BILETA
British & Irish Law, Education and Technology Association

20th BILETA Conference:
Over-Commoditised; Over-Centralised; Over-Observed: the New Digital Legal World?

April, 2005, Queen's University of Belfast

Open Science: Open Source Software Licenses and Scientific Research

Andrés Guadamuz
University of Edinburgh

1. Introduction

Recent years have witnessed an increase in the quantity and quality of research dedicated to the subject of the economics of research and development for science and technology,[1] with particular interest paid to the economic study of the impact of intellectual property rights in the fostering of innovation.[2] Intellectual property has generally been considered as one of the most important drivers of new innovation in science and technology because researchers and inventors can recoup their investment in the shape of limited monopolies to their ideas.[3] However, some authors have raised concerns that enhanced intellectual property protection may actually have adverse effects in the development of future research.[4] Basic research had usually not been considered to be subject to protection, and up until recently it was generally offered to the public in the shape of peer-reviewed journals. However, there is a growing trend towards excessive commercialisation and protection of scientific data, usually considered outside the realm of protection, as illustrated in the case of the growing protection of the human genome.[5]

Because access to scientific data has become a requisite of modern research and development (R&D), there is growing concern that the trend towards commercialisation will translate into less available public academic research, which would therefore reduce the overall scientific output. These worries have prompted several studies and reports that attempt to address the problem of the dissemination of academic scientific research.[6] The area of biotechnology has been deemed to be of particular concern, as this area of study has significant economic potential; therefore it has been subjected in recent years to a patenting rush of unprecedented proportions.[7] This phenomenon has prompted the release of genetic information in the public domain, which has also prompted fears of the misuse of the publicly available data by unscrupulous users, who will use this information to close and commodify research through excessively general patents.

These problems have motivated some to call for the devising and utilisation of new ways of protecting basic scientific research from potentially damaging commodification of knowledge.[8] One proposed solution is to use a new intellectual property licensing model that has been successful in software development, generally known as open source software. This system uses intellectual property protection to ensure the wider dissemination of software, by maintaining the copyright protection over a work. The software is then distributed using a licence that allows further copying and redistribution of the work, ensuring that the wider community will have access to the software's source code and allow its modification and dissemination. There are several open source and free software licensing models, but the common denominator in most of them is to allow access to the source code and to allow users to disseminate the code without restrictions.[9]
In particular, it is regarding access to scientific research and innovation that the possibility of translating some of these open source models to the scientific research arena comes into play. The initial application of the open source model has been in the adoption of a scientific publishing model often referred to as open access (OA). The OA movement can be best exemplified by the publication

of scientific outputs and other materials online.[10] These results are offered free of subscription, allowing the wider scientific community access to high-quality content with the click of a button.

However, open access is not enough to ensure access to scientific works because OA generally covers only those materials that are subject to copyright protection, such as journal articles. If scientists want to distribute their works using the open source model, then there would need to be some sort of licence that allows the distribution of patented works, or works contained in scientific databases.

The solution would appear to be a simple matter of translating existing licences to protect patented research, but this has proven to be much harder than previously expected.[11] It is very interesting that while there are new open access and open source licences created every day, an open science licence that protects research through patents and database rights has yet to materialise, despite the obvious enthusiasm from commentators, and extensive political will to generate such a licence.[12]

There are many reasons for the difficulties encountered. Some have pointed out that the open source model does not work best with patented works,[13] because the model appears to be in conflict with the public interest justifications for patents, which imply that inventors are expected to recoup the investment they have incurred. It has also been remarked that the open source model works best with copyright works because they protect creations that are immediately awarded protection, while patented research requires a specific application to the research, making its dissemination through open licences a more difficult endeavour.

The present article tries to respond and contribute to these developments by the examination of the existing scheme to determine the efficacy of the movement and its application to patents. Then, the paper will present a suggestion for a new licensing model for patentable scientific research that allows access and dissemination to diverse fields of endeavour.

2. What is open source?

There is considerable discussion about the different definitions and variations of what is generally understood as open source software, particularly because there is currently a divergence of opinion between different camps in whether one should use open source or free software to define the movements implicit in the permissive distribution of software. This is not the place to solve this dispute,[14] but it should be said that agreeing in the terms is of significant importance to the nascent movement. Suffice it to say, there are different terms that can be used to describe the movement: Free Software (FS), Open Source Software (OSS), Free Open Source Software (FOSS), Free Libre Open Source Software (FLOSS), Open Code,[15] and non-proprietary software.[16] The reason behind the many different terms and definitions is mostly historical, and comes from the fact that each denomination, particularly FS and OSS, have become attached with a specific philosophies and ideologies, and what is more, each of these definitions will usually inform the type of licences used to distribute the work.[17] This work will use the term open source software when talking specifically about the many different licences used in software development.

In its widest possible sense, OSS is used to define a computer programme that allows later modifications by the user or other developers by providing access to its source code through the use of a permissive licence. In this light, non-proprietary software is considered such if it "*is released with a license that would permit others to "fork" the software and release their own modified versions without onerous restrictions, even though the copyright may remain in the hands of a single individual. At least in theory, control has been conceded.*"[18]

Beyond this basic definition, there are some few differences between the other terms, but they are generally referring to some core principles. In the strictest sense, the FS concept is centred on the idea of developing programs and distributing them freely.[19] It is vital to note that the meaning of the word "free" in FS does not mean free as in having no price, but rather free as in "freedom".[20]

Stallman defines free software as having four basic freedoms: the freedom to run the program; the freedom to study how the program works by giving access to the source code; the freedom to redistribute copies; and the freedom to improve and distribute improvements to the public.[21] As understood by the proponents of free software, programmers and other developers can charge for the software if it is their desire to do so, but the same underlying freedom behind the software must exist either if it is acquired for a fee or if it is not. The user must still be able to have all of the freedoms described, with access to the source code as the most basic requisite.[22] The Free Software Foundation (FSF) goes as far as to state that:

> *"The freedom to use a program means the freedom for any kind of person or organization to use it on any kind of computer system, for any kind of overall job, and without being required to communicate subsequently with the developer or any other specific entity."[23]*

This freedom is protected by the adoption of a restrictive licensing model that makes use of existing copyright legislation that guards the source code from proprietary software developers who want to copy it, adapt it and include it in their own programmes. This licensing model is exemplified in the General Public License (GPL).[24]

Open source is closely related to Free Software development, but it does contain a different emphasis about the freedoms involved. The term open source was coined during a strategy meeting in February 1998 in Palo Alto California by a group of software developers with links to the Linux operating system.[25] The need to create a new term to define this viewpoint had become evident because, until then, the prevalent way to describe all output produced by the non-proprietary approach was by using the expression "free software", based mostly on the FS philosophy described. It was apparent to many software developers that this movement had a tarnished reputation in the business world as a result of the more radical ideas held by people linked to the FSF. Open Source then is the opposite of "closed source", the traditional proprietary approach to software development in the commercial world. Closed source is software *"in which the customer gets a sealed block of bits which cannot be examined, modified, or evolved."*[26] The main idea behind open source is to provide software for which the source is available for examination, modification and peer-review. The official definition of OSS came out of the original meeting, and was based on the Debian Free Software Guidelines, a licensing model that accompanies the Debian GNU/Linux system, a Linux distribution.[27] This has generated an Open Source Definition (OSD), which includes a recommended set of clauses that an OSS licence should contain.[28] These licences are exemplified by the Berkley Software Distribution (BSD), the Apache License and the Mozilla Public License (MPL).[29]

3. The Open Science movement

It has become increasingly common to see to the term "open source" used to describe all sorts of fields of study outside of the software arena that gave rise to the concept.[30] The application of this term to other fields could be loosely described as the open licensing movement, which can be defined as the distribution of works protected through intellectual property with the use of permissive "some rights reserved" licences based on the principles reflected in open source licences. Given the varied choice of terminology encountered in software development, the new licensing scheme could also receive different names, such as non-proprietary licences, free licences or commons licences. However, the preferred word to describe this disparate movement seems to be the use of the term "open". Superficially, there seems to be a good argument to choose the term "open source", as it is the one that is more readily identifiable by the public as a description of non-proprietary software models.[31] However, the term "open source" is problematic because the open source paradigm may not translate well into other fields because the original term was used to describe the availability of a computer programme's source code.[32] Therefore, open source should not be used to identify licensing schemes that do not refer to software at all, and where there is no source code to be open. For example, a recent article in The Economist asks:

> *"What does it mean to apply the term "open source" in fields outside software development, which do not use "source code" as a term of art? Depending on the field in question, the analogy with source code may not always be appropriate."[33]*

Despite these misgivings, there would appear to be almost universal agreement about using the word "open" to describe a philosophical the movement that shares the principles and objectives of the two main non-proprietary software camps. The use of these ideals in the area of scientific research presents the birth of a new movement that could be called "open science". This movement could be defined as the application of open source licensing principles and clauses to protect and distribute the fruits of scientific research. This can be done by applying the OSS model to protect other works in areas as varied as biotechnology, biodiversity databases, traditional knowledge and medical research. Non-proprietary and open access models would be an excellent option to maintain a body of technological knowledge that can be shared without fear of misappropriation by commercial interests, facilitating technology transfer to developing countries. This can be understood as open science. According to Maurer:

> *"Open science is variously defined, but tends to connote (a) full, frank, and timely publication of results, (b) absence of intellectual property restrictions, and (c) radically increased pre- and post-publication transparency of data, activities, and deliberations within research groups."[34]*

The suggested definition of open science can be used to cover the many different types of scientific outputs described, but there are two main areas of output that are being discussed in the literature as subject to the potential adoption of open licences. These are the scientific publishing and the scientific output, such as databases and patented inventions. The first is embodied in the open access movement; the second is better exemplified in the so-called open biotechnology movement.

