

# Publishing Open-Access Biomedical Data: Legal Challenges

**Sara Hugelier**

*Interdisciplinary Centre for Law and ICT, KU Leuven, Sint-Michielsstraat 6, Leuven 3000, Belgium*  
[www.law.kuleuven.be/icri/en](http://www.law.kuleuven.be/icri/en)

## ARTICLE INFO:

RECEIVED: 20 Jun 2014  
REVISED: 04 Sep 2014  
ACCEPTED: 12 Nov 2014  
ONLINE: 06 Jan 2015

## KEYWORDS:

Open Access  
Privacy  
Biomedical Research  
Data Protection  
Intellectual Property

## ABSTRACT

This review article gives a brief overview of some of the key legal challenges the OpenScienceLink, <http://opensciencelink.eu> (*hereafter OSL*) project faced with publishing biomedical research data online on an open access and re-use basis. Firstly, we give an overview of some of the key privacy challenges that the OSL project faced. Secondly, we analyse the intellectual property challenges that research data possesses and how this type of data could be protected, considering that purely factual elements are generally not protected under intellectual property laws. In this respect, this article also looks at a number of licensing options for these specific types of data.



Creative Commons BY-NC-SA 4.0

© 2015 Hosting by Procon Ltd. All rights reserved.

## Introduction

*The value of data lies in their use. Full and open access to scientific data should be adopted as the international norm for the exchange of scientific data derived from publicly funded research.*<sup>1</sup>

In our current ‘information economy,’ knowledge is a source of competitive advantage. Researchers, research institutions and the general public depend on the wide dissemination of innovative ideas. Depending on the novelty of this idea and the reputation of the source of circulation, this knowledge will be picked up and in turn accelerate further scientific discoveries. When these scientific discoveries are then taken up by the industry, boundless opportunities exist for further research and the development of new applications and products. Ultimately, the wide dissemination will

not only accelerate economic growth but it will also spur our understanding of society.

Unfortunately, despite the current cascade of modern research data and new emerging internet-infrastructures, in most cases the ability to access, use and re-use this data is lagging. In this respect, our global science system stands at a critical juncture.

There are a vast number of scientists and policy-makers who already grasp the importance of this data-driven research and the importance of open data, particularly regarding scientific data. Initiatives in the policy field include the OECD principles and guidelines for access to research data<sup>2</sup> and the European Commission's most recent public consultation on Open Research Data.<sup>3</sup> As the Commission observed in its Communication of 2004:<sup>4-5</sup> “Open access constitutes a key tool to foster collaborative exchanges of ideas and people in a way that catalyses science and innovation.” To

optimise this system is to catalyse economic growth.

Despite these efforts, it seems that many stakeholders have not yet caught up with the rapidly changing research environment. To date, access to raw scientific data underpinning many scientific papers is restricted or not made available at all. This tendency towards 'closed' scientific data in order to maintain professional acclaim was addressed during the 2013 stakeholder consultation mentioned above.<sup>3</sup> There it was argued that to restrict access, use and re-use of this data opposes directly the rationale of many medical trial patients, who donate data for the benefit of society rather than for the use by a closed select few. On the other hand, many medical trial patients also fear for abuse of their medical records and for their privacy. It was recently reported that in the US from Washington to New York they are putting privacy at risk by selling records that can be used to link a person's identity to medical conditions using public information.<sup>6</sup> Another issue is the 'big science' versus 'small science' gap.<sup>7</sup> At the core of knowledge advancement are still the small groups and research endeavours of single researchers while today's focus is often on 'big science.' It is crucial that small science scholars are also able to archive their data and enable other researchers to build upon their knowledge.<sup>8</sup>

Currently, one of the main obstacles pertaining to the organisation and dissemination of research data under open access condition are the legal barriers. Expanding the open data infrastructure to 'research data' presents us with a broad range of legal challenges quite distinct from those emanating from the publication and dissemination of scientific monographs or papers.

First, a serious challenge to the concept of open research data is presented by concerns regarding personally identifiable information or data. Regarding this issue, the open research data infrastructure stands at a crux. On the one hand, data anonymization is currently our most important safeguard in protecting the patients who contribute to the research or trials. On the other hand, it has been argued that the greater and more in-depth the access to the data, the higher the utility of this data and, consequently, higher the chances of spurring scientific progress.<sup>9</sup> Protecting the privacy of patients thus comes at a scientific cost. Hence, in an open access regime, the core challenge will be not to overstretch the data protection rules while at the same time refraining from unduly restricting the concept of personal data. This is a difficult balance to strike, but one,

which is pivotal to the development of an open research data infrastructure.

