

Legal Approaches for Open Access to Research Data

Catherine Doldirina, Anita R. Eisenstadt, Harlan Onsrud, and Paul F. Uhler

Authors of this paper in alphabetical order are: Catherine Doldirina, Independent Consultant, via Niccodemi 8, Apt 19, 20156 Milan, ITALY; Anita Eisenstadt, IDA Science and Technology Policy Institute, 1899 Pennsylvania Avenue, NW, Suite 520, Washington, DC 20006, USA; Harlan Onsrud, University of Maine, School of Computing and Information Science, Orono, ME 04469, USA; and Paul F. Uhler, National Academy of Sciences, 4544 State Route 17B, Callicoon, NY 12723, USA. The views expressed here are those of the authors and not necessarily those of their employing institutions. We would like to thank the following individuals for their generous assistance. Reviewers who provided their insights on an informal basis included: Hal Abelson, Jorge Contreras, Willi Egloff, Brett Frischmann, Lucie Guibault, Michael Madison, Thomas Margoni, Jerome Reichman, Charles Schweik, and Peter Suber.

We also would like to thank Linda Garlet for her help with the editorial work. Finally, we want to acknowledge the staff at [journal publication] for their support in publishing this article.

Table of Contents

CONTENTS

1	INTRODUCTION	4
1.1	Background and Objectives.....	4
1.2	Key Terms and Concepts Used in this Article	9
2	PROTECTION OF DATA BY INTELLECTUAL PROPERTY LAWS	13
2.1	Copyright.....	14
2.1.1	<i>Summary of Key International Principles</i>	14
2.1.2	<i>Differences in National Implementation of Copyright Principles</i>	16
2.1.3	<i>Application of Copyright to Factual Data</i>	18
2.1.4	<i>Pragmatic Copyright Considerations for Research Data</i>	19
2.2	<i>Sui Generis</i> Database Rights and Other Forms of Intellectual Property Rights	20
2.3	Other Forms of Intellectual Property Rights	22
3	REMOVING INTELLECTUAL PROPERTY LAW RESTRICTIONS ON THE USE OF RESEARCH DATA.....	23
3.1	Governmental Mechanisms	24
3.1.1	<i>Treaties and Agreements</i>	24
3.1.2	<i>Policies of Intergovernmental Organizations</i>	26
3.1.3	<i>National Legislation, Regulations, and Research Data Policies</i>	28
3.2	Voluntary “Self-Help” Approaches.....	31
3.2.1	<i>Standard Waivers of Rights</i>	32
3.2.2	<i>Standard Common-Use Licenses</i>	32
3.2.3	<i>Private Contracts and Agreements</i>	34
3.2.4	<i>Institutional Policies</i>	35
4	CONCLUSIONS AND RECOMMENDATIONS	37
4.1	Legal Mechanisms that Promote Open Access and Legal Interoperability	37
4.1.1	<i>Government Mechanisms</i>	40
4.1.2	<i>Mechanisms for Non-Governmental Institutions</i>	41
4.2	Recommendations for Open Access Research Data Repositories	42
4.3	Recommendation for a Global Research Data Commons.....	43

Appendix A: Compendium of International Declarations, Principles, and Policies Promoting Open Access to Data..... 45

Appendix B: Selected Government Instruments and Policies Promoting Open Access to Data in the United States and European Union..... 51

1 INTRODUCTION

1.1 Background and Objectives

The digital network revolution has enabled unprecedented approaches to better accomplish many tasks in all facets of our daily lives.¹ New business models, social networks, modes of collaborating in science and technology, and boundless other forms of knowledge production are now being used by large segments of society.² The explosion of data and information, often labeled “big data,” has created extraordinary possibilities for sharing data across boundaries—whether geographic, disciplinary, or institutional.³

The sharing of information, and particularly data, has resulted in countless benefits.⁴ Making data more readily available to researchers and other users worldwide enables the data to be used in new and innovative ways.⁵ It has accelerated scientific discovery and innovation.⁶ Data sharing furthers the scientific principles of transparency and replicability by enabling scientists to replicate experiments and verify the results. Data sharing also promotes new forms of collaborative knowledge production such as citizen science and crowdsourcing.⁷

¹ For example, the OECD’s Data-Driven Innovation Report finds that the generation and use of data are redefining our knowledge capacity and that data-driven innovation has become a key pillar of 21st century growth, with the potential to enhance productivity, economic development, and social wellbeing. OECD, DATA-DRIVEN INNOVATION: BIG DATA FOR GROWTH AND WELL-BEING (OECD Publishing 2015), <http://www.oecd.org/sti/data-driven-innovation-9789264229358-en.htm>.

² For a general discussion of such networked strategies, *see, e.g.*, JOE WEINMAN, DIGITAL DISCIPLINES: ATTAINING MARKET LEADERSHIP VIA THE CLOUD, BIG DATA, SOCIAL, MOBILE, AND THE INTERNET OF THINGS (Wiley 2015). *See also*, YOCHAI BENKLER, THE WEALTH OF NETWORKS—HOW SOCIAL PRODUCTION TRANSFORMS MARKETS AND FREEDOM (Yale University Press 2006), http://www.benkler.org/Benkler_Wealth_Of_Networks.pdf.

³ For an analysis of the big data phenomenon, *see* VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK (Eamon Dolan/Mariner Books 2014). *But see*, CHRISTINE L. BORGMAN, BIG, DATA, LITTLE DATA, NO DATA (The MIT Press 2015), <https://mitpress.mit.edu/big-data>.

⁴ For a broad overview of the many benefits of open data that summarizes and references dozens of reports and articles on this theme, *see* PAUL F. UHLIR, THE VALUE OF OPEN DATA SHARING, REPORT FOR THE GROUP ON EARTH OBSERVATIONS (CODATA 2015), https://www.earthobservations.org/documents/geo_xii/GEO-XII_09_The%20Value%20of%20Open%20Data%20Sharing.pdf,

⁵ THE FOURTH PARADIGM: DATA-INTENSIVE SCIENTIFIC DISCOVERY (Tony Hey, Stewart Tansley, and Kristin Tolle eds. Microsoft Research 2009), <http://research.microsoft.com/en-us/collaboration/fourthparadigm/>. *See also*, THE ROYAL SOCIETY, SCIENCE AS AN OPEN ENTERPRISE (2012), https://royalsociety.org/~media/royal_society_content/policy/projects/sape/2012-06-20-saoe.pdf; and Paul F. Uhlir, *The Emerging Role of Open Repositories as a Fundamental Component of the Public Research Infrastructure*, in OPEN ACCESS: OPEN PROBLEMS, 59 (G. Sica, ed., Polimetrica International Scientific Publisher 2006), <http://eprints.rclis.org/9656/1/OpenAccess.pdf>.

⁶ Using the Human Genome Project as an illustrative example, *see* Heidi L. Williams, *Intellectual Property Rights and Innovation: Evidence from the Human Genome*, JOURNAL OF POLITICAL ECONOMY 121:1 (2013), <http://economics.mit.edu/files/8647> and SIMON TRIPP & MARTIN GRUEBER, ECONOMIC IMPACT OF THE HUMAN GENOME PROJECT (Battelle 2011), http://www.battelle.org/docs/default-document-library/economic_impact_of_the_human_genome_project.pdf.

⁷ *See, e.g.*, LIOR ZOREF, MINDSHARING: THE ART OF CROWDSOURCING EVERYTHING (Penguin Random House 2015). *See also*, JAMES ROBBINS, *CROWDSOURCING FOR THE BIRDS*, N.Y. TIMES, AUG. 19, 2013, http://www.nytimes.com/2013/08/20/science/earth/crowdsourcing-for-the-birds.html?pagewanted=all&_r=0. For a strategic investment initiative in citizen science and crowdsourcing supported by NSF and the White House, *see*

Similarly, openly networked data have enabled individuals to improve their daily lives, access educational opportunities, and become more informed consumers.⁸ Open data have enhanced trust and greater transparency of governance, and enabled more active participation in the political process.⁹ A policy based on openness to public information resources provides policymakers with more comprehensive access to factual data as a foundation for better informed planning.¹⁰

Open data sharing also supports broad economic benefits and growth. In the economic sphere, new businesses have been created using public information to develop value-enhanced products and to use data in ways that were never envisioned by their original producers.¹¹ Numerous studies have examined these economic and other advantages of open and unrestricted access to government and publicly funded data.¹² In short, the open and unrestricted availability of data in the public sphere leverages investments for the benefit of all and furthers the public interest.

There is a discernable and concomitant global trend towards adopting open data policies.¹³ A major impetus for this increase in international sharing of research data is its critical role in

<https://www.performance.gov/content/invest-strategically-public-participation-science-technology-engineering-and-mathematics?view=public#overview>.

⁸ For example, open Earth observation data from several U.S. government organizations have supported the education of students in schools. *See, e.g.*, the Globe Program at <https://www.globe.gov/>.

⁹ *See, e.g.*, PAUL F. UHLIR, POLICY GUIDELINES FOR THE DEVELOPMENT AND PROMOTION OF GOVERNMENTAL PUBLIC DOMAIN INFORMATION (UNESCO 2004), http://portal.unesco.org/ci/en/ev.php-URL_ID=15863&URL_DO=DO_TOPIC&URL_SECTION=201.html.

¹⁰ *See, E.g.*, JANICE S. NELSON, USGS, U.S. GEOLOGICAL SURVEY EARTH RESOURCES OBSERVATION AND SCIENCE (EROS) CENTER—FISCAL YEAR 2010 ANNUAL REPORT. Open-File Report 2011–1057 (USGS 2011), <http://pubs.usgs.gov/of/2011/1057/pdf/of2011-1057.pdf>; and government use of data from the Global Infrastructures for Supporting Biodiversity Research (GLOBIS-B) project at <http://www.globis-b.eu>.

¹¹ GEOFF SAWYER & MARC DE VRIES, ABOUT GMES AND DATA: GEESE AND GOLDEN EGGS: A STUDY ON THE ECONOMIC BENEFITS OF A FREE AND OPEN DATA POLICY FOR GMES SENTINELS DATA (European Association of Remote Sensing Companies 2012), http://ears.org/file_download/134/Open+Data+study+Final+report.pdf and PETER WEISS, BORDERS IN CYBERSPACE: CONFLICTING PUBLIC SECTOR INFORMATION POLICIES AND THEIR ECONOMIC IMPACTS (NOAA 2002), http://www.epsiplus.net/reports/borders_in_cyberspace.

¹² Some of these reports include: NEIL BEAGRIE & JOHN HOUGHTON, THE VALUE AND IMPACT OF DATA CURATION AND SHARING: A SYNTHESIS OF THREE RECENT STUDIES OF UK RESEARCH DATA CENTERS (JISC 2014), http://repository.jisc.ac.uk/5568/1/iDF308_-_Digital_Infrastructure_Directions_Report%2C_Jan14_v1-04.pdf; DELOITTE, MARKET ASSESSMENT OF PUBLIC SECTOR INFORMATION (U.K. Department for Business Innovation and Skills, 2013), https://www.gov.uk/government/uploads/systems/uploads/attachment_data/file/108905/bis-13-743-market-assessment-of-public-sector-information.pdf; REBOOTING PUBLIC SERVICE DELIVERY: HOW CAN GOVERNMENT DATA HELP TO DRIVE INNOVATION (OECD 2016), <http://www.oecd.org/gov/Rebooting-Public-Service-Delivery-How-can-Open-Government-Data-help-to-drive-Innovation.pdf>; and GRAHAM VICKERY, REVIEW OF RECENT STUDIES ON PSI RE-USE AND RELATED MARKET DEVELOPMENTS (European Commission 2011), http://ec.europa.eu/information_society/newsroom/cf/document.cfm?doc_id=1093.

¹³ Many international policies, declarations, and statements that promote open access to public research data, though not necessarily promoting public domain or commons approaches, have been promulgated over the past two decades. A compendium of such soft-law instruments can be found in Appendix A. Some are broad in scope and apply to all types of government or public-sector data, while others are focused on publicly funded scientific research data. Some are limited in scope to data regarding a specific scientific discipline, whereas others apply to all fields of scientific research funded by the organization or its members. A range of legal approaches for removing intellectual property restrictions on the use of research data are reviewed in Part 3, below and recommended

addressing global challenges.¹⁴ These challenges transcend national boundaries and cannot be resolved by a single nation. They include the mitigation, adaptation and resiliency to climate change; the conservation of biodiversity; combatting infectious diseases; environmentally sustainable development; and natural disaster management.

The year 2015 was pivotal in mobilizing the global community to address these pressing challenges. In March, the Third UN World Conference on Disaster Reduction adopted the Sendai Framework for Disaster Risk Reduction 2015-2030,¹⁵ in September, the UN Sustainable Development Summit adopted goals to promote economic prosperity, social development, and environmental protection¹⁶; in November, the Group on Earth Observations Ministerial Summit affirmed the GEO principles of open access to data to address socioeconomic and environmental global challenges,¹⁷ and in December, the UN Framework Convention for Climate Change adopted a climate change agreement (the Paris Agreement) to reduce greenhouse gas emissions.¹⁸ International data sharing is either encouraged or highlighted in each of these and will be essential to achieve or monitor progress.

There has been a great deal of effort to promote the open sharing of publicly funded data through formal and informal mechanisms at international, national, and sub-national levels. Some open access efforts focus on data produced or collected by governments, whether in the developed or developing world, while others also apply to research data funded by governmental entities or not-for-profit organizations in the public interest. More narrowly, some open access policies are discipline specific, in fields such as Earth and environmental science, genomics and other life sciences, and astronomy.

Even the private sector has promoted the sharing of data by making data resulting from their own funding, publicly available, with or without restrictions.¹⁹ The International Serious Adverse Events Consortium is a recent example of a private sector initiative to expand access to DNA data to promote biomedical research in drug-induced serious events. U.S., European, and Japanese private pharmaceutical companies and the Wellcome Trust partner with over 50 academic and clinical groups to make their data and research results available to the public free of any patent encumbrances.²⁰

mechanisms are set forth in Part 4.

¹⁴ Treaties and many international agreements have been established to address these and other global problems, *see infra* section 3.1.1.

¹⁵ SENDAI FRAMEWORK FOR DISASTER RISK REDUCTION 2015-2030 (UN 2015)
http://www.unisdr.org/files/43291_sendaiframeworkfordrren.pdf

¹⁶ *See* <https://sustainabledevelopment.un.org/post2015/summit>.

¹⁷ *See* <http://www.earthobservations.org/dswg.php>; and GEO STRATEGIC PLAN 2016-2025: IMPLEMENTING GEOSS (GEO 2015),

http://www.earthobservations.org/documents/GEO_Strategic_Plan_2016_2025_Implementing_GEOSS.pdf.

¹⁸ PARIS AGREEMENT (United Nations Framework Convention on Climate Change 2015),

http://unfccc.int/files/home/application/pdf/paris_agreement.pdf.

¹⁹ Some early examples include The SNP Consortium (TSC), a voluntary association of companies and institutions that produced a public resource of single nucleotide polymorphisms (SNPs) in the human genome with open source software (description available at <http://www.ncbi.nlm.nih.gov/pubmed/12519964>), and Google Earth, which provides Earth mapping and imagery and geographic information to users,

<http://www.google.com/earth/explore/products>.

²⁰ *See, e.g.,* Contreras JL, Floratos A, Holden AH, *The International Serious Adverse Events Consortium's Data*

To realize the full potential of data sharing on a free and open basis, however, a number of barriers may need to be overcome including legal, policy, technical, management, institutional, economic, or sociocultural sources.²¹ There also can be a variety of legitimate restrictions on access and use of publicly generated data, such as the protection of national security, public safety, personal privacy, confidentiality, proprietary interests for commercial exploitation in the private sector, and defined terms of exclusive use by the researcher (embargo periods).²² In some cases, data are withheld because their release could cause some specified harm, such as the precise locations of endangered species habitats or archeological artifacts, or the protection of some indigenous peoples' rights.

While there have been an increasing number of national and international initiatives to promote open access to data, develop best practices for data management, and achieve technical interoperability of data,²³ less attention has been given to enabling the legal interoperability of data.²⁴ This article focuses on a subset of legal issues that are important to promote the sharing

Sharing Model, 31 NATURE BIOTECH. 17 (2013). Members of the iSAEC use a protective patent strategy modeled on the SNP Consortium. U.S. patent applications claiming DNA markers are filed with the intention of abandoning the patent after publication.