3.1 Open access

The term open access (OA) has become prevalent in the literature in recent years to identify works that are freely available over the internet (using free in the "liberty" sense). These works will generally be distributed by maintaining their copyright – although the term should be generic enough to define works that have been released into the public domain. Open access then will be any work that has been offered under a permissive licence that allows the redistribution of the work.

More specifically, open access has gained some very specific connotations because it is used to refer to academic journals and some forms of academic publication through the use of such licences. This is evidenced by the many different conferences and symposia that have been organised to provide a theoretical framework to OA, which has resulted in the influential Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities,[35] the Budapest Open Access Initiative (BOAI),[36] the Bethesda Statement on Open Access Publishing,[37] and also the European Cultural Heritage Online (ECHO) Charter.[38] Of these, one of the most cited definitions is that of the Berlin Declaration, which defines open access thus:

> *"We define open access as a comprehensive source of human knowledge and cultural heritage that has been approved by the scientific community. […] Open access contributions include original scientific research results, raw data and metadata, source materials, digital representations of pictorial and graphical materials and scholarly multimedia material."*

This definition is very narrow, as it only accepts as open access those works approved by the scientific community, which seems to imply that a form of peer-review is required. In similar fashion, the BOAI defines open access in light of peer-reviewed and scholarly publications, but it allows for the publishing of materials that have not been reviewed for the purpose of comments. The BOAI states that open access cover literature which published in "*peer-reviewed journal articles, but it also includes any unreviewed preprints that they might wish to put online for comment or to alert* *colleagues to important research findings*."[39]

Still, most of the definitions in the declarations tend to be very narrow.[40] Philosopher and open access advocate Peter Suber proposes a more open definition, which states that:

> *"Open access" (OA) is free online access. OA literature is not only free of charge to everyone with an internet connection, but free of most copyright and licensing restrictions. OA literature is barrier-free literature produced by removing the price barriers and permission barriers that block access and limit usage of most conventionally published literature, whether in print or online.[41]*

This definition tends to be more in line with the sharing ethos[42] that gives birth to open source software, and therefore it is more in line with the intellectual and ideological parent of open access. While it must be said that there is an argument to be made in favour of quality control of scholarly research, the movement could benefit from further dissemination of other content offered online.

The growth of open access journals is undeniable. At the moment of writing, the Directory of Open Access Journals (DOAJ)[43] listed 1375 open access journals in all categories of scientific research, with estimates that at least 30 new journals are being added to the DOAJ every month. Figures such as these tend to indicate that this model may very well be the future of academic publishing, particularly for journals.

Initially, one could be suspicious about the academic and economic viability of the model, but this is being disputed by the existing research. Some studies for example, indicate that journals that are available online have wider circulation and are more cited than more prestigious journals. A study of 119,924 conference articles in computer science found that the most cited articles were significantly most likely to come from journals available online than from offline journals by an average 336%.[44] Another study in the United States has found that online journal publishing is economically sustainable under the present system because the revenue obtained by each published article from the publisher is equal to the cost of producing the article, which removes the economic recuperation justification. The study points out that "*The monetary cost of the time that scholars put into the journal business as editors and referees is about as large as the total revenue that publishers derive from sales of the journals*."[45] This statement is corroborated by more recent research that concludes that open source journals are increasing in numbers because they are able to be financed by different types of funding sources, including author fees, conference hosting and the provision of value-added services.[46]

3.2 Open biotechnology

To understand the application of open source to the field of biotechnology, one must understand the race to sequence the human genome.[47] During the 1990s, there were several groups attempting to decode the human genome, but most of the public efforts were brought together in 1996 with the creation of the International Human Genome Sequencing Consortium (the Consortium), a collection of researchers from around the world.[48] These efforts were geared towards the principle of sharing the information obtained by the participants with the common goal of classifying the totality of human genetic sequences, exemplified by the "Bermuda Principles".[49] The Principles clearly specified that the results of the research would be placed in the public domain as soon as possible. The Human Genome Organisation (HUGO) was made responsible for coordinating the data and for using the internet for its release.

Despite the early spirit of sharing exemplified by the Consortium and by the Bermuda Principles, some firms started looking at the possibility of commercialising the results. In 1998 a member of the Consortium created the company Celera Genomics, which set off immediately to finish a sequence of the human genome before the Consortium did.[50] Celera eventually would fail in achieving the full genome first, even though there was growing suspicion that they were using parts of the publicly made material by the Consortium in order to boost their research. In the year 2000, a joint statement by the two participants was made announcing that there was an initial draft of the full human genome,[51] and the results from the public sector have already been made public in several websites.[52] However, in 2001 the contending parties published their respective results at the same time, while the Consortium claimed that Celera had copied their published results, and Celera refuted the claims.[53]

Without analysing these claims, it is clear that the race itself worried some public-sector researchers about the possible abuse of publicly available information that could be used later on to make broad patent claims and commodify the biological data offered.[54] This fear seemed corroborated by facts. By 1999, Celera had applied for the patenting of 6,500 human gene sequences, and by 2000 it had been awarded 300 patents.[55]

The state of affairs in biotechnology patenting generates considerable problems for those involved in this area, as excessive patenting threatens to hinder collaboration and research considerably because it generates an environment that lives in constant fear of litigation. A study conducted in 2002 has found that researchers working in the area of genetics have reduced significantly normal academic collaboration practices due to fears about patent infringement.[56] Similarly, overly broad gene patents could be used to attempt to gain a foothold in the market and stifle competition in the nascent biotechnology industry. Small research centres, educational institutions and individual researchers may find it difficult to conduct research for fear of becoming involved in a patent suit. Moreover, even if a biotechnology patent has been erroneously granted, stakeholders and researchers would still need to get involved in a lengthy procedure to cancel the invalid patent, further stifling research.[57] A decrease in the practice of sharing biotech research could have nefarious consequences to the field, as the exchange of data held in different databases could be hindered.[58]

This is where open biotechnology has been suggested as a possible tool to foster the exchange of research and the transfer of technology amongst researchers all over the world. The general idea behind open biotechnology is to protect the fruits of scientific research by using non-proprietary licences – particularly copyleft ones. The research would be made available to the public online with an attached licence that allows further uses of the material, but forbids the commercialisation of the research by threatening to enforce the intellectual property rights that protect them. This strategy would be compatible with the existing ethos of sharing research that exists in the scientific community. Talking about the possible use of the open source model in the field of bioinformatics, scientist Ewan Birney from the European Bioinformatics Institute commented that "*For us, it's straight scientific principles. If you want to be a scientist, open up your data and open up the code that helps you work with that data*."[59]

The first effort to implement this idea was undertaken by Tim Hubbard of the Sanger Institute in the UK, which was involved in the Human Genome Project. Hubbard became interested in open source and open content licences, until one day he realised that the model could be used to protect human genome research.[60] Although Hubbard drafted a licence, the idea was never implemented by the Sanger Institute because all of the materials were being released into the public domain and could not be licensed. John Sulston, a prominent voice in the genetic research community, has provided some sobering comments about the fact that protecting scientific works intended for public dissemination with a licence is contrary to the ethos behind such undertaking.[61] The idea is to make the works available to the public, not to tie them up in legal battles and complex patent suits.

After this initial disappointment, there have been some few other suggestions about the use of the OSS model to protect the public results of the biotechnology research, although the implementation of such ideas has been minimal.[62] The failure to produce a viable movement is made more evident when the open biotechnology idea is contrasted with the aforementioned open source software and open access models, which have been hugely successful by all measures. In contrast, the implementation of an open biotechnology or open health licence has been slow. There is no shortage of suggestions and positive press about the possibilities of open biotechnology, but these have not produced particularly concrete efforts.[63]

Nevertheless, there is one particular area of open biotechnology that has been successful, and that is the area of bioinformatics. Bioinformatics is the application of information technology to solve biological problems.[64] Bioinformatics projects can use all sorts of software, but it has been increasingly likely to see open source software used.[65] This really means that the success of bioinformatics has more to do with the success of open source software than with the application of open biotechnology.

