What Computational Scientists Need to Know About Intellectual Property Law: A Primer

Victoria Stodden[1]

Abstract

As computation takes a central role in scientific research, the production of digital scholarly objects has new implications for Intellectual Property Law. The goal of this chapter is to provide a "rough guide" to IP Law for scientists who work with computers and are making available datasets and code, as well as the published article. These digital scholarly objects are potentially subject to copyright and patent law, and this chapter attempts to disentangle the various options available to scientists who practice really reproducible research and share their code and data. The basics of copyright law are explained in the context of scientific research, and options such as the various Creative Commons licenses, open licensing for software, and permissioning for dataset re-use. This chapter addresses the three primary digital research outputs in turn, the manuscript (including open access publishing), the code, and the data. It ends with citation recommendations for each of these, in particular for software and data.

Introduction

Data and code are becoming as important to research dissemination as the traditional manuscript. For computational science the evidence is clear: it is typically impossible to verify scientific claims without access to the code and data that generated published findings. Gentleman and Lang [1] introduced the notion of the "Research Compendium" as the unit of scholarly communication, a triple including the explanatory narrative, the code, and the data used in deriving the results. One of the reasons for including the code and data is to facilitate the production of *really reproducible research*, a phrase coined by Jon Claerbout in 1991[2] to mean research results that can be regenerated from the available code and data. Claerbout's approach was paraphrased by Buckheit and Donoho [2] as follows:

> The idea is: *An article about computational science in a scientific publication is **not** the scholarship itself, it is merely **advertising** of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions which generated the figures.*

[1] Victoria wishes to thank an anonymous reviewer for many extremely helpful comments.

[2] See http://sepwww.stanford.edu/doku.php?id=sep:research:reproducible for the Stanford Exploration Project's pioneering recommendations for reproducible research.

Enabling computational replication typically means supplying the data, software, and scripts, including all parameter settings, that produced the results. [3, 4]. This approach runs headlong and unavoidably into current Intellectual Property law, which creates a stumbling block rather than an impassable barrier to the dissemination of really reproducible research. In this chapter I describe these Intellectual Property stumbling blocks to the open sharing of computational scientific knowledge and present solutions that coincide with longstanding scientific norms. In Section 1, I motivate scientific communication as a narrative with a twofold purpose: to communicate the importance of the findings within the larger scientific context and to provide sufficient information that the results may be verified by others in the field. Sections 2 and 3 then discuss Intellectual Property barriers and solutions that enable code and data sharing respectively. Each of these three research outputs, the research article, the code, and the data, require different legal analyses and action in the scientific context as described below. The final section discusses citation for digital scholarly output, focusing on code and data.

A widely accepted scientific norm, as labeled by Robert K. Merton, is *Communism* or *Communalism* [5]. By this Merton meant that property rights in scientific research extend only to the naming of scientific discoveries (Arrow's Impossibility Theorem for example, named for its originator Kenneth Arrow), and *all other intellectual property rights* are given up in exchange for recognition and esteem. This notion, at least in the abstract, underpins the current system of publication and citation that forms the basis for academic promotion and reward.

Computational science today is facing a credibility crisis: without access to the code and data that underlie scientific discoveries, published findings are all but impossible to verify [4]. Reproducible computational science has attracted attention since Claerbout wrote some of the first really reproducible manuscripts in 1992.[3] More recently, a number of researchers have adopted reproducible methods [2, 6, 7] or introduced them in their role as journal editors [8, 9, 10]. This chapter discusses how Intellectual Property Law applies to data in the context of communicating scientific research.

1. *Publishing the Research Article*

Scientific publication has taken the well-recognized form of the research article since 1665 with the first issue of the Philosophical Transactions of the Royal Society of London.[4] This section motivates the sharing of the research paper, and discusses the clash that has arisen between the need for scientific dissemination and modern intellectual property law in the United States.

