

## ***Blood, Sweat and Grants*** **‘Honest Jim’ and the European database-right**

JASPER A. BOVENBERG

### **Abstract**

Access to detailed, up-to-date and available bioinformatics databases has been identified by the Commission of the European Union as a pillar for the harvesting of the potential of life-sciences and biotechnology. Unconditional access to research data, however, is squarely at odds with the primary interest of every scientist to be the first to make a discovery. This classical dilemma is specifically pressing in the data-driven field of biomedical research, where data-quantity has become a quality on its own, where speed matters and patients can't wait. The dilemma urges a consideration of the principle, the practice and the law regarding access for academic researchers to unpublished research data. The consideration will include the presentation of the outcome of a global and a national survey among biomedical researchers on the accessibility of 'their' research data. The principled arguments pro unconditional access and the laws and practical considerations contra unconditional access offer conflicting perspectives and the resulting situation is compounded by the uncertainty created by the European database-right as to who holds legal title to the databases. Therefore, it is explored whether the two opposing concerns – unconditional access vs. legitimate restrictions – can be accommodated by the adoption and implementation of a general policy for access to biomedical data that greases the wheels of (European) research.

### **Prologue**

In 1953 Watson and Crick won the race to discover the molecular structure of DNA. They described the double helix in an article in *Nature*. The second last sentence of the article reads:

*We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M.H.F. Wilkins and Dr. R. E. Franklin and their co-workers at King's College London.*<sup>1</sup>

How appropriate this gesture was, appears from a paragraph in *The Double Helix*, the personal account of the race to discover the DNA-structure, published fifteen years later by 'Honest Jim' Watson:

*Rosy [Franklin], of course, did not directly give us her data. For that matter no one at King's realized they were in our hands.*<sup>2</sup>

Watson and Crick had come upon Franklin's data via Max Perutz, the leader of their unit, to whom the data had been reported as a member on a committee appointed by the Medical Research Council to look into the research activities of Wilkins' lab. Confronted with this alleged breach of faith on his part, Perutz later replied that the incident inaccurately pictured Wilkins and Franklin as jealously trying to keep their

data secret and Watson and Crick as getting hold of crucial data in an underhand way. Perutz pointed out that, first, the remit of the committee on which he had served was not to look into the research activities of Wilkins but to establish contact between the groups of people working for the MRC; second, that the MRC report was not confidential and, third, that the one important piece of information it contained might have been known by Crick a year earlier if Watson had taken notes at the seminar where Franklin had presented the data.<sup>3</sup>

## Introduction

Watson's revelations illustrate the dilemma of every scientist: the race to be the first *versus* the need to stand on the shoulders of colleagues and competitors. Fifty years after the description of the double helix, this dilemma is more urgent than ever before, especially in the area of biomedical research. The principle is simple: both verification of scientific findings and the public nature of its funding demand unconditional access to research data. In real life, things ain't that simple. After Internet and the ICT-revolution, GRIDS are increasingly providing a new research tool. GRIDS make possible a whole new type of experiments and enable researchers to work jointly and simultaneously on the same datasets. As a result, biomedical research is more 'data-driven' than ever before and 'unique datasets' have become a researcher's 'life-line'. Which data can he claim as his own? Is he under an obligation to give access to his unpublished data? Why? Which data? To anybody? For what purpose? With or without conditions attached?

Over these complex questions hangs, as from March 11, 1996, as a *deus ex machina*, the sword of the European database-right.<sup>4</sup> This novel, *sui generis* right aims to protect the blood, sweat and tears that go into producing a database, against unauthorised use by third parties. To that end, it provides for an exclusive right to use and re-utilize databases that demonstrate a substantial investment. *Prima facie* this right is at odds with the scientific imperative of the *free flow of information*. However, depending on who actually owns the right, it is a double-edged sword that could cut either way. In the hands of a public funding body, it is an instrument to secure academic scientists access to the research data produced by their colleagues. In the hands of a scientist it may be used as a weapon to legitimise and enforce data-exclusivity.

A recent and urgent case study to analyse the above issues is presented by the Netherlands National Genomics Strategy.<sup>5</sup> This strategy is built on the premise that the Netherlands have a number of unique biomedical databases and population cohorts, which may yield a competitive edge globally. Part of this strategy is to open up these databases for further, large-scale research into the 'nature and nurture' of common complex diseases. For this strategy to succeed, however, the scientists that build these collections must be willing to give access to their data, and, if not, such access should be legally enforceable. As we will see, this discussion is inextricably interwoven with and has implications for similar issues at the European level, where similar initiatives exist such as COGENE and the GenomEUtwin project. In fact, the European Commission has identified 'access to detailed, up-to-date and available bioinformatics databases and open access to knowledge' as two of the three pillars for the harvesting of the potential of life-sciences and biotechnology.<sup>6</sup>

This Article discusses the ‘essential tension’ between the need for academic researchers to have unconditional access to unpublished research data and the exclusive database-right. The sharing ethos in academia is paradigm, as is the assertion that the European database-right constitutes a violation of that ethos. However, for these issues too, ‘context is King’. To do justice to this context, I took the following approach. Part I sets forth the *Principle* underlying access to data: *Communism*. Revisiting the Four Commandments for scientific behaviour postulated by Robert K. Merton, this Article will argue that there is a need to formulate a Fifth Commandment: unconditional access to unpublished research data. Part II will examine the current *Practice* governing access to research data in biomedical academia: *Capitalism*. It will first present and discuss the outcome of two surveys into the actual willingness of biomedical researchers to grant access to ‘their’ research data. One survey was held among human geneticists worldwide; the second survey was held among all principal investigators in the area of specific biomedical research in the Netherlands. Second, Part II will review a variety of reasons why, in practice, access may be denied or subjected to specific conditions. We will see that the principled arguments *pro* unconditional access (Part I) and the laws and practical considerations *contra* (unconditional) access (Part II) offer divergent if not downright conflicting perspectives. Therefore, Part III will analyse whether this Gordian knot can be cut by operation of the *Law*, *i.e.* by analysing who *owns* scientific research data. As the European database-right seems the most obvious avenue for this legal analysis, we will apply the European database-right to a concrete biomedical database, which has both a Dutch and a European dimension: the Netherlands Twin Registry that forms part of the GenomEU-Twin project. Finally, Part IV will discuss *Policy*: how can the *principle*, the *practice* and the *law* pertaining to access to unpublished research data be converted into an effective data-access policy designed to grease the wheels of publicly funded biomedical research.

## I. In principle: Communism

### I.1 Merton’s Four Commandments

*The goal of science.* The scientific *mores* have been formulated by science sociologist Robert K. Merton.<sup>7</sup> Merton first determines the institutional goal of science as the ‘extension of certified knowledge.’<sup>8</sup> The technical method employed toward this end is empirical research. The scientific *mores* both derive from and serve the goal and method of science: universalism, communism, disinterestedness and organised skepticism.<sup>9</sup>

*Universalism.* This imperative aims to guarantee that ‘truth-claims are to be subjected to *pre-established impersonal criteria*’. What works in a Tokyo laboratory must also work in a laboratory in Leiden. Thus, universalism enables scientists from diverse backgrounds and personal beliefs to contribute to the universal tree of knowledge.

*Communism.* ‘*Communism*’ is meant by Merton as the communal ownership of ‘intellectual goods’. ‘Secrecy is the anti-thesis of this norm: full and open communication its enactment’. The communal nature of science is incompatible with exclusive (intellectual) property rights.<sup>10</sup> Scientific findings constitute a ‘*common heritage*’ in which ‘the equity of the individual producer is limited’. It follows from

the institutional goal and method of science that the scientist's sole property right in 'his' findings is that of recognition and esteem. This emphasis on recognition forces a scientist to be the first to make a scientific discovery. According to Merton this results in a system of 'competitive co-operation'.<sup>11</sup> The products of this competition are communized, with the original contributor being rewarded only with recognition and eternal fame.<sup>12</sup> Merton emphasises that this competitive co-operation cannot prejudice the status of scientific knowledge as common good. The communist nature of science is also illuminated and reinforced by the *mos* to acknowledge in publications that the individual contribution of the author to the tree of knowledge was made possible by the contributions of many others.<sup>13</sup>

*Disinterestedness and organised skepticism.* The imperative of disinterestedness means that the pursuit of scientific truth should always prevail over the pursuit of the scientists' personal interests. In particular, scientists should not serve political interests.<sup>14</sup> The norm of 'organised skepticism', means that the scientific community should subject any findings of its members to empirical and logical testing before accepting them as true.<sup>15</sup>

*The reward structure of science.* In addition to formulating the Four Commandments of science, Merton also elucidated how these norms are reinforced by the reward system of science.<sup>16</sup> Scientific communism is rewarded by recognition.<sup>17</sup> Recognition takes many forms: eponymy, - Pythagoras' axiom, Kepler's laws, Huntington's disease, Factor V Leiden; prizes and premiums, – the Nobelprize, the Spinoza prize, and membership of academies – the Royal Society, the Royal Netherlands Academy of Sciences.<sup>18</sup> Recognition of course requires publication and here the goal, method and reward system interrelate. Publication serves the norms of universalism and disinterestedness and enables independent replication and thus verification of truth claims. Publication of original findings also serves the institutional goal of science: the extension of certified knowledge.

## **1.2 The Fifth Commandment: unconditional access to research data**

Merton's communism demands and implies that researchers have access to each other's publications. The question, however, is whether access to published findings is sufficient. 'Honest Jim's account of the race to discover the structure of DNA illustrates the importance of access to the *'unpublished experimental results and ideas'*. Following is an inventory of arguments to formulate a Fifth Commandment: unconditional access to unpublished research data.

