

CIP Competitiveness and Innovation Framework Programme

CIP-Pilot Actions, 2007-2013

CIP-ICT-PSP-2012-6

Project CIP-Pilot 325101 / OpenScienceLink

Deliverable **D8.3**Distribution **Public**



http://opensciencelink.eu

Best Practices, Blueprints and Policy Guidelines for Open Access to Scientific Information

Laurens Naudts, Freyja van den Boom, Todor Tagarev, Petya Tagareva, George Tsatsaronis, Michael Schroeder

Status: Final (Version 1.0)

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://projects.biotec.tu-dresden.de/OpenScienceLink/

Project start February 2013
Project duration 36 months

EC Project Officer

Deliverable D8.3

Deliverable type Report
Distribution level Public

Deliverable Number

Deliverable title Report Title

Contractual date of delivery M36

Actual date of delivery

Relevant Task(s) WP8 (WP3)
Partner Responsible PROCON
Other contributors KU Leuven

Number of pages 70

Author(s) Laurens Naudts, Freyja Van Den Boom

Internal Reviewers TUD, NTUA KU Leuven, Status & version Version 1 (Final Version)

Keywords Open Access, Legal framework, Intellectual Property

Rights, Data Protection, Privacy, Licensing, Best

Practices, Guidelines, Blueprints.





Abstract

Open access data publishing is the act of making data freely available online so that they can be downloaded, analysed, re-used and cited by other people than the creators of the data. This can be achieved in various ways. In the broadest sense, any upload of a dataset onto a freely accessible website could be regarded as "data publishing".

This deliverable serves as an open-access guide and is not merely addressed at users who wish to make use of the OpenScienceLink platform to upload their data. It outlines best practices for the publication and sharing of research data in the context of OpenScienceLink and beyond. In addition, the final section of this deliverable aims to provide a list of policy guidelines and best practices in an effort to increase the potential of open access initiatives as a whole. These guidelines are aimed at open access initiatives such as OSL, as well as open access policy makers.

There are several issues to be considered during the process of making scientific information freely available. This guide serves the following purposes:

- Promote Open Access by detailing relevant aspects and definitions related to Open Access Publishing.
- Provide additional insight to those unfamiliar with open access to scientific information initiatives.
- Provide an overview of potential legal caveats from the researcher's and platform's perspective
- Formulate guidelines and best practices for Open Access initiatives and Open Access Policy makers in an effort to encourage the further development and improvement of open access to scientific information.

Disclaimer

This Guide does not override institutional policies on data management or publication. Researchers must ensure their decisions about data publication align with operate within the policies of their institution and/or funding arrangement. The Guide indicates when additional information needs to be sought and when to check institutional policies.¹

¹ The Registry of Open Access Repository Mandates and Policies (ROARMAP) is a searchable database providing links to the open access policies of *inter alia* universities: http://roarmap.eprints.org/





Table of Contents

A	bstrac	t	iii
T	able of	Contents	iv
1	Int	roduction	8
	1.1	What is the OpenScienceLink platform?	8
2	Ope	en Access to Scientific Information: F.A.Q.s	10
	2.1	What is Scientific Information?	10
2.2 2.3		What is "Research Data"?	10
		What is metadata?	10
	2.4	Why add metadata?	11
	2.5	What is a Database?	11
	2.6	What is a "Data Paper"?	11
	2.7	What are the benefits of publishing data/data paper?	12
	2.8	Can a Data Paper be indexed and cited?	13
	2.9	What is Open Access?	13
	Wh	at does 'open' mean?	13
	Wh	at does 'Open Access' mean?	13
2.10		Why Open Access?	14
	2.11	How to Provide Open Access? (Green and Gold Open Access)?	15
	2.12	Where should data be published?	15
	2.13	How should data be shared/structured?	16
3	Spe	ecific OpenSciencelink Open Access Policy - FAQs	18
	3.1	How do I submit a data paper?	19
	3.2	How should I cite my data?	19





	3.3	What are the criteria for a repository to be accepted by OSL?	19
	3.4	Is there any specific format I need to follow to submit a dataset?	19
	3.5	Wat Metadata Model does the OpenScienceLink platform use?	20
	3.6	From which locations will my dataset become available for download?	20
	3.7 the r	How does the review process differ from an ordinary research article review? What eviewers look at when it comes to a dataset article?	
4	Ма 23	aking Scientific Research and Data Open Access – Intellectual Property Rights Guidel	lines
	4.1	What is Copyright?	23
	4.2	What rights are granted to the holder of copyright?	23
	4.1	.1. Do Bare Facts – Research Data enjoy copyright protection?	24
	4.1	.2. Do Metadata enjoy copyright protection?	24
	4.1	.3. Do datasets enjoy copyright protection?	24
	4.1	.4. Does a database enjoy copyright or any other protection?	25
	4.1	5. What are the exclusive rights of the original database's creator?	25
	4.1	.4. Who owns copyright?	25
		1) Joint ownership	26
	:	2) Works created in the course of employment	26
	4.3	What is the Sui Generis Database Right?	26
	4.4	Can I re-use other people's material?	27
	4.5	Can I copy and publish substantial parts of databases other than OpenScienceLink?	28
	4.6	Can I link my data to other research publications?	28
	4.7	Is there a copyright/sui generis exception for scientific research?	29
5	Ma	aking Scientific Research and Data Open Access – Licensing Guidelines	30
	5.1	Introduction: License and agreements	30
	5.2	What are publishing agreements?	30





	5.2.1.	How to deal with a Publisher/Publisher Licenses?	31
	5.2.2.	What is an embargo period?	32
	5.2.3.	Under what license will the Open Science Link Platform publish my data?	32
	5.2.4.	Why should open access works be accompanied by a license?	32
	5.2.5.	Why the choice for a CC.04 BY-NC-SA License?	33
	5.2.6.	What does the CC.04 BY-NC-SA License entail?	33
	1)	BY – Attribution:	34
	2)	NC – Non Commercial:	34
	3)	SA – Share Alike:	34
	5.2.7.	What are the consequences of the Share Alike clause?	34
	Wh	at is Adapted Material?	35
	5.2.8.	What are the conditions to Share Alike?	35
	5.2.9.	Does the Creative Commons license apply to the Sui Generis Database Right?	35
	5.3 H	low can I ensure License compatibility with the OSL License?	36
	More	information on intellectual property rights	37
6	Makin	ng Scientific Research and Data Open Access – Data Protection Guidelines	38
	6.1 P	ersonal Data and Data Protection	38
	6.2 V	When should I take the Data Protection Framework into account?	38
	6.3 V	Vhat is data processing?	39
	6.4 V	Vhat is personal data?	39
	6.5 Is	s biomedical data personal data?	39
		oes anonymized data constitute personal data?	
	6.6.1.	Pseudonymization	40
	6.6.2.	Does the OSL Platform require anonymization?	40
	6.6.3.	What is anonymization?	
	6.6.4.	How can research data be anonymized?	





	6.7	What are my other responsibilities as a researcher?4	3
	6.8	Can I re-use the data on the OSL platform (data protection)?4	3
	6.9	Does the OpenScienceLink Platform Guarantee the Protection of my personal data?4	4
7	Gui	delines for Open-Access Initiatives and Open-Access Policy Makers4	5
	7.1 acces	Promote and encourage the inclusion of effective review processes based on opes to scientific information4	
	7.2	Further facilitate research collaboration4	6
	7.3	Further promote the development and integration of novel trend analysis methods4	7
	7.4	Further encourage research evaluation processes4	8
	7.5	Further reduce the costs of open – access data journal development5	0
	7.6	Take into account the legal requirements - Legal Checklist for OA Platforms5	2
	Dat	ta Protection Requirements5	2
	Dat	ta Protection Requirements – The GDPR Framework5	5
		vacy and Data Protection guidelines on the open access to scientific information wit	
	Int	ellectual Property Requirements6	1
	7.7	Further guidelines for Open Access Policy Makers6	4
	Tal	ke into account open access policy research6	4
	Inc	rease the effectiveness of OA policies6	4
8	Ref	Gerences6	6
9	Use	efull links6	8





1 Introduction

Open access data publishing is the act of making data freely available online so that they can be downloaded, analysed, re-used and cited by other people than the creators of the data. This can be achieved in various ways. In the broadest sense, any upload of a dataset onto a freely accessible website could be regarded as "data publishing".

The OpenScienceLink project aims to provide a universal well-structured repository of scientific and research data – currently focusing on biomedical and clinical research - for experimentation and benchmarking of pertinent research works in a given thematic area. Moreover, the OpenScienceLink web-based platform will enable and provide the publishing and sharing of publications and experimental datasets, as well as their link with researchers and scholars.

The OpenScienceLink platform also aims to contribute to the overall open-access environment however, by providing open-access stakeholders and those willing to embark on an open-access quest with the necessary information for doing so successfully. This guide is not merely addressed at users who wish to make use of the OpenScienceLink platform to upload their data. It outlines best practices for the publication and sharing of research data in the context of OpenScienceLink and beyond. In addition, section 7 of this deliverable aims to provide a list of policy guidelines and best practices in an effort to increase the potential of open access initiatives. These guidelines are aimed at open access initiatives such as ourselves, as well as open access policy makers.

There are several issues to be considered during the process of making scientific information freely available. This guide serves the following purposes:

- Promote Open Access by detailing relevant aspects and definitions related to Open Access Publishing.
- Provide additional insight to those unfamiliar with open access to scientific information initiatives.
- Provide an overview of potential legal caveats from both the researcher's and platform's perspective.
- Formulate guidelines and best practices for Open Access initiatives and Open Access Policy makers in an effort to encourage the further development and improvement of open access to scientific information.

1.1 What is the OpenScienceLink platform?

OpenScienceLink is an EU-funded project aimed at introducing and piloting a holistic approach to the publication, sharing, linking, review and evaluation of research results, based on the open access to scientific information. OpenScienceLink will pilot a range of novel services that could alleviate the lack of structured data journals and associated data models, the weaknesses of the review process, the poor linking of scientific information, as well as the limitations of current research evaluation metrics and indicators.

Five pilot services will be integrated and piloted in particular:

 Data journals development based on semantically-enabled research dynamics detection,

- A novel open, semantically-assisted peer review process,
- Services for detection and analysis of research trends,
- Services for dynamic researchers' collaboration based on non-declared, semantically-inferred relationships, and,
- A set of scientific field-aware, productivity- and impact-oriented enhanced research evaluation services.

More information on the OpenScienceLink plaform can be found following this link: http://opensciencelink.eu/objectives/

For detailed information on the 5 Pilot Scenarios from an end-user's perspective, please follow this link:

http://opensciencelink.eu/wp-

 $content/uploads/2015/02/OSL_D3_1_DetailedSpecification of OpenScienceLinkPilotScenarios and Services_REVISION_1.pdf$





2 Open Access to Scientific Information: F.A.Q.s

In the following section we aim to provide the answers to frequently asked questions concerning open access scientific research. Although the considerations relate mostly to questions asked by scientific researchers, they are also informative for research institutions and publishers wishing to enter the open access sphere. These questions mostly pertain to the open access on scientific information in general and could therefore be used as a template for those who wish to inform their end-users concerning open access policies.

2.1 What is Scientific Information?

In general, the term 'scientific' may refer to all academic disciplines, not only hard sciences. According to the European Commission, 'scientific information' in the context of research and innovation may fall under two broad categories²:

- Peer-reviewed scientific research articles (published in scholarly journals) or
- Research data (data underlying publications, curated data and/or raw data).

2.2 What is "Research Data"?

Research data constitutes all data from an experiment, study or measurement, including the metadata and processing data.³ Under the EU Horizon 2020 track 'research data' refers to information, in particular facts or numbers, collected to be examined and considered and as a basis for reasoning, discussion, or calculation. In a research context, examples of data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images. The focus is on research data that is available in digital form.

2.3 What is metadata?

Metadata is data about your data, i.e. a description of your data. A common example of metadata is a library catalogue record. The London School of Economics notes that metadata is usually a formally agreed set of standards often with controlled fields and vocabularies, enabling the facilitation of data preservation, discovery and citation.⁴ The metadata that a researcher needs to provide may depend upon the chosen repository or archive for data publication.

⁴ http://www.lse.ac.uk/library/usingTheLibrary/academicSupport/RDM/doing/metadata.aspx





² European Commission, *Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020* (30 October 2015), p. 2.

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

³ Sarah Hugelier, 'Publishing Open-Access Biomedical Data: Legal Challenges', *Biomed Data J. 2015; 1(1):* 43-51

2.4 Why add metadata?

Adding structured metadata to your publications provides many advantages. Smart metadata can facilitate the discovery of your work as well as allow others to comprehend and evaluate your data on a basic level without requiring access to the datasets themselves.

The purpose of OpenScienceLink - as well as most other open access initiatives - and the basis for its selection of data and services is the creation of a repository of well-structured and semantically linked datasets. The role of metadata herein is crucial. Therefore metadata descriptions should be based on existing and established domain ontologies.

2.5 What is a Database?

In general terms, a database can be defined as a usually large collection of data organized especially for rapid search and retrieval, for instance by a computer.⁵ The Oxford Dictionary defines a database as a structured set of data held in a computer, especially one that is accessible in various ways.⁶

On the European level, a database has been legally defined as a collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means. A database should be understood to include *inter alia* scientific works, collections of works or collections of other material such as text, images, numbers, facts, and data.

