

---

# Open Bioinformation in the Life Sciences as a Gatekeeper for Innovation and Development

Roberto Caso and Rossana Ducato

---

## Abstract

Despite the increasing advocacy towards the “openness” of science and research data, it is still far from being a widespread practice. The goal of this paper is to identify the most pressing obstacles (in terms of funding, technology, Intellectual Property Rights, contracts, data protection, and social norms) which are hindering the development of Open Science and Open Research Data, with particular attention to the situation of developing countries. The innovative aim of this paper, which is the first essay of a broader research, is to prepare the epistemological basis for a Law and Technology theory of “Open Bioinformation” (OB), where bioinformation stands for research data in life sciences. We argue that so far the literature has addressed the promotion of openness in science and research data only in a sectorial manner, taking into account just one or a few of the factors affecting openness as if they were not related or mutually influenced. Therefore, the suggested solutions are limited to a single perspective and fail to consider the dynamics of information control. In our view, a holistic approach, that tries to zoom out from the specific disciplines and take into account the whole picture, would contribute to determining an effective policy for promoting OB. For this reason, we have to consider the technological, legal,

---

Roberto Caso is the author of paragraphs 3 and 6, and co-author of paragraphs 1 and 7; Rossana Ducato is the author of paragraphs 2, 4, 5, and co-author of paragraphs 1 and 7.

---

R. Caso (✉)

Faculty of Law, University of Trento, Trento, Italy

e-mail: roberto.caso@unitn.it

URL: <http://www.lawtech.jus.unitn.it/index.php/people/roberto-caso>

R. Ducato

e-mail: rossana.ducato@unitn.it

URL: <http://www.lawtech.jus.unitn.it/index.php/people/rossana-ducato>

and sociological aspects, in order to assess whether and how changes in one domain might affect the others.

---

## 1 Introduction

Everyone says that data sharing is an imperative in science. Everyone agrees that the free and immediate access to genetic information and medical data is crucial for the progress of life sciences research. Paraphrasing James Boyle in one of his most famous writings, such statements are so obvious that we should be able to make them in a law article without having to add footnotes.<sup>1</sup>

Open access (OA) to research data, as a gatekeeper for innovation and development, is of paramount importance in the so-called “Global South” (GS). In the field of medical and biotechnological research, developing countries face a considerable delay, which is exacerbated by the chronic lack of funds for the creation of research infrastructures and investment in education and training, as well as by widespread recourse to the practice of secrecy and/or the application of strong Intellectual Property Rights (IPRs), which hinders access to and circulation of scientific knowledge.<sup>2</sup>

A possible way out of this situation has been identified in the open models for sharing the building block of life sciences, i.e. research data. In particular, in the paper we refer to such a heterogeneous category as “bioinformation” in order to subsume a composite set of digitized biological information, relating to an individual, and which is commonly used for the purposes of biomedical and biotechnological research.<sup>3</sup> The following section will be specifically dedicated to framing the definition of bioinformation and explaining the importance of sharing for current scientific research. The third section will be devoted to understanding the dynamics, the content, and the tools of “openness” in life sciences. Open models for the free access and reuse of scientific knowledge are commonly catalogued under the labels of “Open Science” (OS) and “Open Research Data” (ORD), but their meaning is still vague and polysemantic in the literature. For this reason, we will try to untangle some ambiguities, by clarifying the terms of our discourse and presenting the legal transposition of OS and ORD.

---

<sup>1</sup>The reference is to Boyle (1997, p. 87).

<sup>2</sup>Stiglitz (2008) and Henry and Stiglitz (2010), who argue that poorly designed intellectual property regimes can impede innovation. According to them, there are alternative ways of organizing research—i.e. providing funding and incentives—that can help promote innovation and disseminate its results in a more efficient manner. See also the reflections in Nelson (2004).

<sup>3</sup>The term “bioinformation” has not yet been used in the meaning that is proposed here. An analysis of the literature shows it in three occurrences. In a first sense, it is used in bioinformatics to describe biological information in living organisms (see Paton 1996; Kanguane 2009); in a second sense, it is used in forensics to allude to DNA and fingerprints (Nuffield Council of Bioethics 2007); finally, in a third and generic sense, it is mentioned as a synonym of gene/genome (Milosavljevic 2000; Parry (2004)) or as information about the human body, Rose (2001).

Our analysis grew out of the realisation that despite the increasing advocacy towards the “openness” of science and research data, it is still far from being a widespread practice.<sup>4</sup> The goal of this paper is to identify the most pressing obstacles (in terms of funding, technology, IPRs, contracts, data protection, and social norms), which are blocking the development of OS and, in particular of ORD, with particular attention to the situation of developing countries.<sup>5</sup> The innovative aim of this paper, which is the first essay of a broader research, is to prepare the epistemological basis for a Law and Technology theory of “Open Bioinformation” (OB), where bioinformation stands for research data in life sciences. We argue that so far the literature has addressed the promotion of openness in science and research data only in a sectorial manner, taking into account just one or a few of the factors affecting openness as if they were not related or mutually influenced. Therefore, the suggested solutions are limited to a single perspective and fail to consider the dynamics of information control. In our view, a holistic approach, that tries to zoom out from the specific disciplines and take into account the whole picture, would contribute to determining an effective policy for promoting OB. For this reason, we have to consider the technological, legal, and sociological aspects, in order to assess whether and how changes in one domain might affect the others.

Once the causes of the problem have been identified, we will recommend some strategies and solutions that could make OB a more viable option. In particular, we will discuss two examples (“open through licenses” and “open through social norms”) where openness can be realized thanks to the combination of different strategies and legal tools.

---

## 2 There’s Something About Bioinformation: A Short Premise on Research Data for Life Sciences

If information is the blood and fuel of our world, indeed bioinformation is the vital principle of the current research methods in life sciences.<sup>6</sup> “Bioinformation” is an umbrella term we use to refer to information that is: (a) *biological*, i.e. of cellular and molecular human origin; (b) related to the *βίος*, the existential sphere of a person’s life; (c) *bioinformatic*, since computer programming is applied to the processing of biological data, which are digitized or born-digital; (d) *bioMedTech*, in the sense that it can be used for the purpose of medical or biotech research. This includes all information derived from biological samples or consisting of data generated by the individual or other subjects involved in the care/research process (physicians, researchers, nurseries, etc.). This can be, *inter alia*, data relating to the molecular or biochemical characteristics of the sample, genetic information, data generated in clinical trials, diagnosis, prescriptions, medical history, eating habits, etc.

---

<sup>4</sup>David and Foray (2002) and Pampel and Dallmeier-Tiessen (2014).

<sup>5</sup>From a comparative perspective, we must specify that no particular geographic area will be the object of the analysis: we will mention some general trends shared by the countries of the GS.

<sup>6</sup>Quoting James Gleick: “We can see now that information is what our world runs on: the blood and the fuel, the vital principle” (Gleick 2011).

The availability of this data is not only crucial for personalized medicine, but also a fundamental resource in many fields of bioscience research, since by linking genomic data or biochemical interactions with environmental factors and information relating to the illness' long-term course, we can improve our understanding of the causes or development of certain diseases (it is the idea currently behind research methods, for example, in genome-wide association studies, drug discovery, cancer research, translational medicine, pharmacogenomic investigations, etc.).<sup>7</sup>

Advances in technology and the convergence of different disciplines—computer science, biology, engineering, mathematics, and medicine—have helped to shape this kind of information as a new commodity<sup>8</sup>: nowadays, genome sequencing is faster and cheaper than at the end of the Human Genome Project<sup>9</sup>; data are more accurately annotated and can be stored in more widely available high-quality tools—such as computers, smartphones, and wireless devices; infrastructures like the new generation of research biobanks linked to electronic health records allow for professional and systematized collection<sup>10</sup>; the huge amount of data generated can be gathered in new kinds of storage spaces like the cloud<sup>11</sup>; data and information can be easily copied and transferred through digitization,<sup>12</sup> and so on.

Technology has greatly contributed to the potential of scientific progress, developing tools and infrastructures that allow for more and better information. Nevertheless, data collected by a researcher or a single institution, even a large one, are not sufficient to conduct a genome-wide association study or an evidence-based medicine project<sup>13</sup>: firstly, because data-intensive scientific discovery needs a huge amount of information from diverse sources; secondly, such investigations are intrinsically interdisciplinary, thus requiring collaboration from experts from different disciplines; thirdly, the skills, equipment and know-how are shared among stakeholders in both the public and private sector, making it necessary to overcome the traditional boundaries between the different players and build new forms of partnerships.<sup>14</sup>

Thus, progress in research requires a vast pool of scientifically reliable data, as well as expertise from different fields of knowledge and industry. Such a need has made data sharing, rather than an option, a categorical imperative for promoting

---

<sup>7</sup>West (2006).

<sup>8</sup>On the commodification of information caused by the expansion of the IPRs domain and the new possibilities opened up by technology, see Boyle (2003) and Hess and Ostrom (2003). With a specific focus on developing countries, Forero-Pineda (2006).

<sup>9</sup>The Human Genome Project (<http://www.genome.gov/10001772>) was a collaborative research program started in 1990 and aimed at sequencing the entire human genome. The first draft was published in 2001 (International Human Genome Sequencing Consortium: Lander et al. 2001), while the complete sequence was released in April 2003. At the end of the Human Genome Project the cost of the sequencing was around \$100 million and in 2014 was estimated at \$5,000. See Hayden (2014).

<sup>10</sup>Kohane (2011), Jensen et al. (2012), Scott et al. (2012) and Guarda (2013).

<sup>11</sup>Rosenthal et al. (2010) and Stein (2010).

<sup>12</sup>Topol (2013).

<sup>13</sup>Floca (2014, p. 298).

<sup>14</sup>In drug discovery, the collaboration among industries, academia, and other funders has been supported by Weigelt (2009). See also, Krumholz et al. (2014).

scientific progress (in the public interest) and, at the same time, for surviving in a highly specialized and competitive market (in the interest of private companies).<sup>15</sup> This is confirmed by the creation of networks of international research consortia that adopt collaborative policies and open access rules. The latter were codified in some soft law instruments, such as the Bermuda Principles (1996),<sup>16</sup> the Fort Lauderdale Agreement (2003),<sup>17</sup> the Amsterdam Principles,<sup>18</sup> or the Toronto Statement.<sup>19</sup> Many other initiatives from governments, international organizations and civil society have been supporting OA to scientific data over the last few years. To mention a few of them: the OECD *Principles and Guidelines for Access to Research Data from Public Funding* (2007)<sup>20</sup>; the EU Commission Communication on *Scientific information in the digital age: access, dissemination and preservation* (2007)<sup>21</sup>; the Panton Principles (2010)<sup>22</sup>; the Royal Society *Science as an open enterprise report* (2012)<sup>23</sup>; the UNESCO *Policy guidelines for the development and promotion of open access* (2012)<sup>24</sup>; the EU Commission Communication *Towards better access to scientific information: Boosting the benefits of public investments in research* (2012)<sup>25</sup>; the EU Commission Recommendation *on access to and preservation of scientific information* (2012)<sup>26</sup>; the *Open Research Data Pilot* in Horizon 2020.<sup>27</sup>

Despite the spread of an “open culture” and the common understanding of the need for data sharing in science, there is still confusion around terms like “Open Science” and “Open Research Data”. Actually, they are not clearly defined from a legal perspective. The next section aims at providing a coherent legal framework for such concepts.

---

<sup>15</sup>Hagedoorn et al. (2000) and Edwards et al. (2009).

<sup>16</sup>[http://web.ornl.gov/sci/techresources/Human\\_Genome/research/bermuda.shtml](http://web.ornl.gov/sci/techresources/Human_Genome/research/bermuda.shtml). Accessed 18.10.2014.

<sup>17</sup><http://www.genome.gov/Pages/Research/WellcomeReport0303.pdf>. Accessed 18.10.2014.

<sup>18</sup>Rodriguez et al. (2009).