*1.2.1 Unconditional access is necessary for scientific revolutions.* The primary goal of unconditional access is to enable verification of scientific findings by others. Thus, this imperative helps create a barrier against 'fraud, theft and self-deception'.<sup>19</sup> However, verification also serves a more remote goal. While Merton and later on Popper assume that scientists are continuously engaged in improving respectively falsifying each other's findings, Kuhn has pointed at the confirmatory behaviour of scientists.<sup>20</sup> Scientists are inclined to mainly confirm and refine established theories or 'paradigms'.<sup>21</sup> Kuhn has named that process 'normal science'. He compares 'normal science' to solving puzzles, in that both are not aimed at producing major novelties, neither conceptual nor phenomenal.<sup>22</sup> The solution of a puzzle is anticipated and not intrinsically interesting or important; the challenge is to find a new

way to achieve that solution.<sup>23</sup> Inadvertently, Watson aptly illustrates this point in his description of Sir Lawrence Bragg, then director of the Cavendish laboratory where Watson and Crick did their work on DNA:

*For almost forty years Bragg, a Nobel Prize winner and one of the founders of crystallography, had been watching X-ray diffraction methods solve structures of ever-increasing difficulty. The more complex the molecule, the happier Bragg became when a new method allowed its elucidation.*<sup>24</sup>

In contrast, Kuhn points out that the ‘really pressing problems, e.g. a cure for cancer (..) are often not puzzles at all, largely because they have no solution.’<sup>25</sup> Kuhn emphasises, however, that paradigms are necessary for the conduct of ‘normal science’. Paradigms indicate which data to select, which experiments should be done and which refinements are necessary to perfect a theory.<sup>26</sup> However, if scientists spend more time to adjust facts to theory rather than the other way around,<sup>27</sup> a crisis will emerge in case too many facts no longer fit the paradigm. Such a period of crisis may lead to the formulation of new theories, the so-called paradigm shift or ‘scientific revolution’.<sup>28</sup> In a revolution those who challenge the scientific establishment will scrutinize existing paradigms by:

*handling the same bundle of data as before, but placing them in a new system of relations with one another by giving them a different framework.*<sup>29</sup>

However, the revolution will not take place or be delayed, unless there is unconditional access to the data. It is precisely in a period of crisis that such access is crucial. If Kuhn is right, his theory provides a powerful argument for unconditional access to research data. If scientists are in a position to deny access to their research data, they will be inclined to grant access exclusively to scientists who adhere to the same paradigms.<sup>30</sup> This inclination may be reinforced by the existence of symbiotic, informal networks of like-minded scientists.<sup>31</sup>

*1.2.2 Unconditional access is necessary for the ‘Invisible Hand of Research’.* Related to the paradigm issue, is the importance of individualism and independence in science.<sup>32</sup> Traditional academic freedom entails that scientists are free to choose their research objectives and method. This freedom will be curtailed if science is centrally co-ordinated, whether by non-scientists or by scientific peers. Co-ordination seems a logical and necessary instrument to avoid duplication of efforts and a waste of resources. However, co-ordination can be fatal to individual initiatives. In *The Republic of Science*, Michael Polyani has argued that independent initiatives, undertaken by competing scientists, are the most efficient way to organise scientific research.<sup>33</sup> As long as a scientist keeps an eye on the work of the others, he will take their efforts into account when formulating his own research questions.<sup>34</sup> Polyani has labelled this system a system of self ‘co-ordination by way of mutual adjustments of independent initiatives.’<sup>35</sup> This co-ordination should not be centralised, but rather be guided by an ‘invisible hand’. Just as the ‘invisible hand’ in a free market-economy helps producers and consumers make supply meet demand and thus achieve maximum welfare, the ‘invisible hand of research’ will guide scientists to achieve maximum progress of science.<sup>36</sup>

*The Jigsaw puzzle.* Polanyi illustrates his point using the metaphor of the ‘Jigsaw puzzle’, a fitting metaphor in view of Kuhn’s description of ‘normal science’ as puzzle-solving. Solving a Jigsaw-puzzle requires, just as solving complex scientific problems, the contributions of multiple persons.<sup>37</sup> Each person will, watched by the others, focus on a particular section of the puzzle and make use of the insights which develop successively as other sections are being solved. The effectiveness of the group will exceed that of its isolated members ‘to the extent to which some member of the group will always discover a new chance for adding a piece to the puzzle more quickly than any one isolated person could have done by himself.’<sup>38</sup> Now imagine, Polanyi argues, that the activities of the group are co-ordinated by a single authority. That would reduce the joint effectiveness of the group as a whole to the effectiveness and insights of this co-ordinator. In the alternative model, the ‘Invisible Hand’ will prevent this from happening.

Polanyi’s model however, requires that each scientist knows what the other puzzle-solvers are doing. For the ‘Invisible Hand’ to succeed, every helper also needs access to all pieces of the puzzle. According to Polanyi, mutual adjustment of independent initiatives will take place by scientists’ taking note of the published results of other scientists. Publications thus play the role in science, which the prices of goods and services play in the economy, as a means to ‘make supply meet demand’.<sup>39</sup> However, Polanyi fails to complete his analogy with Adam Smith’s ‘Invisible Hand’. He fails to mention that the ‘Invisible Hand’ cannot operate, if the market forces do not have complete information or if one party has a monopoly.<sup>40</sup> A similar market-failure is likely to paralyse the operation of the ‘Invisible Hand’ of research, if a scientist has no knowledge of, or no access to, unpublished research data. In other words, if Polanyi is right, his model of the ‘Invisible Hand’ is also a powerful argument for unconditional access to research data.

*1.2.3 Unconditional access is necessary to enable new research.* This argument has a financial and a scientific component. First of all, unconditional access will save costs. Absent exclusive rights on data, other researchers need not negotiate terms of access and save both transaction costs and real costs, such as royalties and other forms of compensation. Thus, the original data-producers subsidize ‘second comers’.<sup>41</sup> Though this may look like a ‘free ride’ for the second comer on the blood, sweat and tears of the original researcher, the system has an element of ‘distributive justice’. Most likely, the original researcher has had the benefits of the efforts put in by his predecessors. And the system works two ways, of course. The potential ‘free ride’ may be offset by reciprocal access or access to data elsewhere. The scientific component of this argument cannot be better formulated than by Newton’s aphorism: If I have seen farther, it is by standing on the shoulders of giants.

*1.2.4 Unconditional access is required because of public funding.* Another argument is the fact that the research has been paid for out of public means. In this view, the government should not fund the same research twice, since that would be a waste of taxpayer’s money. On closer investigation, however, this position may be too simple. As Merton made clear, duplication of effort does not always amount to a waste of resources.<sup>42</sup> First, the replication of research, for purposes of validating or falsifying truth claims, is an essential element of the scientific method. Second, as Polanyi made clear, chances are that a particular problem is solved more quickly if worked upon by

multiple researchers. Third, if multiple researchers reach the same conclusions, the results are more likely to be accepted. Fourth, multiple researchers may have diverse approaches, give diverse interpretations and see diverse implications for subsequent research. A fifth, more prosaic consideration is that the data are not always fit for use by others. To that end, they must be made accessible and stored, which may require an *extra* investment. This extra investment is typically not funded by the funding bodies, since they typically only fund the original research. In this view, a researcher may legitimately reject a request for access to ‘his’ data on the ground that such access requires an investment which has not been publicly funded.

*1.2.5 Unconditional access is necessary because research is ‘data-driven’.* Since Merton formulated his Four Commandments, a number of developments have heightened the need of access to data.<sup>43</sup> Most recently, after Internet and the ICT-revolution, scientists increasingly make use of GRIDS. GRIDS are operating systems for virtual networks of computers which enable the bundling of hardware and information – such as computational power, data-storage and measurements devices.<sup>44</sup> GRIDS enable a quintessentially novel type of experiments and enable scientists to work on the same data simultaneously. The ‘permanent data-revolution’ has created ‘oceans of data’ in biomedical research, such as the sequence of the human genome, a database of three billion base-pairs.<sup>45</sup> The sequence illustrates how the boundaries between ‘data’ and research results have become blurred. The sequence is both milestone and a starting point for further research.<sup>46</sup>

One of the consequences of the data-explosion is that contemporary biomedical research has become *data-driven* rather than *hypothesis-driven*. One of the challenges in the post-genomic era, for example is the research into the “nature and nurture” of common complex disorders.<sup>47</sup> This research involves the correlation of millions of data-points by correlating molecular measurements of thousands of genes, proteins and metabolites with all sorts of clinical data (such as blood pressure, cholesterol, diabetes and MRI-scans). Subsequently or sometimes initially, these data are aggregated with data from databases resulting from comparable research data stored in other databases.<sup>48</sup> Any identified patterns and correlations can be used to formulate new hypotheses. This type of research may be aided by the creation of bio-banks, large-scale collections of “nature and nurture” information in the form of human biological material and associated health information. Such a bio-bank can be constructed by linking pre-existing molecular and clinical databases.<sup>49</sup> As we saw in the Introduction, the opening up of existing collections is part of the Netherlands National Genomics Strategy and one of the pillars of the EU for the harvesting of the potential of life-sciences and biotechnology. Obviously, this type of research and this strategy is to a large extent dependent on unconditional access to appropriate datasets.

*1.2.6 Unconditional access is necessary as science is used for policy-making: policy-embedded research.* In *Science and the Social Order*, Merton mentions a dinner for scientists in Cambridge. The scientists are said to toast:

*[t]o pure mathematics and may it never be of use to anyone!*

Science however, has come to be used as both an impetus and a justification for most, if not all kinds of far-reaching policy-making. As to public health policy, virtually all decision-making and priority setting nowadays must be ‘evidence-based’, meaning

based on scientific evidence. For example, as part of its 6th Framework Programme for research and development, the EU has established a special programme to enhance research for policy support in the area of public health, with an indicative budget of EUR 590 million. As ‘policy-enhancing’ instruments, scientific ‘truth-claims’ have assumed an impact that goes far beyond the traditional goal of extending certified knowledge for its own sake. This impact provides an extra argument for the replication of the scientific evidence that is used to justify a specific policy. The recent controversy over climate change and global warming illustrates this point. The Kyoto protocol aimed to address the environmental concerns has been ratified, without any replication of the original research findings underlying its policies. It was not until a couple of researchers undertook to replicate this research, that it became clear that it might be seriously flawed<sup>50</sup>. Obviously, replication of ‘policy supporting’ science requires unconditional access to research data.

*1.2.7 Unconditional access is necessary because speed matters and patients can't wait.* A final argument does, strictly speaking, not derive from the institutional goal or method of science, but is no less principled. It concerns the actual object of biomedical research: the patient. Ultimately, it is his problem science is trying to solve. Scientists however, are inclined to appropriate problems. Watson and Crick's decision to devote themselves to the discovery of the molecular structure of DNA, was against the academic etiquette. DNA was Wilkins' problem, not to be touched by others. As Watson put it:

*At this time molecular work on DNA in England was, for all practical purposes, the personal property (sic) of Maurice Wilkins, a bachelor who worked in London at King's College'. It would have looked very bad if Francis (Crick) had jumped in on a problem that Maurice had worked over for several years.<sup>51</sup>*

Wilkins, in turn, was having difficulties with Rosalind Franklin, his assistant, who claimed DNA as *her* problem.<sup>52</sup> Such an appropriation of problems is undesirable in biomedical research, at least to the extent the interests of patients are at stake. Unlike, let's say, the fruit fly, the object of this type of research, the patient, has a stake in the outcome. Even more so, the outcome may not only be vital for himself, but also for his family, his progeny, an entire population and, where globally spread diseases such as AIDS and SARS are concerned, the global community. Should researchers nevertheless refrain from ‘jumping in’ on their colleagues' problems, patients and their relatives will stand up to enforce unconditional access and sharing. For them ‘speed matters’; they cannot and will not wait until the researcher who claims *their* problem as *his* problem, has actually solved the problem. Recent examples are patient initiatives in the United States to bundle forces, collect material and data of their peers worldwide and make these available for the research community.<sup>53</sup> A group of parents of autistic children, for example, has, frustrated by the unwillingness of the researchers to share their data, created their own database, the Autism Genetic Resource Exchange. This database is accessible for every researcher and has so far yielded some 18 publications.<sup>54</sup> As a related, additional benefit, unconditional access obviates the need for a researcher to re-contact a patient or volunteer and subject him to repeat investigations.