Another success story in open science is the Center for the Application of Molecular Biology to International Agriculture (CAMBIA),[66] which is an organisation that is attempting to solve many of the problems faced by open science by the judicious use of intellectual property "work-arounds" in the areas of agriculture, food security, biotechnology and the environment.[67] This is done through several IP specific proposals:

- A portal called Biological Innovation for the Open Society (BIOS), which brings together a number of open biotechnology efforts.
- The BIOS patent database includes 1,500,000 life-science patents from the different jurisdictions, which will allow researchers to look for possible patented materials in their area of research, allowing them to avoid costly litigation at a later date.
- CAMBIA has applied and obtained twelve patents of biologic material in different patent offices around the world.[68]
- Bioinformatics tools offered through a research portal called BioForge.[69] This repository will host diverse projects that operate in a similar manner to open source software projects, providing a place to bring together researchers.

This is an encouraging step that will hopefully reverse the relative slow e rate of progress in open biotechnology. This is because there are now some workable licences within the BIOS project. These will be discussed in detail in the next section.

4. The licensing paradigm

One of the problems exposed so far in the open access debate is that it has become clear that there is significant misuse and misunderstanding of the terms and definitions involved. It is common to read terms such as free software, commons, open source and public domain used interchangeably. There must be an understanding that besides the ideological and philosophical connotations of each term, the heart of the movement is the distribution of intellectual works through permissive licences.

The open licensing model is centred on the licences; without them, the movement is just a project management technique. The free software and open source software movements have shown the way to follow regarding licensing agreements. The starting point for non-software licences will be to learn from the experiences in non-proprietary software development.[70] However, there should be an understanding that these licences are just a starting point, as OSS licences tend to be specific to software development, and in many instances they have been drafted by software engineers with little or no intervention of the legal community.[71] Furthermore, the some software developers appear to display considerable reluctance about external intervention in the decision process regarding licensing decisions.[72] This section looks closer at the attempts to translate the OSS ideals to scientific research.

4.1 Open access licences

Open access publishing tends to use "some rights reserved" licences in order to distribute the academic materials involved. These may include the use of standard licences, or in some instances it may include the use of customised licences. At the time of writing, the most prevalent open access licences are one of the many different Creative Commons (CC) licences.[73]

The Creative Commons project attempts to create so-called "*intellectual property conservancies*", [74] separating a block of human knowledge offered for the benefit of the public, but still protected by intellectual property licences. This is analogous to nature conservation areas that exist for the wider social benefit, but have restrictions on certain uses. In the Creative Commons, the goal of intellectual property conservancies is achieved through the offering of a wide variety of licences to protect creative works from misuse. This is done through the application of open source principles, where the work retains its copyright protection, but it is distributed freely[75] as long as the conditions contained in the licence are met. The interesting part of the CC licensing environment is that it empowers users because there is a wide range of licences to choose from. Creators and authors need only to go to website and select from different options offered in some few drop-down menus; the system then chooses the licence that fits the parameters entered. These licences range from offering the work straight to the public domain, to more restrictive licences with restrictions as to the commercial distribution of the work and the use of licences in such distributions.

Creative Commons licences maintain a minimum set of standards that are met by all of their offered legal documents, with the exception of the one that offers the work to the public domain. This could be called the Creative Commons Definition, but it is generally known as the CC Baseline Rights.[76] All CC licences will provide these baseline rights:

- Licensors retain their copyright; this explains why the baseline rights do not apply to public domain offerings.
- The licences announce that fair use rights are not affected by the licence. This is a curious statement, as it should be assumed that any clause that erodes acquired fair use or fair dealing rights should be specified in the licence.
- Licensees will have to obtain specific permission to perform one of the acts restricted by the licence. For example, if the licence does not allow modification or adaptation of a work, this action could only be performed with the permission of the owner. This seems to be a redundant statement, as this is an action that is usually understood in all licences.
- Copyright notices should not be removed from all copies of the work.
- Every copy of the work should maintain a link to the licence.
- Licensees cannot alter any terms of the licence. This seems to be yet another redundant clause, as it should be understood that this is common licensing practice.
- Licensees cannot use technology to restrict access to the work. This baseline right specifically forbids the use of technical protection measures.[77]
- Licensees are granted the right to copy, distribute, display, digitally perform and make verbatim copies of the work into another format.
- The licences have worldwide application, have lasts for the entire duration of copyright (unless otherwise specified), and are irrevocable.

It is important to note that the baseline definition of CC licences does not mention anything about modification or adaptation of a work; does not deal with copyleft-like clauses requiring the use of similar licences to distribute the work; does not mention attribution; and does not deal with the distribution of copies for commercial purposes.

This makes the basic Creative Commons definition more alike to the open source ideals than to the free software principles exemplified by the GPL.[78] Nevertheless, creators can choose a CC licence that maintains all of the restrictions mentioned, from all of the options offered. Authors then can choose from the following options to generate their licence:

- **Attribution:** The work is made available to the public with the baseline rights, but only if the author receives proper credit.
- **Non-commercial:** The work can be copied, displayed and distributed by the public, but only

if these actions are for non-commercial purposes.

- **No derivative works:** This licence grants baseline rights, but it does not allow derivative works to be created from the original.
- **Share-Alike:** This is based on copyleft principles. Derivative works can be created and distributed based on the original, but only if the same type of licence is used, which generates a viral licence.[79]

It is possible to have licences that combine different of these options. The strongest (and most popular) CC licence is the Attribution-NonCommercial-ShareAlike License,[80] which is the licence that most resembles the strongest copyleft software ones (such as the GPL). All CC licences are presented in three formats: the first is a short and easy to read "Commons Deed", which explains the terms and conditions of the licence in a simple manner; the second format is the "Legal Code", which is the full licence; the third is the "Digital Code", which provides an HTML version of the licence[81] that can be read by search engines and makes it easier to list the content in the Creative Commons directory.
Creative Commons presents a very positive step towards the wider distribution of non-proprietary technology. It is innovative, thoroughly planned and greatly implemented. CC delivers open access licences in the digital domain with scalability, adaptability and ease of use for those unfamiliar with the legal issues involved in licensing. CC also offers jurisdiction-specific versions of their licences to make them more valid in an international environment and to respond to legal requirements in a given country.

The other major open content licence is the GNU Free Documentation License (GFDL),[82] which is the FSF's non-software licence, and it is generally used to protect manuals and other literature related to the FS movement. However, the GFDL is also used in other open access projects, such as the free online encyclopaedia Wikipedia. The GFDL is could be classified as an open access licence because it allows the copying, distribution and adaptation of a work provided the author complies with the conditions included. These can be found in section 2 of the licence, which states:

> *You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License.*

This is an important point of the licence, because it specifies that it allows for the commercial use of the works. The GFDL allows for the modification and translation of the work, provided some specific sections are maintained or deleted; and all derivative works must be licensed using the GFDL.[83] This clearly means that this is a copyleft licence, perpetuating itself through this viral clause. The viral nature of the licence exists in section 4, which states that *"[y]ou may copy and distribute a Modified Version of the Document […] provided that you release the Modified Version under precisely this License…*" This is different to the share-alike element in CC licences, as these require only that the work is released with a licence that contains similar clauses and rights.

The viral nature of the GFDL can be seen in practice through the wide copying and dissemination of Wikipedia articles, which are being used by many other open content providers. For example, Wikipedia articles are now used in many other content providers, such as The Free Dictionary.[84] The articles found in this online resource have to be licensed through the GFDL, allowing yet another third party to copy them and use them in their website, provided that they use the GFDL.

With so many creative works that may be subject to be protected by open access licences, it should come as no surprise that there has been a recent proliferation of licences that allow commercial and non-commercial content creators to adopt the non-proprietary open access model. One such project is the OpenContent Licence (OPL), a collaborative effort that sets a copyleft licence, ensuring that shared works will continue to remain free to subsequent users.[85] Similar efforts also include music creation via the Open Audio Licence (OAL)[86] the SCRIPT-ed Open Licence (SOL),[87] and even Open Cola, the world's first copyleft fizzy drink.[88]

Looking at the vibrant nature of the open access movement, it is clear that there are enough reasons to believe that the publication of scientific materials in journals or through other online means that at least this part of the open science movement is doing quite well through the creation of growing body of work that is easily accessible to researchers around the world. However, what happens with scientific research that is not subject to publication? Can scientific databases, archives, repositories and patented research be protected through open licences?

4.2 Open science licences

All of the aforementioned licences have one thing in common; they protect only works that are subject to copyright protection. This is valuable when one considers that a significant amount of scientific works are protected by copyright, particularly academic journal articles and other literary works. This raises the question as to whether there can be open licences that protect other types of work in commercially viable fields like biotechnology and health research – which are usually available through databases or patents. This has proven to be more difficult because, although there have been many scientists and researchers advocating the implementation of open licensing models to the scientific arena, it is difficult to find a scientific equivalent to the GFDL or the Creative Commons licences that apply to those fields.