<sup>19</sup>Toronto International Data Release Workshop Authors (2009).

<sup>20</sup><http://www.oecd.org/sti/sci-tech/38500813.pdf>. Accessed 18.10.2014.

<sup>21</sup>COM(2007)56, [http://ec.europa.eu/research/science-society/document\\_library/pdf\\_06/communication-022007\\_en.pdf](http://ec.europa.eu/research/science-society/document_library/pdf_06/communication-022007_en.pdf). Accessed 18.10.2014.

<sup>22</sup>Murray-Rust et al. (2010).

<sup>23</sup><https://royalsociety.org/~media/policy/projects/sape/2012-06-20-saoe.pdf>.

<sup>24</sup><http://unesdoc.unesco.org/images/0021/002158/215863e.pdf>. Accessed 18.10.2014.

<sup>25</sup>COM(2012) 401, [http://ec.europa.eu/research/science-society/document\\_library/pdf\\_06/era-communication-towards-better-access-to-scientific-information\\_en.pdf](http://ec.europa.eu/research/science-society/document_library/pdf_06/era-communication-towards-better-access-to-scientific-information_en.pdf). Accessed 18.10.2014.

<sup>26</sup>C(2012) 4890 final [https://ec.europa.eu/research/science-society/document\\_library/pdf\\_06/recommendation-access-and-preservation-scientific-information\\_en.pdf](https://ec.europa.eu/research/science-society/document_library/pdf_06/recommendation-access-and-preservation-scientific-information_en.pdf). Accessed 18.10.2014.

<sup>27</sup>The pilot was announced in 2013, [http://europa.eu/rapid/press-release\\_IP-13-1257\\_en.htm](http://europa.eu/rapid/press-release_IP-13-1257_en.htm). Accessed 18.10.2014; see Article 43 of the Regulation (EU) No 1290/2013 of the European Parliament and of the Council of 11 December 2013 laying down the rules for participation and dissemination in “Horizon 2020—the Framework Programme for Research and Innovation (2014–2020)” and repealing Regulation (EC) No 1906/2006; see also the Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020, version 1.0, 11 December 2013, [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf). Accessed 18.10.2014.

### 3 “Open Science” and “Open Research Data”: Finding the Definitions

Open science is a very popular concept in the current scientific debate, but its meaning seems to be defined and interpreted in different nuances.<sup>28</sup> An oft-cited definition comes from Stephen Maurer, who described the OS features around three pillars: “(a) full, frank, and timely publication of results, (b) absence of intellectual property restrictions, and (c) radically increased pre- and post-publication transparency of data, activities, and deliberations within research groups”.<sup>29</sup> More broadly, OS has been described as: “not only accessibility to research objects such as articles, data, code, protocols and workflows that people are free to use, re-use and distribute without legal, technological or social restrictions, but also the opening up of the entire research process—right from agenda—setting, data generation and data analysis, to dissemination and use”.<sup>30</sup>

For an overview of the OS phenomenon, it is useful to refer to the study by Fecher and Friesike, who from a literature review have identified at least five “schools of thought”<sup>31</sup>: (1) the so-called “Public School” emphasizes the need to make science understandable for the general public and the research process accessible to scientists; (2) the “Democratic School” stresses the importance of gaining access to the products of research (not only publications and data, but namely source materials, digital representations, multimedia materials); (3) the “Pragmatic School” promotes OS as a mechanism for making research more efficient; (4) the “Infrastructure School” deals with the challenges raised by the technical infrastructures that enable collaborative research projects through the web; (5) the “Measurement School” argues in favour of alternative and specific scientific impact factors for the digital age.

To adopt a strict notion of OS would be useless by definition, also considering the “open” nature of such a concept. Rather than five parallel lines, we imagine the different schools outlined by Fecher and Friesike as diverse points of view on the same phenomenon, showing us various ways of approaching it. They necessarily complement each other. The argument behind “openness” finds its root in the idea of Mertonian communalism,<sup>32</sup> but OS can alternatively be justified in light of utilitarian theories (it is better because it is more efficient). The promotion of sharing and collaboration among researchers shall be enabled through suitable online (and common) platforms and infrastructures. At the same time, such a system of sharing and dissemination of results can only withstand if scientists are given the right incentives. An open and wide diffusion of science materials is not only beneficial to professionals, but has to engage society more generally,

---

<sup>28</sup>Grubb and Easterbrook (2011) and Frischmann et al. (2014).

<sup>29</sup>Maurer (2003). In the same sense, Nielsen (2011).

<sup>30</sup>Open Knowledge Foundation (2014, p. 15).

<sup>31</sup>Fecher and Friesike (2014).

<sup>32</sup>Merton (1942).

empowering citizens. Sharing shall not be confined to scientific publications or materials, but extend to research data. The latter, in particular, are the object of the “Open Research Data” movement, a subcategory of the broader OS. Research data, such as those previously outlined as bioinformation, “form[s] the basis for the quantitative analysis underpinning many scientific publications”,<sup>33</sup> and they represent the fundamental building block of basic research.<sup>34</sup>

OS and ORD have certainly emerged as extra-legal phenomena, but they have begun to take on a legal dimension. Therefore, it is crucial to understand how they fit in the legal categories and what sources of law can be found in this field.

We can find some general normative indicators in the mandate to share scientific knowledge and benefits derived from them, affirmed by Article 27 of the Universal Declaration of Human Rights (1949), Article 15 of the Covenant on Economic, Social and Cultural Rights (1966), Articles 2 and 19 of the UNESCO Declaration on the Human Genome and Human Rights (1997), and Articles 2, 15, and 24 of the UNESCO Universal Declaration on Bioethics and Human Rights (2005), which explicitly take into consideration the importance of scientific data sharing for developing countries<sup>35</sup>; meanwhile, the relevance of a broad access to biological materials and genetic data has been affirmed by Articles 18 and 19 of the UNESCO International Declaration on Human Genetic Data (2003).

Despite the principles they affirm, these international documents only have a programmatic value. Their provisions are declamations and not binding and operative rules. Furthermore, they are not decisive for our discussion because they do not solve the main critical tension, that is the balance between free access to the benefits flowing from scientific knowledge and the exclusive rights granted by intellectual property law: to use the terms of the UNESCO Declaration on the Human Genome and Human Rights, such soft law statements echo, but do not

---

<sup>33</sup>European Commission (2012), point 3.

<sup>34</sup>The definition of research data is hard to find in the literature. According to some authors, because there is no consensus on the notion of data itself, it would be preferable to adopt a very broad approach: the term *research data* shall “include any kind of data produced in the course of scientific research, such as databases of raw data, tables, graphics, pictures or whatever else”. Dietrich and Wiebe (2013, p. 17). In the same sense, also the EU Guidelines on *Open Access to Scientific Publications and Research Data in Horizon 2020* which state that: “*Research data* refers to information, in particular facts or numbers, collected to be examined and considered and as a basis for reasoning, discussion, or calculation. In a research context, examples of data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images” (footnote 5, p. 3). See also Leonelli (2013b), according to whom: “scientific data can be defined as material artifacts that are collected and used as empirical evidence for the plausibility of claims about the nature of reality (‘the earth revolves around the sun’) and/or the efficacy of specific interventions (‘500 milligrams of paracetamol help to relieve headache’)”.

<sup>35</sup>Caulfield et al. (2012).

prejudice, the international instruments governing the IPRs framework.<sup>36</sup> The conflict with TRIPs Agreements was pointed out in the Report on Ethics, Intellectual Property and Genomics, issued by the International Bioethics Committee (IBC) in 2002.<sup>37</sup> It is interesting to note that this document explicitly mentions the term “open science”—understood in a narrow sense as the antithesis to a strong intellectual property rights protection on some pharmaceutical developments which is able to affect the right to life and health of millions of people, especially in the South of the World—but we should also note that the concerns expressed in it were not implemented in the subsequent UNESCO declarations.

The top-down approach does not solve our problem of finding the legal definitions. In fact, we should note that *openness* started to become familiar in the legal discourse from the bottom, and, in particular, with the advent of open source software and, later on, the open access movement.<sup>38</sup>

Open source software, born in the computer programming environment, is characterized by a decentralized production and a collaborative effort among everyone who wants to contribute to the programming of a piece of software.<sup>39</sup> Openness here concerns the source code of the software (i.e. the human-readable language), which is freely distributed. In this way, the program can be: run for any purpose; studied and modified as desired; redistributed as such; distributed with the modifications.<sup>40</sup> In order to keep the code open, a *viral* license is applied, which allows software to be freely used, modified, and shared, but both the code and any enhancement or derivative work must be shared on the same license terms.<sup>41</sup>

Open Access refers to research publications and its core has been recognized (and shaped) by the “Three Bs”, three declarations issued between 2002 and 2003, and resulting from three different initiatives: the Budapest Open Access Initiative Declaration (2002), the Bethesda Statement on Open Access Publishing (2003), and the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (2003). Its main features have been effectively summarized by Peter Suber, who has described OA as a literature that is “digital, online, free of charge,

---

<sup>36</sup>Namely, the Berne Convention for the Protection of Literary and Artistic Works of 9 September 1886 and the UNESCO Universal Copyright Convention of 6 September 1952, as last revised at Paris on 24 July 1971, the Paris Convention for the Protection of Industrial Property of 20 March 1883, as last revised at Stockholm on 14 July 1967, the Budapest Treaty of the WIPO on International Recognition of the Deposit of Micro-organisms for the Purposes of Patent Procedures of 28 April 1977, and the Trade Related Aspects of Intellectual Property Rights Agreement (TRIPs) annexed to the Agreement establishing the World Trade Organization, which entered into force on 1 January 1995.

<sup>37</sup><http://unesdoc.unesco.org/images/0013/001306/130646e.pdf>. Accessed 18.10.2014. See Kuppuswamy (2009, p. 137 and ff).

<sup>38</sup>Caso and Ducato (2014).

<sup>39</sup>See Di Bona and Ockman (1999), Raymond Raymond (2000) and Stallman (2002).

<sup>40</sup>These are the four fundamental freedoms established by the General Public License manifesto: <https://www.gnu.org/gnu/manifesto.en.html>. Accessed 18.10.2014.

<sup>41</sup>Probably the best known example is the GNU GPL license, created by Richard Stallman. Stallman (1998).

and free of most copyright and licensing restrictions”.<sup>42</sup> OA philosophy, thus, recognizes an unrestricted access and reuse of e-contents through the Internet, and contractual tools as the operative solution for doing so. In this sense, the Creative Commons licenses, developed since 2002, have been a valuable instrument for supporting the implementation of OA in a concrete way.<sup>43</sup> Through these modular and user-friendly licenses, the way of sharing digital content—not necessarily creative (it is possible for example to waive the *sui generis* right on databases)—has been radically changed, because the author is free to choose their copyright settings.<sup>44</sup>

Filtering the precipitate of both open source and open access in order to infer a legal meaning of the concept of *openness*, we can observe at least two aspects in which the law can operate: firstly, openness involves a limitation of IPRs; secondly, accessibility to a specific resource is managed through licenses or contracts.

The foregoing observations hopefully clarify the terms of our analysis. We can now proceed to analyse the dynamics of ORD in its operational reality, paying particular attention to the situation of developing countries.

---

## 4 Open Bioinformation in the Developing World: An Overview

If ORD is crucial for the promotion of innovation and development in technologically advanced countries, it is even more so for the developing world, where openness is now considered a possible way for lifting the traditional barriers between the North and the South of the globe.<sup>45</sup>

In the public health field, the promotion of access to scientific information as a means for overcoming the inadequate institutional, infrastructural and regulatory capacity to conduct high-quality investigations in Africa has been strongly affirmed by the Algiers Declaration on “Narrowing the Knowledge Gap to Improve Africa’s Health” (2008).<sup>46</sup>

Besides the declamations, some projects based in the developing world are starting to promote collaborative science and ORD/OB in a concrete way: it is the case of the Human Heredity and Health in Africa (H3Africa) Consortium,<sup>47</sup> which aims at building a network for engaging African countries in the genomic

---

<sup>42</sup>Suber (2012). For a complete overview of the OA movement, see Frosio (2014); meanwhile for a specific focus on academic publications, Moscon (2015).