*Conclusion Part I.* In addition to Merton's Four Commandments, scientific revolutions, the ‘Invisible Hand of Research’, public funding, the data-driven nature



of research and the interests of patients and the public at large, jointly and separately provide powerful arguments to postulate a Fifth Commandment: unconditional access to unpublished research data. The next part will examine the current *Practice* governing access to research data in biomedical academia.

## II In practice: Capitalism

### II.1 'Honest Jim'

As was to be expected, neither Merton's communist ideal nor the ideal of unconditional access are standard practice in the real life of science. In *The Double Helix* Watson reveals a number of incidents of 'capitalist' behaviour.<sup>55</sup> The quote in the Introduction already revealed that Watson and Crick had been 'stimulated by a general knowledge of the unpublished experiments and results' of Wilkins and Franklin, unbeknownst to the latter. Also, Wilkins, who was to share the Nobelprize with Watson and Crick, was said to have rejected a request by his competitor Linus Pauling for a copy of Wilkins' X-ray photos of DNA. Watson asserts that Wilkins wanted to study the data in more detail before releasing them.<sup>56</sup> And when Watson and Crick find out that Pauling made a mistake in his published description of the structure of DNA, they go to the pub to toast to the 'Pauling failure'.<sup>57</sup>

Watson's revelations struck the scientific community like a bombshell.<sup>58</sup> That was, however, not because of the departures from Merton's Four Commandments, but because Watson had violated the community's *omerta*. Violations of the sharing ethos were known to occur, but, whereof one cannot speak, thereof one must be silent. Since the publication of *The Double Helix*, however, a number of commentators have pointed at the fact that Merton's science ethos is not always adhered to.<sup>59</sup> In his review of *The Double Helix*, Merton asserted that, historically, Watson's competitive behaviour was not unique and a necessary corollary of the institutional emphasis on scientific priority.<sup>60</sup> Merton could not resist noting the intense and successful co-operation of Watson and Crick with other scientists. Thus, he managed to squeeze Honest Jim's behaviour within the 'competitive cooperation', he had described in his Four Commandments.

### II.2 Data-access in Biomedical Academia

*II.2.1 The Bermuda Principles.* The evidence presented in *The Double Helix* is, obviously, anecdotal. In contrast, a number of recent large-scale 'community resource projects' have achieved their objective of pre-publication release of the raw data they produced, suggesting adherence by the scientific community to the Five Commandments. Namely, the 1996 Bermuda principles, encouraging producers of large scale DNA sequencing assemblies to release prior to publication their data immediately for free and unrestricted use by the scientific community<sup>61</sup>, have been reinforced by various funding agencies, including the US National Human Genome Research Institute (NHGRI). To date, these policies have secured open access to at least 548<sup>62</sup> public genetic databases worldwide, available on the internet, including the large international nucleotide databases (EMBL and GenBank).<sup>63</sup> In 2003, the attendees of a meeting organized by the Wellcome Trust, reaffirmed the Bermuda principles of pre-publication release and recommended their extension to other existing and future 'community resource projects', such as the SNP-consortium and

the International HapMap Project<sup>64</sup>. However, the meeting also considered that beyond those large-scale ‘community resource projects’, many valuable small-scale data sets could come from other sources. Since those resources emerge from research efforts whose primary goal is not resource generation, the meeting considered that the contribution of these data to the public domain is more a voluntary matter. That raises the question to which extent such contributions are standard practice.<sup>65</sup>

*II.2.2 Global survey.* With the assistance of the Human Genome Organisation, the American Society of Human Genetics and the European Society of Human Genetics, I conducted a global survey among their members in the fourth quarter of 2003. The web-survey was completed by a total of 118 human geneticists.<sup>66</sup> While this was only a fraction of the sample taken and therefore not representative, the responses offer an interesting peek in the practice of data-access in academic genetics in a range of countries (over 15 different countries) and at a diversity of institutions (universities, hospitals and research institutes). A majority of 51% of the respondents responded that they did not grant access to their own databases to non-commercial institutions. Notably, in view of the predominantly public funding of the databases, only one third (35%) of the respondents reported to grant access to their databases to non-commercial institutions for free. These percentages were only slightly different for those databases that were labelled ‘unique’ or single-source.

*II.2.3 United States survey.* In the US, Campbell et al. conducted the first national survey on data-withholding among American geneticists. Their study yielded, *inter alia*, the following results.<sup>67</sup> In a 3 year period, 47% of the respondents had denied at least one request for additional information, data or material concerning published results. Most denials concerned requests for biomaterial (35%), sequence information (28%), applicable findings (25%), phenotype information (22%) and additional information concerning laboratory techniques. Due to these denials, 28% was unable to verify published findings.

*II.2.4 The Netherlands survey.* In a 2004 websurvey, Dutch biomedical researchers were asked whether they granted access to ‘their’ data to other scientists. This and related questions formed part of a survey I conducted as part of a survey of Dutch population-cohorts and patient-databases, conducted by the Royal Netherlands Academy of Sciences.<sup>68</sup> Its specific purpose was ‘to investigate whether the opening up of the banked material and the linking of existing datasets is possible’, which might help erect another pillar supporting the Netherlands National Genomics Strategy.<sup>69</sup> The survey was held among all principal investigators in the Netherlands in the area of general practitioner’s medicine and epidemiology, as well as the principal investigators in the area of asthma, Alzheimer’s disease, breast cancer, lymphoma cancer, rheumatoid arthritis and multiple sclerosis. The response rate was 60%. Asked whether they granted access to ‘their’ research data, the respondents answered as follows. 15% did not grant access, 9,3 % granted access on demand, without conditions, 1,5% made their data publicly available (anonimised) and 73,3% only gave access on special terms and conditions.<sup>70</sup> Admittedly, this type of survey has many limitations and additional empirical data is to be gathered before any definitive conclusions can be drawn. Nevertheless, the responses clearly suggest that a large majority does not grant unconditional access to their research data. Their attitude is at odds with the Fifth Commandment, postulated in Part I. As a result, it may deter or delay scientific revolutions, paralyse the ‘Invisible Hand’, delay patient cure and care,

result in suboptimal use of ‘unique datasets’ for the Netherlands National Genomics Strategy and undermine two of the three European pillars for the harvesting of the potential of the life sciences and biotechnology.

### II.3 Reasons for ‘capitalist’ behaviour

*II.3.1. Reasons for data withholding.* The respondents in the global and Netherlands surveys have not been asked to state the reasons for denying or imposing conditions on access. However, the Netherlands survey was designed so as to preclude a number of reasons that are typically advanced to justify the withholding of data, *i.e.* privacy<sup>71</sup>, technical inaccessibility of the data and/or its limited potential of the collection for ‘secondary use’. Specifically, 79% of the respondents confirmed that, assuming compliance with privacy rules, ‘their’ database could be used for research other than the research for which the data had originally been collected. And 76% indicated that, assuming again privacy-compliance, their database was technically (in terms of hard- and software) accessible for additional research by others.

In general, a variety of reasons could be advanced to explain why scientists depart from the sharing ethos: personal motives, financial interests, legal constraints, institutional incentives and uncertainty as to which data should be accessible.

#### II.3.2 Personal motives, financial interests and informal networks

*The Nobelprize.* Watson was rather blunt about his motives; honour, the desire to accomplish something to brag about against friends and, especially, girlfriends, the desire to defeat giants like Linus Pauling and, ultimately, to win the Nobelprize. As we have seen above, these competitive notions form part, to a certain extent, of Merton’s normative structure.

*Blood, sweat and subsidies.* The American survey revealed as an important motive for data-withholding the time and costs involved in meeting the requests. Another major motive was the desire to use the data for new research.<sup>72</sup> Having to share data may also lead to a drop in ‘authorships’, another yardstick for government and public funding bodies to assess and reward productivity. Apart from being a resource for new research and new publications, the data can also be a potential source for new patent applications and a new round of public (or private) funding, as illustrated by the Netherlands National Genomics Strategy. The Dutch government acknowledges the unique character of certain Dutch datasets and takes that into consideration when deciding on grant proposals. If the primary data-producers would be under an obligation to grant unconditional access to ‘their’ datasets, their opportunities to capitalize on their resource would be diluted to the extent other scientist would have a ‘free ride’ on their original efforts.<sup>73</sup> Such a ‘free ride’ might also prove a disincentive to start producing such databases in the first place.<sup>74</sup>

*Access by whom and for whom? Informal networks.* Proponents of unconditional access typically assume the standard situation: a secondary researcher requests access to a primary researcher’s data. In complex research, however, this situation will be a rarity. Hilgartner and Rauf have pointed out that researchers, both data-requesters and data-producers, form part of research-networks.<sup>75</sup> The decision whether or not to grant access is not made by a single scientist, but by multiple parties: the research

team, representing various departments or universities and public and private funding bodies<sup>76</sup>. And increasingly, as we saw above, patients have come to claim a say on access issues as well, patients being both data-suppliers and stakeholders. The parties requesting access to data could be competing research groups, complementary research groups, public and private funding bodies<sup>77</sup> and, again, patient interest groups. Factors to consider when deciding on a request for access will include the likelihood of future reciprocity,<sup>78</sup> recognition,<sup>79</sup> the proposed use of the data,<sup>80</sup> potential abuse of the data, loss of control over the data for future own use<sup>81</sup> and the quality of the researcher or research team making the request. Some data-collections have built a reputation which may suffer from publications of inferior research. And finally, the decision may be affected by the existence of symbiotic and informal networks between like-minded scientists.

### II.3.3 Legal constraints

*II. 3.3.1 Privacy.* To the extent biomedical research data contains personal data, the access to such data is subject to an array of intersecting privacy laws. In the UK for example, a Wellcome Trust report has identified the following current and proposed legislation governing public health researchers accessing existing collections of research data and biological samples:

- the Data Protection Act 1998;
- the common law of confidentiality;
- the Human Tissue Act;
- the Human Rights Act;
- the stance of local ethics committees;
- the application of Section 60 of the Health and Social Care Act 2001;
- the requirements of clinicians' regulatory bodies regarding patient confidentiality; and
- the Medicines for Human Use (Clinical Trials) Regulations 2003<sup>82</sup>.

