

# The International Journal of Digital Curation

Issue 1, Volume 1 | Autumn 2006

## Digital Curation, Copyright, and Academic Research

Andrew Charlesworth,  
Senior Research Fellow  
Centre for IT & Law, School of Law/  
Dept. of Computer Science  
University of Bristol

### Abstract

A defining characteristic of contemporary copyright law is the willingness of governments to accept the argument that the impact of digital technologies requires copyright owners to be given ever greater control over the use of their works, regardless of the detriment to the copyright regime's 'public interest' elements. Yet a one-size-fits-all 'all rights reserved' copyright regime clearly fails to meet the requirements of many rightsholders. One response has been the Creative Commons movement which seeks, through licences based on existing copyright laws, to provide a simple mechanism for rightsholders to disseminate their works under less restrictive conditions. The Creative Commons' initial success has led to suggestions that its principles could be equally applied to scientific research outputs, such as publications, licensing of research materials, and datasets. This article argues that the Science Commons approach, if based on the Creative Commons model, and premised at its root on utilitarian copyright law, will both fail to address contemporary policy drivers in research, or to provide researchers with the type of rights that they actually want. It suggests that constructing an appropriate set of rights for the Science Commons, particularly for datasets, will require a willingness to step outside the utilitarian model and look to the Continental copyright tradition, which sets less store in economic rights and gives greater weight to moral rights.

## Introduction

The interaction between the intellectual property regime, notably copyright law, and the dissemination of research information, is clearly a controversial one. As reviewed below, recent years have seen considerable debate about the effects of changes to the copyright regime, which have been driven themselves primarily by the changes wrought upon existing understandings of the purpose of copyright law by digital technologies. From this debate, and from the concerns about the possible negative impacts upon information dissemination of recent amendments to copyright laws globally, have arisen a number of initiatives, such as the Open Source movement, and the Creative Commons. These aim to provide mechanisms by which authors, creators and rightsholders can utilise national copyright laws to achieve a more targeted and appropriate set of rights over their works than the perceived “all or nothing” approach of standard copyright law. Such initiatives have achieved some success in raising public awareness of the implications of ‘strong’ copyright for public access to information; encouraging debate about the nature, scope and application of copyright; and bringing about freer dissemination of information in at least some areas of activity.

This paper develops an argument suggesting that while the aspirations of those initiatives are laudable, there are key obstacles obstructing their successful adoption by academic researchers. Firstly, in the academic context, there are still significant drivers which militate against practical implementation of their ‘weak’ copyright methodologies. Secondly, the initiatives themselves have been lacking in imagination in adopting essentially a ‘one size fits all’ approach to works, entirely structured around the utilitarian copyright regime. The rights that academic researchers appear to want in their works and datasets simply do not match the types of rights available under that regime. It is suggested that, in this context at least, the nature of the rights required is more attuned to those often available in the civil law jurisdictions - the concept of *droit d’auteur*, including the elements of *droit moral*.<sup>1</sup>

The paper will first consider the socio/economic context of the debate, and the obstacles to the type of legal changes that would facilitate greater access to data and datasets. It will then consider some of the approaches to overcome those obstacles adopted by the Open Source movement, and the Creative Commons, with particular reference to the developments around the Science Commons. Finally it will discuss what are likely to be the underlying implications of those approaches for the development of future ‘law-influencing’ strategies for those working in the data/digital curation – a role perhaps ideally suited to the UK Digital Curation Centre.<sup>2</sup>

It is important to note that this article is not calling for an end to the existing copyright regime, nor even making an argument for significant changes to that regime – there are many existing works that fulfil the needs of that debate. What it *is* calling for is a more creative use of licensing than has been currently envisaged, even by campaigning organisations such as the Creative Commons. It suggests that imaginative use of licence conditions could be used to reposition certain moral rights principles as key elements of a future Science Commons. Those principles, some examples of which are suggested in the table at the end of this paper, would be based not on simple utilitarian grounds, but rather upon a common set of norms and expectations that could be expected to be familiar to academic researchers.

<sup>1</sup> Often translated in English as ‘moral rights,’ though as some commentators (Holderness 1998, Vaver 1999) note, this is perhaps an unsatisfactory, not to say misleading, approximation of the meaning of the concept.

<sup>2</sup> The Digital Curation Centre  
<http://www.dcc.ac.uk/>

## Understanding the Context

The key to understanding how the law of intellectual property in the UK and the US (as opposed to the law in Continental Europe) has developed, lies in an understanding of the utilitarian aims of that law. In broad terms, the Anglo-American utilitarian view of copyright is generally understood to be based on the concept that by granting certain rights over their works to authors/creators, and thus ensuring them a 'fair economic return' for their labour, those individuals are encouraged to engage in creative/inventive processes, which advance science and knowledge, thereby befitting society generally.<sup>3</sup> (Deazley, 2004; Rose, 1993)

The granting of quasi-monopolistic rights to authors/creators to limit the copying of private material for a set period of time, and to police and punish infringement of that right, is theoretically balanced by the grant to the public of various fair use/fair dealing related defences with regard to the use of copyrighted material: the creation of doctrines such as exhaustion;<sup>4</sup> the setting of fixed terms of protection after which private domain material enters the public domain; and the right to free use of material in the public domain. The nature of this balance, sometimes referred to as the copyright 'social contract', has been the subject of debate from the beginning. (Charlesworth, 2005) Rightsholders constantly seek to increase the scope of rights over their works, to extend those rights to hitherto unforeseen types of work, and to oppose technological innovations that threaten the effective exercise of those rights; whilst other interest groups, both public and commercial, seek to expand the scope of fair use/fair dealing rights, to oppose extensions to terms of protection, and to exclude new technologies from their scope. (See further (Bettig, 1996; Goldstein, 1994; May, 2000; Perelman, 2002; Thierer & Crews, 2002; Vaidhyanathan, 2001).)