<sup>43</sup><https://creativecommons.org/>. Accessed 18.10.2014.

<sup>44</sup>Lessig (1999), Carroll (2006) and Goss (2007).

<sup>45</sup>Rahman (2012, p. 7).

<sup>46</sup>The Algiers Declaration was issued by the ministers of health and heads of delegation of African countries, during the Ministerial Conference on Research for Health in the African Region, held in June 2008.

<sup>47</sup>Ramsay et al. (2014).

revolution; the MalariaGEN,<sup>48</sup> a data-sharing community studying malaria by integrating epidemiology with genomics; the Gambian National DNA Bank,<sup>49</sup> the first biobank created in Africa in collaboration with the Jean Dausset Foundation-CEPH that promotes the sharing of collected information; the Malaysian Oral Cancer Database and Tissue Bank System (MOCDTBS),<sup>50</sup> which makes data and specimens available to researchers; or the *Datos Científicos Abiertos* Program,<sup>51</sup> launched by the *Comisión Nacional de Investigación Científica y Tecnológica* (CONICYT) of Chile for promoting best practices and the creation of a policy for sharing scientific data.<sup>52</sup> Several initiatives are coming up from the bottom. A paradigmatic example is the OpenSciDev Group, attributable to the Open Knowledge Foundation, whose goal is to set the agenda for the realization of an open and collaborative science in the developing world.<sup>53</sup>

OS, ORD and OB in particular, are becoming extremely popular because they can potentially solve some age-old problems of the GS, *in primis* the availability and the equal distribution of information and knowledge. As pointed out by the OpenSciDev Group with reference to publications, academic and commercial journals are inaccessible to most of the researchers and institutions in developing countries due to the high cost of subscription.<sup>54</sup> Such a situation creates a vicious circle, because limited access to research resources reduces the chances of authors from the GS of being published in international journals, and their underrepresentation implies at least two important consequences:<sup>55</sup> (1) a limited visibility and a low impact factor of developing-country (DC) researchers (and, as a result, they have little chance of spreading their ideas, being quoted, being involved in collaborative research projects, having access to opportunities for training abroad, etc.); (2) a reduced ability for institutions both of the North and the South to know the research generated in a certain DC, thus preventing both North-South and South-South collaborations.

The same applies to scientific data produced in life sciences research. Those fields are highly expensive, requiring a huge amount of investment for the gathering of samples, data and analysis: the cost of laboratories, chemicals, reagents,

---

<sup>48</sup><http://www.malariagen.net/>. Accessed 18.10.2014. For an overview of their data-release policy, see Parker et al. (2009). MalariaGEN is a network that includes several participants from different countries, thus enacting a North-South collaboration.

<sup>49</sup>Sirugo et al. (2004).

<sup>50</sup>Zain et al. (2013).

<sup>51</sup><http://datoscientificos.cl/>. Accessed 18.10.2014.

<sup>52</sup>Muñoz Palma (2012).

<sup>53</sup><http://opensci.dev.com/>. Accessed 18.10.2014. One of the most interesting things is the *modus operandi* adopted by such a group, which shares all its documents online via googledocs. So, everyone who wants to contribute to the project and working papers can suggest some edits and comment on the files. All documents are licensed under CC-BY 4.0.

<sup>54</sup>Open Knowledge Foundation (2014, p. 17 and ff). See also Chan et al. (2005) and Dulle et al. (2013).

<sup>55</sup>Chan et al. (2005), Czerniewicz and Goodier (2014) and Veldsman and Gevers (2014).

machinery, equipment, specialised and trained personnel is unaffordable for most DCs.<sup>56</sup> To give an overview of the costs, we can mention the well-known example of the Human Genome Project.<sup>57</sup> The US government invested about \$2.7 billion from 1990 to 2003 in the collaborative research program aimed at the sequencing of the entire human genome. Just to have an idea of the scale, the cost of a single research project is approximately equivalent to the GDP of Burundi in 2013.<sup>58</sup>

Almost all research in DCs is conducted with scant public funding and the partnership with industry is not well implemented, so the sharing of bioinformation is of paramount importance for carrying out data-intensive research in those DCs that would otherwise be cut off from the research net.<sup>59</sup> An open approach, supported by a decent ICT infrastructure and sufficient expertise, could offer a cost-effective solution for performing research with limited resources.<sup>60</sup>

OB can also foster participation and engagement in a research project.<sup>61</sup> This is of particular importance in life sciences research, where the success of an investigation may depend on the collaboration, in some cases, of a group of people or an entire population. A democratization of the whole process, the so-called “partnership governance”,<sup>62</sup> incorporating research participants and giving them decision-making power, would allow citizen empowerment and increase trust in the organization conducting the research.<sup>63</sup>

Thus, strong altruistic and economic arguments support the promotion of OB, but there is a further point to consider, which has ethical implications. A great number of DCs represent a sort of new ‘goldmine’ for biotechnologically advanced countries. Populations from low-income countries can be the source of a valuable pool of data, because of the genetic peculiarities of a certain ethnic group or, sadly,

---

<sup>56</sup>Hardy et al. (2008).

<sup>57</sup><http://www.genome.gov/10001772>. Accessed 18.10.2014.

<sup>58</sup>According to the data of the World Bank. See [http://data.worldbank.org/indicator/NY.GDP.MKTP.CD?order=wbapi\\_data\\_value\\_2013+wbapi\\_data\\_value+wbapi\\_data\\_value-last&sort=asc](http://data.worldbank.org/indicator/NY.GDP.MKTP.CD?order=wbapi_data_value_2013+wbapi_data_value+wbapi_data_value-last&sort=asc). Accessed 18.10.2014.

<sup>59</sup>Hardy et al. (2008), Gómez and Bongiovani (2012), Muñoz Palma (2012, p. 24) and Rahman (2012, p. 15).

<sup>60</sup>Open Knowledge Foundation (2014, p. 20).

<sup>61</sup>Ibid., p. 22; Tindana et al. (2007).

<sup>62</sup>Winickoff (2009). The model for realizing such a partnership governance could be found in the charitable trust, according to Winickoff and Winickoff (2003).

<sup>63</sup>According to Frischmann, Madison, and Strandburg: “commons governance offers a defense against potential privatization of commonly useful shared resources and the possibility that an individual IP rights owner would “hold up” the enterprise as a whole. Examples of such arrangements might include “open source” commons constructed for basic biological building blocks such as the Single Nucleotide Polymorphism (SNP) consortium or the publicly available databases of genomic sequences that are part of the Human Genome Project. Formal licenses and related agreements assure that participants become part of what amounts to a mutual nonaggression pact that is necessary precisely because of the possibility that intellectual resources may be propertized” (Frischmann et al. 2014, p. 26).

because patients affected by the diseases are based there.<sup>64</sup> After the collection of biological samples and information, research is conducted in developed countries and the results (new drugs, treatments, diagnostic methods, vaccines, etc.) are not always granted back to research participants, thus raising several ethical and benefit-sharing concerns.<sup>65</sup> It would be fair and compliant with the international principles mentioned above to make freely available at least the data and the analysis generated from the screening of DCs' population, allowing local scientists to reuse them for the needs and priorities of local research.<sup>66</sup>

Even though OB represents a new hope for the GS, it is not a common practice and it is facing several obstacles. From a literature review, we have counted six variables that affect the openness of data, and, in particular, bioinformation: (1) public investment; (2) technology; (3) intellectual property; (4) contracts; (5) privacy; (6) social norms.

- (1) The origin of every problem related to OA can be traced back to funding and sustainable business plans for the long term.<sup>67</sup> In the GS, basic research is carried out with an insufficient amount of public money.<sup>68</sup> As already outlined, the lack of public-private partnerships does not help overcome such an impasse. This can result in inadequate lab equipment, resources, and libraries, the lack of educational and training programs for specialised personnel, a weak ICT infrastructure, etc.<sup>69</sup>
- (2) OB may be hampered by technology: the lack of ICT infrastructures or their inability to share and re-use information, hindering the database interoperability or data portability, constitutes a serious weak point in the very possibility of data sharing.<sup>70</sup> The process of integrating data depends on the adoption of standards which ensure the source (metadata) and the data curation.<sup>71</sup> In the GS the problem is exacerbated by poor digitization of information and limited access to the Internet.<sup>72</sup>

---

<sup>64</sup>Sgaier et al. (2007).

<sup>65</sup>Costello and Zumla (2000), Cambon-Thomsen (2004), Dickenson (2004), Knoppers (2005) and Parker et al. (2009). For an overview of the main critical issues of such a practice, see also de Vries et al. (2011).

<sup>66</sup>Knoppers (2000).

<sup>67</sup>Bastow and Leonelli (2010). The study by Halla Thorsteinsdóttir, Uyen Quach, Abdallah S. Daar and Peter A. Singer shows that political will and public investments have been crucial for the development of health biotechnology in seven developing countries (Brazil, China, Cuba, Egypt, India, South Africa, and South Korea), which have been taken into account as case studies (Thorsteinsdóttir et al. 2004).

<sup>68</sup>Muñoz Palma (2012), Mboera (2012) and Inyang (2012).

<sup>69</sup>Sirugo et al. (2004), Hardy et al. (2008), Mboera (2012) and Rahman (2012, p. 8).

<sup>70</sup>De Roure et al. (2003), Altunay et al. (2010) and Leonelli (2013a).

<sup>71</sup>Ankeny and Leonelli (2015).

<sup>72</sup>Kahn (2012), Mboera (2012), Leonelli (2013b) and Open Knowledge Foundation (2014, p. 37).

- (3) The complex landscape of intellectual property rights and the uncertain legal status of data are a serious disincentive to collaborative research.<sup>73</sup> The commodification and enclosure of data may appear in the guise of copyright and *sui generis* database right protection. Such IPRs, although designed for databases, ultimately end up affecting the contents of the database itself.<sup>74</sup> In particular, the *sui generis* right has been strongly criticized for its potentially negative consequences, such as the danger of creating monopolies, the increase of transactions costs, the interference with data aggregation, and the negative impact on the cooperative ethos.<sup>75</sup>
- (4) The private control of bioinformation is indeed exercised through contracts, as in the case of Data Transfer Agreements (DTA). These can be effectively enforced through technological measures that are designed to manage and protect the rights of access and use of digital contents, including through the immediate and timely sanction of any violation of the contract conditions.<sup>76</sup> Mastering the jungle of the terms of agreements is far from a trivial task, and it inevitably involves transactional costs,<sup>77</sup> which are incompatible with the timelines of scientific research.<sup>78</sup>
- (5) The rationale of OB is potentially in conflict with the right to privacy and confidentiality.<sup>79</sup> Just to mention the two biggest legal models for data protection, in Europe, Directive 95/46/EC<sup>80</sup> and Directive 2002/58/EC<sup>81</sup> frame the general rules, which will be profoundly affected by the new Regulation, with particular reference to the treatment of personal data for scientific research<sup>82</sup>; meanwhile, the US has sector-specific federal legislation (the HIPAA; the Federal Drug and Alcohol Confidentiality Statute; the Common

---

<sup>73</sup>Guibault and Wiebe (2013). See also, Reichmann and Uhlir (2003).

<sup>74</sup>Trosow (2004) and Davison and Hugenholtz (2005).

<sup>75</sup>Reichman and Samuelson (1997), Reichman and Uhlir (1999), David (2000), Reichman and Uhlir (2003, pp. 396 and ff.), David (2004) and Trosow (2004).

<sup>76</sup>Dussollier (2002), Caso (2004) and Ginsburg (2005).

<sup>77</sup>Guibault and Margoni (2013).

<sup>78</sup>Reichman and Uhlir (2003, pp. 402–404), Streitz and Bennett (2003) and Margoni (2013).