Second, other daunting challenges arise from the intellectual property management of open research data. It needs to be ascertained whether the research data would qualify for intellectual property rights protection. In principle, neither EU copyright law nor US copyright law extends copyright protection to facts per se. In general, it is required that the work distinctively be the product of its author's intellectual efforts.<sup>10</sup> For instance, as stated in *Feist*,<sup>11</sup> the sine qua non of US copyright is originality. Another example is the German legislation, where the work has to be a 'personal intellectual creation.'<sup>12</sup> Where is the originality in research facts? Moreover, should compilations of research data be protected by an intellectual property right, does this not conflict with the whole rationale of open data?

Another challenge rises from the compilation of different types of data. It is possible that some data is protected under intellectual property rights, for instance the database that holds the datasets, and that other data is not protected at all. To whom do the respective rights and obligations of these data belong and on which data?

Finally, sharing research data on an open-access basis also presents us with a number of ethical challenges. These issues include the unintended secondary use, misappropriation and commercialisation of research data, unequal distribution of scientific results and disproportionate impacts on scientific freedom as well as other economic, social and scientific costs.<sup>13</sup>

This paper will give the reader a 'basic' overview of some of the key legal challenges that were encountered when setting up the OSL platform, <http://opensciencelink.org>. This editorial constitutes in no way a comprehensive overview presenting clear-cut legal solutions.<sup>14</sup>

## Research Data

The first essential question that has to be tackled when looking at the different EU policies, regulations and initiatives concerning the Open Access sharing of research data is how to define 'research data.' On 2 July 2013, the European Commission held a public consultation on open research data in Brussels. This conference was attended by a variety of stakeholders from the research community, industry, funders, libraries, publishers, infrastructure developers and others. One of the first questions on which the debate focused was: "How can we define research data and what types of data should be open?"

This question was addressed by all groups of contributors and the definitions that were given varied accordingly. In general, researchers agreed that research data constitute all data from an experiment, study or measurement, including the metadata and processing details. It was also acknowledged by some that some data should be restricted, such as sensitive data and commercial data. A specific example constitutes that of genome sequencing. Both the European Public Health Alliance (EPHA) and the International Council for open Research and Open Education (ICORE) supported this broad definition of research data.

Given the fact that the OSL project focuses for the most part on researchers as their target audience for the platform, we will use this definition when talking about research data further in this paper. In particular, we focused on biomedical research data.

### **Data protection and privacy concerns**

Despite a desire to see innovation and progress in research, the public remains very concerned about potential loss of privacy. How to use and disseminate data taken from human subjects during clinical or psychological research has long been the focus of debates.<sup>15</sup> Over the years we learned that the confidentiality of personal data is not paramount in all circumstances. Sometimes, it may be appropriate and lawful to provide open access to personal data. However, in all cases of processing personal data the requirements of international, European and national data protection laws must be met.

A number of approaches exist when it comes to sharing scientific data. Traditionally, access to personal data should be restricted with a presumption of non-disclosure. An example of this is the classic doctor-patient relationship. This obviously constitutes a major impediment to data sharing.

Another approach is the balancing of confidentiality with socially useful sharing and disclosures of data using generally accepted rules for doing so. In this approach the subject of the research data is usually involved in the data sharing decision via their consent. The consent should be an explicit and informed one, so that the patient is aware of the purpose of the research, identification of the data controller(s), the duration of the research project, and identification of the processor(s).

