

Licensing Research Data under Open Access Conditions

Dr. Lucie Guibault

Chapter to be published in: D. Beldiman (ed.), *Information and Knowledge: 21st Century Challenges in Intellectual Property and Knowledge Governance*, Cheltenham, Edward Elgar, upcoming 2013.

1. Introduction

Innovative scientific research plays a crucial role in addressing global challenges, such as healthcare, environmental, and security issues, while research in social sciences and the humanities occupies a key function in understanding emerging social phenomena. The speed and depth of scientific research, understood in its broadest sense, depends on fostering collaborative exchanges between different communities and assuring its widest dissemination. This, in turn, is fundamental for the constant evolution of science and human progress. Access to research data not only increases the returns from public investment in this area, but it also reinforces open scientific inquiry. It encourages diversity of opinions, promotes new areas of work and enables the exploration of topics not envisioned by the initial investigators. Collaborative exchanges help avoid unnecessary duplication of research and gives insight into the methodology followed. Timely and cost efficient access to scientific research therefore contributes in increasing the general economic and social welfare.¹

On July 17, 2012, the European Commission published its Communication to the European Parliament and the Council entitled 'Towards better access to scientific information: Boosting the benefits of public investments in research'.² As the Commission observes, 'discussions of the scientific dissemination system have traditionally focused on access to scientific publications — journals and monographs. However, it is becoming increasingly important to improve access to research data (experimental results, observations and computer-generated information), which form the basis for the quantitative analysis underpinning many scientific publications'.³ The Commission believes that through a more complete and wider access to scientific publications and data, the pace of innovation will accelerate, researchers will collaborate so that duplication of effort will be avoided. Moreover, open research data will allow subsequent researchers to build on previous research results, as it will involve citizens and society in the scientific process.

The Communication marks an official new step on the road to open access to publicly funded research results in science and the humanities in Europe. Scientific publications are no longer the only elements of an open access policy: research results upon which publications are based must now also be made available to the public. To implement this policy, the European Commission set up a pilot initiative on Open Access (OA) to peer reviewed research articles in its Seventh Research Framework Programme (FP7), otherwise known as the OpenAire project, to ensure that the results of the research it funds are disseminated as widely and effectively as possible to guarantee maximum exploitation and impact in the world of researchers and beyond.⁴ OpenAire pursues three objectives: to

¹ OECD Principles And Guidelines For Access To Research Data From Public Funding, Paris, OECD, 2007; Communication of the European Commission on scientific information in the digital age: access, dissemination and preservation, COM(2007)56, Brussels, 14.2.2007.

² Brussels, 17.7.2012 COM(2012) 401 final

³ Id., p. 3.s

⁴ <https://www.openaire.eu/en/support/faq>

build support structures; to establish and operate an electronic infrastructure; and to manipulate research datasets. All of this under open access conditions.

As noble as the open access goal is, however, the expansion of the open access policy to publicly funded research data raises a number legal and policy issues that are often times distinct from those concerning the publication of scientific articles and monographs. Since open access of research data – unlike publications – is a relatively new policy objective, less attention has been paid to the specific features of research data. This paper therefore gives an overview of the main legal issues and policy considerations involved in licensing research data under open access conditions from a European perspective.

The paper is further divided in two parts. First, it examines the rights status of research data, which, contrary to most scientific publications that are eligible to copyright protection, is not always easy to ascertain. As a consequence, the respective rights and obligations of the parties involved in producing and disseminating the research output are unclear. Second, should collections of scientific data be indeed protected by an intellectual property right, under which licensing terms and conditions should the research institution best release the data to the public? Would existing standard form licences, like the Creative Commons licensing suite, be adapted to the specific features of research data? Would their use meet the objectives of open access and the special needs of all parties involved? For the purposes of this paper, ‘research data’ is to be understood in its broadest sense as any kind of data produced as output of research in science and the humanities, including experimental data, survey data, tables/charts, graphics, metadata etc. Note that although privacy issues form an important aspect of the publication of research data, these are left to other more knowledgeable authors in the field and remain outside the scope of this paper.

2. ‘Research data’ and Intellectual Property Protection

2.1. Copyright

Whereas scientific publications virtually always attract copyright protection under the copyright laws of the Member States of the European Union, the individual research data and the datasets containing them may not so easily fall under the copyright regime. Since copyright does not protect mere facts and ideas, but rather attaches to the original expression of ideas, research data is not likely to qualify as protectable subject matter for lack of originality. The concept of originality in copyright law has been harmonized at the European level with respect to software⁵, databases⁶ and photographs⁷, criterion which was recently extended to all kinds of works through the interpretation of the Court of Justice of the European Union (CJEU).⁸ The *Infopaq* decision⁹, as later followed in *Bezpenostní*

⁵ Directive 2009/24/EC of the European Parliament and of the Council of 23 April 2009 on the legal protection of computer programs (Codified version) (Text with EEA relevance) OJ L 111, 5.5.2009, p. 16–22, art. 1(3).

⁶ Directive 96/9 of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, OJ L 77, 27.3.1996, p. 20–28, art. 3(1).

⁷ Directive 2006/116/EC of the European Parliament and of the Council of 12 December 2006 on the term of protection of copyright and certain related rights (codified version), OJ L 372, 27.12.2006, p. 12–18, art. 6.

⁸ M. van Eechoud, *Along the Road to Uniformity - Diverse Readings of the Court of Justice Judgments on Copyright Work*, *JIPITEC : Journal of Intellectual Property, Information Technology and E-Commerce Law*, 2012-1, p. 60-80.

⁹ *Infopaq International A/S v Danske Dagblades Forening*, Case C-5/08, Judgment of the Court, 16 July 2009.

*softwarová asociace*¹⁰ and other subsequent cases¹¹, established that a work is original if it is the 'author's own intellectual creation'. In the *Infopaq* decision, the Court further clarified that the originality of a work must be assessed through its components:

*Regarding the elements of such works covered by the protection, it should be observed that they consist of words which, considered in isolation, are not as such an intellectual creation of the author who employs them. It is only through the choice, sequence and combination of those words that the author may express his creativity in an original manner and achieve a result which is an intellectual creation.*¹²

To be eligible for copyright protection, collections of data, tables and compilations must therefore show a sufficient degree of originality in their selection and arrangement.¹³ Whether collections of scientific research data meet the criterion of originality is a question of fact to be determined on a case-by-case basis. However, if the selection and arrangement of the contents of a scientific database are dictated by technical factors or imperatives of accuracy and exhaustiveness, then the author can exercise little to no creativity or originality in the choice, sequence and combination of the data in the collection. Scientific databases are therefore in most cases not likely to meet the threshold for copyright protection.

2.2. Sui Generis Database Right

Collections of scientific data may be protectable subject matter under the European *sui generis* database right, however. Through article 7 of the Database Directive, as implemented in the legislation of the Member States, the maker of a database showing a substantial investment (assessed qualitatively and/or quantitatively) in either the obtaining, verification or presentation of its contents has the exclusive right to prevent the extraction and/or re-utilization of the whole or of a substantial part, evaluated qualitatively and/or quantitatively, of the contents of that database. Like copyright protection, the *sui generis* database right arises automatically, without any formal requirement, at the moment the database is completed or disclosed to the public.

In the context of scientific research, the wording of article 7 of the Database Directive raises two main questions: first, does the scientific database show a substantial investment relating to the 'obtaining, verification or presentation' of the content of that database? And second, should the answer to the first question be positive, does a research institution or the individual researcher qualify as the maker of the database, and therefore as a rights owner, if such database has been put together through public funding?