The advent of digital technologies has intensified the debate over the nature of the 'social contract'. Advances in information technology have massively increased the value of information and material held by various rightsholder interest groups, notably content providers, such as the music and film industries. This has led them to be more aggressive about utilisation of intellectual property rights in their works, more ambitious in attempting to 'enclose' publicly-created works,<sup>5</sup> and increasingly willing to expend significant resources exerting influence over the arbiters of that 'social contract' - national governments and national legal systems. Given the finance dedicated to ever more persistent lobbying at both national and supranational levels, in combination with other factors, such as the increasing role played by corporations in the development of the international intellectual property agenda, e.g. in the World Intellectual Property Organisation (WIPO) and the World Trade Organisation (WTO) (Drahos & Braithwaite, 2002; Sell, 2003), it is perhaps not surprising that governments have been willing to accede to, or unable to resist, the rightsholders' demands. As such, the last 15 years have seen:

<sup>3</sup> For example, Article I, Section 8, Clause 8 of the US Constitution states, "the Congress shall have power . . . to promote the progress of science and useful arts, by securing for limited times to authors and inventors the exclusive right to their respective writings and discoveries."

<sup>4</sup> The doctrine of exhaustion permits a legitimate purchaser of a copyrighted good to sell or pass it onwards without first seeking the rightsholder's permission – hence, the existence of second-hand book and CD shops, and the fact that a person can lend a book they've purchased to a friend, or give it away to a charity shop. The doctrine has come under particular attack in the digital environment through use of licence agreements purporting to exclude onward passage of original digital works by a legitimate purchaser, and by digital rights management (DRM) techniques that tie access to a legitimately purchased work to a specific device, such as a PC.

<sup>5</sup> For example, West Group's attempt to 'privatise' US judicial opinions by adding pagination and other additional materials, then asserting copyright in the resulting works, which consisted predominantly of material created by the US courts. Works by the US government, including the courts, are not eligible for US copyright protection. See *Matthew Bender & Co. v. West Publishing Co.*, 42 *United States Patent Quarterly* 2d 1930 (US District Court, Southern District of New York 1997), decision affirmed, 158 *Federal Reporter* 3d 674 (US Court of Appeals, 2nd Circuit 1998), certiorari petition to the Supreme Court denied, *West Publishing Co. v. HyperLaw, Inc.*, 119 *Supreme Court Reporter* 2039 (1999).

- the extension of the copyright regime to new types of works;<sup>6</sup>
- significant reductions in the scope of fair use/fair dealing defences and other public rights;<sup>7</sup>
- significant extensions in the term of protection for copyright;<sup>8</sup> and
- the expansion of copyright law to include new elements favourable to rightsholders, for example, the protection of the use by rightsholders of technical measures, such as digital rights management (DRM) adding a further layer of protection to their works.<sup>9</sup> (Bechtold, 2004; Charlesworth, 2005).

These considerable, and often very public, changes to copyright law, are perceived by many commentators as unbalancing the copyright 'social contract' to the considerable detriment of the general public (and, indeed, to the disadvantage of technology innovators who find their innovations subject to levies, restrictions, and bans in the marketplace). This perceived inequity has led to, or given impetus to, a number of innovative approaches designed to provide the author/rightsholder with means to ensure their copyrighted works might be made more readily available to the public, without simply dedicating them to the public domain. The risk with simply dedicating material to the public domain is that it may be subject to 'enclosure' by third parties, (Gellman, 1995; Lessig 2004, pp.147-153), or the benefits of future work based on it may be privatised in a way that is alien to the wishes of the original author/rightsholder. The need is, therefore, for the development of more effective mechanisms that provide the author/rightsholder with sufficient rights to permit them to control the types of uses of their work, such as preventing its 'misuse' or misappropriation, but which at the same time permit the fullest public access to that work.

It is clear that rightsholders do not all hold to a homogenous set of interests. As a result, many creators and other rightsholders have aims which are disserved by current developments in national and international copyright policy. It is these rightsholders that the alternative copyright movements, in particular the Open Source Software and Creative Commons communities, are largely aimed at. Using existing national copyright laws, these groups have developed a set of tools, in the form of copyright licences, which permit creators of works and other rightsholders to disseminate those works in a manner which protects a less restrictive set of rights than those automatically granted by copyright law. The licences can, in certain circumstances, also ensure that any derivative work created from those works is also disseminated under those less restrictive conditions. In the words of the Creative Commons:

"We use private rights to create public goods: creative works set free for certain uses. Like the free software and open-source movements, our ends are cooperative and community-minded, but our means are voluntary and libertarian. We work to offer creators a best-of-both-worlds way to protect their works while encouraging certain uses of them — to declare "some rights reserved."

Thus, a single goal unites Creative Commons' current and future projects: to build a

<sup>6</sup> e.g. UK *Copyright (Computer Programs) Regulations 1992*  
[http://www.opsi.gov.uk/si/si1992/Uksi\\_19923233\\_en\\_1.htm](http://www.opsi.gov.uk/si/si1992/Uksi_19923233_en_1.htm), and

*UK Copyright and Rights in Databases Regulations 1997*.  
<http://www.opsi.gov.uk/si/si1997/19973032.htm>

<sup>7</sup> e.g. UK *Copyright and Related Rights Regulations 2003*, ss.9-14  
<http://www.opsi.gov.uk/si/si2003/20032498.htm>

<sup>8</sup> e.g. UK *Duration of Copyright and Rights in Performances Regulations 1995*  
[http://www.opsi.gov.uk/si/si1995/Uksi\\_19953297\\_en\\_1.htm](http://www.opsi.gov.uk/si/si1995/Uksi_19953297_en_1.htm)

<sup>9</sup> e.g. UK *Copyright and Related Rights Regulations 2003*, s.24

layer of reasonable, flexible copyright in the face of increasingly restrictive default rules.”<sup>10</sup>

The Open Source community is perhaps less cohesive than the Creative Commons in its approach to the use of the copyright system to protect rights in software source code. There are clear ideological divisions within its ranks. Some developers aim to use open source software licences to ensure that their software code is ‘free’ in the sense that anyone can use it, and no-one can appropriate either it or any derivative works based on it. Others wish to ensure that their code remains open, but do not necessarily require that derivative works based on that code not be appropriable by third parties. (See further (DiBona et al., 1999; Rosen, 2004; van Wendel de Joode, 2003).)