²¹ See, e.g., CHRISTINE BORGMAN, *BIG DATA, SMALL DATA, NO DATA* (The MIT Press, 2014). See also, Paul Uhlir and Peter Schröder, *Open Data for Global Science*, *Data Science Journal* 6:OD36 (2007), <http://datascience.codata.org/articles/abstract/10.2481/dsj.6.OD36/> (provides access to full article download)

²² *Id.* For examples of an intergovernmental organization's public data release policies, see two policy recommendations of the Organisation for Economic Co-operation and Development (OECD) titled OECD PRINCIPLES AND GUIDELINES FOR ACCESS TO RESEARCH DATA FROM PUBLIC FUNDING (OECD, 2007), <http://www.oecd.org/science/sci-tech/38500813.pdf> and OECD RECOMMENDATION OF THE COUNCIL FOR ENHANCED ACCESS AND MORE EFFECTIVE USE OF PUBLIC SECTOR INFORMATION [C(2008)36] (OECD, 2008), <http://www.oecd.org/internet/ieconomy/40826024.pdf>. Although such OECD policies are not legally binding upon the Member States, they are adopted by consensus of the world's most economically developed nations. Both sets of OECD recommended data policies recognize that national governments may restrict access to data due to national security considerations, privacy concerns, or the protection of private proprietary rights in certain circumstances. For example, the OECD Guidelines for Access to Research Data from Public Funding provide that "[a]ccess to or use of data may be restricted to safeguard the privacy of individuals, protect confidentiality, proprietary results or national security." (p. 13) The OECD Guidelines to Promote Access and Use of Public Sector Information assume openness in public sector information as a default rule but recognize that governments may define "... grounds of refusal or limitations, such as for protection of national security interests, personal privacy, preservation of private interests, for example where protected by copyright, or the application of national access legislation and rules" (p. 5). Similar restrictions on access to data can be found in many international agreements, and national legislation and policies. Although these restrictions promote legitimate interests and achieve compatibility with other legal requirements, an overly broad application of such restrictions undermines open access principles for public data and the benefits that such a policy can confer.

²³ G-7 Science and Technology Ministers' Tsukuba Communiqué, Tsukuba, Ibaraki, Japan, May 15-17, 2016

²⁴ The geospatial community has focused on the issue of legal interoperability. The Open Geospatial Consortium, a non-profit organization composed of governments, commercial organizations, NGO's academic and research organizations, is focused on making quality open standards for the global geospatial community. Its standards are used in domains spanning the environment, defense, health, agriculture, meteorology, and sustainable development. See <http://www.opengeospatial.org>. Another geospatial data activity focused on the issues of legal interoperability of publicly-funded data was the Group on Earth Observations' (GEO) Data Sharing Working Group and its subgroup on Legal Interoperability. That subgroup wrote two white papers, the most recent of which was published by GEO titled MECHANISMS TO SHARE DATA AS PART OF GEOSS DATA-CORE (GEO 2014), http://www.earthobservations.org/documents/dswg/Annex%20VI%20-%20Mechanisms%20to%20share%20data%20as%20part%20of%20GEOSS%20Data_CORE.pdf [hereinafter

and unfettered reuse of data. It does not, however, explore all legal issues associated with the open sharing of data, such as those related to liability or privacy law challenges in research.²⁵ Rather, it addresses intellectual property law related to open data sharing and examines various legal mechanisms that enable access, use, and reuse of data with a particular emphasis on data of value to not-for-profit science and other public interests.²⁶ Government generated or funded data are the primary focus of this article, but some of these mechanisms can also be used by privately funded data producers, although likely on a more limited basis.

As has been documented in this introduction, collective experience across the globe shows that the open use and sharing of scientific and technical data accelerate scientific discoveries and stimulate economic growth. The overall goal of this article is to recommend to governments, public-interest organizations, various kinds of public research consortia, and individual researchers across the globe legal mechanisms that facilitate the open sharing and use of data in support of achieving public-interest societal objectives.

We therefore have structured the rest of this article as follows. The final section of the Introduction provides a number of key terms that are essential to understanding this topic. In Part 2 we describe how copyright and other forms of intellectual property rights may protect research data and certain related data. We examine differences in national implementation of copyright principles and how these differences impact the trans-boundary sharing of data.

We then focus in Part 3 on several government mechanisms that can be utilized to remove intellectual property rights restrictions on the use and reuse of research data to achieve their maximum legal interoperability. International approaches include treaties, agreements, and policies. On a national level, they include legislation, regulations, and national policies. The article primarily cites to legal and policy sources from the jurisdictions of the United States of America and the European Union²⁷ to substantiate our points and recommendations. The citations to these two jurisdictions is linked to the expertise of the authors, the significant role these jurisdictions have played in the evolution of data access laws and policies, and limited resources for a global review of national approaches.²⁸ In addition to government mechanisms to promote open access to research data, in the last section of Part 3, we discuss the use of existing private law instruments and mechanisms that individual researchers and their institutions can use

GEOSS DATA-CORE WHITE PAPER 2014]. More recently, the Research Data Alliance–CODATA Interest Group on Legal Interoperability of Research Data has published *PRINCIPLES AND IMPLEMENTATION GUIDELINES ON THE LEGAL INTEROPERABILITY OF RESEARCH DATA* (RDA-CODATA 2016), forthcoming [hereinafter RDA-CODATA Interest Group]. That Interest Group coordinated its activities closely with this article and had common membership with one of the co-authors, Paul Uhler.

²⁵ For an illustrative project pursuing the ability to protect privacy in research data while making the sharing of that protected data easier and more efficient, consult *PRIVACY TOOLS FOR SHARING RESEARCH DATA*, <http://privacytools.seas.harvard.edu/>

²⁶ Even in instances where there are no legal restrictions imposed on a dataset itself, linked intellectual property and contract-related restrictions in tools (such as software) and platforms (both software and hardware) may still limit the use of such data. This article does not address these related challenges.

²⁷ Theories and disputes regarding the nature of the European Union as an institution aside, this article treats the EU as a single jurisdiction with respect to regulatory sources that the EU has adopted and implemented.

²⁸ A comparative analysis of the relevant regulatory and policy mechanisms in these two jurisdictions is not the goal of this article.

to eliminate intellectual property restrictions and ensure full legal interoperability for their data.²⁹ We also look at research data consortia and research programs more generally that can create their own data release policies and make their data available to the world with no restrictions on reuse by means of private agreements and policies.

We conclude the article in Part 4 with a summary of our findings, and present a set of conclusions and recommendations for policy makers, publishers, librarians, data center managers, university administrators, public interest organizations, and even individual researchers, as well as their legal counsel. The Appendices A and B provide a list of relevant organizations and research data policies in the context of Part 3.

1.2 Key Terms and Concepts Used in this Article

It is important at the outset to provide working definitions of the most relevant terms that are used throughout the paper and are important elements of the analysis provided.³⁰ Rather than listing the terms in alphabetical order, core terms are listed earlier and closely related terms are grouped together.

Legal Interoperability: By legal interoperability, we mean the ability to combine data from two or more sources without conflicts among restrictions imposed by data providers (i.e., support of one restriction inherently negates support of another) and without having to seek authorization from the data providers on a case-by-case basis. More specifically:

Legal interoperability among multiple datasets from different sources occurs when:

- the legal use conditions are clearly and readily determinable for each of the datasets typically through automated means;³¹
- the legal use conditions imposed on each dataset allow creation and use of combined or derivative products; and

²⁹ The genesis of this article was the *GEOSS Data-CORE White Paper*, 2014, at note 20. The vision of GEO, an intergovernmental entity, is to “realize a future wherein decisions and actions, for the benefit of mankind, are informed by coordinated, comprehensive and sustained Earth observations and information.” (<http://www.earthobservations.org/vision.php>). GEO established a GEOSS Data Collection of Open Resources for Everyone (Data-CORE) to promote access to Earth observation datasets and enable use and reuse of the data without legal restrictions. The original White Paper focused on sharing Earth observation data in Data-CORE. This article substantially expands upon the scope of the White Paper to encompass the international and interdisciplinary sharing of data in all fields of publicly-funded and non-profit funded research. Authors of the White Paper also consulted with the RDA-CODATA Interest Group on Legal Interoperability of Research Data.

³⁰ This section introduces several of the listed terms in the context of their common usage meanings within the technical database practitioner community, with some accompanying discussion of their legal relevance. More extensive legal analysis follows in later sections.

³¹ For an explanation of automated approaches to enable finding of machine-readable licenses attached to research data, see *Mechanisms for licensing data* in Digital Curation Centre guide on How to License Research Data (<http://www.dcc.ac.uk/resources/how-guides/license-research-data>) [hereinafter DCC Guide on How to License Research Data].

- users may legally access and use each dataset without seeking authorization from data creators on a case-by-case basis assuming that the accumulated conditions of use for each and all of the datasets are met.³²

Legal interoperability implies that the search for and tracking of legal conditions for datasets will occur online through automated methods. When data are combined from multiple sources, the resulting dataset will incorporate the accumulated restrictions imposed by each and every source.³³ Therefore, any restrictions need to be tracked. To be interoperable, dataset restrictions also need to be compatible. Compatibility of datasets means that conditions of use of datasets do not inherently negate the conditions of use of other data sets as long as the accumulated restrictions of all may be carried forward in the derivative dataset. Comparison of standard open data legal instruments enables identification of legally interoperable terms that despite differences in wording, constitute the same or similar conditions of use. The fewest restrictions contained in parent datasets results in the fewest restrictions in derivative datasets. The simplest cases for tracking and legal interoperability occur when datasets are affirmatively identified as having no legal restrictions.³⁴

Interoperability: Interoperability, as typically used by the computer and information science communities, means the ability of information systems to work with each other because their interfaces are completely understood even when individual components are technically different and managed by different organizations. Interoperability includes technical, syntactic and semantic interoperability. Technical interoperability is usually associated with hardware and software components, systems and platforms that enable machine-to-machine communication. Syntactic interoperability is usually associated with data formats and provides for the exchange of clearly defined classes of data. Semantic interoperability is concerned with ensuring that the precise meaning of exchanged information is understandable by any other application and is the ability to automatically interpret the information exchanged meaningfully and accurately in order to produce useful results.³⁵

³² Definition derived from Harlan J. Onsrud, *Legal Interoperability in Support of Spatially Enabling Society*, in SPATIALLY ENABLING SOCIETY: RESEARCH, EMERGING TRENDS AND CRITICAL ASSESSMENT 163-172 (Abbas Rajabifard et al., eds. Leuven University Press 2010), <https://soc.kuleuven.be/io/english/research/publication/book-spatially-enabling-society-research-emerging-trends-and-critical-assessment>. See that article for a more extensive discussion of legal interoperability. For a complementary perspective, see *GEOSS Data-CORE White Paper*. See also Jorge L. Contreras & Jerome H. Reichman, *Sharing by design: Data and decentralized commons Overcoming legal and policy obstacles*, SCIENCE 350:6266 (11 December 2015)

³³ Use conditions may be imposed by a license, or through a contract or agreement, or by action of national laws.

³⁴ See also the Principles and Implementation Guidelines of the RDA-CODATA Interest Group on Legal Interoperability of Research Data (2016), available at: [http://\[TBC\]](http://[TBC]).

³⁵ Derived from H. Kubicek et al., *Chapter 7, Layers of Interoperability*, in ORGANIZATIONAL INTEROPERABILITY IN E-GOVERNMENT (Springer-Verlag 2011), DOI 10.1007/978-3-642-22502-4_7. Interoperability has been sought by the technical community through such approaches as open formats, formal ontologies, non-proprietary software and algorithms, open metadata, standardization of tools, and other essential functions and features that enable sharing of data from different sources. Technical and syntactic interoperability challenges have often been easier to address because solutions have been more apparent frequently and have received significant attention from funders and researchers alike. Resolving semantic interoperability remains a substantial research challenge significantly inhibiting the sharing and integration of data. Specific vocabularies and criteria used for categorizations used by one

Facts: Facts exist in the world independently or depict the world as it is and no original creative expression attaches to them. Facts are typically discoverable, observable and often objectively verifiable. “The first person to find and report a particular fact has not created the fact; he or she has merely discovered its existence....”³⁶ Facts may not be copyrighted and are part of the public domain available to every person.³⁷

Data: “Facts, measurements, recordings, records, or observations about the world collected by scientists and others, with a minimum of contextual interpretation. Data may be in any format or medium taking the form of writings, notes, numbers, symbols, text, images, films, video, sound recordings, pictorial reproductions, drawings, designs or other graphical representations, procedural manuals, forms, diagrams, work flow charts, equipment descriptions, data files, data processing algorithms, or statistical records.”³⁸ Although data are often born digital in contemporary society or may be converted to a digital form for convenience, data may be non-digital as well. Some data may be eligible for copyright because it is creative original expression and some data, including that which is merely facts, may be eligible for other appropriate or applicable forms of intellectual property protection, for example by virtue of *sui generis* database rights, depending on its jurisdiction of collection or creation.

Research Data: Data that are used as primary sources in accomplishing technical or scientific enquiry, research, or scholarship, and that are used as evidence in the research process and/or are commonly accepted in the research community as necessary to validate research findings and results. All other digital and non-digital content have the potential of becoming research data.³⁹ Research data may be experimental data, observational data, operational data, third party data, public sector data, monitoring data, processed data, or repurposed data.⁴⁰

Dataset: The term “dataset” in this paper refers to a collected, selected, coordinated, or arranged set of data elements in electronic form consisting often of observed, discovered, or derived values. A dataset might consist of a brief ad hoc table of values or might consist of millions of values arranged using a standard format in order to facilitate electronic updating and processing.

Database: The term “database” in this paper refers to a collection of information that is organized so that it can easily be accessed, managed, updated. It might consist typically of a collection of schemes, tables, queries, reports, observations, and other data objects organized to

community of experts are often not interchangeable or unable to be resolved or integrated with those of another. Similarly, meanings for words and phrases across languages often differ and can be difficult to translate. These semantic interoperability challenges persist as well for legal interoperability.

³⁶ *Feist Publications Inc. v. Rural Telephone Service Co., Inc.*, 499 U.S. 340, 347 (1991) [hereinafter *Feist*].

³⁷ *Id.*, at 348. Even though the EU doesn’t have jurisdiction to regulate the issue of copyrightability, its acceptance of creativity as the basis for granting protection to work can be inferred from Recital 9 Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society. Official Journal L 167 , 22/06/2001 P. 0010 – 0019.

³⁸ DICTIONARY.CASRAI.ORG, <http://dictionary.casrai.org/Data>.

³⁹ In modern society, large amounts of data are gathered for purposes other than research, but may also be very useful for scientific inquiries. Examples include the tracking of millions of movements, activities, and interactions of vehicles, objects, humans, animals, and natural phenomena for commercial, governmental, and personal (e.g., crowd-sourcing) purposes.

⁴⁰ This definition from DICTIONARY.CASRAI.ORG, http://dictionary.casrai.org/Research_data.

model aspects of reality to support processes. Databases are sometimes classified according to types of content: bibliographic, full-text, numeric, or images.⁴¹

Database Management System: A database management system (DBMS) is a computer software application that interacts with the user (typically through a structured query language), other applications, and the database itself to select and manipulate data. A general-purpose DBMS is designed to allow the definition, creation, querying, update, and administration of databases. Well-known DBMSs include MySQL, PostgreSQL, Microsoft SQL Server, Oracle, Sybase, and IBM DB2.⁴²

Public Domain: As used in this article, public domain is material "... that is not covered by intellectual property rights. Material might be in the public domain because it was never capable of being owned. Examples would be the English language or the formulae of Newtonian physics. Alternatively, something might be in the public domain because rights have expired. The works of Shakespeare or the patents over powered flight are examples."⁴³ Public domain works are owned by the public and are available for use without permission.⁴⁴

Open Access: Open access means "... digital, online, free of charge, and free of most copyright and licensing restrictions."⁴⁵ "The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited."⁴⁶ Datasets within the *public domain* and the *open commons* qualify as open access data (or, just "open data").

Commons: Commons refers to formal and informal institutions for managing access to a shared resource.⁴⁷ Hess and Ostrom state a core distinction as "the difference between property regimes that are open access, where no one has the right to exclude anyone from using a resource, and common property, where members of a clearly defined group have a bundle of legal rights including the right to exclude nonmembers from that resource"⁴⁸ Benkler presents a fuller scope

⁴¹ Derived from SEARCHSQLSERVER.TECHTARGET.COM, <http://searchsqlserver.techtarget.com/definition/database> and WIKIPEDIA.ORG, <https://en.wikipedia.org/wiki/Database>. Similarly, the EU Database Directive states ... "the term 'database' should be understood to include literary, artistic, musical or other collections of works or collections of other material such as texts, sound, images, numbers, facts, and data; whereas it should cover collections of independent works, data or other materials which are systematically or methodically arranged and can be individually accessed; ..."

⁴² Derived from WIKIPEDIA.ORG, <https://en.wikipedia.org/wiki/Database>.