2.2.1. Substantial investment

With respect to the first question, the terminology of the Database Directive has given rise to numerous legal disputes before the national courts, which have led in turn to a number of

¹⁰ *Bezpečnostní softwarová asociace v. Ministerstvo kultury*, C-393/09, Judgment of the Court (Third Chamber) of 22 December 2010.

¹¹ *Eva Maria Painer v. Standard Verlag GmbH*, C-145/10, Judgment of the Court (Third Chamber), 1 December 2011; *Football Dataco v. Yahoo UK Ltd.*, C-604/10 Judgment of the Court (Third Chamber), 1st March 2012.

¹² *Infopaq International A/S v Danske Dagblades Forening* Case C-5/08, Judgment of the Court (Third Chamber), 16 July 2009, para. 45.

¹³ T.-E. Synodinou, The Foundations of the Concept of Work in European Copyright Law, in: T.-E. Synodinou (ed.), *Codification of European Copyright Law – Challenges and Perspectives*, The Hague, Kluwer Law International, 2012, pp. 93-113, p. 101.

references to the CJEU for preliminary rulings on the interpretation of the Directive's provisions. Note that so far disputes have involved almost exclusively commercial entities, with here and there a public entity, but only one decision involved an educational institution.¹⁴ Regarding the interpretation of the criteria for protection of article 7, the CJEU has given a narrow interpretation of the Directive's requirement that the database show a substantial investment, assessed qualitatively or quantitatively, in either the obtaining, verification or presentation of the contents. In the landmark *British Horseracing Board*¹⁵ and *Football Fixtures* cases¹⁶, the Court ruled that term 'obtaining' excludes the costs incurred in the creation of new data (such as generating fixtures lists) from being considered relevant to satisfy the requirement of the substantial investment. Paragraph 31 of the *British Horseracing Board* decision reads as follows:

[t]he expression 'investment in...the obtaining...of the contents of a database must, as William Hill and the Belgian, German and Portuguese Governments point out, be understood to refer to the resources used to seek out existing independent materials and collect them in the database, and not to the resources used for the creation as such of independent materials. The purpose of the protection by the sui generis right provided for by the directive is to promote the establishment of storage and processing systems for existing information and not the creation of materials capable of being collected subsequently in a database.

Although the costs incurred for creating data are not considered as part of the substantial investment in making a database, the costs necessary for the verification of the accuracy of the data and for the presentation of such data to third party users do count in the assessment of whether the investment was substantial.¹⁷ The results of a practical application of the CJEU principles are particularly complex regarding the distinction between obtaining and creating data and regarding the concrete determination of the investment necessary to trigger the protection. This remains an evaluation that must be made on a case-by-case basis. This is particularly evident when dealing with scientific data, such as meteorological data or genetic sequences, which are not easily ascribable to one of these two categories.¹⁸ Applying the criteria developed by the CJEU to scientific databases, it is unclear whether the majority of research databases meet the formal requirements for the *sui generis* right. Many collections of data may arguably remain outside the scope of protection because the materials constituting the database are merely created – and not obtained from already existing sources – and the threshold of substantial investment is not reached by further investing either in the verification or the presentation of such contents.

2.2.2. The beneficiary of the protection

¹⁴ *Directmedia Publishing GmbH v Albert-Ludwigs-Universität Freiburg* C-304/07, [2009] 1 C.M.L.R. 7.

¹⁵ *British Horseracing Board Ltd v William Hill Organization Ltd* (BHB decision), C-203/02, [2004] ECR I-10415

¹⁶ *Fixtures Marketing Ltd v Svenska AB* (Svenska), C-338/02, [2004] ECR I-10497; *Fixtures Marketing Ltd v Organismos Prognostikon Agonon Podosfairou EG* (OPAP), C-444/02, [2004] ECR I-105449; *Fixtures Marketing Ltd v Oy Veikkaus Ab* (Oy Veikkaus), C-46/02, [2004] ECR I-10365.

¹⁷ See Annemarie Beunen, *Protection for databases – The European Database Directive and its effects in the Netherlands, France and the United Kingdom*, Nijmegen, Wolf Legal Publishers, 2007, p. 137.

¹⁸ See Mark J. Davison and P. Bernt Hugenholtz, *Football fixtures, horseraces and spinoffs: the ECJ domesticates the database right*, *EIPR*, 2005-3, p. 113-118, p 115.

Should research datasets indeed show a level of substantial investment in the ‘obtaining, verification or presentation’ of the data to qualify for protection, the second question to be addressed is who qualifies as the maker of the database. More specifically, does a research institution, or the individual researcher, benefit from the protection even if the investment is made possible only through the injection of public funding?

Recalling the wording of recital 41 of the Directive, the maker of the database is the person who takes the initiative and the risk of investing. Whether the research institution or the individual researcher himself is to be considered as the ‘maker of the database’ may be a question of fact. It may depend on such factors as the financing structure of the university, the department or the research project, the conditions of the grant or subsidy and the legal relationship between the researcher and his institution. It may also depend on how recital 41 of the Directive has been implemented in national legislation, if at all, and whether it puts the accent on the initiative taker, the risk of investing or on another element. For instance, the beneficiary of the protection has not been defined in Dutch legislation. There is therefore no reference in the Dutch Database Act to any kind of risk of investing. Dutch commentators seem to agree, however, that the *sui generis* right “should only be conferred as a reward for the risks taken by the innovating industry which actually produce the databases, not the commissioning parties.”¹⁹ Other commentators posit that employees, subcontractors or anyone else executing the work without bearing the financial responsibility for the end product cannot be considered as database right holder.²⁰

The Database Directive is silent on the issue of the source of the funding or the role of public money in the acquisition of rights. The Netherlands is so far the only Member State to have explicitly regulated the exercise of the *sui generis* rights by public sector bodies. Article 8 of the Dutch Database Act denies a public authority the right to exercise its exclusive database rights unless the right is reserved explicitly by a general mention in an act, order or ordinance, or in a specific case by notification on the database itself or while the database is made available to the public. The specific question of the impact of public funding on the nature of the investment made to produce a database was the object of a ruling from the Raad van State (Dutch Council of State) in the *Landmark* case²¹. In first instance, the District court of Amsterdam ruled that the City Council did not qualify as a ‘producer of a database’ and therefore did not own any database right in the information it gathered, since the collection of data occurred in the performance of its public task and with the support of governmental subsidies; the database was indeed realized thanks to public funds, partially coming from the Ministry of Housing, Spatial Planning and Environment. Referring to recital 41 of the Database Directive, the Court considered that City Council did not qualify as a producer of the database because it did not actually bear the risk of the investment since public funding and government subsidies could not be equated to an investment that needed to be recouped on the market.²² This ruling was confirmed on 29 April 2009 by the Raad van State.

¹⁹ Annemarie Beunen *supra* note 17, p. 150.

²⁰ J.H. Spoor, D.W.F. Verkade and D.J.G. Visser, *Auteursrecht*, 3rd ed., Deventer, Kluwer, 2005, p. 625.

²¹ ABRvS 29 April 2009, n 07/786, AMI 2009-6 (*College B&W Amsterdam/Landmark*; with annotation from M. Van Eechoud).