The reason for this is that the open licence model works best with copyright than with patents or databases. There are two main reasons for this fact. Firstly, copyright subsists in an original work as soon as it is fixed in tangible form.[89] Secondly, copyright does not require any sort of registration to initiate protection, which means that copyright "*flows from the nib of a pen*",[90] making it much easier and cheaper to distribute through an open licence as soon as it is originated. On the other hand, works that require some sort of registration to be subject to protection – such as patentable scientific research – will be more difficult to distribute through an open licence, as there are several steps that are required to be able to distribute it as an with some sort of permissible licensing model. Although the specific difficulties to provide a software solution will be discussed in more detail later, there has been some success in applying some limited open science solutions for databases and patented works. These efforts are listed next.

4.2.1 Database licences

The importance of scientific databases for research is an issue that has been well explored in the literature.[91] There can be little doubt that in the information age, access to the vast amount of scientific data stored in databases is of utmost importance for researchers around the world. Access to a large number of databases is offered for a fee by a vast array of service providers and institutions.[92] The growth in the number and the economic importance of scientific databases has been accompanied by increased concerns about the reuse of the data to provide further works of commercial value. A report by the National Research Council in the United States points out that:

> *Currently many for-profit and not-for-profit database producers are concerned about the possibility that significant portions of their databases will be copied or used in substantial part by others to create "new" derivative databases. If an identical or substantially similar database is then either redisseminated broadly or sold and used in direct competition with the original rights holder's database, the rights holder's revenues will be undermined, or in extreme cases, the rights holder will be put out of business. Besides being unfair to the rights holder, this actual or potential loss of revenue may create a disincentive to produce and then maintain databases, thus reducing the number of databases available to others.[93]*

This is of particular worry for those who are releasing genetic data into the public domain, as described earlier. With publicly available databases, commercial providers would find large sections of readily available information that can be repackaged and resold as part of a commercial database. This possibility is precisely what has prompted the calls to protect databases through open source licences.[94]

Despite the suggested application of the open source licences and ideals to databases, the actual application has not been met with the enthusiasm that it deserves, which can be explained by two main reasons. Firstly, most non-commercial information – particularly in the field of biotechnology – most of the information is released into the public domain.[95] This type of release is extremely useful for future researchers, but it does little to curb the further commercialisation of the data.

Secondly, the legal protection of databases is a subject that is not fully harmonised at the international level, while different jurisdictions apply a wide range of legal figures and levels of protection to this type of intellectual work. For example, the United States[96] has been struggling with the application of copyright law to the subject of databases by extending the definitions of originality of a work. Earlier cases declared that the mere rearrangement of information was not enough to prove originality.[97] Despite the originality requirement, there are circumstances where the courts will award copyright protection to databases due to the fact that there is enough originality in what is done to the data.[98] Europe has followed a different path by embracing a "sweat of the brow" approach, where the work and investment that goes into the gathering and arranging of the data is rewarded, even if the data itself is not original.[99] This is particularly evident with the European Directive on the legal protection of databases.[100] The Directive awards a *sui generis* right to databases in which there has been a quantitative and qualitative investment in obtaining or verifying the data contained.[101] However, the picture is made more complex by the recent rulings from the European Court of Justice,[102] which has eroded the database right considerably.

It is precisely this complicated legal landscape that makes the possible application of open source software models so difficult for databases. It would seem possible that providers of scientific data contained in a database compiled in a country that provides for the copyright protection to databases may be able to issue their work through an open access licence, perhaps even a Creative Commons licence. This would be possible because in most jurisdictions databases are protected as literary works.[103] However, those providers would first have to be able to be awarded protection in the first place, which is not always the case, as evidenced by the originality standards prevalent in the United States. In countries with a *sui generis* right, the licensing would have to meet with the requirements of the recent cases and the directive, which is not an easy task. According to Waelde and McGinley:

> *Suffice it to say many questions remain over the extent to which scientific databases might qualify for the sui generis right. Whereas at first blush it might have appeared that many might fall outwith the necessary criteria, […] it is far too early to argue that the contents of scientific databases fall into the public domain as a result of the ruling, however much that might benefit scientists and the progress of science.[104]*

Taking into consideration all of these difficulties, it should come as no surprise that a large amount of online scientific data is released into the public domain. Nevertheless, there are indications that the solution may not be found in database protection, but in contractual law. This is best evidenced by the International HapMap Project Public Access License (HPPAL),[105] which is part of the HapMap genetic database project. Unlike all other open licences, the HPPAL does not assign any intellectual property rights; it is an end-user agreement. The data can only be accessed through following a registration process, where the user is required to agree to terms and conditions before gaining access to the certain parts of the HapMap genetic database.

The wording of the HPPAL makes it appear to be an intellectual property assignment of rights (it is after all, called a licence), but it is not entirely clear what rights are held over the data that is being offered.[106] The HPPAL is very careful not to assign intellectual property rights, so it must be assumed that this is just a user agreement where the author enters into the obligation to comply with the terms and conditions set out in the document. Specifically, paragraph 3 of the licence states that: "*You may not access, copy, modify, sublicense, distribute or otherwise use the Genotype Database or the data contained in it except as expressly provided under this License*."

The most relevant part of the HapMap licence is with regards to future patent applications. Paragraph 2(b) of the licence does not allow the patenting of genetic information from the database, with the exception of particular uses of sequences, provided that the patent allows further use of the information obtained from the database. The paragraph reads:

> *…you shall not file any patent applications that contain claims to particular uses of any SNP, genotype or haplotype data obtained from the Genotype Database or any SNP, haplotype or haplotype block based on data obtained from, the Genotype Database, unless such claims do not restrict, or are licensed on such terms that that they do not restrict, the ability of others to use at no cost the Genotype Database or the data that it contains for other purposes;*

This is an attempt to provide a viral or share-alike element to the agreement, as it the freedoms protected by this licence are protected in the future licensing of patented material. The HapMap licence offers an ingenious way of getting around the problems of database protection enumerated above, as it relies on contractual obligations rather than on intellectual property protection, and may prove to be the way to go as far as database licences are concerned.

4.2.2 Patent licences

If the application of open licensing to scientific databases has been minimal, the porting of OSS licences to patented research has been almost inexistent and doubly problematic for reasons that will be explored in more depth in the next section. Nevertheless, there have been a handful of attempts to provide a workable licence for patented material, including a recent draft licence.

One of the most promising efforts to provide a licence has been put forward by the Creative Commons project. Because Creative Commons licences are geared specifically towards creative works subject to copyright protection, a new concept has been designed to accommodate scientific research. This concept is the Science Commons project, [107] which has been created to generate licences that will deal with intellectual property works that are not covered by existing CC licences. The project is ongoing at the time of writing, and it has yet to produce a licence draft available to the public.

In a more thorough effort, a paper by Hubbard and Love explores some alternative models of pharmaceutical research and development to produce new medicines.[108] Their proposal uses the existence of free software as an illustration that alternative business models are viable, but unfortunately it fails to make the point of how to translate OSS licensing ideals into the pharmaceutical industry. Although Hubbard and Love's argument may not connect directly with open access models, their suggestion is important because they propose workable ways to fund the basic research and to generate incentives to companies to distribute their intellectual property to the public, which would be released through open licences. But the question remains, which open licences?

The answer to the question of licences may lie in the aforementioned CAMBIA project. One of the most important parts of the objectives of CAMBIA is the use of "open source" ideals to generate a protected commons for researchers in the life sciences. This is done through the CAMBIA BIOS Licence (BIOS Licence). According to the project, the licence works like this:

> *In lieu of royalties, the legally binding conditions that BIOS licensees must agree to, in order to obtain a license, are that improvements are shared, and that licensees cannot appropriate the fundamental "kernel" of the technology and improvements exclusively for themselves. Licensees may obtain access to improvements and other information, such as regulatory and*

> *biosafety data, shared by other licensees in a confidential "protected commons". They cannot prevent other licensees from using the technology in the development of different products. [109]*

The core concept of the BIOS Licence is that it is going to be used to provide patented research with a permissive licence that operates with OSS and open access principles. However, this is easier said than done, and the final result demonstrates just how difficult it is to word an open source patent licence.[110]

An important point in the BIOS Licence is that it covers some specific patented technologies, defined as "*Crop Molecular Enabling Technologies and associated patents, patent applications, knowhow, data, materials, and business, technical, economical and manufacturing information*". [111] This specifies that the licence covers the field of biotechnology, but there doesn't seem to be any reason why this definition should not be broadened to include other types of patented technologies. This technology can be owned by the licensor, who retains control over it. There is also a possibility that the licensor could be a licensee himself. The reason for this is because the BIOS Licence contains a viral clause that allow licensees to sub-licence the material, as long as the same rights that are contained in the licence are preserved in the vertical agreement. Paragraph 2.5 reads:

> ***2.5** Licensor also grants to Licensee, so long as it retains rights under this Agreement, the right to issue sublicenses to third parties (Sublicensees) under the rights and licenses granted to Licensee in Paragraph 2.1 (with no right of the Sublicensees to sublicense such rights to others).#*

Paragraph 2.1 is the licence grant, which gives licensees a non-exclusive, royalty-free right to make and use the technology in order to develop it, improve it or sell licensed products from it. This means that improvements to the patented technology are allowed as long as those are communicated to licensor within sixty days. When compared to existing open source and open access licences, the elements and clauses present in the BIOS Licence make it akin to a CC Attribution-Commercial-ShareAlike licence.