Currently, knowing that biomedical research is being transformed substantially through the ap-

plication of information technologies that allow for data to be shared on an unprecedented scale, it is believed that these approaches described above do not keep in pace with the changes in our research capability. In an era when information is shared digitally at the global level, mechanisms of informed consent remain static, paper-based and organised around national boundaries and legal frameworks.<sup>16</sup> Therefore, researchers have developed an alternative approach: dynamic consent. This is a bi-directional, ongoing, interactive process between patients and researchers. The benefit of this interface is that it enables individuals to exercise their autonomy by giving informed consent for new types of research in real time rather than being asked to give a broad consent at the beginning of the research process. The benefits for the research process are that recruitment is easier, less costly and more efficient; the legal and ethical requirements of consent can be met with ease; there is greater transparency and accountability in the research process and research findings can be returned to research participants as part of a personalised medicine approach. Dynamic consent has the potential to enhance patient confidence and enable long-term patient-researcher collaborations in research.<sup>17</sup>

Overall, choosing to release biomedical research data containing personal data for open access purposes, keeping into account article 7 and 8 of the EU Data Protection Directive, can despite the above-described approaches still be a lengthy and cumbersome process. Large research projects that rely on the permission to obtain certain data and the permission to publish this data will occupy much of a researcher's time, and it will take as much time for the data to be released.<sup>18</sup> When the datasets also concern sensitive data such as medical records, it should even be considered whether it is really appropriate and practicable to meet data protection requirements and provide open access to personal data.<sup>19</sup>

Ultimately, the OSL platform choice for a fourth approach: the anonymization of data, especially in the light of its open access purpose and the scope of the platform, i.e. biomedical research data. 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, such as informed consent, are no longer required in this situation.<sup>20</sup>

Many authors have argued that today however, in our "big data" era, anonymization is no longer a solution nor is our individual notice and

consent (the privacy statement).<sup>21</sup> Anonymization works in a world of “small data” but with big data, re-identification is increasingly easy. Examples are not too hard to find. AOL researchers released a massive dataset of search queries, anonymized by scrubbing user IDs and IP addresses. Later, Netflix provided a huge database of movie recommendations available for study with the personal identifiers removed. Despite scrubbing the identifiable information from the data such as username and IP address, computer scientists were able to identify individual users in both datasets.<sup>22</sup> Ultimately, it often comes down to a judgment call as to whether the data is truly anonymized. Admittedly, anonymization might not be fully risk-proof but as long as our legislation is not up-to-speed with the latest developments in our high-tech society, it remains a good option if combined with a good sense of judgment and expert advice.<sup>23</sup>

### Intellectual property rights management

When we talk about biomedical research data, we talk about bare ‘facts.’ This is a subject matter, which is generally not protected by any intellectual property legislation. However, a compilation of data, which is a product of intellectual creation, might be entitled to protection. It is the form that will qualify for protection under certain circumstances; the content i.e. the ‘bare facts’ will still be free. This is logical to a certain extent. Since researchers and scholars do not have exclusive rights to utilise the data that they discovered or collected, they cannot monopolize ‘facts.’

The collection of research data can fall under two levels of protection:

- A. Intellectual property rights can attach to the original expression of the facts in a particular form via copyright.
- B. Collections of scientific data can also be protectable under the European *sui generis* database right if the maker of the database showed a substantive investment.<sup>24</sup>

#### A. Copyright

To obtain copyright protection, an “intellectual creation” is required. Copyright law will only protect an “original expression” or works that show ‘creativity’ or ‘originality.’ Scientific publications virtually always attract copyright protection under copyright legislation. Writing a book or a scientific article almost always involves personal creativity. The individual research data however and the datasets containing them prove to be more complex.

A database which is defined in the legislation as “a collection of independent works, data or other materials which are arranged in a systematic or methodical way and are individually accessible by electronic or other means” can be protected by copyright but only the database as such, not the individual content.<sup>25</sup>

According to Article 3 of the Database Directive, for a database to receive legal protection it must be ‘original,’ i.e. the author’s ‘own intellectual creation’ by reason of the selection or arrangement of the contents.<sup>26</sup> This level of ‘originality’ is the same as in Article 1 (3) of the Software Directive and Article 6 of the Terms of Protection Directive.<sup>27</sup> This standard clearly constitutes a typical European compromise criterion out of the standards in the different Member States, ranging from ‘skill and labour’ in the UK,<sup>28</sup> through the requirement of the ‘individual character and the personal stamp of the author’ in France and Belgium, to a demanding requirement of the ‘print of the author’s personality that rises above average’ in Germany.<sup>29</sup> Considerable variety exists in the national Courts’ approaches to the requirement of originality.<sup>30</sup>