2.6 What is a "Data Paper"?

A data paper is a publication that is designed to make other researchers aware of data that is of potential use to them. A data paper includes a description on the methods used to create the data set, a description of the structure of the data set and an elaboration on the dataset's reuse potential. A link to the data paper's and data set's location in a repository should also be included. The important to note that a data paper does not replace a research article, but rather supplements it. When mentioning the data behind a study, a research paper should reference the data paper for further details. Similarly, the data paper should contain references to the research papers associated with the dataset.

In general, the purpose of a data paper is to:

⁸ http://openhealthdata.metajnl.com/about/#q2





⁵ Merriam-Webster Dictionary

⁶ Oxford Advanced Learner's Dictionaty

⁷ Hrynaszkiewicz I, Norton ML, Vickers AJ, Altman DG. Preparing raw clinical data for publication: guidance for journal editors, authors, and peer reviewers. BMJ. 2010

- Describe the data in a structured human-readable form, and
- Bring the existence of the data to the attention of the scholarly community
- Provide a citable journal publication which allows the data publisher to be credited.

The description should include several important elements documenting how the dataset was collected, who owns the data, details of which software was used to create the data, how the data can be viewed, etc.

2.7 What are the benefits of publishing data/data paper?

Data publishing has become increasingly important and has had an effect on the policies of funding frameworks and organizations. Moreover, shifting towards the free online access of the research results of publicly-funded research is also a core strategy of the European Commission. Or the European Commission.

Furthermore, data papers mean that data you have released can be cited and that those citations can be tracked. This is not only an indirect measure of impact and therefore important for career progression, but it can also help you understand who is using your data. In turn, this can lead to new collaborations.

It is now widely recognized that making research results more accessible to all societal actors contributes to better and more efficient science, and to innovation in the public and private sectors. The European Commission encourages all EU Member States to put publicly-funded research results in the public sphere in order to strengthen science and the knowledge-based economy.¹¹

Other incentives for authors and institutions to publish data include:

- The idea that data produced using public funds should be openly published and made available for inspection, interpretation and re-use by third parties.
- Data-collecting efforts and associated costs will be reduced by avoiding duplication of work.
- Open data increases the potential for interdisciplinary research, and for re-use in new contexts not envisaged by the data creator; Open data increases transparency and the overall quality of science; published datasets can be re-analyzed and verified by others.
- Published data can be cited and re-used in the future, either alone or in association with other data.
- Published data can be indexed, made discoverable and searchable.

 $^{^{11}}$ RECOMMENDATIONS COMMISSION RECOMMENDATION of 17 July 2012 on access to and preservation of scientific information (2012/417/EU





⁹ See for example Riding the Wave: How Europe Can Gain From the Rising Tide of Scientific Data report submitted to the European Commission in October 2010

¹⁰ http://ec.europa.eu/research/openscience/index.cfm?pg=openaccess

Data creators, and their institutions and funding agencies, can be credited for their work of data creation and publication through the conventional channels of scholarly citation. Priority and authorship is achieved in the same way as with a publication of a research paper.

2.8 Can a Data Paper be indexed and cited?

Data Papers can be indexed and cited like any other research article, thus bringing registration of priority, a permanent publication record, recognition and academic credit to the data creators.

2.9 What is Open Access?

The European Commission has defined open access as the "practice of providing online access to scientific information that is free of charge to the end-user and that is re-usable"¹². Scientific information usually refers to peer-reviewed scientific research articles and scientific research data. Even though Open Access policies exist the question whether scientific information should be made open, arises only after the researcher has decided to publish his or her results.¹³

What does 'open' mean?

The Open Knowledge Foundation has defined open content as "a piece of content or data that is free to use, reuse, and redistribute by anyone — subject only, at most, to the requirement to attribute and share-alike."

What does 'Open Access' mean?

The goal of OpenScienceLink is to publish research data and datasets under conditions of Open Access.¹⁴ Open Access entails "the free availability of scientific literature on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from

<https://www.iprhelpdesk.eu/sites/default/files/newsdocuments/Open_Access_in_H2020.pdf>
14 Under Horizon 2020 'Research data' refers to information, in particular facts or numbers, collected to be examined and considered and as a basis for reasoning, discussion, or calculation. In a research context, examples of data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images. The focus is on research data that is available in digital form.





¹² European Commission, *Guidelines on Open Access to Scientific Publications and Research Data in Horizon* 2020 (30 October 2015), p. 2.

 $http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf\\$

¹³ IPR Help Desk, 'Open Access to Publications and data in Horizon 2020: Frequently Asked Questions' (May 2014);

gaining access to the internet itself. Relevant in the context of research data is that 'scientific' refers to all scholarly disciplines. ¹⁵

The only constraint on reproduction and distribution, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited."¹⁶

Open Access according to the Budapest furthermore requires that:

- (1) The author and right holder grant(s) to all users a free, irrevocable, worldwide, right of access to, and a license to copy, use, distribute, transmit and display the work publicly and to make and distribute derivative works, in any digital medium for any responsible purpose, subject to proper attribution of authorship, as well as the right to make small numbers of printed copies for their personal use.
- (2) A complete version of the work and all supplemental materials, including a copy of the permission as stated above, in an appropriate standard electronic format is deposited (and thus published) in at least one online repository using suitable technical standards (such as the Open Archive definitions) that is supported and maintained by an academic institution, scholarly society, government agency, or other well-established organization that seeks to enable open access, unrestricted distribution, inter-operability, and long-term archiving.

2.10 Why Open Access?

The rationale given by many policy makers and scholars for pushing open access conditions is that 'opening up the data allows for new knowledge to be discovered through comparative studies, data mining and so on. It allows greater scrutiny of how research conclusions have been reached, potentially driving up research quality."¹⁷ This rationale is also the idea behind the OpenScienceLink project: "to foster the widest access and re-use of scientific publications and data".

When done correctly, the making of scientific information freely available online, has the potential to benefit all stakeholders within the Open Access value chain, such as researchers, research organisations, publishers, editors and funding authorities.

Open data can be reused by the wider public for a range of purposes including teaching, journalism and citizen science projects. Making research outputs available for others to work with and build upon is part of the social contract of academia. Moreover, some studies claim

¹⁷ For example: Ball, A., (2012) and Guibault, L (2012), (166), and Communication from the Commission towards better access to scientific information, COM (2012) 401 final (1)





 $^{^{15}}$ Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020 Version 1.0 11 December 2013

¹⁶ Budapest Open Access Initiative (2002). Retrieved at http://www.budapestopenaccessinitiative.org/read

that open access publications are more likely to be cited than subscription publications. ¹⁸ This Open Access citation advantage should be considered by researchers who wish to increase the impact of their works.

2.11 How to Provide Open Access? (Green and Gold Open Access)?

Concerning the open access publication of scientific peer-reviewed publications, such as data papers, two main open access approaches exist: green and gold open access.

- Green Open Access (Self-archiving): The scientific work (e.g. data paper) is archived by the researcher in an online repository, such as OpenScienceLink before, after or in tandem with its publication. Access to the article is often subject to an embargo period because publishers wish to retain some type of exclusivity in order to recoup their investments.¹⁹
- Gold Open Access (Open Access Publishing): The scientific work is immediately provided in an open access fashion by the publisher. The costs associated to the publication do not longer reside with the readers. Instead, they can be charged to the university or research institute or to the research funder.²⁰

2.12 Where should data be published?

Data are stored in data repositories. These are databases or archives created to collect, disseminate and preserve scientific output and make them freely available.²¹ There are many data repositories to choose from. Repositories can be institution-, discipline- or content-specific, or general. A list of international repositories can be found at databib.org.

Some things to consider when choosing a repository are:

- Does your employer, funder, or publisher recommended or mandate certain repositories?
- Does your research discipline have conventions concerning where to publish data?
- What metadata and format of data does the repository require?

²¹ https://www.openaire.eu/support/faq





¹⁸ Alma Swan, 'The Open Access Citation Advantage: Studies and Results to Date' (2010);

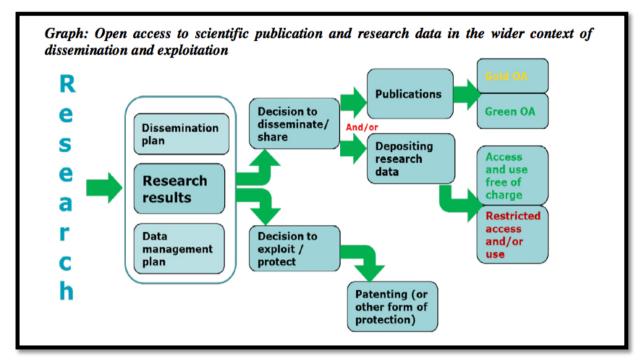
http://eprints.soton.ac.uk/268516/2/Citation_advantage_paper.pdf

¹⁹ IPR Help Desk, 'Open Access to Publications and data in Horizon 2020: Frequently Asked Questions' (May 2014), p. 2;

https://www.iprhelpdesk.eu/sites/default/files/newsdocuments/Open_Access_in_H2020.pdf

- Does the repository enable tracking of data citations by allocating a unique identifier?
- Does the repository manage conditional access to data, and is conditional access managed by the repository or the data owner?

As shown in the graph below the decision whether or not to protect the research results is important and needs to be discussed at an early stage with the research and development department of a researcher's institution.



Source: Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020

2.13 How should data be shared/structured?

For data to be exploited to its maximum potential it is necessary for it not just to be accessible but also intelligible and searchable. This is where standards for data preservation are required. Standards cover what should be included in the dataset, 'ontologies' or controlled vocabularies for annotating datasets, and exchange formats, for facilitating sharing.

Pragmatic and technical guidance on how to go about preparing your data suitably is available from various sources. A few are listed below.

<u>UK Data Archive</u>	Managing and Sharing Data (2011) is 'designed to help
	researchers and data managersproduce highest quality
	research data with the greatest potential for long-term use'.





<u>Digital Curation Centre</u>	The <u>DCC</u> provides advice on how to store, manage and protect digital data. Their site includes <u>tools and applications</u> , <u>MRC data plan FAQs</u> , information on <u>data management plans</u> , a list of <u>funders policies</u> , <u>legal information</u> and a developing series of ' <u>how-to</u> ' guides.
BioSharing	The site includes a growing <u>catalogue of standards</u> to help ensure that 'experiments are reported with enough information to be comprehensible and (in principle) reproducible, compared or integrated'.
Wellcome Trust	Provides <u>Guidance</u> for researchers: <u>developing</u> a <u>data</u> management and sharing plan
UK Medical Research Council	Provides <u>tools and resources</u> , including their <u>Data and tissues</u> <u>toolkit</u> and their <u>Cohort dataset directory</u> , plus a short <u>glossary</u> of common data-sharing terms.
National Cancer Research Institute	The NCRI Informatics Initiative 'supports the development of data standards and promotes a culture of data-sharing to facilitate storage and dissemination of research data'.
US National Institutes of Health	Resources include <u>examples of data sharing plans</u> alongside <u>more general policy documents</u>



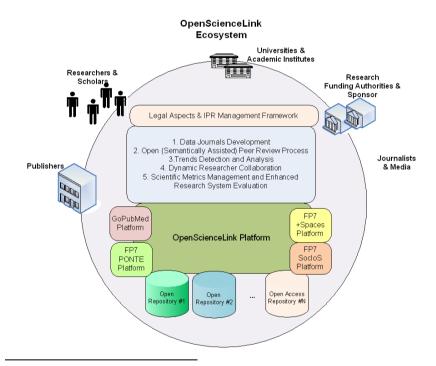


3 Specific OpenSciencelink Open Access Policy - FAQs

OpenScienceLink provides immediate open access to its content on the principle that making research freely available to the public supports a greater global exchange of knowledge. Authors of data made available on the platform remain the copyright holders and grant OpenScienceLink and any other third parties the right to use, reproduce and share the article according to the Creative Commons license agreement. ²²

All data on the OpenScienceLink is made available under the following license: <u>CC BY-NC-SA 4.0.</u>²³ This means that anyone who wants to submit research data onto the platform has to understand and agree for the work to be made available under this license. Any user of the research data has to conform to the use requirements under this license. To upload the work the author must confirm to have permission to do so by clearing the copyrights. The data may not include any personal information under the data protection directive.

The following sections will provide guidance for using the OpenScienceLink platform.²⁴



²² For more detailed information about these requirements we refer to D3.2 Legal and IPR Management Framework Specification





²³ The human readable license is available at https://creativecommons.org/licenses/by-nc-sa/4.0/

²⁴ For more detailed information about these requirements we refer to D3.2 Legal and IPR Management Framework Specification

3.1 How do I submit a data paper?

The <u>submission guidelines</u> for data papers for the BioMedical Journal OpenScienceLink platform can be found on the webpage of the Biomedical Data Journal:

http://biomed-data.eu/content/submission-guidance

Given the abundance of scientific journals and repositories, there is no set standard on how to submit a scientific article, including a data paper. The researcher himself must verify the submission guidelines of the parties where he wishes to submit the article.

3.2 How should I cite my data?

As a good practice you must always cite the data you use. This can also be an obligation depending on the license that is attached to the scientific work (e.g. Creative Commons licenses require proper attribution to be given, whereby the correct citation method is often indicated within the work).

