.7	Summary of Services	44
5	Conclusion.....	48

List of Tables

Table 1: Summary of OpenScienceLink Platform Functionalities	16
Table 2: Association of Pilot 1 Requirements and Functionalities	18
Table 3: Association of Pilot 2 Requirements and Functionalities	21
Table 4: Association of Pilot 3 Requirements and Functionalities	21
Table 5: Association of Pilot 4 Requirements and Functionalities	25
Table 6: Association of Pilot 4 Requirements and Functionalities	30
Table 4: Summary of OpenScienceLink Platform Services	47

1 Introduction

This Deliverable comprises the outcome of the work which was held in WP3, and more specifically Tasks 3.1-3.4, during the first iteration of the OpenScienceLink project. It comprises the bridge between the specified requirements in WP2 and the pure technical specifications in WP4 at platform and component level. The presented analysis has been conducted on a per Pilot basis. For each Pilot, the collected requirements were analysed and the respective Scenario was developed. Based on this, the Platform functionalities for allowing the realisation of each scenario were extracted. For coherence and completion reasons, each specified Functionality was further analysed in terms of and mapped to the user requirements it corresponds to. Following this, the Services to be offered to the users and the models required for their realisation and provision were specified on a per pilot basis. Similarly to the Functionalities analysis process, each Service was further analysed and mapped to the respective Platform Functionality/ies.

2 OpenScienceLink Pilot Scenarios and Functionalities

Taking into account the Pilot Requirements elicited in D2.1, in this Section the Pilot Scenarios are presented from the end user's perspective, describing their step-by-step interaction with the Platform and connecting the user actions with the respective technical processes. The relevant Functionalities are then elicited for each step, and further analyzed into Services in Section 4.

2.1 Research Dynamics-aware Open Access Data Journals Development

This Pilot aims at providing the means for developing, publishing and managing data journals including a multitude of openly accessible data about experiments, trials and validation processes. The availability of data journals will facilitate researchers in accessing scientific datasets in the scope of their research, while at the same time supporting them in relevant benchmarking processes (on the basis of specific datasets). A key characteristic of the data journals development process will be its ability to leverage semantic technologies/capabilities enabling the instant identification of the most appropriate datasets for researcher wishing to use or augment existing data journals. These semantic capabilities will also empower scholars in the identification of research dynamics associated with specific datasets. Such dynamics will greatly facilitate them in planning their research.

As implied in the above Pilot description, it should cover the complete lifecycle of a data journal; from establishment and submission to evaluation and publishing.

2.1.1 New journal issue suggestion

The publisher logs in the Platform in order to identify scientifically hot topics for the creation of a new journal issue. S/he specifies a scientific field of interest, and the Platform suggests possible topics for the new journal issue, based on the trend detection and analysis that it performs on that field. Alternatively, the publisher directly enters a specific topic, requests for and is presented with its trend analysis measurements. At this stage the publisher is also able to see from the pool of datasets that have been submitted, which datasets are related to the designated topic. This way the publisher is aware of: (i) the significance of the issue to be released, based on the trend detection and analysis of topics, and, (ii) the coverage that these topics have in the uploaded/submitted datasets by the researchers.

Functionalities: user log-in, trend detection, trend analysis, relevant datasets detection

2.1.2 Dataset submission

A researcher registers at the Platform and then logs in, in order to submit a relevant data set for the specific call. Alternatively, the researcher may submit a dataset pertaining to any field covered by the active journals at any given time. In this case the submitted dataset will be published once a relevant issue or volume of the data journal is released, which can include the specified dataset. By using an appropriate user interface, the researcher specifies the dataset's parameters and metadata (e.g. creator, date or period of creation, overall description, purpose, conditions, features or parameters measured, level of noise, references to related works that used this dataset, how the data could be used, other notes or remarks that are important to bring into the attention of future users of the dataset), and uploads the dataset to the Platform, where it is stored for further evaluation. The dataset format will follow the OpenScienceLink data model (Section 4.1.2.1), and the file may be in any file format. Once the dataset is uploaded it becomes indexed and searchable from all Platform users, but not downloadable, which requires that the dataset is first reviewed and then published by the data journal.

Functionalities: user registration, user log-in, dataset with metadata submission, research work semantic linking

2.1.3 Dataset peer review

After the datasets have been submitted, they must be peer reviewed in order to be evaluated in terms of a series of factors, such as rationale, detail level, possible impact, etc. for inclusion in the journal issue. For this purpose, the dataset is primarily connected using semantic enabled technologies, such as text annotation with ontology concepts, with related literature, topics, and authors of the life sciences domains. The publisher can then identify the appropriate reviewers and then may assign certain datasets to certain reviewers for peer-reviewing. The reviewers evaluate the datasets according to different criteria, aided by the tools provided by the Platform. When the review is finished, the Platform publicizes the reviewers' comments, which can be viewed and discussed upon by other Platform users. The peer review process and its functionalities are described in detail in Section 2.2.

Functionalities: user log-in, research work semantic linking, relevant reviewers identification, dataset reviewing, comments publication, review commentary, semantic search with filtering and ranking

2.1.4 Publishing

Once the review of the dataset is over, and the dataset is accepted for publication, it enters into a special status within the dataset pool, flagged as 'ready to be published', and enables the publisher to select it as part of any future volume or issue for publication in the data journal. The candidate datasets for publishing comprise all of the peer-reviewed datasets. The rest of the uploaded datasets are either in the 'pending to be assigned reviewers' stage, or in the 'under review' stage. If further recommendations or clarifications are demanded through the review process for any given dataset, the dataset will still be in 'under review' status, until all conditions posed by the reviewers are met. At any given stage, the researchers who uploaded the dataset will be informed for the process, and will be able to follow the status of their dataset's reviewing process.

The publisher uses the Platform in order to publish the selected datasets in a predefined manner, according to particular specifications. The datasets and corresponding metadata can then be viewed and downloaded by the issue readers. Once a dataset is published, it is indexed with appropriate issue/volume numbers and pages, and turns into the 'published' stage. From this point on, the dataset is available to all of the Platform users for downloading and re-using in their research.

Functionalities: user log-in, dataset selection, dataset publication

2.1.5 Assessment and evaluation

The Platform provides the publisher with certain metrics for the assessment and evaluation of each dataset and the journal issue. This functionality is described in detail in Section 2.5. The Platform also keeps track of papers that are referring to or have used each dataset. Additionally, the semantic relationships among the medical concepts in the ontologies in the Platform allow for semantic linking services to exploit the classification of the medical terms and thus properly link the datasets with topics/work/researchers at different levels of granularity. The semantic linking of the datasets with the underlying ontology concepts will allow in principle better searching strategies for the datasets through the Platform. The main lever of semantically connecting the datasets with the rest Platform data entities (papers, authors, topics) is the list of the metadata and their values provided by the researchers during the uploading of the dataset. With this linking at hand, the publisher will be able to see and review statistics such as: number of times dataset appeared as result in queries, number of times dataset was downloaded, trendy

topics that the dataset is related with, and other metrics, that can help the shaping of the datasets' evaluation, and significance in any life science field.

Functionalities: user log-in, dataset assessment and evaluation, research work semantic linking

2.2 Novel open, semantically-assisted peer review process

This pilot aims at offering a novel peer-review process methodology, which consists of the following steps: Automated reviewers suggestion, reviewing support and actual review process, post-review discussion

2.2.1 Paper-based reviewers suggestion and selection

The publisher (or the responsible editor) logs in the Platform in order to initiate the review process of submitted papers or datasets. The first step is to identify the most appropriate reviewers for evaluating each paper or dataset, i.e. researchers who are highly scientifically related to the specific research topic the scientific paper or dataset resides in and thus could potentially serve as its reviewers, based on their research activity and output. The Platform proposes a list of potential reviewers, who are retrieved by using two different sources: (a) existing literature and published papers, (b) scientific interests and respective communities.

The next step for the publisher is to filter out from the proposed list of reviewers (for a specific research work) the scientists who directly or indirectly relate to the authors of the work under review. The Platform performs this evaluation and notifies the publisher in cases that conflict of interest is detected, e.g., through shared affiliations, if these are available. Next, the publisher or editor selects a number of the suggested reviewers to be invited for the reviewing process. In case the candidate reviewer is already registered with the Platform, the Platform notifies electronically the user that s/he has been selected as a candidate reviewer, and the user can then accept or decline the invitation. In the case that the candidate reviewer is not a registered user of the Platform, the Platform sends a similar invitation, with additional links so that the user can create in a simplified manner a user profile with the Platform. The publisher or editor is notified each time a candidate invited reviewer accepts or declines an invitation.

Functionalities: user log-in, research work semantic linking, relevant reviewers identification through author networks, relevant reviewers identification through literature, conflict of interest detection, reviewer pool suggestion, reviewer selection and invitation, user registration

2.2.2 Review support

If the candidate reviewer has accepted the invitation to review a paper or a dataset, s/he logs in the OpenScienceLink Platform in order to review the assigned work, or first registers, in case s/he is not a registered user of the Platform yet. The Platform aids the reviewer by providing a number of useful tools. First of all, the Platform provides a list of papers which are semantically related to the paper or the dataset under review. These papers serve as state of the art work to which the paper should be compared in terms of similarity and advancement. Additionally, the reviewer can suggest papers from the list for inclusion as references to the research work under review, given their semantic relevance to the research topic the work deals with. The reviewer can perform, at any time, semantic searches across a variety of open access data sources, in order to identify relevant literature, which will help him/her evaluate the paper or dataset under review.

The reviewer further uses the Platform's Semantic Filtering mechanism, which enables the semantic filtering of the initial semantic search results s/he is presented with. This filtering takes place through the ontologies incorporated in the OpenScienceLink Platform. In fact, the ontologies serve as a dynamic table of contents for each result set with which the reviewers are presented. Hence, the reviewer can view the top terms met across the results set and select a specific ontology term in order to narrow the results set and retrieve more accurate results.