⁴³ JAMES BOYLE, *THE PUBLIC DOMAIN: ENCLOSING THE COMMONS OF THE MIND* 38 (Yale University Press 2008)

⁴⁴ Some definitions of public domain do not focus on *works as a whole*, but instead include as part of the definition "reserved spaces of freedom inside intellectual property" (*Id.*). Such definitions include exceptions in the use of works provided through legislation such as fair use, fair dealing and concepts developed by case law, such as for example, the freedom to read, first sale doctrine, copyright merger doctrine, and similar legal exceptions. An even more expansive and human rights based aspirational definition for public domain is advocated through the Public Domain Manifesto. (See *The Public Domain Manifesto*, in *The Digital Public Domain: Foundations for an Open Culture* p.xx (Melanie Rosnay & Juan De Martin, eds. Open Book Publishers 2012)).

⁴⁵ Peter Suber, *Open Access* 4 (MIT Press 2012),

https://mitpress.mit.edu/sites/default/files/9780262517638_Open_Access_PDF_Version.pdf

⁴⁶ *Id.* 7

⁴⁷ See generally, BRETT FRISCHMANN, MICHAEL MADISON, & KATHERINE STRANDBURG, EDs., *GOVERNING KNOWLEDGE COMMONS* (Oxford Univ. Press 2014)

⁴⁸ Charlotte Hess & Elinore Ostrom, *Ideas, Artifacts and Facilities: Information as a Common Pool Resource*, *LAW & CONTEMP. PROBS.* III. 66:121 (2003)

for the commons and argues that the theory of the commons needs to include both "local, non-state-based institutional design for sustainable governance of resources as to which a defined set of claimants ... lay claim in commons" and "resource sets in modern market economies governed so that more or less anyone may use the resource set and no one, or no group, has exclusive rights as against anyone else."⁴⁹ It is this second or "open commons" sense in which we typically use the term *commons* in this article and also the sense in which Creative Commons licenses apply. Individuals and organizations might place content or intellectual work into an open commons sharing environment through action of a private self-help legal instrument such as by license, but may do so only up to the greatest extent allowed by the laws of their jurisdiction. As such, private acts of entities in many jurisdictions are incapable of dedicating their works in their entirety to the public domain. Intellectual works transferred through *Creative Commons* licenses and similar common-use licenses become *commons* resources with certain restrictions imposed, while works in the *public domain* are commons resources with no intellectual property conditions imposed.

Information: As used in this article, information includes any form of expression stored on a physical medium including creative and non-creative expression.⁵⁰

Metadata: Metadata are data about data. Metadata describe the data content of datasets through a systematic standard means and are typically included as a critical component of many scientific and technical databases. The widespread use of metadata makes it much easier to find datasets appropriate to specific needs.

Intellectual Property Rights: As used in this article, "intellectual property rights" means copyright, *sui generis* database rights, patent, trademark, trade secret, and other similar laws that establish a regime for access, use or reuse of data, metadata, or data products. It does not include other restrictive information laws based on the protection of national security, public safety, personal privacy, confidentiality, and other specific reasons for withholding research data from the public.

2 PROTECTION OF DATA BY INTELLECTUAL PROPERTY LAWS

As noted in the Introduction, it is highly desirable that datasets produced by or used in public research should have unrestricted re-use terms and conditions in order to make them most beneficial for research and other public-interest purposes. That said, various laws limit or restrict access to, use and re-use of data and information due to different reasons and interests.⁵¹ This

⁴⁹ Yochai Benkler, *Between Spanish Huertas and the Open Road* in GOVERNING KNOWLEDGE COMMONS 70 (Brett Frischmann, Michael Madison, & Katherine Strandburg, eds. (Oxford Univ. Press 2014).

⁵⁰ This common-usage definition for *information* varies from that of many communication theory and dictionary definitions. We use the term generally to refer to information objects that may be transferred electronically. A more typical or traditional definition would include both oral and written communications and state that information consists of one or more statements received by a human that have worth to the human. That is, information does not exist until it is received in the mind of another (WARREN WEAVER & CLAUDE ELWOOD SHANNON, THE MATHEMATICAL THEORY OF COMMUNICATION (Univ. of Illinois Press. 1963)).

⁵¹ Restrictive laws and governmental policies include protection of national security, privacy, confidentiality, and intellectual property, among other more narrowly-based concerns. Moreover, in some countries freedom of

section looks only at the restrictions based on intellectual property (IP) laws, which, unlike other statutory restrictions, can sometimes be waived or modulated with less restrictive standard common-use licenses, or with statutory exclusions or governmental policies.

As a general proposition, all kinds of datasets, including those produced or used in research, may be subject to copyright or related intellectual property rights (e.g., database protection) in accordance with existing national legislation. Certain types of use or re-use (e.g., copying and subsequent distribution with or without alteration) of datasets typically require the explicit permission of the rights holder.

2.1 Copyright

2.1.1 Summary of key international principles

The Berne Convention for the Protection of Literary and Artistic Works sets forth a common international framework of minimum standards for copyright, but does not create legally enforceable copyright protection.⁵² It is the oldest and the most accepted source of international law that codifies the fundamental principles of copyright protection.⁵³

An international copyright protection regime that would automatically protect rights in creative content on an identical basis throughout the world, however, does not exist. National copyright protection therefore varies across jurisdictions, depending upon the laws of each country and their interpretation by the national courts.

The first principle we wish to emphasize is one of *originality* and *creativity*: only works that are the original intellectual creations of their authors are eligible for copyright protection.⁵⁴ The traditional interpretation of this principle excludes protection of subject-matter solely on the basis of the input of time, labor, or financial resources. It emphasizes instead the primacy of creativity and by extension the need to recognize the author's work. A work must contain a

information legislation provides a mechanism for access to government-maintained datasets, although it is frequently, slow, costly, incomplete, and can even be dangerous to the one requesting the information. Such legislation is not discussed further in this article because it is not typically used by researchers to obtain research datasets.

⁵² Berne Convention for the Protection of Literary and Artistic Works, Sept 9, 1886 (hereinafter the Berne Convention), completed at Paris on May 4, 1896, revised at Berlin on November 13, 1908, completed at Berne on March 20, 1914, and revised at Rome on June 2, 1928, at Brussels on June 26, 1948, at Stockholm on July 14, 1967 and at Paris on July 24, 1971, http://www.wipo.int/treaties/en/text.jsp?file_id=283698 (Paris Text 1971).

⁵³ There are other international treaties that regulate copyright: the Universal Copyright Convention, the Agreement on Trade-Related Aspects of Intellectual Property Rights, and the WIPO Copyright Treaty. However, since all of them are complementary to the Berne Convention and we present just a summary of the main aspects of copyright law, only the Berne Convention is discussed here.

⁵⁴ Berne Convention, art. 2. Article 2 Berne Convention doesn't use the terms originality and creativity. It says "literary and artistic works" are protected. In its section 5 it states "Collections of literary or artistic works such as encyclopaedias and anthologies which, by reason of the selection and arrangement of their contents, constitute intellectual creations". In addition, section 8 denies copyright to news of the day or to miscellaneous facts having the character of mere items of press information. From such principles, originality and creativity may be deduced as the basis for protection.

certain minimum amount of authorship in the form of original literary, pictorial, or graphic expression to qualify for copyright protection.⁵⁵

A second and related principle of copyright protection is that only the *expression* of a work, and not the ideas on which it is based, are protected.⁵⁶ WIPO states this concept as “Copyright protection extends to expressions and not to ideas, procedures, methods of operation or mathematical concepts as such.”⁵⁷ Furthermore, protection of expression, such as the freedom to communicate facts and ideas under Article 19 of the United Nations International Covenant on Civil and Political Rights, excludes the possibility to protect the ideas and facts themselves.⁵⁸ Facts or events reported in a creative expression such as a book, newspaper, video, or report are not protected under copyright from being reported by others in their own creative expressions.

A third principle is that the work may have to be fixed on a tangible medium if required by national law.⁵⁹ Fourth, copyright protection is automatic and does not require any formal registration.⁶⁰ Finally, a central tenet of copyright is that it must be limited in time.⁶¹ The minimum term of protection under the Berne Convention, as amended, is life of the author, plus 50 years.⁶²

Even though an author’s rights under copyright typically are transferrable or able to be waived under copyright, moral rights are an exception to this general principle. The Berne Convention states that “[i]ndependently of the author’s economic rights, and even after the transfer of said rights, the author shall have the right to claim authorship of the work and to object to any distortion, mutilation or other modification of, or other derogatory action in relation to, the said work, which would be prejudicial to his honor or reputation.... The means of redress for safeguarding the rights granted by this Article shall be governed by the legislation of the country where protection is claimed.”⁶³ Moral rights protect the personal reputation of the creator, which may include the right to attribution, the right to the integrity of the work, or the right to have the work published anonymously. Many countries enforce such moral rights and some do not allow

⁵⁵ See the U.S. Copyright Office’s webpage at <http://www.copyright.gov/circs/circ34.pdf>.

⁵⁶ See e.g., Berne Convention, art. 2(1) (implied); ART. 2, WIPO Copyright Treaty (expressed). December 20, 1996, 2186 U.N.T.S. 121.

⁵⁷ Cf. WIPO Copyright Treaty, art. 5

⁵⁸ WHAT COPYRIGHT DOES NOT PROTECT, <https://ucomm.wsu.edu/what-copyright-does-not-protect/>.

⁵⁹ See Berne Convention, art. 2(2). A “tangible medium” might include analog or digital print, musical recording, motion picture, and the like.

⁶⁰ Berne Convention, art 5(2).

⁶¹ For example, U.S. CONST. art. 1 § 8, cl. 8 states: “The Congress shall have the 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...”

⁶² Berne Convention, art. 7(1). However, many jurisdictions, notably the United States and the countries of the European Union, have adopted an even longer copyright term of protection lasting the life of the author, plus 70 years.

⁶³ Berne Convention, art. 6bis (1) & (3). Moral rights is a translation from the French term “droit moral” and refers to the ability of a creator to control the fate of its works. *Betsy Rosenblatt*, MORAL RIGHTS BASICS (1998), <http://cyber.law.harvard.edu/property/library/moralprimer.html>.

them to be waived.⁶⁴ Moral rights are traditionally associated with literary and artistic works, and therefore less likely to be invoked by creators of research data.

In addition, no copyright protection regime is complete without some designated limitations and exceptions to the author's rights in favor of the users of copyrighted works.⁶⁵ Limitations and exceptions can be based on the status of the user, the type of use, its extent, the type of protected works, or other factors.

Although limitations and exceptions to copyright, *sui generis* database rights, and other legal rights would be relevant to many uses of research data, whether commercial or not, their applicability is determined on a case-by-case basis. Such limitations and exceptions typically are narrowly drawn and situation-dependent.⁶⁶ Thus, we believe they do not represent a legally adequate solution to implement the broad reuse and interoperability requirements for public research data generally.

2.1.2 Differences in national implementation of copyright principles

Codification of the norms of international law regarding principles of copyright protection has resulted in acceptance of a set of minimum requirements throughout the world in almost all countries. However, national law determines the exact scope of the protection granted and its implementation in any given jurisdiction.

There are several areas where the differences can be very significant. The criteria for copyright protection delineated in the Berne Convention can be interpreted differently, depending on the jurisdiction. This is true particularly for the criterion of creativity: its minimum threshold is not interpreted homogeneously. The consequences of such differences are that some subject-matter may be eligible for protection in one jurisdiction, but not in another. In addition, the exact scope of copyright protection is also laid down in national regulatory regimes. For instance, under the U.S. Copyright Act copyright applies only to the expression: "Copyright protection subsists ... in original works of authorship fixed in any tangible medium of expression.... In no case does copyright protection for an original work of authorship extend to any idea, procedure, process, system, method of operation, concept, principle, or discovery."⁶⁷ While in Germany, "Copyright protects the author in his intellectual and personal relationships to the work and in respect of the

⁶⁴ Commonwealth countries will generally allow a waiver of moral rights while civil law countries will not. Some civil law countries, however, will enforce an author's promise not to assert his/her moral rights. See ANDRES GUADAMUZ, WIPO, COMPARATIVE ANALYSIS OF NATIONAL APPROACHES ON VOLUNTARY COPYRIGHT RELINQUISHMENT (2014), http://www.wipo.int/edocs/mdocs/mdocs/en/cdip_13/cdip_13_inf_6.pdf; CELIA LERMAN, CREATIVE COMMONS, UNWAIVABLE MORAL RIGHTS AND THE NEED FOR NEW INTERNATIONAL SUBSTANTIVE MINIMA (2012), <http://dx.doi.org/10.2139/ssrn.2159697>. The United States has legislatively created moral rights solely for works of visual art. See 17 U.S.C. § 106A (1990), <https://www.law.cornell.edu/uscode/text/17/106A>. In the United States, however, moral rights are not recognized more generally as part of copyright law.

⁶⁵ For an overview of copyright and other IP law limitations and exceptions in the context of research data, see Jerome H. Reichman and Ruth L. Okediji, *When Copyright Law and Science Collide: Empowering Digitally Integrated Research Methods on a Global Scale*, MINN. L. REV. 96:1362 (2012) [hereinafter Reichman and Okediji 2012], http://www.minnesotalawreview.org/wp-content/uploads/2012/08/ReichmanOkediji_MLR1362.pdf.

⁶⁶ *Id.*

⁶⁷ United States Copyright Act, 17 U.S.C. § 102 (b) (1976) [hereinafter U.S. Copyright Act].

use of the work. It shall also serve to ensure equitable remuneration for the exploitation of the work.”⁶⁸

The term of protection can also vary. The minimum term of protection set forth in the Berne Convention is life of the author plus 50 years. Many countries, including the United States, have extended the term of protection to life of the author plus 70 years.⁶⁹ Also in the United States, works-for-hire, usually in an institutional setting, attract a 95-year term of protection and are highly relevant to the production of both commercial and non-commercial datasets.⁷⁰ These differences too may lead to situations in which the same subject-matter is still protected in one state, but is already in the public domain in another.

The third important area that increases the risk of different protection of the same subject-matter is legislation in some jurisdictions that introduces IP protection different from copyright, like the *sui generis* database rights in the European Union. Introduction and implementation of new IP categories can create barriers for exchanging and sharing IP subject-matter across borders. Some of these forms of IP protection are discussed in Section 2.3.

The fourth important area that often contains differences among regulatory sources in various jurisdictions and the one particularly relevant to users of copyrighted subject-matter, including whenever appropriate research datasets, is that of exceptions and limitations to the rights of authors. In the United States, the main set of limitations is referred to as “fair use,”⁷¹ while in some other common law jurisdictions,⁷² a similar, but less permissive, limitation is known as “fair dealing.”⁷³ In most continental European jurisdictions, the system of limiting copyright is set up through a closed list of exceptions that allows for use of copyrighted works without prior authorization or remuneration.⁷⁴ Such differences in approaches to treating and codifying exceptions and limitations may result in identical acts permitted by the regime of one jurisdiction, but not tolerated by another.

These differences raise issues, particularly in the Internet or online environment that is often characteristic for retrieval and use of data today. In particular, the seemingly borderless Internet space potentially poses challenges to the territorial nature of the IP rights regulatory mechanisms both on the international level and within individual jurisdictions.⁷⁵ One of these issues that may be most crucial when it comes to retrieval and use of data in practice is the correct determination

⁶⁸ Copyright Act of 9 September 1965 (Federal Law Gazette Part I, p. 1273), as last amended by Article 8 of the Act of 1 October 2013 (Federal Law Gazette Part I, p. 3714). Translated version available at http://www.gesetze-im-internet.de/englisch_urhg/englisch_urhg.html#p0007.

⁶⁹ U.S. Copyright Act, note 67, § 302 (a).

⁷⁰ *Id.* § 302 (c).

⁷¹ Codified in Section 107 of the U.S. Copyright Act of 1976 that recognizes a limitation and exception to the exclusive rights granted by copyright law for a set of purposes such as “criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship or research.”

⁷² Fair Dealing is found in Australia, Canada, India, New Zealand, Singapore, South Africa, and the United Kingdom.

⁷³ A discussion of Fair Dealing under UK law may be found at <https://www.gov.uk/exceptions-to-copyright>.

⁷⁴ Article 5, Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society. OJL 167 (22/06/2001).

⁷⁵ See A. Lester, *The Implications of the Google Books Search Dispute for the Private International Law of Intellectual Property*, 16 J. TECH. L. & POL’Y 121 FF. (2011). See also Teresa Scassa & Robert J. Currie, *New First Principles? Assessing The Internet's Challenges To Jurisdiction*, 42(4) G. J. INT’L L. (2011).

as to which jurisdiction’s law applies when data are created in one jurisdiction and used in others. The typical default rule is that the jurisdiction in which the intellectual property rights are created governs in the absence of an agreement to the contrary.⁷⁶ However, this becomes more complicated when data are pooled from many users in different jurisdictions and generated products are distributed worldwide.⁷⁷ These questions can be addressed by an agreement or license issued by the rights holder, but ideally some clear default rules should be established and adhered to coherently by multiple actors in order to avoid problems regarding legal interoperability of data.