²² The Court also made reference to the ECJ *Directmedia* decision, (Case C-304/07) Judgment of the Court (Fourth Chamber) of 9 October 2008) in which the Court declared that the purpose of the database right is ‘to guarantee the person who has taken the initiative and assumed the risk of making a substantial investment in terms of human, technical and/or financial resources in the obtaining, verification or presentation of the contents of a database a return on his investment by protecting him against the unauthorized appropriation of the results of that investment by acts

The Dutch position is so far unique within the EU. In fact, the CJEU recently rendered decision in the *Compass-Datenbank* case²³ in which it indirectly confirmed that public entities are entitled to own and exercise *sui generis* database rights. The referring Austrian court asked, in essence, whether the activity of a public authority in permitting interested persons to search for data and in providing them with print-outs thereof in return for payment, while prohibiting any other use of that data by relying on its *sui generis* database right, constitutes an economic activity of an undertaking within the meaning of Article 102 TFEU. The Court answered in the negative. Incidentally, the Court declared that:

*a public entity which creates a database and which then relies on intellectual property rights, and in particular the abovementioned sui generis right, with the aim of protecting the data stored therein, does not act, by reason of that fact alone, as an undertaking. Such an entity is not obliged to authorise free use of the data which it collects and make available to the public.*²⁴

It would appear from this ruling that the Court sees no ground in preventing public bodies from qualifying as makers of a database due to the mere fact that the substantial investment is made possible through public funding. The CJEU exercised judicial restraint in the *Compass-Datenbank* case and refrained from examining the level of investment that might have given rise to protection. So the story does not tell whether the investment made by the Austrian government demonstrated the required 'substantial investment' to qualify for protection. Logically however, provided 'substantial investment' is directed to obtaining, verifying and presenting' the data, the fact that scientific databases are produced with public funding should not deprive research institutions or researchers, depending on who bears the risk of investment, from benefiting from the *sui generis* protection under the European Directive (with the exception, perhaps, of the Netherlands).

2.2.3. Scope of protection

Where the 'obtaining, verification or presentation' of research datasets does manifest the substantial investment necessary to qualify for protection, the *sui generis* protection confers two transferable rights on the maker of a database: the right of extraction and the right of re-utilization of substantial parts of the database, which are respectively defined as follows: '(a) 'extraction' shall mean 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; (b) 're-utilization' shall mean 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'. These two concepts have received a broad interpretation from the CJEU.²⁵ In the *Directmedia* case, the Court found that an act of 'extraction' occurs when all or part of the contents of the database concerned are transferred to another medium, whether of the same or of a different nature. Such a transfer implies that all or part of the contents of a database can be found in a medium other than the original database.²⁶ In the view of the ECJ, it is immaterial whether the transfer is based on a technical process of copying the contents of a protected database or on a simple manual process; similarly, it is irrelevant that the transfer of the contents of the database may lead to an arrangement of

which involve in particular the reconstitution by a user or a competitor of that database or a substantial part of it at a fraction of the cost needed to design it independently'.

²³ *Compass-Datenbank*, C-138/11, Judgment of the Court (Third Chamber), 12 July 2012.

²⁴ *Id.*, para. 47.

²⁵ *Directmedia Publishing GmbH v Albert-Ludwigs Universität Freiburg*, C-304/07, [2009] 1 C.M.L.R. 7.; *Apis – Hristovich EOOD v Lakorda AD*, C-545/07 [2009] ECRI-1627.

²⁶ *Directmedia* decision, para 36.

the elements that is different from that in the original one. The Court adds that the transfer of material from a protected database to another database following an on-screen consultation of the first database and an individual assessment of the material contained in that first database is also capable of constituting an extraction.

In the British Horseracing Board decision, the Court ruled that the concept of 're-utilisation' must 'be understood broadly, as extending to any act, not authorised by the maker of the database protected by the *sui generis* right, of distribution to the public of the whole or a part of the contents of the database'. The Court observed that the nature and form of the process are of no relevance in this respect.²⁷

The Database Directive does not provide any definition of the terms 'substantial'/'insubstantial' or 'qualitative'/'quantitative'. Intuitively, the first method to assess the substantiality of the part extracted and/or re-utilized consists in comparing quantitatively the amount of data taken with the total amount of data contained in the database as a whole. Accordingly, the ECJ decides that a quantitatively substantial part corresponds to 'the volume of data extracted and/or reutilized and must be assessed in relation to the volume of the contents of the whole of that database'. Furthermore, the Court observes that 'if a user extracts and/or reutilizes a quantitatively significant part of the contents of a database whose creation required the deployment of substantial resources, the investment in the extracted or re-utilized part is, proportionately, equally substantial'.²⁸ It has been argued that the substantiality of a part can also derive from its economic value, namely from the price that would be paid for such a part, which would raise proportionately with the rate of investments incurred in the obtaining, verification and/or presentation of the part's contents. But such a direct correlation cannot always be traced, and the Directive recognizes a *sui generis* protection on the basis of the investment made in the database's production rather than in its economic value.

The protection under the *sui generis* right lasts for 15 years from the first of January of the year following the date on which the database was completed. The term of protection for a database may start anew under two conditions, both dealing with the term 'substantial'. The first one is represented by a substantial modification of the contents of the database, evaluated either qualitatively or quantitatively, which can consist of additions, deletions or alterations (including rearrangement of the contents). Secondly, this substantial modification must represent a substantial investment, evaluated qualitatively or quantitatively. This is one of the most controversial and criticized provisions of the Directive since it apparently offers grounds for a perpetual protection of the databases.²⁹

Finally it is worth pointing out that, according to article 11 of the Database Directive, only natural persons who are nationals of a Member State or who have their habitual residence in the territory of the EU can benefit from the database right. Furthermore, companies and firms are also entitled to such protection if they are formed according to the law of a Member State and have their registered office, central administration or principal place of business within the EU. Article 11.2 clarifies that in case a company or a firm has a registered office only in the territory of the EU, its operations must be substantially and durably linked with the economy of a Member State. In other words, the protection of the *sui generis* database right is not only unique to Europe in that it is conferred only on EU nationals,

²⁷ BHB decision, para 61.

²⁸ BHB decision, para. 70.

²⁹ J.H. Reichman and R. L. Okediji, When Copyright Law and Science Collide: Empowering Digitally Integrated Research Methods on a Global Scale, 2012, Minn. L. Rev. Vol. 96, p. 1362, 1451.

whether natural or legal persons, but also because there is no real comparable regime of protection for non-original databases outside the European Union.³⁰

The complexity of the rights status of research data arguably has the potential to adversely affect the re-use opportunities of the collections of scientific data, given the difficulty – both for research institutions making the database available and for prospective re-users – in determining each time whether a certain database is covered by the *sui generis* right and in which measure re-utilization and extraction can take place freely.

3. Open Access Licensing of Research Data

The uncertainty arising from the complex rights status of scientific research data under European law can, to some extent, be alleviated through the use of contracts. Setting standardized contractual conditions of use lowers transaction costs between rights owner and users and eases the re-use of information, even in cases where the determination of rights can be problematic. On the other hand, if no rights attach to a scientific database or its content, then there is no ground for licensing at all. This certainly explains why American research institutions, like the Harvard-MIT Data Center, do not display any licensing term in relation to the data they make available for public use: section 102(2) of the US Copyright Code expressly states that ‘in no case does copyright protection for an original work of authorship extend to any idea, procedure, process, system, method of operation, concept, principle, or discovery, regardless of the form in which it is described, explained, illustrated, or embodied in such work.’ Moreover, ‘Compilations’ (or databases) are protected pursuant to section 101 of the Code only insofar as the elements assembled are selected, coordinated, or arranged in such a way that the resulting work as a whole constitutes an original work of authorship. US law does not recognize any equivalent to the European *sui generis* database right.³¹

Since the legal status of scientific databases and their content is murkier under European law, the use of standard licenses eliminates the need for the user to look for the rights owner and to negotiate the terms of use. To foster the widest access and re-use of scientific publications and data, policy makers are now pushing for open access conditions. What do open access conditions entail and how can they be put in practice effectively?