That having been said, both of these communities have developed out of the perceived need for a limited form of copyright protection, which national copyright frameworks have long failed to explicitly cater for. As such, they provide compelling evidence that not only are those copyright frameworks failing to adequately reflect the public interest in freer access to copyright works but, for a growing category of creators and other rightsholders, particularly in the digital environment, the law is failing to provide an appropriate and efficient means of ensuring the effective dissemination of their works. Further examples of the desire for a break from traditional forms of copyright licensing are also developing. The rise of sharing websites, such as Flickr, the photosharing website,<sup>11</sup> has been, at least in part, aided by the public’s acceptance of the possibilities inherent in the Creative Commons licences. Musicians are also seeing how flexible use of licensing can allow them to successfully break away from music industry business models that favour distributors over creators.<sup>12</sup> In the words of UK judge and intellectual property law commentator, Mr Justice Laddie, for many creators the current state of the Anglo-American copyright system is that it is “over-strength, over-regulated and over-rated.” (Laddie, 1996)

This charge - “over-strength, over-regulated and over-rated” – seems to have particular resonance when one considers the outputs of much academic and other publicly funded research, whether these be publications or datasets. Research in all disciplines relies heavily on the ability to access existing works, in order that researchers can: familiarise themselves with the nature and scope of work already carried out, or underway; learn about new techniques, observations and theories; test the work of other researchers; and, ultimately, build upon the existing store of knowledge. Obstacles to the free flow of this information stream are damaging to both publicly and privately funded researcher alike, resulting in, amongst other difficulties: duplication of effort with concomitant waste of resource; inability to adequately referee or confirm published research; and delays in the creation and distribution of valuable public goods. Lack of access to research outputs may hinder public discussion of important social issues, or may result in debate taking place without the benefit of full access to relevant research, fostering inadvertent or deliberate misrepresentation of its implications, and encouraging polemic rather than reasoned discussion. (See, for example Kahlon (2005).)

<sup>10</sup> About the Creative Commons  
<http://creativecommons.org/about/history>  
<http://www.creativecommons.org.uk/>

<sup>11</sup> Flickr  
<http://www.flickr.com/>

<sup>12</sup> See, for example, Unsigned Band Web  
<http://www.unsignedbandweb.com/>

## An Alternative Approach – The Science Commons

The founders of the Creative Commons have targeted some aspects of this issue by creating a project called the Science Commons.<sup>13</sup> The stated aim of this project is

‘... to encourage stakeholders to create areas of free access and inquiry using standardized licenses and other means: a ‘Science Commons’ built out of voluntary private agreements.’

At present, the project has three main areas of activity: publishing, licensing, and data. The publishing area is concerned with the interface between modern technologies for publishing and distribution, such as the World Wide Web and developments in copyright law. Its aim is to ‘reduce the legal and logistical effort involved in managing copyrighted scientific publications.’ The type of work envisaged as being undertaken

‘...include pre-print and post-print standard commercial publisher licenses, supporting author self-archiving through education, outreach, and technology, institutional archive licenses, and the impact of collaborative annotation and text mining on copyright management in institutional or corporate archives.’<sup>14</sup>

This type of work may sound familiar to those working in the area of digital curation – it certainly mirrors the kinds of developments that have been taking place within the Higher education sector in the UK.<sup>15</sup>

The licensing area is concerned with the licensing of research materials, and concerns the extent to which current transaction costs can obstruct research work, even before it begins. Carrying out research, particularly in biological sciences may require the licensing of existing research materials, such as cell lines, model animals, DNA constructs, and screening assays. This can be a complex and expensive process, and may delay or even prevent certain types of research. The Science Commons’ aim is to provide solutions to the complexity of research licensing across a range of research materials, by providing a standard, open framework for managing material transfer. The initial work aims to answer questions such as:

How can funding organizations, corporations and universities utilize standard, open agreements to facilitate licensing of intellectual property and exchange of materials?

Which existing agreements are good targets for standardization?

What value would standard, open legal approaches bring for orphan diseases and health products in the developing world?<sup>16</sup>

<sup>13</sup> ‘Welcome to Science Commons’  
<http://sciencecommons.org/>

Despite its name, the Science Commons appears to have set its sights on a wider range of disciplines than ‘the Sciences’ might imply - an early development has been the Open Access Law Program. See  
<<http://sciencecommons.org/literature/oalaw>>

<sup>14</sup> Science Commons - Scientific Publishing  
<http://sciencecommons.org/literature>

<sup>15</sup> See the work of Project RoMEO <http://www.lboro.ac.uk/departments/ls/disresearch/romeo/> the SHERPA project <http://www.sherpa.ac.uk/>, and the Zwolle Group <http://www.surf.nl/copyright/>

<sup>16</sup> Science Commons - Licensing  
<http://sciencecommons.org/licensing>

The US Food and Drug Administration defines an ‘orphan disease’ as a condition that affects fewer than 200,000 people in the US. Orphan diseases and third world diseases are often not seen as ‘economically viable’ topics for research by pharmaceutical companies as revenues from treatments are unlikely to cover development costs. However, the cost of licensing the necessary research materials may prevent other non-profit organisations, such as universities, charities and other NGOs, from undertaking the research. A more flexible licensing might help to overcome some of the licensing problems.