The Platform allows the reviewer to apply a variety of ranking factors on the retrieved results, such as relevance, timeliness, quality of publications' fora, "authority" of author, etc., according to his/her preference.

Functionalities: user log-in, research work semantic linking, relevant papers retrieval, missing references suggestion, semantic search with filtering and ranking

2.2.3 Review form submission

After the reviewer has finished the paper or dataset evaluation s/he submits the review comments through an appropriate reviewing form. The reviewing form contains appropriate fields, scores, and text areas, where the reviewer can justify the assigned scores and comment on the paper's or dataset's contents, or suggest improvements.

The eventual decision of accepting, rejecting, or suggesting revisions for a submitted dataset or paper is taken by the publisher or the editor, who is notified each time a review form comes into the Platform.

Functionalities: user log-in, review form submission

2.2.4 Open-identity post-review discussion

Depending on the publisher's permissions, other users of the Platform are now able to access the reviewers' comments, as well as the corresponding paper or dataset, and discuss them, by making new comments, depending on the permissions set by the publisher or editor.

The comments submitted during the open-identity post-review discussion are annotated with terms from the ontologies incorporated into the OpenScienceLink platform. This way, the editor is able to easily navigate through the comments and filter them based on the ontology terms, and have a more efficient and fast way of having a clear view on the scientific community's feedback on the submitted reviews.

The Platform also helps the publisher evaluate the confidence of the comments received during the open-identity post-review discussion, based on the researcher's relevance and impact in the field. The latter functionality is described in detail in Section 2.5.

Functionalities: user log-in, review commentary, comment annotation, comment evaluation

2.3 Data Mining for Biomedical and Clinical Research Trends Detection and Analysis

This pilot emphasizes the ability of the OpenScienceLink Platform to discover and analyze research trends. The identification and analysis of such trends is essential for the allocation of research funding (by private sponsors and governmental agencies), as well as for the overall planning of research strategies. The pilot is primarily empowered by the advanced data mining capabilities of the Platform, on the basis of data for the biomedical and clinical research field.

2.3.1 Trend Mining

All users that log into the Platform are able to view current research trends in the biomedical domain, including topical research areas and areas with the highest growth. The users can see per community, research field and subfields thereof, the measured trendiness over a pre-specified timeslot. The users are also able to view in a ranked list and respective graphs the highest trends, e.g., the scientific subfields that have accumulated the largest mass of scientific interest based on publications of papers and datasets in the field. Alternatively, the user enters a specific topic, requests for and is presented with its trend analysis over the pre-specified timeslot. The timeslots can be user defined, among a list of options including year analysis, and 5-year analysis.

Functionalities: user log-in, trend detection, trend analysis

2.4 Data Mining for Proactive Formulation of Scientific Collaborations

This Pilot aims at enabling the networking and collaboration of researchers and scholars working on similar fields and sharing similar interests. It facilitates the creation and tracking of networks of researchers, in a way similar to social networks. However, a key distinguishing characteristic of this Pilot from conventional research information systems is the ability to infer relationships between researchers and research groups, including (in several cases) non-obvious, non-declared relationships. The Pilot is therefore expected to open new horizons in the collaboration of researchers, scholars, research groups and research organizations. The Pilot will be primarily empowered by the advanced social networking and data mining capabilities of the OpenScienceLink Platform.

2.4.1 Dynamic and automated identification of researcher's relationships

A researcher logs in the Platform in order to identify researchers working on similar fields or sharing similar interests that s/he can collaborate with. The Platform uses two different paths in order to infer relationships between researchers including (in several cases) non-obvious, non-declared relationships, and propose new collaborations: a) social networks through researcher's participation in groups/communities and b) literature through researcher's publications. The researcher filters the results according to their personal preferences (e.g. relevance to the field, excellence in the field, distance of residence, affiliation).

Functionalities: user log-in, researchers connection identification through social networks, researchers connections identification through literature, researchers collaboration filtering and proposition

2.4.2 Dynamic and automated identification of groups of relevant researchers (around specific areas or interests)

The researcher also wishes to use the Platform in order to identify research groups interested at the same scientific area and working on similar fields in order to possibly collaborate with them. The Platform combines the information from two different types of sources in order to infer relationships between the researcher and research groups including (in several cases) non-obvious relationships, and propose new collaborations: a) social networks for the provision of group characteristics, b) literature for the provision of author characteristics. The researcher filters the results according to his/her personal preferences (e.g., base location of group, research group importance in the field, etc.).

Functionalities: user log-in, research groups identification through social networks and literature, research groups collaboration filtering and proposition

2.5 Scientific field-aware, Productivity- and Impact-oriented Enhanced Research Evaluation Services

The main objective of this Pilot is to introduce, produce and track new objective metrics of research and scientific performance, beyond conventional metrics, associated with conventional indices and impact factors. This new type of sciento-metrics enables research sponsors, funding authorities and governmental agencies to shape their research strategies, researchers to be evaluated based on a multitude of factors representative of their productivity, impact and domain rather than through simplified means such as the number of citations within a time period, important research work, in terms of potential, to be brought forward, among others.

The Pilot uses mechanisms and Services that exploit the different metadata from the research data sources and the social networks in order to provide representative and complete evaluation of research outputs (articles, data sets, etc), researchers, institution and communities.

2.5.1 Research work evaluation

A user logs in the Platform in order to evaluate a research work (paper or dataset) by choosing among various metrics provided by the Platform.

One of them is connected with the dynamics of the field that the research work resides in, i.e. whether the research field is in its infancies, at its peak or in decline. According to this metric, the impact of the research work is weighted depending on the research potential and dynamics of the specific field.

Another metric, which gives a qualitative flair to the strictly quantitative evaluation of research, is through the number of the citations the research work has. Additionally, the timeliness and density of the citations can be used as an evaluation metric.

The measured impact of the research work is also affected by the level of difficulty for other researchers to have access to this work. Self-archiving authorization, publisher archiving provisions, copyright provisions, abstracting and indexing services, and reference linking, are some of the factors that affect this metric.

The work can finally be evaluated based on the potential interest about it, the number of other researchers who downloaded it and their own research evaluation.

Based on the weighted combination of the above metrics, the Platform calculates the overall evaluation of a research work. The weights may vary depending on the type of research work.

Functionalities: user log-in, research field dynamics evaluation, source of citations evaluation, citation dynamics evaluation, quality and broadness of dissemination evaluation, potential interest evaluation, overall research work evaluation

2.5.2 Researcher evaluation

The user uses the Platform in order to evaluate a researcher. The researcher's evaluation is based on the combination of the overall evaluation of all of their known research works, stemming from the previous step. Alternatively, the individual metrics for research work evaluation can be used in order calculate the researcher's evaluation by utilizing different weighting than before.

Functionalities: user log-in, researcher evaluation

2.5.3 Research group or community evaluation

The user uses the Platform in order to evaluate a research group or community. Each research group or community consists of a number of member researchers. By using the researcher evaluation (previous step) for the members of a scientific group or community, the Platform calculates the group's or community's evaluation. For the calculation of a research group's evaluation, the evaluation of the group's leader has a greater weight than of the rest of its members. On the other hand, for the evaluation of a research community only its top members are considered, but with equal weighting.

Functionalities: user log-in, research group evaluation, research community evaluation

2.5.4 Journal evaluation

The user uses the Platform in order to evaluate a journal. The journal is handled as an assortment of research works. The Platform calculates the overall evaluation of each of the research works that have been published in the journal individually, and combines them in order to provide the journal evaluation. Alternatively, the individual metrics for research work evaluation can be used in order calculate the journal's evaluation by utilizing different weighting than before.

Functionalities: user log-in, journal evaluation

2.5.5 Institution evaluation

The user uses the Platform in order to evaluate an institution. The institution's evaluation is based on the combination of the evaluation of the researchers stemming from that institution, and is calculated per research area (domain). Only researchers whose name appears first in the authors list of a research work are considered. The overall evaluation of an institution is calculated based on the individual domains evaluation.

Functionalities: user log-in, institution evaluation per domain, overall institution evaluation

2.5.6 Country evaluation

The user uses the Platform in order to evaluate a country. The evaluation of a country is the result of the combination of the evaluation of each of the researchers in that country, and is calculated per research area (domain). The overall evaluation of a country is calculated based on the individual domains evaluation.

Functionalities: user log-in, country evaluation per domain, overall country evaluation

2.6 Summary of scenarios and functionalities

In the above sections, the Pilot scenarios were analyzed and the OpenScienceLink Platform Functionalities were elicited.

Some of these functionalities are encountered in more than one Pilots and thus comprise common Functionalities of the Platform and are given a common-category id. They mainly have to do with user handling and security, as well as the semantic linking of the research work, all which are of fundamental functionalities of the Platform shared among Pilots.

Some other functionalities are encountered twice, but this is due to the fact that Pilot 1 encompasses some of the Functionalities of the other Pilots. In this case, the functionalities are given an id that corresponds to the Pilot where they originally stem from.

The following table summarizes and numbers the elicited Functionalities.