3.3 Open Access Principles

The term ‘open access’ was first formally defined at a meeting in Budapest in early December 2001. Out of that meeting came the so-called Budapest open access Initiative³² and ‘open access’ was defined as 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 gaining access to the internet itself. The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited.’ The Budapest open access Initiative was followed up some fifteen months later by the Bethesda Statement³³, which came out of a one-day meeting of

³⁰ Among the countries outside the European Union that recognize some protection on non-original databases are South-Korea, Japan.

³¹ *Feist Publications, Inc. v. Rural Telephone Service Co.*, 499 U.S. 340 (1991).

³² Available at: <http://www.opensocietyfoundations.org/openaccess/>.

³³ Available at: <http://www.earlham.edu/~peters/fos/bethesda.htm/>.

scientists, funding agencies, librarians, scientific societies and publishers, held in April 2003. In October of the same year, the Max Planck Society in Germany convened a meeting on 'open access to Knowledge in the Sciences and Humanities.' This meeting widened the discussion to include the humanities and produced the 'Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (October 2003)'.³⁴

Open access contributions include original scientific research results, raw data and metadata, source materials, digital representations of pictorial and graphical materials and scholarly multimedia material. On the basis of these statements and initiatives, the three following essential characteristics of open access emerge: free accessibility, further distribution, and proper archiving.³⁵ The Berlin Declaration gives a definition of which contributions qualify as open access, e.g. those that satisfy the following two conditions:

1. *The author(s) and right holder(s) of such contributions 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 (community standards, will continue to provide the mechanism for enforcement of proper attribution and responsible use of the published work, as they do now), 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, interoperability, and long-term archiving.*

Since the lack of access to complete literature can seriously impede advances in knowledge, the main thrust of the Declaration lies in the creation of a new 'open access' business model for scientific publishing or, absent this, of institutional repositories where all scientific and scholarly publications are to remain freely accessible. According to the Berlin Declaration, the only main constraint on reproduction and distribution of articles should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited. The principles set out in the Berlin Declaration are primarily aimed at governments, universities, research institutions, funding agencies, foundations, libraries, museums, archives, learned societies and professional associations. Remarkably, neither publishers nor authors are listed in this enumeration.

Admittedly, because the Berlin Declaration's main goal is to increase access to scientific knowledge, whether through the 'Green' or 'Golden' road of OA,³⁶ less emphasis is put in

³⁴ Available at: <http://oa.mpg.de/lang/en-uk/berlin-prozess/berliner-erklarung/>.

³⁵ Open Society Institute, *Open Access Publishing and Scholarly Societies – A Guide*, New York, OSI, 2005, p. 6; see also: P. Suber, *Open Access*, MIT Press, 2012.

³⁶ The 'Green Road' to OA is based on a 'subscriber-pays-model' where pre-print/post-prints of articles are deposited in an institutional repository often after an embargo period, while the 'Golden Road' to OA is based on a 'author-pay-model' – See L. Guibault, *Owning the Right to Open Up Access to Scientific Publications*, in: L. Guibault and C. Angelopoulos (ed.), *Open Content Licences : From Theory to Practice*, Amsterdam, Amsterdam University Press, 2011, pp. 137-167, 157.

practice on the possibility for scientific researchers to re-use the material they access. It is therefore often unclear to what extent a subsequent researcher is or ought to be allowed to extract and re-utilize a substantial part of the content of a database for the purpose of creating a new database. The Green Road may well meet the three minimum OA requirements, namely free access, possibility to copy, use, distribute the work and permanent archiving, but releasing research results along the Golden Road ensures a better access, clearer re-use possibilities, visibility and 'findability' of research output on the internet.

Of course, the absence of universally accepted definition of open access does not make things easier. But in other contexts, like the software, cultural heritage and governmental sectors, the possibility to re-use the licensed material is considered to be one of the main characteristics of the 'open' ideology. Principle 3 of the Open Source Definition states that 'the license must allow modifications and derived works, and must allow them to be distributed under the same terms as the license of the original software'.³⁷ Strongly inspired by definition of 'free software' developed by the Free Software Foundation,³⁸ the Definition of the Free Cultural Works declares that:

*works of authorship should be **free**, and by freedom we mean:*

- *the **freedom to use** the work and enjoy the benefits of using it*
- *the **freedom to study** the work and to apply knowledge acquired from it*
- *the **freedom to make and redistribute copies**, in whole or in part, of the information or expression*
- *the **freedom to make changes and improvements**, and to distribute derivative works³⁹*

In other words, Free Cultural Works are defined 'as works or expressions which can be freely studied, applied, copied and/or modified, by anyone, for any purpose'. The European Directive on the re-use of public sector information, although not applicable to scientific information, states that 'licences should not unnecessarily restrict possibilities for re-use or be used to restrict competition'.⁴⁰ In comparison to these movements, the Green Road's implementation of the principles of the Berlin Declaration usually remains vague regarding the freedom to make changes and improvements and to distribute derivative works.

Establishing open access as a worthwhile procedure ideally requires the active commitment of each and every individual producer of scientific knowledge. To date, the Berlin Declaration has been signed by more than 400 organizations worldwide.⁴¹ But the OA principles have so far been applied mainly to scientific publications. They are meant to be translated in practice as a set of licensing terms that allow not only the dissemination of scientific articles free of charge, but also greater flexibility in terms of re-use of these articles. Several existing standard form licences could probably be used in the context of OA publishing, including the Creative Commons Licences, the Open Data Commons and the Digital Peer Publishing License. Would they all meet the requirements of the OA principles? Which of these would be the most suitable for the dissemination of scientific information? Are restrictions, such as a prohibition to use the work for commercial purposes and the

³⁷ See: <http://opensource.org/docs/osd>

³⁸ See: <http://www.gnu.org/philosophy/free-sw.html>

³⁹ See: <http://freedomdefined.org/Definition>

⁴⁰ Directive 2003/98/EC of 17 November 2003 on the reuse of public sector information, OJ L 345/90, 31.12.2003.

⁴¹ See: <http://oa.mpg.de/lang/en-uk/berlin-prozess/signatoren/>

'share-alike' clause contained in some of these licenses compatible with the OA principles and with the sharing needs of the scientific community?