If you use data from a repository that has been released under an open license you should include a reference to the data paper describing the data, followed by a reference to the data in the repository itself.

It is essential that the citations are in the references section of the article and include the Digital Object Identifiers (DOIs) or any other identifier the repository might use. Citation methods differ depending on your active field of research, however each citation must include basic elements that allow the dataset to be identified in the future, such as:

- Title
- Author
- Data
- Version
- Persistent Identifier, such as the Digital Object Identifier.

3.3 What are the criteria for a repository to be accepted by OSL?

Data must be made available via a suitable repository which should meet the following criteria:

- Allow open licences
- Sustainable to ensure the long-term preservation of the data
- Suitable for the type of data involved
- Able to provide persistent identifiers (e.g. DOI, handle, ARC etc.)

3.4 Is there any specific format I need to follow to submit a dataset?

Within OpenScienceLink, the dataset format follows the OpenScienceLink data model and the file may be in any file format. Nevertheless, depending on the chosen repository, formats may alter.





3.5 Wat Metadata Model does the OpenScienceLink platform use?

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.

3.6 From which locations will my dataset become available for download?

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.

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.

3.7 How does the review process differ from an ordinary research article review? What do the reviewers look at when it comes to a dataset article?

Published datasets are subject to a novel review process. The publisher (or the responsible editor) logs in the Platform in order to initiate the review process of submitted papers or datasets by creating a new review call. 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.

Submitted research papers, data papers, along with accompanying data sets, and data-based modelling and models' papers will be subjected to rigorous peer review. More information on the review process and the responsibilities of the partners involved, can be found through the following links:

Ethical Responsibilities of Authors:





http://www.biomed-data.eu/content/ethical-responsibilities-authors

Ethical Responsibilities of Reviewers:

http://www.biomed-data.eu/content/ethical-responsibilities-reviewers





4 Making Scientific Research and Data Open Access - Intellectual Property Rights Guidelines

The following section aims to provide researchers with guidelines on how to handle copyright. The main goal is to make researchers aware of the responsibilities that follow out of copyright. These requirements mainly arise when using other people's material during the creation of a scientific work. Before uploading scientific material to the OpenScienceLink portal, the researcher must take into account his responsibilities under relevant national copyright law.

With regard to the re-use of material that has been originally published on the OpenScienceLink portal, we refer to the guidelines concerning our licensing practices. Content uploaded to the OpenScienceLink platform is licensed to platform users under the Creative Commons 4.0 – BY-NC-SA license. The two main consequences of this license for the researcher are:

- 1. When uploading material via the OpenScienceLink portal, the author agrees that his material will be shared to others under the CC.04 BY-NC-SA License.
- 2. Material originally published via the OpenScienceLink portal can be re-used under the conditions of the CC.04 BY-NC-SA License.

The use of scientific articles via external databases accessible via the portal is not captured by this license. For instance, GoPubMed enables the access to articles from the Lancet and Open Access journals. In order to re-use these works, the licensing conditions of those journals, and of the authors that have written the articles therein, must be taken into account.

The following section does not only provide information to the researcher as to which material he can use. It also informs him concerning his rights. Nevertheless, whilst still retaining his exclusive rights, when uploading material via the OpenScienceLink portal, the author agrees to publish his work under the aforementioned creative commons license. Therefore allowing others to use his material under the conditions of that license.

4.1 What is Copyright?

Copyright aims to protect the rights of authors by ensuring that they receive recognition, payment and protection for their works. A copyrighted work could include a production in the scientific domain, regardless the mode or form of its expression. However, to enjoy copyright protection, works must meet the criteria for copyright protection: the work has to be "an original expression", i.e. it must be the result of the free and creative choices of the author. In principle, scientific publications, including data papers, will enjoy copyright protection.

4.2 What rights are granted to the holder of copyright?

Copyright grants the right holder several exclusive rights with regard to the actions that can be performed with the copyrighted work. As a researcher building upon other people's work, the exclusive rights of the original author must be respected.





- The reproduction right: the exclusive right to authorize or prohibit direct or indirect, temporary or permanent reproduction by any means and in any form, in whole or in part. This right also includes the right to adapt. The latter is
- The communication to the public right: the exclusive right to authorise or prohibit any communication to the public of the copyrighted work, by wire or wireless means, including the making available to the public of this work in such a way that members of the public may access them from a place and at a time individually chosen by them.

Due to the exclusive nature of copyright, the reproduction and public communication of protected works requires the permission of the right holder.

4.1.1. Do Bare Facts - Research Data enjoy copyright protection?

Bare facts do not enjoy copyright protection: they belong to the domain of knowledge, which is a public good.²⁵ Consequently, biomedical research data per se do not enjoy copyright protection as well.

Research data may still be copyright-protected if they have been expressed in a tangible form with a sufficient level of originality. The latter however does not protect the data as such, but rather the work in which the data has been incorporated.²⁶

4.1.2. Do Metadata enjoy copyright protection?

When metadata refer to descriptions of biomedical data using standardised keywords and terms, they are unlikely to be protected by copyright due to a lack of creative freedom in the choice, sequence and combination of the terms. However copyright in metadata cannot be ruled out.

Nevertheless, the purpose of OpenScienceLink - as well as most other open access initiatives - and the basis for its selection of data and services is the creation of a repository of well-structured and semantically linked datasets. The role of metadata herein is crucial and descriptions should be based on existing and established domain ontologies.²⁷ Indeed, any freedom in the description and definitions of terms may lead to less discoverability of the data and should therefore be avoided.

4.1.3. Do datasets enjoy copyright protection?

Scientific datasets are unlikely to attract copyright when they are the result of research. This can be concluded from the fact that the level of freedom required for a researcher *to express his creativity* is fairly limited.

²⁷ (D4.1 Opensciencelink consortium, 2015)





²⁵ Joris Deene, 'The Legal Status of Research Data (Copyright, Database Right)' (June 2015).

²⁶ Joris Deene, 'The Legal Status of Research Data (Copyright, Database Right)' (June 2015).

However it is the researcher who *'Has to make sense of the data that have been collected by exploring and interpreting them'*. ²⁸ When a researcher has sufficient freedom and made personal choices in how to present the results of his data collection activities the resulting dataset will be protected under copyright.

4.1.4. Does a database enjoy copyright or any other protection?

A database may enjoy copyright protection when it is considered 'original', i.e. when it is the author's own intellectual creation by reason of the selection or arrangement of the contents. A database selection or arrangement, which purely depends on technical factors or aims to achieve accuracy and exhaustiveness, will also not be protected by copyright. Scientific databases, which constitute bare facts and therefore limit creative freedom, are generally not considered to fall under copyright protection.²⁹

The creator of a database may also enjoy a *sui generis* right for the investment he has made when creating the database (*see section What is the Database Rights?*)

4.1.5. What are the exclusive rights of the original database's creator?

The author of the database has the exclusive right to carry out or authorise the:

- a) Temporary or permanent reproduction by any means, in any form, in whole or in part;
- b) Rights of adaptation, translation, arrangement and any other alteration;
- c) Any form of distribution to the public of the database or of copies thereof (subject to Community exhaustion); and
- d) Any communication to the public, display or performance to the public;

Scientific datasets and works originally provided to the OSL platform are licensed under the CC.04 – BY-NC-SA license. This license allows the re-use of the works within the OSL repository under the conditions of the license (see section Making Scientific Research and Data Open Access – Licensing Guidelines).

4.1.4. Who owns copyright?

In general this is the author or creator of the work. However there are a few instances where permission needs to be obtained from other parties. It depends on national legislation how this permission should be given and from whom.

²⁹ Lucie Guibault and Andreas Wiebe (Eds.) Safe to be open, Study on the protection of research data and recommendations for access and usage





²⁸ In quantitative research, data analysis often only occurs after all or much of data have been collected. However, in qualitative research, data analysis often begins during, or immediately after, the first data are collected, although this process continues and is modified throughout the study.

1) Joint ownership

If a work has two authors there is joint copyright for both authors. When a research project has multiple researcher's institutions there is joint copyright for all researchers/institutions

If research material is derived from existing data enjoying copyright and the newly created work also enjoys copyright there is joint copyright.

2) Works created in the course of employment

The vast majority of scholarly works is made in the course of employment either with a research institution, an enterprise or with universities. In some countries copyright in works created during the course of employment will vest in the employer.³⁰ However since this may differ, it is important to always consult national copyright regulations. Academic institutions and funding bodies may waive copyright in research materials and publications and assign ownership to the researchers. But the opposite might also be true: the rights of the researcher may be assigned to the research institution. It is therefore very important do check with your institution or funding body what applies to you.

It is important to check the copyright policy of your institution.

4.3 What is the Sui Generis Database Right?

The sui generis database right is an intellectual property right for the maker of a database that is granted when it can be demonstrated that the database shows there has been a substantial investment (qualitative and/or quantitative) in either the obtaining, verification or presentation of the contents of the database.

The right was designed to safeguard the position of makers of databases against misappropriation of the results of their investments, financial resources, time, effort and energy. To enjoy this protection, this investment must have been substantial. The investment that has been made in researching and creating the data included within the database does not count towards this assessment (very often spin-off databases may therefore not be protected).

The Sui Generis right allows the database producer to prevent extraction and/or re-utilization of the whole or of a substantial part of the database. This means that the owner of a database's consent is required if a researcher wishes to retrieve substantial portions of the database or reuse substantial portions of the database. The substantiality of extraction and re-utilization is evaluated qualitatively (e.g. how much effort went in creating the part of the database that has been extracted?) and/or quantitatively (e.g. how much content has been extracted from the database?). This analysis must be made on a case-by-case basis.

³⁰ In some countries, e.g. Belgium, a transfer of rights in the form of an explicit agreement between an employer and employee may be necessary for the employer to obtain copyright.





Retrieval of non-substantial parts from databases, e.g. some individual items, is allowed and does not require prior permission from the relevant right holder.

For more information on the sui generis database rights, we refer to the OSL Legal and IPR Management Framework Specification, found here:

http://opensciencelink.eu/wp-content/uploads/2013/06/OSL D3 2 LegalAndIPRManagementFrameworkSpecification.pdf

4.4 Can I re-use other people's material?

In principle, the following actions require the permission of the relevant right holder:

- (1) If the re-use includes creating a new physical fixation of the copyright protected work the right of *reproduction* prevents this without authorization.
- (2) As long as the work is used as is and not translated or rewritten there will be no illegal act of adaptation.
- (3) The act of mere linking existing datasets, publications and associated datasets or sets of raw data will however not constitute a prohibited act as long as these data are not duplicated.

With regard to protected data, as a general rule of thumb, following actions may still be performed without the original author's permission³¹:

- Incorporation of factual data in an original scientific work.
- Making a copy of the research for research purposes only.
- Citing research data

The author's permission would still be required for other actions performed with copy-right protected data.

The OpenScienceLink platform provides services for researchers to publish and re-use publications and datasets. Tools for the re-use of data include the analyses of data for trend reports but also combining different datasets for new research publications. In order to use a copyrighted work, permission from the author is required.

Content originally uploaded to the OSL platform is made available under a Creative Commons BY-NC-SA license (see section Making Scientific Research and Data Open Access – Licensing Guidelines). This enables the re-use of data under the conditions of that license. The use of scientific articles and datasets found in external databases accessible via the OSL portal is not

³¹ Joris Deene, 'The Legal Status of Research Data (Copyright, Database Right)' (June 2015).





captured by this license. For instance, GoPubMed enables the access to articles from the Lancet and Open Access journals. In order to re-use these works, the licensing conditions of those journals, and of the authors that have written the articles therein, must be taken into account. Publisher copyright policies can be found via the website of Sherpa/Romeo.

Besides legal requirements, authors of scientific material also have ethical responsibilities. In this regard, the <u>Biomedical Data Journal</u> submission guidelines may provide assistance. For instance, authors should always give proper attribution to their sources.

A general checklist concerning the use of research data can be found following this link:

https://www.fosteropenscience.eu/sites/default/files/pdf/1401.pdf

4.5 Can I copy and publish substantial parts of databases other than OpenScienceLink?

When copying and publishing substantial parts of other databases containing research data, it is always necessary to obtain consent for these actions from the right holders of these pre-existing databases. The only exception to this principle is when the pre-existing database is not protected by intellectual property rights.³²

Even though the OSL platform is licensed under the CC 4.0 BY-NC-SA license, this license does not pertain to the external databases accessible via the platform. For the extraction and reutilization of substantial parts of these databases, the researcher must respect that database's proper license.

4.6 Can I link my data to other research publications?

For Open Access Data Journals the linking between research literature and primary datasets is important. An OpenScienceLink researcher is able to generate links between his research publications and associated datasets either stored in the OpenScienceLink repository or in external repositories and/or open data journals.

Mere linking does not infringe an author's exclusive rights if and to the extent that this work is freely accessible on the website to which the Internet link leads. In this case, there is an assumption that the copyright holder has, when they permitted this communication, considered all Internet users as the public of his communication.

Where a clickable link makes it possible for users of the site on which that link appears to circumvent restrictions put in place by the site on which the protected work appears, it must be assumed that an infringing act takes place. This will equally be the case where a hyperlink would

³² OSL, Requirements, Use Cases and KPIs (2013); http://opensciencelink.eu/wp-content/uploads/2013/06/OSL_D2_1_ReportonStakeholdersMainUseCasesKPIsandDataSourcesAssessment.pdf





be applied to content which has not been lawfully made available (hyperlinks to illegal source materials).