The data area is concerned with considering ways in which access to scientific data can be made more open. It is suggested that a number of factors are at play in hindering open access:

- developments in the international intellectual property regime, not least the recent creation of the concept of database rights,<sup>17</sup> threatens to lock up research data, make it more expensive or place access to it under restrictive licensing agreements. (Waelde & McGinley, 2005)
- the effects of new technologies for data creation and dissemination, in combination with intellectual property rights, risk the compartmentalisation of data created in the course of mapping and archiving knowledge of the world's land, oceans, forests and historic sites into "all rights reserved" (full use of copyright with excessive control over data) and "no rights reserved" (denial of copyright, with attendant loss of valuable attributes of copyright protection) categories.
- a growing tendency towards a wasteful data economy, where costly-to-produce research data that could be reused or reanalysed, possibly for different ends than that for which it was originally produced, is either withheld or lost due to individual researchers fearing loss of control over their data, or institutional requirements mandating non-disclosure. It is suggested, perhaps controversially, that "Implicit in data sets are answers to questions the researcher perhaps did not specify - answers that are a consequence of the throughput of the experiment."<sup>18</sup>

The aim here is primarily exploratory, with a focus on:

‘Evaluat[ing] and draft[ing] open, voluntary and interoperable legal solutions for databases and genomes in the sciences, based on the "some rights reserved" philosophy of Creative Commons.’<sup>19</sup>

## Analysing the Science Commons Concept

Of the three areas, it is the last, that of data, that appears to hold the most interesting questions, but which will contain potentially the most contentious and the most intractable issues. Even before considering the issues of the rights in data produced in the course of research, the Science Commons’ stated approach to the nature of research data, and its possible uses and reuses, is unlikely to garner unanimous support. In this regard, the responses to the Biotechnology and Biological Sciences Research Council (BBSRC) online consultation on “Developing a BBSRC Data Sharing Policy” are illuminating. (Biotechnology and Biological Sciences Research Council [BBSRC], 2005) While there was qualified support for data sharing, concerns were raised about issues such as:

- Data quality and interpretation
  - incomplete data sets and lack of well-managed meta-data;
  - difficulties in supplying suitable warranties of quality of data; and,
  - potential liability for the inadvertent misinterpretation or misrepresentation by experimenters of data leading to wrong conclusions/decisions by others;
- Financial and career cost implications

<sup>17</sup> See, for example, EU Directive 96/9/EC on the legal protection of databases (‘the Database Directive.’) <http://europa.eu/scadplus/leg/en/lvb/l26028.htm>

<sup>18</sup> Science Commons – Data  
<http://sciencecommons.org/data>

<sup>19</sup> *Ibid.*

- the cost of managing data and datasets, notably the need for greater input and curation activities at a local level before submission to a publicly accessible repository, with the attendant opportunity costs for researchers generating the data;
- the need for publishing, contributing to, data sharing with, and curating of, appropriate databases to have similar merit in a science career as a paper publication to incentivise such activities, and to ensure that researchers who work to make datasets available for sharing are not placed at a career disadvantage to those who choose not to;
- Data sharing infrastructure
  - the need not just for open data sharing, but also for the software needed to access the data made available to be openly available and not expensive proprietary software.

It is also worth noting that, despite the altruistic nature of the Science Commons' (and indeed the Creative Commons') approach as regards the use of the intellectual property regime, this approach to data and datasets may not accurately reflect the norms of other cultures, or the aims and objectives of the developing world. Certainly, after centuries of exploitation of the resources of the developing world by the developed world, it may be a somewhat optimistic to expect that a bright new policy of research data sharing suddenly emanating from the developed world will not be considered, in some circles at least, as just another mechanism for ensuring that the developed nations maintain their ability to abstract value from the developing world, without necessarily contributing to the economies of those nations.

On the issue of rights, it will be clear from the discussion in the context section above that the Science Commons approach to data and datasets is (in the US and UK at least) going to run head-on into a utilitarian intellectual property environment which is essentially antithetical to it. While the Science Commons aims to use a similar principle of "some rights reserved" to the Creative Commons, which is underpinned by the existing intellectual property regime, simply enlisting intellectual property law in the development of the Science Commons is not, in and of itself, going to be sufficient. The contextual social and economic environment within which that law is interpreted is also going to have to change, and this is going to be rather more difficult to achieve than simply writing Creative Commons or Science Commons licenses that individuals and organisations could use *in principle* – there is going to have to be considerable thought and effort devoted to ensuring that they have rational incentives to actually do so *in practice*.

The precedents here are mixed. An example of the difficulties of translating the theoretical benefits of sharing academic research outputs into practical application is demonstrated by the situation in the more developed area of Open Access publishing. Open access to academic outputs is widely seen as desirable, but there are numerous policy drivers against it; and, in the UK at least, some of them currently arise (largely inadvertently, it must be said) from the activities of the very organisations - research councils and government - that wish to facilitate the development of Open Access. To quote from the recent JISC Rights in Digital Environments Report (JISC, 2005):

The rumoured focus in some panels of the UK Research Assessment Exercise on 'top journals' or 'top publishers', as well as the AHRB/AHRC's proposed exercise to create "top ten" lists of the most important journals in a bid to 'establish new performance measures for arts and humanities research,' risks seriously hindering the development of academic open access publishing, and forcing academics into situations where they have to sign unduly onerous assignments or licence agreements in order to be published in the 'right' journals, or by the 'right' publisher. The inadvertent creation of such artificial pressures on academic authors to permit the commercial enclosure of their works is symptomatic of the failure of both the

government and the Research Councils to:

- adequately assess the likely impacts of new research audit practices upon individual and institutional behaviour as regards rights management in the academic environment;
- understand possibilities of the digital environment for enhancing the dissemination of UK academic research and publishing to an international audience.