In terms of adopting a uniform regulatory mechanism in this regard, a protocol complementing existing relevant IP rights conventions based on good practice and coordinated law-making efforts of states could be a solution. Its primary focus and purpose would be the resolution of the conflict of laws in this field of using data.⁷⁸

2.1.3 *Application of copyright to factual data*

Factual data exist in the world independently or depict the world as it is and no original creative expression attaches to them. The reported observations by scientists and other researchers of myriad physical and natural characteristics often fall much closer to the legal concept of being discovered facts rather than being created. Measurements, counts, and standard descriptors (e.g., gender, age, and height) as often contained in digital datasets that indicate *what is* or *what happened* are facts. Even though there may be significant disagreement over the accuracy or precision of the observations, the observed and discovered facts are not protected by copyright.

The stylistic rendering of facts is copyrightable. However, when there are only a limited number of ways to express a concept or idea and there is little difference between the idea and expression, the doctrine of copyright merger⁷⁹ may apply. Thus, under copyright one may not maintain a monopoly over the discovered shape and dimensions of a microscopic cell or of an

⁷⁶ For a short overview of largely UK and some EU case law and practice in this regard see K. Lipstein, *Intellectual Property: Jurisdiction or Choice of Law?* 61 CAMBRIDGE L.J. 295 (2002). See also a more in-depth and EU-focused paper S. Neumann, *Intellectual Property Rights Infringements in European Private International Law: Meeting the Requirements of Territoriality and Private International Law*, 7(3) J. PRIVATE INT’L L. 583 (2011).

⁷⁷ For example, it is often the case with satellite Earth observation data. For an analysis of the potential issues regarding application of the IP right laws to this type of data, with specific focus on the principles of international space law and the applicable US regulatory regime and enforcement practices, see J.D. Cromer, *How on Earth Terrestrial Laws Can Protect Geospatial Data*, 32 J. SPACE L. 267 (2006) particularly at 263-268.

⁷⁸ The preparatory work for such a document could greatly benefit from, for example, the relevant work of the European Max Planck Group for Conflict of Laws in Intellectual Property, which developed Principles for Conflict of Laws in Intellectual Property. See EUROPEAN MAX PLANCK GROUP ON CONFLICT OF LAWS IN INTELLECTUAL PROPERTY (CLIP), PRINCIPLES ON CONFLICT OF LAWS IN INTELLECTUAL PROPERTY (2013)

⁷⁹ The merger doctrine is derived in the United States from *Baker v. Selden*, 101 U.S. 99, 25 L. ED. 841 (1879). *Toro Co. v. R & R Products Co.*, 787 F. 2D 1208, 1212 (8TH CIR. 1986) states “[u]nder the copyright law doctrine of merger, a close cousin to the idea/expression dichotomy, copyright protection will be denied to even some expressions of ideas if the idea behind the expression is such that it can be expressed only in a very limited number of ways. The doctrine is designed to prevent an author from monopolizing an idea merely by copyrighting a few expressions of it.” For a more recent case addressing the doctrine, see *Oracle America, Inc. v. Google, Inc.*, 750 F. 3D 1339 (2014).

unearthed building from an archeological site since the non-interpretive illustrative expressions of the size and shape are likely *one with the facts*.

Facts, once legally accessed, may be used, re-used, or re-disseminated by anyone for any purpose absent other legal protections. A definitive answer is dependent upon the circumstances of each separate case and the characteristics of the shared subject matter.

Datasets or collections of facts or data products may be subject to copyright protection as “compilations” or “collections” of information, even if they consist entirely of individually non-copyrightable facts, if their “selection, coordination, or arrangement” is achieved through some human creativity or originality.⁸⁰ Thus, the classification, coding, formats, and interpretations of data in a compilation may fall under copyright protection. In such cases, however, copyright only encompasses the creative arrangement or selection of the elements of a compilation (dataset), while its “parts” (items, materials...) are either not protected by copyright, or protected independently from the protection granted to the dataset (compilation).

Compilations of facts and their ancillary information in this category are generally interpreted to have “thin” copyright that protects against wholesale, verbatim copying and against copying any of the creative expression in the original selection, coordination, or arrangement. Compilations, particularly of factual material, that are arranged for ease of use, to comply with standards in some disciplinary or business context, or in some obvious, routine, or mechanical ways (e.g., in alphabetical order or adhering to a technical interoperability standard), generally are not protected by copyright.

As mentioned previously, *metadata* summarize basic information about a dataset, are often organized through a formal standard arrangement, and make searching for and finding data meeting specific criteria much easier. In illustrating some of the concepts just discussed, standardized metadata field structures are commonly not covered by copyright protection. That is, data suppliers typically select a standard and well-accepted metadata structure in order to facilitate utilitarian search and interoperability needs. Because of this, the standard structure is not creative and quite common. The structure used may have the legal equivalent arrangement status of alphabetical order and thus may not be protected by copyright. Further, observations contained within the metadata structure about the dataset such as size, theme, geographic coverage, and similar attributes may be held as merely factual observations. Thus, the individual attributes or measures may not be covered by copyright.

In contrast, a particular selection and gathering of metadata elements to meet a specific purpose might be held to reach the level of being “creative” and thus justify copyright protection. Negating the creative selection argument for protection is that automated search and processing seldom depends on the specific selection supplied by the creator to achieve the users’ technical needs. In short, whether a specific instance of data or metadata is protected by copyright will often require a case-by-case analysis. This lack of legal clarity places substantial burdens on both suppliers and users of data.

⁸⁰ *Feist*, note 36; WIPO Copyright Treaty, art. 5; Council Directive 96/9 of 11 March 1996 on the Legal Protection of Databases, art 3(2) 1996 O.J. (L. 77) 20 [hereinafter EU Database Directive, 1996].

2.1.4 Pragmatic Copyright Considerations for Research Data⁸¹

The preceding copyright discussion has numerous practical ramifications for research scientists. To recapitulate through examples, the discovery and identification of a plant species or bacteria at a particular spatial location and time or the discovery of a geologic feature at a particular depth below the surface would likely be held as facts by the courts. This is true regardless of the form in which a discovered fact “is described, explained, illustrated, or embodied in [a] work.”⁸² Many empirical observations and measurements in science fall much closer to the legal understanding of being “discovered” rather than being “created or authored” regardless of whether these observations or measurements stand alone, are among millions of observations within a single sensor observation program, or are included within a compilation of observations drawn from many sources.

This is also true whether an observation comes raw from a sensor or whether the observation has been processed in an attempt to arrive at a value closer to truth. Whether the volume of a lake is observed to be 46 km³ or 46.01083 km³,³ both of these numerical value estimates obtained by measurement have been “discovered” rather than “created.” Both are “facts” regardless of the time, knowledge, effort, and expense to arrive at each value and regardless of their precision. “Facts, whether alone or as part of a compilation, are not original and therefore may not be copyrighted.”⁸³ Similarly, events reported in science such as geologic, weather, extinction, and magnetic events, or reported data from controlled experiments or clinical trials often would be held as factual data as opposed to being legally authored or created.⁸⁴ Upon publication or other lawful access, copyright does not prevent use of such facts by others.

Even when the selection and arrangement of facts are protected by copyright, the typical software application in extracting data from one or more databases will seldom copy creative elements of the original datasets since the software application will extract its own algorithmically defined selection fit for a particular purpose.⁸⁵ Version tracking of datasets is relatively widespread and encrypted hidden identifiers may be carried by datasets, but there is no widely deployed technology currently available to comprehensively track extractions from datasets used in subsequent derived datasets. There is thus no way at this time to automatically carry forward the license and contract legal conditions nor the applicable national laws that

⁸¹ Much of the discussion in this section applies as well to challenges in enforcing misappropriation under unfair competition law and related legal approaches for enforcing license and contract terms.

⁸² 17 U.S.C. § 102 (B) as referenced in *Feist*, note 36, at 356, <http://www.copyright.gov/title17/92chap1.html>.

⁸³ *Feist*, note 36, at 351.

⁸⁴ See, e.g., NATIONAL RESEARCH COUNCIL, A QUESTION OF BALANCE: PRIVATE RIGHTS AND THE PUBLIC INTEREST IN SCIENTIFIC AND TECHNICAL DATABASES (National Academy Press, 1999) [Hereinafter A QUESTION OF BALANCE], <http://www.nap.edu/catalog/9692/a-question-of-balance-private-rights-and-the-public-interest>.

⁸⁵ As an example, web mapping service software might remotely search thousands of other web-hosted databases to find data that meet the attribute query needs of an active user, extract only the limited data required by the query from one or multiple remote sites, and then generate and display the query-defined map. That query-defined map will seldom express in its final visualization the selection and arrangement expression of any of the datasets from which the map was constructed. Note also however that while a dataset might contain only factual data objects, a database typically contains also schemes, tables, reports and other data objects that if original and copied might result in copyright violations.

might be germane to those extractions. For these pragmatic reasons, widespread legal enforcement of attribution and other licensing provisions as applied to contributions to derivative datasets is typically very difficult in the current technological environment.⁸⁶ Because of national variations in copyright protection of datasets, the extent of copyright protection afforded to the use of entire, or just portions of, datasets depends also upon the jurisdiction of origination of the dataset and the copyright (and other) rules adopted and implemented there. Thus, the challenges in enforcing copyright controls over the downstream uses of research data are substantial.

2.2 *Sui Generis* Database Rights

One of the alternative and complementary forms of IP protection to copyright is the *sui generis*⁸⁷ database rights legislation adopted in the European Union and some other countries around the world. In the EU, the Directive on the legal protection of databases was enacted in 1996,⁸⁸ and subsequently transposed into the national legislation of all EU Member States. Several other countries (e.g., Mexico,⁸⁹ Russia,⁹⁰ and South Korea⁹¹) have adopted similar legislation. Such laws protect the information compiled in databases, even mere facts that form more than an “insubstantial part” of the database, defined either quantitatively or qualitatively, as long as the database is the result of a “substantial investment in obtaining, verifying and presenting” its contents.⁹² The protection goes far beyond copyright as it protects the *contents* of the database and confers to the database rights holder the right to “prevent extraction and/or re-utilization” of all or substantial parts of the contents of the database.⁹³

The codified *sui generis* database right has not been easy to apply consistently and authoritatively within the jurisprudence so far—including the European Court of Justice, which is the only body that can authoritatively interpret EU legislation. In a number of its decisions, the Court has clearly stated that the resources spent on the generation of the *contents* of a database may not be counted as resources spent on the creation and maintenance of the database itself.⁹⁴

⁸⁶ For more details about this uncertainty as it affects common-use licensing, *see infra*, Section 3.2.2.

⁸⁷ “*Sui generis*” means “of its own kind” or unique.

⁸⁸ Official Journal of Legislation of the European Union, OJ L 077 (27/03/1996) AT 20–28, <http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A31996L0009>.

⁸⁹ *See* WORLD INTELLECTUAL PROPERTY ORGANIZATION, WIPO SUMMARY ON EXISTING LEGISLATION CONCERNING INTELLECTUAL PROPERTY IN NON-ORIGINAL DATABASES (2002), http://www.wipo.int/meetings/en/doc_details.jsp?doc_id=2296. *See also* INTERNATIONAL ASSOCIATION FOR THE PROTECTION OF INTELLECTUAL PROPERTY, SUMMARY REPORT QUESTION Q182 DATABASE PROTECTION AT NATIONAL AND INTERNATIONAL LEVEL (2004), 3–4, <https://www.aippi.org/download/committees/182/SR182English.pdf>.

⁹⁰ Arts. 1333–1336, Civil Code of the Russian Federation, 18 December 2006 N 230-FZ.

⁹¹ WORLD INTELLECTUAL PROPERTY ORGANIZATION, SUMMARY ON EXISTING LEGISLATION CONCERNING INTELLECTUAL PROPERTY IN NON-ORIGINAL DATABASES SCCR/8/3 (2002), http://www.wipo.int/meetings/en/doc_details.jsp?doc_id=2296.

⁹² Article 7(1), EU Database Directive.

⁹³ Article 7(2)-(5), EU Database Directive.

⁹⁴ The ECJ in Case C-203/02, *British Horseracing Board Ltd. and Others v. William Hill Organization Ltd.*, para. 42 stated:

The expression ‘investment in ... the obtaining ... of the contents’ of a database in Article 7(1) of the directive must be understood to refer to the resources used to seek out existing independent materials and collect them in the database. It does not cover the resources used for the creation of materials

Substantial expenditure on the latter is the only basis to acquire the *sui generis* database protection.⁹⁵

Such an interpretation may translate into non-availability of this type of protection for the generators of many kinds of research data. More specifically, this potentially could be the case when in a particular activity the expenditure on the generation of the data themselves far exceeds the minimal or insubstantial costs of setting-up and maintaining the database in which these data are stored.

The legal merits of the *sui generis* exclusive property right that protects mere investment in factual compilations are not analyzed in this article.⁹⁶ What is important to understand in the present context is that such database protection legislation confers additional statutory rights to database rights holders, whether in the private sector or not, which can be used to enforce additional protections (as discussed further below) in those jurisdictions that have enacted such legislation.

Research data can additionally be protected under legal mechanisms other than copyright or a *sui generis* database right.⁹⁷ Statutory protections in most jurisdictions include laws addressing confidential information, privacy, trade secrets, patents, and unfair competition law. The use of these statutory mechanisms is typically to prevent access to and unrestricted use of the resources without explicit permission.

2.3 Other Forms of Intellectual Property Rights Protection⁹⁸

which make up the contents of a database.

For the details of the Court's reasoning, *see* paras 34-41. OJ C 6, 08.01.2005 at 10. *See also* ECJ Case C-444/02, *Fixtures Marketing Ltd v. Organismos prognostikon agonon podosfairou AE (OPAP)*, para 41, and paras 39-40, online: <http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:62002CJ0444&qid=1467449998481&from=EN>.

⁹⁵ ECJ Case C-203/02, para. 31-32. *See also* FIRST EVALUATION REPORT OF DIRECTIVE 96/6/EC ON THE LEGAL PROTECTION OF DATABASES DG Internal Market and Services Working Paper IP/05/1567 (Brussels, December 12, 2005), para. 4.1.4, which highlights some of the decisions of the European Court of Justice. Online: http://ec.europa.eu/internal_market/copyright/docs/databases/evaluation_report_en.pdf.

⁹⁶ For an analysis of the European Directive on the legal protection of databases in the context of scientific research data, *see* the official review of the EU Database Directive that was prepared by the EC in 2005 (COMMISSION OF THE EUROPEAN COMMUNITIES, DG INTERNAL MARKET AND SERVICES WORKING PAPER, FIRST EVALUATION OF DIRECTIVE 96/9/EC ON THE LEGAL PROTECTION OF DATABASES (2005), http://ec.europa.eu/internal_market/copyright/docs/databases/evaluation_report_en.pdf); *see also*, J. H. Reichman and Paul F. Uhlir, *Database Protection at the Crossroads: Recent Developments and Their Impact on Science and Technology*, 14 BERKELEY TECH. L.J. 793 (1999), <http://scholarship.law.berkeley.edu/btlj/vol14/iss2/14>. For a review of different legal options for protecting scientific databases, while taking due account of research uses of such data, *see* A QUESTION OF BALANCE, at note 69. All these analyses of the EU Database Directive of 1996 were critical of the law as written and implemented, especially for public research.

⁹⁷ M. W. Carroll, *Sharing Research Data and Intellectual Property Law: A Primer*, 13(8) PLOS BIOL (2015), <http://dx.doi.org/10.1371/journal.pbio.1002235>

⁹⁸ There are other legal bases for protecting databases, such as misappropriation under competition law and trespass under property law, but these do not fall under the rubric of intellectual property law and we do not address them here.

There are several other types of IP rights that should be mentioned. Patents are an intellectual property mechanism that potentially pose an obstacle to the sharing of research data. Patents are exclusive rights in inventions. Unlike copyright, patents only arise if the inventor applies for and is granted a patent by a public entity. Premature disclosure of a patent can impair one’s ability to acquire patent protection for an invention. Most research data are not eligible for protection as inventions so sharing of data is less likely to jeopardize a patent application than the publication of an associated article.⁹⁹ In his article “Sharing Research Data and Intellectual Property Law: A Primer,” Professor Carroll notes that there are two cases in which disclosure of facts might impact a patent application: (1) the published article does not describe the invention, but the data do; or (2) the data disclosure fills a gap in other researchers’ knowledge so that inventions that arise from the research are not described by the data, but are rendered “obvious” to one skilled in the art by the disclosure.¹⁰⁰

Some universities invoke trade secret protection (also referred to as proprietary or confidential information) to maintain the confidentiality of their research data.¹⁰¹ In such cases, researchers may not share the data until the research results are published and the data upon which the publication is based becomes accessible to others for verification and reproducibility purposes. According to international standards, national laws may treat information as a trade secret if it is not known to the public, generates economic value from not being publicly accessible, and reasonable measures have been taken to keep it secret.¹⁰² Public disclosure of the information removes any trade secret protection because the information has become generally known.