<sup>79</sup>Kaye (2012), Hoffman (2014) and Mascalzoni et al. (2014).

<sup>80</sup>Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, published in the Official Journal L 281, 23.11.1995, pp. 0031–0050.

<sup>81</sup>Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications), published in the Official Journal L 201, 31.07.2002, pp. 0037–0047.

<sup>82</sup>See, in particular, Article 83 of the Draft of the General Data Protection Regulation [COM(2012) 11 final, 25.01.2012]. The text of the proposal is available at the following link: [http://ec.europa.eu/justice/data-protection/document/review2012/com\\_2012\\_11\\_en.pdf](http://ec.europa.eu/justice/data-protection/document/review2012/com_2012_11_en.pdf). Accessed 18.10.2014.

Rule; the GINA Act, etc.).<sup>83</sup> The basic principle in both jurisdictions is the obtainment of the data subject's informed consent. Such rules, designed to protect a fundamental right, pose de facto (legitimate) restrictions and exemptions to OB.<sup>84</sup>

Another fundamental point stressed in the literature is that such protection only comes into play if information relates to an identified or identifiable person (*ex multis*, Article 2, Directive 95/46/EC; Article 5, § LXXII, Brazilian Constitution; Article 2, *Ley de Argentina* 25326/2000; Article 4(d), *Ley de Uruguay* 18331/2008; Article 3(b), *Ley de Costa Rica* 8968/2011; Chap. 1, § 55, South Africa Act 4/2013), or protected health information (PHI) that “does not identify an individual” or allow “a reasonable basis to believe that the information can be used to identify an individual” (HIPAA). Such an objective scope is critical because it is failing on that technological premise on which all data protection legislations have relied for reaching a balance between the protection of the individual and the free movement of information: anonymization.<sup>85</sup> Several studies show the increasing possibility of re-identifying individuals from anonymized data,<sup>86</sup> suggesting that anonymization is a promise that cannot be maintained in absolute terms in the digital environment.<sup>87</sup>

- (6) Finally, it is fundamental to take into consideration social norms, and, in this case, the scientific ethos. Despite the Mertonian principles,<sup>88</sup> researchers are not ontologically inclined to share their data for a number of reasons<sup>89</sup>: creating a dataset costs time, money and labour and they are not willing to share it without some form of compensation; sharing would eliminate the competitive advantage; the quality of a dataset might determine how grants are awarded, with consequent benefits in terms of career advancement and livelihood of the research group.<sup>90</sup> The lack of adequate economic or reputational incentives risks inhibiting the informal exchange of information within the scientific community. In the GS, such a problem seems to be one of the most difficult to address, since several scholars denounce the practice of secrecy as a common behaviour in the community of peers and the lack of a culture of sharing.<sup>91</sup>

---

<sup>83</sup>Health Insurance Portability and Accountability Act [(45 C.F.R. § 160–164 (2002)]; Federal Drug and Alcohol Confidentiality Statute (42 U.S.C. § 290dd-2); the Common Rule [45 C.F.R. § 46.101 (2005)]; Genetic Information Nondiscrimination Act (42 U.S.C. § 2000ff et seq.).

<sup>84</sup>Schwartz (1994), Solove (2004) and Floca (2014).

<sup>85</sup>Ohm (2010).

<sup>86</sup>Gymrek et al. (2013).

<sup>87</sup>Lunshof et al. (2008).

<sup>88</sup>Merton (1942).

<sup>89</sup>Borgman (2007).

<sup>90</sup>Gitter (2013).

<sup>91</sup>Mboera (2012) and Rahman (2012, p. 8).

All these factors should be considered in order to design an effective policy for OB, because they mutually influence each other. In order to provide a preliminary analysis of these complex dynamics, we will examine two cases in which we can observe the interactions among some of the abovementioned variables for achieving openness of bioinformation: the first one touches upon the limitation of IPRs through licenses and social norms, while the second one focuses on how to shape the social norms of the scientific community by using incentives and legal tools.

---

## 5 IPRs in Data?

In order to solve the first set of issues, a premise is needed: we have to understand which type of IPRs can be applied to data. In contrast to secret information or the end-product (publications or inventions), the application of an exclusive right in factual data is highly problematic. There is no legal definition of data nor a specific regulation for them.<sup>92</sup> The word “data” (*datum*, in Latin) comes from the ancient Greek *dedomena*, that literally means “difference”. According to a general notion, they are uninterpreted variables not processed by a cognitive intervention.<sup>93</sup> If there is no human intervention, strictly speaking, the necessary precondition for intellectual or industrial property is missing.<sup>94</sup>

Nevertheless, IPRs can indirectly affect data management and circulation through the legal regime applicable to the collections of data. Compilations and databases, in fact, can be protected by copyright and, in some jurisdictions, also by the so-called *sui generis* right (SGR).<sup>95</sup> Collections of data are eligible for copyright protection if they constitute, as a whole, an original work of authorship, whose creativity is expressed through the selection, coordination or arrangement of data and materials.<sup>96</sup>

---

<sup>92</sup>The only one legally described and expressly regulated is personal data, which is protected in accordance with national and international data protection rules.

<sup>93</sup>Floridi (2010, pp. 25–28).

<sup>94</sup>As is well known, copyright protects original works of authorship, but not facts or ideas; meanwhile, patent law grants the temporary monopoly for an invention that is new, involves an inventive step and is susceptible of industrial application. A right of property in data can be detected also in the provisions regarding the protection of a certain type of information, as in the case of know-how (see Article 39 TRIPS).

<sup>95</sup>For a general overview, see Derclaye (2014).

<sup>96</sup>Such a principle is valid on both sides of the Atlantic. The US system, in fact, protects compilations “as a work formed by the collection and assembling of pre-existing materials or of data that are selected, coordinated, or arranged in such a way that the resulting work as a whole constitutes an original work of authorship” (17 U.S.C. § 101); meanwhile, Directive 96/9/EC on the legal protection of databases states that “databases which, by reason of the selection or arrangement of their contents, constitute the author’s own intellectual creation shall be protected as such by copyright. No other criteria shall be applied to determine their eligibility for that protection” (Article 3). The case law has confirmed the legislative component in the leading case *Feist v. Rural*, 499 U.S. 340 (1991) for the US system and in the ECJ Case C 5/08 *Infopaq*

In some jurisdictions, non-creative databases can also be protected<sup>97</sup>: this is the case of the *sui generis* right recognized in the EU and Mexico, and the sweat of the brow doctrine accepted in South Africa.<sup>98</sup>

The EU legal system grants a 15-year protection period of protection to “the maker of a database which shows that there has been qualitatively and/or quantitatively a substantial investment in either the obtaining, verification or presentation of the contents to prevent extraction and/or re-utilization of the whole or of a substantial part, evaluated qualitatively and/or quantitatively, of the contents of that database” (Article 7, Directive 96/9/EC).

The substantial investment, and not creativity, is the precondition for the exercise of the SGR; furthermore, the work of the maker of the database must be directed to the gathering, validation or presentation of data. The interpretation of such requisites has given rise to some contrasts. In particular, the meaning of obtaining and verification has been at the centre of a hermeneutical dispute before the European Court of Justice (ECJ). In *British Horseracing Board v. William Hill*

---

(Footnote 96 continued)

International [2009] ECR I 6569 and Case C-604/10 *Football Dataco Ltd and Others v Yahoo! UK Ltd and Others* [2012] ECDR 7 for the EU.

<sup>97</sup>For historical accuracy, we have to mention that before the introduction of Directive 96/9/EC, a similar right, namely the “catalogue rule”, already existed in Scandinavian countries (Karnell 1997). Also the US and the Australian systems used to protect the non-creative databases, applying the sweat of the brow doctrine, according to which copyright rewards the efforts and work that go into a compilation of facts. Such a principle was rejected in the US since the notorious case *Feist v. Rural* (1991), where the Court affirmed: “Without a doubt, the ‘sweat of the brow’ doctrine flouted basic copyright principles. Throughout history, copyright law has ‘recognized a greater need to disseminate factual works than works of fiction or fantasy’. *Harper & Row*, 471 U.S., at 563. [...] But ‘sweat of the brow’ courts took a contrary view; they handed out proprietary interests in facts and declared that authors are absolutely precluded from saving time and effort by relying upon the facts contained in prior works. In truth, ‘it is just such wasted effort that the proscription against the copyright of ideas and facts... [is] designed to prevent’ [...] Protection for the fruits of such research... may in certain circumstances be available under a theory of unfair competition. But to accord copyright protection on this basis alone distorts basic copyright principles in that it creates a monopoly in public domain materials without the necessary justification of protecting and encouraging the creation of ‘writings’ by authors”. For a comment, see Fulwood (1991), Ginsburg (1992) and Strong (1994). For the sake of completeness, it should be noted that after the enactment of Directive 96/9/EC, the US Congress tried to re-insert an exclusive right model for database protection similar to the SGR with some legislative proposals in 1996 and 2000. See Reichman and Uhler (2003). The Australian jurisprudence arrives at the same conclusion in the cases *IceTV Pty Ltd. v. Nine Network, Australia Pty Ltd.* (2009) and *Telstra Corporation Limited v Phone Directories Company* (2010) (Lindsay 2012).

<sup>98</sup>For a general overview of the *sui generis* right in Europe, see Stamatoudi (1997) and Derclaye (2008), (2014). An introduction to the Mexican provisions regarding the legal protection of databases can be found in Ovilla Bueno (1998), Caballero Leal (2000) and De La Parra Trujillo (2004).

*Organization Ltd* (2004)<sup>99</sup> and the three *Fixtures* cases (2004),<sup>100</sup> the European Court distinguished between “obtaining” and “creation”: the database is eligible for the *sui generis* protection only if the aim of the investment is to “seek out existing independent materials and collect them”,<sup>101</sup> but not if the effort is directed at the “resources used for the creation as such of independent materials”.<sup>102</sup> The activity of verification implies the ensuring of the reliability of the information contained in a database. Thus, according to the ECJ, the substantial investment has to be evaluated only with regard to those resources used “to monitor the accuracy of the materials collected when the database was created and during its operation”<sup>103</sup> and not those “used for verification during the stage of creation of data or other materials which are subsequently collected in a database”,<sup>104</sup> because they are resources used during a database creation.<sup>105</sup> In other words, the ECJ tried to “domesticate”<sup>106</sup> the SGC recalling the utilitarian rationale of the directive, that is the protection of data storage and the encouragement of processing system development and not the creation of new informational resources like data and materials.<sup>107</sup>

The exclusive right attributed to the maker of the database is particularly pervasive because it helps prevent a lawful user of the database from extracting and/or re-utilizing substantial parts of its contents, evaluated qualitatively and/or quantitatively, and impede the repeated and systematic extraction and/or re-utilization of insubstantial parts of the contents of the database if in conflict with a normal exploitation of that database or with the legitimate interests of the maker of the database (see, in particular Articles 7 and 8, Directive 96/9/EC).

The directive contains a temperament for the abovementioned control by the maker of the database, allowing Member States to implement specific exceptions to SGR, such as in the case of extraction for the purposes of illustration for teaching or

---

<sup>99</sup>ECJ Case C-203/02, *British Horseracing Board v. William Hill Organization Ltd* (BHB) [2004], ECR I-10415.

<sup>100</sup>ECJ Case C-338/02, *Fixtures Marketing Ltd v. Svenksa AB* (Svenska), [2004] ECR I-10497; ECJ Case C-444/02, *Fixtures Marketing Ltd v. Organismos Prognostikon Agonon Podosfairou EG* (OPAP), [2004], ECR I-105449; ECJ Case C-46/02, *Fixtures Marketing Ltd v. Oy Veikkaus Ab* (Oy Veikkaus), [2004] ECR I-10365.

<sup>101</sup>*British Horseracing Board v. William Hill*, para 31.

<sup>102</sup>*Ibidem*.

<sup>103</sup>*Ibid*, para 34.

<sup>104</sup>*Ibidem*.