This is a worthwhile effort to create a viable "open source" licence of patented materials. However, even in its draft stages it is easy to see that the language seems stretched and unclear in many instances – something that could turn away some potential licensors who could find the complex explanation of the terms and conditions difficult to navigate. Another question that arises from the draft is it is not very clear what would be the role of the initial licensor and his place in a lengthy chain of sub-licensees. Paragraph 2.5 states that "*For the purposes of this Agreement, the operations of all Sublicensees shall be deemed to be the operations of the Licensee, for which the Licensee shall be responsible*."

Does this really mean that the sub-licensee has the same rights and obligations as the licensor? What about the obligation to notify the licensor of any improvements? It would seem that this would fall into the obligations of the sub-licensee. Is the sub-licensee in any obligation to notify the original licensor?

It has been pointed out that the BIOS Licence is a work in progress and requires more fine tuning to be ready for consumption. Nevertheless, some of these problems cited serve to demonstrate the difficulties of translating open source into patents. The reasons for this will be discussed next.

4.2.3 Trouble with patents

The difficulties with the one existing patent licence and the lack of other open science licences is that there appears to be an inherent problem in porting a licensing model that has been designed to work with copyright into a system that would have to work with patents.

There can be little doubt that patents offer the strongest short-term protection of technologies emanating from costly scientific research.[112] It has several advantages in order to protect certain technologies, particularly because some types of research may produce outputs that would not be suitable for copyright protection. While this is precisely how the commercial world operates, those interested in making their work available through the public under some sort of open licence model will have to do it by protecting their work, as the entire system is based on the threat of infringement suits brought against those who had not shared the work following the required clauses.
The problem then for some institution wanting to do this is that they will have to obtain a patent in order to licence it, and this can prove to be an expensive endeavour. Some studies estimate that an average biotechnology patent application could cost an average $7,500 USD in the United States alone.[113] Because patents must be applied separately in each jurisdiction where they will be exploited, the costs for a small research institution could be prohibitive. Even when the patent has been obtained, the enforcement of patents is where the costs are steeper. The cost of defending a patent in the United States where the dispute is less than a million USD can range from $300,000 to $750,000 USD. This means that even if a research institution obtains a patent to protect their research, the right holders would find it extremely expensive to defend their intellectual property against misuse – particularly considering that those likely to use open source licences may be small research institutions, or even to individual researchers.[114] The problem would be more pronounced for researchers in developing countries, as they would possibly have to enforce patents abroad.

The sheer costs involved could be enough to dissuade small and medium research facilities to stay clear of the potential liabilities involved with the patent system, and continue releasing information through more traditional means. Nevertheless, there may be a viable solution for the problem of the enforcement of patents held by individual organisations. The problem of enforceability of OSS licences is similar to what has been described in the previous paragraphs. In software, many small software developers do not have the resources to enforce their copyrights. For that purpose, the FSF recommends that all those programmers using their licences should assign copyright of their works to the FSF because in that way they can enforce the licence better in case of infringement.[115] This scheme could be replicated in open science licences, and may be a role of NGOs like CAMBIA, who are pushing for this model. Therefore, collective organisations could be in charge of the enforcement of research held by individuals.

Another possible problem for the use of open licences of patented technology is that it could be argued that open licences are incompatible with prevalent patent policy goals. An often stated goal of a patent system is to encourage the distribution of inventions through the utilitarian justification that allows for the economic reward.[116] An open licence model might clash with this objective because it would stop inventors from being able to economically recuperate investments in future research related to the patented one, particularly if a non-commercial viral licence is used. For example, imagine a patented gene sequence that has been licensed through a copyleft licence containing non-commercial viral restrictions, Researchers who would want to patent improvements based on the licensed sequence and then exploit them commercially may find that they would be in breach of their licences, as they would not be able to sub-license their work commercially because of the viral restrictions. But what if those researchers incurred in considerable expense to produce the improvements? They would have to license their research using a non-commercial clause as well, which would defeat the utilitarian justification for patents. The BIOS Licence does not appear to have this problem because it allows commercial use of the patented technology, so there may be a case that there cannot be a non-commercial open licence for patents.

The apparent incompatibility of the patent system with open source is well known in the groups that advocate for its implementation. The statement of purpose of the Science Commons project enumerates some of the problems faced by the project in trying to translate the CC ideals into working licences for works that rely on patent protection.[117] The Science Commons proposal goes as far as to point out that "*Many of the things that we have learned in forming the Creative Commons* *do not translate completely to the world of science policy. We dealt primarily with copyright - here the issues would also involve patent and trade secret*."[118]

The potential incompatibility between patents and open source licences is difficult to solve. Even with the early draft of the CAMBIA Licence, the nature of the patent system seem to offer insurmountable obstacles to the possible adoption of a viable open science licence dealing with patented technology. Those industries that are willing to incur in the cost of expensive research will want to see their efforts rewarded, and a large patent portfolio will give research-heavy institutions an excellent bargaining position when dealing with other competitors within the industry. The race to decode the human genome has served to demonstrate that there are significant economic interests at play, and this is a fact that will not go away overnight. This is why the best option for smaller research facilities and public interest oriented institutions is to disseminate works through releases into the public domain. This type of dissemination has the effect of widening the accessibility of the research by other small industries. Another effect of the release of materials into the public domain would be to pre-empt future patent applications, because the research has been already made public. Eisenberg explains this tactic:

> *In addition to making it difficult for publicly-funded investigators and their institutions to file timely applications for patents, the Bermuda rules also lead to the prompt creation of "prior art" that could potentially defeat patent claims based on similar DNA-sequencing efforts in the private sector. No one can get a patent on something that was already publicly disclosed before the patent claimant discovered it.[119]*

Despite this seemingly watertight solution, there is still a real potential that the information that has been made available for free could be copied and then used to make patent applications about that same material. The chaotic state of patent application in areas such as software and biotechnology – particularly in the United States[120] – provides a warning that patent offices cannot be trusted in identifying whether a patent application is innovative, or if it is based on significant prior art.

It would be fair to assume that there are too many problems, and it would be fair to suggest that perhaps open science should be scrapped – at least in open science for patents. The lack of licences at the moment makes the possible implementation difficult, while the few efforts that have been proposed still seem to fall short. Even the strongest proponents of open science and open biotechnology recognise that the movement cannot go forward without viable licences. Janet Hope comments that:

> *Key issues for advancing the open source biotechnology analysis will be developing open source patent licences and other licences appropriate for biotechnological subject matter, assessing the importance of higher capital costs in biotechnology development and establishing whether or not there exist secondary markets for biotechnology services or other commercial offerings that might support business models along the lines that have proved successful in the software context.[121]*

It is perhaps a time to look at different options. Those who believe in ensuring wider access to scientific research and technology should not be daunted by the difficulties encountered, as there may be other solutions that can provide a viable manner to harness the creative and development strengths of the open source model with other ways of dissemination.

5. Specific proposals

Apart from the CAMBIA Licence, the tackling of the patenting problem is short in suggestions, as most of the proponents of the open science and open biotechnology solutions for scientific research usually fail to even tackle the question of the potential problems presented by patents.[122] The solution to the problem may not be in the drafting of new complicated licences, but in looking elsewhere for inspiration as to new manners of allowing wider access to patented technologies.

In this line, Cukier has suggested that this is not an issue of licences, but rather suggests that the patenting problems could be overcome through the use of policy by applying existing national interest patent defences that are already in use in the United States in the area of defence and health. [123] He comments that:

> *… US-funded research enables the government to use the resulting technology on a royalty free basis. In the case of the Bayh-Dole Act, the government has "march-in" rights to take control of a patent it does not believe it being sufficiently exploited. More broadly, the US and its contractors can't be prohibited from using patented technology as a matter of law… [124]*

Another novel solution would be to continue using the tried and tested open access and open source software licences, but to change the clauses to read more generically. For example, instead of using “copyright”, the licences could use “intellectual property”, which would cover patents. However, this seems like an ad hoc patch that fails to provide a real answer to the problems highlighted. This could be solved by using existing software licences that mention patents. There is one such licence: the Apache Licence (version 2.0),[125] which contains a patent assignment clause that reads:

> *Subject to the terms and conditions of this License, each Contributor hereby grants to You a perpetual, worldwide, non-exclusive, no-charge, royalty-free, irrevocable (except as stated in this section) patent license to make, have made, use, offer to sell, sell, import, and otherwise transfer the Work, where such license applies only to those patent claims licensable by such Contributor that are necessarily infringed by their Contribution(s) alone or by combination of their Contribution(s) with the Work to which such Contribution(s) was submitted. If You institute patent litigation against any entity (including a cross-claim or counterclaim in a lawsuit) alleging that the Work or a Contribution incorporated within the Work constitutes direct or contributory patent infringement, then any patent licenses granted to You under this License for that Work shall terminate as of the date such litigation is filed.*