Pilot	Scenario Step	F-ID	Functionality
Research Dynamics-aware Open Access Data Journals Development	New journal issue suggestion	F3.1	Trend detection
		F3.2	Trend analysis
		F1.1	Relevant datasets detection
	Dataset submission	F1.2	Dataset with metadata submission
	Dataset peer review	F2.1, F2.2, F2.3, F2.4, F2.5	Relevant reviewers identification
		F2.6, F2.7	Dataset reviewing
		F2.8	Comments publication
		F2.9	Review commentary
	Publishing	F1.3	Dataset selection
		F1.4	Dataset publication

	Assessment and evaluation	F5.1, F5.2, F5.3, F5.4, F5.5, F5.6	Dataset assessment and evaluation
Novel open, semantically-assisted peer review process	Paper-based reviewers suggestion	F2.1	Relevant reviewers identification through author networks
		F2.2	Relevant reviewers identification through literature
		F2.3	Conflict of interest detection
		F2.4	Reviewer pool suggestion
		F2.5	Reviewer selection and invitation
	Review support	F2.6	Relevant papers retrieval
		F2.7	Missing references suggestion
	Review form submission	F2.8	Review form submission
	Open-identity post-review discussion	F2.9	Review commentary
		F2.10	Comment annotation
		F2.11	Comment evaluation
Data Mining for Biomedical and Clinical Research Trends Detection and Analysis	Trend Mining	F3.1	Trend detection
		F3.2	Trend analysis
Data Mining for Proactive Formulation of Scientific Collaborations	Dynamic and automated identification of researcher's relationships	F4.1	Researchers connections identification through social networks
		F4.2	Researchers connections identification through literature
		F4.3	Researchers collaboration filtering and proposition
	Dynamic and automated identification of groups of relevant researchers (around specific areas or interests)	F4.4	Research groups identification through social networks and literature
		F4.5	Research groups collaboration filtering and proposition
Scientific field-aware, Productivity- and Impact-oriented Enhanced Research Evaluation	Research work evaluation	F5.1	Research field dynamics evaluation
		F5.2	Source of citations evaluation
		F5.3	Citation dynamics evaluation
		F5.4	Quality and broadness of dissemination evaluation

		F5.5	Potential interest evaluation
		F5.6	Overall research work evaluation
	Researcher evaluation	F5.7	Researcher evaluation
	Research group or community evaluation	F5.8	Research group evaluation
		F5.9	Research community evaluation
	Journal evaluation	F5.10	Journal evaluation
	Institution evaluation	F5.11	Institution evaluation per domain
		F5.12	Overall institution evaluation
	Country evaluation	F5.13	Country evaluation per domain
		F5.14	Overall country evaluation
All	-	F0.1	User registration and profile
		F0.2	User log-in
		F0.3	Research work semantic linking
		F0.4	Semantic search with filtering and ranking

Table 1: Summary of OpenScienceLink Platform Functionalities

3 OpenScienceLink Functionalities and User Requirements

In this section the Functionalities are presented in regard with the Pilot Requirements identified in D2.1. The Pilot Requirements are mapped to the Functionalities, showing the close relationship between the user and technical requirements. As a matter of fact, all of the core and essential Pilot Requirements are tackled by the identified Functionalities, as well as most of the desired ones. There is one section for every Pilot.

3.1 Research Dynamics-aware Open Access Data Journals Development

The following table presents the Pilot 1 Functionalities and their relation to the User Requirements identified in D2.1.

F-ID	Functionality	Pilot Requirement	PR-ID
F3.1	Trend detection	Identify trends, emerging research fields, and fields in decline;	F_PR1.10
F3.2	Trend analysis	Identify trends, emerging research fields, and fields in decline.	F_PR1.10
		Identify and prioritize potential members of the Editorial Board based on initial and emerging research fields to be covered by the journal	F_PR1.11
F1.1	Relevant datasets detection	Identify potential contributors to the journal	F_PR1.12
F1.2	Dataset with metadata submission	The registered researchers should be able to upload their scientific data for sharing and publication under an open-access data journal	F_PR1.3
F2.1, 2.2, F2.3, 2.4, F2.5	Relevant identification reviewers	Accept or reject an invitation for review	F_PR1.8
		Identify and prioritize candidate reviewers per submitted contribution while eliminating potential conflict of interests	F_PR1.13
F2.6, F2.7	Dataset reviewing	Researchers should be given the ability to peer review research works and openly discuss and express opinions on the reviews and the review results.	F_PR1.4
		Retrieve related bibliography covering the topics or the datasets	F_PR1.6
		Retrieve related datasets to suggest comparisons	F_PR1.7
		Facilitate the review process	F_PR1.14
		Researchers should be given the ability to peer review research works and openly discuss and express opinions on the reviews and the review results.	F_PR1.4

		Maintain the reviews within the platform for further processing, e.g. discussion with colleague researchers.	F_PR1.9
F2.8	Comments publication	Researchers should be given the ability to peer review research works and openly discuss and express opinions on the reviews and the review results.	F_PR1.4
		Maintain the reviews within the platform for further processing, e.g. discussion with colleague researchers.	F_PR1.9
F2.9	Review Commentary	Researchers should be given the ability to peer review research works and openly discuss and express opinions on the reviews and the review results	F_PR1.4
F1.3	Dataset selection	For an uploaded dataset, retrieve the metadata and the data.	F_PR1.5
		Automate the publication management process	F_PR1.21
F1.4	Dataset publication	Provide permanent access to the data journal and its published work, quickly and in long-term.	F_PR1.17
		Automate the publication management process	F_PR1.21
F5.1,F5.2, F5.3, 5.4, F5.5, F5.6	Dataset assessment and evaluation	Automate the publication management process	F_PR1.21
		Automate the maintenance process	F_PR1.22
F0.1	User registration and profile	The researchers should be able to register and create a profile based on their academic/research interests.	F_PR1.1
F0.3	Research work semantic linking	Retrieve related bibliography covering the topics or the datasets	F_PR1.6
		Link data journals to other relevant publications and company products and services	F_PR1.24
F0.4	Semantic search with filtering and ranking	Retrieve related bibliography covering the topics or the datasets	F_PR1.6
		Retrieve related datasets to suggest comparisons	F_PR1.7
		Search across literature for papers related to the dataset	F_PR1.25
		Filter the retrieved search results based on domain specific terms	F_PR1.26
		Rank the retrieved search results based on relevance, authority of author and timeliness	F_PR1.27

Table 2: Association of Pilot 1 Requirements and Functionalities

**D3.1 Detailed Specification of
OpenScienceLink Pilot Scenarios and Services**

3.2 Novel open, semantically-assisted peer review process

The following table presents the Pilot 2 Functionalities and their relation to the User Requirements identified in D2.1.

F-ID	Functionality	Pilot Requirement	PR-ID
F2.1	Relevant identification of reviewers through author networks	The publisher submits a request to retrieve potential evaluators based on an article text. Also unregistered researchers (e.g. authors) are retrieved	F_PR2.10
F2.2	Relevant identification of reviewers through literature	The publisher submits a request to retrieve potential evaluators based on an article text. Also unregistered researchers (e.g. authors) are retrieved	F_PR2.10
F2.3	Conflict of interest detection	The publisher submits a request to retrieve potential evaluators based on an article text. Also unregistered researchers (e.g. authors) are retrieved	F_PR2.10
F2.4	Reviewer pool suggestion	The publisher submits a request to retrieve potential evaluators based on an article text. Also unregistered researchers (e.g. authors) are retrieved	F_PR2.10
F2.5	Reviewer selection and invitation	The researcher receives an invitation to be a reviewer	F_PR2.4
		The user uses a form to accept or reject the invitation. In the later case s/he should put a recommendation of a colleague that could do the review	F_PR2.5
F2.6	Relevant papers retrieval	The evaluator can submit a request to get related information for the uploaded article, including other publications of the author	F_PR2.9
		The evaluator retrieves the article text via the platform	F_PR2.11
		Retrieve related bibliography covering the topics or the datasets	F_PR2.17
F2.7	Missing references suggestion	Retrieve related bibliography covering the topics or the datasets	F_PR2.17
		Retrieve cited bibliography	F_PR2.18
F2.8	Review form submission	The evaluator submits a review to the platform for the authors / researchers, editor and funder to access.	F_PR2.6
		The evaluator uses a form to put the review results	F_PR2.7
		The publisher sets up a form in which the evaluator can place his/her review	F_PR2.8

		Researchers should be given the ability to peer review research works and openly discuss and express opinions on the reviews and the review results	F_PR2.15
		Maintain the reviews within the platform for further processing, e.g. discussion with colleague researchers	F_PR2.21
F2.9	Review commentary	The registered researchers should be able to comment/discuss previously published research works or datasets	F_PR2.14
		Researchers should be given the ability to peer review research works and openly discuss and express opinions on the reviews and the review results	F_PR2.15
		View reviews of other evaluators on already published work	F_PR2.20
		Evaluators are anonymous w.r.t. the authors / researchers and colleague evaluators.	F_PR2.23
F2.10	Comment annotation	View reviews of other evaluators on already published work	F_PR2.20
F0.1	User registration and profile	The evaluator has to set up a profile with links to his/her articles. This will enable recruitment by editors and funders	F_PR2.2
		The evaluator includes major topics s/he is working on into the profile	F_PR2.3
		Evaluators provide a valid email address, so they can be contacted by editors and funders	F_PR2.24
F0.2	User log-in	The user (evaluator, editor, publisher) has to login to authenticate	F_PR2.1
F0.3	Research work semantic linking	The evaluator can submit a request to get related information for the uploaded article, including other publications of the author	F_PR2.9
F0.4	Semantic search with filtering and ranking	The evaluator can further filter by topics and other biomedical concepts, i.e. semantically search	F_PR2.12
		The evaluator can search through literature for related scientific work and advancements in the field of the research work under review.	F_PR2.13
		Retrieve related bibliography covering the topics or the datasets	F_PR2.17
		Retrieve cited bibliography	F_PR2.18
		Retrieve related datasets to suggest comparisons	F_PR2.19

		The evaluator can filter and refine their literature searches through pre-defined domain specific terms.	F_PR2.25
		The evaluator can rank the results of their searches across published scientific work in terms of relevance to search, authority of authors and timeliness.	F_PR2.26

Table 3: Association of Pilot 2 Requirements and Functionalities

3.3 Data Mining for Biomedical and Clinical Research Trends Detection and Analysis

The following table presents the Pilot 3 Functionalities and their relation to the User Requirements identified in D2.1.