Finally, commercial contracts, also referred to in commerce as “end-user licensing agreements” (EULAs) are used widely in the marketplace for limiting distribution and use of data. Such legal mechanisms lack the quality of the legislated copyright that grants rights against all third parties, and usually bind only the two parties (the rights holder of the dataset and the other contracting party). In addition, commercial licenses and contracts used in the marketplace as a rule contain restrictive conditions that limit further use and dissemination of the data products.¹⁰³ For these reasons, the typical marketplace forms and uses of contracts and licenses are not discussed further. Rather, these legal instruments are discussed in the following section only in the context of facilitating widespread sharing of research data.

3 REMOVING INTELLECTUAL PROPERTY LAW RESTRICTIONS ON THE USE OF RESEARCH DATA

The most effective way to make data and datasets available without restrictions is to ensure their status within or, where possible, transfer to the public domain. Public domain status enables the unrestricted re-use, re-dissemination, and legal interoperability of data. Moreover, unlike copyright, which lasts for the life of the author plus 50 (or even 70) years, the duration of public

⁹⁹ Carroll MW (2015) Sharing Research Data and Intellectual Property Law: A Primer. PLoS Biol 13(8): e1002235. doi:10.1371/journal.pbio.1002235

¹⁰⁰ *Id.*

¹⁰¹ *Id.*

¹⁰² *Id.*

¹⁰³ *See also* Section 3.3.3 below.

domain status is unlimited. In cases in which datasets do not arise within or enter the public domain by action of law, other legal options can enable the unfettered sharing of research data.

In Section 1.2, we defined public domain as material that is not covered by intellectual property rights.¹⁰⁴ The public domain may be preserved formally through specific national legislation or regulation that expressly ensures that that material cannot be removed from the public domain. This is accomplished by excluding certain subject matter from protection under copyright.¹⁰⁵ The public domain may also be preserved or expanded through establishment of national policies requiring that information created or funded by the government be retained in the public domain. Government contracts or grants may also contain provisions requiring research data and results to be dedicated to the public domain.¹⁰⁶ In each of these instances, while copyright and other intellectual property rights would be disallowed for private parties in the resultant data, other laws, policies, or contract terms could still apply to the public domain works in order to restrict distribution of all or portions of the works in appropriate instances. Appropriate instances might be to protect security, privacy, confidentiality of human subjects, and similar concerns.

Public domain status may be attained as well when the protection of the eligible subject-matter has exceeded the statutory term of protection. This is impossible for all copyrightable datasets which have been “born digital” since they are too recent to have exceeded the term of copyright protection. Data within the public domain may be labeled as such in its metadata or otherwise by government or another entity using a volunteer self-help instrument (e.g., a Creative Commons Public Domain Mark¹⁰⁷) to provide continuing clarity as to its legal status.

Voluntary, private self-help legal instruments may be used by a specific individual or entity (including government agents if within their authority to do so without resort to passage of new laws) to place a specific information object under a legal status that closely approximates the public domain. That is, such instruments waive, dedicate or transfer all intellectual property rights in content to the *commons* to the greatest extent allowed by the law under the controlling legal jurisdiction.

In Section 3.1 we discuss governmental means for achieving public domain or close approximations of public domain status for research data. Section 3.2 discusses private self-help legal instrument methods for achieving commons legal status for research data that similarly closely approximates public domain status for research data.

3.1 Governmental Mechanisms

¹⁰⁴ See J. H. Reichman and Paul F. Uhlir, *A Contractually Reconstructed Research Commons for Scientific Data in a Highly Protectionist Intellectual Property Environment*, 66 DUKE UNIVERSITY LAW SCHOOL, LAW AND CONTEMPORARY PROBLEMS 315, 318-319 (2003). The data may be restricted by “other means” that are not based on any intellectual property rights (e.g., laws protecting national security, confidentiality, and privacy) and that cannot be waived or reduced, *see* notes 21 and 22 and the accompanying text for identification and discussion of such restrictions.

¹⁰⁵ See, e.g., WIPO Copyright Treaty art 2, and the discussion of public domain in Section 1.2 above.

¹⁰⁶ For example, the United States Geological Survey’s National Geospatial Data Center’s data acquisition policy requires that all geospatial data arising from its partnerships with state government and business contain a clause that places resulting data in the public domain.

¹⁰⁷ See CREATIVE COMMONS, CC PUBLIC DOMAIN MARK, <https://creativecommons.org/publicdomain/mark/1.0/>

3.1.1 Treaties and Agreements

Intergovernmental treaties or agreements constitute one broad approach used by governments for placing scientific data in the public domain. Both types of agreements are binding upon the parties under international law. In the United States, however, the domestic procedures for the U.S. to become a party to treaties and agreements differ.¹⁰⁸ Ratification of treaties requires the advice and consent of the U.S. Senate under Article II, Section 2, clause 2 of the Constitution in order for the treaties to be brought into force. International agreements that are brought into force on a Constitutional basis other than with the advice and consent of the Senate are "international agreements other than treaties," often referred to as "executive agreements."¹⁰⁹

Treaties or intergovernmental agreements may be bilateral or multilateral in nature. For example, many countries have bilateral science and technology agreements providing that scientific data arising from cooperative activities under the agreement shall be made publicly available in accordance with accepted scientific practice.¹¹⁰ Another example is in the case of environmental treaties.¹¹¹ For instance, the 1959 Antarctic Treaty established, *inter alia*, that "scientific observations and results from Antarctica shall be exchanged and made freely available."¹¹² Other international treaties in the field of utilization of natural resources and protection of environment that have similar provisions include the 1984 Convention on the Law of the Sea,¹¹³ the 1985 Vienna Convention on the Protection of the Ozone Layer,¹¹⁴ and the 1998 Aarhus Convention.¹¹⁵ The member nations of the Arctic Council are currently negotiating a binding agreement on scientific cooperation to increase research capability and sharing of Arctic scientific data.¹¹⁶

¹⁰⁸ For a discussion of the difference between treaties and executive agreements, see the U.S. State Department webpage, <http://www.state.gov/s/l/treaty/faqs/70133.htm>.

¹⁰⁹ *Id.*

¹¹⁰ For example, Article 10 of the U.S.-France Science and Technology Agreement which entered into force on November 13, 2010, provides that "science and technology information of a non-proprietary nature under the Agreement shall be made, unless otherwise agreed by the Parties, to the world scientific community through customary channels and in accordance with the applicable laws of the Parties and normal procedures of the relevant governmental agencies.", <http://www.state.gov/documents/organization/184717.pdf>.

¹¹¹ For a discussion of these issues in the context of environmental research, see generally Paul F. Uhlir, Robert S. Chen, Joanne Irene Gabrynowicz & Kathleen Janssen, *Toward Implementation of the Global Earth Observation System of Systems Data Sharing Principles*, 35 J. SPACE L. 201 (2009), <http://www.spacelaw.olemiss.edu/jsl/pdfs/articles/35JSL201.pdf>

¹¹² The Antarctic Treaty of 1959, art. III, June 23, 1961, provides that scientific observations and results from Antarctica shall be exchanged and made freely available, <http://www.ats.aq/e/ats.htm>.

¹¹³ United Nations Convention on the Law of the Sea (UNCLOS), (Montego Bay, Jamaica, 10 December 1982, 1833 UNTS 3), http://www.un.org/Depts/los/convention_agreements/texts/unclos/closindx.htm.

¹¹⁴ Vienna Convention on the Protection of the Ozone Layer (Vienna, Austria 22 March 1985, 1513 UNTS 293.), https://treaties.un.org/pages/ViewDetails.aspx?src=TREATY&mtdsg_no=XXVII-2&chapter=27&lang=en.

¹¹⁵ UNECE Convention on Access to Information, Public Participation in Decision-making and Access to Justice in Environmental Matters (Aarhus, Denmark, 25 June 1998, 2161 UNTS 447). The Aarhus Convention entered into force in 2001, https://treaties.un.org/pages/ViewDetails.aspx?src=TREATY&mtdsg_no=XXVII-13&chapter=27&lang=en.

¹¹⁶ The Arctic Council's Task Force for Enhancing Scientific Cooperation in the Arctic, which is negotiating the text of the agreement, met for the sixth time in Copenhagen in August 2015, <http://www.arctic-council.org/index.php/en/our-work/2/8-news-and-events/360-sctf-copenhagen>. The agreement draft text is not publicly available but the final text should be finalized and made publicly available by 2017.

International agreements establishing large-scale international scientific facilities, such as astronomical observatories¹¹⁷ or ocean observing systems¹¹⁸ typically provide that the scientific data generated at the facility will be made publicly available for use by researchers worldwide, perhaps after some defined embargo period. Some of these agreements are binding under international law while others are non-binding and rely upon the political good will and mutual interest of the partners to carry out their obligations under the agreement.

Most of the treaty and executive agreement examples cited above do not eliminate all intellectual property rights in the research data covered by the formal agreements. Rather they are illustrative of approaches that have been used to achieve close approximations of public domain status for research data in the past. Similar treaty and international agreement devices could be used to completely achieve public domain or “no rights reserved” commons status for research data.

Even though they have limitations, as discussed further below, agreements among intergovernmental organizations may be used in a similar fashion as a means for bringing data into alignment with public domain and commons principles. This might occur when two or more government parties contractually agree that the results of their joint or cooperative efforts will be placed into an open access legal framework. This is typically done in the research data setting by institutions that enter into a Memorandum of Understanding (MoU), which establishes data-release policies and practices.

3.1.2 Policies of Intergovernmental Organizations

Intergovernmental organizations can also play an important role by developing guidelines or principles to promote the dissemination of scientific data, either broadly for all kinds of research data or in a certain discipline or application area. International governmental organizations and intergovernmental organizations are primarily or exclusively composed of sovereign state members. Both can be established by a treaty or charter that confers the organization with international legal personality.¹¹⁹

¹¹⁷ See, e.g., the Atacama Large Millimeter-Submillimeter Array, an international astronomy facility located in Chile is a partnership of the European Southern Observatory, the U.S. National Science Foundation, and Japan’s National Institutes of Natural Science, in cooperation with Canada, Korea, Taiwan, and Chile Its data sharing policy can be found at <https://almascience.eco.org>. Data is hosted in three places: an archive data, a calibrator catalogue and a science verification database. After a proprietary period for project teams to analyze their data and publish their research results, data are available to the broader global research community. CERN, the European Organization for Nuclear Research, launched its Open Data Portal in October 2014. For the first time, data produced by experiments at the Large Hadron Collider were made publicly available to the global research community. Open source software is also available to read and analyze the data. The data are shared under the Creative Commons CC0 instrument and the datasets are assigned unique identifiers so that they can be cited in articles. Available at: <http://home.cern/about/updates/2014/11/cern-makes-public-first-data-lhc-experiments>.

¹¹⁸ See, e.g., The National Ocean Discovery Program is an international research program to bring together the Earth, ocean, Atmospheric and the life sciences. The United States and Japan each provide two permanent ocean drilling platforms to collect data from the oceans. The program has adopted Principles of Scientific Investigation, International Ocean Discovery Program which include data sharing. The Program provides open data to all expedition samples and data once the science party has had an opportunity to complete their initial work (typically one year). Available at: <http://www.iodp.org/principles-of-scientific-investigation>.

¹¹⁹ For a list of intergovernmental organizations, see the YEARBOOK OF INTERNATIONAL ORGANIZATIONS,

An international approach to data sharing can lead to more coherent and harmonized implementation mechanisms. It can promote a more uniform interpretation of the major terms and concepts regarding data sharing and promote technical and legal interoperability, regardless of whether the sources of the data are from the public sector, in academia or government.

One example of an intergovernmental body that does have an open access data policy is the Group on Earth Observations (GEO).¹²⁰ GEO was established in 2005 to promote international access to international Earth observation data. In 2010, GEO created an open category of data, the Group on Earth of Observation System of Systems Data Collection of Open Resources for Everyone (Data-CORE).¹²¹ To promote access to these datasets and enable use and reuse of the data without restrictions, GEO adopted fundamental principles for data sharing, expanding the worldwide trend towards open data worldwide. Updated in 2015, the GEOSS Data Sharing Principles provide that data, metadata and products will be shared as open data by default, by making them available as part of Data-CORE without charge or restrictions on reuse, subject to the contributing agency conditions of registration and attribution when the data are reused.¹²²

The Global Biodiversity Information Facility (GBIF), an intergovernmental initiative of countries, economies, and international organizations collaborating to promote free and open access to biodiversity data, is another example of use of MOUs to organize publication and use of biodiversity according to common standards. Participants sign a GBIF Memorandum of Understanding that sets forth terms for participation, including the provision and use of biodiversity data.¹²³ GBIF also offers its participants' data through its portal with the use of an agreement that stipulates unfettered reuse of data as a default rule, subject only to the condition of attribution.

The International Arabidopsis Genome Research Project also illustrates how a highly effective research data sharing environment may be accomplished through informal coordination and collaboration among government entities and the international research community with few to no legally enforceable agreements among the cooperating organizations. In this project, several national funding agencies and the international scientific community partnered to advance the scientific field of plant genomics, developed enabling technologies for sequencing, and made the sequencing data publicly available.

Commencing in the late 1990s, funding agencies in the United States, United Kingdom (funded by the European Commission), France, and Japan began sequencing the arabidopsis genome to advance fundamental understanding of flowering plants. Arabidopsis was chosen because it did not have significant commercial value as a crop, but could serve as a model for studies of all plants,

<http://www.uia.org/yearbook>.

¹²⁰ The intergovernmental Group on Earth Observations has 101 member countries and the European Union, but does not have its own international legal personality. For more information about GEO, *see*, <https://www.earthobservations.org/index.php>.

¹²¹ GROUP ON EARTH OBSERVATIONS (GEO), GEO DATA SHARING PRINCIPLES IMPLEMENTATION, https://www.earthobservations.org/geoss_dsp.shtml.

¹²² The revised 2015 GEOSS Data Sharing Principles may be found at <https://www.earthobservations.org/dswg.php>

¹²³ Global Biodiversity Information Facility (GBIF), available at: <http://www.gbif.org>.

including rice and maize.¹²⁴ A Multinational Arabidopsis Steering Committee was established. The Committee, in conjunction with the research funding agencies, created public repositories and globally disseminated the genetic sequencing information. This project moved the field of plant biology to the age of systems biology and led to the development of better tools to sequence the function of plant genomes. These coordinated activities have been carried out for over 20 years without any official signed agreements. Research funding agency grant conditions, which are discussed in greater detail in Section 3.1.3, played a significant role in making the arabidopsis sequencing data open by requiring the grantees to make their data publicly available.

Although the examples just discussed do not actually achieve public domain legal status for research data, they do enable the greater sharing of research data. The policies of intergovernmental organizations may be used to achieve even greater legal clarity to minimize the potential for future ownership and use challenges and thereby promote even higher levels in the use and sharing of research data.

3.1.3 National Legislation, Regulations, and Research Data Policies

Two primary approaches can be used by governments to place their data in the public domain.¹²⁵ The first is the option of international “intervention” whereby norms of international public law codify what data shall be made available to users without any restrictions on subsequent use. These international obligations or norms can then be implemented through national legislation and subsequent administrative regulations. Such an approach may result in a more coherent treatment of data in a certain subject matter area or activity and facilitate global sharing of the data in that context.

The other option is for countries to develop their own open access data management norms, irrespective of any international treaty or executive agreement, to adopt legislative, regulatory, or policy measures that make certain types or categories of data publicly available, free of encumbrances on their use.¹²⁶ Such national leadership on open access policies may influence other countries to follow suit. At a minimum, such approaches can have substantial impact in the nations in which they are enacted.

There is collective experience and a growing realization among many national governments that open access to government and government funded data increases transparency in government, accelerates scientific discovery, and stimulates economic growth in society.¹²⁷ National governments providing free and widespread open access to their data view such data as a public asset whose value increases with increased use by both scientific and commercial users. To achieve their public-interest goals, formal national legislation, regulations, and policies are being used to make government data more readily accessible.

¹²⁴ *Mini case study: International Arabidopsis Genome Research Project*, in OECD, MEETING GLOBAL CHALLENGES THROUGH BETTER GOVERNANCE: INTERNATIONAL CO-OPERATION IN SCIENCE, TECHNOLOGY AND INNOVATION 239 (OECD PUBLISHING 2012), <http://dx.doi.org/10.1787/9789264178700-en>.