Scientific results are described in the research manuscript, including their derivation and context, and this manuscript is typically published in an established academic journal. It is of primary importance that the body of scientific knowledge, comprised of journal publications, have as little error as possible. This is in part accomplished through peer review, and in part through the very act of publication and permitting a wide audience access to the work. The recognition that the scientific

[3] See http://sepwww.stanford.edu/doku.php?id=sep:research:reproducible

[4] For a brief history see http://rstl.royalsocietypublishing.org/ including an image of the first issue with the endearing title "Philosophical Transactions Giving Some Account of the Present Undertakings, Studies, and Labours of the Ingenious in Many Considerable Parts of the World."

research process is error prone, that error can creep in anywhere and from any source, is central to the scientific method and wider access to the findings increases the chances that errors will be caught.

The second reason property rights have been eschewed in scientific research is the understanding that scientific knowledge about our world, such as physical laws, mathematical theorems, or the nature of biological functions, are to be discovered, not invented created, and this knowledge belongs to all of humanity. This is not to say scientific discovery is not a creative act, quite the contrary, but that the underlying scientific fact is a public good, a facet of our world not subject to ownership. This is the underlying rationale behind U.S. federal government grants of over $50 billion dollars for scientific research in 2012 [11]. This vision is also reflected both in the widespread understanding of scientific facts as "discoveries" and not "inventions," and in current intellectual property law which does not recognize a scientific discovery as rising to the level of individual ownership, unlike an invention or other contribution. We will see this notion rise again in the discussion on scientific data.

Copyright law in the United States originated in the Constitution, stating that "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."[5] Through a series of subsequent laws, copyright has come to assign a specific set of rights to authors of original expressions of ideas *by default*. In the context of scientific research, this means that the written description of a finding is automatically copyrighted by the author(s) (how copyright applies to data and code is discussed in the following two sections). Copyright secures exclusive rights vested in the author to both reproduce the work and prepare derivative works based upon the original. There are exceptions and limitations to this power, such as Fair Use, but these do not provide for an intellectual property framework for scientific knowledge that matches longstanding scientific norms of openness, access, and transparency.

Intellectual Property law, and its interpretation by academic and research institutions, means that authors have copyright over their research manuscripts. Copyright can be transferred to others and the copyright holders can grant permissions for use to others as they see fit. In a system established many decades ago journals typically request that copyright be assigned to the publisher for free, rather than remain with the authors, as a condition of publication. Many journals have a second option for authors if they request it, where copyright remains with the author but permission is granted to the journal to publish the article.[6] If copyright was transferred, access to the published article usually involves paying a fee to the publisher. Typically scientific journal articles are available only to the privileged few affiliated with a university library that pays the journal subscription fees, and articles are otherwise offered for a surcharge of about $30 each. Authors of scientific articles, and the owners of copyright, typically transfer copyright to publishers as a condition of publication.

Publishing scientists today have other options. A transformation is underway that has the potential to make scientific knowledge openly and freely available. The *open access movement* has established ways of publishing that secure long term public access to the research article. This may still involve the journal requesting a transfer

[5] U.S. Const. art. I, §8, cl. 8.

[6] See for example Science Magazine's alternative license at http://www.sciencemag.org/site/feature/contribinfo/prep/lic_info.pdf (last accessed January 29, 2013).

of copyright to them, and it usually involves an upfront fee to compensate the journal for the loss of revenue from library subscriptions and article purchases.

This transformation started in 1991 when Paul Ginsparg, Professor of Physics at Cornell University, set up an open repository called arXiv.org (pronounced "archive") for physics articles awaiting journal publication. In the biosciences, the Public Library of Science, PLoS, was launched 2000.[7] They publish under a new model, open access publishing, which publishes scientific articles by charging the authors the costs upfront, typically about $2000 per article, and making the published papers freely available online.[8]

On balance openly available articles appear to be cited at higher rates that those behind subscription paywalls [12, 12a]. There are steps a researcher can take when publishing a manuscript, to help maximize the future access to their article. First, a researcher can request the alternative copyright agreement, that gives the journal permission to publish the article but leaves copyright with the author. Another approach is to use the SPARC addendum, to retain rights to post the article on the author's webpage, in scholarly repositories, or more widely on the Internet.[9] The SPARC addendum, for example, ensures the right of the author to retain:

(i) the rights to reproduce, to distribute, and to publicly display the Article in any medium for noncommercial purposes;
(ii) the right to prepare derivative works from the Article; and
(iii) the right to authorize others to make any non-commercial use of the Article so long as Author receives credit as author and the journal in which the Article has been published is cited as the source of first publication of the Article. For example, Author may make and distribute copies in the course of teaching and research and may post the Article on personal or institutional Web sites and in other open-access digital repositories.