<sup>105</sup>Although the ECJ seems to make a clear distinction, in several cases it can be very hard to find a difference between the obtaining and creation of scientific data. The terms of the debate can be efficiently summarized by referring to the two points of view expressed by Derclaye (2004) and Davison and Hugenholtz (2005).

<sup>106</sup>Davison and Hugenholtz (2005).

<sup>107</sup>As the Court motivates, in fact: “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”. *British Horseracing Board v. William Hill*, para 34.

scientific research, as long as the source is indicated and to the extent justified by the non-commercial purpose to be achieved (Article 9, Directive 96/9/EC). This (shiny) attempt at openness has not been transposed across the whole Union in a uniform way, remaining a dead letter in many legal systems like Italy and Spain.<sup>108</sup>

Furthermore, we have to consider the duration of the SGR: it arises automatically from the date of completion of the database but the period of protection begins to run afresh after any substantial change, evaluated qualitatively or quantitatively, to the contents of a database, including any substantial change resulting from the accumulation of successive additions, deletions or alterations, which would result in the database being considered to be a substantial new investment, evaluated qualitatively or quantitatively. In that case, the database resulting from that investment shall qualify for its own term of protection (Article 10, Directive 96/9/EC).

The vagueness of the European *sui generis* right and of its scope have raised several concerns from a legal point of view. The “rolling” duration, the difficulties in distinguishing in practice between “obtaining” and “creation” of data, the unclear policy about publicly funded databases,<sup>109</sup> and the limited scope of the SGR exceptions make such a right “one of the least balanced and most potentially anti-competitive intellectual property rights ever created”.<sup>110</sup>

Similar policy considerations can be made with reference to the Mexican SGR, although we should point out that such a legal model has shortcomings and has been poorly developed. Article 108 of the *Ley Federal del Derecho de Autor* (1996) only states that: “Las bases de datos que no sean originales quedan, sin embargo, protegidas en su uso exclusivo por quien las haya elaborado, durante un lapso de 5 años”. Interpreting in a systematic way such a provision, we can infer that all non-creative databases, regardless of any evaluation of the effort for establishing them, are protected by the Mexican SGR for a period of 5 years. Furthermore, in contrast to the European solution, the SGR cannot be cumulated with copyright: original database are covered by the *derecho de autor*, meanwhile non-original databases can be protected through the SGR.<sup>111</sup> Even though the duration is shorter than the European SGR, the objective requirements are broader and the SGR extends to all non-creative Mexican databases, without taking into account any further conditions.

---

<sup>108</sup>Ducato (2013) and Guibault and Wiebe (2013).

<sup>109</sup>Only The Netherlands has explicitly denied a public authority the ability of exercising the SGR (Article 8, Dutch Database Act). See, Guibault (2013). Although not expressly recognized by the legislative component, also in the Italian legal system it is possible to reach the same conclusion. Legal scholars have, in fact, observed an irresolvable contradiction between the industrial or commercial rationale protected by the Directive and the public goals pursued by a public administration, rejecting the application of the SGR to publicly funded databases. See, Cardarelli (2002). The same principle has been confirmed also by the case law and precisely by Tribunale di Roma, Sez. IP, ordinanza 5 giugno 2008, Edizioni Cierre s.r.l. v. Poste Italiane s.p.a., in AIDA, 2010, 688.

<sup>110</sup>Reichman and Samuelson (1997).

<sup>111</sup>See De La Parra Trujillo (2004).

In South Africa the sweat of the brow doctrine is still a cornerstone of copyright protection.<sup>112</sup> Contrary to the holding of the *Feist* case, the South African High Court has recently affirmed the copyright infringement in the case *Board of Healthcare Funders v. Discovery Health Medical Scheme and Others* (2012), since the latter used, published and adapted the contents of applicants' Practice Code Numbering System ("PCNS"). The PCNS is a database that includes personal data related to medical practitioners (name, address, bank account details, preferred payment methods, etc.) and codes for medical service providers, attributing to such information a unique identifying number. In stating the violation of the Copyright Act, the South African Court interpreted the originality requirement adopting a very low standard: "There is little doubt if regard be had to the work and energy put in over the three phases of the development of the PCNS that indeed while some of the component parts may not necessarily be original in its totality the work could be said to be original. It would be cynical to suggest that no effort or skill was expended in the development of the system over the years and in my view the respondents' stance that the work lacks originality must be dismissed in the light of the meaning that has come to be attached to the concept of originality in the case law developed over the years".<sup>113</sup>

## 5.1 Open Through Licenses

The limits imposed by IPRs in scientific databases, through the long arm of the control offered by copyright and SGR, but also the uncertainty about the legal status of a dataset (as seen in the case of Europe, Mexico, and South Africa) may hinder both the regional and the transnational circulation of information. In these circumstances, a viable solution towards open models can be pursued through a legal agreement: "since the legal status of scientific databases and their content is more difficult to assess [...], the use of standard licenses would eliminate the need for the user to look for the rights owner and to negotiate the terms of use".<sup>114</sup>

Several models of standard licenses, in the form of user-friendly web tools, have been developed over the last few years allowing the exercise of IPRs on digital content according to the needs and wishes of the author. Probably the most well-known example are the Creative Commons (CC) licenses.<sup>115</sup> Such legal

---

<sup>112</sup>Pistorius (2008).

<sup>113</sup>The Court here refers to *Bosal Africa (Pty) Ltd v Grapnel (Pty) Ltd & Another* 1985 4 SA 482 (C); *Payen Components SA Ltd v Bovic CC and Others* 1995 4 SA 441; *CCH Canadian Ltd v Law Society of Upper Canada* [2004] 1 SCR 339; *Haupt t/a Soft Copy v Brewers Marketing Intelligence (Pty) Ltd and Others* 2006 4 SA 458 (SCA).

<sup>114</sup>Guibault and Margoni (2013, p. 148). See also, Aliprandi (2011) and Leucci (2014).

<sup>115</sup>Creative Commons (CC) is a charitable corporation that promotes the sharing and circulation of knowledge in compliance with copyright law. Although it offers standardized models, its modular licenses (attribution, non-commercial, no derivative works, share alike) and their combinations can provide flexibility in setting the interests of the parties. <http://creativecommons.org/>.

instruments, created by the ingenuity of Lawrence Lessig, offer both professionals and laymen a simple way to manage copyright and, as far as we are concerned, also database rights. CC licenses are, in fact, designed in three main layers: (1) the Legal Code, that is the full text of the license; (2) the Common Deed, or the “human-readable” version that summarizes in an effective way (also through the use of icons) the main conditions of the license; (3) the “machine-readable” version of the license, which is written in a software format that computers can understand.<sup>116</sup>

There are essentially three types of CC license that can promote the principles of data openness in different nuances<sup>117</sup>:

- CC0 (“No Right Reserved”).<sup>118</sup> Rather than a license, it is a waiver according to which the author dedicates the work to the public domain by giving up all of his or her rights to the work worldwide.<sup>119</sup> In our case, it means that, for example, everyone can copy, modify, or distribute a substantial part of a database, even for commercial purposes, without asking permission and before the expiration of the 15-year period.
- CC-BY-4.0 (“Attribution”).<sup>120</sup> Solving a gap affecting the previous ones, the latest version (4.0) of this license applies also to data, since it expressly includes the copyright on database and the SGR.<sup>121</sup> Under the terms of this agreement, the licensor grants a worldwide, royalty-free, non-sublicensable, non-exclusive, irrevocable license to reproduce and share his/her creation, in whole or in part, and to produce, reproduce, and share any modification of the same. The only obligation of the user is to give credit to the creator in any reasonable manner requested by the licensor, provide a link to the license, and indicate if changes were made.<sup>122</sup> Another innovation of the version 4.0, attractive for researchers, relates to the attribution requirements: in addition to the obligation of indicating the URI (Uniform Resource Identifier), to the extent reasonably practicable, the new CC-BY includes also the possibility of indicating the hyperlink to the licensed material. In this way, credit attribution is flexible and allows an easier compliance especially in the case of datasets.<sup>123</sup>

<sup>116</sup><https://creativecommons.org/licenses/>.

<sup>117</sup>Creative Commons provides two other options, namely “non-commercial” and “no-derivatives”. See, Guibault (2013).

<sup>118</sup><http://creativecommons.org/about/cc0>.

<sup>119</sup>Aliprandi (2011, p. 33).

<sup>120</sup><http://creativecommons.org/licenses/by/4.0/>.

<sup>121</sup>Guibault (2013). For a critical analysis of the previous exclusion of the database SGR from the scope of the CC licenses, see Guibault (2011).

<sup>122</sup><http://creativecommons.org/licenses/by/4.0/legalcode>.

<sup>123</sup>On the other hand, such a possibility carries on the problem of the links’ expiration, which de facto is able to cross the attribution obligation. For a general overview of the problem for digital publication, see Kling and Callahan (2003).

- CC-BY-SA-4.0 (“Attribution-Share Alike”).<sup>124</sup> In addition to the clauses already seen for the CC-BY, the Share-Alike adds to the license the so-called viral effect: every modification, remix or transformation of the original work should be licensed under the BY-SA conditions or under any compatible license.

Another set of licenses—specifically crafted for the management of the bundle of rights on databases—has been created by the Open Data Commons (ODC) project.<sup>125</sup> The standard agreements developed by it are: (1) the ODC Public Domain Dedication and Licence (PDDL)<sup>126</sup>; (2) the Open Data Commons Attribution License (ODC-By)<sup>127</sup>; (3) the Open Data Commons Open Database License (ODbL).<sup>128</sup> Their function and content mirrors that of the CC described above, with two main differences: ODC licenses do not cover every genre of intellectual work but only databases, and they are not expressed in the “machine-readable” form.<sup>129</sup>

Even though the ODC licenses are database-specific and should be considered as the more customized legal tool for data, some authors have found the Achilles heels of such agreements exactly in their sectoriality. Considering that they cover just databases and not the content itself, a research repository should necessarily use different types of licenses (one for the scientific publication and another for the dataset supporting that publication), thus creating inconsistencies within the system.<sup>130</sup>

<sup>124</sup><http://creativecommons.org/licenses/by-sa/4.0/>.

<sup>125</sup>The Open Data Commons was one of the first projects in drafting a specific open license for database in 2008 (<http://opendatacommons.org/>). ODC is now part of the Open Knowledge Foundation, a not-for-profit organization whose associative goal is the promotion of the openness and the sharing of knowledge in its every form. See Pollock and Walsh (2012).

<sup>126</sup>The ODC-PDDL is an irrevocable dedication to the public domain through which the rightholder waives all rights and claims in copyright or *sui generis* database rights over a certain database built in every possible media and formats now known or created in the future. In case the waiver is not valid in a particular jurisdiction, the PDDL includes a worldwide, royalty-free, non-exclusive licence to use the work for any purpose for the duration of any applicable copyright and database rights. See more at: <http://opendatacommons.org/licenses/pddl/1.0/>.

<sup>127</sup>The ODC-By allows users to freely share, modify, and use the database subject only to the attribution requirements in the manner specified in the license. According to the license, the rights of the user consist in the: (1) extraction and re-utilisation of the whole or a substantial part of the Contents; (2) creation of derivative databases; (3) creation of collective databases; (4) creation of temporary or permanent reproductions by any means and in any form, in whole or in part, including any derivative databases or as a part of collective databases; (5) distribution, communication, display, lending, making available, or performance to the public by any means and in any form, in whole or in part, including any derivative database or as a part of collective databases. Even if tailored on database rights, such a license resembles the contents and the aim of the CC-BY. See: <http://opendatacommons.org/licenses/by/1.0/>.

<sup>128</sup>The ODC-ODbL is a license agreement intended to allow users to freely share, modify, and use a database while maintaining this same freedom for others. This is realized through the following clause: “4. Any Derivative Database that You Publicly Use must be only under the terms of: i. This License; ii. A later version of this License similar in spirit to this License; or iii. A compatible license”. See: <http://opendatacommons.org/licenses/odbl/1.0/>.