As the resulting situation is awfully complicated, only a few basics will be discussed here.<sup>83</sup> Implementing the EU Data Protection Directive, the Dutch Data Protection Act<sup>84</sup> provides that personal data may be processed for research purposes, provided that the processor has taken appropriate security measures so as to ensure that processing is limited to this specific purpose.<sup>85</sup> Processing personal *health* data however, without that person's consent, is prohibited. Yet, this prohibition does not apply to processing for research purposes, provided that the research serves a public interest, the processing is necessary for the research concerned, asking explicit consent is either impossible or requires a disproportionate effort and the research provides for guarantees so as to minimise potential harm to the data-subject's privacy.<sup>86</sup>

*II.3.3.2 Informed consent for research.* A related legal impediment that may provide a legitimate reason to withhold access is the requirement of informed consent. For example, to the extent research data contains patient-data, the Dutch Act on the physician patient relationship provides that a physician may only grant access to patient data for scientific research in the limited events set forth in the Act and provided that the statutory conditions be met.<sup>87</sup> In all other events, any processing of

personal health data requires the patient's prior informed consent. A controversial issue in this respect is whether researchers are required to obtain specific re-consent for each new research project. While this is the case under the current construction of the informed consent requirement, there is a trend towards acceptance of simplifying existing specific informed consent requirements for previously unanticipated research use, which would obviate the need for *re*-consent for each new research project.<sup>88</sup> Notably, the global survey mentioned above revealed that most respondents (81%) indicated that they asked their patients or research subjects for their consent to use their data and material not only for the initial diagnosis, treatment or research, but also for future, unspecified research purposes. The Dutch survey revealed that 54% of the respondents asked consent for a specific research question, while 34% asked consent for research in general.<sup>89</sup>

*II.3.3.3 Quality, liability and a wrongful life.* Another legal constraint concerns the quality of biomedical data and the related potential for their producers to incur liabilities for the publication of inaccurate or incomplete data. For example, for a couple of decades now, geneticists have been collecting so-called mutations, changes in DNA comprising just one base-pair. A mutation can cause an inheritable disease or cancer. A major number of mutations has been discovered, described and included in the so-called 'Locus Specific Databases' ('LSDB's'). Prior to inclusion in the LSDB, the data have been verified by a curator. The trouble is that the data in these databases is not only used for fundamental research, but also for clinical applications, such as diagnosis. Obviously, it is vital that the mutations data are accurate and complete. In spite of a number of built-in safeguards and quality-controls, 100% accuracy and completeness can never be guaranteed. As a result, a mutation may wrongfully be held as harmful and thus give ground for, for example, an abortion or prophylactic surgery, or *vice versa*.<sup>90</sup> Depending on the circumstances, the original data-producer may be liable for any resulting damages, which liability may even extend, as per a recent Dutch Supreme Court ruling, to liability for a wrongful life. To limit his potential exposure, a researcher or his institution, may want, in addition to a proper disclaimer<sup>91</sup>, to impose conditions for access to the data, or only supply the data for a specific purpose or only grant class access, *i.e.* for accredited academics only.

*II.3.3.4 Proprietary claims and commercialisation of research.* Finally, access may be limited by proprietary claims, as is often, but not necessarily, the case for drug research sponsored by industry. Industry normally allows publication of the research results, subject to a right to comment and a sixty-day waiting period to allow for patent applications.<sup>92</sup> However, industry is likely to make a proprietary claim to the underlying research data as such, if only to satisfy the requirements for obtaining a market authorisation for the product concerned. In addition to proprietary claims by industry, academia itself is also increasingly staking its own proprietary claims. Researchers are being pushed, if not obligated under the terms of their grant, to commercialise the results of their research. Part of the 'mission' imposed by the Netherlands National Genomics Strategy is to have grantees "sell genomics-knowledge to the business community".<sup>93</sup> Given the ever increasing value of biomedical data, this mission is likely to extend to not only the filing of patent applications but also the exploitation of datasets, as they are a major source of patentable inventions. Arguably, the exploitation of these databases requires the licensing of exclusive rights to the data concerned, which, in turn, is likely to

compromise unconditional access by other academic researchers, who are likely to be equally pressed to capitalise on the results of their own publicly funded research.

### II.3.4 Which data, actually?

II. 3.4.1 *Typology of data.* Another reason that might explain data-withholding is the issue *which data* ought to be accessible. Proponents of unconditional access base their claim on the traditional model: a scientist produces ‘data’ (output), disseminates these data by means of a publication, which publication makes the published findings available for other scientists as ‘input’ for further research. In practice, however, things are not that simple.<sup>94</sup> First, a lot of data cannot be used by third parties without a proper explanation by the original data-producer, if only an explanation of the inclusion criteria used to determine the research sample. Second, as part of the scientific process, scientists record a host of divergent data: lab notebooks, raw data, derived variables, preliminary analyses, draft articles, grant applications etc.<sup>95</sup> And if the research involves human subjects, the Good Clinical Practice guidelines prescribe<sup>96</sup> a research protocol<sup>97</sup>, an Investigator’s brochure<sup>98</sup>, source documents<sup>99</sup>, ‘case-report forms’<sup>100</sup> and a clinical research report.<sup>101</sup> Should all these data be accessible? Or, for example, only the clinical research report?

II. 3.4.2 *Data-units or data-streams?* A more fundamental question is whether ‘data’ as such actually exist. According to Hilgartner and Rauf<sup>102</sup> there are only data-streams, heterogeneous collections of data, assembled in various formats. They give protein-crystallography as an example. Protein-crystallography comprises many forms of data, including ‘clones’ for the production of protein samples, protein crystallisation techniques, atomic models, algorithms to construct these models, atomic co-ordinates and computer generated images of molecular structures.<sup>103</sup> What data streams really are about, are the complex connections between these forms, the so-called ‘assemblages’. A second characteristic of data-streams is the heterogeneity of their factual status. The meaning and usefulness of some data are beyond reasonable doubt, while the meaning and usefulness of other data are so uncertain that they are even questioned by the data-producers themselves. Many data are on a gliding scale somewhere in between those two extremes. Their status also changes during the research process. Scientists are continuously analysing and interpreting data and that may explain their reluctance to release them too early.<sup>104</sup> This may explain, for example, Wilkins’ refusal to send his DNA photos to Pauling. In addition, data are constantly being processed. X-ray photos are converted into numbers, numbers into tables, graphics, models and images which may finally be published. This processing may change content, format and usefulness of the data.<sup>105</sup>

In brief, data are no pre-packaged units, capable of being shared or published. On the contrary, there is a continuous stream, which can be split, shared and published at a number of points and intervals, depending on the conventions of the specific research discipline concerned, according to Hilgartner and Rauf.<sup>106</sup> Their concept of data-streams applies *a fortiori* to contemporary biomedical research. As we have seen in Part I, this research involves the correlation of millions of data-points by relating molecular measurements concerning thousands of genes, proteins, metabolites with all sorts of clinical data (such as blood pressure, cholesterol, diabetes and MRI-scans).<sup>107</sup> This research is not about data-streams; it is about waves of data.

*Conclusion Part II.* With the notable exceptions of the publicly available raw data produced by a number of recent large-scale community resource projects, such as the HGP, the ideal world of unconditional access is not always the real world. This may be due to personal interests, scientific interests, financial interests, legal constraints such as privacy-concerns, informed consent requirements, quality issues, potential exposure to liability, proprietary claims, and conceptual doubt as to what research data exactly are. The principled arguments *pro* unconditional access (Part I) and the personal, institutional and legal arguments *contra* (unconditional) access (Part II) provide divergent, if not mutually exclusive and potentially conflicting perspectives. Part III will analyse whether this Gordian knot can be cut by operation of the law, in particular the double-edged sword of the European database-right. This right should provide an answer to the question who actually owns the research data.

### III The Law: who owns the research data?

#### III.1 *The European database right.*

*III.1.1 Introduction.* The question who owns the research data will be analysed on the basis of the EU database-right. This right was introduced in 1996 by the European Union and entered into force in the Netherlands in 1999. The right vests an exclusive right in the producer of a database to grant permission to extract and re-utilize the contents of the database.<sup>108</sup> If owned by the government it could be used as an instrument to enforce access; if owned by a researcher it could be used as an instrument to legitimise and enforce data-exclusivity. This Part first examines to which extent the database right applies to research databases and flags a number of complications. It then discusses the issue of allocation: who actually owns the database right in a research database.

*III.1.2 Case study: the Netherlands Twin Registry.* Due to the potentially unlimited diversity of biomedical research databases, any analysis of the above issues ‘*in abstracto*’ may get lost in assumptions. To move beyond assumptions and generalities, the application of the database-right to research data will be analysed using a concrete example: the Netherlands Twin Registry (“NTR”). In many ways, the NTR is representative for the databases that are the object of the Netherlands National Genomics Strategy. As part of the European GenomEUtwin program, it may also be representative for the databases that are part of the European biomedical research infrastructure. The NTR was incorporated at the Amsterdam Free University, for purposes of scientific research aimed at elucidating to which extent differences between individuals are determined by heritable and environmental factors.<sup>109</sup> The NTR comprises a large number of families having twins and contains *inter alia* data on birth weight, pregnancy, physical abnormalities, health and behaviour as well as physiological data, as blood pressure and cholesterol level.<sup>110</sup> The NTR also contains blood samples from which genetic data can be derived.

#### III.2 The database-right and research data

*III.2.1 Database-right and research data.* The database-right aims to protect collections of data which meet the statutory definition of a database. The definition reads as follows:

*A compilation of works, data or other elements, systematically or methodically arranged and independently accessible by electronic means or otherwise and of which the creation, control or presentation of the contents demonstrates in quantitative or qualitative respects a substantial investment.*<sup>111</sup>

*Prima facie* this right is squarely at odds with Merton's communism and the Fifth Commandment of unconditional access to research data. Because of its adverse implications for science, the right has been criticised by many observers of scientific research.<sup>112</sup> Indeed, the theoretical justification of the introduction of this novel right was clearly not the advancement of science. Rather, the database-right has been created with the explicit goal to promote the (European) database-*industry*. This triggers the question whether the database-right actually applies to scientific research databases. Neither the Dutch law, nor the European Directive, nor their respective legislative histories, however, provide for a ground to exempt research databases from the application of the database-right. On the contrary, the research exemption and the definition of research data in the preamble to the Directive<sup>113</sup> imply that this right does also apply to collections of scientific research data.<sup>114</sup> So far, the issue has not provoked any lawsuits before the Dutch, German or French courts. Notably, critics in the science community have not questioned its applicability and simply assume that it does apply to research databases. The *communis opinio* of legal commentators on the database-right in general is that the statutory definition of a database practically encompasses all types of data imaginable.<sup>115</sup> It is to be assumed therefore, that the database right could extend to collections of research data.