The same upholds for embedding material.

4.7 Is there a copyright/sui generis exception for scientific research?

Many EU Member States have adopted a research exception to the right of reproduction and to the right of communication to the public when the copyrighted work is used "for the sole purpose scientific research. As not all member states have implemented this exception, the researcher is responsible for verifying the existence thereof. Research exceptions usually have the following conditions:

- 1. Works are used for the sole purpose of illustration for teaching or scientific research;³³
- 2. The source, including the author's name, is indicated, unless this turns out to be impossible;
- 3. Works are used to the extent justified by the non-commercial purpose to be achieved. Furthermore, the exception can only be applied to the extent that
- 4. The use of the works does not conflict with a normal exploitation of the work or other subject-matter and does not unreasonably prejudice the legitimate interests of the right holder (three-step test).

As is the case for copyright, some Member States have implemented an exception to the sui generis database right for scientific research.³⁴

In general, the extraction of substantial parts of a database will not infringe the exclusive rights of the database owner when:

- (i) The database is used for the sole purpose of scientific research;
- (ii) The sources are indicated;
- (iii) The database is used to the extent justified by the non-commercial purpose to be achieved.

The researcher is responsible for verifying the existence of this exception within his national legislation.

³⁴ Recital (36) explains that the term scientific research covers both the natural sciences and the human sciences.





³³ According to the Explanatory Memorandum, paragraph 36: "the term 'scientific research' within the meaning of this Directive covers both the natural sciences and the human sciences".

5 Making Scientific Research and Data Open Access - Licensing Guidelines

5.1 Introduction: License and agreements

By uploading your scientific work to the OSL platform the scientific work and data are made available under the Creative Commons 4.0 – BY-NC-SA license. The author keeps the copyright but grants OpenScienceLink and its users a license for use under the requirements *attribution*, *non-commercial* and *share alike*. Once selected, user licenses are non-revocable. We recommend authors to check whether their funding body allows for such licensing or rather requires a specific license.

The user of the OpenScienceLink platform must confirm to have all the relevant permissions needed prior to sharing his material (scientific articles or datasets) on the platform. The Creative Commons license has two main consequences:

- The author agrees that his work is licensed under the CC.04 BY-NC-SA license conditions (see section Making Scientific Research and Data Open Access Licensing Guidelines)
- The re-use of scientific works published under this license must take into account the CC.04 BY-NC-SA licensing conditions, *e.g.* the re-use of material originally uploaded the OSL platform.

However, the CC.04 BY-NC-SA license only pertains to scientific works that are or have been uploaded originally to the OSL platform, e.g. articles and datasets published in the Biomedical Data Journal. The CC license does not apply to works found in external databases accessible via the OSL platform.

5.2 What are publishing agreements?

When a research paper or underlying dataset has been or is to be published in a scientific journal, the question whether that material can also be rightfully uploaded to the OpenScienceLink platform will often depend upon the existence of a publisher agreement. Although more and more publishers offer open access, not all agreements allow for the data to be shared under the OpenScienceLink license.

Before uploading to the OpenScienceLink researchers need to verify if the agreements are compatible with OSL's creative commons license. ³⁵

³⁵ See for example Elsevier 's Authors who publish in Elsevier journals can share their research by posting a free draft copy of their article to a repository or website. Researchers who have subscribed access to articles published by Elsevier can share too. There are some simple guidelines to follow, which vary depending on the article version you wish to share. : http://www.elsevier.com/about/company-information/policies/sharing





- A list of publishers and their policies are available via the SHERPA/ROMEO site at http://www.sherpa.ac.uk/romeo/
- A list of open access journals can be found at **DOAI** (Directory of Open Access Journals)
- A list of compatible licenses is available at http://creativecommons.org/science

User license*	Read, print and download	Redistribute or republish the final article (e.g. display in a repository)	Text and data mine	Translate the article	Reuse portions or extracts from the article in other works	Sell or re-use for commercial purposes
CC BY 4.0	Yes	Yes	Yes	Yes	Yes	Yes
CC BY-NC-ND 4.0	Yes	Yes	Yes	Yes For private use only and not for distribution	Yes	No
Elsevier user license	Yes	No	Yes	Yes	No	No

Source: http://www.elsevier.com/about/company-information/policies/open-access-licenses

5.2.1. How to deal with a Publisher/Publisher Licenses?

The dissemination of scientific articles, including data papers, usually happens via the publication of the article through a scientific journal. In order to be published, the publisher has to obtain from the author a permission to publish the scientific work. Usually the author grants the publisher this permission by signing over his copyright to the publisher. By doing so, the publishers gains full rights and control over the article. If an author wishes to make the article available via open access at a later stage, he or she will thus have to ask permission from the publisher to do so.

As an author of an article or data paper that has been accepted for publication, there are different scenarios to consider before you are able to upload your paper to the OSL open access repository. The OpenAire Open Access platform has recognized the difficulties authors face in this regard and distinguishes between several scenarios:

- 1. The Publisher Agreement Requires The Transfer of Copyright
 - a. Sign the agreement as such
 - b. Do not sign, but provide the publisher a license to publish³⁶
 - c. Provide an addendum enabling open access publication.
 - i. This may relate to the data, the data paper or both data and data paper.
- 2. A publishing agreement has already been signed: does the agreement leave the rights to publish data or academic publications in an Open Access repository?

³⁶ http://copyrighttoolbox.surf.nl/copyrighttoolbox/authors/licence/





- a. If yes, you may provide open access to your article
- b. If no, ask your publisher to sign an addendum allowing you to provide open access

The OpenAire Open Access platform also provides a useful template letter for authors who wish to request a copyright change in order to enable open access:

https://www.openaire.eu/template-letter-for-authors-request-for-copyright-change-doc

Datasets can also be published in a journal as supplementary material to the article. If the journal requires copyright to be transferred to the publisher, it may be recommended to separate the supplementary data from the article itself and to ensure that the transfer of copyright for data is not required as a publishing condition.³⁷ Even though copyright protection is not granted on data gathered for research purposes, by doing so, no subsequent copyright claims should arise out of the research data.³⁸

5.2.2. What is an embargo period?

Some scientific researchers may allow open access publishing but only after an embargo period has lapsed. This means that the article cannot be made open access until a certain period has passed after the publication of the work in the scientific journal. These periods vary by journal and scientific domain.

Authors should respect the embargo period before uploading their work to an Open Access Platform

5.2.3. Under what license will the Open Science Link Platform publish my data?

The OpenScienceLink platform publishes and makes available data under the CC04 – BY-NC-SA license.

5.2.4. Why should open access works be accompanied by a license?

Even though open access policies have as their primary purpose for works to be made freely available for anyone, the true value from open access stems from the further use of freely available information. If a scientific work does not carry license information, subsequent users do not know what they can do with the work.

Licenses set out the conditions for re-use of a protected work. Thus they assure future users that they can use these works without recourse. Within an open access context, the existence of

³⁷ Preparing raw clinical data for publication: guidance for journal editors, authors, and peer reviewers *BMJ* 2010;340:c181 available at http://www.bmj.com/content/340/bmj.c181.full ³⁸ Ibid.





licenses also facilitates the understanding of text- and data mining possibilities.³⁹ They allow platforms to understand what actions can be performed on copyrighted scientific information.

5.2.5. Why the choice for a CC.04 BY-NC-SA License?

According to a UNESCO report concerning Open Access, the Creative Commons licensing scheme provides the following advantages over custom licenses:

- 1. For researchers, a ready-made license is more likely to suit a publisher's requirements, saving time and effort in drawing up a custom license.
- 2. Because they are commonly used and understood, a potential reader and re-user of a work will immediately understand the conditions of the license.
- 3. CC licenses have machine-readable metadata. This simplifies processes where applications such as harvesters and text-mining tools carry out automated tasks, including trend analysis. Subsequently, these tools can recognize, which content they are permitted to gather and work upon.

As a globally recognized license, the CC.04 BY-NC-SA License brings not only instant recognition and assurance of the legal status of the license, it also assures some interoperability of the content. Researchers in the EU are assured that their rights and permissions will be recognized more easily by users outside the EU and vice-versa. In addition, the CC04 – BY-NC-SA License includes the sui generis database right, increasing its interoperability. As such, it constitutes a suitable license under which the research data included in the OSL platform can be used.⁴⁰

5.2.6. What does the CC.04 BY-NC-SA License entail?

The CC.04 BY-NC-SA License allows others to remix, tweak, and build upon the licensed material on a non-commercial basis, as long as proper credit is given and new creations are licensed under identical terms as the original license.⁴¹ The following sections will detail the different components of the licensing: attribution, non-commercial and share alike.

- By uploading material to the OSL platform, the author agrees to license his work under the CC.04 BY-NC-SA license and thus allows his work to be used under these conditions (see below)
- The author using material from the OSL platform must use this material in accordance with the conditions of the CC.04 BY-NC-SA license.

⁴¹ http://creativecommons.org/licenses/





³⁹ UNESCO, Policy Guidelines for the Development and Promotion of Open Access (2012), p. 38;

http://unesdoc.unesco.org/images/0021/002158/215863e.pdf

⁴⁰ Ibid, p. 39.

1) BY – Attribution:

This condition refers to the attribution that must be made to the original authors of the work. The author and other parties designated to receive attribution must, in subsequent use, be named according to the manner requested by the licenser. If the licensor has supplied additional information that must be provided, this information must be retained. In addition, once the licensed work has been modified, it must be indicated that the work is modified. Considering proper attribution must be made, the licensee may not create the impression that their use is endorsed by the licenser or any other party designated for attribution.⁴²

2) NC - Non Commercial:

The original work can be used and adapted, but the original and derived work shouldn't be used for commercial purposes.

3) SA – Share Alike:

The licensed work can be adapted. However, the adapter is bound by the terms and conditions pertaining to the original license. The share-alike clause also requires that the adapter's license (the license that pertains to the derived work) is either the same license as the original license or any later version of that license. In addition, another CC license can be chosen for the derived work, as long as it contains the same features as the original license. In other words, the use of the modified work cannot be restricted or broadened (e.g. commercial use) by additional terms and conditions.

Works published in the OSL database are licensed under an SA License. Datasets created modifying the OSL datasets or including a substantial part of the OSL datasets have to be licensed under the same or compatible conditions as the CC BY-NC-SA license.

5.2.7. What are the consequences of the Share Alike clause?

ShareAlike licenses were designed to guarantee that the freedoms associated with the licensed work survive as the work is adapted by others and that the freedoms attach to adaptations of those work as well. The main reasoning behind the implementation of a ShareAlike clause is to prevent the 'monopolisation' of the original content through their modification. Considering the ShareAlike clause targets the publication of adapted works, in order for the clause to come

https://www.unesco.de/fileadmin/medien/Dokumente/Kommunikation/Open Content A Practical Guid e to Using Open Content Licences web.pdf (last accessed, 6-12-2015), p. 53.





⁴² Ibid, p. 30

⁴³ https://wiki.creativecommons.org/wiki/ShareAlike compatibility

⁴⁴ Till Kreutzer, *Open Content – A Practical Guide to Using Creative Commons Licenses* (German Commission for UNESCO, Wikimedia Deutschland and North Rhine-Westphalian Library Service Centre, 2013); available at:

into effect, there must first be an adapted work and that work must subsequently be shared. As such, the SA stipulation will not apply when someone adapts a work and keeps it to himself.⁴⁵

What is Adapted Material?

Adapted material refers to material subject to copyright and similar rights that is derived from or based upon licensed material and in which the licensed material is translated, altered, arranged, transformed, or otherwise modified in a manner requiring the permission under copyright and similar rights held by the licensor.⁴⁶

5.2.8. What are the conditions to Share Alike?

In accordance with the Creative Commons Legal Code, if you share the adapted material you have produced, the following conditions also apply⁴⁷:

- 1. The adapter's license you apply must be a creative commons license with the same License elements of the original CC 4.0 BY-NC-SA license (current version or later) or a BY-SA Compatible License.
- 2. The adapted work must include the text of, or the URI or hyperlink to the Adapter's License applied.
- 3. The adapted work must not offer or impose any additional or different terms or conditions on, or apply any effective technological measures to the Adapted Material that restrict exercise of the rights granted under the Adapter's License applied.

5.2.9. Does the Creative Commons license apply to the Sui Generis Database Right?

Yes. One of the reasons why the OSL platform chose the CC 4.0 BY-NC-SA license, was that it also applies to the sui generis data base rights. When the CC 4.0 BY-NC-SA license concerns a database, the following conditions apply:

- a. The license grants the right to extract, reuse, reproduce, and share all or a substantial portion of the contents of the database;
- b. if all or a substantial portion of the database contents are incorporated in a database in which you have sui generis database rights, then the database in which you have sui generis database rights (but not its individual contents) is Adapted Material
- c. One must comply with the conditions of Attribution, Non Commercial Use and Share Alike if all or a substantial portion of the contents of a database are shared.