Whether collections of biomedical scientific research data such as the ones in OSL will meet the criterion of ‘originality’ depends on the interpretation of the various national courts. Essential for copyright to apply is that the databases show a certain character or a certain creative or intellectual effort by the author. A database selection or arrangement which purely depends on technical factors or imperatives of accuracy and exhaustiveness will not be protected by copyright. There is no room for the author to exercise creativity or originality in the choice, sequence and combination of data in the collection.<sup>31</sup>

The OSL database arrangement is effectively based on the standard scientific practice, the database will thus most likely fail to meet these above described requirements of “intellectual creation.” It is unlikely that the database will be protected by copyright legislation. Therefore, the *sui generis* database right will be the most important one for the OSL project and databases in the field of research data in general. Nonetheless, copyright legislation will still be of importance when looking at e.g. the full text articles, papers and tables within the OSL database.<sup>32</sup>

The following actions fall under copyright protection.<sup>33</sup>

- *Reproduction*: The right of the copyright owner to prevent others from making copies of his works without his authorization is the most basic

right protected by copyright legislation. A reproduction can be temporary or permanent, partially or the whole work, in analogue or digital form. Examples include the printing of the data, the copying of the data on your computer, on CDs, DVDs, USB-sticks but it does not include the mere viewing of the data on the OSL platform.

- *Distribution*: In parallel with the right to reproduction, the copyright owner can grant users the distribution of copies made with the copyright owner's consent. It is essential that the data be given to users on 'material storage media' for the right of distribution to come into play. The right of distribution usually is terminated upon first sale or transfer of ownership of a particular copy.

- *Communication to the public*: The concepts of communication and public are to be interpreted in a broad manner, insomuch as public communication can exist solely by making available the work or service to an indeterminate number of potential users in such a way that they may access it, even if they do not make use of that possibility. Hence, the creation of a link that can be clicked and that can redirect a user to a protected work or service constitutes an act of making available to the internet public and, as such, an act of communication to the public.<sup>34</sup>

- *Adaptation*: Adaptation is generally understood as the modification of a work, for example, as far as the OSL database would be copyright protected, the authors of the database have the right to authorise any adaptation of this database. When it comes to the individual elements in that database (in the case of raw data, this is unlikely, *supra*), it will depend on national copyright legislation whether an adaptation is allowed or not.

### B. *Sui generis* database protection

The *sui generis* database right does not, unlike copyright, protect the 'original' result of an 'intellectual creation.' When developing the Database Directive, the European authorities were aware that the originality requirement would always remain a major impediment for the protection of a lot of databases. Most databases take money, time and effort to create and maintain, but in most cases they do not require 'originality.' The database right is specifically intended for someone who invested substantially to recoup his investment by exploiting the database.<sup>35</sup>

It is with these non-original databases in mind which entail 'a substantial investment' of money, time and effort that the European authorities developed the *sui generis* database right. The *sui generis* protects in fact 'the sweat of the brow' of

the database producer, rather than its 'originality.' According to recital 7 of the Directive, the *sui generis* right was developed because "the making of databases [requires] the investment of considerable human, technical and financial resources while such databases can be copied or accessed at a fraction of the cost needed to design them independently."

The *sui generis* right works independently from the copyright protection of the database itself as well as the copyright of its content. A database can at the same time enjoy the protection of copyright as well as the *sui generis* right. In case the content of the database enjoys copyright protection, then the database maker should ensure that he has the authorisation of the copyright owners to use these works for the purpose of the database.

The criterion for databases to be eligible for *sui generis* protection is the 'substantial investment' made by the database producer. The Directive does not provide many indications as to when a database involves a substantial investment. According to Article 7 of the Directive the database must show a qualitatively and/or quantitatively substantial investment in the obtaining, verification or presentation of the contents. In practice, most databases will probably result from a quantitative investment, involving the deployment of financial resources and/or the expanding of time, effort and energy (recital 40 of the Directive).<sup>36</sup>

According to Vanhees, the development and commercialization of a database demands a considerable investment if: (1) data has to be collected; (2) permission has to be obtained for their storage and use in the database; (3) the collected data has to be selected, processed and organized; (4) instruments have to be made to find separate elements of the database; and (5) the contents of the database have to be verified and updated.<sup>37</sup>

A database comprising biomedical research data will not necessarily involve a substantial investment within the meaning of the European Database Directive. In a number of landmark cases it was decided that 'obtaining' excludes the costs incurred in the creation of new data (such as generating fixtures lists) from being considered relevant to satisfy this requirement.<sup>38</sup> The ECJ in fact discounts any investment in collecting data that is indivisibly linked to its creation.<sup>39</sup> What can be taken into account are the costs necessary for the verification of the accuracy of the data and for the presentation of such data to third party users.