This seems like a viable possibility, as Apache is the dominant web server around the world, where 70% of all websites on the net served by Apache software.[126] The data suggests that a direct translation of the Apache licence to the realm of patented technology would be possible. However, there should be a cautionary word regarding the Apache Licence, and it is the fact that it is not the predominant open source licence. Out of more than 64,000 open source projects listed in the SourceForge portal, only 344 use this licence.[127]

Recent developments have suggested that there may be another way, and that strict licences are not needed to provide a common pool of accessible scientific data and technology. IBM has made the headlines of every major technology-related publication by stating that it will not enforce 500 software patents that it owns if they are used by open source software projects.[128] This unprecedented move has been achieved through a clever use of contract law. IBM has published a legally-binding promise not to enforce a number of their patents to those software projects that are released to the public through a licence approved by the Open Source Institute.[129] This element of IBM’s pledge is very important, as it gives a tight definition of what will be an open source project. The definition reads:

> *Open Source Software is any computer software program whose source code is published and available for inspection and use by anyone, and is made available under a license agreement that permits recipients to copy, modify and distribute the program’s source code without payment of fees or royalties. All licenses certified by opensource.org and listed on their website as of 01/11/2005 are Open Source Software licenses for the purpose of this pledge.[130]*

The document goes on to promise that IBM will not assert any of the listed patents in the united

States, or its counterparts worldwide, against open source projects, defined as above.[131] The document ends with a list of the 500 patents. This announcement should be met with some scepticism, as IBM has a considerable software patent portfolio, and was awarded more than 3,000 patents in 2004 alone.[132] One should also be sceptical about the possible legal validity of such promise.

The main question about the validity of the pledge is centred on the question of its classification within contractual law figures. In this document, IBM is making a unilateral promise that stands on the assumption that it can be met by those who qualify as an open source developer. This promise does not require an obligation per se, it simply promises not sue a group of people that fulfil certain characteristics. In this manner, it is not so different than a retailer that promises to give free CDs to those who bring them a coupon. The issue of unilateral promises is an area of the law that varies from one jurisdiction to another. In some Common Law systems, the question of unilateral promises has often been dealt with as an issue of contract formation and consideration.[133] However, the landmark case of *Carlill v. Carbollic Smoke Company[134]* establishes that a unilateral promise that is accepted through the performance of an act is valid. Scotland does not have a problem with acceptance; therefore unilateral promises are much less of an issue and have to be considered generally valid.[135] Other European countries have different rules for the acceptance of unilateral promises,[136] but countries like Germany[137] and France[138] allow some models of obligations arising from unilateral promises. This tends to give strength to the validity of IBM's promise.

IBM's non-enforcement promise is a very practical and seemingly valid solution that can be applied to all other sorts of patents, and it could prove to be an effective tool to solve the problematic application of OSS models to patented technologies. This could work for individual scientists or research institutions that are interested in maintaining their intellectual property, but that want to allow access to their patented material to specific recipients. These institutions could publish their own promise not to assert their patent portfolio, or a selected list of patents, as long as the users fall into a specified category of beneficiaries. It is important that the patent owner identifies clearly the intended users of the technology, and defines it unequivocally in the document. A promise that allows use in developing countries would be useless, unless it is accompanied by a clear definition of what constitutes a developing country, and what sort of users in those countries would benefit from the pledge. A possible clause could read thus "*This pledge will benefit researchers based in a Least Developed Country as defined by the United Nations Conference for Trade and Developmen*t"; or "*For the purpose of this promise, developing country will be defined as a country that is listed as having low human development in the 2004 Human Development Report by the United Nations Development Programme*."

Another important element to add is that the patent owner could generate a web form where users that fall into the definition could register as such, which would have the added benefit that the owner would have a better idea of who is using the technology.
One of the main advantages of the use of a unilateral promise is that it helps to focus the access to scientific research to those who the patent owner would not consider to be a commercial threat or potential competition, which would erase some of the concerns about the possible incompatibility of open source models with the expenses and commercial value of research. This solution is not a licensing scheme; therefore it eliminates some of the more complex contractual chains of distribution that can be found in viral contracts. Researchers could also gain in the knowledge that there will be a certain amount of knowledge that can be used without fear of infringement.

6. Draft promise[139]

Preamble
[This space can be used to indicate the purpose and the rationale behind the promise]
Definitions
[This space will contain a series of strict definitions of the beneficiaries. Some examples are provided]

"Technology": the list of patents included in the Annex.
"Owner": The patent owner, [INSERT NAME].
"Beneficiary": Any organisation that is a listed participant of BioForge projects as of [INSERT DATE].

Promise

The Owner hereby promises not to enforce any of the listed Technology against Beneficiaries that have registered their intent to use the Technology at the following address [INSERT FORM ADDRESS].

Warranty

The Owner certifies that the Technology is owned by him/her and that [or that it has specific permission to issue the promise]. The Owner also certifies that the Technology is not subject to litigation as of {INSERT DATE]. The Owner presents the Technology "as is", and makes no warranty as to the accuracy of the information contained in the patent application.

Limitation of Liability

Subject to any liability which may not be excluded or limited by law, the Owner will not be held liable for incidental, consequential or indirect loss or damage howsoever and whenever caused to the Beneficiary.

Annex

[Table of patents, listing patent number, awarding office, beneficiary and title (or brief description)]

7. Conclusion

The issue of access to scientific research is becoming one of the most important issues of our time. The direction of the flow of knowledge rests greatly on the problem of the ownership of the technology. One of the grandest ideas in recent years is the use of intellectual property tools to protect certain parts of human knowledge, something that is managing to generate shared knowledge, a common pool of technology and research that can be accessed by all; a common space where the information flows with fewer restrictions than in an entirely proprietary model.

This common space has already been experimented and explored within the free software and open source software communities. The non-proprietary software experiment has demonstrated that open development models are viable and sometimes even commercially successful. Amongst these models, one of the most interesting licences is that offered by so-called copyleft licenses, those licences that allow software to be transferred with the insurance that the source code will remain open, with the caveat that anyone who redistributes the software, with or without changes, must pass along the freedom to further copy and change it.

However, software development is not the only area in which this licensing model could be applied. The viral nature of copyleft licenses has generated a considerable amount of interest in circles that transcend software development. The idea of sharing materials is not new, and has been made more evident by the chaotic and sometimes anarchic nature of the internet. However, shared materials tend to suffer from the possibility of third parties that use the freely acquired information to turn them into proprietary works. That is why many different organisations are turning to the copyleft model to protect works that are being freely shared online.

This article has explored the application of non-proprietary software licences to scientific research – particularly academic publications, scientific databases and patented technologies. To do this, several different licensing models have been explored. It is clear that copyright materials are well suited for this experiment, and the area of scientific publications shows special promises for the future. Unfortunately, other types of scientific outputs present more challenges to those involved. Costly research and development have produced entire fields of study that are not suitable to adopt the open source ideals. Although the trailblazing efforts of CAMBIA, HapMap and the BIOS Licence must be applauded and recognised, the author feels that there is much to be done to ensure access to expensive technology to the widest possible audience. IBM's unilateral promise gives researchers a model to emulate to ensure this objective. It is fasible to apply this document to almost all investigation efforts that result in a patent.

The author recognises that this is just a draft proposal, but it is hoped that others can continue to add to this idea if it is found to be useful. In the best spirit of the Bazaar, and paraphrasing Linus Law, given enough eyeballs, all licence bugs are shallow.

---

[1] For example, see: Feldman M.P., Link A.N. and Siegel D.S. *The Economics of Science and Technology: An Overview of Initiatives to Foster Innovation, Entrepreneurship, and Economic Growth,* London: Kluwer Academic (2002).

[2] See: Granstrand O. *The Economics and Management of Intellectual Property : Towards Intellectual Capitalism,* Cheltenham: Edward Elgar (1999); and Maskus K.E. *Intellectual Property Rights in the Global Economy*, Washington, D.C.: Institute for International Economics (2000).

[3] Particularly since Mansfield E, "Patents and Innovation: An Empirical Study", 32 *Management Science* 175 (1986).

[4] Bunk S. "Researchers Feel Threatened by Disease Gene Patents", 13 (20) *The Scientist* 7 (1999); and Reichman J.H. and Ulhir P.F. "Database Protection at the Crossroads: Recent Developments and Their Impact on Science and Technology", 14(2) Berkeley *Technology Law Journal* (1999).

[5] Sulston J. "Intellectual Property and the Human Genome", in Drahos P. and Mayne R. (eds) *Global Intellectual Property Rights: Knowledge, Access and Development,* Basingstoke: Palgrave Macmillan (2002) pp.61-73.

[6] *An Economic Analysis of Scientific Research Publishing,* Wellcome Trust, (2003); and *Scientific Publications: Free for All?* House of Commons Science and Technology Committee, HC 399-I (2004).

[7] Dutfield G. *Intellectual Property Rights and the Life Science Industries: A 20th Century History,* Aldershot: Ashgate (2003).