These are valuable rights authors likely wish to retain so they can re-use their own work and share with others, and this can be accomplished by using the SPARC addendum with the traditional publisher's agreement.

A second option is choosing to publish in Open Access journals. This is a personal decision for the authors as journal impact factor is often tied to career advancement, but open access journals like PLoSONE have been gaining in prestige.[10]

When publishing in an open access journal, authors are sometimes asked to designate a Creative Commons license for their article. Authors can also find themselves confronted with this choice when depositing to a repository, or even when posting the article on their own webpage, depending on the downstream use they wish to permit. Creative Commons licenses are very useful for researchers, and I will discuss their various licensing options. In the Creative Commons sense, "license" is the term used to mean that an owner gives advance permission for use of his or her copyrighted works. Although related this is a different sense of the term than, say, a software license or patent license that is paid for and permits use of the software or

[7] See http://blogs.plos.org/plos/2011/11/plos-open-access-collection-%E2%80%93-resources-to-educate-and-advocate/ for a collection of articles on Open Access.

[8] See http://www.plos.org/publish/pricing-policy/publication-fees/ for up-to-date pricing information.

[9] See http://www.arl.org/sparc/author/addendum.shtml

[10] See http://scholarlykitchen.sspnet.org/2011/06/28/plos-ones-2010-impact-factor/ for recent impact factor information.

patent for a period of time. In this case license refers to the granting of certain uses by the copyright holder in advance – without charge to anyone – so there is no need to contact the copyright holder to request permission.

Creative Commons has provided documents (licenses) that encode certain terms of use in formal legal language, making it easy for researchers and others to grant permission for use of their work if they happen to want what these licenses provide. The most basic Create Commons license is "CC-BY" and, essentially, it permits unrestricted downstream use so long as attribution is given to the original author. Note that in this case the author is also the copyright holder. Licensing options that grant permission for use can only be applied by the copyright holder (or with the copyright holder's permission), so think carefully before signing your copyright over to other entities, such as journals.

CC-BY is the closest permission structure to that which scientists and researchers are used to – essentially saying, use my work however you wish, but make sure you credit me.[11] Creative Commons, however, designed licenses with a broader community in mind and offers other licensing options. For certain specialized scientific research these may be useful, so I touch on them here for completeness, but each option adds further restrictions over CC-BY that I believe should be outweighed by their benefits over CC-BY. Creative Commons has licenses that restrict downstream use to noncommercial purposes only (NC), that forbid the creation of derivative works (ND), and direct downstream users as to what license they must use on their work (SA). The simplest choice that matches scientific community norms is CC-BY.

With broader sharing of publications, scientific knowledge could be spread more widely, more mistakes caught, and the rate of scientific progress improved. In addition, more downstream activity would be encouraged, such as technological development, industry growth, and further scientific discoveries [13, 14]. Open archiving is mandated by the National Institutes for Health, where published articles arising from NIH funded research must be deposited in PubMed Central[12] within 12 months of publication. On February 22, 2013, this was extended to all federal funding agencies through an Executive Memorandum released by the Office of Science and Technology Policy in the Whitehouse.[13] To maximize access we need a streamlined and uniform way of managing copyright over scientific publications, and also copyright on data and code, as discussed in the next section.

Section 2. *Publishing Scientific Software, Code, and Tools*

The computational steps taken to arrive at a result are often complex enough that their complete communication is prohibitive in a typical scientific publication. This is a key reason for releasing the code that contains all the steps, instructions, data calls, and parameter settings that generated the published findings. Of the three digital scholarly objects discussed in this chapter, code has the most complex interactions with Intellectual Property Law since it is both subject to copyright and patent.

---

[11] See http://creativecommons.org/licenses/by/3.0/

[12] PubMed Central is located at http://www.ncbi.nlm.nih.gov/pmc/ .