¹²⁵ See RDA-CODATA Interest Group, at note 24. Principle Five on “Harmonization” has summary recommendations of “top-down” and “bottom-up” approaches. *Id.*

¹²⁶ *Id.*

¹²⁷ See Section 1.1.

From a legal standpoint, datasets may be dedicated to the public domain most effectively through national legislation. National governments can enact laws that provide exemptions from intellectual property protection. An example is Section 105 of U.S. Copyright Law, which prohibits federal government agencies and employees from claiming copyright in government works.¹²⁸ This has very far-reaching effects because of the huge volume of data and information that the U.S. government produces.¹²⁹

Other pieces of U.S. legislation that are germane in this context are the Paperwork Reduction Act of 1980,¹³⁰ which established the Office of Information and Regulatory Affairs (OIRA) at the White House Office of Management and Budget (OMB), and authorized it to oversee the federal government's collection of information from the public and to establish information policies. With regard to the latter, OIRA established OMB Circular A-130 in 2000 developing policy for the management of Federal information resources, including federal research data, consistent with the 1976 Copyright Act and other laws.¹³¹ The OMB includes procedural and analytic guidelines for implementing specific aspects of these policies as appendices.¹³²

Some European countries have also revised their copyright legislation to enhance access to publicly funded research. Germany amended its Act on Copyright and Related Rights on October 1, 2013 to allow publicly funded researchers to retain the legal right to post their publications on line, even if they have transferred their rights to publishers, twelve months after publication by the publisher.¹³³ The United Kingdom passed several amendments to its copyright legislation that slightly modified the statutory exceptions to copyright to reflect the digital environment. The revisions, which took effect in 2014, allow greater reuse of copied materials for educational and non-commercial research purposes, including data and text mining analysis. Finland and Australia are also considering revisions to their copyright laws to increase access to publications reporting on results of publicly funded research.¹³⁴

Regulations issued by ministries or agencies of national governments are used to implement legislation. Some legislation may be specific to open access of data or consistent with the original “organic act” that established the governmental organization and set out its purpose.¹³⁵

¹²⁸ 17 U.S.C. § 105 (1976), <http://www.copyright.gov/title17/92chap1.html#105>.

¹²⁹ See e.g. www.data.gov, the U.S. Government's website for open data. To date, more than 180,000 Federal datasets and collections have been made available on [data.gov](http://www.data.gov). Impact Report: 100 Examples of President Obama's Leadership in Science, Technology and Innovation, June 21, 2016.

¹³⁰ 44 U.S.C. §§ 3501–3521 (1980), as amended in 1995 in 44 U.S.C. Chapter 35.

<https://www.gpo.gov/fdsys/granule/USCODE-2008-title44/USCODE-2008-title44-chap35-subchapI-sec3501>.

¹³¹ https://www.whitehouse.gov/omb/circulars_a130_a130trans4/.

¹³² *Id.*

¹³³ See Article 38(4) of the German Copyright Act. A description of the German amendment is found at <http://www.ip-watch.org/2013/07/02/german-parliament-passes-changes-to-national-copyright-law/>.

¹³⁴ OECD, MAKING OPEN SCIENCE A REALITY, OECD SCIENCE, TECHNOLOGY, AND INDUSTRY POLICY PAPERS, NO. 25 13-14 (OECD Publishing Paris 2015) <http://dx.doi.org/10.1787/5jrs2f963zs1-en>.

¹³⁵ See

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/448269/Exceptions_to_copyright_-_An_Overview.pdf;

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/375954/Research.pdf.

In the United States, federal departments and agencies all issue myriad regulations consistent with the legislation set forth in the preceding subsection, other pertinent legislation, and their mission.

The European Union both obligates its institutions and the Member States (via their public authorities) to share information they produced. On the European Union level, the relevant document is Commission Decision on the reuse of Commission documents¹³⁶ that lays down “conditions for the reuse of documents held by the Commission or on its behalf by the Publications Office of the European Union with the aim of facilitating a wider reuse of information, enhancing the image of openness of the Commission, and avoiding unnecessary administrative burdens for reusers and the Commission services alike” (Article 1). The framework imposed on the Member States is provided in the Public Sector Information (PSI) Directive¹³⁷ that encourages the Member States to make as much information available for re-use as possible. It provides rules regarding charges for PSI and its re-use, prohibits enforcement of discriminatory conditions of re-use, encourages use of standard licenses that promote re-use and value-adding activities. Its provisions are currently being updated. In addition, INSPIRE Directive and regulatory documents, Copernicus (GMES) Regulation, Directive on public access to environmental information,¹³⁸ and many other pieces of *acquis communautaire* contribute to strengthening of the trend towards open data and adopting it as a default standard for access and use of specific types of data and information.

In 2014, the European Union adopted the European Commission’s delegated regulation supplementing Regulation (EU) No 911/2010 of the European Parliament and of the Council on the European Earth Monitoring Programme (GMES) by defining criteria for access to GMES data.¹³⁹ The regulation places data from Copernicus, the European Union’s Earth observation program, in the public domain. Article 3 states that “GMES dedicated data and GMES service information” shall be available for users to access freely, openly and fully. Recital 7 of the regulation confirms the commitment of full compatibility of the Copernicus data sharing regime with the GEOSS Data Sharing Principles. The regulation states that “...GMES open dissemination should be fully compatible with the GEOSS data sharing principles.” Thus as a condition for full participation, the regulations are determinative. For contributing missions from third parties, however, where data comes not from Sentinel satellites, the data policy will be determined by the provider.

Depending upon their national legal framework, government can place government-generated or funded data in the public domain through the adoption of national policies. In the United States, for example, several White House policies have made open data the default for government data.¹⁴⁰ On May 9, 2013, President Obama signed an executive order that made open and

¹³⁶ Commission Decision on the reuse of Commission documents of 12 December 2011, <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2011:330:0039:0042:EN:PDF>.

¹³⁷ Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information. OJL 345/90, <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2003:345:0090:0096:EN:PDF>.

¹³⁸ Directive 2003/4/EC of the European Parliament and of the Council of 28 January 2003 on public access to environmental information.

¹³⁹ No. 1159/2013 of 12 July 2013. OJ L 309/1 (19.11.2013).

¹⁴⁰ Executive Order 13642, Making Open and Machine Reusable the New Default for Government Information,

machine-readable data the new default for government information.¹⁴¹ A February 2013 White House Office of Science and Technology Policy Memo “Increasing Access to the Results of Federally Funded Scientific Research” promotes access to research data arising from federally funded research.¹⁴² The policy requires Federal agencies that spend over \$100 million per year on research and development to develop plans and policies to make unclassified peer-reviewed publications, metadata, and data resulting from their funding publicly accessible within a year of publication.¹⁴³

National policies can also focus on making particular types of research data openly available. The U.S. White House issued a National Strategy for Earth Observations on April 19, 2013, to increase the efficiency and effectiveness of the Earth observation systems in the United States.¹⁴⁴ Consistent with U.S. global climate research program data principles dating back to 1991,¹⁴⁵ the National Strategy reaffirms that Earth observations data should be fully and openly available to all users promptly, in a non-discriminatory manner, and generally free of charge whenever possible.¹⁴⁶ As more countries adopt open access data sharing policies for their Earth observation data, there will be far fewer legal interoperability challenges in the use of such data.

National funding agencies may also require recipients of contracts, cooperative agreements, or grants to place their data in the public domain and deposit it in a national repository as a condition of the funding instruments. For example, the United States Geological Survey’s National Geospatial Data Center’s data acquisition policy, which requires that all geospatial data arising from its partnerships with state government and business contain a clause that places resulting data in the public domain.¹⁴⁷

3.2 Voluntary “Self-Help” Approaches

Many organizations and agencies—and even individual researchers—may not need to wait for government entities to authorize removal of intellectual property law restrictions on the use of

May 9, 2013; OMB Memorandum to Federal Agencies, Open Data Policy-Management Information as an Asset (M-13-13, May 9, 2013).

¹⁴¹ *Id.*

¹⁴² See https://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf.

¹⁴³ *Id.*

¹⁴⁴ See http://www.whitehouse.gov/sites/default/files/microsites/ostp/nstc_213_earthobsstrategy.pdf

¹⁴⁵ The White House Bromley Principles were adopted in 1991 to promote sharing of global change Earth observation data.

¹⁴⁶ At one time, the United States charged for Landsat images. An attempt to commercialize Landsat data resulted in higher prices and diminished use of the images. See, for example, NATIONAL RESEARCH COUNCIL, BITS OF POWER: ISSUES IN GLOBAL ACCESS TO SCIENTIFIC DATA 121-124 (National Academy Press, 1997). Since 2008, the Landsat data archived in the United States, have been available without any cost or restrictions, worldwide. For an assessment of the benefits of this policy change for Landsat data, see Janice S. Nelson, comp. *U.S. Geological Survey (USGS) Earth Resources Observation and Science (EROS) Center—fiscal year 2010 annual report*. U.S. Geological Survey Open-File Report 2011-1057 (2011). Available at: <http://pubs.usgs.gov/of/2011/1057/pdf/of2011-1057.pdf>.

¹⁴⁷ TIM SAULTZ, MARK GEWINNER & JIM ALMEKINDER, USGS, USGS GEOSPATIAL LIAISON GUIDE TO THE USGS GEOSPATIAL PRODUCT AND SERVICES CONTRACT (ver. 2.2 2012) http://thor-f5.er.usgs.gov/GPSC/Draft_State_Liaison_Guide_to_GPSC.pdf.

their data by others. If the current restrictive policy is an internal policy developed and implemented by the agency or organization and there are no overriding government regulations or legislation forcing the organization to follow the current restrictive policy, the organization may have the option of simply altering the policy through internal processes.

Various self-help approaches that the agency, organization, or individual researcher might consider include the use of waivers, licenses, contracts or agreements, and institutional policies. Below is a brief discussion of the applicability of these approaches to research datasets, and the benefits and drawbacks of each approach.¹⁴⁸

3.2.1 *Standard Waivers of Rights*

Waivers of rights are an express written statement by the rights holder that no intellectual property rights will be enforced by that rights holder in the dataset or other work protected by copyright or a *sui generis* database right. This is the most favorable condition for the user of the dataset, since it provides a status closely equivalent to the public domain in many jurisdictions, and allows complete freedom for any user to integrate, re-use, and disseminate further all or a portion of the dataset. It provides full interoperability with no restrictions, except for those that cannot be waived by law. It retains no protection for the dataset provider, including no legally enforceable attribution or any other requirement. The lack of a legally enforceable attribution requirement within a standard waiver may not have much practical adverse effect in most cases, since attribution and citation are normative and ethical practices in the academic and scientific communities. Also, many jurisdictions do not allow the waiver of all intellectual property rights, because the author's moral rights, if applicable, cannot be fully waived.¹⁴⁹ A standard legal instrument in broad use across the globe that uses waiver as one of its critical elements is contained within the Creative Commons No Rights Reserved (CC0) instrument (*see* Table 1 in Part 4, below).

3.2.2 *Standard Common-Use Licenses*

Licenses and contracts¹⁵⁰ are used if the data provider wishes to retain some rights and control the use(s) of the data in some way. There is a popular misconception, however, that licenses and contracts are the same thing. They are not.¹⁵¹ The differences between the two will become clear from the discussion below.

Standard common-use licenses, which are voluntarily adopted by the licensor, are based upon existing statutory intellectual property rights for implementation.¹⁵² They are applied automatically and do not depend on a negotiated agreement between the rights holder and the

¹⁴⁸ For more detail about voluntary waivers of rights, common-use licenses, and contracts *see* Sarah Hinchliff Pearson, *Three Legal Mechanisms, in* FOR ATTRIBUTION: DEVELOPING ATTRIBUTION AND CITATION PRACTICES AND STANDARDS, 71-75 (Paul F. Uhlir, ed., 2012), (hereinafter Pearson 2012). *See also* <https://creativecommons.org>.

¹⁴⁹ *See* Section 2.1.1 at note 53.

¹⁵⁰ Contract and agreement are used interchangeably in this article.

¹⁵¹ *See, e.g.,* Christopher M. Newman, *A License is Not a 'Contract Not to Sue': Disentangling Property and Contract in the Law of Copyright Licenses*, 98:2 Iowa L. Rev, (2013).

¹⁵² For a discussion of licenses, *see* Pearson 2012, note 148.

user(s). They do not extend to facts or materials already in the public domain, because there is no underlying statutory protection for that material, but can apply to datasets or protectable portions of datasets, although the uncertainty of enforcing the rights when necessary remains. Finally, common-use licenses can be used to decrease the levels of protection within the scope of the statutory protection. Because such licenses depend on statutory foundations, the effect of the language of a license may vary across jurisdictions.

It is important to note that an attribution requirement may not be legally enforceable for all data used in all jurisdictions. In those cases where it is not, attribution may be seen as a standard community practice or norm, or a moral or ethical imperative. Although it is not exactly the same as a legally enforceable requirement, proper attribution is the accepted standard practice in the scientific and larger academic community.¹⁵³ Failure to provide proper attribution may rise to the level of plagiarism, a form of research or academic misconduct and possibly an infringement of copyright.

There are many kinds of standard-form licenses with the primary differences being the amount of rights that are granted to the licensee by the licensor. Some standard common-use licenses permit only a restricted number of uses, while others grant the user greater freedom to reuse the work, with only some rights reserved by the licensor. The latter standard licenses are broadly referred to as “common-use” licenses. Moreover customized licenses, which we do not endorse for research data because they typically inhibit the widespread sharing of such data, can be created by any provider with any mix of terms and conditions, depending on the type of data and jurisdiction.¹⁵⁴

The most widely used and prevalent set of common-use licenses is the set offered by Creative Commons (CC).¹⁵⁵ According to the CC website all Creative Commons licenses have three layers of description. One is a “machine readable” version which is tagged to the licensed work for online use in CC Rights Expression Language (CC REL). Another is called the “Commons Deed,” which is a “human readable” version of the license that summarizes the most important terms and conditions for non-legal experts. The final layer is the “legal code,” a traditional legal tool in language that lawyers understand.¹⁵⁶ The CC licenses have been reviewed in over 70 countries and are in use throughout the world. Although the exact number of works that use a CC license, mark or instrument is not known, it is estimated that there are over one billion online.¹⁵⁷

Creative Commons licenses are intended for use only with original and creative works. Because many datasets have no original and creative elements, and because of the pragmatic issues

¹⁵³ For example, the Group on Earth Observations (GEO) has adopted GEOSS Data Citation Guidelines that can help in this regard. *See*: http://www.gstss.org/library/GEOSS_Data_Citation_Guidelines_V2.0.pdf. Other organizations dealing with research data in different jurisdictions have similarly adopted such guidelines.

¹⁵⁴ *E.g.*, according to Directive 2009/24/EC of the European Parliament and of the Council of 23 April 2009 on the legal protection of computer programs, Article 5(2) “the making of a back-up copy by a person having a right to use the computer program may not be prevented by contract.”

¹⁵⁵ *See* <https://creativecommons.org>.

¹⁵⁶ *See* <https://creativecommons.org/licenses/>.

¹⁵⁷ Presentation by Puneet Kishor at the Research Data Alliance conference, in Dublin, Ireland, 28 March 2014. The website flickr.com alone has over 350 million photographs tagged with CC licenses.

discussed under section 2, we do not recommend the use of the Creative Commons Attribution License or more restrictive common-use licenses for general use with datasets even if the data content, or its selection and arrangement, includes some minimal creativity. Even minimal restrictions raise legal interoperability challenges due to accumulating restrictions caused by the common day-to-day extraction of data from multiple databases and recombination for scientific purposes. Not knowing or resolving the legal status of the resulting data compilations results in substantial impediments to the sharing and use of research data.¹⁵⁸ Actually tracking the legal status of millions of datasets on an Internet-wide basis such that searches would be capable of automatically determining the legal conditions under which each data selection is available, constraining data searches to user-desired legal conditions and then carrying the legal conditions forward in derivative works has not yet been achievable and would create substantial technological and bureaucratic overhead with very little return in value.¹⁵⁹

If attribution is desired by providers for research data, as is usually the case, the citation to parent datasets as a normative practice in the academic and scientific communities can fulfill an equivalent role.¹⁶⁰ Data citation may be promoted through development of efficient technological means for facilitating the practice, inclusion of the practice in codes of conduct, and incorporation into the requirements of research publication outlets.¹⁶¹

3.2.3 *Private Contracts and Agreements*

In contrast to common-use licenses, negotiated contracts are based upon the express agreement, or assent, of the parties. Such contractual agreements usually require formal offer, acceptance, consideration, and (usually) written terms.¹⁶² Formal offer and acceptance for datasets and other digital information products are typically made with “click through” agreements online, or by means of “shrink wrap” agreements, although their validity can vary across jurisdictions.¹⁶³

Unlike common-use licenses supported by intellectual property laws, these contracts are not dependent on their enforcement for an underlying statute, although of course they must not be made for an illegal purpose. Also unlike common-use licenses, they can, depending on the

¹⁵⁸ For a more detailed discussion of challenges of each of various standard open access instruments when used with research data, see *DCC Guide on How to License Research Data* at note 31,

¹⁵⁹ The authors thus do not recommend for use with data most of the licenses listed as Conformant Licenses (<http://opendefinition.org/licenses/#Data>) linked by the 2010 Panton Principles for open data in science (P. MURRAY RUST, ET.AL., PANTON PRINCIPLES: PRINCIPLES FOR OPEN DATA IN SCIENCE, <http://pantonprinciples.org/>), but concur in the strong recommendation for use of CC0 or PDDL.