F-ID	Functionality	Pilot Requirement	PR-ID
F3.1	Trend detection	The registered users should be able to see the advances (emerging fields and trendiness) in their fields of interest.	F_PR3.1
		Retrieve information on the scientific direction and trends of a specific topic	F_PR3.4
F3.2	Trend analysis	The registered researchers should be able to see the top authors, research institutes and countries in the field of their study and interest.	F_PR3.2
		Retrieve information on the scientific direction and trends of a specific topic	F_PR3.4
		Identify trends, emerging research fields, and fields in decline;	F_PR3.6
F0.4	Semantic search with filtering and ranking	The registered users should be able to perform free-text searches across published research work.	F_PR3.3
		Filter search results based on domain specific terms selected by the user	F_PR3.9
		Rank retrieved search results based on their relevance to the submitted search, the authors' authority and in chronological order	F_PR3_10

Table 4: Association of Pilot 3 Requirements and Functionalities

3.4 Data Mining for Proactive Formulation of Scientific Collaborations

The following table presents the Pilot 4 Functionalities and their relation to the User Requirements identified in D2.1.

F-ID	Functionality	Pilot Requirement	PR-ID
F4.1	Researchers connection identification through social	The user (researcher or publisher) can request for suggestions of scientific collaborations with researchers	F_PR4.1

	networks	The suggested collaborations among researchers should be on the basis of the degree of relevance of the research topics and fields they work on, as indicated across their published work and/or their participation in research communities. They should also take into consideration their former and existing collaborations and social-media declared relationships	F_PR4.3
F4.2	Researchers identification through literature connections	The user (researcher or publisher) can request for suggestions of scientific collaborations with researchers	F_PR4.1
		The suggested collaborations among researchers should be on the basis of the degree of relevance of the research topics and fields they work on, as indicated across their published work and/or their participation in research communities. They should also take into consideration their former and existing collaborations and social-media declared relationships	F_PR4.3
F4.3	Researchers proposition collaboration	The user (researcher or publisher) can view the generated suggestions of scientific collaborations with researchers in the form of a list of names	F_PR4.2
		Each of the suggested scientific collaborations with other researchers is accompanied by an indicative list of the scientific papers which are highly related to the user's work (if researcher, then the published scientific work and if publisher, then the journal's topics)	F_PR4.7
		Each of the suggested scientific collaborations with other researchers is further accompanied by: <ul style="list-style-type: none"> a. (in case of a researcher initiating the request) the scientific topics, fields and/or areas in which they share common interests with the researcher b. (in case of a publisher/editor initiating the request) the scientific topics, fields and/or areas that the researchers work on which overlap with the ones of the publisher's journal/issue 	F_PR4.8
		Each of the suggested scientific collaborations with researchers is accompanied by a percentage indicating how strong the suggested collaboration is	F_PR4.9

		The researcher can request for receiving regular notifications about suggestions of scientific collaborations with researchers, research groups and/or communities via e-mail	F_PR4.12
		The researcher can receive and view e-mail notifications about suggestions of scientific collaborations with researchers, research groups and/or communities regularly (e.g., once biweekly)	F_PR4.13
F4.4	Research group identification through social networks and literature	The researcher can request for suggestions of scientific collaborations with research groups and/or research communities.	F_PR4.4
		The suggested collaborations between researchers and research groups and/or communities should be on the basis of the degree of relevance of the research topics and fields they deal with, as indicated across the researchers' published work and/or their participation in research communities and the communities' and groups' scientific topics and members profile (if available). They should also take into consideration the researchers' existing collaborations with research groups and communities	F_PR4.6
		The researcher can request for receiving regular notifications about suggestions of scientific collaborations with researchers, research groups and/or communities via e-mail	F_PR4.12
F4.5	Research group collaboration proposition	The researcher can view the generated suggestions of scientific collaborations with research groups and/or scientific communities in the form of a list of groups and/or communities highly related to their work.	F_PR4.5
		Each of the suggested scientific collaborations with research groups and/or communities is accompanied by the scientific topics, fields and/or areas that they cover and are common with the researcher's interest.	F_PR4.10
		Each of the suggested scientific collaborations with research groups and/or communities is accompanied by a percentage indicating how strong the suggested collaboration is.	F_PR4.11
		The researcher can receive and view e-mail notifications about suggestions of scientific collaborations with researchers, research groups and/or communities regularly (e.g., once biweekly).	F_PR4.13

F0.1	User registration and profile	The suggested collaborations among researchers should be on the basis of the degree of relevance of the research topics and fields they work on, as indicated across their published work and/or their participation in research communities. They should also take into consideration their former and existing collaborations and social-media declared relationships	F_PR4.3
		The suggested collaborations between researchers and research groups and/or communities should be on the basis of the degree of relevance of the research topics and fields they deal with, as indicated across the researchers' published work and/or their participation in research communities and the communities' and groups' scientific topics and members profile (if available). They should also take into consideration the researchers' existing collaborations with research groups and communities.	F_PR4.6
		Each of the suggested scientific collaborations with other researchers is accompanied by an indicative list of the scientific papers which are highly related to the user's work (if researcher, then the published scientific work and if publisher, then the journal's topics)	F_PR4.7
		Each of the suggested scientific collaborations with other researchers is further accompanied by: <ul style="list-style-type: none"> a. (in case of a researcher initiating the request) the scientific topics, fields and/or areas in which they share common interests with the researcher b. (in case of a publisher/editor initiating the request) the scientific topics, fields and/or areas that the researchers work on which overlap with the ones of the publisher's journal/issue 	F_PR4.8
		Each of the suggested scientific collaborations with research groups and/or communities is accompanied by the scientific topics, fields and/or areas that they cover and are common with the researcher's interest.	F_PR4.10
		The researcher, who is among top scientist in a specific field (as indicated by pilot 3 and/or pilot 5), is suggested with leading a new, automatically formed research community in the research topic/field s/he excels at.	F_PR4.14

		Researchers are informed about new, automatically formed research communities within their research topic(s)/field(s).	F_PR4.15
		The profile should be used to automatically suggest possible research collaborations at the international and domestic level.	F_PR4.16
F0.4	Semantic search with filtering and ranking	Search through published scientific work for specific pieces (papers, articles, datasets, etc) of a researcher and for research work related to a topic	F_PR4.18
		Filter search results based on domain specific terms selected by the user	F_PR4.19
		Rank retrieved search results based on their relevance to the submitted search, the authors' authority and in chronological order	F_PR4.20

Table 5: Association of Pilot 4 Requirements and Functionalities

3.5 Scientific field-aware, Productivity- and Impact-oriented Enhanced Research Evaluation Services

The following table presents the Pilot 4 Functionalities and their relation to the User Requirements identified in D2.1.

F-ID	Functionality	Pilot Requirement	PR-ID
F5.1	Research field dynamics evaluation	The user is presented with all the available evaluation metrics applied in each case (paper, data set, journal, etc).	F_PR5.5
		The user is able to choose among the available evaluation metrics based on which s/he prefers to view evaluation results in each case (paper, data set, journal, etc).	F_PR5.6
F5.2	Source of citations evaluation	The user is presented with all the available evaluation metrics applied in each case (paper, data set, journal, etc).	F_PR5.5
		The user is able to choose among the available evaluation metrics based on which s/he prefers to view evaluation results in each case (paper, data set, journal, etc).	F_PR5.6
F5.3	Citation dynamics evaluation	The user is presented with all the available evaluation metrics applied in each case (paper, data set, journal, etc).	F_PR5.5
		The user is able to choose among the available evaluation metrics based on which s/he prefers to view evaluation results in each case (paper, data set, journal, etc).	F_PR5.6
F5.4	Quality and broadness of dissemination evaluation	The user is presented with all the available evaluation metrics applied in each case (paper, data set, journal, etc).	F_PR5.5

		The user is able to choose among the available evaluation metrics based on which s/he prefers to view evaluation results in each case (paper, data set, journal, etc).	F_PR5.6
F5.5	Potential interest evaluation	The user is presented with all the available evaluation metrics applied in each case (paper, data set, journal, etc).	F_PR5.5
		The user is able to choose among the available evaluation metrics based on which s/he prefers to view evaluation results in each case (paper, data set, journal, etc).	F_PR5.6
F5.6	Overall research work evaluation	The user can request for viewing the evaluation for a specific paper, data set, journal and/or researcher s/he chooses.	F_PR5.1
		The user is able to choose a specific research topic, field or area and view the evaluation of its top 20 subjects and objects (i.e., researchers, papers, data sets, etc) based on the available metrics.	F_PR5.7
		The user is able to rank the evaluation subjects and objects (i.e., researchers, papers, data sets, etc) in a specific research topic, field or area based on each evaluation metric.	F_PR5.8
		The researcher can access, at any time, their own evaluation along with the current evaluation of their published papers and data sets as well as of their research group and/or institution.	F_PR5.9
		The publisher can access, at any time, the evaluation of their own journals, journal issues, published papers and data sets.	F_PR5.10
		The user may choose to follow the evaluation of specific papers, data sets, journals, researchers, research groups, communities and institutions.	F_PR5.11
		Researchers should be able to have access to the results of research evaluation at all levels – journal, individual researcher, research group, research institute/university, country	F_PR5.13
F5.7	Researcher evaluation	The user can request for viewing the evaluation for a specific paper, data set, journal and/or researcher s/he chooses.	F_PR5.1
		The user is presented with all the available evaluation metrics applied in each case (paper, data set, journal, etc).	F_PR5.5
		The user is able to choose among the available evaluation metrics based on which s/he prefers to view evaluation results in each case (paper, data set, journal, etc).	F_PR5.6