[13] See http://www.whitehouse.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research

Software is considered an original expression of an underlying idea, and therefore it is subject to copyright. As discussed in the previous section copyright adheres by default – a programmer who does nothing other write software will produce code copyrighted to herself.[14] The algorithm or methods that the code implements are not subject to copyright themselves, but copyright adheres to the code that implements the algorithm or methods. The effect of copyright in this case is the prohibition on others to reproduce or modify the code.[15] (See Box 1)

Box 1 Inset: Copyright in a Nutshell

The original expression of ideas falls under copyright by default (text, code, figures, tables, original selection and arrangement of data)

Subject to some exceptions and limitations, copyright secures exclusive rights vested in the author to:

1. reproduce the work,
2. prepare derivative works based upon the original

Copyright is of limited but long duration, generally life of the author plus 70 years.

Copyright works counter to longstanding scientific norms that encourage re-use and verification of results. This means running the code on a different system (reproducing) or adapting the code to a new problem (re-using). Authors must grant permission to others to use their code in these ways. The Creative Commons licenses discussed in the previous section were created for digital artistic works and they are not suitable for code, and so cannot solve our problem. There are, however, a great number of open licenses for software that permit authors to permission the code for replication and re-use. Software exists primarily in two forms, source and compiled, and transmission of the complied form alone is not sufficient for scientific purposes. Communication of the source code, whether intended to be compiled or not, is essential to understanding and re-using scientific code. In the context of scientific research, source code is often in the form of scripts, for example in MATLAB or Python, that execute in association with an installed package and are not compiled.

There are several open licenses for code that place few restrictions on re-use beyond attribution, creating an Intellectual Property framework resembling conventional scientific norms. The (Modified) Berkeley Software Distribution (BSD) license for example permits the downstream use, copying, and distribution of either unmodified or modified source code, as long as the license accompanies any distributed code and the previous authors' names are not used to promote any modified downstream software. The license is brief enough it can be included here:

[14] Although the exception in academic research, the copyright can initially go to an employer or commissioning party under the "work made for hire" doctrine.

[15] There are exceptions and limitations to copyright, such as Fair Use, but these do not extend to scientific scholarly objects and how researchers would typically use them. From a computational researcher's perspective, these exceptions and limitations should not be relied on to provide sufficient access and affirmative steps such as licensing should be taken. For more on Fair Use see http://www.copyright.gov/fls/fl102.html and [15, 16].

Copyright (c) <YEAR>, <OWNER>
All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

- Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
- Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
- Neither the name of the <ORGANIZATION> nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

This text is followed by a disclaimer releasing the author from liability for use of the code. The Modified BSD license is very similar to the MIT license, with the exception that the MIT license does not include a clause forbidding endorsement. The Apache 2.0 license is also commonly used to specify terms of use on software. Like the Modified BSD and MIT licenses, the Apache license requires attribution but it differs in that it permits users to exercise patent rights that would otherwise only extend to the original author, so that a patent license is granted for any patents needed for use of the code (probably a fairly obscure situation for academic research). The Apache 2.0 license further stipulates that the right to use the software without patent infringement will be lost if the downstream user sues the licensor for patent infringement. Attribution under Apache 2.0 requires that any modified code carries a copy of the license, with notice of any modified files and all copyright, trademark, and patent notices that pertain to the work be included. Attribution can also be done in the notice file. The *Reproducible Research Standard* [17, 18] recommends using one of these three licenses or a similar attribution license for scripts and software released as part of a scientific research compendium.

Patents are a second form of intellectual property that can create a barrier to the open sharing of scientific codes. Columbia University for example states in its Faculty Handbook that,

> … the University and a member of the faculty may expect and require of one another cooperation in the development and exploitation of conceptions … In particular, the University will advise a faculty member about securing a patent, and will participate with him or her in seeking patent protection, in every way compatible with their several capacities and common interests. … The obligations of a faculty member include the execution of an assignment or a patent, and of rights thereunder, in appropriate circumstances.[16]

There are exceptions, but this expectation of patenting is typical in academic research institutions.

Patenting is often viewed as a method of enabling access, especially by institutional technology transfer offices, to technology that would otherwise remain inaccessible in academic institutions and research journals. In the case of software, patents add a

[16] See http://www.columbia.edu/cu/vpaa/handbook/appendixd.html (last accessed Feb 12, 2013).

layer of complexity, and possible fees, to the scientific notion of reproducibility of results. Reproducibility implies the open availability of the software that permits replication along with the published results (Gentleman and Lang's *Research Compendium* introduced previously in this chapter). Researchers seeking a patent appear to be reluctant to release their code publicly, possibly for fear of creating "prior art" and thus creating a barrier to patent granting, or a perceived loss of revenue from researchers who would like to use their software for research purposes [19].