3.2 Creative Commons Licences

The Creative Commons (CC) licensing system is the most widely used set of licenses because it offers a series of easy to use, standardized and automated licenses that authors can affix to their work in order to indicate under which conditions it may be used. Thanks to these licenses, it is no longer necessary for users to contact the rights holder prior to every use of the work to find out what can or cannot be done with the work. The work is, therefore, made available to everyone in accordance with the conditions of the chosen CC license. Besides the four core stipulations (Attribution (By), Non-Commercial (NC), No-Derivatives (ND) and Share Alike (SA)), a number of fundamental principles lie at the basis of each CC license. Taking into account the conditions of the chosen license, the licensor grants the user a worldwide, non-exclusive, perpetual (for the duration of the applicable right) license to reproduce, display, perform, communicate and distribute copies of the work. All rights may be exercised in all media and formats whether now known or subsequently devised. The above rights include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. In principle, all rights not expressly granted by the licensor are reserved. All CC licenses are irrevocable. This means that the moment the work is distributed under CC a license on the Internet, the author can no longer change his mind or withdraw the license. In addition, the user is required to add a copy of, or the uniform Resource Identifier for, the applicable CC license to each copy of the work that he distributes, communicates or makes available to the public.⁴²

It is also important to note that, in principle, the CC license system makes no distinction between works in digital or non-digital format, or between several types copyright relevant acts, such as the act of reproduction or communication to the public. Article 2 of each CC license provides that nothing in the license is intended to reduce, limit or restrict any uses free from copyright or rights arising from limitations or exceptions that are provided for in connection with the copyright protection under copyright law or other applicable laws. Moreover, the licensor may not apply any effective technological measures to the work that restrict the ability of a recipient of the work to exercise the rights granted under the terms of the license.

Creative Commons licences are widely used to disseminate scientific publications under OA principles. For example, all publications made available on through the OAPEN Library⁴³, Hindawi Publishing⁴⁴, the Public Library of Science (PLOS),⁴⁵ or in BioMed Central⁴⁶ are distributed under either one of the six core CC licences. BioMed Central's summary of the agreement states that 'anyone is free: to copy, distribute, and display the work, to make derivative works, to make commercial use of the work, under the following conditions: the original author must be given credit for any re-use or distribution; it must be made clear to others what the license terms of this work are'. A joint Creative Commons Nederland/SURFdirect report recommended the use of the CC Attribution 3.0 License above all other licences or other combinations of CC licences, for this licence is the most in line

⁴² Creative Commons Attribution 3.0 Unported License available at: <http://creativecommons.org/licenses/by/3.0/legalcode>.

⁴³ See: <http://www.oapen.org/home>.

⁴⁴ See: <http://www.hindawi.com/>.

⁴⁵ See: <http://www.plos.org/>.

⁴⁶ See: <http://www.biomedcentral.com/>.

with the principles of OA, while placing the fewest obstacles on the further re-use of research results, and still requiring that proper attribution be given to the author and that derivative works be identified as such.⁴⁷ The application of a No-Derivative clause would meet the Berlin Declaration's requirements of free access, further distribution and proper archiving, but it would fall short of what the open community considers as a real 'free work'. Similarly the application of a Non-Commercial and a Share-Alike clause would put an unnecessary burden and obstacles on the re-use of scientific publications and would generate potential incompatibilities between scientific projects.

The use of CC licences in relation to scientific databases and their contents is even more problematic. The main reason lies in the Creative Commons organisation's firm belief that 'scientific data should be freely available to everyone', as a result of which the application of the core six CC licences to databases is, from a European perspective, far from optimal.⁴⁸

CC licenses are translated and adapted to the laws of a maximum of jurisdictions in the world (to date in more than 50 jurisdictions worldwide). National jurisdictions are able to 'port' the CC licences to their local legal system based on 'unported' licenses, which are in principle jurisdiction-agnostic: they do not mention any particular jurisdiction's laws or contain any sort of choice-of-law provision. While versions 1.0, and 2.0 of the 'unported' licence (previously known as the 'generic' licence) were based on the provisions of the U.S. Copyright Act, version 3.0 of the 'unported' licenses is instead based on the provisions of the Conventions of Berne and Rome.⁴⁹ Since the database right is a purely European phenomenon, it is not surprising to note that databases are only indirectly covered by the unported Creative Commons licenses version 3.0. The definition of "Work" under the licences includes the "literary and/or artistic work offered under the terms of this License including without limitation any (...) compilation of data to the extent it is protected as a copyrightable work". No explicit reference is made to the European database right.

When porting the CC licences to their national law, several European jurisdictions took it upon themselves, for the sake of completeness, to include databases as a subject-matter of the licenses. This is the case of the Netherlands, Germany, France and Belgium where version 2.0 also added 'extraction and reutilization' of substantial parts of a database in the version 2.0 rights grant, as the equivalent to the right of reproduction, performance and distribution for works covered by copyright and neighbouring rights. This European initiative was not seen favourably by the founders of the Creative Commons licences: first, because the licences are said to protect the fruits of creative effort and not merely investment; second, because since the database right is purely European, its inclusion in the licences could lead to legal uncertainty for database makers residing outside of Europe; and third, because there was fear that some licensors would try to contractually claim protection on databases, thus 'importing' the database right, in jurisdictions that do not recognise it.⁵⁰

Consequently, a compromise was reached before version 3.0 was to be ported anywhere in Europe: the *sui generis* database right was to be waived in all European licences. For example, the Dutch definition of 'Work' still covers "the copyrightable work of authorship

⁴⁷ P. Keller and W. Mossink, *Hergebruik van materiaal in onderwijs- en onderzoekomgevingen*, Utrecht/ Amsterdam: Creative Commons Nederland and SURFdirect, March 2009, p. 31 available at http://www.creativecommons.nl/downloads/090323SURFCC_Hergebruik_van_materiaal.pdf.

⁴⁸ See: <http://creativecommons.org/science..>

⁴⁹ Catharina Maracke, Creative Commons International, *JIPITEC* : Journal of Intellectual Property, Information Technology and E-Commerce Law, 2010, Vol 1, Issue 1, p. 4-18.

⁵⁰ *Id.*, p. 10.

put at disposal under the terms of this License. For the purposes of this License a Work should also be taken to mean the phonogram, the first recording of a film and the (broadcasting) programme in the sense of the Neighbouring Rights Act and the database in the sense of the Database Act, insofar as such phonogram, first recording of a film, (broadcasting) programme and database is protected under the applicable law within the User's jurisdiction". However, the license elements requirements (Attribution, Non-Commercial, No-Derivatives, and Share-Alike) are no longer applied to database rights. This follows from article 4 subparagraph e) of the European transposition of the licence, which reads:

For the avoidance of doubt, it must be noted that the aforementioned restrictions (paragraph 4(a), paragraph 4(b), paragraph 4(c) and paragraph (d) do not apply to those parts of the Work that are deemed to fall under the definition of the 'Work' as stated in this License solely on account of compliance with the criteria of the sui generis database law under national law implementing the European Database Directive.

Under version 3.0, database rights have been effectively removed from the scope of the licenses as a result of which the optional license elements lose their effect and not be applied to databases, insofar as they are protected under the *sui generis* regime.⁵¹ Thus, the licensor of a database licensed under an Attribution Share Alike Netherlands 2.0 license will expect derivatives to carry the Share Alike element and stay in the Commons. However, the Share Alike interoperability clause allows that any derivative of the database may be relicensed under a license which may state that the licensing restrictions, including Share Alike, cannot be applied to a database. Therefore, the second derivative will not be shared with the Share Alike element, and the original licensor's expectation will be disappointed as far as Attribution, No-Derivative, Non-Commercial and Share-Alike are concerned: these restrictions will not be applied.

In principle, all six core CC licences meet the OA principles of free accessibility, further distribution, and proper archiving. All licences also guarantee proper attribution of the author, in application of the mandatory 'Attribution' licence element; and the No-Derivative licence element allows the author to control the integrity of his work, as the author so requires. At this time, however, the core Creative Commons licensing suite is not the appropriate instrument to license scientific databases and their content, even for those who wish to use the most liberal licence (CC-Attribution Licence). The fact that the *sui generis* database right is waived takes away all usefulness of the licences for the purpose of disseminating scientific research results under specific conditions. Scientific research institutions who wish to allow the re-use of a substantial part of their database on the condition that they receive attribution for the use of their data are frustrated in their intention, for version 3.0 explicitly disallows this possibility. The same holds true for the wish of the database maker to limit third parties from making a commercial use of or a derivative product from a substantial part of the database.