<sup>129</sup>Aliprandi (2011, pp. 35–36), Guibault and Margoni (2013, p. 155) and Leucci (2014, p. 12).

<sup>130</sup>Guibault and Margoni (2013, p. 158).

“Open” licenses are a paradigmatic example of the interaction among different variables: they fit into the copyright and *sui generis* database right domain, but they allow a customization of the right-holder preferences. Thus, such legal tools help in managing the shortcomings of a strong and totalizing IP protection.

Furthermore, they internalize, in a simple and standardized way, some norms of the scientific community: in particular, the option “attribution” reflects a form of reputational reward. This is particularly important, considering that one problem with current credit attribution mechanisms is that they are essentially based on authorship of journal articles.<sup>131</sup>

Thanks to their user-friendly features, open licenses have been successfully adopted for several data access and sharing policies. For example, the Personal Genome Project (PGP),<sup>132</sup> created in 2005 at Harvard Medical School, has set up a scientific database collecting genomic, environmental and human trait data and has licensed the repository under CCO conditions.<sup>133</sup>

---

## 6 The Role of the Researcher: Dr. Jekyll or Mr. Hyde?<sup>134</sup>

The use of the Internet and Web 2.0 has also affected scientific culture, enhancing the possibilities of information disclosure and networking. Nowadays, a researcher has a number of tools—such as blogs, thematic social networks, wikis, etc.—which enable a real-time sharing of his/her thoughts, datasets, analysis, small and negative findings with potentially everybody and without waiting for a traditional publication in a scientific journal.<sup>135</sup> This can produce several advantages: data can circulate more broadly and faster than in the paper-based context, partial results can be cross-checked and validated by several experts, communication enhances the possibility of receiving feedback from a larger community, the disclosure of the so-called “blind-alleys” (negative findings), which of course are never published because unproductive of results, can guide other scientists in their investigations or, at least, avoid the duplication of research in the same deadlock field.<sup>136</sup>

However, a favourable attitude towards sharing is not widespread among researchers, especially in the GS.<sup>137</sup> We have probably to dismiss the Mertonian idea of an investigator moved by high values and/or the public benefit, and the concept of the scientist as a rational individual acting in the interest of the scientific

---

<sup>131</sup>Ankeny and Leonelli (2015).

<sup>132</sup><http://www.personalgenomes.org/>. Accessed 18.10.2014.

<sup>133</sup><http://www.personalgenomes.org/organization/sharing>. Accessed 18.10.2014.

<sup>134</sup>Many of the considerations developed in this paragraph were already expressed in Caso and Ducato (2014).

<sup>135</sup>Bartling and Friesike (2014, p. 8) and Rinaldi (2014).

<sup>136</sup>Boggio (2008, p. 10) and Bartling and Friesike (2014, p. 9).

<sup>137</sup>Mboera (2012).

body.<sup>138</sup> In a more cynical way, we have to admit that building a dataset requires huge intellectual efforts, and in the end those data constitute the scientist's "little treasure", which will be used for publishing any significant result. Sharing such information will mean losing a significant competitive advantage and run the concrete risk of favouring the priority of someone else's publication or invention.<sup>139</sup>

If at all, sharing has been practised in the scientific community as a means for ensuring a relationship among two researchers or labs.<sup>140</sup> Put in other words, it has been conceived as a "gift relationship"<sup>141</sup>: a courtesy occasionally made inside a small community of peers, presumably hoping to be reciprocated in the hour of need.

The lack of openness has been in some way challenged by the data sharing policies adopted by several public funding bodies in Europe and the US.<sup>142</sup> Many grant agreements obligate researchers to "grant back" their results and to make their dataset available for re-use. Such conditions are generally fulfilled by uploading research data into a public repository. These policies are an important recognition of the value of data collecting; nevertheless, they only have a limited scope (see for example the opt-out mechanisms in the Horizon 2020 Open Data Pilot) and face a gigantic problem of enforcement. The lack of strict controls and effective sanctions lead to a dilution of the innovative significance of such an institutional effort.<sup>143</sup>

We argue that one of the possible solutions for encouraging data sharing lies in creating special incentives for researchers, which internalize reputational rewards.<sup>144</sup> As Ankeny and Leonelli have outlined, current credit attribution mechanisms are shaped around the traditional outcome of a research: the publication.<sup>145</sup> Traditional metrics fail to measure the value of efforts spent in data collection and sharing, leaving this type of work out of their evaluation grids.<sup>146</sup>

In the end, why should a researcher be forced to share his/her dataset with someone else? Why should he/she compromise his/her career? The labour behind such tasks is far from an automated one: it requires time and professional skills, but

---

<sup>138</sup>As in the Polanyi's view (1962).

<sup>139</sup>Borgman (2007, 2010, 2012), Pisani and AbouZahr (2010), Gitter (2013), Pampel and Dallmeier-Tiessen (2014) and Ankeny and Leonelli (2015).

<sup>140</sup>Ankeny and Leonelli (2015). See also, Reichman and Uhler (2003, p. 453 and ff).

<sup>141</sup>Ibidem.

<sup>142</sup>Leonelli (2013b). The leading case is represented by the Bermuda Principles, developed in 1996 for fostering the sharing of DNA sequences along the Human Genome Project (Collins et al. 2003). Among the latest examples of data sharing policies, see the "Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020" ([http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf)) or the NIH "Genomic Data Sharing Policy" (<http://gds.nih.gov/03policy2.html>). Accessed 18.10.2014.

<sup>143</sup>Ankeny and Leonelli (2015). The lack of enforcement and control mechanisms is critically linked by Frischmann, Madison and Strandburg for the sustainment of a knowledge commons (Frischmann et al. 2014, p. 35).

<sup>144</sup>Caso and Ducato (2014).

<sup>145</sup>Ankeny and Leonelli (2015).

<sup>146</sup>Carlson (2011, p. 293).

nowadays it does not receive any recognition. If we want to make “openness” effective and fair for its players, the challenge is to think up novel mechanisms that bring to the surface the “undeclared work” of: (1) collecting reliable data; (2) sharing them.

In the context of biobanks, for example, Anne Cambon-Thomsen has proposed the creation of a BRIF (*Bioresource Research Impact Factor*), a special citation impact factor for biorepository.<sup>147</sup> Such metrics should “trace the quantitative use of a bioresource, the kind of research using it and the efforts of the people and institutions that construct it and make it available”,<sup>148</sup> giving credit to those who created and maintained a valid resource.

In life sciences research, which depends also on the possibility of access to biological samples (which are a scarce resource), we have proposed to think of a “sharing-index”,<sup>149</sup> measuring the contribution of a scientist in making his/her dataset available worldwide and rewarding him/her with priority access to the material resources of a biorepository or total/partial waiver of the cost recovering fees.

The recognition of the contribution in creating a dataset would also be functional for accountability purposes, in view of assessing potential responsibilities.<sup>150</sup> Information is valuable only if truthful,<sup>151</sup> so evaluating the accuracy and integrity of data would push competition towards the top and generate as a positive externality the improvement of the general quality of informational resources.

Also, contractual tools can play a role in this context: for example, the use of a license, which recognizes the attribution, would serve to build a reputational reward for the researcher who has decided to share his/her data collection. In this sense, we argue that in drafting data sharing policies it would be preferable to adopt a CC-BY-4.0 or an ODC-By instead of a CC0 or a PDDL. The latter, in fact, would not allow the original contributor to gain credits.

---

## 7 Conclusions

With the advancements in science and technology over the last few years, bioinformation has acquired an unprecedented importance and value, not only for the individual to whom it relates for the possible consequences in terms of personalized medicine, but also for the different stakeholders who are interested in something that has become an exploitable resource: the current methods of research in life sciences are characterized, in fact, by a massive and cross-oriented analysis of bioinformation, which is collected, indexed, verified, made available or sold, like a new commodity. The boundaries of IPRs have gradually extended. Scientists tend

---

<sup>147</sup>Cambon-Thomsen et al. (2011). It represents the evolution of the BIF, Biobank impact factor proposed by Cambon-Thomsen (2003). See also, De Castro et al. (2013).

<sup>148</sup>Cambon-Thomsen et al. (2011, p. 503).

<sup>149</sup>Caso and Ducato (2014).

<sup>150</sup>Ankeny and Leonelli (2015).

<sup>151</sup>Only truthful information generates new knowledge, according to Floridi (2010).

to protect with secrecy and IPRs those resources that until a few years ago were informally exchanged. The privatization of bioinformation is critical because the enclosure movement tends now to encompass the “raw material” of every investigation. Considering the cumulative nature of knowledge, such a commodification can create a dangerous impasse for the scientific progress.

In the current information economy the possibility of accessing and using such data is crucial for innovation and development, but it is even more important for developing countries. For the latter, openness means the possibility of access to a resource that they are not able to create due to the lack of funds, the chance of not being cut off from the international research net, and the hope of decreasing the knowledge gap with the North of the world.

However, the open philosophy is universally preached, but little practiced. As outlined in the paper, there are legal, technological and social obstacles that can explain such a situation: (1) the lack of public investments; (2) the absence of ICT infrastructures or their inability to share and re-use information, hindering the database interoperability or the data portability; (3) a pervasive private control of data through strong IPRs, contracts and technological protection measures; (4) the potential conflict between data protection and open access to bioinformation; (5) the lack of adequate economic or reputational incentives to share information within the scientific community and society in general. The construction of an effective OB policy must consider the interaction of all these factors in order to create a virtuous circle of sharing and a new knowledge commons.<sup>152</sup> We have presented examples of how two obstacles (pervasive private control of data and the lack of adequate economic or reputational incentives to share information) can be mitigated by using combined solutions from different domains.

This is just the first contribution of a more complex study and we are aware of the fact that the most difficult Gordian knot to untie is the interaction with privacy rules. We believe that future efforts of legal scholars must aimed at addressing a fair balance among the interests of data subjects, researchers and the society.

We are not so naive as to argue that the differences between the North and the South of the world would be solved with the adoption of a policy for open access to research data: nevertheless, we believe that this could be a first step in narrowing the current gap and making the informational resources more equally accessible.

---

## References

- Aliprandi S (2011) Open licensing e banche dati. *Informatica e diritto* (1–2): 25-43
- Altunay M et al (2010) A science driven production cyberinfrastructure—the open science grid. *J Grid Comput* 9(2):201–218
- Ankeny RA, Leonelli S (2015) Valuing data in postgenomic biology: how data donation and curation practices challenge the scientific publication system. In: Stevens H, Richardson S (eds) *PostGenomics*. Duke University Press, Durham

---

<sup>152</sup>As defined by Hess and Ostrom (2006) and Suber (2006).