[8] Maurer S. "New Institutions for Doing Science: From Databases to Open Source Biology", *European Policy for Intellectual Property Conference on Copyright and database protection, patents and research tools, and other challenges to the intellectual property system*, University of Maastricht (November 24-25 2003).

[9] Guadamuz A. "Viral Contracts or Unenforceable Documents? Contractual Validity of Copyleft Licenses", 26 (8) *European Intellectual Property Review* 331 (2004).

[10] *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*, 20-22 October 2003, Berlin. @: <http://www.zim.mpg.de/openaccess-berlin/berlindeclaration.html>

[11] Cukier K. "Open Source Biotech: Can a Non-Proprietary Approach to Intellectual Property Work in the Life Sciences?" 1 (3) *The Acumen Journal of Life Sciences* (2003) @: <http://www.cukier.com/writings/opensourcebiotech.html>

[12] Burk D. "Open Source Genomics", 8 *Boston University Journal of Science and Technology Law* 254 (2002).

[13] Cooper Feldman R. "The Open Source Biotechnology Movement: Is It Patent Misuse?" *SSRN Working paper* (2004) @: <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=545082>

[14] To see one of the arguments in the debate, see: Stallman R. *Why “Free Software”' is better than `”Open Source”,* (1998). @: <http://www.fsf.org/philosophy/free-software-for-freedom.html>

[15] This is the term preferred by Lessig to avoid the FS/OSS debate. See: Lessig L. *Code and Other Laws of Cyberspace*, New York: Basic Books (2000), p.7.

[16] This term is the term preferred by the author. See: Guadamuz, supra note 9, at p.332.

[17] Rosen L.E. *Open Source Licensing : Software Freedom and Intellectual Property Law,* 1st ed, Upper Saddle River, N.J.: Prentice Hall PTR (2004), pp.51-69.

[18] “Proprietary Software*”, Wikipedia*, 2002. @: <http://www.wikipedia.com/wiki/proprietary+software>

[19] Stanco T. *We are the New Guardians of the World*, (2001). @: <http://lwn.net/daily/guardians.php3>

[20] Or as it is often stated in OS and FS circles, free must be understood as in freedom, not as in beer.

[21] Stallman R. *The Free Software Definition*, (1996). @: <http://www.fsf.org/philosophy/free-sw.html>

[22] Stallman R. *Selling Free Software*, (1996). @: <http://www.fsf.org/philosophy/selling.html>

[23] Stallman, supra note 21.

[24] The text of the GPL can be found here: <http://www.gnu.org/copyleft/gpl.html>

[25] Open Source Initiative, *History of the OSI*, (2001) @: <http://www.opensource.org/docs/history.html>

[26] Raymond E. *Keeping an open mind*, (1999). @: <http://tuxedo.org/~esr/writings/openmind.html>

[27] The guidelines can be found here: <http://www.debian.org/social_contract.html#guidelines>

[28] The OSD can be found here: <http://www.opensource.org/docs/definition.php>

[29] A list of approved licences can be found here: <http://www.opensource.org/licenses/>

[30] See for example: Boettiger S. and Burk D. "Open Source Patenting", 1 (6) *Journal of International Biotechnology Law* 11 (2004); and Burk, supra note 12.

[31] As a measure of the prevalence of open source over other terms, Google throws 337,000 results for "non-proprietary", 19,000,000 results for "free software"; and 23,600,000 for "open source".

[32] Source code is the programming statements in a programming language that exists before the programme is compiled into an executable application. The executable form of the software is generally known as the object code, and can only be read by the machine.

[33] "An open-source shot in the arm?" *The Economist*, June 10, 2004. @: <http://www.economist.com/displaystory.cfm?story_id=2724420>

[34] Maurer, supra note 8.

[35] Full text of the declaration can be found here: <http://www.zim.mpg.de/openaccess-berlin/berlindeclaration.html>

[36] More about the initiative in this site: <http://www.soros.org/openaccess>

[37] <http://www.earlham.edu/~peters/fos/bethesda.htm>

[38] <http://www.ling.lu.se/projects/echo/contributors/charter.html>

[39] BOAI, supra note 36.

[40] Another narrow definition can be found in the Directory of Open Access Journals (DOAJ). See: <http://www.doaj.org/articles/questions#definition>

[41] Suber, P. *Creating an Intellectual Commons through Open Access*, May 28, 2004. @: <http://dlc.dlib.indiana.edu/archive/00001246/01/suberrev052804.pdf >

[42] See: Guadamuz, A "The 'New Sharing Ethic' in Cyberspace", 5(1) *Journal of World Intellectual Property* 129 (2002).

[43] <http://www.doaj.org>

[44] Lawrence, S. "Free online availability substantially increases a paper's impact", *Nature*, May 31 (2001). @: <http://www.nature.com/nature/debates/e-access/Articles/lawrence.html>

[45] Odlyzko, A. "The Economics of Electronic Journals", 2(8) *First Monday* (1997). @: <http://firstmonday.org/issues/issue2_8/odlyzko/index.html>

[46] Willinsky J. "Scholarly Associations and the Economic Viability of Open Access Publishing", 4(2) *Journal of Digital Information* 177 (2003).

[47] For a comprehensive account of the race for the human genome, see: Sulston J. "Intellectual Property and the Human Genome", in Drahos P. and Mayne R. (eds) *Global Intellectual Property Rights: Knowledge, Access and Development,* Basingstoke: Palgrave Macmillan (2002) pp.61-73.

[48] These included the Wellcome Trust, the UK Medical Research Council, the U.S. National Center for Human Genome Research, the German Human Genome Programme, the European Commission, the Human Genome Organisation and the Human Genome Project of Japan.

[49] For a look at a summary of both meetings, see: <http://www.ornl.gov/hgmis/research/bermuda.html>

[50] Sulston, supra note 5, pp.64-65.

[51] Human Genome Project. *International Human Genome Sequencing Consortium Announces "Working Draft" of Human Genome*, Human Genome Project press release, June 2000. @: <http://www.genome.gov/10001457>

[52] A good collection of these can be found here: <http://www.ensembl.org/genome/central/>

[53] To contrast both views, see: Waterson R; Lander E. and Sulston J. "On the sequencing of the human genome", *Proceedings of the National Academy of Science*, vol.99, no.6 2002, pp.3712-3716; and Myers E; Sutton G. et al. "On the sequencing and assembly of the human genome", *Proceedings of the National Academy of Science*, vol.99 no.7, 2002, pp.4145-4146.

[54] Howard K. "The Bioinformatics Gold Rush" 283(1) *Scientific American* 58 (2000).

[55] P Kent H. "Benefits of genetic research must be shared, international genome organization warns", 162(12) *Canadian Medical Association Journal* 1736 (2000).

[56] Blumenthal D. et al, "Data Withholding in Academic Genetics", *JAMA* 477 (2002).

[57] Andrews, L. "The Gene Patent Dilemma: Balancing Commercial Incentives With Health Needs", *Houston Journal of Health Law*

and Policy 65 (2002).

[58] Thompson N. “May the Source Be With You”, *Washington Monthly*, July/August 2002. @: <http://www.washingtonmonthly.com/features/2001/0207.thompson.html>

[59] Cited by Williams S. “I Hack the Body Electric”, *O’Reilly Network*, July 25 2002. @: <http://www.oreillynet.com/lpt/a/2580>

[60] Cukier, supra note 11.

[61] Sulston, supra note 5, p.64.

[62] Burk, supra note 12.

[63] See for example, Carlson R. “Open-Source Biology and Its Impact on Industry”, *Spectrum Online*, May 2004. @: <http://www.spectrum.ieee.org/WEBONLY/resource/may01/spea.html>

[64] “Bioinformatics”, *Wikipedia*. @: <http://en.wikipedia.org/wiki/Bioinformatics>

[65] For example, see: Dudoit S., Gentleman R.C. and Quackenbush J. "Open source software for the analysis of microarray data", 1 *Biotechniques* 45 (2003); and Hubbard T., Barker D. et al. "The Ensembl genome database project" 30(1) *Nucleic Acids Research* 38 (2002).

[66] The centre can be found here: <http://www.cambia.org>

[67] Dennis C. "Biologists launch 'open-source movement'", 431 *Nature* 494 (30 September 2004).

[68] CAMBIA, Intellectual Property. @: <http://www.cambia.org/cambia_ip.html>

[69] Based on the successful open source software project repository and portal called SourceForge. BioForge is found here: <https://www.bioforge.net/>

[70] For some literature that deals directly with FOSS licences, see: Gomulkiewicz R. "De-Bugging Open Source Software Licensing", 64 *University of Pittsburgh Law Review* 75 (2002); Ravicher D. "Facilitating Collaborative Software Development: The Enforceability of Mass Market Public Software Licences", 5 *Virginia Journal of Law & Technology* 11 (2000); Kennedy D. “A Primer on Open Source Licensing Legal Issues: Copyright, Copyleft and Copyfuture” 20 *St. Louis University Public Law Review* 345 (2001),; and Nadan C. "Open Source Licensing: Virus or Virtue", 10 *Texas Intellectual Property Law Journal* 349 (2002).