Of course, when developing a research database, considerable research was necessary to collect the research data, but it is the database itself

which requires the investment. For databases to fall within the scope of the *sui generis* right, the sole criterion will be whether a substantial investment was made in “obtaining, presenting and verifying” the data.<sup>40</sup> This means that collecting and presenting the research data should not merely be a by-product of the time and effort that went into “creating” the data.<sup>41</sup>

As far the OSL database is protected by *sui generis* database rights, the following rights of the right holder are exclusively protected:

- *Extraction*: The database right holder has the exclusive right of retraction, which is defined as “the permanent or temporary transfer of all or a substantial part of the contents of a database to another medium by any means or in any form.” The right pertains to the downloading, copying, printing, or any other reproduction in whatever (permanent or temporary) form. Important to note here is the fact that also ‘temporary’ transfers are included under the exclusive rights of the database-right-holder.

- *Re-utilization*: The database-right-holder also has the exclusive right of ‘re-utilization.’ This is defined in Article 7 as “any form of making available to the public all or a substantial part of the contents of a database by the distribution of copies, by renting, by on-line or other forms of transmission.”<sup>42</sup>

### C. Expected usages of the OSL platform

The OSL project essentially created an integrated e-infrastructure, the OSL platform, where biomedical research data linked with the relevant publications is published digitally and is openly accessible after registration by any user. As part of this platform, OSL offers dataset publications including an abstract, keywords and metadata. The data file can both be a data repository in the web (provided by a link) or provided as a table in excel spreadsheet format or other software. The metadata file will be a detailed machine-readable file that provides a structured description of the dataset, including key features of the experimental samples and the techniques used to generate the data (e.g. organism, tissue source, technologies).

Copyright protection will be particularly relevant for the scientific publications, articles and probably also the abstracts. The respective databases will usually be protected by the *sui generis* database right.

Within the OSL platform, we expect a number of specific usages by the registered users as well as the data controllers:<sup>43</sup>

- *Access*: When it comes to copyright protectable data, the act of accessing could infringe upon the right of reproduction, the right of distribution as well as the right of communication to the public (*supra*). A mere onscreen view is not enough, however, printing or copying data in the working memory of the computer could qualify as an infringement on these exclusive rights of the copyright-holder. When it comes to the *sui generis* database right protected databases, the right-holder needs to consent to opening up the data on an open-access basis. This is especially relevant for the copying of or making available databases/datasets that belong to external third parties.

- *Linking*: The OSL platform will not only contain uploaded datasets, metadata and publications, but in some cases it will also provide hyperlinks towards other scientific publications and selected datasets. National courts are generally divided as to whether hyperlinking constitutes a communication to the public. However, on the 13<sup>th</sup> of February 2014, the European Court of Justice (ECJ) ruled the following: “[...] *the provision on a website of clickable links to works freely available on another website does not constitute an ‘act of communication to the public,’* [...]” Basically, the ECJ ruled that an ordinary “clickable” hyperlink makes a work available to the public. However, if the link is to a publically available portion of a website used by the rights holder to make work available to the same public as the link, it is not made available to a new public and the right is not infringed.<sup>44</sup>

- *Mining*: Text and data mining activities may infringe the rights owner’s copyright and/or database right if done without prior authorisation. There is a research exception contained in the Database Directive as well as in the Information Society Directive but this exception has not been implemented properly in all Member States which creates substantial uncertainty within the European research community. This will also bring about negative consequences for researches to engage in mining activities on a cross-border basis.<sup>45</sup>

- *Re-use*: The main objective of the OSL platform is to allow the scientific community to share biomedical research datasets and publications in order to enhance research results and stimulate innovation. This means that researchers might want to modify existing research datasets or add their own research results to an existing dataset. If the data could qualify as copyright protectable, then the re-use could infringe upon the right of reproduction, public communication as well as upon the right of adaptation. First, the right of re-

production will be infringed from the moment the relevant datasets are merely loaded into the working memory of a computer without the consent of the right owner. Second, since the goal of OSL is to publicly communicate biomedical research data, OSL could infringe upon third party's copyright by making external copyright-protected data available to the public. Third, the right of adaptation could be infringed when users would modify the datasets or publications. This is however not the aim of the OSL platform itself. The *sui generis database* right will be infringed if a substantial part of the database or the database itself is transferred or re-used. So, the data should not be copied or made available publicly without consent of the rights owner.<sup>46</sup>