⁴⁷ Section 3.b Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International Public License Legal Code; available at: https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode





⁴⁶ Section 1.a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International Public License Legal Code; available at: https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode

For more information on these conditions, we refer to the creative commons website and the <u>creative commons legal code</u>. ⁴⁸ To gain further insight into the general functioning of the creative commons framework, we refer to Till Kreutzer's *Open Content – A Practical Guide to Using Creative Commons Licenses*. ⁴⁹

5.3 How can I ensure License compatibility with the OSL License?

The Share Alike clause that the OSL license includes, requires those adapting the work to relicense the modified material under the same conditions as the original SA license. When using material licensed under the CC 4.0 BY-NC-SA license, the adapted material must be licensed under the same conditions.

However, the modified material that has been created, may have been created in part by the use of other material subject to a different license. In those instances, the Share Alike clause may result in license incompatibility: the user can comply with only one or two or more conflicting licence obligations.⁵⁰

As OSL is a platform, the responsibility when uploading licensed material lies with the person uploading the material. The author must check whether the license that will be granted by the OSL platform is compatible with the original license pertaining to the material used. Please also keep in mind that not all acts performed with a copyrighted work require the original author's permission (see section Making Scientific Research and Data Open Access – Intellectual Property Rights Guidelines).

⁵⁰ Ibid, p. 54

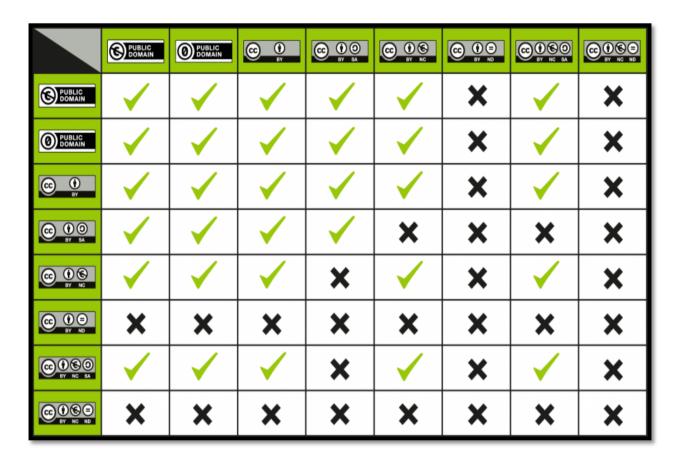




⁴⁸ https://creativecommons.org/licenses/by-sa/4.0/legalcode

⁴⁹ Till Kreutzer, *Open Content – A Practical Guide to Using Creative Commons Licenses* (German Commission for UNESCO, Wikimedia Deutschland and North Rhine-Westphalian Library Service Centre, 2013); available at:

https://www.unesco.de/fileadmin/medien/Dokumente/Kommunikation/Open Content A Practical Guid e to Using Open Content Licences web.pdf (last accessed, 6-12-2015)



Source: Creative Commons. This figures provides an overview of the CC 4.0 BY-NC-SA license with other creative commons licenses.

https://wiki.creativecommons.org/wiki/ShareAlike_compatibility

https://wiki.creativecommons.org/wiki/ShareAlike_compatibility_process_and_criteria

https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode

https://www.unesco.de/fileadmin/medien/Dokumente/Kommunikation/Open Content A Practical Guide to Using Open Content Licences web.pdf

More information on intellectual property rights

For more information on intellectual property rights, or if you have any other question related to copyright, the Intellectual Property Helpdesk provides free, first-line advice concerning IPR to beneficiaries of EU funded research projects and EU SMEs.

More information on the European IPR Helpdesk can be found following this link:

https://www.iprhelpdesk.eu/





6 Making Scientific Research and Data Open Access - Data Protection Guidelines

6.1 Personal Data and Data Protection

One of the main goals of the OpenScienceLink platform is to facilitate the capabilities of stakeholders within the open access value chain in accessing a multitude of openly accessible data sources, while at the same time empowering their semantic linking and data processing. Open data policies differ from general open access policies in that they also require the end-user to take into account privacy considerations. If data concerning individuals are published online, privacy and data protection legislation must be taken into account. Within Europe, the main regulatory instrument for preserving the privacy and data protection of individuals is the European Data Protection Directive⁵¹ and, in the future, the European Data Protection Regulation.

6.2 When should I take the Data Protection Framework into account?

The European Data Protection framework applies to "the processing of personal data wholly or partly by automatic means, and to the processing otherwise than by automatic means of personal data which form part of a filing system or are intended to form part of a filing system".

In order for the data protection ruleset to be applicable, the following conditions must be fulfilled:

- a) There has to be a data processing activity (see section What is data processing?) and
- b) The data that is being processed must be related to an identified or identifiable individual (*see section What is personal data?*).

When collecting the data and when creating and publishing the resulting datasets, the researcher will have to comply with the data protection framework if the abovementioned criteria are met. However, in order to retain the functionality of data, whilst also guaranteeing an adequate level of protection for individuals whose data were processed in the context of scientific research, the OpenScienceLink platform has chosen to only allow the uploading of datasets that have been anonymized. The responsibility to anonymize data lies with the researcher. For more information on how to anonymize data, we refer to the section *How Can Research Be Anonymized*. The researcher is also responsible for the application of the data protection framework for data processing actions concerning personal data prior to the anonymization of that data.

⁵¹ http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML





For more information on the legal framework concerning privacy and data protection, we kindly refer to D3.2 of the OpenScienceLink project.⁵²

Researcher guidelines on how to appropriately collect personal data can be found on the website of EU project Foster:

https://www.fosteropenscience.eu/

6.3 What is data processing?

Data processing refers to any operation or set of operations which is performed upon data, whether or not by automatic means, such as collection, recording, organization, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, blocking, erasure or destruction. Generally speaking, the act of data processing extends to almost every action performed on data.

6.4 What is personal data?

Personal data refers to any information relating to an identified or identifiable natural person ('data subject'). An identifiable person is defined as a person who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identify.

The OpenScienceLink platforms requires personal data to be anonymized before it is uploaded in order to fully guarantee the rights of the individuals who have been involved in the creation of the research data sets.

6.5 Is biomedical data personal data?

Depending on the type(s) of biomedical data a researcher uses, reliable information concerning an individual's identity can be inferred. Moreover, if the research data can be linked to an identified or identifiable individual, data concerning the health of an individual is considered sensitive. The processing of sensitive data should be subject to higher scrutiny.

The OpenScienceLink platforms requires personal data to be anonymized before it is uploaded in order to fully guarantee the rights of the individuals who have been involved in the creation of the research data sets.

6.6 Does anonymized data constitute personal data?

The principles of data protection do not apply to anonymous information. Anonymous information is data which does not relate to an identified or identifiable natural person or to data rendered anonymous in such a way that the data subject can no longer be identified and re-identified.

⁵² OSL, *Legal and IPR Management Framework Specification* (2015), http://opensciencelink.eu/wp-content/uploads/2013/06/OSL_D3_2_LegalAndIPRManagementFrameworkSpecification.pdf





The data protection rules do not concern the processing of anonymous information, including for the processing of anonymous data for statistical and research purposes.⁵³ Anonymization of personal data will put the data outside data protection regulations as the data can thus no longer be linked to an individual and therefore cannot be considered personal data. All obligations stemming from data protection regulations, are no longer required in this situation.⁵⁴

Considering its open access nature and the emphasis of the platform, i.e. biomedical research data, yet realizing the value of biomedical data for research, the OSL platform requires all data to be anonymized before it is uploaded in order to fully guarantee the rights of the individual.

It is important to note that the process of anonymizing personal data is considered a data processing activity of personal data. All stages of processing prior to the anonymization of data, as well as the anonymization itself, thus requires data protection principles to be taken into account. It is the responsibility of the researcher to ensure the correct application of these principles.

6.6.1. Pseudonymization

Particularly important to note for scientific research is that pseudonymity of data cannot be equated to anonymised data.⁵⁵ Pseudonymisation still allows an individual data subject to be singled out and linkable across different data sets. Because pseudonomity is still likely to allow for identifiability, pseudonomysed data is considered to fall inside the scope of the data protection framework.

6.6.2. Does the OSL Platform require anonymization?

For the OpenScienceLink project, anonymity is a key concern when the platform services involve the making available of clinical research data. After evaluation of the platform services and legal framework it was decided that at this stage the platform should only allow research data when it is properly anonymized. Thus researchers are required to properly anonymize their research results before uploading them to the platform.

⁵⁵ WP29, Opinion 05/2014 on Anonymisation Techniques (10 April 2014), p.10; available at: http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf





⁵³ According to Recital 26 of the Directive, "the principles of protection shall not apply to data rendered anonymous in such a way that the data subject is no longer identifiable."

⁵⁴ Sarah Hugelier, 'Publishing Open-Access Biomedical Data: Legal Challenges', *Biomed Data J. 2015; 1(1): 43-51.* When it comes to the lengths and efforts the data controller should go through to anonymize the data, it is recommended that an extra effort should be made to truly remove any potential identifying information. When anonymization techniques fail, the data control-ler again has to rely on article 7 or 8 of the Directive for legitimate grounds of processing the data.

A case-by-case analysis is necessary to verify whether the identification of a person based on 'all the means likely reasonable' is indeed no longer possible.⁵⁶

6.6.3. What is anonymization?

In order to fully respect the privacy and data protection rights of individuals, the OSL requires all data to be anonymized. In its opinion on personal data, the Article 29 Working Party has defined anonymous data as "any information relating to a natural person where the person cannot be identified whether by the data controller or by any other person, taking account of all the means likely reasonable to be used either by the controller or by any other person to identify that individual." The latter refers to the situation whereby the individual information may not in itself lead to the identification of an individual, but the combination of that information with other direct or indirect identifiers may lead to the identification of the individual.

Check: Can my data be confidentialised?58

Some data cannot be modified to sufficiently minimise the risk of harm to participants or subjects; or, if modified to this extent would render the data largely invaluable. This is often because the context in which the data were collected makes recovery of identity possible. For example, data from in-depth personal interviews, and some kinds of cultural or historical data for which identity is an intrinsic aspect of the data. In the former case, data may also reveal identities of other individuals who did not participant (and thus consent) in the study.

In the majority of cases, data, which cannot be confidentialised, cannot be published without the explicit consent from research participants and without the approval of the research institute's Research Ethics Committee. This does not necessarily preclude publication of a metadata record if that record cannot be used to identify the persons or subjects involved in the study.

6.6.4. How can research data be anonymized?

Taking into account the nature of the data, the researcher will have to carefully consider what information can be left out without the data losing its value. The researcher will also have to consider whether or not data can be linked back to an identified or identifiable individual through the use of reasonable means. The following considerations may be used as guidance.

The US Health Insurance Portability and Accountability Act of 1996 (HIPAA) has provided a list of 18 items that need to be removed from patient information in order for it to be considered anonymous for the purposes of sharing information between the "covered entities" specified in the act. The BMJ has also provided practical guidance for those involved in the publication

⁵⁸ http://ands.org.au/guides/sensitivedata.html





⁵⁶ Article 29 Working Party (2013). Personal data, pseudonymous data and anonymous data: an Article 29 Working Party Future of Privacy subgroup discussion paper. *ICO*; Data Protection Commissioner. Data Protection Guidelines on research in the Health sector, (9). Retrieved from http://www.dataprotection.ie/documents/guidance/Health research.pdf.

⁵⁷ Article 29 Working Party (2007). (21) 26

process of medical data by proposing a minimum standard for anonymizing (or de-identifying) data for the purposes of publication in a peer reviewed biomedical journal or sharing biomedical data with other researchers, either directly, where appropriate, or via a third party.⁵⁹ Moreover, The BMJ considers there to be a risk of identification when a data set includes three or more indirect identifiers.

Data can be anonymized using the following methods. Please consider the following to be guidelines. A case-by-case analysis will still be necessary:⁶⁰

• For quantitative data

- remove direct identifiers e.g. names, address, institution, photo
- reduce the precision/detail of a variable through aggregation e.g. birth year vs. date of birth, occupational categories, area rather than village
- generalize meaning of detailed text variable e.g. occupational expertise
- restrict upper lower ranges of a variable to hide outliers e.g. income, age
- combining variables e.g. creating non disclosed rural/urban variables from place variables

1. For qualitative data

- plan or apply editing at time of transcription
- avoid blanking out; use pseudonyms or replacements
- avoid over-anonymizing removing/aggregating information in text can distort data, make them unusable, unreliable or misleading
- Consistency within research team and throughout project.
- Identify replacements, e.g. with [brackets]
- keep anonymization log of all replacements, aggregations or removals made keep separate from anonymized data files

Datasets containing personal data cannot be made available under the creative commons license on the platform! Therefore researchers must ensure that their data sets do not contain personal data. If the original data set did contain personal data, all data must be properly anonymized before uploading the content to the OpenScienceLink Platform.

Pseudonymity should not be equated to anonymity. Pseudonymized datasets cannot be uploaded to the OSL platform!

⁶⁰ Research Data Management Support Services UK Data Service University of Essex April 2014





⁵⁹Preparing raw clinical data for publication: guidance for journal editors, authors, and peer reviewers *BMJ* 2010;340:c181 available at http://www.bmj.com/content/340/bmj.c181.full

6.7 What are my other responsibilities as a researcher?

During the research process, personal data may be collected and processed by the researcher. If personal data is being processed during the research phase, the data protection requirements fall under the responsibility of the research institution and researcher carrying out the research.