		The user is able to choose a specific research topic, field or area and view the evaluation of its top 20 subjects and objects (i.e., researchers, papers, data sets, etc) based on the available metrics.	F_PR5.7
		The user is able to rank the evaluation subjects and objects (i.e., researchers, papers, data sets, etc) in a specific research topic, field or area based on each evaluation metric.	F_PR5.8
		The researcher can access, at any time, their own evaluation along with the current evaluation of their published papers and data sets as well as of their research group and/or institution.	F_PR5.9
		The user may choose to follow the evaluation of specific papers, data sets, journals, researchers, research groups, communities and institutions.	F_PR5.11
		Researchers should be able to have access to the results of research evaluation at all levels – journal, individual researcher, research group, research institute/university, country	F_PR5.13
F5.8	Research group evaluation	The user can request for viewing the evaluation for a specific research group and/or community s/he chooses.	F_PR5.2
		The user is able to choose a specific research topic, field or area and view the evaluation of its top 20 subjects and objects (i.e., researchers, papers, data sets, etc) based on the available metrics.	F_PR5.7
		The user is able to rank the evaluation subjects and objects (i.e., researchers, papers, data sets, etc) in a specific research topic, field or area based on each evaluation metric.	F_PR5.8
		The researcher can access, at any time, their own evaluation along with the current evaluation of their published papers and data sets as well as of their research group and/or institution.	F_PR5.9
		The user may choose to follow the evaluation of specific papers, data sets, journals, researchers, research groups, communities and institutions.	F_PR5.11
		Researchers should be able to have access to the results of research evaluation at all levels – journal, individual researcher, research group, research institute/university, country	F_PR5.13
F5.9	Research evaluation community	The user can request for viewing the evaluation for a specific research group and/or community s/he chooses.	F_PR5.2

		The user is presented with all the available evaluation metrics applied in each case (paper, data set, journal, etc).	F_PR5.5
		The user is able to choose among the available evaluation metrics based on which s/he prefers to view evaluation results in each case (paper, data set, journal, etc).	F_PR5.6
		The user is able to choose a specific research topic, field or area and view the evaluation of its top 20 subjects and objects (i.e., researchers, papers, data sets, etc) based on the available metrics.	F_PR5.7
		The user is able to rank the evaluation subjects and objects (i.e., researchers, papers, data sets, etc) in a specific research topic, field or area based on each evaluation metric.	F_PR5.8
		The user may choose to follow the evaluation of specific papers, data sets, journals, researchers, research groups, communities and institutions.	F_PR5.11
		Researchers should be able to have access to the results of research evaluation at all levels – journal, individual researcher, research group, research institute/university, country	F_PR5.13
F5.10	Journal evaluation	The user can request for viewing the evaluation for a specific paper, data set, journal and/or researcher s/he chooses.	F_PR5.1
		The user is presented with all the available evaluation metrics applied in each case (paper, data set, journal, etc).	F_PR5.5
		The user is able to choose among the available evaluation metrics based on which s/he prefers to view evaluation results in each case (paper, data set, journal, etc).	F_PR5.6
		The user is able to choose a specific research topic, field or area and view the evaluation of its top 20 subjects and objects (i.e., researchers, papers, data sets, etc) based on the available metrics.	F_PR5.7
		The user is able to rank the evaluation subjects and objects (i.e., researchers, papers, data sets, etc) in a specific research topic, field or area based on each evaluation metric.	F_PR5.8
		The publisher can access, at any time, the evaluation of their own journals, journal issues, published papers and data sets.	F_PR5.10

		The user may choose to follow the evaluation of specific papers, data sets, journals, researchers, research groups, communities and institutions.	F_PR5.11
		Researchers should be able to have access to the results of research evaluation at all levels – journal, individual researcher, research group, research institute/university, country	F_PR5.13
F5.11	Institution evaluation per domain	The user can request for viewing the evaluation for a specific academic and/or research institution and country s/he chooses for a specific domain.	F_PR5.4
		The user is presented with all the available evaluation metrics applied in each case (paper, data set, journal, etc).	F_PR5.5
		The user is able to choose among the available evaluation metrics based on which s/he prefers to view evaluation results in each case (paper, data set, journal, etc).	F_PR5.6
		The researcher can access, at any time, their own evaluation along with the current evaluation of their published papers and data sets as well as of their research group and/or institution.	F_PR5.9
F5.12	Overall institution evaluation	The user can request for viewing the evaluation for a specific academic and/or research institution and country s/he chooses.	F_PR5.3
		The user is able to choose a specific research topic, field or area and view the evaluation of its top 20 subjects and objects (i.e., researchers, papers, data sets, etc) based on the available metrics.	F_PR5.7
		The user is able to rank the evaluation subjects and objects (i.e., researchers, papers, data sets, etc) in a specific research topic, field or area based on each evaluation metric.	F_PR5.8
		The user may choose to follow the evaluation of specific papers, data sets, journals, researchers, research groups, communities and institutions.	F_PR5.11
		Researchers should be able to have access to the results of research evaluation at all levels – journal, individual researcher, research group, research institute/university, country	F_PR5.13
F5.13	Country evaluation per domain	The user can request for viewing the evaluation for a specific academic and/or research institution and country s/he chooses for a specific domain.	F_PR5.4

		The user is presented with all the available evaluation metrics applied in each case (paper, data set, journal, etc).	F_PR5.5
		The user is able to choose among the available evaluation metrics based on which s/he prefers to view evaluation results in each case (paper, data set, journal, etc).	F_PR5.8
F5.14	Overall country evaluation	The user can request for viewing the evaluation for a specific academic and/or research institution and country s/he chooses.	F_PR5.3
		The user is able to choose a specific research topic, field or area and view the evaluation of its top 20 subjects and objects (i.e., researchers, papers, data sets, etc) based on the available metrics.	F_PR5.7
		The user is able to rank the evaluation subjects and objects (i.e., researchers, papers, data sets, etc) in a specific research topic, field or area based on each evaluation metric.	F_PR5.8
		Researchers should be able to have access to the results of research evaluation at all levels – journal, individual researcher, research group, research institute/university, country	F_PR5.13
F0.1	User registration and profile	The researcher can access, at any time, their own evaluation along with the current evaluation of their published papers and data sets as well as of their research group and/or institution.	F_PR5.9
		The publisher can access, at any time, the evaluation of their own journals, journal issues, published papers and data sets.	F_PR5.10

Table 6: Association of Pilot 4 Requirements and Functionalities

4 OpenScienceLink Pilot Services and relevant Models

In this section the Services per Pilot are presented, based on the analysis conducted in the previous sections of the document. In specific, the Services required for the implementation of each of the previously elicited Functionalities are defined and detailed. There is one sub-section for each Pilot, as well as a common sub-section, where the Services that apply to all Pilots are described. Wherever applicable, relevant Models are also presented, analysed and extended, in order to be of value within the scope of the Pilot.

The Platform should make the Services available to its users / stakeholders through web-based user interfaces (WebUI), as required. These user interfaces should be described in detail in D4.1.

4.1 Research Dynamics-aware Open Access Data Journals Development

This Section describes the Services and models needed to enable the Pilot 1 related Functionalities.

4.1.1 Services

4.1.1.1 Trend detection

- **Trend detection Service(s)**
Service described in Section 4.3.1.1.

4.1.1.2 Trend analysis

- **Trend analysis Service(s)**
Service described in Section 4.3.1.2.

4.1.1.3 Relevant datasets detection

- **Relevant datasets detection Service**
Filters the pool of submitted datasets according to the topic and presents the relevant datasets to the user.

4.1.1.4 Dataset with metadata submission

- **Dataset storage Service**
Saves the dataset in the Platform.
- **Metadata storage Service**
Saves the accompanying metadata in the Platform.

4.1.1.5 Relevant reviewers identification

- **Relevant reviewers identification Service(s)**
Services described in 4.2.1.1, 4.2.1.2, 4.2.1.3, 4.2.1.4 and 4.2.1.5.

4.1.1.6 Dataset reviewing

- **Dataset reviewing Service(s)**
Services described in 4.2.1.6, 4.2.1.7 and 4.2.1.8.

4.1.1.7 Comments publication

- **Comments publication Service(s)**
Services described in 4.2.1.8.

4.1.1.8 Review commentary

- **Review commentary Service(s)**
Services described in 4.2.1.9

4.1.1.9 Dataset selection

- **Dataset retrieval Service**
Retrieves a specific dataset form the OpenScienceLink Platform.
- **Dataset selection Service**
Flags the dataset as selected for inclusion in the specific journal issue.

4.1.1.10 Dataset publication

- **Dataset publication Service**
Organizes the presentation of datasets and metadata according to certain specifications and publishes them, making use of the OpenScienceLink dataset model (Section 4.1.2.1).
- **Dataset download Service**
Retrieves a dataset and sends it to the users browser

4.1.1.11 Dataset assessment and evaluation

- **Assessment and Evaluation Service(s)**
Services described in Section 4.5.1.

4.1.2 Models and indicators

This section presents models and indicators which are relevant to Pilot 1 and should be taken into account when designing the corresponding services.

4.1.2.1 Metadata models for raw data

The OpenScienceLink metadata model will be based on and comprise an extension of the Unidata's Common Data Model (CDM) and the Dryad Metadata Application Profile, including parameters such as dataset source type (real-world vs. synthetic), level of noise, popularity (e.g., number of views, number of downloads), associated research topics, citation source (i.e., which researchers have used this dataset), among others.

- **Unidata's Common Data Model (CDM)**

The Common Data Model (CDM) is an abstract data model for scientific datasets. It merges the netCDF, OPeNDAP, and HDF5 data models to create a common API for many types of scientific data.

CDM has three layers, which build on top of each other to add successively richer semantics:

1. The data access layer, also known as the syntactic layer, handles data reading and writing.
2. The coordinate system layer identifies the coordinates of the data arrays. Coordinates are a completely general concept for scientific data; we also identify specialized georeferencing coordinate systems, which are important to the Earth Science community.
3. The scientific feature type layer identifies specific types of data, such as grids, radial, and point data, adding specialized methods for each kind of data.