Neither of these reasons should prevent a patent-seeking researcher from making his or her code publicly and openly available. Under U.S. law, an inventor or rights holder can apply for a patent on a published invention, so long as it is within one year of disclosure.[17] A dual system of patent licensing for industry application can co-exist with openly downloadable software for academic research purposes. If a researcher feels inclined to pursue a patent on software, he or she should ensure that academic researchers are able to openly and easily download the software, without going through a patent licensing process (even one without a fee) in accordance with the Principle of Scientific Licensing, which states [17]:

> ***Principle of Scientific Licensing****: Legal encumbrances to the dissemination, sharing, use, and re-use of scientific research compendia should be minimized, and require a strong and compelling rationale before application.*

Code can be made available in a dedicated code repository such as GitHub, BitBucket, SourceForge, or RunMyCode [7].[18] All will provide links to the stored code, permitting it to be associated with the manuscript and data. This theme of accessibility of research compendia continues in the next section with a discussion on publishing the data associated with scientific findings.

Section 3. *Publishing Datasets and "Raw Facts"*

Data is understood as integral in the communication of computational findings, part of the *Research Compendium* introduced earlier in the chapter. Data can refer to an input into scientific analysis, such as a publicly available dataset like those at Data.gov[19] or those gathered by researchers in the course of the research, or it can refer to the output of computational research, as is the case in computational simulations. In short, it is typically an array of numbers or descriptions, to which analysis and interpretation is applied. It does not include computer code, discussed in the previous section.

In 1991 the U.S. Supreme Court held in Feist v. Rural Telephone Service Co. that raw facts are not copyrightable but the original "selection and arrangement" of these raw facts may be.[20] [21] The Supreme Court has not made a ruling concerning

---

[17] This is known as a "statutory bar" to an otherwise valid patent.

[18] See https://github.com/ , https://bitbucket.org/ , http://sourceforge.net/ , and http://www.runmycode.org/

[19] See https://explore.data.gov/

[20] Copyright does extend to databases under European Intellectual Property Law. This is a key distinction between European and U.S. Intellectual Property systems in the context of scientific research.

[21] See Feist Publications v. Rural Telephone Service Co., 499 U.S. 360 (1991).

Intellectual Property in data since and modern computational research may create a residual copyright in a particular dataset, if original selection and arrangement of facts takes place. Collecting, cleaning, and readying data for analysis is often a significant part of scientific research and arguably could be considered "original selection and arrangement" in the sense of Feist.

The *Reproducible Research Standard* recommends therefore releasing data under a Creative Commons CC0, or "no rights reserved" publication, in part because of the possibility of such a residual copyright existing in the dataset.[22] The public domain certification means that as the dataset author, and potential copyright holder, you will not exercise any rights you may have in the dataset that may derive from copyright (or any other ownership rights). A public domain certification also means that as the author you are relying on downstream users to cite and attribute your work appropriately. For this reason, a specific citation recommendation should be included with the dataset, suggesting to downstream users that they cite any use of the dataset itself.

Datasets may have barriers to re-use and sharing that do not stem from Intellectual Property Law, such as confidentiality of records, privacy concerns, and proprietary interests from industry or other external collaborators that may assert ownership over the data. Good practice suggests planning for maximal data release at the time of publication at the beginning of a research collaboration, whether it might be with industrial partners who may foresee different uses for the data than supporting reproducible research, or with scientists subject to a different Intellectual Property framework for data, such as those in Europe.