It has become clear that if Open Access publishing is to become more prevalent, particularly in non-science subjects, it will be necessary for greater thought to be given to the types of environmental disincentives that exist and how they can be overcome. It remains the case in some of the more conservative subject disciplines that publication of an academic's work outside a traditional grouping of subject journals may result in the academic failing to obtain tenure, or being passed over for promotion due to adverse internal RAE grading. While such disincentives, whether actual or perceived (Oppenheim, 2005), remain unchallenged, the ability of some academics to publish in Open Access formats may remain a theoretical, rather than practical, option. The Open Access process suggests that without a full understanding of both the operational context, a strategy for ensuring that the relevant policy actors are adequately involved, and clear reassurances to 'early adopter' researchers that they will not be inadvertently penalised, the Science Commons is likely to struggle to make significant headway.

The development of a Science Commons will also have to take into account the inherent tensions in the UK academic sector between other strands of research policy. It is a common requirement amongst the UK Research Councils that data and datasets arising from funded research must be provided for archiving and public access at the end of projects. At the same time, however, there is increasing political pressure for research institutions to be seen to be producing exploitable works and developing spin-off companies from the research they undertake, in order to increase their operating revenues. Institutions are also urged to actively seek out commercial sponsorship and funding for research projects. While these policy goals are not wholly incompatible, there appears to be a lack of comprehension, at the political level, about the implications of the increasing commercialisation of academic research for concepts such as open access to research data, broad peer review, and academic freedom.

A further problem facing academic researchers is the enthusiasm in both US and UK governments for devolving power and decentralizing services from central government to third parties, including non-profit agencies and private firms (Rhodes, 1994). In some areas, this has led to the creeping privatisation of formerly publicly accessible data and datasets produced in the public sector. An example of this is noted by David (2004) who describes the transfer of the US Landsat system of remote sensing satellites from the National Oceanic & Atmospheric Administration (NOAA) to the Earth Observation Satellite (EOSAT) Company, a joint venture of Hughes and RCA, which then became the exclusive marketer for the satellites' images. This led to a huge increase in image prices, which allowed EOSAT to obtain large profits from commercial and federal government contracts, but effectively shut out many academic and independent researchers. In such circumstances, large swathes of publicly-funded research data can vanish behind proprietary 'fencing' without ever being made accessible to a 'Science Commons'.

Examination of the above issues suggests that, at present, there is little coherent strategy on the part of government as regards either the future publication of research, or future mechanisms for encouraging researchers to share data, and ensuring that data is in a suitable form to be usefully shared. In some cases, it seems not so much that the left hand doesn't know what the right hand is doing, as much as the left hand doesn't know the right hand even exists - there is

clearly much work to be done in providing effective solutions - and here the Science Commons project may help drive consideration of the issues forward. However, a further key issue remains - does the Science Commons, based as it is on “the “some rights reserved” philosophy of Creative Commons”, which in turn is based on the utilitarian model of copyright, really offer researchers the rights that they want as regards their research and its outputs?

## What Do We Really Want (and How Do We Get It)?

It appears that there is considerable support for concepts like the Science Commons in the academic/publicly funded research environment. What is not clear is whether the socio/economic context in which institutions currently operate will ever permit such concepts to translate into widespread practice. To achieve the broader goals of open access to literature, research materials, data and datasets is going to take more than clever use of existing intellectual property laws. Indeed, it is probably going to take more than tinkering with the laws themselves, even if sufficient motivation can be provided to governments (and possibly more importantly in the current environment, to corporate-dominated international intellectual property forums) to achieve legislative change. There are, perhaps, lessons to be learned from the content industries about effective, concerted and targeted lobbying of policymakers. Researchers, their institutions and representative organisations will need to play a much more high-profile hand than they have been prepared to (and for) in the past. Certainly the *examples* of the Science Commons and Creative Commons are helpful in this regard, even if in practical terms, they prove unable to garner sufficient support to achieve a ‘tipping point’ in research, whereby Science Commons/Creative Commons licences come to be the dominant model in their respective fields.

Yet the role of the current Science Commons/Creative Commons approaches may be rather less valuable, if they are not in a position actually to offer those positive benefits that researchers, as authors and data creators, are likely to want from their research outputs. Reading comments about the sharing of data and datasets from researchers, like those submitted to the BBSRC consultation mentioned above, suggests that the utilitarian copyright approach, even as augmented in the UK by the addition of a limited set of moral rights, remains an inadequate reflection of those wants and needs. The concept that has come to dominate utilitarian copyright doctrine, that what rightholders want from their works is an exclusively economic return, does not begin to cover the diverse range of concerns researchers have about the use and reuse of their data. For example:

- the need for mechanisms to ensure that data producers receive appropriate acknowledgment of their input by those using the data;
- the desirability of reuse being with the knowledge of the originators in order to avoid mistakes based on misinterpretation or misrepresentation;
- that reuse mechanisms offer the opportunity for true collaboration;
- the desirability of notice that other researchers might be about to offer different positions to those of the original researchers based on data reinterpretation. (BBSRC, 2005)

In fact, in many regards, a more appropriate base model for addressing those concerns might be drawn from the example of the Continental *droit d’auteur*. This emphasises both the justification of property rights in a work as the right to the fruit’s of one’s labour, and the notion that a work of creation is intimately linked with its creator, entitling him/her to a considerable degree of control over it. (Holderness, 1998) Of course, the immediate problem for data and datasets is that the *droit d’auteur* has traditionally required a higher test of ‘originality’ than the utilitarian model which can be considered as ‘a degree of creativity reflecting the individuality of the author’s personality’ (Davies, 2002, p.330) a quality which a journal article or a book is

likely to have, but which a dataset may struggle to achieve. It is also clear that this model has itself been far from immune from the influence of the marketplace, with suggestions that the economic element of the *droit d'auteur* has been allowed to override the implicit understanding that, in exercising their rights, authors must also respect the social interests in the dissemination of their work. (Beldiman, 2005, pp.41-43 citing Geiger 2004)

However, even if an approach based on contemporary understandings of the *droit d'auteur* doesn't of itself provide an ideal solution, then perhaps developing a wider view of 'researcher rights' - one that might plausibly be incorporated into a normative moral rights approach to research outputs - could usefully be developed as a strand of our thinking about future strategies for encouraging and supporting open access. The question is then, of course, what form should those 'researcher rights' take?