*III.2.2 Independent data.* The next element of the definition is that the data be 'independent', *i.e.* they must have a meaning of their own, regardless of the rest of the contents of the database. Whether that is the case, will vary from database to database. Conceptually, it is possible to have a database of which the individual data only have meaning when taken together, just as the various chapters making up a novel only make sense when read as a whole. In fact, the very goal of any research on the data will be to gain insight in the relationships between the individual data. The concept of data-streams even suggests that independent data never have a meaning of their own. Rather, there are many forms of data, which are constantly processed and their meaning may vary from experimental (*e.g.* Rosy's data) to paradigm. This conceptual approach however, cannot take away from the fact that usually subsets of a database can be independently studied for separate research questions, suggesting that the individual data in the database have independent significance. The NTR data, for example, are frequently used to create small subsets to allow more specific research.

*III.2.3 Systematically arranged and separately accessible.* This requirement holds that the database must have a search function allowing for direct retrieval of specific data.<sup>116</sup> It is generally held that digitally stored data will, by definition, meet this requirement.<sup>117</sup> Thus, even a collection of unstructured (clinical) data could qualify, as long as the collection is searchable for specific items, without the user having to scroll over all the data.<sup>118</sup> Obviously, the NTR satisfies this requirement and it seems safe to assume that most other biomedical research databases do so as well. 76% of the respondents in the KNAW-survey confirmed that their databases were accessible (qua hard- and software) for additional research by other researchers.



*III.2.4 Substantial investment.* The goal of the database-right is to promote investment in database-production. Only those database are eligible for protection that demonstrate a *substantial* investment. It is held that it is not necessary that the database has been produced for exploitation purposes. Databases produced for internal purposes or produced by volunteers are said to demonstrate substantial investment.<sup>119</sup> An investment may take the form of monetary investment, time, effort and energy, or ‘blood, sweat and tears’. Both the Directive and Netherlands law are silent as to when an investment qualifies as substantial. This will require a case-by-case analysis. Given the amount of time and money required to build the NTR, this database most likely demonstrates a substantial investment.

### III.3 Complications of the database-right

*III.3.1 Complications.* While many research databases will satisfy the various elements of the statutory definition, the application of the database-right to research databases nevertheless gives rise to three complicated questions. First, is a database eligible for protection, even if it is a mere ‘spin off’ from another investment. Second, does the database-right create a monopoly on ‘unique data’? Third, how does the right relate to the public domain?

*III.3.1.1 Complication 1: the ‘spin off’-theory.* As we have seen, the goal of the database-right is to promote investment in (European) database-production. Given this goal, it has been argued that those databases which are not a direct product of a specific investment but rather a spin-off from an investment that was made in another activity, do not qualify for protection.<sup>120</sup> This question was raised during the debate in the Dutch Parliament over the implementation of the database-right into Dutch law, when members of Parliament asked the following question. ‘The discovery of a new solar system requires a substantial investment in a telescope or a journey into space. Among other things, it will result in obtaining a collection of stars with names attached to them. Is such a collection a protected database?’<sup>121</sup>

In response to this question, the Dutch Minister replied that this collection does not satisfy the statutory definition of a database, because the investment was not *aimed* at obtaining the list of stars but concerned research by way of a telescope or a journey into space.<sup>122</sup> Commentators by and large concur, be it that some have argued that this could be different if observations and measurements are made for the specific purpose of creating a catalogue of stars.<sup>123</sup> While the Dutch lower courts and even the Dutch Supreme Court seem split on the issue, the ‘spin off-theory’ has recently been adopted by the European Court of Justice.<sup>124</sup> The ECJ construed the statutory requirement of an investment in obtaining the *contents* of a database as to require an investment made to *obtain* the existing elements *and* to put them in a specific database. For purposes of assessing whether a database demonstrates a substantial investment, investments that were made to merely create the elements that form the contents of a database do not suffice.<sup>125</sup>

The ECJ-cases concerned databases which had been produced by the commercial sports industry and subsequently used in the gambling-industry. It remains open to question therefore, whether the ‘spin off’-theory also applies to public investments in scientific research. On the one hand, no single public investment in science is, strictly speaking, *aimed* at creating a database. As we have seen in Part I, the institutional

goal of science is ‘the extension of certified knowledge’. The goal of the NTR, for example is not to create a twin registry, but to extend the knowledge concerning the nature and nurture of differences between individuals. As a matter of principle, given academic freedom, it is even questionable whether a public funding body can impose that its funding be applied for a specific research goal, such as building a database. And as a practical matter, we saw in Part I that the public funding of research in many cases does not include the costs of creating and maintaining accessible databases.

On the other hand, as Kuhn made clear, most research methods consist to a large extent of collecting, verifying and presenting data, incidentally the terms used in the Dutch statutory definition of a database. The explicit goal of the Human Genome Project, for example, was to create a database containing the fully annotated sequence of human DNA. The effort has produced a (downloadable) public database which constitutes both an ‘extension of certified knowledge’ *and* a database for future research. And for contemporary *data-driven* biomedical research, creating databases is a crucial element. If we were to adopt this approach, any investment in research amounts to an investment to *obtain* existing elements *and* to collect these in a database, and thus such investment would meet the spin-off criterion as adopted by the ECJ.

*III.3.1.2 Complication 2: unique data.* The application of the EU database-right is further compounded by the phenomenon of ‘unique data’. The preamble to the EU database-right Directive explicitly provides that the database-right may not give rise to the creation of a new right on data proper.<sup>126</sup> A database-right is only an exclusive right in the database as such. It does not pre-empt anyone from collecting the same data to create his own database. However, this framework does not apply in the event the collected data are one of a kind or ‘unique’. For example, a specific twin’s weight or length at a specific age can be measured only once. Conceptually, then, as a corollary of the preamble to the Directive, the database-right should simply not apply to databases containing unique data. The Directive, however, does not provide for such a ban, but purports to offer alternative solutions to secure access to unique data.

First, if unique data make up only an *insubstantial* part of the database, they can be extracted and re-utilized without the consent of the owner of the database-right.<sup>127</sup> Second, the Directive provides that the exercise of a database right is at all times subject to the competition laws. Under these laws, a data-monopolist abusing his monopoly can be forced to grant access to the data on the terms of a compulsory license. In practice, this is not a viable option for most scientists since competition proceedings are a costly and time-consuming affair that public funding bodies are unlikely to fund. Third, the database-right provides for a ‘research exemption’, which allows the lawful user to use the database for research purposes, without the consent of the owner of the database. This solution too, however, is rather limited.<sup>128</sup> It requires a ‘lawful user’, *i.e.* a user who has legitimate access to the database, *e.g.* under the terms of a license. Also, the exemption only applies if the database has been made available, one way or the other, to the public. This latter requirement may be satisfied once a researcher has published his findings, as most journal policies demand that authors make the data underlying their published findings available to others for purposes of verification.<sup>129</sup> The research exemption is also limited in that it is not compulsory; it can be excluded in the license of a lawful user. And the exemption only covers the use and not the re-utilisation of the database, although this restriction

is not necessarily an impediment to research, as research typically only requires the right to use the data and not the right to make the data available to the public.<sup>130</sup> Furthermore, the exemption limits access to the extent justified by the non-commercial purpose, a line which, in contemporary, 'profit driven' academia, will be hard to draw. Finally, the exemption as provided in the Directive was optional. The review of the implementation of the Directive in the Member States revealed that France and Italy had not implemented the research exemption.<sup>131</sup>

*III.3.1.3 Complication 3: public databases and the public domain.* A third complication, at least under Dutch law, is that the database-right does not apply to databases which have been produced by the 'public authority', unless the database-right has been reserved either by statute, regulation or decree, or in a specific case by means of an explicit notice attached to the database itself or an explicit reservation when the database was made available to the public.<sup>132</sup> This provision is based on article 13 of the Directive which provides that the database-right does not prejudice existing rights to access public documents. The rationale behind this provision is that public works, produced by the public authorities should in principle be part of the public domain.<sup>133</sup> 'Public authorities' include public institutions, to the extent they are operating within the scope of their public remit and/or within their public competences.<sup>134</sup> Obviously, public research funding bodies will usually satisfy this criterion. The resulting situation is rather awkward. While these bodies are likely to qualify as the producers of the database (as we will discuss next), the database-right does not apply, except if it has been reserved. Whether it has been reserved is to be judged on a case-by-case basis, as the database may be funded by a variety of public funding bodies. The terms and conditions of the Netherlands Organisation for Research Funding (NWO), for example, provide that its grants *may* be subject to specific conditions concerning ownership rights in respect of the databases that are produced in the course of the funded research.<sup>135</sup> In the event NWO elects not to reserve its database-right, the databases generated by the funded research seem to form part of the public domain. While that outcome may be intuitively appealing, whether it is desirable for scientific research-databases, is open for debate. Apart from privacy concerns, a number of legitimate concerns presented in Part II argue for placing limitations on unconditional access. It is unclear which party is in a position to impose such limitations if the database forms part of the public domain.

#### **III.4 Who actually owns the database-right?**

*III.4.1 The producer: the funding body or the researcher?* The database-right vests in the producer of the database. The producer is the person who bears the *risk* of the investment in the database. An investment may comprise a financial or a professional investment.<sup>136</sup> However, for purposes of determining who bears the risk, only financial risks will be taken into account.<sup>137</sup> Employees and contractors<sup>138</sup> who do not bear a financial risk, cannot be the owners of the database-right.<sup>139</sup> Departing from the Directive, the Dutch legislator considers it also irrelevant who took the *initiative* to create the database.<sup>140</sup> As long as the initiative and the funding are provided by one and the same person or entity, this should not be a problem. However, in the case of publicly funded research, the initiative and the responsibility for the database will typically vest in the researcher, whereas the funding body bears the financial risk. In that event, it is unclear which person qualifies as the producer,<sup>141</sup> although the financial component is probably decisive. On the one hand, this seems a desirable

outcome: the sword of the database-right would thus provide an instrument for public funding bodies to secure all scientists access to the data produced by their peers. This outcome would also underpin the pillars for the harvesting of European life sciences and biotechnology research. On the other hand, being a producer may have undesirable implications for a public funding body. As the owner of the database-right, it will have to handle requests for access and impose conditions to meet the legitimate objections against unconditional access advanced in Part II. This will require expert knowledge of the area of research in question which the body may not have and/or a balancing of interests the body may not be capable of handling.