- Bartling S, Friesike S (2014) Toward another scientific revolution. In: Bartling S, Friesike S (eds) *Opening science. The evolving guide on how the Internet is changing research, collaboration and scholarly publishing*. Springer, Heidelberg, pp 3–15
- Bastow R, Leonelli S (2010) Sustainable digital infrastructure. *EMBO Rep* 11(10):730–734
- Boggio A (2008) Transfer of samples and sharing of results: requirements imposed on researchers. <http://ssrn.com/abstract=1020734>. Accessed 18 Oct 2014
- Borgman CL (2007) *Scholarship in the digital age*. MIT Press, Cambridge
- Borgman CL (2010) Research data: who will share what, with whom, when, and why? <http://works.bepress.com/borgman/238/>. Accessed 18 Oct 2014
- Borgman CL (2012) The conundrum of sharing research data. *J Am Soc Inform Sci Technol* 63(6):1059–1078
- Boyle J (1997) A politics of intellectual property: environmentalism for the net? *Duke Law J* 47:87–116
- Boyle J (2003) The second enclosure movement and the construction of the public domain. *Law Contemp Prob* 66:33–74
- Caballero Leal JL (2000) Protección jurídica de las bases de datos y protección sui generis para las bases de datos no originales. In: Antequera Hernández R, Palacios López MA (eds) *Propiedad intelectual. Temas relevantes en el escenario internacional*, Guatemala, SIECA/USAID, p 327
- Cambon-Thomsen A (2003) Assessing the impact of biobanks. *Nat Genet* 34:25–26
- Cambon-Thomsen A (2004) The social and ethical issues of post-genomic human biobanks. *Nat Rev Genet* 5:866–873
- Cambon-Thomsen A, Thorisson GA, Mabile L (2011) The role of a bioresource research impact factor as an incentive to share human bioresources. *Nat Genet* 43:503–504
- Cardarelli F (2002) Le banche dati pubbliche: una definizione. *Diritto dell'informazione dell'informatica* 2:321–341
- Carlson D (2011) A lesson in sharing. *Nature* 469:293
- Carroll MW (2006) Creative commons and the new intermediaries. *Mich St L Rev* 45
- Caso R (2004) *Digital rights management. Il commercio delle informazioni digitali tra contratto e diritto d'autore*. CEDAM, Padova
- Caso R, Ducato R (2014) Intellectual property, open science and research biobanks. The Trento Law and Technology Research Group Research Paper Series 22. <http://eprints.biblio.unitn.it/4374>. Accessed 18 Oct 2014
- Caulfield T, Harmon SH, Joly Y (2012) Open science versus commercialization: a modern research conflict? *Genome Med* 4(2):17
- Chan L, Kirsop B, Arunachalam S (2005) Open access archiving: the fast track to building research capacity in developing countries. [https://tspace.library.utoronto.ca/bitstream/1807/4415/1/Open\\_Access\\_Archiving.pdf](https://tspace.library.utoronto.ca/bitstream/1807/4415/1/Open_Access_Archiving.pdf). Accessed 18 Oct 2014
- Collins FS, Morgan M, Patrinos A (2003) The Human Genome Project: lessons from large-scale biology. *Science* 300(5617):286–290
- Costello A, Zumla A (2000) Moving to research partnerships in developing countries. *Br Med J* 321(7264):827–829
- Czerniewicz L, Goodier S (2014) Open access in South Africa: a case study and reflections. *S Afr J Sci* 110(9/10):1–9
- David PA (2000) A tragedy of the public knowledge ‘commons’? *Global Science, Intellectual Property and the Digital Technology Boomerang*. SIEPR Discussion Paper 00-002, pp 1–41. <http://web.stanford.edu/group/siepr/cgi-bin/siepr/?q=system/files/shared/pubs/papers/pdf/00-02.pdf>
- David PA (2004) Can “open science” be protected from the evolving regime of IPR protections?. *Journal of Institutional and Theoretical Economics* 160:9–34, preprint available at: <http://philob.at/wiki/images/David-openscience-ipr-ann.pdf>
- David PA, Foray D (2002) An introduction to the economy of the knowledge society. *Int soc sci J* 54:9–23

- Davison M, Hugenholtz B (2005) Football fixtures, horse races and spin-offs: the ECJ domesticates the database right. *Eur Intell Prop Rev* 27(3):113
- De Castro P, Calzolari A, Napolitani F, Maria Rossi A, Mabile L, Cambon-Thomsen A, Bravo E (2013) Open data sharing in the context of bioresources. *Acta Inform Med* 21(4):291–292
- De la Parra Trujillo E (2004) El derecho sui generis sobre las bases de datos en México y la Unión Europea. *Derecho Comparado de la Información* 3:101–124
- De Roure D, Jennings NR, Shadbolt NR (2003) The semantic grid: a future e-science infrastructure. In: Berman F, Fox J, Hey T (eds) *Grid computing: making the global infrastructure a reality*. Wiley, Chichester, pp 437–470
- De Vries J et al (2011) Ethical issues in human genomics research in developing countries. *BMC Med Ethics* 12(1):5
- Derclaye E (2004) Databases sui generis right: should we adopt the spin off theory. *Eur Intell Prop Rev* 26(9):402–413
- Derclaye E (2008) *The legal protection of databases: a comparative analysis*. Edward Elgar Publishing, Cheltenham-Northampton
- Derclaye E (2014) The database directive. In: Stamatoudi I, Torremans P (eds) *EU copyright law: a commentary*. Edward Elgar Publishing, Cheltenham-Northampton, pp 298–354
- Di Bona C, Ockman S (1999) *Open sources: voices from the open source revolution*. O'Reilly Media, Sebastopol
- Dickenson D (2004) Consent, commodification and benefit-sharing in genetic research. *Dev World Bioethics* 4(2):109–124
- Dietrich N, Wiebe A (2013) Definition of research data. In: Guibault L, Wiebe A (eds) *Safe to be open. Study on the protection of research data and recommendations for access and usage, OpenAIRE+*. Universitätsverlag Göttingen, p 17
- Ducato R (2013) 'Adiós Sui Géneris': a study of the legal feasibility of the sui generis right in the context of research biobanks. *Revista de Derecho y Genoma Humano/Law Hum Genome Rev* 38:125–146
- Dulle F, Kaane S, Nyamboga C (2013) Application of information and communication technologies for documentation and dissemination of scholarly output among Inter University Council for East Africa member Institutions. *Mousaion* 31(3):127–144
- Dusollier S (2002) Exceptions and technological measures in the European copyright directive of 2001—an empty promise. *IIC* 34(1):62–75
- Edwards AM, Bountra C, Kerr DJ, Willson TM (2009) Open access chemical and clinical probes to support drug discovery. *Nat Chem Biol* 5:436–440
- European Commission (2012) Recommendation on access to and preservation of scientific information. Bruxelles, 17.7.2012 C (2012) 4890 final. [http://ec.europa.eu/research/science-society/document\\_library/pdf\\_06/recommendation-access-and-preservation-scientific-information\\_en.pdf](http://ec.europa.eu/research/science-society/document_library/pdf_06/recommendation-access-and-preservation-scientific-information_en.pdf). Accessed 18 Oct 2014
- Fecher B, Friesike S (2014) Open science: one term, five schools of thought. In: Bartling S, Friesike S (eds) *Opening science. The evolving guide on how the Internet is changing research, collaboration and scholarly publishing*. Springer, Heidelberg, pp 17–47
- Floca R (2014) Challenges of open data in medical research. In: Bartling S, Friesike S (eds) *Opening science. The evolving guide on how the Internet is changing research, collaboration and scholarly publishing*. Springer, Berlin, pp 297–307
- Floridi L (2010) *Information: a very short introduction*. Oxford University Press, Oxford
- Ferero-Pineda C (2006) The impact of stronger intellectual property rights on science and technology in developing countries. *Res Policy* 35(6):808–824
- Frischmann BM, Madison MJ, Strandburg KJ (2014) Governing knowledge commons. In: Frischmann BM, Madison MJ, Strandburg KJ (eds) *Governing knowledge commons*. Oxford University Press, Oxford, pp 1–39
- Frosio GB (2014) Open access publishing: a literature review. <http://www.create.ac.uk/wp-content/uploads/2014/01/CREATe-Working-Paper-2014-01.pdf>. Accessed 18 Oct 2014

- Fulwood S (1991) *Feist v. rural*: did the Supreme Court give license to reap where one has not sown. *Comm Law* 9:15
- Ginsburg JC (1992) No “Sweat”? Copyright and other protection of works of information after *feist v. rural telephone*. *Columbia Law Rev* 338–388
- Ginsburg JC (2005) Legal protection of technological measures protecting works of authorship: international obligations and the US experience. *Columbia Law J Law Arts* 29:11–37
- Gitter DM (2013) The challenge of achieving open source sharing of biobank data. In: Pascuzzi G, Izzo U, Macilotti M (eds) *Comparative issues in the governance of research biobanks*. Springer, Heidelberg, pp 165–189
- Gleick J (2011) *The information: a history, a theory, a flood*. Fourth Estate, London
- Gómez N, Bongiovani P C (2012) Open access and A2K: collaborative experiences in Latin America. [http://www.degruyter.com/dg/viewbooktoc.chapterlist.resultlinks.fullcontentlink:pdf/eventlink/\\$002fbooks\\$002f9783110263121\\$002f9783110263121.343\\$002f9783110263121.343.pdf?t:ac=product/129022](http://www.degruyter.com/dg/viewbooktoc.chapterlist.resultlinks.fullcontentlink:pdf/eventlink/$002fbooks$002f9783110263121$002f9783110263121.343$002f9783110263121.343.pdf?t:ac=product/129022). Accessed: 18 Oct 2014
- Goss AK (2007) Codifying a commons: copyright, copyleft, and the Creative Commons project. *Chi-Kent Law Rev* 82:963
- Grubb AM, Easterbrook SM (2011) On the lack of consensus over the meaning of openness: an empirical study. *PLoS ONE* 6(8):e23420
- Guarda P (2013) Biobanks and electronic health records: open issues. In: Pascuzzi G, Izzo U, Macilotti M (eds) *Comparative issues in the governance of research biobanks*. Springer, Heidelberg, pp 131–141
- Guibault L (2011) Creative Commons licenses: what to do with the database right? *Comput Law Mag* 6:1–4
- Guibault L (2013) Licensing research data under open access conditions. In: Beldiman D (ed) *Information and knowledge: 21st century challenges in intellectual property and knowledge governance*. Edward Elgar, Cheltenham, pp 63–92
- Guibault L, Margoni T (2013) Analysis of licensing issues. In: Guibault L, Wiebe A (eds) *Safe to be open. Study on the protection of research data and recommendations for access and usage*. Universitätsverlag Göttingen, pp 143–160
- Guibault L, Wiebe A (2013) *Safe to be open. Study on the protection of research data and recommendations for access and usage*. Universitätsverlag Göttingen
- Gymrek M et al (2013) Identifying personal genomes by surname inference. *Science* 339(6117):321–324
- Hagedoorn J, Link AN, Vonortas NS (2000) Research partnerships. *Res Policy* 29:567–586
- Hardy BJ, Séguin B, Goodsaid F, Jimenez-Sanchez G, Singer PA, Daar AS (2008) The next steps for genomic medicine: challenges and opportunities for the developing world. *Nat Rev Genet* 9:S23–S27
- Hayden EC (2014) Technology: the \$1,000 genome. *Nature* 507:294–295
- Henry C, Stiglitz JE (2010) Intellectual property, dissemination of innovation and sustainable development. *Global Policy* 1(3):237–251
- Hess C, Ostrom E (2003) Ideas, artifacts, and facilities: information as a common-pool resource. *Law Contemp Prob* 66:111–145
- Hess C, Ostrom E (2006) *Understanding knowledge as a common: from theory to practice*. MIT Press, Cambridge
- Hoffman S (2014) Citizen science: the law and ethics of public access to medical Big Data. *Berkeley Technol Law J*. <http://ssrn.com/abstract=2491054>. Accessed: 18.10.2014
- International Human Genome Sequencing Consortium. Lander E S et al (2001) Initial sequencing and analysis of the human genome. *Nature* 409:860–921
- Inyang HI (2012) Production and access to scientific data in Africa: a framework for improving the contribution of research institutions. In: Mathae KB, Uhilir PF (eds) *The case for international sharing of scientific data: a focus on developing countries, proceedings of a symposium*. National Academies Press, Washington, D.C., pp 115–117