Datasets should be made available in recognized repositories for the field, if they exist, and conform to any established standards for formats, meta-data, or exposition. If recognized repositories don't exist, both The DataVerse Network and Dryad will host datasets from any field, for example, and provide association with the manuscript and code through persistent links.[23] They are able to accommodate access restriction on the datasets, due to privacy concerns or other constraints. A number of federal funding agencies have data sharing requirements in their grant guidelines. The National Science Foundation grant guidelines state that "Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants.[24] Similarly, the National Institutes for Health grant guidelines state that "The NIH expects and supports the timely [no later than the acceptance for publication of the main findings from the final data set] release and sharing of final research data from NIH-supported studies for use by other researchers."[25] These guidelines have been minimally enforced but this may change. The February 22, 2013 Executive Memorandum mentioned above requires federal funding agencies to develop enforceable open data plans.

Citation

[22] See http://creativecommons.org/about/cc0 for further details on the CC0 license.

[23] See http://thedata.org/ and http://datadryad.org/ .

[24] See http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/aag_6.jsp .

[25] The NIH data sharing guidelines apply to grants greater than $500,000. See http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html .

The research article, code, and data are shared with the hope that they will be used by other researchers. Citation of data and software use is not standard in the computational sciences and must become so. Aside from being a plagiarism violation [20], using uncited code and data is poor scientific practice and it impedes both transparency in research and rewards for scientific contributions [21]. When sharing code or data, it is helpful to provide citation information both to guide downstream users and to remind users that citation is expected.

Throughout this chapter the use of open attribution-only licensing has been recommended, but it is worth commenting on the relationship between this legal concept and traditional academic citation. They are not identical. In the case of open software licensing as discussed in this chapter, attribution generally refers to listing contributions and authors in a file that accompanies the software. This is important for provenance and transparency, but doesn't satisfy citations standards used in academic rewards. Some open licenses require this type of attribution, but it must be noted that additional, not legal, citation should take place to satisfy scientific norms. Any software use should receive a scientific citation in the list of references, on a par with referenced publications. A footnote mentioning the software use is not adequate. The content of this citation should include, at minimum: the author(s); the software version; the location of the code on the Internet; the date of software release, and the data of software access. If the authors suggest further citation information, for example a report describing the software, this should be cited.

In the case of Creative Commons attribution licensing, the two concepts lie slightly closer. Section 4(b) of the CC-BY 3.0 license states that,

> If You Distribute … the Work or any Adaptations …, You must … keep intact all copyright notices for the Work and provide, reasonable to the medium or means You are utilizing: (i) the name of the Original Author … (ii) the title of the Work … (iii) to the extent reasonably practicable, the URI, if any… and (iv) … in the case of an Adaptation, a credit identifying the use of the Work in the Adaptation. … The credit required by this Section 4 (b) may be implemented in any reasonable manner…[26]

Arguably, what is "reasonable to the medium" in the research context is scientific citation. The CC-BY license is most likely to be applied to the research paper itself, for which citation practices exist, but if applied to text describing data selection and arrangement, for example, it could be interpreted as requiring standard scientific citation. Hopefully the research community quickly adopts practices that include code and data citation as standard, and legal requirements remain a last resort.

Conclusion

The current set of scientific norms evolved over hundreds of years to maximize the integrity of our stock of scientific knowledge. They espouse standards of independent verification and transparency, and publication of research findings to disseminate the knowledge widely. Current scientific practice has not kept up with technological advancement, meaning much of the published computational findings are unreplicable

[26] Note that CC-BY 4.0 has now been publicly released for comment. http://creativecommons.org/weblog/entry/36713 .

since the source code and data are not made conveniently and routinely available. To make reproducibility possible in today's computational research environment, the communication of new types of scholarly objects, for example a digital research paper, code, or data, requires engaging Intellectual Property law. In this chapter I have traced how Intellectual Property Law interacts with digital scholarly communication, through both the relevant aspects of the copyright and patent systems, for scholars sharing really reproducible computational research.

For broad re-use, sharing, and archiving of code to be a commonly accepted practice in computational science, it is important that open licenses be used that minimize encumbrances to access and re-use, such as attribution only licenses like the MIT license or the Modified BSD license, or the Creative Commons attribution license. A collection of code with an open licensing structure permits archiving, persistence of the code, and research on the code base itself, just as is the case for collections of research articles. For these reasons, as well as the integrity of our body of scholarly knowledge, it is essential to address the barriers created by current Intellectual Property Law in such a way that access and re-use are promoted and preserved, and future research encouraged.