Since the porting of version 3.0 in Europe, Creative Commons' treatment of the *sui generis* database right has been criticized on two points: First, because if someone applies an international license to a database from a country that recognizes *sui generis* database protection, there is a possibility they are not granting licensees any rights to use the database in a way that implicates the *sui generis* database right. Licensees may not realize

⁵¹ L. Guibault, Creative Commons Licenses: What to Do with the Database Right?, *Computers and Law Magazine of SCL*, 2011, Vol. 21, Issue 6, p. 3.

that they could need extra permissions to use a substantial portion of the licensed database, in the EU and a few other jurisdictions. Second, it appears that in practice, a number of European institutions have refrained from using CC licenses because they do not affirmatively license the *sui generis* database right on par with copyright (i.e., without waiving conditions when only the *sui generis* database right exists). Therefore the mandatory waiver of the *sui generis* database right has to some extent operated as an obstacle to the widespread use of the CC licensing system in Europe.

In view of the criticism and the obstacles perceived to exist in Europe due to the approach taken regarding the licensing of the *sui generis* database right, the Creative Commons organisation decided to address, among other issues, the specific question of the database right in the discussions leading to a new version 4.0 of the licensing suite. Stepping back from its previous position, Creative Commons declared its intention to license database rights from now on, under the same terms and conditions as copyright. The organisation feels that "this is the best way to ensure that database rights are not a barrier to would-be licensors seeking to exercise those rights, or to those using of CC-licensed works, who might otherwise need to ask for separate permission to use the work as intended".⁵² An effort is being made to clarify the fact that database rights (as with all licensed rights) are only within the scope of the license if they are held by Licensor and apply to use of the licensed work by the licensee. As one of the goals for version 4.0 is to craft a license suite that allows use of the work consistent with the expectations of both licensors and licensees (and rights closely related to copyright may impede that), the permissions granted by the licenses may need to account for other laws that grant copyright-like rights in a particular subset of jurisdictions (such as *sui generis* database rights). "Copyright-like" rights are those rights that overlap with the exclusive rights of copyright and are exclusively held by the same person as the copyright. The drafting process of version 4.0 started in September 2011 and a final text of version 4.0 is expected to be officially launched in the course of 2013. The proposed changes regarding the *sui generis* database right should make the CC licence suite suitable to license not only scientific publications but also the related research data, all with one single instrument.

Until such time that version 4.0 of the core CC licensing suite is up and running, rights owners on works and other types of information, like scientific research data, have the possibility to relinquish their rights through another legal tool: the Creative Commons Zero Universal Dedication 1.0 (CC0).⁵³ CC0 is a legal tool that operates as a waiver of copyright and related or neighbouring rights (including *sui generis* right and moral rights) to the fullest extent permitted by law. Applying CC0 to a work or any other type of protected subject-matter:

overtly, fully, permanently, irrevocably and unconditionally waives, abandons, and surrenders all of Affirmer's Copyright and Related Rights [including database rights] and associated claims and causes of action, whether now known or unknown (including existing as well as future claims and causes of action), in the Work (i) in all territories worldwide, (ii) for the maximum duration provided by applicable law or treaty (including future time extensions), (iii) in any current or future medium and for any number of

⁵² See: http://wiki.creativecommons.org/4.0/License_subject_matter.

⁵³ See: http://creativecommons.org/about/cc0_

*copies, and (iv) for any purpose whatsoever, including without limitation commercial, advertising or promotional purposes (the ‘waiver’).*⁵⁴

Consequently, anyone can use the information released under a CCO Dedication in any way and for any purpose – including commercial. If the waiver is not effective or legally valid for any reason, CCO acts as a license from the affirmer granting everyone an unconditional, irrevocable, non-exclusive, royalty-free license to use the work for whatever purpose. Especially in cases where the determination of the rights status of a database is difficult, CCO can be a useful tool for clarifying to the general public that the affirmer is committed to relinquishing protection to the broadest extent possible.

The strong position adopted in the past by CC towards the *sui generis* database right has left the field open for the elaboration of ‘competing’ licenses that specifically address the database right and its peculiarities, and also allowing for restrictions.

3.3 Open Data Commons

A more recent project in the realm of open access is the Open Data Commons (ODC)⁵⁵ led by the Open Knowledge Foundation.⁵⁶ A set of three licences was launched between 2008 and 2010: the Public Domain Dedication and License (PDDL),⁵⁷ the Attribution license (ODC-By), and the Open Database License (ODC-ODbL). Strongly inspired by the Creative Commons licences in their structure and wording, the ODC licences are designed to specifically provide for the licensing of databases and their contents that are protected under copyright and the European *sui generis* database right. The licences are meant to cover all types of databases, whether produced by commercial entities, public sector institutions or anyone else. They are therefore not explicitly crafted to apply to scientific research results. The ODC licences are much less known or used than the Creative Commons licences. The best-known ‘major adopter’ is the OpenStreetMap project, which switched from a CC-By-SA to the ODbL due to the waiver of the database right under the CC licences.⁵⁸ And because the ODC licences are not provided in a ‘machine-readable’ format, it is much more difficult to estimate their level of use. Be that as it may, and although not devoid of imperfections, all three ODC licences meet the OA principles of free accessibility, further distribution, and proper archiving. The two later licences also guarantee proper attribution of the author, e.g. database maker.

The PDDL⁵⁹ most closely resembles the CCO Dedication. The document is aimed at placing the database and its contents in or as close as possible to the public domain and is intended to apply to databases or their contents, either together or separately. It is intended to allow

⁵⁴ Art 2 of CCO 1.0 Universal further states that ‘Affirmer makes the Waiver for the benefit of each member of the public at large and to the detriment of Affirmer’s heirs and successors, fully intending that such Waiver shall not be subject to revocation, rescission, cancellation, termination, or any other legal or equitable action to disrupt the quiet enjoyment of the Work by the public as contemplated by Affirmer’s express Statement of Purpose.’

⁵⁵ See: <http://opendatacommons.org/>

⁵⁶ See: <http://okfn.org/> The Open Knowledge Foundation is a non-profit organization founded in 2004 and acts as one of the main international leaders in the promotion of open knowledge in many different forms.

⁵⁷ Open Data Commons was created in December 2007 as a platform for the drafting of the first ‘open’ database license, the Public Domain Dedication and License, which was written by Jordan Hatcher and Dr Charlotte Waelde. In January 2009 the Open Data Commons project was transferred to the Open Knowledge Foundation. The Foundation is in charge of the daily administration of the project, whilst its Advisory Council is responsible for the drafting and management of the licenses. See <http://opendatacommons.org/about/>.

⁵⁸ See: <http://www.openstreetmap.org/copyright>.

⁵⁹ See: <http://opendatacommons.org/licenses/pddl/1.0/>

users to freely share (copy, distribute and use), create (produce derivative works from the database) and adapt (modify, transform and build upon) the work – meaning either or both the database (covered by copyright or *sui generis* right) and its contents (defined as the ‘data’ and including ‘information, independent works or other material collected into the database’) – for any purpose and without any kind of restrictions, permanently and irrevocably. In fact, users can use the databases or their contents commercially and apply technical protection measures and they are neither required to attribute the creator of the data or database nor to provide further users with a copy of the license. Should the relinquishment or waiver of rights not be valid in a particular jurisdiction, the PDDL document contains a fall-back license, comparable to that of the CC0.