Let us imagine that a UK researcher has created a new dataset in the sciences, as part of the process of undertaking a particular piece of research. She has written several papers based on her evaluation of the material within the dataset, and these have been published in a selection of journals. Our researcher has been conscientious about ensuring open access to her articles and has placed them with publishers that allow her to engage in both preprint and postprint 'self-archiving' - publishers on the 'green' road to Open Access. (Goodman, 2004; Guedon, 2004) She is aware that by doing so she increases the likelihood of other researchers accessing and, in due course, perhaps citing, her work. (Harnad et al., 2004) She had considered publishing in Open Access journals - the 'gold' road - but was afraid this might count against her in the UK RAE, as the Open Access journals in her field are not long established.

This leaves the question of the potential dissemination of the underlying dataset. Making the dataset freely publicly accessible will permit other researchers to evaluate her interpretations of the data and, in some disciplines, to see if they can replicate her experimental results. As such, the researcher has an incentive to make the data set publicly available, both in terms of a creator's benefit to her - provision of further and more detailed evidence of her research and its value; and in terms of a societal benefit - the data being made available to other researchers. In such circumstances (and for the sake of this example, absent a sponsor or employer keen to maintain secrecy in, or proprietary rights over, the dataset, in order to make private use of the data for further institutional research, or to propertize the dataset by selling it to other researchers), there are two key questions:

- what are the 'rights' the researcher might legitimately consider she requires; and,
- after balancing her moral right to exercise control over her data against the impact the exercise of that right would have on social interests, is it appropriate to grant those 'rights' to her?

Thus, it is argued, it is appropriate to consider those underlying principles noted as the theoretical basis for the original formulation of *droit d'auteur* - society's understanding that a creator should be able to exercise certain controls over the use of their work (and that some of those controls could, or perhaps, should be inalienable), outside of the purely economic rights of the utilitarian approach; balanced by an obligation on the creator to have regard to the social interest in the dissemination of their work. Then perhaps, it is possible to start to address in a coherent fashion, the types of concerns voiced by researchers outlined above, in terms of determining whether there might be a defensible set of researchers' rights suitable for incorporation into a future Science Commons licensing scheme. The following table examines some possible rights that those planning the 'Science Commons' might consider as falling within that set.

**Table 1: Researchers' rights that might be considered for adoption by the Science Commons**

Right that might be claimed by a researcher	Justification	Example	Counterpoint	Potential cost to the 'public interest'
Right to ensure that dataset creators receive appropriate acknowledgment of their input by those reusing the data (acknowledgement, attribution or paternity right)	Allowing creator(s) of datasets to be identified as such (or not) increases likelihood of openness with datasets where reputation gain	Reuse of dataset in order to create new publications	None	None. It is arguable that this should, in the context of research data, be an inalienable right, regardless of whether the research was, for example, carried out in an employment context. <sup>20</sup>
Right to be informed of reuse of the data (information right - subject to a reasonableness test)	Allowing creator to know that their creation is being reused can help to avoid mistakes based on misinterpretation/misrepresentation	Reuse of data for research purposes other than that for which the data was originally collected/created	Might 'chill' reuse of data by 3 <sup>rd</sup> parties who may wish to produce proprietary outputs, or to engage in 'speculative' research.	Risk of loss of valuable proprietary products or 'blue sky' and innovative research.
Right to ensure that the data is reused appropriately (integrity right - subject to a reasonableness test) <sup>21</sup>	Allowing creator to prevent : <ul style="list-style-type: none"> <li>deliberate misrepresentation of data - protects creator's reputation, increases openness</li> <li>use without appropriate metadata - protects creator's reputation, increases openness</li> <li>use in inappropriate circumstances - protects creator's reputation, increases openness</li> </ul>	Data misused by 3 <sup>rd</sup> parties to support particular conclusions  Data 'mashed' with other data out of original context  Data 'mashed' with other data in ways that breach research ethics or the law	Difficult in any of the circumstances to administer a system which would permit the creator and reusers to interact in a timely manner.  May be used by the creator to suppress 'blue sky' research, research critical of the creator's research methodology, and data interpretation, or research which the creator did not see as a consequence of the throughput of the original research	Cost of administering a new 'rights' system.  Indefinite period of legal uncertainty as creators, lawyers and judiciary adapt to new 'rights' system.  Risk of loss of valuable proprietary products or 'blue sky' and innovative research.

<sup>20</sup> The author of a copyright literary, dramatic, musical or artistic work already has the right to be identified as the author of that work (s.77), where their right is asserted (s.78) under the UK Copyright, Designs and Patents Act 1988, but this is subject to a list of exemptions (s.79), including where the work is produced in an employment context (s.79(3)).