¹⁶⁰ See the DATA CITATION SYNTHESIS GROUP: JOINT DECLARATION OF DATA CITATION PRINCIPLES (M. Martone ed., 2014), <https://www.force11.org/group/joint-declaration-data-citation-principles-final>. See also NATIONAL RESEARCH COUNCIL, FOR ATTRIBUTION: DEVELOPING ATTRIBUTION AND CITATION PRACTICES AND STANDARDS, (Paul F. Uhlir ed., 2012), <http://www.nap.edu/catalog/13564/for-attribution-developing-data-attribution-and-citation-practices-and-standards>. See generally, the Data Cite organization’s website at <https://www.datacite.org/node>.

¹⁶¹ *Id.* See also CODATA-ICSTI Task Group on Data Citation Standards and Practices, *Out of Cite, Out of Mind: The Current State of Practice, Policy, and Technology for the Citation of Data* (Yvonne M. Socha, ed) 12 *Data Science Journal CIDCR1–CIDCR75* (2013), https://www.jstage.jst.go.jp/article/dsj/12/0/12_OSOM13-043/_pdf.

¹⁶² Restatement (Second) of Contracts (1981).

¹⁶³ For an overview of these practices, see, e.g., MARGARET JANE RADIN, *BOILERPLATE: THE FINE PRINT, VANISHING RIGHTS AND THE RULE OF LAW* (2013).

jurisdiction, apply to data otherwise unprotected by statute (e.g., to factual material in the public domain).¹⁶⁴ Contracts are only valid for the agreeing parties, so others who may obtain the dataset are not bound by the terms of the original agreement. This makes contracts susceptible to leakage and they can therefore be an uncertain mechanism for rights holders. Finally, in contrast to common-use licenses, contracts and agreements are not standard. They may be lengthy and frequently are so confusing that their terms are misunderstood or even ignored by the user.¹⁶⁵ It is for all these reasons that we discourage the use of private contracts by single researchers or organizations.

However, private contracts or agreements also may be used as a means to enhance access to research data in instances of research projects or consortia that involve multiple participating organizations or researchers. For example, those contributing data to a collaborative project in citizen science or for public benefit might be required to click a contractual agreement that they are contributing their data subject to use by others under a common-use license.¹⁶⁶ In other cases, collaborating organizations might strive to require through contract that all research data generated in a mutual project should be made available to the world under, for example, a CC0 No Rights Reserved instrument within a specified period of time. Other research data sharing contracts among research organizations have been used to provide privileges to consortium members as well as access to data for non-consortium members.¹⁶⁷ Thus, in certain circumstances, negotiated non-standard contracts have played and continue to play a substantial role in enhancing access to research data.

Even though organizations might agree to provide open and efficient access to research data through a federated effort, researchers are often confronted with a battery of individual organizational “click licenses” or “conditions of use” that conflict with the federation agreement and each other when the researchers attempt to actually extract and combine data from multiple

¹⁶⁴ For the law on this is the United States, for example, *see, e.g.*, Public Domain Sherpa, available at: <http://www.publicdomainsherpa.com/license-agreement.html>.

¹⁶⁵ *See generally* Pearson 2012, note 148.

¹⁶⁶ For a review of illustrative citizen science projects, *see* ANNE BOWSER AND LEA SHANLEY, *NEW VISIONS IN CITIZEN SCIENCE* (Woodrow Wilson Center, 2013), <http://www.wilsoncenter.org/sites/default/files/NewVisionsInCitizenScience.pdf>. For a government perspective on the future importance of citizen contributions to science, *see* <https://www.performance.gov/content/invest-strategically-public-participation-science-technology-engineering-and-mathematics?view=public#overview>. For an example of crowdsourcing initiated to benefit the public, but also now of great benefit to science, *see* Open Street Map at <https://www.openstreetmap.org/>.

¹⁶⁷ One example is an agreement controlling access to the Atacama Large Millimeter/Submillimeter Array (ALMA), an international astronomy facility located in Chile, funded by the U.S. National Science Foundation, the European Southern Observatory, and Japanese National Institutes of Natural Sciences (NINS), in cooperation with Chile. A trilateral agreement signed in Tokyo, Japan, on December 15, 2015 establishes the framework for the next 20 years of ALMA’s operation. Article 18.2 governing data policies provides that: “All data taken by ALMA shall be jointly owned by the Parties. Ownership shall not impact the free access to the data for use by observers and the astronomical community according to a policy set by the Board.” The Board allows principal investigators from any country to submit proposals to use the observatory facilities. Principal Investigators have exclusive rights to use their observational data for one year, after which the data becomes publicly accessible. Although this agreement is among government actors, the same legal device approach may be used by nongovernmental consortia or research programs for the data they produce and desire to make available to external users. *See* Atacama Large Millimeter-Submillimeter Array at <https://almascience.eco.org>.

sites identified and linked through the federated resource.¹⁶⁸ Adhering to or resolving the conflicting license provisions and conditions of use are often far greater challenges for researchers than that of resolving the technological interoperability challenges. This greatly diminishes the use and value of the federated resource for the research community.

3.2.4 Institutional Policies¹⁶⁹

Research universities can adopt policies to make the data resulting from research conducted by their researchers (faculty and students) publicly available and assist them in negotiating with journals to keep their works publicly available. University policies and employment contracts typically allocate IP rights arising from research conducted by their faculty. These formal policy and contract terms vary across universities but can be written to promote open access to research data. Universities can also encourage agreements among researchers to place their data in a public repository, i.e., as a default rule.¹⁷⁰ For example, Columbia University's Center for International Earth Science Information Network (CIESIN) has a data policy that offers unrestricted access to and use of data without charge unless otherwise specified in the documentation accompanying particular data. The policy provides that data users should acknowledge CIESIN as the source of the data¹⁷¹

Most organizations have the ability to impose policies on members of their organizations. While such policies may or may not rise to the level of imposition of legal obligations, breach of such policies may result in sanctions on a breaching party up to and including dismissal from the organization. By example, many universities have developed formal codes of conduct for employees and students, intellectual property policies, and human research protection requirements to which employees and students are required to adhere. Some of these affect the handling of research data. Funding agencies similarly often impose research data dissemination and access policies in their proposal solicitation guidelines. It is common knowledge within the science community that violations of such funding agency requirements will often result in pragmatic sanctions such as unfavorable reviews or negative reactions towards future funding as opposed to legal sanctions. Such university and funding agency institutional policies for faculty and students are often far more effective for promoting the open sharing of research data than would be copyright enforceability of such provisions.¹⁷²

¹⁶⁸ See *GLOBAL EARTH OBSERVATION SYSTEM OF SYSTEMS PORTAL* (GEOSS PORTAL), http://www.geoportal.org/web/guest/geo_home_stp. See also, the Australian Urban Research Infrastructure Network (AURIN), described in Richard O. Sinnott, et al., *The Australia Urban Research Gateway*, 27 (2) CONCURRENT COMPUTATION: PRACTICE AND EXPERIENCE 358 (2015).

¹⁶⁹ For analogous institutional policies around open access publications, see Jorge L. Contreras, *Confronting the Crisis in Scientific Publishing: Latency, Licensing and Access*, SANTA CLARA LAW REVIEW 53:491 (2013)

¹⁷⁰ For example, the international community's agreement to share gene sequencing of the plant *Arabidopsis* is more fully described in *Appendix B, Case Study on International Arabidopsis Genomic Research Project*, in OECD, MEETING GLOBAL CHALLENGES THROUGH BETTER GOVERNANCE: INTERNATIONAL CO-OPERATION IN SCIENCE, TECHNOLOGY AND INNOVATION (2012), <http://dx.doi.org/10.1787/9789264178700-en>.

¹⁷¹ CIESIN's Data Policy may be found at <http://www.ciesin.org/documents/CIESINDataPolicy.pdf>.

¹⁷² Reviewer comments have suggested that a lack of copyright or contract enforceability may result in failure to achieve desired behavior regarding open access to research data and data citation practices in the academic community. Experience in the normative behavior of the academic community strongly suggests otherwise, however. *On the Origin of the Species* by Charles Darwin is in the public domain (i.e. has no IP or contract

Many universities and research institutions have established policies promoting open access to the published scholarly literature of science. However, far fewer have developed policies promoting open access to the core data upon which the reported research results are based.¹⁷³ This situation makes it far more difficult to verify, critique or replicate the past scientific work or to apply further analyses that original investigators did not foresee. This policy gap should be addressed.¹⁷⁴

The ease or ability for an academic researcher or scholar to place the data sets upon which their research conclusions are based into a long-term open access data repository is highly dependent on the science domain and institution with which their work is affiliated. The same holds true for graduate students completing thesis or dissertation work. Some science domains support the ability of scholars to deliver research data to and extract them from one or more open access repositories, while others do not. Some universities support effective open access repositories for the work of their scholars or are members of consortia supporting such repositories, while other universities are not. Some scholarly journals in some domains provide facilities for hosting supplemental materials such as research data files, while others do not. However, even if available through a science domain, university or scholarly publisher, few of these facilities provide the full set of minimum operational capabilities recommended at the end of this article. These availability and capability gaps should be addressed.

4 CONCLUSIONS AND RECOMMENDATIONS

Open sharing of research data has great benefits for science and for improving societal conditions globally.¹⁷⁵ The foregoing analysis leads to a number of conclusions and recommendations.

Many governments and funders of research; publicly funded research organizations, such as universities and data centers; and individual scientists throughout the world are producing research data that meet high technical standards and that are openly available online for all who may want to access those data. This action alone, however, is insufficient to allow efficient and effective use by others. As described in this article, research data found openly available on the web acquire any statutory intellectual property rights of the jurisdiction, such as copyright and *sui generis* database rights, that can greatly hinder their use by others. Such protections

protection) and yet if another scholar were to claim the work or the ideas as their own, they would be likely dismissed from their employment and ridiculed by the broader academic community. Violations of community norms and institutional policies against plagiarism and failure to meet grant conditions are easy to discover through the Internet and growing easier as technology advances.

¹⁷³ Out of the 770 open access policies of universities and research institutions listed in the Registry of Open Access Repository Mandates and Policies (ROARMAP) at <http://roarmap.eprints.org/cgi/search/advanced>, only 51 universities are listed as including data within their open access policies (last visited Feb 15, 2016).

¹⁷⁴ One group addressing this gap is the Scholarly Publishing and Academic Resources Coalition (SPARC). The SPARC Research Data Sharing Policy Initiative “is developing a channel for productive, ongoing collaboration among funders and stakeholders to shape academy friendly research data sharing policies for research data across and between disciplines as they are being developed and ultimately, implemented.” (<http://sparcopen.org/our-work/research-data-sharing-policy-initiative/>)

¹⁷⁵ See Section 1.1, at notes 1-18.

arise automatically unless expressly excluded, waived, or modified. In order to relieve users of the burden of clearing any potential intellectual property rights in data prior to their use, those governments, organizations, and individuals making data accessible should ensure that both the providers and the users of those data have clear and effective legal rights to use their data.

4.1 Legal Mechanisms that Promote Open Access and Legal Interoperability

Public domain status is the best legal option for enabling and securing unrestricted re-use, further dissemination, and legal interoperability of data and information. On an international level, for data collected or created formally under treaties or executive agreements or other governmental mechanisms that confer public domain status are the preferred option because they provide the most certainty for users and apply to all users equally within those jurisdictions.

Datasets already having public domain status should include a notice in their metadata or on the database owner's server informing potential users of their public domain status. The Creative Commons Public Domain Mark, noted below, serves this purpose. That notice can help to overcome the incorrect assumption by some potential users that the data are subject to protection and have attendant restrictions on reuse. Such a notice will thereby promote the further use of the data and legal interoperability.

Many datasets, however, do not have public domain status and are protected in whole or in part under statutory intellectual property laws. In such cases, use of the Creative Commons No Rights Reserved Instrument (CC0) or the Open Data Commons Public Domain Dedication and License (PDDL), as appropriate, would make the data publicly available for reuse, enhance distribution, and ensure legal interoperability. It is ideal to use standard legal instruments such as these that are valid under the laws of as many jurisdictions as possible, are clear and understandable to the data provider and user, and are machine-readable to enhance searches and linkages to the data affiliated with the license.

Table 1 summarizes the recommended legal mechanisms that governments, organizations, and individual scientists may use to provide open access with no restrictions on reuse to research data and datasets. As shown, the open status of research data from the intellectual property perspective can best be achieved through public law mechanisms such as intergovernmental treaties and agreements, or national legislation, administrative regulations, or policies (including government contract or grant provisions). Private-law equivalents can also be used by the data rights holder in the form of standard widely-used legal instruments incorporating open-access legal provisions to ensure that any instrument used is applicable in as many jurisdictions as possible. Some entities, such as national governments, may have all of the options available to them.

Table 1: Summary of Legal Mechanisms that Promote Open Access and Legal Interoperability

Name of Legal Mechanism	Summary Description
A. Government Mechanisms	
Intergovernmental Treaties and Agreements	Governments can enter into treaties or international agreements (multilateral or bilateral) that create binding obligations among governments to exclude government generated or funded data from copyright protection, place research data arising from joint research publicly in the public domain or place data collected in a particular location (i.e., Antarctica) in the public domain.
Intergovernmental Policies	Intergovernmental organizations can adopt policies to make certain types of publicly funded research data publicly available without restriction on use or reuse.
National Legislation, Regulations, or Policies	National governments can enact laws, issue regulations, or adopt policies to place publicly funded research data in the public domain. Government research funding agencies can include requirements in grants or contracts to make resulting research data publicly available without restrictions on use or reuse.
B. Private Law Equivalent Mechanisms	
Creative Commons Public Domain Mark ^a	The CC Public Domain Mark is used to mark datasets over which copyright has expired, and thus are already in the public domain, enabling their more ready identification in global web searches. Except for data arising in the public domain (e.g. facts) and datasets in historic documents, few datasets should likely have this mark applied. ^b
Creative Commons No Rights Reserved Instrument (CC0) ^c	To the extent possible under law across the world, the person or authority that associates CC0 with the work waives all copyright and related or neighboring rights to the work. ^d
Open Data Commons Public Domain Dedication and License (PDDL) ^e	The PDDL allows the database user to “copy, distribute and use the database”; “produce works from the database”; and “modify, transfer and build upon the database.”
Nongovernmental Agreements and Policies	Nongovernmental organizations ^f can implement policies to retain no rights in research data or conclude agreements among institutional members of a consortium to retain no rights in research data.

^a See <http://creativecommons.org/choose/mark/>.

^b Creative Commons does not recommend use of the Public Domain Mark for works with “limited, hybrid public domain status” such as works of the U.S. Federal Government. U.S. Copyright law states: Copyright protection under this title is not available for any work of the United States Government (17 U.S. Code § 105 - Subject matter of copyright: United States Government works). Such works are not completely in the public domain because it is not entirely clear that the lack of availability of copyright in U.S. federal government works would necessarily apply in some international dispute contexts. However, Creative Commons plans to develop and make available a Public Domain Mark for those works with limited, hybrid public domain status. Use of such marks would make it more difficult for others to restrict access to data placed in the public domain by data producing countries.

^c Online at <http://creativecommons.org/choose/zero/>; explanation at <http://creativecommons.org/about/CC0>. CC0 applies to all content, including data.

^d CC0 is actually a three-tiered instrument consisting first of a comprehensive *waiver*, followed by a fall-back *public license*, and a further fall back *non-assertion of action provision*. See <https://wiki.creativecommons.org/wiki/CC0>.

^e This license is intended to be used with databases and may be used to cover both the database and its data content. It may also be used to cover just the database with the data left to be covered by another license. In this context “data” is defined as the contents of the database, which includes the information, independent works, or other material collected into the database, while “database” is defined as a collection of data arranged in a systematic or methodical way and individually accessible by electronic or other means. See <http://www.opendatacommons.org/licenses/pddl/1-0/>.

^f Nongovernmental organizations include commercial and non-commercial entities.