It is necessary that the researcher is transparent towards and provides adequate information to the individuals involved in the research. First and foremost, participants in a study should have agreed to participate in that study. Their agreement should be written. Moreover, consent should be provided for all activities that make up the data life cycle. This means that consent must not only be given for the data subject's engagement in the research process, but also for the dissemination and sharing of the data. This requirement upholds even when research data is anonymized at a later stage. Thus, a researcher must carefully consider the future uses of the research data.⁶¹ Furthermore, it is important to note that the consent of research participants should be informed. This requires that the researcher actively provides information concerning *inter alia*⁶²:

- The project and the researchers that are involved
- The data subject's participation in the study
- The use of information within and beyond the project
- The rights of the data subject, e.g. the right to withdraw from participation.

In addition, the researcher should best regulate the access to the personal data he has collected. Access to personal data should only be provided to those individuals that require access.

At the moment, OpenScienceLink only allows anonymized data to be uploaded to the platform. The anonymization process remains the responsibility of the platform user: he must confirm before uploading the data to the platform that the research results do not include personal data.

More information on the confidentiality of personal health information used for research can be found following this link: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1513443/

The EU project Foster also provides researchers with guidelines concerning the use of personal data in research: https://www.fosteropenscience.eu/

6.8 Can I re-use the data on the OSL platform (data protection)?

Yes, the data that has been uploaded to the OSL platform can be re-used by the researcher.

⁶² Sebastian Netscher, *Research Ethics and Legal Compliance: Informed Consent and Data Licensing* (CESSDA Training at the Data Archive for the Social Sciences)





⁶¹ Veerle Van den Eynden, 'Research Data Management and Open Data: Legal and Ethical Issues – Consent, Anonymisation, Access Regulation' (ECPR Summer School 2015).

However, the open-access datasets provided by the Biomedical Data Journal might contain anonymised data, i.e. datasets pertaining to previously identifiable individuals. In accordance with recommendations by the Article 29 Data Protection Working Party (Dir 95/46/EC), license-holders are prohibited from re-identifying any individual. License-holders are also prohibited from using the data to take any measure or decision with regard to the individuals concerned. When the license-holder detects that the individuals involved can be or have been re-identified, the license-holder is obliged to send a notification thereof to the licensor of the dataset.

A dataset shall be considered compromised as soon as the data has become re-identifiable, i.e. as soon as the data subjects to which the data relate have become re-identifiable. Upon notification that the dataset might be compromised, the licensor shall be granted the right to recall the 'compromised' dataset. The licensor shall be given the right to suspend or terminate the accessibility of the data, for instance by removing the file from the OpenScienceLink platform and the BMDJ website.

6.9 Does the OpenScienceLink Platform Guarantee the Protection of my personal data?

The use by OpenScienceLink of data provided in your 'user profile' will constitute a form of 'processing'. The OpenScienceLink Consortium aims to respect end-users' concerns with regard to privacy and values its relationship with the end-user. Our Privacy Policy can be found following this link:

http://opensciencelink.eu/opensciencelink-platform/opensciencelink-platform-privacy-notice/





7 Guidelines for Open-Access Initiatives and Open-Access Policy Makers

The following section aims to provide open-access initiatives and open-access policy makers with guidelines and best practices in an effort to increase the overall potential of the open access to scientific information. These practices and guidelines have been developed taking into account the specific perspective of the OpenScienceLink platform. They build upon the insights gained from research performed under Work Package 8, i.e. the stakeholder, technical and financial evaluation of the platform's pilot services. Thus the guidelines relate to the pilot services developed under the OSL framework. Therefore, these guidelines may seem to target open access initiatives similar to OSL only. Nevertheless, these guidelines aim to increase the overall potential of open access on scientific information in order to actualize innovative research. These are best practices and policy blueprints that, when taken into account, may prove to be beneficial for all stakeholders of the open access value chain. Of course, not all guidelines formulated below can be integrated by every stakeholder. For instance, the reduction of costs concerning open-access data journal development primarily targets journal publishers. Yet, the reduction of costs may eventually spill over to university researchers. Furthermore, Member States seeking to encourage open access initiatives, might find inspiration in these guidelines for the improvement of national open-access policies. Finally, our guidelines also include information on the benefits that certain open-access actions may have on the stakeholders of the open-access value chain. These benefits could be something to strive for: if all benefits have been achieved, your open access initiative may be considered a success.

7.1 Promote and encourage the inclusion of effective review processes based on open access to scientific information

The OSL platform aims to provide a novel open, semantically-assisted peer review process. The inclusion of effective review processes within open access initiatives should nonetheless be further encouraged. Indeed, peer-reviews guarantee that scientific research upholds a certain qualitative standard and they safeguard the credibility of scientific research. Therefore, peer-review mechanisms should also be an integral component of making scientific information open access.

Moreover, smart and integrated peer-review systems do not only benefit overall research quality. They may also be presented as an added incentive for stakeholders within the open-access value chain. For more information on potential business models, we refer to D8.2.3 Stakeholders, Technical and Financial Evaluation.

Stakeholder	Benefits
Researchers	Gain recognition and career rewards for reviewing; Gain experience as reviewers.
Publisher	





	Speed up and ensure high quality of the review process.
	Improve own submission platforms by using services that support a semi-automated review process.
	Introduce new models in reviewing process.
Editor	Gain time in findings more reliably competent reviewers for their journals. Increase acceptance rate of invited reviewers. Increase quality of reviewing process.
Research Sponsors and Funding Authorities	Seek guarantees that published results (e.g. datasets) of sponsored research have been subject of rigorous review of high quality.

The benefits of effective peer-review processes for open-access stakeholders

7.2 Further facilitate research collaboration

Our analysis has shown that the importance of collaboration should be estimated extremely highly.⁶³ The provision of collaboration tools for researchers and research institute by Open Access initiatives should thus be further promoted. Indeed, in sciences, the subjects get more and more complex every day. This means that one subject can no longer be adequately tackled by one person alone. The average number of authors of high-end publications, such as Science and Nature, continues to increase. Therefore, it is paramount that open access initiatives – who have as one of their primary goals to enable researchers to build upon other people's works – provide tools for collaboration, especially in an internet age where research links, for instance through social networking capabilities, can be easily set up. Analytics should enable platforms to draw conclusions, make recommendations and suggest hypothesis to teams.⁶⁴

When done right, the advantages of setting up good collaboration tools should not be underestimated as they contribute to the core activities of stakeholders involved within the Open Access value chain. For more information on potential business models, we refer to D8.2.3 Stakeholders, Technical and Financial Evaluation.

Stakeholder	Benefits
Researcher	Broaden up collaborations with other researchers and/or research groups and communities.

⁶³ OSL, D8.2.3. Stakeholders, Technical and Financial Evaluation (February 2016)

⁶⁴ OSL, D8.2.2. Stakeholders, Technical and Financial Evaluation (February 2016)





	Increase potential for research synergies and visibility for job positions.
	Find collaborators, who are highly relevant to the topic of their interest, fast and effortlessly.
	Be presented with the opportunity to join dynamically formed research communities and, thus, remain involved in activities within their field of interest.
	Be easily pulled out from the research "crowd".
Publisher	Fast and effortless discovery of potential editors for their journals whose research activity fits well the journals' topics.
Editor	Increased and/or more targeted invitations to researchers for submission of articles journals and/or journal issues.
Research Sponsors and Funding Authorities	Efficient and effective discovery of potential evaluators for research work which has been applied for funding.

The benefits of collaboration tools for open-access stakeholders

7.3 Further promote the development and integration of novel trend analysis methods.

Trend analysis is one of the key tools for decision making today and especially in the future. The availability and – more – important the accessibility of big data volumes ask for tools to interpret available data. Trend analysis also promises big financial potential as it allows open access initiatives to tailor their products to the demands of the end-users. Initiatives providing targeted content when they are needed the most should be considered a unique selling point. Moreover, the integration of trend analysis tools could also provide additional benefits to the various stakeholders within the open access value chain. For instance, they may greatly reduce the transaction costs associated with scientific research. Nevertheless, caution is required as well. Implementation of trend analysis tools should not deter others from seeking innovative research areas themselves. The ability to follow trends is valuable, yet scientific research also requires trends to be set. For more information on potential business models, we refer to D8.2.3 Stakeholders, Technical and Financial Evaluation.

Stakeholder	Benefits
Researcher	Avoid starting or reconsider continuing working on research topics

⁶⁵ OSL, D8.2.3. Stakeholders, Technical and Financial Evaluation (January 2016)





	which are in decline.
	Discover the "hot" research areas to upgrade your research.
Universities/Academic Institutes/Scientific Societies	Facilitate effective distribution of academic resources to the research efforts. Assist in the organization of scientific congresses and meetings to include "hot" topics and invite top researchers.
Editor	Facilitate the reviewing of proposals about the organisation of special issues Support the decision process concerning the organization of special issues in fields with rapidly rising interest. Detect top authors in a field that could be invited to contribute to a special issue or a new journal.
Publishers	Tool for re-evaluating the topics that their journals focus on as well as for deciding upon the issuing of a new journal (aiming at a rising and promising topic). Detect top researchers in a topic or field and invite them to become editors or members of the editorial board.
Enterprises with R&D departments (e.g., pharma)	Access to a series of tools and services which can significantly assist building their research strategy and allow for more effective positioning of their resources in research efforts.
Research Sponsors and Funding Authorities	Use of the platform for consultation when preparing research agendas and calls for research proposals. Assist in decision about proposals for research grants.
Media	Improvement of the media coverage of scientific areas, fields and topics which are considered as uprising or reviving and their progress could potentially be of interest to the public. Fast and effortless identification of top or rising scientists in specific topics, fields or areas for whom documentaries could be prepared or interviews be taken.

The benefits of trend analysis tools for open-access stakeholders.

7.4 Further encourage research evaluation processes

The evaluation and scoring of existing research can be an attractive component for open access initiatives. This seems especially the case when it can be combined with trend analysis. Indeed, after detecting relevant research trends, research evaluation can be used to detect the best





research in the particular field. This requires however, reliable and scientific correct measures: doing analysis on the basis of mediocre designed indicators is rather senseless. Thus, it is important that reliable indicators are established in order to increase the potentialities of open access. In a similar fashion, flexibility in search terms should enable third-party-users to find better results. Instead of finding 'words', what should be found is 'meaning'. Revertheless, considering the impact these tools might have on the researchers and research institutes, it is important to provide information and be transparent concerning the parameters used within the evaluation calculation. Indeed, receiving a bad evaluation can be detrimental to a researcher's career and can thus deter a researcher from making scientific research available. For more information on potential business models, we refer to D8.2.3 Stakeholders, Technical and Financial Evaluation.

Among the benefits gained from smart research evaluation are:

Stakeholder	Benefits
	Access to research evaluation of high customisability and granularity; from researcher and research group to institutions and countries and from papers and data sets to journals and publishers for:
Researcher	 efficient management of their own scientific curriculum vitae having an instant view of the evaluation of: researchers within their field in order to direct their collaboration interest or follow their work scientific works in their field so that they remain always updated with important research results and efforts available publishing means in their field and/or specific topic so that they make informed decisions about the journals and conferences through which they will publish their work available entities (both academic and business ones) in their field – or related ones – so that they can investigate their jobrelated options with as much information available as possible.
Publisher	Are given the possibility to have access at any time to the evaluation of their publishing means (including journals, journal issues, special issues, etc) from a variety of angles and be able to detect and highlight the one(s) in which they are best at (e.g., academic impact, business penetration, etc) for their dissemination and communication purposes. Instant access to the evaluation of researchers for making informed decisions when choosing editors and building the editorial board or when planning to invite researchers to submit papers to their journals.
Research Sponsor and/or	

⁶⁶ OSL, D8.2.3. Stakeholders, Technical and Financial Evaluation (January 2016)





Funding Authority	Access to the evaluation of researchers who apply for funding allowing for a more informed and less time-consuming decision process.
Academic and/or Research Institution	Continuous monitoring of the evaluation of their institution, schools, departments, groups, researchers from various aspects rather than relying on yearly analysis of their positioning among other institutions which is quite often based on partial data and very specific evaluation metrics.
Media	Improvement of the media coverage of biomedical articles which publish high quality research and of scientists whose research is of great value to their field rather than relying on the media's assessment of the scientific research value(with its main driver quite often being potential appeal to readers). Increased credibility and improved reputation of the media means (newspaper, magazine, electronic news portal, etc) by avoiding publishing of scientific articles which are of low scientific validity and may lead to confusion, cause false hope or unsubstantiated worries to the public.

The benefits of research evaluation for open-access stakeholders

7.5 Further reduce the costs of open – access data journal development

Open access data publishing can be an expensive endeavor.⁶⁷ In order to meet the cost of data publishing, authors publishing within open-access journals are often asked to pay an open access fee or article-processing charge (APC) per published paper. Although this cost is usually covered by the research funds from the departments that contribute to this work, it may nonetheless remain a burden on the shoulders of the individual authors. Consequently, article-processing charges could inhibit the promotion of open access to scientific research.

In an effort to partially remove this burden, some publishers have created different *Membership programs* to alleviate associated publishing costs. In general, a Membership Program enables academic and research institutions, societies, groups, funders and corporations to actively support open access in scholarly publishing, and help ensure the most widespread dissemination of the work published by their researchers or members. Contingent upon the type of Membership program chosen by the institution, Member institutions may cover some or all of

⁶⁷ For instance costs associated with: policies and business processes need to be developed, legal issues, website and social media management, journal management, printing and mailing (if print is chosen), advertising, overhead, etc.





the publication costs for their individual researchers when they submit to a certain publisher. Thus the use Membership Programs may be beneficial to the overall open-access environment.