<sup>21</sup> The author of a copyright literary, dramatic, musical or artistic work has the right, in certain circumstances, not to have their work subjected to derogatory treatment (s.80), under the UK Copyright, Designs and Patents Act 1988, but this is subject to a list of exemptions (s.81), and this right is considerably weaker than those found elsewhere in Europe. (Holderness, 1998)

Right to notice of differing interpretations of the data (interpretation right - subject to a reasonableness test)	Allowing creator to prevent publication of research based on their data without notice to them allowing for a 'right to reply'	Reinterpretation of data to criticize creator's research methodology, and data interpretation	May be used by the creator to attempt to suppress research critical of the creator's research methodology, and data interpretation	May slow down speed of research, reduce reuse of existing data, impede or prevent valuable research
Right to be involved in reuse of the data (collaboration right - subject to a reasonableness test)	Allowing the creator to be involved in future development of original data use	Involvement of creator permits them to provide specific knowledge of dataset variables	May produce problems such as conflict of interests, funding difficulties, personality clashes etc.	May be unattractive to would-be users of the data, reducing reuse of existing data sets.
Right to withdraw the data (withdrawal right - subject to a reasonableness test) <sup>22</sup>	Allowing the creator to withdraw the data from public use in cases where continuing reuse damages the creator's reputation	There are serious flaws in the dataset	Removal of the dataset will make it difficult, if not impossible, to assess the value or accuracy of works based upon it	Loss of a dataset, even a flawed one, may impede existing or future research.
Right to decide whether, when, and how to release the data to the public (divulgence right- subject to a reasonableness test) <sup>23</sup>	Allowing the creator to decide when data is ready for release, with potential, for example, to override employer/employee or sponsor/funder powers to enforce such release	The creator can complete their research and publish, prior to release of data to third parties	A creator may with hold data without good cause, or long after the need for privacy/secretcy is required.	Risk that if misused this may discourage research investment or significantly impede further research in particular areas. <sup>24</sup>
Right to a reasonable share in any financial benefit arising from further exploitation of the dataset <sup>25</sup> (economic right)	Encourages research and further development / reuse of datasets, may increase openness of datasets.	Where a creator's dataset is reused for proprietary purposes, and a financial benefit is obtained, the creator is rewarded, both for the original research, and for making that research data available	An unnecessary extension of the utilitarian copyright model into the moral rights sphere.	May be unattractive to would-be users of the data, reducing reuse of existing data sets.

<sup>22</sup> A similar right (*droit de retrait ou de repentir*) exists for authors in France, Germany, Italy and Spain, and 'probably' in Belgium (Holderness, 1998; Dietz, 1995; Liemer, 2005)

<sup>23</sup> A similar right (*droit de divulgation*) exists for authors in France (Holderness, 1998; Dietz, 1995; Eagles, 2004; Liemer, 2005).

<sup>24</sup> Although in some scientific disciplines, such as palaeontology, an 'unofficial' version of such a right appears to already exist - consider the case of Erik Jarvik and the fossil *Ichthyostega*. Jarvik was one of a team that discovered the fossil in the early 1930s, but only produced his final publication on it in 1996. During that period "... palaeontology's unwritten rules demanded that no-one publish any new information about it until he did". (Zimmer, 1998, pp.50-51)

<sup>25</sup> Perhaps akin to the concept of artists' resale royalty rights (*droit de suite*) which exists in the laws of, amongst other countries, Belgium, France, Germany, Italy, Portugal and Spain, and are incorporated into Article 14ter of the Berne Convention. (Reddy, 1995)

## Conclusion

It will be clear from Table 1 above that developing a ‘Science Commons’ in datasets, which protects the kinds of right which researchers are currently suggesting that they might wish to avail themselves of, whilst providing an open environment for the dissemination of data, might not be an easy task. However, there is no overriding reason why a ‘Science Commons’ should slavishly follow the utilitarian model of copyright espoused by the Common Law tradition and place economic rights at the very centre of the rights it wishes to protect, if this fails to meet the primary requirements of those engaged in the creation of new digital datasets. Many of the rights suggested here have counterparts either in existing UK law - the acknowledgement or paternity right; or in one or more European national legal systems - the right to ensure a work’s integrity, the right to withdraw a work, and the right to decide when to release the work. (Holderness, 1998) There is no reason, when planning the legal infrastructure for a set of international ‘Science Commons’ licences, not to choose to draw from the Civil Law tradition as well. To fail to at least consider doing so, is to ignore the possible benefits of incorporating a legal tradition that, perhaps, corresponds most closely to the way in which researchers aspire to work and collaborate.

## Acknowledgements

This paper is based on a short paper entitled “Bringing Digital Curation into the Legal Limelight” delivered by the author at the 1st International Digital Curation Conference, Bath, UK - 29-30 September 2005.

The Centre for IT & Law is sponsored by Herbert Smith LLP, Vodafone Group Services Ltd, Hewlett Packard Laboratories, Barclaycard, and the Law Society Charitable Trust.

## References

- Bechtold, S. (2004). Digital rights management in the United States and Europe. *American Journal of Comparative Law*, 52, 323-382.
- Beldiman, D. (2005). Book Review. Fundamental rights, author's right and copyright-commonalities or divergences? Review of Geiger, C. Droit d'auteur et droit du public a l'information-Approche de droit compare. (2004). *Columbia Journal of the Law and Arts* 29(1), 39-62.
- Bettig, R.V. (1996). *Copyrighting culture: The political economy of intellectual property*. Boulder, Co: Westview Press.
- Biotechnology and Biological Sciences Research Council (BBSRC). (2005). Developing a BBSRC Data Sharing Policy. - Read comments. Retrieved June 22, 2006, from: [http://www.bbsrc.ac.uk/society/dialogue/consultations/data\\_sharing\\_policy/read.html](http://www.bbsrc.ac.uk/society/dialogue/consultations/data_sharing_policy/read.html)