### Licensing framework

Research data is usually not protected by copyright legislation. However, the databases containing the data and datasets as well as the publications that are linked to this data will be protected by the *sui generis database* right – if a sufficient substantial investment has been made and the respective copyright legislation treats that accordingly.

After examining the different types of expected usages of the OSL platform from registered users and data curators, it became clear that a number of these actions could potentially infringe intellectual property legislations. This is for example the case in the situation where users substantially copy the OSL database or where OSL would link to a database which is not yet publically available, and without the rights owner's consent. This is also the case for most text and data mining activities even within the research community.

Allowing these activities to take place for research purposes without infringing intellectual property rights could be achieved in several ways. The OpenAire project,<sup>47</sup> as well as the Expert Group on text and data mining,<sup>48</sup> already presented the European legislator with a number of suggestions regarding legislative action, i.e. introduction of a new and broader mandatory research exception at a European level or through a revised normative interpretation of the 'reproduction' right.

For the OSL platform to work and research datasets to be uploaded without infringing intellectual property legislations, the OSL platform decided to use a well-known licensing agreement, i.e. the Creative Commons 4.0 license.<sup>49</sup> Unlike its predecessor, the Creative Commons 3.0 version, CC 4.0 does include the *sui generis* database right

in its scope. CC 4.0 also allows for a maximum of legal interoperability with machines and above all, it is a cross-border license which is well known all over the world. The OSL consortium included the database and all of its content in the CC 4.0 license in order to avoid intellectual property infringements or any legal uncertainty about the use and re-use of the research data. In order to promote and push for the open exchange of biomedical research data within the scientific community, it is essential that projects like OSL have policies that clarify the legal status of the data that is published while allowing society to gain full benefit from this data.<sup>50</sup>

### Acknowledgements

The preparation of this manuscript was supported by the EU-funded project OpenScienceLink (Grant agreement # 318652).

### References

- <sup>1</sup> National Research Council, Committee on Issues in the Transborder Flow of Scientific Data. *Bits of Power: Issues in Global Access to Scientific Data*. Washington, D.C.: National Academy Press; 1997. Quote on p. 10.
- <sup>2</sup> OECD. *OECD Principles and guidelines for Access to Research Data from Public Funding*. 2007.
- <sup>3</sup> European Commission, *Public Consultation on Open Research Data*. 2013. Available at <http://ec.europa.eu/digital-agenda/en/news/ec-public-consultation-open-research-data-2-july-brussels>.
- <sup>4</sup> European Commission (2007). *Communication from the Commission to the European Parliament, the Council and the European Economic and Social Committee on scientific information in the digital age: access, dissemination and preservation*, p.56.
- <sup>5</sup> *Communication from the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: a reinforced European Research Partnership for Excellence and Growth*, COM (2012) 392. *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions towards better access to scientific information: boosting the benefits of public investments in research*, COM(2012)401 and *Commission recommendation on access to and preservation of scientific information*, COM (2012) 4890.
- <sup>6</sup> Robertson J. *Your Medical Records Are for Sale*. *Bloomberg Business Week*. 2013 Aug 8. Available from: [www.businessweek.com/articles/2013-08-08/your-medical-records-are-for-sale](http://www.businessweek.com/articles/2013-08-08/your-medical-records-are-for-sale).
- <sup>7</sup> Onsrud H, Campbell J. *Big opportunities in access to small science data*. *Data Science Journal*. 2007;6: 58-59.

Available from: [http://works.bepress.com/harlan\\_onsrud/5/](http://works.bepress.com/harlan_onsrud/5/).

<sup>8</sup> Swan A, Brown S. To share or not to share: publication and quality assurance of research data outputs. Research Information Network, 2008 Jul. Available from: [http://eprints.soton.ac.uk/266742/1/Published\\_report\\_-\\_main\\_-\\_final.pdf](http://eprints.soton.ac.uk/266742/1/Published_report_-_main_-_final.pdf).