- Jensen PB, Jensen LJ, Brunak S (2012) Mining electronic health records: towards better research applications and clinical care. *Nat Rev Genet* 13:395–405
- Kahn M (2012) Implementing a research data access policy in South Africa. In: Mathae KB, Uhlir PF (eds) *The case for international sharing of scientific data: a focus on developing countries, proceedings of a symposium*. National Academies Press, Washington, DC, pp 21–23
- Kanguane P (2009) *Bioinformation discovery, data to knowledge in biology*. Springer, Heidelberg
- Karnell G (1997) The Nordic catalogue rule. In: Dommering E, Hugenholtz PB (eds) *Protecting works of fact, copyright freedom of expression and information law*. Kluwer Law and Taxation Publishers, Denver, p 67
- Kaye J (2012) The tension between data sharing and the protection of privacy in genomics research. *Annu Rev Genomics Hum Genet* 13(1):415–431
- Kling R, Callahan E (2003) Electronic journals, the Internet, and scholarly communication. *Ann Rev Info Sci Tech* 37:127–177
- Knoppers BM (2000) Population genetics and benefit sharing. *Commun Genet* 3:212–214
- Knoppers BM (2005) Of genomics and public health: building public “goods”? *Can Med Assoc J* 173(10):1185–1186
- Kohane IS (2011) Using electronic health records to drive discovery in disease genomics. *Nat Rev Genet* 12:417–428
- Krumholz HM, Gross CP, Blount KL, Ritchie JD, Hodshon B, Lehman R, Ross JS (2014) Sea change in open science and data sharing: leadership by industry. *Circ Cardiovasc Qual Outcomes* 7:499–504
- Kuppuswamy C (2009) *The international legal governance of the human genome*. Routledge, New York
- Leonelli S (2013a) Integrating data to acquire new knowledge: three models of integration in plant science. *Stud Hist Philos Biol Biomed Sci* 44(4):503–514
- Leonelli S (2013b) Why the current insistence on open access to scientific data? *Big Data, knowledge production and the political economy of contemporary biology*. *Bulletin of Science, Technol Soc*. <http://bst.sagepub.com/content/early/2013/08/19/0270467613496768>. Accessed 18 Oct 2014
- Lessig L (1999) *Code and other laws of cyberspace*. Basic books, New York
- Leucci S (2014) Preliminary notes on open data licensing. *J Open Access Law* 2(1):1–24
- Lindsay DF (2012) Protection of compilations and databases after IceTV: authorship, originality and the transformation of Australian Copyright Law. *Monash Univ Law Rev* 38(1):17–59
- Lunshof JE et al (2008) From genetic privacy to open consent. *Nat Rev Genet* 9(5):406–411
- Margoni T (2013) The roles of material transfer agreements in genetics databases and bio-banks. In: Pascuzzi G, Izzo U, Macilotti M (eds) *Comparative issues in the governance of research biobanks*. Springer, Heidelberg, pp 231–249
- Mascalzoni D et al. (2014) International charter of principles for sharing bio-specimens and data. *Eur J Hum Genet* 1–8
- Maurer SM (2003) New institutions for doing science: from databases to open source biology. [http://www.epip.eu/papers/20031124/200411\\_conference/papers/maurer\\_paper.pdf](http://www.epip.eu/papers/20031124/200411_conference/papers/maurer_paper.pdf). Accessed 18 Oct 2014
- Mboera LEG (2012) The management of health and biomedical data in Tanzania: the need for a national scientific data policy. In: Mathae KB, Uhlir PF (eds) *The case for international sharing of scientific data: a focus on developing countries, proceedings of a symposium*. National Academies Press, Washington, D.C., pp 27–28
- Merton RK (1942) Science and technology in a democratic order. *Legal Political Sociol* 1:115. Reprinted as Merton RK (1973) *The normative structure of science*. In: Storer NW (ed) *The sociology of science: theoretical and empirical investigations*. Chicago University Press, Chicago, pp 267–278
- Milosavljevic A (2000) The economic value of bioinformation. *Bioinformatics* 17(7):571–572

- Moscon V (2015) Academic freedom, copyright, and access to scholarly works: a comparative perspective. In: Caso R, Giovanella F (eds) *Balancing copyright law in the digital age—comparative perspectives*. Verlag-Berlin, Springer, pp 99–135
- Muñoz Palma P (2012) Access to research data and scientific information generated with public funding in Chile. In: Mathae KB, Uhler PF (eds) *The case for international sharing of scientific data: a focus on developing countries, proceedings of a symposium*. National Academies Press, Washington, D.C., pp 24–26
- Murray-Rust P, Neylon C, Pollock R, Wilbanks J (2010) *Panton principles, principles for open data in science*, <http://pantonprinciples.org/>. Accessed: 18 Oct 2014
- Nelson RR (2004) The market economy, and the scientific commons. *Res Policy* 33:455–471
- Nielsen M (2011) *Reinventing discovery: the new era of networked science*. Princeton, Princeton University Press
- Nuffield Council on Bioethics (2007) The forensic use of bioinformation: ethical issues. <http://nuffieldbioethics.org/wp-content/uploads/The-forensic-use-of-bioinformation-ethical-issues.pdf>. Accessed 18 Oct 2014
- Ohm P (2010) Broken promises of privacy: responding to the surprising failure of anonymization. *UCLA Law Rev* 57:1701–1777
- Open Knowledge Foundation (2014) Open and collaborative science for development. <https://groups.google.com/forum/#!forum/openscidedev>. Accessed 18 Oct 2014
- Ovilla Bueno R (1998) La protección jurídica de las bases de datos en México. De los lineamientos internacionales a la nueva Ley Federal del Derecho de Autor. In: Becerra Ramírez M (ed) *Estudios de derecho intelectual en homenaje al profesor David Rangel Medina*. UNAM, México, p 313
- Pampel H, Dallmeier-Tiessen S (2014) Open research data: from vision to practice. In: Bartling S, Friesike S (eds) *Opening science. The evolving guide on how the Internet is changing research, collaboration and scholarly publishing*. Springer, Heidelberg, pp 213–224
- Parker M et al (2009) Ethical data release in genome-wide association studies in developing countries. *PLoS Med* 6(11):e1000143
- Pary B (2004) *Trading the genome: investigating the commodification of bio-information*. Columbia University Press, New York
- Paton R (1996) Metaphors, models and bioinformation. *Biosystem* 38:155–162
- Pisani E, AbouZahr C (2010) Sharing health data: good intentions are not enough. *Bull World Health Organ* 88(6):462–466
- Pistorius T (2008) The IP protection of electronic databases: copyright or copywrong? In: Venter HS, Eloff MM, Eloff JHP, Labuschagne L (eds) *Proceedings of the ISSA 2008 innovative minds conference*. ISSA, Pretoria, pp 63–80
- Polanyi M (1962) The republic of science: its political and economic theory. *Minerva* 1:54–74
- Pollock R, Walsh J (2012) Open knowledge: promises and challenges. In: Dulong de Rosnay M, De Martin JC (eds) *The digital public domain: foundations for an open culture*. Open Book Publishers, Cambridge, pp 125–132
- Rahman A (2012) Why is international scientific data sharing important? In: Mathae KB, Uhler PF (eds) *The case for international sharing of scientific data: a focus on developing countries, proceedings of a symposium*. National Academies Press, Washington, D.C., pp 7–14
- Ramsay M, de Vries J, Soodyall H, Norris SA, Sankoh O (2014) Ethical issues in genomic research on the African continent: experiences and challenges to ethics review committees. *Hum Genomics* 8:15
- Raymond ES (2000) The cathedral and the bazaar. <http://www.catb.org/~esr/writings/homesteading/cathedral-bazaar>. Accessed 18 Oct 2014
- Reichman JH, Samuelson P (1997) Intellectual property rights in data? *Vanderbilt Law Rev* 50:52–166
- Reichman JH, Uhler PF (1999) Database protection at the crossroads: recent development and their impact on science and technology. *Berkeley Technol Law J* 14:793–838

- Reichman JH, Uhler PF (2003) A contractually reconstructed research commons for scientific data in a highly protectionist intellectual property environment. *Law Contemp Prob* 66:315–462
- Rinaldi A (2014) Spinning the web of open science: social networks for scientists and data sharing, together with open access, promise to change the way research is conducted and communicated. *Eur Mol Biol Organ Rep* 15(4):342–346
- Rodriguez H, Snyder M, Uhlén M, Andrews P, Beavis R, Borchers C et al (2009) Recommendations from the 2008 International summit on proteomics data release and sharing policy: the Amsterdam principles. *J Proteome Res* 8(7):3689–3692
- Rose H (2001) The Commodification of bioinformation: The Icelandic Health Sector Database. [http://www.wellcome.ac.uk/stellent/groups/corporatesite/@msh\\_grants/documents/web\\_document/wtd003281.pdf](http://www.wellcome.ac.uk/stellent/groups/corporatesite/@msh_grants/documents/web_document/wtd003281.pdf). Accessed 18 Oct 2014
- Rosenthal A, Mork P, Hao Li M, Stanford J, Koester D, Reynolds P (2010) Cloud Computing: a new business paradigm for biomedical information sharing. *J Biomed Inform* 43(2):342–353
- Schwartz PM (1994) European Data Protection Law and restrictions on international data flows. *Iowa Law Rev* 80:471–496
- Scott CT, Caulfield T, Borgelt E, Illes J (2012) Personal medicine—the new banking crisis. *Nat Biotechnol* 30:141–147
- Sgaier SK, Jha P, Mony P, Kurpad A, Lakshmi V, Kumar R, Ganguly NK (2007) Biobanks in developing countries: needs and feasibility. *Science* 318(5853):1074–1075
- Sirugo G et al (2004) A national DNA bank in the Gambia, West Africa, and genomic research in developing countries. *Nat Genet* 36(8):785–786
- Solove D J (2004) *The digital person: technology and privacy in the information age*. New York University Press, New York
- Stallman R (1998) Copyleft: pragmatic idealism, <http://www.cyberkuhinja.com/kitchenmedialab/download/activism/Copyleftpragideia.doc>. Accessed 18 Oct 2014
- Stallman R (2002) Free software, free society. <http://www.gnu.org/doc/fsfs-ii-2.pdf>. Accessed 18 Oct 2014
- Stamatoudi I (1997) The EU Database directive: reconceptualising copyright and retracing the future of the sui generis right. *Hellenic Rev Int Law* 50:441
- Stein LD (2010) The case for cloud computing in genome informatics. *Genome Biol* 11(5):207
- Stiglitz JE (2008) Economic foundations of intellectual property rights. *Duke Law J* 57:1693–1724
- Streitz WB, Bennett AB (2003) Material transfer agreements: a university perspective. *Plant Physiol* 33(1):10–13
- Strong W S (1994) Database protection after *Feist v. Rural Telephone Co.* *J. Copyright Soc’y USA* 42:39
- Suber P (2006) Creating an intellectual commons through Open Access. In: Hess C, Ostrom E (eds) *Understanding knowledge as a common: from theory to practice*. MIT Press, Cambridge, pp 171–208
- Suber P (2012) Open access overview. <http://legacy.earlham.edu/~peters/fos/overview.htm>. Accessed 18 Oct 2014
- Thorsteinsdóttir H, Quach U, Daar AS, Singer PA (2004) Conclusions: promoting biotechnology innovation in developing countries. *Nat Biotechnol* 22:DC48–DC52
- Tindana PO et al (2007) Grand challenges in global health: community engagement in research in developing countries. *PLoS Med* 4(9):e273
- Topol E (2013) *The creative destruction of medicine: how the digital revolution will create better health care*. Basic Books, New York
- Toronto International Data Release Workshop Authors (2009) Prepublication data sharing. *Nature* 461:168–170
- Trosow SE (2004) Sui generis database legislation: a critical analysis. *Yale J Law Technol* 7: 534–642
- Veldsman S, Gevers W (2014) Increased visibility and discoverability of South African health-related research. *S Afr Med J* 104(4):287

- Weigelt J (2009) The case for open-access chemical biology. *Eur Mol Biol Organ Rep* 10(9): 941–945
- West M (2006) Embracing the complexity of genomic data for personalized medicine. *Genome Res* 16:559–566
- Winickoff DE (2009) From benefit sharing to power sharing: partnership governance in population genomics research. In: Kaye J, Stranger M (eds) *Principles and practice in Biobank Governance*. Ashgate, Surrey, pp 53–66
- Winickoff DE, Winickoff R (2003) The charitable trust as a model for genomic biobanks. *N Engl J Med* 349(12):1180–1184
- Zain RB et al (2013) An oral cancer biobank initiative: a platform for multidisciplinary research in a developing country. *Cell Tissue Banking* 14(1):45–52