[71] Richard Stallman is generally attributed as the author of the GPL, but Professor Eben Moglen provided legal assistance. See: Moody G, *Rebel Code: Linux and the Open Source Revolution,* London: Penguin (2002), p.26.

[72] Tsiavos, P. “The (dis)illusions of a rebel: A reappraisal of the General Public License through techno-organizational analysis”, *BILETA Annual Conference*, Durham, 25-26 March 2004.

[73] For example, the Public Library of Science (PLoS), one of the largest open access journal collections, is published through a CC licence.

[74] Creative Commons. *Legal Concepts*. @: <http://creativecommons.org/learn/legal/>

[75] In the Free Software sense.

[76] Creative Commons. *Baseline Rights*. @: <http://creativecommons.org/learn/licenses/fullrights>

[77] For more about technical protection measures, see: Dusollier S. “Electrifying the fence: the legal protection of technological measures for protecting copyright”, 21(6) *European Intellectual Property Review* 285 (1999).

[78] Guadamuz, supra note 9, pp.333-334.

[79] For more about the concept of viral contracts, see: Radin M.J. "Humans, Computers, and Binding Commitment", 75(4) *Indiana Law Journal* 38 (2000).

[80] Version 2.0 can be found here: <http://creativecommons.org/licenses/by-nc-sa/2.0/>

[81] To be more specific, the code uses Resource Description Framework (RDF) metadata. For more about RDF, see: <http://www.w3.org/RDF/>

[82] The full text of the licence can be found here: <http://www.gnu.org/copyleft/fdl.html>

[83] GFDL, para 4.

[84] <http://encyclopedia.thefreedictionary.com/>

[85] The licence can be found here <http://www.opencontent.org/opl.shtml>

[86] This can be found here: <http://www.eff.org/IP/Open_licenses/eff_oal.php>

[87] <http://www.law.ed.ac.uk/ahrb/script-ed/sol.htm>

[88] Lawton, G. “The Great Giveaway”, *New Scientist*. @: <http://www.newscientist.com/hottopics/copyleft/copyleftart.jsp>

[89] Berne Convention on Literary and Artistic Works, Art. 2.

[90] Cohen, P. and Ryan, T. *Copyright Law and the Internet*. @: <http://info.utas.edu.au/docs/info/utas88/Peter.Cohen.html>

[91] See for example, Gelbart W.M. "Databases in Genomic Research", 282(5389) *Science* 659 (1998); and Lawrence S. and Giles L. "Accessibility of Information on the Web", 400 *Nature* 107 (1999).

[92] See National Research Council. *A Question of Balance: Private Rights and the Public Interest in Scientific and Technical Databases*. Report by the Committee for a Study on Promoting Access to Scientific and Technical Data for the Public Interest (1999), pp.40-51.

[93] Ibid, p.14.

[94] See supra note 8.

[95] See for example, the GDB Human Genome Database, @: <http://gdbwww.gdb.org/>.

[96] For more complete analyses of U.S. database protection, see: Askanazi J., Caplan G., et al. "The Future of Database Protection in U.S. Copyright Law", *Duke Law & Technology Review* 17 (2001); and Ginsburg, J.C. "Copyright, Common Law, and Sui-Generis Protection of Databases in the United States and Abroad", 66 *University of Cincinnati Law Review* 151 (1997).

[97] In particular, see *Feist Publications v. Rural Telephone Co.* 499 U.S. 340 (1991).

[98] This is made more evident in *CCC Information Services v. Maclean Hunter Market Reports*, 44 F.3d 61 (2d Cir. 1994).

[99] For more about database protection in Europe, see: Colston C. “Sui Generis Database Right: Ripe for Review?” (3) *Journal of Information, Law and Technology* (2001). @: <http://www2.warwick.ac.uk/fac/soc/law/elj/jilt/2001_3/colston/>

[100] Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, OJ L 077, 27.03.1996.

[101] Art 7(1).

[102] These are: *British Horseracing Board v William Hill* C-203/02; *Fixtures Marketing Ltd. v Svenska Spel AB* C-338/02; *Fixtures Marketing Ltd v OY Veikkaus Ab* C-46/02; and *Fixtures Marketing Ltd. v OPAP* C-444/02.

[103] See for example, the UK’s Copyright, Designs and Patents Act 1988, s 3A; and even Art. 2.5 of the Berne Convention.

[104] Waelde C. and McGinley M., "Public Domain; Public Interest; Public Funding: focussing on the ‘three Ps’ in scientific research", 2:1 *SCRIPT-ed* 83 (2005). @: <http://www.law.ed.ac.uk/ahrb/script-ed/vol2-1/3ps.asp>, p.91.

[105] The licence can be found here: <http://www.hapmap.org/cgi-perl/registration>

[106] For more about genetic database protection, see: Baba E. “From Conflict to Confluence: Protection of Databases Containing Genetic Information”, 30 *Syracuse Journal of International Law and Commerce* 121 (2003).

[107] Creative Commons. *Science Commons*. @: <http://creativecommons.org/projects/science/proposal>

[108] Hubbard T. and Love J. “A New Trade Framework for Global Healthcare R&D”, 2(2) *PLOS Biology* (2004). @: <http://www.plosbiology.org/plosonline/?request=get-document&doi=10.1371%2Fjournal.pbio.0020052>

[109] CAMBIA, *About the BIOS licence*. @: <http://www.bios.net/daisy/bios/398>

[110] At the time of writing, the licence is still in draft stages. The comments are for version 1.1.

[111] BIOS Licence version 1.1, Background.

[112] See Mansfield, supra note 3.

[113] For more about the economics of the patent system, see: Jaffe A and Lerner J, *Innovation and Its Discontents,* Princeton NJ: Princeton University Press (2004), pp.64-69.

[114] This assumption is an extrapolation from what is taking place in the use of open licences in software and the creative industries.

[115] Moglen E. Why *the FSF gets copyright assignments from contributors*. @: <http://www.gnu.org/copyleft/why-assign.html>

[116] Cooper Feldman, R. The *Open Source Biotechnology Movement: Is it Patent Misuse?* SSRN Working paper, May 2004. @: <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=545082>

[117] Creative Commons. *Science Commons proposal*. @: <http://creativecommons.org/projects/science/proposal>

[118] Ibid.

[119] Eisenberg R.S. *The Public Domain in Genomics*, (2000). @: <http://www.law.nyu.edu/ili/conferences/freeinfo2000/abstracts/eisengberg.html>

[120] For more about patent abuse in these two fields, see: Gratton E. “Should Patent Protection Be Considered for Computer Software-Related Innovations?” *Computer Law Review & Technology Journal* 223 (2003); Andrews, supra note 57; and Jaffe and Lerner, supra note 113.

[121] Hope J. *Open Source Biotechnology?* @: <http://rsss.anu.edu.au/~janeth/OSBiotech.html>

[122] Some examples have already been mentioned. For other examples, see: Maurer S., Rai A. and Sali A. “Finding Cures for Tropical Diseases: Is open source an answer?” Yarris L. ed, *Biotechnology: Essays From Its Heartland*, BASIC Report (2004), pp.33-37.

[123] See Cukier, supra note 11.

[124] Ibid.

[125] The licence can be found here: <http://www.opensource.org/licenses/apache2.0.php>

[126] According to Netcraft’s web server survey for March 2005. See: <http://news.netcraft.com/archives/web_server_survey.html>

[127] Data gathered from SourceForge’s Software Map, @: <http://sourceforge.net/softwaremap/trove_list.php?form_cat=14>

[128] “IBM frees 500 software patents”, *BBC News* (11 January 2005). @: <http://news.bbc.co.uk/1/hi/technology/4163975.stm>

[129] The list of approved licences is here: <http://www.opensource.org/licenses/>

[130] IBM. *IBM Statement of Non-Assertion of Named Patents Against OSS*, (1 January 2005). @: <http://www.ibm.com/ibm/licensing/patents/pledgedpatents.pdf>

[131] Ibid.

[132] “IBM, Matsushita, Canon and HP received the most US patents in 2004”, *IT Facts*. @: <http://www.itfacts.biz/index.php?id=P2370>

[133] For example, see: Mitchell P. and Phillips J. "The Contractual Nexus: Is Reliance Essential?" 22(1) *Oxford Journal of Legal Studies* 115 (2002), pp.118-124.

[134] [1893] 2 QB 484.

[135] With some few exceptions, such as the requirement that the promise should be in writing. For more about promises in Scots contract law, see: MacQueen H.L. and Thomson J.M. *Contract Law in Scotland,* Edinburgh: Butterworths (2000), pp.63-69.

[136] For more about European applicability of unilateral promises, see: Gordley J. (ed), *The Enforceability of Promises in European Contract Law*, Cambridge, Cambridge University Press (2001).

[137] § 657 BGB.

[138] As shown in *Cass. Civ. 1re, October 16 1995*.

[139] This is a very early draft. Comments are welcome, please send them to a.guadamuz@ed.ac.uk.