*III.4.2 Multiple funding bodies: co-producers.* The previous analysis was based on the rather hypothetical assumption that a research database is funded by a single funding body. In practice, this is unlikely to be the case. The NTR, for example, is only partly funded out of the so-called primary funding, a lump sum amount allocated by the government to the Netherlands universities for further distribution among their departments.<sup>142</sup> Additional funding is provided by the European Union which funds the GenomeEUtwin project of which NTR and a group of European and Australian twin registries form part.<sup>143</sup><sup>144</sup> The GenomeEUtwin project, in turn, forms part of the international consortium ‘P3G’, which has been awarded Canadian and, recently, EU funding.<sup>145</sup> In addition, the NTR forms part of the Centre for Medical Systems Biology (“CMSB”), a consortium of Netherlands research institutes, funded by NWO, as part of the National Genomics Strategy, the so-called secondary funding. And as a condition for this secondary funding, the amounts involved may have been ‘matched’ by the Free University of Amsterdam, so that the Free University could also qualify as a producer. The situation gets even more complicated if these investments took place successively, and each investment has resulted in amendments and modifications of the NTR. Successive investments in an existing database create new database-rights in the databases that have resulted from these investments.<sup>146</sup> The answer, then, to the question who actually owns the database-right in the NTR is that multiple funding bodies are likely to qualify as ‘co-producers’, collectively owning the database,<sup>147</sup> assuming the bodies have reserved the database-right and absent contractual provisions to the contrary.

*Conclusion Part III.* While the database-right does apply to research databases, this application is complicated by the ‘spin off-theory’, unique data monopolies and the requisite reservation by the public authorities. The right belongs to the producer, being the person who bears the risks of the financial investments in the database. To the extent biomedical databases are funded by multiple funding bodies, the respective producers are jointly entitled to the database-rights in the database. The European database-right is an inevitable but awkward fit for databases comprising scientific research data. Apart from the fact that its provisions do not readily apply to the specific contexts of this type of databases, its most striking feature is that it fails to acknowledge the ‘investment’ made by a researcher or his institution in the production of such databases. In conclusion, the European database-right does not seem to be the sharp sword to cut the Gordian knot in which the issue of access to research databases is tied up.

#### IV. Policy: Greasing the Wheels of Research

*IV.1 Policy or No Policy?* If there is no reason to develop policy, then that is a good reason not to develop policy. Existing co-operation between academic research institutes and a number of widely used data-collections suggest that either data-access was no issue to these co-operations or that the data-producers involved were able to strike a deal on the terms of access. However, the principled arguments *pro* unconditional access (Part I) and the laws and practical considerations *contra* (unconditional) access (Part II) offer conflicting perspectives. The resulting situation is compounded by the uncertainty created by the database-right as to who holds legal title to the databases. This lends urgency to the formulation of at least a number of principles and recommendations. Therefore, this final Part will explore whether the two opposing concerns – unconditional access vs. legitimate restrictions – can be accommodated by the adoption and implementation of a general policy for access to biomedical data. As an additional benefit, a well designed policy could remedy the adverse implications of the database-rights, by allocating certain rights to the scientist or his institution or by other means granting appropriate credit. This could also help alleviate the plight of funding bodies that may be surprised to find themselves as the owners of the databases they have funded. The above applies in particular to large, *data-driven* research projects in biomedical research. As we saw in the Introduction, two pillars for the harvesting of the potential of life-sciences and biotechnology are access to detailed, up-to-date and available bioinformatics databases and open access to knowledge. The more access becomes vital for research, the better it may be to have a clear set of rules and principles to guide access decisions. A clear policy could help avoid costly and time-consuming negotiations and prevent stalemates, in an era where speed matters and patients can't wait. This way an access policy could actually grease the wheels of research. A final reason for adopting a policy is the fact that funding agencies in other countries have recently established data-sharing policies, which may extend to forms of multinational research. Most notably, the US NIH adopted in 2003 its Data-Sharing policy, which requires grant applicants to include a plan for sharing final research data.<sup>148</sup>

*IV.2 Accessibility: compulsory or voluntary?* Any policy is eventually dependent on political and scientific choices. The most fundamental choice to be made is whether the leading principle should be that access should be always be voluntary or that access should be compulsory, subject to a predefined set of conditions. The conclusions concerning the database-right indicate that the bodies funding the research own the resulting data and thus they are the entities that ultimately make this decision.

*IV.3 General requirements.* A policy should in any event meet the following requirements:

- Be transparent, flexible and doing justice to the context of specific research, the status of the (unique) data involved and the legitimate interests of both researchers and patients;
- Be compliant with applicable legislation, in particular the privacy rules;
- Be in line with data-sharing policies of national and international journals;

- Be in line with grant-policies; references to the database in publications should be a factor to consider in grant applications and a factor in the performance assessment of the institutions concerned;
- Be enforceable; including extra grants to make and maintain data available;

*IV.4 Policy issues.* The issues set forth below reflect and seek to accommodate the concerns for and against accessibility of research data set forth in Part I and Part II. Some issues draw on the issues listed in the NIH Data Sharing Policy. A policy should in any event address the following issues:

- Definition of the data that should be made accessible, taking into account the concept of data-streams, in particular as regards the type and the status of (unique) data;
- Quality control and peer review;
- The right of the maker of the database to be the first to publish; the moment third parties should get access;
- Mode of accessibility, which may be dependent on the type of data, number of expected requests, sensibility of data etc. The NIH-data sharing policy, for example, mentions the following forms:
  - Publications;
  - Data-archives;
  - ‘Data-enclaves’;
  - Any combination of the above
- Mode of recognition by the requesting party in publications and grant applications, based on the database;
- The grounds on which access may be denied, deferred or subjected to a predefined set of conditions, such as the goal and the quality of the research proposal, and/or the quality of the requesting researcher,<sup>149</sup> approval by medical-ethics committee, reasonable doubt as to status of the data;
- The conditions to which access may/may not be subjected;
- Reimbursement of costs;
- Transfer of data to third parties;
- Allocation of IP rights to the results of the research;
- Liabilities and disclaimers;
- Enforcement and mechanisms for conflict resolution;

## Conclusion

Access to detailed, up-to-date and available bioinformatics databases is identified by the European Commission as a pillar for the harvesting of the potential of life-sciences and biotechnology. Unconditional access to research data, however, is at odds with the primary interest of every scientist to be the first to make a discovery. This classical dilemma forces us to consider the *principle*, the *practice* and the *law* regarding access for academic researchers to unpublished research data. As to *principle*, it is argued that, in addition to the *Communism* postulated in Merton’s Four Commandments for the conduct of science, scientific revolutions, the ‘Invisible Hand of Research’, public funding, the data-driven nature of biomedical research and the interests of patients and the public at large, provide powerful arguments to postulate a Fifth Commandment: unconditional access to unpublished research data. However, in

practice, access to research data is governed by *Capitalism*, due to the egos of scientists, the financial interests of their institutions, public privacy-concerns, proprietary claims and conceptual doubt as to what data exactly are. The *Law*, in the form of the European database-right, potentially addresses the conflicting perspectives on access to data offered by communism and capitalism. However, the European database-right is an awkward fit for databases comprising scientific research data and does not provide a straight answer to the question who owns the database. It thus fails to provide a sharp sword to cut the Gordian knot of principled arguments *pro* unconditional access and the laws and practical objections *contra* (unconditional) access. As a matter of *Policy*, then, at least a number of principles and recommendations should be formulated, if only to create legal certainty. In addition this policy should seek to remedy the failure of the Law to acknowledge the ‘investment’ made by a researcher or his institution in the production of databases, either by allocating certain rights to the scientist or his institution or by other means granting appropriate credit. That could also help alleviate the plight of funding bodies that may be surprised to find themselves as the joint legal owners of the databases produced by the research projects they have funded.

## Epilogue

*The Bermuda Principles.* Apart from guidance provided by existing policies regarding specific issues and existing repositories (*e.g.* pathology archives), guidance may be found in examples and experiences, such as the NIH Data-Sharing policy. A major source of inspiration ought to be, *mutatis mutandis*, the Bermuda principles adopted by the Human Genome Project (HGP). Fifty years after the discovery of the molecular structure of DNA, the HGP completed the assembly of the human DNA sequence. After it had abandoned, at the urge of Watson, its initial drive to patent the cDNA sequences it was producing, the HGP adopted the Bermuda principles to secure rapid and unconditional access to the sequence data. Under these principles all participating scientists were required to release their unpublished DNA-sequence data in public databases.<sup>150</sup> The first director of the HGP was ‘Honest Jim’ Watson.

## Acknowledgments

The author is grateful to the following persons for commenting on an earlier draft of this article: Professor GJB Van Ommen (department of Clinical Genetics, Leiden University Medical Centre and director of the Centre for Medical Systems Biology), Professor D Postma (University of Groningen Medical Centre and Chair of the Commission on Multi-factorial Diseases of the Royal Netherlands Academy of Sciences), and Drs G van der Starre (Director of the Netherlands Genomics Initiative). The author is also indebted to Elaine Strass and Gil Moore of the American Society of Human Genetics who made the global web-survey available for online submission and provided invaluable assistance in processing the responses. I also would like to thank the members of HUGO, the American Society of Human Genetics and the European Society of Human Genetics for participating in the study. Finally, I am indebted to the Netherlands Royal Academy of Sciences for allowing me to participate in the design of their survey of Dutch biomedical research databases, to the respondents to this survey and to Dr Mirian Brink and Jeanet van Zoelen of the Academy for their prompt assistance in processing the resulting survey-data. The study and the surveys presented in this publication were conducted by the author in

his capacity of research fellow of the Netherlands NWO Genomics Initiative and the Centre for Medical Systems Biology.