References

[1] R. Gentleman and D. Temple Lang, "Statistical Analyses and Reproducible Research," 2004. http://biostats.bepress.com/bioconductor/paper2/

[2] D. Donoho and J. Buckheit, "WaveLab and reproducible research," Stanford Department of Statistics Technical Report 474, 1995.

[3] G. King, 1995. Replication, Replication. PS: Political Science and Politics 28: 443–499. Copy at http://j.mp/jCyfF1

[4] D. Donoho, A. Maleki, M. Shahram, I. Ur Rahman, V. Stodden, "Reproducible research in computational harmonic analysis," *Computing in Science and Engineering*, 11, January 2009.

[5] Merton, R. K. (1973), "The Normative Structure of Science", in Merton, Robert K., *The Sociology of Science: Theoretical and Empirical Investigations*, Chicago: University of Chicago Press, ISBN 978-0-226-52091-9, OCLC 755754

[6] D. Donoho, V. Stodden, Y. Tsaig, "About SparseLab," 2007. see http://www.sparselab.edu

[7] V. Stodden, C. Hurlin, C. Perignon, "RunMyCode.Org: A Novel Dissemination and Collaboration Platform for Executing Published Computational Results," eSoN IEEE eScience Conference 2012. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2147710

[8] Journal of Experimental Linguistics, Linguistic Society of America, http://elanguage.net/journals/jel;

[9] Biostatistics, Oxford Press. http://biostatistics.oxfordjournals.org/;

[10] R. Trivers, http//:www.psychologytoday.com:blog:the-folly-fools:201205:fraud-disclosure-and-degrees-freedom-in-science" http://www.psychologytoday.com/blog/the-folly-fools/201205/fraud-disclosure-and-degrees-freedom-in-science," 2012, APA

[11] "The U.S. Science Budget," The ScienceInsider, 14 Feb 2011. Available at http://news.sciencemag.org/scienceinsider/budget_2012/

[12] Y. Gargouri, Hajjem C., Larivière V., Gingras Y., Carr L., et al. (2010) Self-Selected or Mandated, Open Access Increases Citation Impact for Higher Quality Research. PLoS ONE 5(10): e13636. doi:10.1371/journal.pone.0013636

[12a] M. McCabe, "Online Access and the Scientific Journal Market: An Economist's Perspective," commissioned paper for Copyright in the Digital Era: Building Evidence For Policy, National Academies of Science Report, forthcoming. Available at http://sites.nationalacademies.org/PGA/step/copyrightpolicy/index.htm

[13] V. Stodden, "Innovation and Growth through Open Access to Scientific Research: Three Ideas for High-Impact Rule Changes," in Rules for Growth: Promoting Innovation and Growth Through Legal Reform, edited by The Kauffman Task Force on Law, Innovation, and Growth. February, 2011.

[14] V. Stodden, ""Open Science: Policy Implications for the Growing Phenomenon of User-Led Scientific Innovation," Journal of Science Communication, 9(1), 2010.

[15] W. Fisher III, "Reconstructing the Fair Use Doctrine," Harvard Law Review, Vol. 101, No. 8 (Jun., 1988), pp. 1659-1795

[16] Paul A. David, "The Economic Logic of 'Open Science' and the Balance between Private Property Rights and the Public Domain in Scientific Data and Information: A Primer" SIEPR Discussion Paper No 02-30, 2005. http://ideas.repec.org/p/wpa/wuwpdc/0502006.html

[17 -> 12b] Victoria Stodden. Enabling reproducible research: Licensing for scientific innovation. *International Journal of Communications Law and Policy,* pages 1–25, 2009.

[18] V. Stodden, "The Legal Framework for Reproducible Research in the Sciences: Licensing and Copyright", IEEE Computing in Science and Engineering, 11(1), January 2009, p.35-40.

[19] V. Stodden. "The Scientific Method in Practice: Reproducibility in the Computational Sciences," MIT Sloan Research Paper No. 4773-10,

[20] National Research Council. *Responsible Science, Volume I: Ensuring the Integrity of the Research Process*. Washington, DC: The National Academies Press, 1992.

[21] National Research Council. *For Attribution -- Developing Data Attribution and Citation Practices and Standards: Summary of an International Workshop*. Washington, DC: The National Academies Press, 2012.