There are several types of memberships programs that can be chosen to address the costs associated with data publishing. Three different types of Membership programs are described below:

1. Prepay Membership

Enables an institution to cover the whole cost of publishing with no additional fees paid by their authors.

2. **Shared Support Membership**

The cost of publishing is split between the institution and author.

3. **Supporter Membership**

Members pay a flat rate annual Membership fee based on the number of science and medical researchers and graduate students at their institution. A 15% discount on the article-processing charge (APC) is given when publishing in the journals.

In addition, publishers of open access data journals may opt to reduce costs by increasing the efficiency of journal management and journal preparation. This could, for instance, be achieved by:

- 1. Seeking synergies by assigning the same person to manage three or four journals.
- 2. Requesting that authors use a dedicated writing tool (that provides the articles in print-ready form).
- 3. Providing templates in commonly used software tools and request that authors use the one they are used to.
- 4. Requesting that articles be written in a language that does not require copy-editing.
- 5. Providing paid services for formatting submissions to journal requirements (in case a contribution is not formatted according to journal requirements)
- 6. Providing paid copy-editing services for cases when a contribution is not written with acceptable quality of language.

Finally, revenue could be increased by seeking direct support from funding agencies and academic libraries, foundations, corporations, etc.⁶⁸ For more information on potential business models, we refer to D8.2.3 Stakeholders, Technical and Financial Evaluation.

⁶⁸ OSL, D8.2.3. Stakeholders, Technical and Financial Evaluation (February 2016)





7.6 Take into account the legal requirements - Legal Checklist for OA Platforms

Open Access policies increasingly cover not only scientific papers, but also research data. Open data policies tend to differ from Open Access because these policies must also take into account the issues of privacy and other frameworks that prohibit the disclosure of data for other reasons (e.g. confidentiality). The following requirements are written from the perspective of an Open Access Platform such as OpenScienceLink. Legal considerations to be taken into account by the scientific researcher are listed in the previous sections.

Data Protection Requirements

The Data Processing requirements refer to the processing of the personal data of platform users, and do not relate to the research data within data sets. Within this context, the individual's right to data protection must be guaranteed. This might necessitate researchers to anonymize their data (*see section How can research data be anonymized?*). For a more in depth elaboration on these principles, we kindly refer to D3.2. Legal and IPR Management Framework Specification.⁶⁹

R1	Data Protection compliance		
	If personal data is processed within the European union, Open Access initiatives and all		
	entities involved <u>must</u> comply with national and European data protection legislation.		
R2	Allocation of roles and responsibilities		
	Designate the data controller and, if appropriate, the data processor		
	• The <i>controller</i> is the natural or legal person which alone, or jointly with others,		
	determines the purpose and means of the processing of personal data. The		
	controller is the one primarily responsible and liable for the legality of the		
	processing and the fulfillment of obligations towards the National Data Protection		
	Authority (NDPA) and the data subjects.		
	• The <i>processor</i> is the party processing personal data on behalf of the <i>controller</i> .		
	This will typically be a sub-contractor or a specialized third-party company.		
	• The relationship between the controller & processor must be governed by a		
	contract wherein the different obligations of the different parties are well		
	indicated.		
	• In case of more than one data <i>controller</i> and/or <i>processor</i> , indicate a single point		
	of contact (article 17 of the Data Protection Directive)		





R3 Ensure legitimacy of processing

Article 7 and 8 of the Data protection Directive provide an exhaustive list of legitimate grounds on which the processing can be based. For most Open Access Initiatives, the ground for processing <u>should</u> be the *data subject's* <u>consent</u> (e.g. via an electronic or paper-based consent form).

All other potential personal data, such as the research data, which will be published on the platform, <u>must</u> be properly anonymized in order to guarantee the protection of those involved.

R3.1 | Requirements for consent

The consent must be:

- Unambiguous
- Specific, distinctive and intelligible
- Based on accurate, full and understandable information
- Genuine and freely given, absent of any pressure
- Given before any processing occurs
- Given with the option to withdraw his or her consent and stop any further processing of the personal data
- Explicit when it concerns sensitive data in the sense of art. 8 95/46/EC

R4 Data Quality Principles

The personal data and the processing <u>must_comply</u> with the following data quality principles for processing personal data:

Personal data must be:

- Processed fairly and lawfully
- Collected only for specified⁷⁰, explicit and legitimate purposes and not further processed in a way incompatible with those purposes
 - Open Access infrastructures must have a <u>privacy notice</u> that clearly articulates the purpose of the data processing and any subsequent use must be limited to those articulated purposes
 - o Before commencing the processing take the time to (re)consider whether

⁷⁰The purpose(s) of processing MUST be identified in advance (prior to initial collection, transfer, etc.)





all data being processed are both <u>adequate and non-excessive</u> to achieve the purposes of the *Open Acceess* platform.

- o Clearly indicate whether information is optional or required
- Adequate, relevant and not excessive in relation to the purposes for which it is collected and/or further processed
- Deleted or rendered anonymous when no longer necessary for the purpose
 - O Prior to any processing operation of personal data, the storage duration of each data element <u>must</u> be specified, either individually or by category, for every entity that is involved in the processing. Limit storage duration whenever possible yet ensure that the data is readily available to authorized entities as long as it is necessary.
- Securely deleted, i.e. the data should not be retrievable after deletion
- Accurate and up to date.
 - Procedures must be in place on how to report and deal with suspected inaccuracies.

R5 *Transparency and administrative requirements*

The data controller and *Open Access infrastructures* must ensure the data subject's rights.

The data controller must:

- Meet all notification and authorization requirements for data processing that may derive of national law of the competent Member State
- Provide the data subject with <u>sufficient information</u>, <u>at least(1)</u> the identity of the controller, (2) the categories of data to be processed, (3) whether the info is obligatory or voluntary, (4) the purpose of the processing, (5) the recipients of the data and (6) the further rights to access and to rectify. This information <u>could</u> be provided in the <u>privacy notice</u> (cf. R4).
 - Where appropriate, data subjects should get more detailed information as to the processing operations performed upon their personal data (e.g. at what time individual processing operations took place, under which pretext, etc.).
- Provide the data subject with <u>the right to obtain intelligible information</u> from the data controller without expense or excessive delay
- Support the data subject's rights to legitimately rectify, reply, revoke, erase or





	<u>block</u> his or her personal data
	- Once the data controller has been determined, clearly <u>declare</u> this partner as such
	and the applicable jurisdiction in all relevant documents (terms & conditions,
	privacy notice, etc.)
	- Each organization must ensure that is has filed a notification to the <u>National Data</u>
	<u>Protection Authority prior</u> to each activity involving personal data processing. Be
	prepared for any requests by the Commission for additional items of information
	related to the processing.
	- The <u>consent of the data subject</u> must be obtained <i>prior</i> to any processing of his
	personal data. A versioning and archiving system must be in place for the
	informed consents given by data subjects to enable later verification that
	appropriate notice was given.
F6	Security and confidentiality requirements
	Open Access infrastructures shall implement advanced organizational and technical
	security measures to ensure confidentiality, integrity and authenticity.

Data Protection Requirements - The GDPR Framework

The following section provides a revised version of the Data Protection Compliance guidelines, taking into account the most recent version of the General Data Protection Regulation.⁷¹ Once a final text is introduced by the European legislator, the new framework must be adopted by the platform in its entirety.

R1.	Data Protection compliance
	If personal data are processed within the European union, the OSL infrastructure and all entities involved must comply with national and European data protection legislation.
R2.	Allocation of roles and responsibilities

The text used as a basis for our assessment is the final compromise text found at: http://www.statewatch.org/news/2015/dec/eu-council-dp-reg-draft-final-compromise-15039-15.pdf. The final compromise might still be subject to changes, but was at the time of writing the most recent, up-to-date version of the GDPR.





Designate the data controller and, if appropriate, the data processor

- The controller is the natural or legal person which alone, or jointly with others, determines the purpose and means of the processing of personal data. The controller is the one primarily responsible and liable for the legality of the processing and the fulfillment of obligations towards the National Data Protection Authority (NDPA) and the data subjects.
- The processor is the party processing personal data on behalf of the controller.
 This will typically be a sub-contractor or a specialized third-party company.
- The relationship between the controller & processor must be governed by a contract wherein the different obligations of the different parties are well indicated, e.g. subject-matter, duration of the processing, the nature and purpose of the processing, the type of personal data and categories of data subjects, the obligations and rights of the controller (Art. 26 GDPR)
- Where two or more controllers jointly determine the purposes and means of the processing of personal data they shall in a transparent manner determine their respective responsibilities for data protection compliance, in particular as regards the exercising of the rights of the data subject and their respective duties concerning information that needs to be provided, by means of an arrangement between them. The arrangement may designate a point of contact for data subjects. The arrangement shall duly reflect the joint controllers' respective roles and relationships vis-à-vis data subjects, and the essence of the arrangement shall be made available to the data subject (art. 24 GDPR)

Data Protection by Design and by Default (Art. 23 GDPR)

- The controller shall implement appropriate technical and organizational measures, such as pseudonymisation, which are designed to implement data protection principles, such as data minimization, in an effective way and to integrate the necessary safeguards into the processing.
- Implement technical and organizational measures for ensuring that by default only the personal data which are necessary for each specific purpose of the processing are processed.
- R3. | Ensure legitimacy of processing





Article 6 and 9 (sensitive data) of the GDPR provide an exhaustive list of legitimate grounds on which the processing can be based. For the OSL platform, the ground for processing should be the data subject's consent (e.g. via an electronic or paper-based consent form).

All other personal data such as the research data, which will be published on the platform, must be properly anonymized.

R4. Requirements for consent

The consent must be (Art. 7 GDPR)

- Unambiguous
- Specific, distinctive and intelligible
- Based on accurate, full and understandable information
- Genuine and freely given, absent of any pressure
- Given before any processing occurs
- Given with the option to withdraw his or her consent at any time and stop any further processing of the personal data (The withdrawal of consent shall not affect the lawfulness of processing based on consent before its withdrawal).
- Explicit when it concerns sensitive data in the sense of art. 9 GDPR

R5. Data Quality Principles

The personal data and the processing must comply with the following data quality principles for processing personal data:

Personal data must be (Art. 5 GDPR):

- Processed fairly and lawfully and in a transparent manner in relation to the data subject (Art. 5a GDPR)
- Collected only for specified72, explicit and legitimate purposes and not further processed in a way incompatible with those purposes (Art. 5b GDPR)

⁷²The purpose(s) of processing MUST be identified in advance (prior to initial collection, transfer, etc.)





- The OSL infrastructure must have a privacy notice that clearly articulates the purpose of the data processing and any subsequent use must be limited to those articulated purposes
- Before commencing the processing take the time to (re)consider whether all data being processed are both adequate and non-excessive to achieve the purposes of the OSL platform.
- o Clearly indicate whether information is optional or required
- Adequate, relevant and not excessive in relation to the purposes for which it is collected and/or further processed (Art. 5c GDPR)
- Kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed (Art. 5e GDPR)
- Deleted or rendered anonymous when no longer necessary for the purpose
 - O Prior to any processing operation of personal data, the storage duration of each data element must be specified, either individually or by category, for every entity that is involved in the processing. Limit storage duration whenever possible yet ensure that the data is readily available to authorized entities as long as it is necessary.
 - Data must be securely deleted, i.e. the data should not be retrievable after deletion.
- Accurate and up to date.
 - Every reasonable step must be taken to ensure that personal data are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay (Art. 5d GDPR)
- R6. Transparency and administrative requirements(Art. 12 GDPR)

The OSL platform shall take appropriate measures to provide the data subject with the required information relating to the processing of personal data in a concise, transparent, intelligible and easily accessible form, using clear and plain language.

The data controller must:





- Where data are collected from the data subject provide information on inter alia (Art. 14 GDPR)
 - Identity and contact details controller
 - Purposes for processing
 - o Recipients of personal data
 - The existence of the right to access, rectification, erasure, restriction of processing, object and data portability
 - The existence of automated decision making, including profiling, and meaningful information about the logic involved and significance and consequences of such processing
 - Further processing of data for a purpose other than the one for which data were collected prior to that further processing
- Where data has not been obtained from the data subject provide information on (Art. 14a):
 - Inter alia: controller identity, processing purposes, categories of personal data, recipients of personal data
 - o Inter alia: data storage period, existence of data subject's rights.
 - The source from which personal data originate, and if applicable whether it came from publicly accessible sources.
- The controller and the OSL infrastructure must facilitate the exercise of the data subject's rights (art. 12 (1)a GDPR)
 - o Right of access (Art. 15 GDPR)
 - o Right to rectification (Art; 16 GDPR)
 - o Right to erasure (Art. 17 GDPR)
 - Right to restrict processing (Art. 17 GDPR)
 - o Right to Data Portability (Art. 18 GDPR)
 - o Right to Object (Art. 19 GDPR)
 - Right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her. (Art. 20 GDPR)
 - Provide the data subject with the right to obtain intelligible information from the data controller without expense or excessive delay





- Once the data controller has been determined, clearly declare this partner as such and the applicable jurisdiction in all relevant documents (terms & conditions, privacy notice, etc.)
- Each organization must ensure that is has filed a notification to the National Data Protection Authority prior to each activity involving personal data processing. Be prepared for any requests by the Commission for additional items of information related to the processing.
- The consent of the data subject must be obtained prior to any processing of his personal data. A versioning and archiving system must be in place for the informed consents given by data subjects to enable later verification that appropriate notice was given.