- 
- <sup>1</sup> J D Watson, F H C Crick, *Molecular Structure of Nucleic Acids*, 4356 Nature (1953) 737.
- <sup>2</sup> J D Watson, *The Double Helix, A Personal Account of the Discovery of the Structure of DNA*, Norton Critical Editions in the History of Ideas (1968), p 105.
- <sup>3</sup> M. Perutz, Letter to the Editor of Science, Science, June 27, 1969, pp. 1537-1538.
- <sup>4</sup> Directive 96/9/EC of the European Parliament and the Council, adopted March 11, 1996 concerning the protection of databases (OJ L77/20). See J A Bovenberg, *Should Genomics Companies Set Up Database in Europe?* Nature Biotechnology (2000), Vol 18, nr 19, p. 107.
- <sup>5</sup> Official policy adopted by the Dutch Government *Knowledge Base Genomics* dated 16 July 2001, TK 27 866, p. 10; see also National Research Council Genomics, *Strategic Plan 2002-2006*, dated 18 July 2002, p. 6.
- <sup>6</sup> Communication from the Commission to the European Parliament, the Council, the Economic and Social Committee and the Committee of the Regions: *Life sciences and biotechnology - A Strategy for Europe* [COM(2002) 27 final - Official Journal C 55 of 02.03.2002].
- <sup>7</sup> R K Merton, *The Normative Structure of Science*, The Sociology of Science (1973), 267-278.
- <sup>8</sup> Merton, op. cit note 7, p. 270.
- <sup>9</sup> *Ibid.*
- <sup>10</sup> *Ibid.*, p. 275.
- <sup>11</sup> *Ibid.*, pp. 273-274.
- <sup>12</sup> *Ibid.*, pp. 274.
- <sup>13</sup> Compare Newton's aphorism *If I have seen farther, it is by standing on the shoulders of giants.*
- <sup>14</sup> Merton, op. cit. note 7, p. 277.
- <sup>15</sup> *Ibid.*, pp. 277-278
- <sup>16</sup> R K *Priorities in Scientific Discovery* in The Reward System of Science, in The Sociology of Science, pp. 286-324.
- <sup>17</sup> Merton op. cit. note 7 p. 273. See also R S Eisenberg, *Proprietary Rights and the Norms of Science in Biotechnology Research*, Yale Law Journal, (1987), 177, p. 183.
- <sup>18</sup> Merton, op. cit. note 16, pp. 297-302..
- <sup>19</sup> E.g. T E Hedrick, *Justifications for the Sharing of Social Science Data*, 12 Journal of Law and Behavior, (1988) 2, pp. 165-167.
- <sup>20</sup> See R S Eisenberg, *Patents and the Progress of Science: Exclusive Rights and Experimental Use*, 56 U.Chi.L.Rev. 1017, p. 1059.
- <sup>21</sup> T S Kuhn, *The structure of Scientific Revolutions*, (1962), p. 24.
- <sup>22</sup> *Ibid.*, p. 35.
- <sup>23</sup> *Ibid.*, Chapter IV 'Normal Science as Puzzle-Solving'.
- <sup>24</sup> Watson, op. cit note 2, p. 9, footnote Watson omitted.
- <sup>25</sup> Kuhn, op. cit. note 21, p. 36-37.
- <sup>26</sup> Kuhn, op. cit. note 21, specifically Chapter I, *The Route to Normal Science* and Chapter II, *The Nature of Normal Science*.
- <sup>27</sup> Kuhn, op. cit. note 21, pp. 5-6 and p. 24 et subseq. See also T S Kuhn, *The Essential Tension*, p. 233-234.
- <sup>28</sup> Kuhn, op. cit. note 21, Chapters VII, VIII and IX.
- <sup>29</sup> H Butterfield, *The Origins of Modern Science, 1300-1800*, London (1949), quoted by Kuhn, op. cit. note 21 p. 85.
- <sup>30</sup> R S Eisenberg, op. cit. note 20, p. 1059. See also H. Reichman, P F. Uhlir, *A contractually reconstructed research commons for scientific data in a highly protectionist intellectual property environment*, (2003) 66-SPG Law & Contemp. Probs. 315.
- <sup>31</sup> Eisenberg, op. cit. note 20 p. 1059.
- <sup>32</sup> *Ibid.*, pp. 1059-1066.
- <sup>33</sup> M Polanyi, *The Republic of Science: Its Political and Economic theory*, 1 Minerva 54, 56 (1962), published in *Sciences Bought and Sold, Essays in the economics of Science*, Ph. Mirowski en E-M Sent (red.), 465-485, p. 467.
- <sup>34</sup> Polanyi, op. cit. note 33 p. 467. See also A. Rip, *The Republic of Science in the 1990s*, Journal of Higher Education 28: 3-23, 1994, p. 12.
- <sup>35</sup> Polanyi, op. cit. note 33 p. 466.
- <sup>36</sup> *Ibid.*, p. 467.



- <sup>37</sup> *Ibid.*, p. 466-477.
- <sup>38</sup> *Ibid.*, p. 477.
- <sup>39</sup> *Ibid.*, p. 477.
- <sup>40</sup> E.g. R S Pindyck en D. Rubinfeld, *Microeconomics*, 4e ed.(1998), Chapter 17, Markets with asymmetric information, p. 617.
- <sup>41</sup> Eisenberg, op. cit. note 20, p. 1055-1059. Hedrick, op. cit. note 19, p. 166.
- <sup>42</sup> R K Merton, *Multiple Discoveries as Strategic Research Site* in *The Reward System of Science* published in *The Sociology of Science*, p. 381.
- <sup>43</sup> E.g. Committee for a study on promoting Access to Scientific and Technical Data for the Public Interest, National Research Council (US) 2004 *A question of Balance, Private Rights and the Public Interest in Scientific and Technical Databases*, and Committee on Issues in the Transborder Flow of Scientific Data, US National Committee for CODATA, Commission on Physical Sciences, Mathematics, and Applications, National Research Council (US), 1997 *Bits of power: issues in global access to scientific data*.
- <sup>44</sup> GRID Forum Nederland, accessible at <http://gridforum.nl>, date last accessed April 23, 2005.
- <sup>45</sup> [International Human Genome Sequencing Consortium](http://www.internationalhumangenome.org/), Initial sequencing and analysis of the human genome, *Nature* **409**, 860 - 921 (15 February 2001).
- <sup>46</sup> T G Wolfsberg, K A Wetterstrand, M S Guyer, F S Collins & A D Baxevanis, A user's guide to the human genome, *Nature* volume 32 supplement pp 1 – 79.
- <sup>47</sup> F S Collins, E D Green, A E Gutmacher and M S Guyer, on behalf of the US National Human Genome Research Institute, *A vision for the future of genomics research*, *Nature*, vol 422, 840, 24 April 2003.
- <sup>48</sup> Genomics and Healthcare, Value Added by research, education and entrepreneurship (April 2001), position paper for the incorporation of a centre of excellence for Genomics and Healthcare, Leiden University Medical Centre and the Faculty of Mathematics and Physics, p. 84.
- <sup>49</sup> Forum 'Biotechnology and Genetics, position paper on Biobanking, letter dated 16 September 2004 to the Deputy Minister of the Ministry of Healthcare, accessible at <http://www.forumbg.nl/werkgroepen/biobanken/>, date last accessed April 18, 2005.
- <sup>50</sup> S McIntyre, R McKittrick, *Hockey sticks, principal components, and spurious significance*, *Geophysical Research Letters*, Vol. 32, L03710, doi:10.1029/2004GL021750, 2005.
- <sup>51</sup> Watson, op. cit. note 2, p.13.
- <sup>52</sup> Watson, op. cit. note 2, pp. 14-15.
- <sup>53</sup> E.g. J A Bovenberg, *Inalienably Yours? The new case for an inalienable property right in human biological material: empowerment of sample donors or a recipe for a tragic Anti-Commons?* (2004) 1:4 SCRIPT-ED, *Journal of Law and Technology*.
- <sup>54</sup> *Whose DNA Is It, Anyway?* Aaron Zitner, 2003 Los Angeles Times.
- <sup>55</sup> His own behaviour earned him the nickname 'Honest Jim', which was the original title of his account (Watson, op. cit. note 2, pp. 7 and 185).
- <sup>56</sup> Watson, op. cit. note 2, p. 15.
- <sup>57</sup> Watson, op. cit. note 2, p. 95.
- <sup>58</sup> See the reviews in Watson, op. cit. note 2.
- <sup>59</sup> For an overview see N Stehr, *The Ethos of Science Revisited*, in *The Sociology of Science*, J. Gaston ed. (Jossey-Bass, 1978) pp. 172-196, finding that the criticism is largely based on Kuhn. See also N Wade, *Twentieth Century Fund, Inc. The Science Business* (1984), pp. 29-30. For a recent in depth study see J H Reichmann en P F Uhlir, op. cit. note 30.
- <sup>60</sup> R K Merton, *Making It Scientifically*, *New York Times Book Review*, 25 February 1968, published in Watson, op. cit. note 2, p.216.
- <sup>61</sup> The Bermuda Principles, accessible at <http://www.gene.ucl.ac.uk/hugo/bermuda.htm>, date last accessed; April 29, 2005.
- <sup>62</sup> M G. Tyshenko, W Leiss, *Genomics and the Current State of Public Databases*. GE3LS Symposium, Poster with abstract. February, 2004. Vancouver, Canada.
- <sup>63</sup> Reaffirmation and Extension of NHGRI Rapid Data Release Policies: Large-scale Sequencing and Other Community Resource Projects, February 2003, accessible at <http://www.genome.gov/10506537>.
- <sup>64</sup> The Wellcome Trust, *Sharing Data from Large-scale Biological Research Projects: a System of Tripartite Responsibility*, Report of a meeting organized by the Wellcome Trust and held on 14-15 January 2003 at Fort Lauderdale, USA, accessible at <http://www.wellcome.ac.uk/en/1/awtpubrepdat.htm>, date last accessed; May 24, 2005.