- **Dryad Metadata Application Profile**

The Dryad Metadata Application profile is based on the Dublin Core Metadata Initiative Abstract Model, including the Singapore Framework, used to describe multi-disciplinary data underlying peer-reviewed scientific and medical literature. The application profile consists of three modules describing the:

1. The publication, which is an article associated with content in Dryad
2. The data package, which is a group of data files associated with a given publication
3. The data file, which is deposited as bitstream.

4.2 Novel open, semantically-assisted peer review process

This Section describes the Services needed to enable the Pilot 2 related Functionalities.

4.2.1 Services

4.2.1.1 Relevant reviewers identification through social networks

- **Social media information aggregation Service**
Aggregates data from various sources.
- **Topic-based community identification Service**
Detects researchers who are highly interested in specific research topic(s) (e.g. based on their research interests), and forms a virtual community.
- **Researchers evaluation and choice of the most suitable ones Service**
Compares the scientific performance of different researchers on the basis of their openly accessible scientific information, as described in Section 4.5.1.7. Alternatively, evaluates according to social network metrics (number of papers uploaded, number of paper views, number of paper downloads) retrieved.

4.2.1.2 Relevant reviewers identification through literature

- **Relevant (co-)author identification Service**
Searches the available literature and retrieves relevant papers on the specific research topic(s). Searches within these results and identifies authors and co-authors: a) who appear in the paper's author list, b) whose work is being referenced in the paper or c) their name is mentioned in the paper text.
- **Multi-parametric results ranking Service**
Ranks the resulting (co-)author names according to various aspects, such as popularity, references count, etc.

4.2.1.3 Conflict of interest detection

- **Scientific collaboration graph building Service**
For each person from the pool of reviewers, identifies the direct and indirect collaborations between them and other researchers. Produces the network of researchers with whom an author has collaborated for the development of their research work. This network is based on co-authorship, affiliation, declared collaboration (extracted for example from the acknowledgements of their previous and current research work indicating provision of data, research hosting, comments provision etc) and research social networks connection.
Additionally, we may want to check the possibility of building the collaboration graph with information stemming from social media, as follows:
Social Media Information Aggregation Service: Aggregates data from various sources
Social Connections identification Service: Identifies other people the specific author is connected to.

Relationship Assessment Service: For each of the above connections (direct or indirect), identifies the existing interactions and evaluates the type and magnitude of the connection.

- **Conflict of interest evaluation Service**

If there are collaboration(s) identified between the candidate reviewer and one of the authors or co-authors of the paper, considers it to be a conflict of Interest, and proposes the removal of the specific reviewer from the pool of reviewers. Notifies the publisher/editor of the reason for removal.

4.2.1.4 Reviewer pool suggestion

- **Reviewer pool suggestion**

Suggests a list of possible reviewers with ranking, taking into account the relevant reviewers identification results and the conflict of interest evaluation. It aggregates the results from the Services 4.2.1.1, 4.2.1.2 and 4.2.1.3.

4.2.1.5 Reviewer selection and invitation

- **Reviewer selection**

Assigns a specific researcher for the review of a specific scientific work, which is stored in the Platform.

- **Reviewer invitation**

Submits an invitation via the Platform or email to a researcher that has been selected for the review of a specific scientific work. The invitation contents depend on whether the researcher is already registered in the Platform or not.

4.2.1.6 Relevant papers retrieval

- **Relevant papers retrieval Service**

Searches the available literature in order to identify and retrieve papers related with the paper being reviewed. Upon request, sorts the results based on relevance.

- **Semantic filtering Service**

Semantically filters the retrieved papers by using an ontology tree, in order to get more detailed results.

4.2.1.7 Missing references suggestion

- **Relevant papers retrieval Service**

(The same as above)

Searches the available literature in order to identify and retrieve papers related with the paper being reviewed. Upon request, sorts the results based on relevance. The relevance is adjusted by the fact whether a well matching paper is cited or not.

4.2.1.8 Review form submission

- **Review form storage Service**

Stores in the Platform the comments made by the reviewer, which are provided via an appropriate form.

4.2.1.9 Review commentary

- **Comment storage Service**

Stores the comments made by the users in the OpenScienceLink Platform.

4.2.1.10 Comment annotation

- **Text annotation Service**

Semantically annotates the users' comments according to specific ontology concepts

4.2.1.11 Comment evaluation

- **Researcher evaluation**

Presents the evaluation of the researcher who made the comment. This service is described in Section 4.5.2.3.

- **Comment ranking Service**

Depending on the commentator's evaluation, ranks their comments accordingly.

4.3 Data Mining for Biomedical and Clinical Research Trends Detection and Analysis

This Section describes the Services needed to enable the Pilot 3 related Functionalities.

4.3.1 Services

4.3.1.1 Trend Detection

- **Citation graph building**

From the available literature, create a graph representing the paper citations. Each node of the graph represents a paper and holds information such as its title, publishing date, author name, etc. Each edge represents a citation between two papers.

- **Citation graph clustering**

In the formed graph, identify the areas with higher edges density and create clusters. These clusters represent scientific fields with a lot of activity.

- **Term Co-occurrence Index Building**

On each of the above Clusters identify the most frequently appearing pairs or triples of terms.

OR for a specific concept (drug, disease, target), get a list of the 20 most frequently and the 20 least frequently co-occurring terms (specifically: drugs, diseases, targets. Not other ones).

4.3.1.2 Trend Analysis

- **Trend over Time Evaluation**

Perform the above process (Term Co-occurrence Index Building) repeatedly with a predefined frequency (e.g. once a year). Keep and process the results in order to evaluate the trend progress over time.

4.4 Data Mining for Proactive Formulation of Scientific Collaborations

This Section describes the Services needed to enable the Pilot 4 related Functionalities.

4.4.1 Services

4.4.1.1 Researchers connection identification through social networks

- **Social media information aggregation Service**

Aggregates data about communities/groups information and their members from various social media.

- **Research connections identification Service**

Uses social media communities/groups information in order to locate researchers with as many common research interests as possible, i.e. searches for overlaps of people (researchers) in scientific communities. As a result, locates a small set of researchers that maximizes the community overlap. Then, excludes existing collaborators.

4.4.1.2 Researchers connections identification through literature

- **Ontology-based data mining Service**

Searches for all the publications of a given author and the respective co-authors.

- **Research relationship evaluation Service**

Based on the publications information stemming from literature, builds the co-authorship graph for a given researcher. By analysing that graph and following the co-authorship path, identifies possible collaborators with who his existing collaborators have collaborated. Alternatively, by utilizing graph clustering techniques, locates clusters with great similarities (common co-authorship properties) and identifies possible collaborators. Then, removes existing collaborators.

4.4.1.3 Researchers collaboration filtering and proposition

- **Results filtering Service**

Provides the ability to the user to use specific filters in order to filter the collaboration propositions according to his/her preferences.

- **Collaboration proposition Service**

Proposes collaborations, taking into account the above results and the user-set filters.

4.4.1.4 Research groups connection identification through social networks and literature

- **Social media information aggregation Service**

Aggregates data about research communities/groups and their metadata (e.g. description, research topics) from various social media.

- **Ontology-based data mining Service**

Searches for the characteristics of a given author, including research interests.

- **Research group connections identification Service**

Utilizes a similarity checking mechanism, in order to semantically compare the research group topics and researcher's interests and identify groups that would possibly interest the researcher. Then, excludes groups that the researcher is already participating in.

4.4.1.5 Research groups collaboration filtering and proposition

- **Results filtering Service**

Provides the ability to the user to use specific filters in order to filter the collaboration propositions according to his/her preferences.

- **Collaboration proposition Service**

Proposes collaborations, taking into account the above results and the user-set filters.

4.5 Scientific field-aware, Productivity- and Impact-oriented Enhanced Research Evaluation

This Section describes the Services and models needed to enable the Pilot 5 related Functionalities.

4.5.1 Services

4.5.1.1 Research field dynamics evaluation

- **Citation graph building Service**

For a given scientific field, identifiable with specific keywords, creates, from the available literature, a graph representing the paper citations. Each node of the graph represents a paper and holds information such as its title, publishing date, author name, etc. Each edge represents a citation between two papers.

- **Citations over time observation Service**

Divides time into periods of equal length. In the citation graph counts the number of citations that occurred within each period of time and stores the results

- **Research field dynamics evaluation Service**

Based on specific models and depending on the behaviour of the citations over time, calculates the current maturity level (novel, rising, mature, dead) of the research domain.

- **Community lifecycle monitoring Service**

Evaluates the lifecycle of a scientific community interested in a specific research field. This may also provide an indication of the research field lifecycle.

4.5.1.2 Source of citations evaluation

- **Source of citations evaluation Service**

Extracts the authors of each citation and for each one of them it retrieves their evaluation. The higher evaluation is taken into consideration for determining the evaluation of the citations source in this iteration.

4.5.1.3 Citation dynamics evaluation

- **Citation dynamics evaluation Service**

Distributes the paper citations over time and calculates the number of citations per time period (e.g. over one year). Makes a projection (rising, falling, etc) of the citations over time, using a relevant model.

4.5.1.4 Quality and broadness of dissemination evaluation

- **Access journal type Service**

Checks if the paper was published in a Closed or Open Access journal.

- **Count number of websites or repositories Service**

Performs a Web search in order to locate the sources (websites or repositories) the paper is available through, and counts their number. For each source attempt to evaluate if the paper is freely available or not.

- **Count number of social networks Service**

Examines if and how many times the paper was published and available freely through social networks, by inspecting the publications in the author's profile.

4.5.1.5 Potential interest evaluation

- **Count number of downloads Service**

Accesses the information of paper downloads and views through social networks.

- **Readers evaluation Service**

Finds more details about the persons/researchers who have downloaded (read) the paper, as well as the downloaders' own evaluation, and evaluates them accordingly.