<sup>9</sup> Ohm P. Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. *UCLA Law Review*. 2009; 57(6): 1701-77. Available from: <http://www.uclalawreview.org/?p=1353>.

<sup>10</sup> Goldstein P. *International copyright: principles, Law and Practice*. Oxford: Oxford University Press; 2001. p. 161.

<sup>11</sup> US Supreme Court (1991). *Feist Publications, Inc. v. Rural Telephone Service Co.*

<sup>12</sup> Germany, Copyright Act §2(2) (1993).

<sup>13</sup> These ethical challenges were presented in more detail in Deliverable 1.3: "Legal and ethical issues in open access and data dissemination and preservation" of the RECODE Project, GA # 321463. Available from: <http://www.recodeproject.eu>.

<sup>14</sup> For more detailed information regarding the legal framework of OSL see Deliverable 3.2 of OpenScienceLink, Grant agreement # 318652. Available from: <http://opensciencelink.eu>.

<sup>15</sup> Olsen S. *Sharing clinical research data: a workshop summary*. Washington D.C.: National Academies Press; 2001.

<sup>16</sup> Kaye J, Whitley EA, Lund D, Morrison M, Teare H and Melham K. Dynamic consent: a patient interface for twenty-first century research networks. *Eur J Hum Genet*. 2014 May 7. E-pub ahead of print. DOI: 10.1038/ejhg.2014.71.

<sup>17</sup> Kanellopoulou NK. Rapid response to "Broad consent is informed consent" by Sheehan M. *BMJ*. 2011;343. DOI: 10.1136/bmj.d6900. See also Trinidad SB et al. Research Practice and Participant Preferences: The Growing Gulf. *Science*. 2011 Jan 21;331(6015):287-8. DOI: 10.1126/science.1199000.

<sup>18</sup> Peto J, Fletcher O, Gilham C. Data protection, informed consent, and research. *BMJ*. 2004 May 1;328(7447):1029-30. DOI: 10.1136/bmj.328.7447.1029.

<sup>19</sup> Anonymisation: managing data protection risk - code of practice. London: Information Commissioner's Office;2012. Available from: [www.ico.org.uk/~media/documents/library/Data\\_Protection/Practical\\_application/anonymisation\\_code.pdf](http://www.ico.org.uk/~media/documents/library/Data_Protection/Practical_application/anonymisation_code.pdf).

<sup>20</sup> 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 controller again has to rely on article 7 or 8 of the Directive for legitimate grounds of processing the data.

<sup>21</sup> Mayer-Schönberger V, Cukier K. *Big Data: a revolution that will transform how we live, work and think*. Boston: Eamon Dolan;2014. See also Ohm, 2009.

<sup>22</sup> Barbaro M, Zeller T. A face is exposed for AOL searcher no. 4417749. *New York Times*. 2006 Aug 9. Available from: <http://www.nytimes.com/2006/08/09/technology/09aol.html?ex=1312776000>. See also Ohm, 2009, Mayer-Schönberger and Cukier, 2014, and Narayanan A, Shmatikov V. Robust de-anonymization of large sparse datasets. In: 2008 IEEE Symposium on Security and Privacy;2008 May 18-21; Oakland, CA. Available from: [http://www.cs.utexas.edu/~shmat/shmat\\_oak08netflix.pdf](http://www.cs.utexas.edu/~shmat/shmat_oak08netflix.pdf).

<sup>23</sup> For example, the UK Information Commissioner's Office developed an anonymization code of practice, which provides a good amount of examples and recommendations as to how to substantially reduce any risks (see reference 19).

<sup>24</sup> Article 7 of the Database Directive.

<sup>25</sup> Article 1 (2) of the Database Directive.

<sup>26</sup> Article 1 of the Database Directive and C-05/08 *In-fo-paq International A/S v DanskeDagbladesForening*; C-393/09 *Bezpečnostnísoftwarovaasociace v. Ministerstvo kultury* C-145/10 *Eva mariaPainer v. Standard Verlag GmbH* and C-604/10 *Football Dataco v. Yahoo UK Ltd*.

<sup>27</sup> Hugenholtz PB. Implementing the European Database Directive. In: Kabel JJC, Mom GJHM, editors. *Intellectual Property and Information Law, Essays in Honour of Herman Cohen Jehoram*. The Hague: Kluwer Law International;1998.