R7. Security and confidentiality requirements (Art. 30 GDPR)

The OSL infrastructure shall implement advanced organizational and technical security measures to ensure confidentiality, integrity and authenticity.

R8. Privacy Impact Assessment and Prior Consultation

Where a type of processing in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, is likely to result in a high risk for the rights and freedoms of individuals, the controller shall, prior to the processing, carry out an assessment of the impact of the envisaged processing operations on the protection of personal data. (Art. 33 GDPR)

Meet all notification and authorization requirements for data processing that may derive of national law of the competent Member State (Art. 34 GDPR)

Privacy and Data Protection guidelines on the open access to scientific information with regard anonymized data sets.

It is strongly advised that open access data-sets do not contain personal data. Rather, open-access data sets should be anonymized by the researcher uploading the datasets. Where the data contained within open-access data sets previously pertained to (sensitive) personal information, license-holders should be expressly prohibited from re-identifying the anonymized data. If





already re-identified, license-holders should furthermore be prohibited from using the data to take any measure or decision with regard to the individuals concerned. Open access platforms should try notification mechanism, or oblige the license-holders to notify the open access platform as soon as it detected that the individuals involved can or have been re-identified. As soon as the data subjects to which the data relate have become re-identifiable, the dataset should be considered compromised. Upon notification that the dataset might be compromised, the open access platform should retain the right to recall the 'compromised' dataset. For instance, by clarifying to platform users that the platform has the right to suspend or terminate the accessibility of the data.

The advisory board on Article 29 Data Protection Working Party (Dir 95/46/EC) has stated the following with regard to anonymised, open-access datasets: Although formulated in the light of public sector information repositories, these considerations should also be taken into account when building Open Access platforms in general:

"The possibility to alert the licensor of the fact that re-identification has taken place or can take place must be available for all other web-users. When an increased risk of re-identification is discovered by the licensor, a procedure should be foreseen in the license whereby the licensor can 'recall' the 'compromised' dataset. In other words, the data protection clause should give the licensor the right to suspend or terminate accessibility of data (for example, the right to turn off the API or remove the file from the platform). The licensor should make all reasonable efforts to require all re-users to delete all or parts of the datasets that have been compromised (have become re-identifiable). This should include prominent notices on websites such as open data portals and forums/email lists/social media accessed by groups or individuals who are likely to be re-using the data. Requiring registration may be the most effective means of recalling datasets but this should not be encouraged if it will require the collection of new personal data from re-users and would have a general effect of discouraging use of (open access) websites and other services."

Intellectual Property Requirements

R1	National law compliance
	Open Access infrastructures and all entities involved must comply with the laws of the
	competent Member State.
R2	Determine right holder of the work
	There will be copyright on the scientific publications or papers (if held original). Several
	parties can hold this copyright: the initial author, the employer, more than one author

⁷³ WP29, *Opinion on Open Data and Public Sector Information (PSI) reuse* (5 June 2013), p. 25-26; available at: http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp207_en.pdf





	or even publishers.
R3	Establish a clear copyright notice on the OA platform.
	When the <u>copyright holder</u> publishes a copyrighted work on the <i>OA</i> platform, he or she
	has to adhere to the copyright policy of the platform. This policy should give
	information to copyright holders, authors, platform users and third parties on:
	- Intellectual property rights transfer
	- Licensing policy
	- Sufficient technical and organizational measures to avoid intellectual property
	rights infringements
	- Ethical responsibilities (i.e. no copyright infringements, no plagiarism, etc.)
	- Contact point for questions
R3.1	Clear Licensing policy
	An Open Access platform <u>must</u> establish a clear licensing policy.
	The OSL platform will use a CC BY-NC-SA 4.0 license. Copyright holders will publish
	their work on the platform under this license, which means they will directly enter into
	a contract (the CC license) with the users/re-users of their work.
	This policy should also make clear whether the license also applies to metadata.
R3.2	Third parties licensing policy
	In case an Open Access platform also uses content from third parties (such as PUBMED),
	the latter's terms of use <u>must</u> be checked to see their licensing conditions and on what
	basis the platform can use or re-use their content.
R3.3	Licensing policy – commercial use
	The licensing policy that <i>OSL</i> uses, does not support commercial use of the works, i.e.
	against payment. If the end-user desires the possibility to exploit his works
	commercially, a new license should be created between the end user and the copyright
	holder. For this, guidelines and/or a contact form should be provided that could initiate
	negotiations between the end user and the copyright holder.
R3.3.1	Disclaimer
	These guidelines or this form <u>should</u> contain a disclaimer in case the person that posted
	the work is actually not the right holder of the intellectual property rights.

1. Additional Legal Requirements





R1	Terms and conditions
	It <u>could</u> be advantageous for the <i>Open Access</i> platforms to define the terms and conditions
	to which other parties, for example the users, have to adhere to when using the platform.
	At the same time, other cooperating will likely have their own conditions and terms with
	which a platform will have to comply.
R1.2	Terms and conditions/contracts
	The Open Access platform should close contracts with related parties and/or provide
	terms & conditions they have to adhere to.
R1.3	Terms and conditions of other parties
	If the terms & conditions of the other party apply to the platform, the platform $\underline{\text{must}}$
	follow the applicable terms & conditions of the other party
	OR
	<i>The platform</i> may opt to close specific contracts with the exclusion of terms & conditions.
R2	Limitation of liability
	Platforms <u>should</u> only guarantee the provisioning of functions their system can actually
	provide.
	Platforms <u>could</u> limit their liability contractually to the extent legally allowed under
	national laws.
R3	Evidence preservation
	To be able to provide evidence in case of claims, technical or organizational measures
	should be implemented, e.g. logging





7.7 Further guidelines for Open Access Policy Makers

Take into account open access policy research

There are currently several on-going research projects concerning the open access of scientific information. For instance, research under the Pasteur4OA FP7⁷⁴ project investigates how open access policies could be improved in the future and how certain open access elements could benefit the overall openness of scientific research. Their website also provides Open Access policy guidelines for research performing organizations⁷⁵ and research funders⁷⁶.

Their Research may provide valuable insight to research institutions and funding authorities with regard to open access policies.

For Research Organizations:

http://www.pasteur4oa.eu/sites/pasteur4oa/files/resource/INSTITUTIONS POLICY%20GUID ELINES_FINAL.pdf

For Funding Authorities:

http://www.pasteur4oa.eu/sites/pasteur4oa/files/resource/Policy%20effectiveness%20-%20funders%20final.pdf

Increase the effectiveness of OA policies

Even though no standard open access policies exist, the aforementioned Pasteur4OA project did uncover some trends that may increase the effectiveness or promotion of open access in general. According to their analysis, the criteria around which open access policies must align are: a) mandatory deposit, b) impossibility of researchers to waive their deposit and c) establish links between deposit and research evaluation.

According to the first policy guideline, open access policies should strive to make the deposit of research results mandatory. Unsurprisingly, when researchers are mandated to deposit their research results, the deposit rate for Open Access items is likely to increase.⁷⁷ When Open Access deposit is made mandatory, Open Access deposit is more likely. Nevertheless, the PASTEUR

⁷⁷ Although mandatory deposit is positively correlated with both open access and restricted access.





⁷⁴ http://www.pasteur4oa.eu/home

⁷⁵ Victoria Tsoukala and Marina Angelaki, 'Open Access Policy Guidelines for Research Performing Organizations' (Pasteur4OA, September 2015)

http://www.pasteur4oa.eu/sites/pasteur4oa/files/resource/INSTITUTIONS_POLICY%20GUIDELINES_FI NAL.pdf

⁷⁶ Alma Swan, 'Open Access policy effectiveness: A briefing paper for research funders (Pasteur 4OA, September 2015)

http://www.pasteur4oa.eu/sites/pasteur4oa/files/resource/Policy%20effectiveness%20-%20funders%20final.pdf

research indicates that this also decreases the deposit of articles as Restricted Access. When an embargo exists, authors seem thus more likely to wait for the embargo to elapse, after which research results are deposited immediately.

Similarly, if authors or researchers are not given the opportunity to waive the deposit of their results, open access deposit rates are also likely to be higher. Waivers exist within open access policies to provide authors an opportunity to not deposit their scientific items on an open access basis. Usually the authors will then have to provide a reason why does not make the work freely available online. A waiver may for instance be used to accommodate the situation where the author wishes to publish his results in a journal that requires full copyright ownership.⁷⁸ When authors cannot waive open access deposits, the overall deposit rate is lower. This may be explained by the fact that authors are reluctant to ignore a publisher's open access embargo. The research also noted that "there is a small but non-significant increase in Open Access deposit rate if authors are required to retain the rights they need for Open Access and are not permitted to waive this". Because authors retain the necessary rights to make their scientific work open access, they are more confident that they do so in a legal manner.

The third guideline makes research evaluation dependent on open access deposit. As researchers need to be positively evaluated in order to get higher up on the academic ladder, in these instances, the chances of Open Access deposit are higher.

In addition to the general guidelines provided within this section. Open Access policies should also make efforts to clarify some of the aspects related to Open Access publishing. Therefore, policy makers should aim to address the frequently asked questions listed within this document.

⁷⁸ UNESCO, *Policy Guidelines fort he Development and Promotion of Open Access* (2012), p. 38; http://unesdoc.unesco.org/images/0021/002158/215863e.pdf>





8 References

Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020 available

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

OpenAIRE Guide for Project Coordinators in H2020

EC Fact sheet: Open Access in Horizon 2020, 9 December 2013

EC Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020, 16 December 2013

European IPR Helpdesk Fact Sheet Open access to publications and data in Horizon 2020: Frequently Asked Questions (FAQ)

European IPR Helpdesk fact sheet "Publishing vs. patenting"

Laakso, M. (2014). Green open access policies of scholarly journal publishers: a study of what, when, and where self-archiving is allowed. Scientometrics. In press. http://dx.doi.org/10.1007/s11192-013-1205-3

Open Access Guidelines for researchers funded by the ERC (revised in December 2014)

Open Data Online Training Material available at Open Data Online Training Material

http://opendataforum.info/files/Laatste%20docs/Open Data Handboek 20141119.pdf

Butler, Declan; *Investigating journals: The dark side of publishing*. Nature 495, 433–435 (28 March 2013)

M. Necasky, J. Pokorny, P. Moravec (Eds.): Methodologies and Best Practices for Open Data Publication Dateso 2015, pp. 52{64, CEUR-WS.org/Vol-1343 available at http://ceurws.org/Vol-1343/paper5.pdf

Share-PSI: Encouraging open data usage by commercial developers: Report (2015),

https://www.w3.org/2013/share-psi/workshop/lisbon/report

Medical Research Council: Personal information in medical research.

UK Data Archive: Managing and sharing data: a best practice guide for researchers. 2009.

N. Korn and C. Oppenheim. Licensing Open Data: A Practical Guide. http://discovery.ac.uk/businesscase/principles/





Consultation on guidelines on recommended standard licences, datasets and charging for the reuse of public sector information. http://ec.europa.eu/digital-agenda/en/news/consultation-guidelines-recommended-standard-licences-datasets-and-charging-re-use-public

LAPSI 2.0 Thematic Network, Licensing Guidelines. http://www.lapsi-project.eu/sites/la





9 Usefull links

- The DataCite organisation has a growing <u>list of repositories</u> for research data.
- The FP7 project Pasteur4OA website for more information on how to make open access policies more effective.
- The FP7 project Foster website which provides researchers with guidelines concerning open access.
- ORCID aims to provide a registry of unique identifiers and a transparent method of linking research activities and outputs to these identifiers.
- BioSharing includes a catalogue of data sharing policies and standards (reporting requirements, terminologies and exchange formats).
- Open Data Principles (project Open Data)
- Panton principles (OD in Science)
- Digital Agenda for Europe
- **UK Government Digital Strategy**
- **Open Data Policy Guidelines**
- The Next Web of Open, Linked Data
- Horizon 2020 http://ec.europa.eu/programmes/horizon2020/
- How to produce and use datasets: lessons learned at http://mlwave.com/how-toproduce-and-use-datasets-lessons-learned/
- Open access (Science in Society site): http://ec.europa.eu/research/science
- OpenAIRE http://www.openaire.eu/
- The OpenAire Guide for Copyright Issues: https://www.openaire.eu/copyright-issues
- ROARMAP (The Registry of Open Access Repository Mandates and Policies) is a searchable international registry charting the growth of open access mandates and policies adopted by universities, research institutions and research funders that require or request their researchers to provide open access to their peer-reviewed research article output by depositing it in an open access repository.





- UNESCO's <u>Policy Guidelines for the Development and Promotion of Open Access</u> (2012). This also includes example policies for funders and institutions.
- The European University Agency has developed an Open Access checklist for universities when creating open access policies: http://www.eua.be/Libraries/publications-homepage-list/Open_access_report_v3
- https://openpolicynetwork.org/resources/#studies
- http://opendataforum.info/index.php/links