4.5.1.6 Overall research work evaluation

- **Research work evaluation Service**

By using predefined weights for each of the above evaluation metrics, and depending on the type of research work (paper or dataset) calculates an overall evaluation of the research work.

4.5.1.7 Researcher evaluation

- **Researcher evaluation Service**

Evaluates a researcher by taking into account the evaluation of all their research works, stemming from the Research work evaluation Service. Alternatively, uses special weighting on the individual research work evaluation metrics, in order to calculate the researcher evaluation.

4.5.1.8 Research group evaluation

- **Research group evaluation Service**

Evaluates a research group by taking into account the evaluation of the individual researchers who are members of the group, stemming from the Researcher evaluation Service. The evaluation of the group leader has a greater weight on the group's evaluation, in comparison to the evaluation of other group members.

4.5.1.9 Research community evaluation

- **Research community evaluation Service**

Evaluates a research community by taking into account the evaluation of the top researchers who are members of the community, stemming from the Researcher evaluation Service. All top community members are weighted equally.

4.5.1.10 Journal evaluation

- **Journal evaluation Service**

Evaluates a journal by taking into account the evaluation of all the research work that has been published in that journal, stemming from the Research work evaluation Service. Alternatively, uses special weighting on the individual research work evaluation metrics, in order to calculate the journal evaluation.

4.5.1.11 Institution evaluation per domain

- **Ontology-based data mining Service**

Searches the literature for all the researchers from a given institution, whose names appear as first author in a research work, and categorizes them per research area (domain).

- **Institution evaluation per domain Service**

Evaluates an institution per research area by taking into account the evaluation of the individual researchers within that institution working on the specific area, stemming from the Researcher evaluation Service.

4.5.1.12 Overall institution evaluation

- **Overall institution evaluation Service**

By combining the institution evaluation per domain, stemming from the Institution evaluation per domain Service, calculates the overall institution evaluation. Has the ability of using different weights for each per domain evaluation.

4.5.1.13 Country evaluation per domain

- **Ontology-based data mining Service**
Searches the literature for all the researchers from a given country, whose names appear as first author in a research work, and categorizes them per research area (domain).
- **Country evaluation per domain Service**
Evaluates a country per research area by taking into account the evaluation of all the researchers stemming from that country, stemming from the Researcher evaluation Service.

4.5.1.14 Overall country evaluation

- **Overall institution evaluation Service**
By combining the country evaluation per domain, stemming from the Country evaluation per domain Service, calculates the overall country evaluation. Has the ability of using different weights for each per domain evaluation.

4.5.2 Models and indicators

This section presents models and indicators which are relevant to the evaluation of the above entities and should be taken into account before designing the corresponding services.

4.5.2.1 Research work evaluation

A very relevant model for research work evaluation is the one introduced by the Public Library of Science (PLOS)¹, which allows research work to be judged on their individual merits, rather than on the basis of the journal in which they were published. It makes use of Article Level Metrics (ALMs), which are a variety of scores and other information attached to each article. These fall into the following categories:

- **Viewed**
Includes recorded online usage activity data for each article. Robot activity is excluded.
- **Cited**
Includes citation data on each article as computed by several third-party citation measuring services. To attain the most comprehensive view of citations, all services are consulted and results are de-duplicated.
- **Saved**
Includes activity data from common online reference management services which allow individual researchers and groups to bookmark papers, collate references, and share sources with their community.
- **Discussed**
Includes data from tools from which to share and discuss research work, in order to offer a more comprehensive view of the article's impact after publication. Tracks comments about an entire article or specific parts thereof, as well as other online article dissemination activity mechanisms, such as posts in blogs and social networks (e.g. Twitter and Facebook).
- **Recommended**
Includes data (mainly review scores) from sources that capture formal endorsements of research articles, such as online recommendation channels.

¹ Public Library of Science (PLOS) is a non-profit publisher, membership, and advocacy organization with a mission to accelerate progress in science and medicine by leading a transformation in research communication. <http://www.plos.org/>

4.5.2.2 Journal evaluation

Presently there are many different models and indicator types for the evaluation of journal impact. Indicators based solely on the number of citations, measure the popularity, while indicators that include different weighting in the citation analysis can measure the prestige of the journal.

- **Journal Impact Factor (IF)**

The IF is the oldest, most renowned and an extensively used index for measuring the quality of a journal. It is a form of citation analysis that offers an indication of how influential a journal is. The calculation of the IF is based solely on the number of citations, and neither the quality of citations nor the inclusion of self-citations are considered. A journal's IF, which is calculated on the basis of the citation statistics, can only be calculated years after publication of the journal, because those citations only appear in articles which are published later. Journal specific IF indicators are always some sort of average of the Impact Factors of the articles published in them. The variations in IF scores according to discipline are significant, and thus comparing journals from different disciplines is not sensible.

- **Journal Immediacy Index**

The Journal Immediacy index represents the average number of times an article is cited in the year it was published, thus it comprises a measure of speed. The Immediacy Index is particularly useful for comparing journals specializing in cutting-edge research, but it can also be used to evaluate newly added journals, which do not currently have Impact Factor values.

Most Open Access (OA) journals belong to the lower scoring half of the evaluated journal ranking lists in their subject area. OA journals, however, are situated higher in the ranks in the Immediacy Index scores than when evaluated with IF scores. A high Immediacy Index score means that the article is cited soon after it has been published, and that its content is topical at the time of publication.

- **Eigenfactor (EF)**

The Eigenfactor (EF) is an overall rating of the importance of a scientific journal whereby all articles published in a journal during a year are taken into consideration when making the calculation. What counts in the Eigenfactor is the size of the journal, or how many articles a journal publishes. The Eigenfactor is scaled so that all the publications that feature in the Journal Citation Reports (JCR) have a summed EF of one hundred. The calculation of the Eigenfactor for journals is based on the PageRank algorithm, which enables determining the value of the journal that is citing the article, and citations are given different weight according to how high the EF score of the journal is. The EF score is calculated by counting all citations from five years after the article was published, and not just citations from a specific category of journals. In addition to scientific journals, the EF takes into consideration citations in newspaper articles, doctoral theses and books etc. It thus helps getting a fairer estimation for journals situated in the middle ground between scientific disciplines. It also takes into account different citation norms which enhances the comparability of journals from different scientific areas.

- **Article Influence**

The Article Influence score is a journal specific Eigenfactor score. It is the journal's Eigenfactor Score divided by the fraction of articles published by the journal. That fraction is normalized so that the sum total of articles from all journals is 1. It tells the average impact of an article published in a journal from a five year period after the publication of the article.

- **H-index**

The h-index is an alternative to the IF when evaluating the impact of journals based on citation information. It gives a rough estimate on the productivity of a journal over a long time period. The good qualities of the h-index, compared with IF, are its stability and its

balanced combining of quantity and quality, which reduces the over-valuing of journals that publish review articles. Because the h-index is not based on the average number of article citations, it is not influenced by the fact that a few articles might have massive amounts of citations. However, the amount of articles a journal has published will affect the h-index score. The h-index is not calculated for one specific fixed time period like the IF, but it can be calculated for varying time periods, e.g. for one or five years. With the h-index it is still preferable to compare journals from within the same discipline.

- **SCImago Journal Rank (SJR)**

The SJR indicator is calculated based on the referencing data of the Scopus database. When calculating the indicator, weighting is given to the citations the journals have received according to how high the SJR score of the citing journal is.

- **Source Normalized Impact per Paper (SNIP)**

SNIP measures a journal's contextual citation impact, taking into account characteristics of its subject field, especially the frequency at which authors cite other papers in their reference lists, the rapidity of maturing of citation impact, and the extent to which the database used for the assessment covers the field's literature.

4.5.2.3 Researcher evaluation

When evaluating the level of researchers and/or research groups on the basis of publishing activity, the most common measures are: the amount of publications, the amount citations and the quality of the publications used as publishing channels for the research.

- **Productivity**

Is calculated by counting the amount of publications of a researcher in the most important databases to the discipline. It is very important that as many databases as possible be used. Finding monographs is often problematic because often they cannot be found in most databases.

- **Citations: H-index**

The h-index is an indicator used in academic publishing to measure the publishing activity of researchers and their scientific influence. The higher the h-index score the more the researcher in question has received citations from their publications.

- **Citations: G-index**

The g-index is an improvement of the h-index. The biggest fault of the h-index is that it does not take into account the real number of citations of each of the most cited publications, which can be considerable more than h.

- **Background and networks**

When establishing the academic level or influence of a researcher, it is an essential part of the evaluation to examine the researcher's background and his/her networking with other researchers in the same field. The organisations or institutions where he/she has worked also tell lot about the networking of the researcher. In addition, we can gain insights into networking by finding out with whom the researcher has co-authored publications.

- **Quality of publications**

There are several different measures for evaluating the quality of publications. The most common is the Impact Factor.

4.5.2.4 Research group evaluation

Evaluating research groups, especially in Biosciences and Medicine, is more important than evaluating individual researchers, because almost all research is conducted in groups. When evaluating research groups, the measures are the same as when evaluating individual

researchers: publishing productivity, the amount of citations from the publications and the quality of the publication that cites the research.

Evaluating research groups differs from evaluating individual researchers because the indicator scores are not available directly from the databases. With indicators that are based on citation information, the citations of researchers in the group from fellow group members should be excluded. For example, calculating the h-index for a research group differs from the evaluation of individual researchers, where the index score is often calculated for the whole of the researcher's life (sometimes from a more restricted period). However, the h-index of research groups is often calculated from the last three to five years, so the quality assessment of the latest research produced by the groups can be as accurate as possible. Research groups can also only be compared to groups in the same discipline, because the group sizes and numbers of publications vary in different disciplines.

4.5.2.5 Institution evaluation

Various models are used internationally for ranking research institutes and universities. The most relevant models are as follows.