- Charlesworth, A. (2005). DRM: The straw to break the back of Procrustean approaches to copyright?. In F.W. Grosheide, & J.J. Brinkhof (Eds.), *Intellectual property 2004: Articles on crossing borders between traditional and actual* (pp.405-422). Antwerp: Intersentia.
- David, P. A. (2004). Can "Open Science" be protected from the evolving regime of ipr protections?. *Journal of Institutional and Theoretical Economics*, 160(1), 9-34.
- Davies, G. (2002). *Copyright and the public interest* (2nd ed.). London: Sweet & Maxwell.
- Deazley, R. (2004). *On the origin of the right to copy: Charting the movement of copyright law in eighteenth century Britain (1695-1775)*. Oxford: Hart Publishing.
- DiBona, C., et al. (1999). *Open sources: Voices from the open source revolution*. Sebastopol, CA: O'Reilly
- Dietz, A. (1995). The moral right of the author: Moral rights and the civil law countries. *Columbia-VLA Journal of Law & the Arts*, 19(3/4), 199-227.
- Drahos, P., & Braithwaite J. (2002). *Information feudalism: Who owns the knowledge economy?*. London: Earthscan Publications
- Eagles, I., & Longdin L. (2004). Technological creativity and moral rights: A comparative perspective. *International Journal of Law and Information Technology*, 12(2), 209-236.
- Gellman, R. (1995). Twin evils: Government copyright and copyright-like controls over government information. *Syracuse Law Review*, 45(3), 999-1072.
- Goldstein P. (1994). *Copyright's highway: The law and lore of copyright from Gutenberg to the celestial jukebox*. New York: Hill and Wang.
- Goodman, D. (2004). The criteria for open access. *Serials Review*, 30(4), 258-270.
- Guedon, J-C. (2004). The 'green' and 'gold' roads to open access: The case for mixing and matching. *Serials Review*, 30(4), 315-328.
- Harnad, S., Brody, T., Vallières, F., Carr, L., Hitchcock, S., Gingras, Y., et al. (2004). The access/impact problem and the green and gold roads to open access. *Serials Review*, 30(4), 310-314.
- Holderness, M. (1998). Moral rights and authors' rights. The keys to the information age. *The Journal of Information, Law and Technology*, 1998 (1). Retrieved June 22, 2006 from: [http://www2.warwick.ac.uk/fac/soc/law/elj/jilt/1998\\_1/holderness/](http://www2.warwick.ac.uk/fac/soc/law/elj/jilt/1998_1/holderness/)
- Joint Information Systems Committee (JISC). (2005). Rights in Digital Environments v0.6. Retrieved June 22, 2006, from: [http://www.jisc.ac.uk/uploaded\\_documents/JISC%20Rights%20in%20Digital%20Environment%20Report.pdf](http://www.jisc.ac.uk/uploaded_documents/JISC%20Rights%20in%20Digital%20Environment%20Report.pdf)
- Kahlon, M. (2005). Pain for mother, fetus and you, the reader without access. Retrieved June 22, 2006, from: [http://bayosphere.com/blog/mini\\_kahlon/20050830/pain\\_for\\_mother\\_fetus\\_and\\_you\\_the\\_reader\\_without\\_access](http://bayosphere.com/blog/mini_kahlon/20050830/pain_for_mother_fetus_and_you_the_reader_without_access)

- Laddie, J. (1996). Copyright: Over-strength, over-regulated, over-rated. *European Intellectual Property Review*, 18(5), 253-260.
- Liemer, S. P. (2005). How we lost our moral rights and the door closed on non-economic values in copyright. John Marshall Law School *Review of Intellectual Property Law*, 5(1), Fall 2005. Retrieved June 19, 2006 from: <http://www.jmls.edu/ripl/vol5/issue1/liemer.pdf>
- Lessig, L. (2004). *Free Culture: How Big Media Uses Technology and the Law to Lock Down Culture and Control Creativity*. London: Allen Lane. Retrieved June 22, 2006 from: <http://www.free-culture.cc/freeculture.pdf>
- May, C. (2000). *A global political economy of intellectual property rights: The new enclosures?*. London, Routledge.
- Oppenheim, C. (2005). Editorial: Open access and the UK Science and Technology Select Committee Report free for all?. *Journal of Librarianship and Information Science*, 37(1), 3-6.
- Perelman, M. (2002). *Steal this idea: Intellectual property rights and the corporate confiscation of creativity*. Basingstoke: Palgrave Macmillan.
- Reddy, M. B. (1995). The droit de suite: Why American fine artists should have the right to a resale royalty. *Loyola of Los Angeles Entertainment Review* 15, 509-546.
- Rhodes, R. A. W. (1994). The hollowing out of the state: The changing nature of the public service in Britain. *Political Quarterly*, 65(2), 138-151.
- Rose M. (1993). *Authors and owners: The invention of copyright*. Cambridge, MA: Harvard University Press
- Rosen, L. (2004). *Open source licensing: Software freedom and intellectual property law*. Prentice Hall Professional Technical Reference
- Sell, S.K. (2003). *Private power, public law: The globalisation of intellectual property rights*. Cambridge: Cambridge University Press.
- Thierer, A. D., & Crews, W. (Eds.). (2002). *Copy fights: The future of intellectual property in the information age*. Washington, DC: Cato Institute.
- Vaidhyanathan, S. (2001). *Copyrights and copywrongs: The rise of intellectual property and how it threatens creativity*. New York: New York University Press.
- van Wendel de Joode, R. (2003). *Protecting the Virtual Commons: Self-organising open source and free software communities and innovative intellectual property regimes*. Cambridge: Cambridge University Press.
- Vaver, D. (1999). Moral rights yesterday, today and tomorrow. *International Journal of Law & Information Technology*, 7(3), 270-278.
- Waelde, C., & McGinley M. (2005). Public domain; public interest; public funding: Focussing on the 'three Ps' in scientific research. *SCRIPT-ed*, 2(1), 83. Retrieved June 22, 2006 from: <http://www.law.ed.ac.uk/ahrb/script-ed/vol2-1/3ps.asp>
- Zimmer, C. (1998). *At the water's edge: Macroevolution and the transformation of life*. New York: Simon & Schuster.