A ‘work’ under the PDDL can be either, or both, the database and its contents (the data). Such concepts are clearly distinguished in the context of the PDDL document; in particular, right-holders can apply the PDDL scheme both to the database and its contents or, alternatively, they can use the document to cover only the database and leave the contents regulated by other licensing models. The legal rights explicitly covered are copyright and database right.

With respect to moral rights, the licence distinguishes between jurisdictions allowing such a waiver to the fullest extent possible and jurisdictions where such a waiver is not possible, so that the licensor ‘agrees not to assert any moral rights over the work and waives all claims in moral rights to the fullest extent possible by the law of the relevant jurisdiction’. Remarkably this licence, as well as the two others drawn up by Open Data Commons, makes no difference, when dealing with moral rights, between databases protected by copyright and by the *sui generis* right. Indeed, the traditional attributes of moral rights – including the right to object to derogatory treatments affecting the author’s honour or reputation – are deemed to accrue exclusively with reference to copyright-protected databases, and commentators normally exclude moral rights from attaching to a database protected exclusively by virtue of a *sui generis* right.⁶⁰

The second licence offered is the Open Data Commons Attribution License (ODC-By). As its name indicates, it closely resembles the CC-By Licence and is intended to allow users to share, modify and use the database freely subject only to the Attribution requirement. The preamble of the license specifies that ‘Databases can contain a wide variety of types of content (images, audiovisual material, and sounds all in the same database, for example), and so this license only governs the rights over the Database, and not the contents of the Database individually. Licensors may therefore wish to use this license together with another license for the contents’. Through the ODC-BY the licensor grants a worldwide, royalty-free, non-exclusive license to use the database for the duration of any applicable copyright and database rights, and explicitly allowing commercial exploitation. Among the rights granted, the following are mentioned: a) extraction and reutilization of the whole or a substantial part of the contents; b) creation of derivative databases;⁶¹ c) creation of

⁶⁰ See E. Derclaye, *The Legal Protection of Databases. A Comparative Analysis*, Edward Elgar Publishing: Cheltenham, 2008, p. 53 and the same commentator in *Intellectual Property Rights and Human Rights: Coinciding and Cooperating*, in: Paul Torremans, *Intellectual property and human rights*, The Hague, Kluwer Law International, Information Law Series, 2008, p. 143.

⁶¹ Section 1 containing the definitions specifies that ‘derivative database’ is intended as ‘a database based upon the Database, and includes any translation, adaptation, arrangement, modification, or any other alteration of the Database or of Substantial part of the Contents. This includes, but is not limited to, Extracting and Re-utilizing the whole or a Substantial part of the Contents in a new Database’.

collective databases;⁶² d) creation of temporary or permanent reproductions by any means and in any form, in whole or in part, including of any derivative databases or as part of collective databases; and e) distribution, communication, display, lending, making available, or performance to the public by any means and in any form, in whole or in part, including of any derivative database or as part of collective databases.

Although the licence only governs the rights over the database, and not the contents of the database individually, its stated legal effect is to operate a licence of applicable copyright and neighbouring rights; a license of the Database Right; and an agreement in contract between the user and the Licensor. It is unclear what role neighbouring rights could play in this context, knowing that the neighbouring rights that are part of the *acquis communautaire* concern the rights of performing artists, phonogram producers, film producers and broadcasting organisations. The grant of a licence on neighbouring rights is all the more odd that article 2.4 of the licence specifies that:

The individual items of the Contents contained in this Database may be covered by other rights, including copyright, patent, data protection, privacy, or personality rights, and this License does not cover any rights (other than Database Rights or in contract) in individual Contents contained in the Database. For example, if used on a Database of images (the Contents), this License would not apply to copyright over individual images, which could have their own separate licenses, or one single license covering all of the rights over the images.

If a separate licence is necessary to cover the rights on the individual contents of the database, then why license neighbouring rights? Following the same questionable logic, moral rights are waived and extensive wording has been included to deal with (non)waivable compulsory and/or voluntary licensing schemes. The fact is that nowhere in Europe are there any collective rights management schemes, mandatory or even voluntary, concerning rights in databases. This language can therefore only relate to the individual contents, which are not covered by the licence. Nevertheless, the ODC-By licence contains useful indication on how to give proper attribution. It also stresses that the licence is without prejudice to the exercise of any exception or limitation provided for in the law.

The third and last licence is the Open Database Licence (ODbL)⁶³ Attribution and Share-Alike for Data/Databases license. Like the ODC-By, the scope of application of the ODbL is unclear and is likely to create confusion among licensors and users. The preamble states that

Licensors should use the ODbL together with another license for the contents, if the contents have a single set of rights that uniformly covers all of the contents. If the contents have multiple sets of different rights, Licensors should describe what rights govern what contents together in the individual record or in some other way that clarifies what rights apply.

The ODbL grants users (as the ODC-BY) of the database a worldwide, royalty-free, non-exclusive license to use the database for the duration of any applicable copyright and

⁶² A 'collective database' corresponds to the licensed database in unmodified form as part of a collection of independent databases that together are assembled into a collective whole. The ODC-BY license further specifies that a work which constitutes a collective database is not considered, under the terms of the license, as a derivative database.

⁶³ The Human-Readable Summary and the Full Text of the ODbL can be found, respectively, at: <http://opendatacommons.org/licenses/odbl/summary/> and <http://opendatacommons.org/licenses/odbl/1.0/>.

database rights, and explicitly allowing commercial exploitation. The ODbL contains identical language to that of the ODC-By regarding neighbouring rights, moral rights and mandatory or voluntary licensing schemes.

Compared to the ODC-By licence, the ODbL contains an important additional condition of use, namely that licensees are required to make public uses of any derivative database of the original database i) only under the terms of the ODbL, or, alternatively, ii) according to a later version of the same license – equivalent in the spirit to the original one – or iii) according to a (non-further-specified) compatible license. If the original database or a derivative database is incorporated into a collective database, the licensee is not required to apply the ODbL to license the collective database; indeed, as defined in the ODbL, a collective database is not considered as a derivative database. Nevertheless, the license still applies to the original database or to a derivative one as part of the collection.

The licence further specifies that no derivative database is created by using the original database (also incorporated into a collective database) or a derivative database to create a produced work, defined as ‘a work (such as an image, audiovisual material, text, or sounds) resulting from using the whole or a Substantial part of the Contents (via a search or other query) from this Database, a Derivative Database, or this Database as part of the Collective Database’. Additionally, regarding the use of a derivative database or a produced work from a derivative database, the user is required to offer recipients of the derivative database or produced work a machine-readable copy of i) the entire derivative database, or ii) a file with all the modifications applied to the database or the methods for carrying out such alterations, including any additional contents between the original database and the derivative database.

Another remarkable difference with the ODC-BY license is the presence of a specific provision on technological measures and additional terms, which are – in principle – forbidden and can only be imposed on the database, on the derivative database or on the whole or a substantial part of the contents (defined as a ‘restricted database’), in the manner specified in the licence. This latter provision – which admits a scheme resembling one of ‘dual-licensing’ adopted for certain models of free software – states that such restrictions are applicable only if a copy of the original or derivative database is made available to the recipient i) without additional fee, ii) in a medium (defined as an ‘unrestricted database’) that does not alter the terms of the license or limit the possibility of any person to enjoy such rights and iii) subject to the fact that the unrestricted database is at least as accessible to the recipient as the restricted database.