- **QS World University Rankings²**

The QS World University Rankings make use of a number of metrics, such as academic peer review, employer review, percentage of international faculty and students, faculty-student ratio. This model also takes into account the citations per faculty metric, as follows. The citations per teaching staff member are calculated over a five year period and the number of citations is proportioned to the size of the teaching staff. The sources used for the citation data are the Web of Science, Scopus and Google Scholar databases.

- **University Ranking by Academic Performance (URAP)³**

The URAP ranking is being published by the Middle East Technical University. Its main objective is to rank the world universities based on academic performances which are determined by quality and quantity of scholarly publications. The ranking indicators include the total number of published articles, the total number of citations, the total document count (including conference papers, reviews, letters, discussions and scripts in addition to journal articles), the scientific impact which is derived by aggregating the impact factors of journals in which a university published articles, the received citation quality which is based on the impact factors of journals where the citing articles are published, and the international collaboration data which is based on the total number of publications made in collaboration with foreign universities.

- **CWTS Leiden Ranking⁴**

The CWTS Leiden Ranking aims to provide measurements of the scientific impact of universities and organizations directly related to universities, and their involvement in scientific collaboration, by using a sophisticated model comprising of a number of bibliometric indicators. The indicators used (by default size-independent, but also size-dependent on request) are mainly impact indicators (e.g. Mean Citation Score, Mean Normalized Citation Score, Proportion of top 10% publications) and collaboration indicators (proportion of inter-institutional collaborative publications, proportion of international collaborative publications, proportion of collaborative publications with industry, mean geographical collaboration distance). The CWTS Leiden data collection methodology is also quite interesting.

² <http://www.topuniversities.com/university-rankings/world-university-rankings/home>

³ <http://www.urapcenter.org/>

⁴ <http://www.leidenranking.com/>

4.5.2.6 Country specific evaluation

In order to assess the quality and productivity according to country, the most commonly used measures are the numbers of publications and impact indicators that are calculated according to citations. Measuring national productivity with bibliometric indicators has been happening for a long time. Many countries publish national science reports in which the state of the country's research and research infrastructure are analysed using bibliometric indicators.

The Academy of Finland state that the interpretation of bibliometric results must always take account of the limitations associated with international databases and bibliometric methods of analysis. According to their model, all publications in which at least one author has a Finnish address are classified as Finnish publications. Publication numbers are fractionalized between the countries of the contributing authors. Numbers of citations are also fractionalized between the countries contributing to the publication. Self-citations have been removed. The number of citations received has been computed for the year of publication and the following two years.

Different countries are compared according to the amount of publications and the amount of citations publications have received. The scientific activity of countries varies and can be roughly correlated with the economic status of the country. In order to eliminate the differences in the amounts of publications between different countries, the amount of citations per publication indicator can be used.

4.6 Common Services

This section describes the Services which are commonly used by more than one Pilots.

4.6.1 Services

4.6.1.1 User registration and profile

- **User profile creation Service**
Allows user to create a personal profile in the Platform, by providing relevant data and metadata, such as their personal information, scientific fields of interest, scientific work, etc.
- **User profile accessing Service**
Accesses the user profile in order to retrieve various user-specific data.
- **User registration Service**
Stores the user profile and other parameters in the Platform and creates a user account which is accessible by using a unique username and a password.

4.6.1.2 User log-in

- **User authentication Service**
Confirms the validity of a user's identity by checking the correctness of their credentials (username and password) towards the corresponding information stored within the Platform. If the credentials are valid, the user can log in the Platform.
- **User authorization Service**
Enforces the Platform's access policy by allowing or denying user access to specific parts and features of the Platform, depending on their privileges.

4.6.1.3 Research work semantic linking

- **Paper semantic linking Service**
By using the paper's title and abstract links the dataset with related literature, topics, and authors of the life sciences domains, by means of semantic enabled technologies, such as text annotation with ontology concepts.

- **Dataset semantic linking Service**

By using the dataset's metadata links the dataset with related literature, topics, and authors of the life sciences domains, by means of semantic enabled technologies, such as text annotation with ontology concepts.

4.6.1.4 Semantic search with filtering and ranking

- **Semantic search Service**

By using the paper's title and abstract links the dataset with related literature, topics, and authors of the life sciences domains, by means of semantic enabled technologies, such as text annotation with ontology concepts.

- **Results filtering Service**

Refines the search results by semantically filtering them according to user-selected concepts stemming from the underlying ontologies.

- **Multi-parametric results ranking Service**

Ranks the (filtered) search results according to a number of predefined criteria. Changing the criteria reorders the results list.

4.7 Summary of Services

This section summarizes all of the Services per Pilot.

Pilot	Functionality	S-ID	OpenScienceLink Services
Research Dynamics-aware Open Access Data Journals Development	Trend detection	S3.1-S3.3	Trend detection 4.3.1.1
	Trend analysis	S3.4	Trend analysis 4.3.1.2
	Relevant datasets detection	S1.1	Relevant datasets detection
	Dataset with metadata submission	S1.2	Dataset storage
		S1.3	Metadata storage
	Relevant reviewers identification	S2.1-S2.8	Relevant reviewers identification 4.2.1.1-4.2.15
	Dataset reviewing	S2.11, S0.9	Dataset reviewing 4.2.1.6-4.2.1.8
	Comments publication	S.12	Comments publication
	Review commentary	S.13	Review commentary
	Dataset selection	S1.4	Dataset retrieval
		S1.5	Dataset selection
	Dataset publication	S1.6	Dataset publication
		S1.7	Dataset download

	Dataset assessment and evaluation	S5.1-S5.12	Assessment and Evaluation
Novel open, semantically-assisted peer review process	Relevant reviewers identification through author networks	S2.1	Social media information aggregation
		S2.2	Topic-based community identification
		S2.3	Researchers evaluation and choice of the most suitable ones
	Relevant reviewers identification through literature	S2.4	Relevant (co-)author identification
		S2.5	Multi-parametric results ranking
	Conflict of interest detection	S2.6	Scientific collaboration graph building
		S2.7	Conflict of interest evaluation
	Reviewer pool suggestion	S2.8	Reviewer pool suggestion
	Reviewer suggestion and invitation	S2.9	Reviewer selection
		S2.10	Reviewer invitation
	Relevant papers retrieval	S2.11	Relevant papers retrieval
		S0.9	Results filtering
	Missing references suggestion	S2.11	Relevant papers retrieval
	Review form submission	S2.12	Review form storage
	Review commentary	S2.13	Comment storage
	Comment annotation	S2.14	Text annotation
Comment evaluation	S5.13	Researcher evaluation	
	S2.15	Comment ranking	
Data Mining for Biomedical and Clinical Research Trends Detection	Trend detection	S3.1	Citation graph building
		S3.2	Citation graph clustering

and Analysis		S3.3	Term Co-occurrence Index Building
	Trend analysis	S3.4	Trend over Time Evaluation
Data Mining for Proactive Formulation of Scientific Collaborations	Researchers connection identification through social networks	S4.1	Social media information aggregation
		S4.2	Research connections identification
	Researchers connections identification through literature	S4.3	Ontology-based data mining
		S4.4	Research relationship evaluation
	Researchers collaboration filtering and proposition	S0.9	Results filtering
		S4.6	Researchers collaboration proposition
	Research groups identification through social networks and literature	S4.1	Social media information aggregation
		S4.3	Ontology-based data mining
		S4.7	Research group connections identification
	Research groups collaboration filtering and proposition	S0.9	Results filtering
		S4.8	Research group collaboration proposition
	Scientific field-aware, Productivity- and Impact-oriented Enhanced Research Evaluation	Research field dynamics evaluation	S5.1
S5.2			Citations over time observation
S5.3			Research field dynamics evaluation
S5.4			Community lifecycle monitoring
Source of citations evaluation		S5.5	Source of citations evaluation
Citation dynamics evaluation		S5.6	Citation dynamics evaluation
Quality and broadness of dissemination evaluation		S5.7	Access journal type
		S5.8	Count number of websites or repositories
		S5.9	Count number of social networks

	Potential interest evaluation	S5.10	Count number of downloads
		S5.11	Readers evaluation
	Overall research work evaluation	S5.12	Research work evaluation
	Researcher evaluation	S5.13	Researcher evaluation
	Research group evaluation	S5.14	Research group evaluation
	Research community evaluation	S5.15	Research community evaluation
	Journal evaluation	S5.16	Journal evaluation
	Institution evaluation per domain	S5.17	Ontology-based data mining
		S5.18	Institution evaluation per domain
	Overall institution evaluation	S5.19	Overall institution evaluation
	Country evaluation per domain	S5.17	Ontology-based data mining
		S5.20	Country evaluation per domain
	Overall country evaluation	S5.21	Overall country evaluation
	All	User registration and profile	S0.1
S0.2			User registration
S0.3			User profile accessing
User log-in		S0.4	User authentication
		S0.5	User authorization
Research work semantic linking		S0.6	Paper semantic linking
		S0.7	Dataset semantic linking
Semantic search with filtering and ranking		S0.8	Semantic search
		S0.9	Results filtering
		S0.10	Multi-parametric results ranking

Table 7: Summary of OpenScienceLink Platform Services

5 Conclusion

In this deliverable we presented the OpenScienceLink Pilot Scenarios, which are derived directly from the Pilot Requirements identified in D2.1, representing the OpenScienceLink end user needs. On the basis of the Scenarios we elicited the Platform Functionalities, which we further analyzed into Services. The Services were numbered and should serve as a basis for the specification of the Platform's architecture and the individual components' specifications.

In summary, for the five Pilots we identified 40 unique Functionalities, which were analyzed into 65 unique Services. 4 common Functionalities and 10 common Services were identified among them.

The identified Functionalities and Services correspond to the current version of Pilot Requirements and user needs. In the following iterations, we plan to enrich the identified mechanisms, models and data sources according to the newly collected and analysed requirements.