- <sup>65</sup> For empirical research into practical applications of Merton's theory in its entirety, see N Stehr, *The Ethos of Science Revisited*, in *The Sociology of Science*, J. Gaston ed. (Jossey-Bass, 1978) p. 184.
- <sup>66</sup> All data were originally collected by the American Society of Human Genetics and are currently on file with the author who is solely responsible for their processing and interpretation. All data are, in an anonymized format, available on request.
- <sup>67</sup> E G Campbell et al., *Data Withholding in Academic Genetics, Evidence from a National Survey*, *Journal of the American Medical Association* (2002); 287:473-480.
- <sup>68</sup> The other survey results will be published by the Royal Academy of Sciences in the course of 2005.
- <sup>69</sup> Forum Biotechnology and Genetics, Position paper Biobanking, letter dated 16 September 2004 to the Deputy Minister of Health, available at <http://www.forumbg.nl/werkgroepen/biobanken/>, date last accessed April 29, 2005.
- <sup>70</sup> 0,9% did not respond. All data processed by and on file with the Netherlands Royal Academy of Sciences.
- <sup>71</sup> Art.7:458 art.7:457 Dutch Civil Code. See C Ploem, *Tussen Privacy en wetenschapsvrijheid, Regulering van gegevensverwerking voor medisch-wetenschappelijk onderzoek* (diss. 2004).
- <sup>72</sup> Campbell et al. op. cit note 67.
- <sup>73</sup> E.g. Hedrick op. cit. note 30, pp. 168-169, S. Hilgartner en Sherry I. Brandt-Rauf, *Data Access, Ownership, and Control: Toward Empirical Studies of Access Practices*, 15 *Knowledge* 4 (1994), 355, 364. S Storer, *Social System of Science*, p. 129-131. Eisenberg, op. cit. note 17, p. 197.
- <sup>74</sup> B Stanley en M Stanley, *Data Sharing, The Primary Researcher's Perspective*, 12 *Law and Human Behavior* (1988) 2, 173, p. 177.
- <sup>75</sup> Hilgartner & Rauf, op. cit. note 73 pp. 362-363.
- <sup>76</sup> *Ibid.* p.364
- <sup>77</sup> *Ibid.*.
- <sup>78</sup> Campbell et al. *supra* note 67.
- <sup>79</sup> Stanley & Stanley, op. cit. note 74, p. 177.
- <sup>80</sup> *Ibid.*, pp. 175-176.
- <sup>81</sup> *Ibid.*, p. 178.
- <sup>82</sup> Wellcome Trust, *Public Health Sciences: Challenges and Opportunities*, report of the Public Health Sciences Working Group convened by the Wellcome Trust, March 2004, available at [http://www.wellcome.ac.uk/doc\\_WTD003192.html](http://www.wellcome.ac.uk/doc_WTD003192.html), date last accessed July 20, 2005.
- <sup>83</sup> For an in depth study of the UK situation, see W Lowrance, *Learning from Experience, Privacy and the Secondary use of Research Data in Health Research*, The Nuffield Trust, 2002, accessible at <http://www.nuffieldtrust.org.uk/publications/detail.asp?id=0&Pid=45>, date last accessed, July 20, 2005. For a detailed analysis of the Netherlands situation see C Ploem, *Tussen Privacy en wetenschapsvrijheid, Regulering van gegevensverwerking voor medisch-wetenschappelijk onderzoek* (diss. 2004).
- <sup>84</sup> **Data Protection Act of July 6, 2000 (DPA).**
- <sup>85</sup> Art. 9 paragraph 3 DPA.
- <sup>86</sup> Art. 23, paragraph 3 DPA.
- <sup>87</sup> Art.7:458 Dutch Civil Code.
- <sup>88</sup> B M Knoppers, *Biobanks: simplifying consent*, 5 *Nature Reviews Genetics*, (2004), 485.
- <sup>89</sup> 12% provided a different answer.
- <sup>90</sup> R G H Cotton and O Horaitis, *Access to Mutation Databases For Research Purposes*, Background Paper for OECD high-level expert meeting on Large-Scale Human Genetic Research Databases, Tokyo February 2004, OECD report forthcoming.
- <sup>91</sup> E.g. the disclaimer of the Human Gene Mutation Database at the Institute of Medical Genetics in Cardiff: <http://archive.uwcm.ac.uk/uwcm/mg/hgmd0.html>
- <sup>92</sup> The period recommended by the Dutch National METC.
- <sup>93</sup> Official Policy of the Dutch Government, op. cit. note 5, pp. 6 and 9.
- <sup>94</sup> See also J Sieber, *Data Sharing, Defining Problems and Seeking Solutions*, 12 *Law and Human Behavior* (1988) 2, 199, 200.
- <sup>95</sup> *Ibid.*
- <sup>96</sup> Guidelines for 'Good Clinical Practice', (CPMP/ICH/135/95) ('GCP'), incorporated by reference into the EU Directive on Clinical Trials (2001/20/EG).
- <sup>97</sup> GCP 1.44 and 6.
- <sup>98</sup> A compilation of clinical and pre-clinical data on the research product, including data on all relevant pre-clinical studies, GCP 1.36 and 7.

<sup>99</sup> Original documents, data and registries such as hospital records, clinical graphs, lab notes, lab diaries, GCP 1.52 and 8.3.13.

<sup>100</sup> GCP 1.11.

<sup>101</sup> GCP 8.4.8.

<sup>102</sup> Hilgartner & Rauf, op. cit. Note 73, p. 355.

<sup>103</sup> *Ibid.*, pp. 362-363.

<sup>104</sup> *Ibid.*

<sup>105</sup> *Ibid.*

<sup>106</sup> *Ibid.*

<sup>107</sup> Genomics and Healthcare, op. cit. note 48, p. 84.

<sup>108</sup> *Supra* note 4.

<sup>109</sup> Accessible at <http://www.tweelingenregister.org/>, date last accessed July 12, 2005

<sup>110</sup> *Ibid.*

<sup>111</sup> Article 1 paragraph 1 sub (a) of the Act of 8 July 1999, implementing Directive 96/9/EC into Dutch law, *Stb.* 1999, 303, as amended by the Act of 6 July 2004, *Stb.* 336, entered into force 1 September 2004 (“Act on Database-right”).

<sup>112</sup> International Council of Scientific Unions, ‘Position Paper on Access to Databases, prepared by the ICSU/CODATA group on Data and Information’, submitted to ‘WIPO Information Meeting on Database Protection Geneva, September 17-19, 1997 (“ICSU-paper”’, accessible via [www.codata.org/codata/data\\_access/index.html](http://www.codata.org/codata/data_access/index.html). See also Bits of power, op. cit. note 43, pp. 153-154. Five years into the adoption of the Directive, criticism has all but faded, as evidenced by the Review of the transposition of Directive 96/9/EC into the legislation of the EU Member States pursuant to Article 16.3 of the Directive, performed by Nauta Dutilh under Contract ETD/2001/B5-3001/E72, on file with author. See also A question of Balance, op. cit. note 43, pp. 69-72, and the Royal Society’s Report *Keeping Science Open* (5.4: considering the database-right not appropriate in the research context). For further reading see J.H. Reichmann and P. Samuelson, *Intellectual Property Rights in Data*, 50 *Vand. L. Rev.*51, and J H Barton and K E Maskus, *Economic Perspectives on a Multilateral Agreement on Open Access to Basic Science and Technology*, 20041:3 (Script-Ed), available on line at <http://www.law.ed.ac.uk/ahrb/script-ed/>.

<sup>113</sup> *Supra* note 4 Recital 36.

<sup>114</sup> See also the Green Paper of the Commission of the European Community on copyright and technology challenges, COM (88) 172 def, Brussels, 7 June 1988, p. 207 and the Working Program of the Commission concerning copyrights an neighbouring rights, COM(90)584 def 17 January 1991, Chapter 6, p. 18.

<sup>115</sup> Spoor Verkade Visser, Copyright, 3<sup>rd</sup> ed. 2005, p. 609. See also the following German commentators Dreyer, Kotthoff, Meckel, Urheberrecht, (2004), p. 955, Dreier, *Urheberrechtsgesetz*, Kommentar (2004), p. 1011, and the Danish commentator Jens-L. Gaster, *Der Rechtsschutz von Datenbanken* (Carl Heymanns Verlag KG, Koln, Berlin, Munchen), p. 30.

<sup>116</sup> Spoor Verkade Visser, op. Cit. Note 115, p. 611.

<sup>117</sup> *Ibid.* p. 612.

<sup>118</sup> *Ibid.*, p. 611.

<sup>119</sup> *Ibid.*, p. 617.

<sup>120</sup> *Ibid.*, pp. 613-617, Hugenholtz, *De spin-off theorie uitgesponnen*, AMI 2002, 161, Visser, *The Database right and the spin off theory*, in Snijders en Weatherill (red.) E-commerce law, Kluwer Law International 2003, 105-110 and Derclaye, *Databases sui generis right: should we adopt the spin-off theory?* EIPR 2004, 402-412.

<sup>121</sup> TK 1998-1999, 26 108, nr. 5.

<sup>122</sup> TK 26108, nr. 6, p. 5; see also Spoor Verkade Visser, op. Cit. Note 115, p. 616.

<sup>123</sup> Spoor Verkade Visser, op. Cit. Note 115, p. 620.

<sup>124</sup> ECJ, C-46/02, 9 November 2004, *Fixtures Marketing Ltd tegen Oy Veikkaus Ab*, EHJV C-444/02, *Fixtures Marketing Ltd versus Organismos prognostikon agonon podosfairou AE* (OPAP), ECJ C-338/02, *Fixtures Marketing Ltd versus Svenska Spel AB*, ECJ, C-203/02, , 9 November 2004, *The British Horseracing Board Ltd et al. versus William Hill Organization Ltd*.

<sup>125</sup> ECJ, C-203/02, , 9 November 2004, *The British Horseracing Board Ltd et al. vs. William Hill Organization Ltd*.

<sup>126</sup> *Supra* note 4 Recitals 45 and 46.

<sup>127</sup> Art 2. Lid 1 sub Database-right Act, junctis, artt. 2 lid 1 sub a and b Database-right Act.

<sup>128</sup> E.g. ICSU-paper, *supra* note 112, pp. 2-3 and 7-9.

<sup>129</sup> Accessible at <http://www.nature.com/nature/submit/policies/index.html#6>, date last accessed February 25, 2005.

<sup>130</sup> Art. 1 lid 1 sub c Database-right Act.

<sup>131</sup> Review of the transposition of Directive 96/9/EC into the legislation of the EU Member States pursuant to Article 16.3 of the Directive, performed by Nauta Dutilh, *supra* note 112.

<sup>132</sup> Art. 8 lid 2 Database-right Act.

<sup>133</sup> Spoor Verkade Visser, op. cit. note 115, p. 140.

<sup>134</sup> MvT w.v. 26108 nr 3., p. 19.

<sup>135</sup> Grant Conditions NWO, January 2004, art.3.20.

<sup>136</sup> *Supra* note 4, Recital 39.

<sup>137</sup> MvT, TK 26108, nr. 3, p. 9.

<sup>138</sup> *Supra* note 4, Recital 41.

<sup>139</sup> Explanatory Notes to the Dutch Act, MvT TK 26108, nr. 3 p. 9. See also Verkade and Visser, *Inleiding en Parlementaire Geschiedenis van de Databankenwet*, p. 10.

<sup>140</sup> Explanatory Notes to the Dutch Act MvT, TK 26108, nr. 3, p. 9. For a different opinion see Quaadvlieg, BIE 1998, 403. Spoor Verkade Visser concurs, noting however, that the business risk is a major component, op. cit. note 115p. 625.

<sup>141</sup> Spoor Verkade Visser, op. Cit. note 115, p.625.

<sup>142</sup> Art. 1.9 lid 1 Dutch Act on Academic Education and Research, Netherlands Science Budget 2004, TK 29338, nr.1, p.5.

<sup>143</sup> See [http://www.cordis.lu/lifescihealth/genomics/fp5-projects\\_genos.htm](http://www.cordis.lu/lifescihealth/genomics/fp5-projects_genos.htm)QLG2-CT-2002-01254, date last accessed February 21, 2005.

<sup>144</sup> This project is supported by the European Commission under the programme 'Quality of Life and Management of the Living Resources' of 5th Framework Programme (no. QLG2-CT-2002-01254).

<sup>145</sup> See <http://www.p3gconsortium.org/>, date last accessed May 25, 2005.

<sup>146</sup> Art. 6 lid 3 Database-right Act. The implications of involvement of charity and industry funding is beyond the scope of this article.

<sup>147</sup> Gaster, op. cit. note 115, p.126.

<sup>148</sup> NIH Data Sharing Policy, accessible at [http://grants2.nih.gov/grants/policy/data\\_sharing/](http://grants2.nih.gov/grants/policy/data_sharing/), date last accessed July 10, 2005.

<sup>149</sup> Stanley & Stanley, op. cit. note 74p. 179.

<sup>150</sup> For an inside story of the genesis of the Bermuda principle see *The Common Thread, Science, Politics and the Human Genome*, by Nobelprize-laureate Sir John Sulston, UK project leader of the HGP and author of the Bermuda Principles, p. 161-169.