<sup>28</sup> *Maple & Co v. Junior Army and Navy Stores* [1882] 21 Ch.D. 369 and *Copinger* (13th Ed.) §§ 3-36. See also *Victoria Park Racing and Recreation Grounds Co. Ltd v. Taylor* [1937] 58 C.R.R. 479 at 511 (decision of the Australian High Court).

<sup>29</sup> *Inkassaprogram— Bundesgerichtshof*, 9 May 1985. See also Hugenholtz (1998), p.187.

<sup>30</sup> Tritton G. *Intellectual Property in Europe*, Suppl. 1. London: Sweet & Maxwell;1997.

<sup>31</sup> Guibault (2013), p.3.

<sup>32</sup> Dietrich N, et al. *Study on Licensing of Publications and Research Data: Summary of Findings*. OpenAire;2013 Dec 19. Available from [www.openaire.eu](http://www.openaire.eu).

<sup>33</sup> *ibid*. See also revision 2 of OSL Deliverable 3.2 and WIPO, *Understanding Copyright and Related Rights*. World Intellectual Property Organization, pp.8-9. Available from: [http://www.wipo.int/edocs/pubdocs/en/intproperty/909/wipo\\_pub\\_909.pdf](http://www.wipo.int/edocs/pubdocs/en/intproperty/909/wipo_pub_909.pdf).

<sup>34</sup> ECJ C-403/08, *Football Association Premier League*, 4 October 2011 and ECJ C-607/11, *ITV Broadcasting*, 7 March 2013.

<sup>35</sup> Janssen K, Dumortier J. The protection of maps and spatial databases in Europe and the United States by copyright and the sui generis right. *J Marshall J Computer & Info L*. 2006;24(2):Article 2.

<sup>36</sup> Hugenholtz, 1998.



<sup>37</sup> Vanhees H. De juridische bescherming van databanken. Antwerpen: Kluwer; 2001. See also Janssen and Dumortier, 2006.

<sup>38</sup> C-388/02 Fixtures Marketing and C-203/02 British Horseracing board Ltd.

<sup>39</sup> C-46/02 Fixtures Marketing Ltd v OyVeikkaus Ab. See also Davison MJ, Hugentholtz PB. (2005), Football fixtures, horseraces and spin-offs: the ECJ domesticates the database right. *European Intellectual Property Review*. 2005;27(3): 113-118.

<sup>40</sup> C-203/02 British Horse Racing Board Ltd. And others v. William Hill Organisation Ltd.

<sup>41</sup> De Cock Buning M, Ringnalda A, Van der Linden T. The legal status of raw data: a guide for research practice. Utrecht: SURFfoundation; 2009.

<sup>42</sup> Article 7 of the Database Directive.

<sup>43</sup> These types of usages are also described in much more detail in revision 2 of Deliverable 3.2 of the OpenScienceLink project as well as in the OpenAire study on licensing of publications and research data.

<sup>44</sup> Sookman B. Barry Sookman blog [Internet] When hyperlinks infringe copyright: Svensson v Retriever Sverige. 2014 Feb 13. Available from:

<http://www.barrysookman.com/2014/02/13/when-hyperlinks-infringe-copyright-svensson-v-retriever-sverige/>.

<sup>45</sup> European Commission, Standardisation in the area of innovation and technological development notably in the field of Text and Data Mining, Report from the Expert Group. Brussels: European Commission;2014. pp. 52-53. DOI: 10.2777/71122.

<sup>46</sup> These types of usage are also described in much more detail in revision 2 of Deliverable 3.2 of the OpenScienceLink project as well as in the OpenAire study on licensing of publications and research data.

<sup>47</sup> Via the OpenAire study referenced above.

<sup>48</sup> European Commission, 2014.

<sup>49</sup> Creative Commons. Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0). Available from: <https://creativecommons.org/licenses/by-nc-sa/4.0/>.

<sup>50</sup> Quote from Finn R, et al. Legal and ethical issues in open access and data dissemination and preservation. RECODE project, Deliverable 3.1. 2014 Apr. Available from: <http://recodeproject.eu/wp-content/uploads/2014/05/D3.1-legal-and-ethical-issues-FINAL.pdf>.