It can be argued that the structure of the ODbL and of its legal requirements – together with the considerable length of the document itself – appears too complex for non-lawyers to understand and interpret without the advice of an expert, thus resulting in increased transaction costs. Since the ODbL incorporates a share-alike element – providing that derivative works must be distributed under the terms of the same ODbL – problems can arise for users who want to combine data and works from different sources.

While in theory the ODC licences would seem perfectly adapted to license rights on research data and the database containing them, several aspects of the licences makes their use less interesting. The main reason is that the licences only cover the database itself and not its contents. There is therefore no possibility to license the whole by means of one instrument. For example, OpenStreetMap has licensed its database under the OdbL, but was forced to license its copyright protected maps under a CC By-SA licence.⁶⁴ Should a research institution, an institutional repository or a publisher wish to license scientific publications

⁶⁴ See: <http://www.openstreetmap.org/copyright>.

together with the data sets upon which the publications are based, then it would need to use two distinct licences, one for the database and one for its content. This duality of instruments is at best unpractical, but at worst liable to lead to confusion and incompatibilities.

3. Digital Peer Publishing Licence (DPPL)

Contrary to the Creative Commons and the Open Data Commons licences, the Digital Peer Publishing Licence was crafted specifically with scientific publications in mind.⁶⁵ The licences were developed between 2004 and 2008 on commission of the Ministry of Innovation, Science, Research and Technology of the State of North-Rhine Westphalia in Germany by two German scholars Prof. A. Metzger and Dr. T. Jaeger of the Institut für Rechtsfragen der Freien und Open Source Software (IfROSS) (Institute for Legal Issues on Free and Open Source Software).⁶⁶ The DPPL is designed for scholarly content because it covers aspects of authenticity, citation, bibliographic data and metadata, permanent access and open formats. The DPPL is not very well known outside Germany. Through the DiPP project sponsored by the State of North-Rhine Westphalia a little fewer than twenty electronic journals are published under the terms of the DPPL. There might be other users who are less easily traceable.

The DPPL is offered in three modules: the DPPL (akin to the CC-By-ND licence), the free DPPL (akin to the CC-By licence) and the modular DPPL. The modular DPPL allows the rights holder to restrict the possibility of the user to make unauthorized modifications only to those parts of the publication that are indicated as such by the rights holder. Where all three licences specifically refer to the Berlin Declaration in their preamble, all meet the OA principles of free accessibility, further distribution, and proper archiving. All licences also guarantee proper attribution of the author and one iteration allows the full control of the integrity of the original work.

The basic module subjects all documents covered to be read and distributed unchanged by everybody or accessed for downloading. There is no distinction made between scientific or commercial use. Because of the document to be delivered only in electronic format under this licence, all rights concerning a printed version or a version on storage media are not covered. Thus electronic distribution is promoted. On the other hand, the bearer of the rights still has the option to contract with a publisher other versions of his work for commercial distribution, if desired.

This license agreement shall, unrestricted by time and place, allow You to reproduce the Work in electronic form and to pass the Work on to third parties by electronic means, e.g. by e-mail communication, and - particularly by making it available for download - to make it publicly accessible. The license shall comprise all types of use of the work that encompass non-physical reproduction or conveyance of the Work, whether already known or unknown at the time of conclusion of this license agreement. Use in physical form, particularly the distribution of printed Works, and the Use of Altered Versions of the Work shall not be permitted.

The extended modules of the license are designed for share and re-use of published material. The "modular DPPL" and the "free DPPL" allow users changing published material and contain arrangements how to properly cite if changes were made. In the modular DPPL,

⁶⁵ See: http://www.dipp.nrw.de/lizenzen/dppl/mdppl/m-DPPL_v3_en_11-2008.html.

⁶⁶ See: <http://www.ifross.org/>.

only those changes may be performed by recipients that are specifically earmarked. This makes possible, for example, that texts are fixed while images might be changed. In the free DPPL, anything in the publication may be changed pursuant to the terms of the license. As the explanatory document to the licences state:

Scientific researchers are regularly interested in making their findings widely available at a high level of quality. The Modular DPPL Licence therefore permits the passing on of copyrighted material. It furthermore grants permission to modify those parts of a work, marked as alterable by the authors and allows the dissemination of altered versions. While this rule intends to facilitate collaboration between scientists, it leaves the decision up the author, to open only certain parts of his work for interactive scientific collaboration. The aim of this Modular DPPL Licence is to ensure that this process takes place in a fair, transparent and secure way for all those involved.

All three iterations of the DPPL are well crafted and balanced. They would in principle suit the needs of scientific institutions very well, were it not for the fact that none of them seem to cover anything other than copyright protected publications. Where ‘Work’ under all three licences is defined as “The work protected by copyright, to which the rights to Use are granted by this license agreement”, it is highly improbable that the licences can be interpreted to apply to databases and data sets as well.

4 Conclusion

Scientific databases are in most cases not likely to meet the threshold for copyright protection. They can, however, be protected by the *sui generis* database right, provided they show substantial investment in the obtaining, verification and presentation of the data. On the basis of the case law of the Court of Justice of the European Union, it is fair to say that the fact that scientific databases are produced with public funding should not deprive research institutions or researchers, depending on who bears the risk of investment, from benefiting from the *sui generis* protection under the European Database Directive (to the exception perhaps of the Netherlands). Whether scientific databases meet the criteria for protection must be determined on a case-by-case basis. The complexity of the rights status of research data arguably has the potential to adversely affect the re-use opportunities of the collections of scientific data, unless it can be licensed effectively.

More and more pressure is put on research institutions in Europe by governments, funding agencies and the public at large to licence the results of their publicly funded research under open access conditions. For this, the Berlin Declaration offers guiding principles. However, because the Berlin Declaration’s main goal is to increase access to scientific knowledge, notwithstanding the business model of the publisher, it focuses primarily on the free access, further distribution and proper archiving elements, while putting less emphasis on the possibility to make derivative works. It is therefore unclear whether, on the basis of the Berlin Declaration, users of research data are allowed to bring modifications to the data. Open Access is not just about having access to articles for free; more importantly, it should help to further research, which implies being able to build on published research. This goes beyond access for reading.

Various licences provide for different possibilities of use and re-use of copyright protected material under open access conditions. The licences examined in this paper are the Creative Commons (CC) licences, the Open Data Commons (ODC) licences and the Digital Peer

Publishing Licences (DPPL). At this point in time, however, none of the schemes studied are drafted so as to govern research data in a useful manner. Once adapted to cover the *sui generis* database right in version 4.0, Creative Commons licences will have the biggest advantage, compared to other open access licences, of offering a standardised worldwide accepted system of licences, the terms of which can be arranged at the holders' content. In furtherance of the open access ideology, the best option would be a CC Attribution licence, since this licence contributes the most to the dissemination and re-use of information as broadly and as freely as possible. Another very viable option would be to broaden the scope of the DPPL so as to also cover material protected by the *sui generis* database right. This would enable research institutions to licence with one instrument all research results, e.g. all scientific publications with the data upon they are based, under open access conditions. The DPPL has the added advantage of codifying as well the authenticity, citation and bibliographical norms of science.