Currently, researchers often develop datasets for their research by making their own observations and combining that data with extractions drawn from other research datasets that carry underlying restrictions. How are they to know when they have crossed the line of copying too much creativity from one or more other sources, so that their research product(s) might be considered legally a derivative product resulting in substantial copyright breach liability exposure? Merely citing the previous works often may be insufficient under the law. Typically, the researcher cannot acquire reliably the legal clarity desired unless the researcher has drawn the data from sources supporting the legal mechanisms of Table 1.

4.1.1 Governmental Mechanisms

The broadest effects of legal action can be achieved at international levels, whether multilateral or bilateral. Intergovernmental treaties and executive agreements can create binding obligations among governments to revise the default provisions of the Berne Convention and thereby exclude government generated or funded data from copyright protection. They also can carve out exemptions from intellectual property protections for certain other categories of data, such as those produced by a multilateral research program, set of institutions, or discipline. The same types of categories can be the topic of non-governmental international agreements and MoUs. These agreements can also be binding upon the parties, but as private contracts rather than governmental obligations under international law.

Intergovernmental policies that promote public domain or commons approaches to sharing research data can also have significant effect. They presumably signify the official positions of the government members that are signatories of such instruments and are thus indicative of the legal treatment of such digital resources within that member country or organization.

The national governmental level also has great legal effect, although it is limited to the territorial jurisdiction in which the law or policy is enacted. Copyright exemptions, such as in the U.S. Copyright Act of 1976 exempting government works from copyright protection, is an example of such an approach with a pervasive effect within the national jurisdiction.¹⁷⁶ Of lesser legal significance, but still of great value in official guidance, are the national policies, such as executive orders or memoranda, that make public domain or commons approaches to sharing government and government-funded data the default rule. Individual ministries or governmental agencies can also adopt such policies, although these would be narrower in scope and effect.

Government research funding agencies have several options for enhancing access to data arising from their research funding. Research funders can include terms in their funding mechanisms (contracts or grants) that require data management plans,¹⁷⁷ or open access to research data. They can also prescribe a limited amount of time for exclusive use of data prior to its wider dissemination. They can also create public repositories or include provisions in government

¹⁷⁶ U.S. Copyright Law, note 123.

¹⁷⁷ Under the February 22, 2013 OSTP Memorandum, "Increasing Access to the Results of Federally Funded Scientific Research," Federal agencies covered by the policy are required to develop Data Management Plans to increase access to research data resulting from their funding, including obligations for awardees to identify data that will be generated and plans to make the data available. Some of the EU Horizon 2020 Programme projects also require some awardees to prepare data management plans.

funding mechanisms to require data resulting from government funded research to be deposited in an appropriate public repository, such as the U.S. National Institute of Health’s National Center for Biotechnology Information, or make data publicly available within a specified time or embargo period. Finally, they may award funding to cover costs associated with the establishment or maintenance of data depositories.

4.1.2 *Non-government Institutional Mechanisms*

Non-governmental institutions have a major role to play in this regard as well. As a condition of publishing, publishers can require underlying data be openly available with the appropriate legal status.¹⁷⁸ Similarly, research institutions, particularly at universities, and data centers can use private law equivalent mechanisms such as CCO and other internationally recognized standard instruments to make their researchers’ data openly available and not subject to reuse restrictions. They can adopt policies to make the data resulting from research conducted by their researchers publicly available and assist them in negotiating with journals to keep their works publicly available. They also can encourage agreements among researchers to place their data in a public repository, i.e., as a default rule.¹⁷⁹ Institutional credit for open access data sharing will also incentivize such behavior. Training can be provided to educate their researchers about the key laws and concepts governing access and reuse of research data and how to maximize others’ use of their data. Research institutions can review draft commercial sponsorship agreements or clinical trial agreements to ensure that they do not inappropriately restrict the researchers’ rights to disseminate the results of their scientific research.

Stakeholder communities, such as libraries, archives, museums, publishers, e-science infrastructures, or organized members of research disciplines themselves can also seek to harmonize their data practices and policies to achieve legal interoperability.¹⁸⁰ This can include adoption of codes of practice, guidelines, or recommendations. The promotion of community norms for broadly implementing research data citation can largely negate the need for intellectual property and contractual legal sanctions. We therefore encourage funding agencies, universities, scholarly publishers, and other organizations in the research enterprise to include failure to cite research data used from other sources to be a breach of their institutional policies.

For their part, individual researchers can voluntarily adopt CC0 No Rights Reserved legal instruments or PDDL for their datasets. Academic researchers and researcher institutions who are collaborating with commercial sponsors should carefully consider the terms of commercial sponsorship agreements, clinical trial agreements, or similar contracts to ensure, as much as possible, that they do not inappropriately restrict researchers’ rights to disseminate the results of the research the scientists have conducted or to share the data from the research. Any time delays for disclosure of the data should be included in the agreement.

¹⁷⁸ As an illustration of research publisher mandates, all articles submitted to Public Library of Science Journals must adhere to their data availability requirements, <http://journals.plos.org/plosone/s/data-availability>, and materials and software sharing requirements, <http://journals.plos.org/plosone/s/materials-and-software-sharing>.

¹⁷⁹ See text at notes 170 and 171.

¹⁸⁰ For additional recommendations about research community roles and harmonization of approaches to the open access and legal interoperability of data, see, the RDA-CODATA Interest Group, note 34.

Use of language that clearly conveys that the research data can be reused without restriction will further promote legal interoperability. As noted in the RDA-CODATA Legal Interoperability for Research Data Implementation Guidelines, all of the relevant stakeholders should actively participate in the development of laws and policies pertaining to access to and reuse of their research data. The RDA-CODATA Principles and Implementation Guidelines for the Legal Interoperability of Research Data, state that all members of the research community need to be informed about the data laws, policies, and norms pertaining to the data they produce or use, and their representatives should engage directly in the applicable legal, policy, and community process governing research data.¹⁸¹

4.2 Recommendations for Open Access Research Data Repositories

All research data produced by a university or other publicly funded research organization and upon which the scholarly literature produced by the organization is dependent, should be placed in a long-term open access repository at the same time that the scholarly publication is made available.¹⁸² More specifically, we recommend that any open access data repository used in affiliation with such an open access data policy should possess at a minimum the following operational characteristics of a research data commons:

Legal Clarity: All research data contributors as a condition of using the long-term repository for data deposits should be required to adhere to a CC0 No Rights Reserved instrument, agree to an ODC Public Domain Dedication and License, or attest that the material is already within the public domain.

Metadata Creation: All research data contributors are required to provide core metadata (including mandatory minimum core entries, of which one should be legal status) and have the option of being supported in extensive metadata creation through semi-automated information extraction and development methods that enable creation of standards-based, comprehensive cross-science metadata by novice metadata creators.

¹⁸¹ *Id.*, Implementation Guidelines for Principle 2.

¹⁸² Some researchers are loathe to share research data acquired through public funding at the time of publication of their initial research findings because they are afraid they might be “scooped with their own data” by other researchers accomplishing further analysis. Funding agencies have been fighting such attitudes and, as a result, NSF and NIH now require detailed data management plans. By illustration, the current NSF Data Sharing Policy states:

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. Grantees are expected to encourage and facilitate such sharing. (<https://www.nsf.gov/bfa/dias/policy/dmp.jsp>)

NIH expects the timely release and sharing of data to be no later than the acceptance for publication of the main findings from the final dataset. (https://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm#time). Those researchers that do not make data available upon which their findings are based in an effective and expedient manner may be confronted with negative peer reviews resulting in failed current or future funding requests or may be required by agency program directors to make the data expediently and openly available as a condition of receiving funding. If analysis by other scientists of open access data collected through full or partial public research funding results in the more expedient advancement of science, this is to be encouraged rather than discouraged.

Ease of Search: Users from different disciplines are able to readily search for, find and access or download germane research data without the need to become conversant in the specific metadata systems of other disciplines.

Ease of Citation of Data Use: Whenever a user downloads by human action or by an automated method all or a portion of a dataset for potential further use, the registered user receives both a screen indication and an automated direct communication (e.g., email) that indicates the preferred standards-based citation for the data in the event that the material is used in a subsequent analysis, a derivative dataset, or used to prepare a derivative creative work.

Encouragement of Citation Normative Behavior: The communication to the user should also indicate that to use the downloaded or extracted data within other distributed works without providing the appropriate provided citations may be viewed as plagiarism by the participating repository community. While not a copyright infringement, such use may similarly be viewed as plagiarism and against the policies of scholarly publishers, funding agencies, and the users' home institutions.

We recommend that current and future research data repositories supported partially or entirely by public funds pursue the goal of incorporating all of these characteristics. In short, open access policies and the infrastructure enabling optimal use of those policies should cover both scholarly works wholly or partially funded by the public as well as the research data upon which those works are supported.

4.3 Recommendation for a Global Research Data Commons

We also recommend development and deployment of a distributed, but coordinated, *Research Data Commons* supported preferably within the global research library network whereby any graduate student completing a thesis and any researcher or professor publishing a research article could:

- (a) place their research data into a long term archive,
- (b) readily create standards-based metadata for their contribution (with the contributor having little to no knowledge about metadata creation),
- (c) link the data files from their thesis or published article, and
- (d) allow the rest of the world to access and use the data without asking for further permission (but strongly encouraging citation to the data) while accommodating an optional embargo on release of the data to extend no more than a year.

All research data submissions would be submitted under a CC0 No Rights Reserved instrument and research databases under the Open Data Commons Public Domain Dedication and License (PDDL). To best advance science, the funding model should allow published researchers and graduate students from any science domain, any part of the globe, and any institution supporting scholarship in science to be able to upload and make their research data accessible regardless of ability to pay.¹⁸³ Institutional, research subject-domain, and publisher data archives would

¹⁸³ The authors suggest that the initial development of open source software and demonstration projects might be supported best by government research funding agencies or direct legislation while long-term continuing support

continue to be implemented and supported, particularly for big data science efforts, but such a commons focused on data upon which scholarly publications are based and directly tied to that literature should greatly enhance access to and preservation of research data for the future. Universities would be much more willing to require through their policies that their researchers openly deposit research data if a facility supporting the needs of most researchers across most science domains existed.¹⁸⁴ Because of wide-ranging reasons for not making research data available¹⁸⁵, not all scientists would contribute to a Research Data Commons when very easy to do so. However, many would.

All stakeholders discussed in this article can play important roles in promoting the open sharing, unrestricted reuse, and enhanced legal interoperability of research data. If they all share the responsibility in enabling access to and reuse of publicly-funded research data, the societal benefits that arise from such data sharing can be realized more broadly.

might be supported best by the international publicly-funded research library community. One illustrative cross-disciplinary research data repository achieving some of the recommendations in this paragraph is Dryad, <http://datadryad.org/>, which adheres to the Joint Data Archiving Policy, <http://datadryad.org/pages/jdap>.

¹⁸⁴ While the challenges in implementation of such a global research data commons would be substantial, the challenges in pursuit of technological concepts that might negate some of the need for such a commons, such as *linked data* and *semantic web* advances, are perhaps far more daunting. Of 895 experts responding to an invitation to complete a survey from the Pew Research Center, 47% agreed with the statement: “By 2020, the semantic web envisioned by Tim Berners-Lee will not be as fully effective as its creators hoped and average users will not have noticed much of a difference.” (JANNA ANDERSON & LEE PAINE, THE FATE OF THE SEMANTIC WEB 2 (The Pew Research Center 2010). While positive web evolution is certainly occurring (See Amit Sheth, *Semantics Scale Up: Beyond Search in Web 3.0*, IEEE INTERNET COMPUTING 15 (6): 3 (Nov-Dec 2011). <http://doi.ieeecomputersociety.org/10.1109/MIC.2011.157> and B. Glimm and H. Stuckenschmidt, *15 Years of Semantic Web: An Incomplete Survey* KUNTLISCHE INTELLIGENZ 30 (2): 117 (June 2016). <http://link.springer.com/article/10.1007/s13218-016-0424-1>), a global scientific research data commons along the lines suggested could be highly complementary to advances in linked data, ontologies, and semantic web implementations in research.

¹⁸⁵ See Borgman (2015), at note 2, Chapters 1, 8, and 9.

Appendix A

Compendium of International Declarations, Principles, and Policies Promoting Open Access to Data

International/Intergovernmental Organizations

- OECD Principles and Guidelines for Access to Research Data from Public Funding (2007)¹⁸⁶
- OECD Recommendation of the Council for Enhanced Access and More Effective Use of Public Sector Information (2008)¹⁸⁷
- UNESCO Policy Guidelines for the Development and Promotion of Open Access (2012).¹⁸⁸
- WMO Resolution 8.1(2), WMO Policy for International Exchange of Climate Data and Products to Support the Implementation of the Global Framework for Climate Services. (May 2015). The Seventeenth WMO Congress adopted Resolution 8.1(2) to promote sharing of climate data and products among WMO Members and facilitate the implementation of the Global Framework for Climate Services.
- WMO Resolution 25, Exchange of Hydrological Data and Products (1999). Members shall provide on a free and unrestricted basis hydrological data and products necessary for the provision of services in support of the protection of life and property and for the well-being of all peoples.¹⁸⁹
- World Meteorological Organization (WMO)¹⁹⁰ Resolution 40, Policy and Practice for the Exchange of Meteorological and Related Data and Products (1995). This WMO policy includes guidelines on relationships in commercial meteorological activities and provides that WMO members shall provide on a free and unrestricted basis essential meteorological and related data and products necessary for the provision of weather forecasting services.¹⁹¹

Intergovernmental Bodies

- G7 Science and Technology Minister's Tsukuba Communiqué, Tsukubai, Ibaraski, Japan, May 15-17, 2016. The Communiqué states that open science can change the way research and development is undertaken with emerging findings leading to greater global collaboration and encouraging a broader range of participants and stakeholders. The Ministers acknowledge the need for common international principles for open science and to put these principles into practice through open access to scholarly publications and open data. In addition, the Ministers agreed to promote open science and the improvement of the global data sharing infrastructure to ensure the discoverability,

¹⁸⁶ See OECD PRINCIPLES AND GUIDELINES at note 22.

¹⁸⁷ See OECD RECOMMENDATION OF THE COUNCIL at note 22.

¹⁸⁸ See <http://unesdoc.unesco.org/images/0021/002158/215863e.pdf>.

¹⁸⁹ *Id.*

¹⁹⁰ The United Nation's World Meteorological Organization has issued three resolutions to promote sharing of data.

¹⁹¹ See https://www.wmo.int/pages/about/Resolution40_en.html.

accessibility, and interoperability of a wide range of ocean and marine data. Ministers agreed to promote open science and sharing of the results of publicly funded research e.g. data and publications in the field of brain science linked to age-related problems, neglected tropical diseases and poverty-related infectious diseases. They also agreed to promote open science and the improvement of the global data sharing infrastructure to ensure the discoverability, accessibility, and interoperability of a wide range of ocean and marine data. They supported establishment of a working group on open science with the aims of sharing open science policies, exploring supportive incentive structures, and identifying good practices for promoting increasing access to the results of publicly funded research, including scientific data and publications, coordinating as appropriate with the OECD, RDA, and other relevant groups.

- G-8 Open Data Charter, June 2013.¹⁹² This open data charter sets out five strategic principles to promote transparency, innovation, and accountability in government data, including open data by default, quality and quantity, useable by all, releasing data for improved governance, and releasing data for innovation.
- GEOSS Data Sharing Principles and White Paper: GEO adopted data sharing principles as a foundation for the Group on Earth Observations System of Systems (GEOSS). The 2005 principles provide for the full and open exchange of data, metadata and products shared in within GEOSS, recognizing relevant international instruments and national policies and legislation. All shared data, metadata and products should be free of charge or no more than cost of reproduction will be encouraged for research and education.¹⁹³
- Global Biodiversity Information Facility (GBIF), Data Sharing Agreement. The goals and principles of making biodiversity data openly available to the greatest extent possible are delineated in the GBIF Memorandum of Understanding (MoU) and Data User Agreement. The signatories of the MoU agree to make biodiversity data available through their nodes to foster scientific research development internationally and to support the public use of these data. While open access and reuse is the default, GBIF seeks to ensure that the source of data is acknowledged and should request that such attribution be maintained in any subsequent use of the data.¹⁹⁴

International Non-Governmental Organizations (NGOs)

- Committee on Data for Science and Technology (CODATA) and Research Data Alliance (RDA)'s Interest Group on Legal Interoperability of Research Data is currently developing principles and guidelines to promote legal interoperability of research data. This document sets forth six key principles to promote legal interoperability of data:
 - One:** Facilitate the lawful access to and reuse of research data.
 - Two:** Determine the rights to and responsibilities for the data.
 - Three:** Balance the legal interests.

¹⁹² See <https://www.gov.uk/government/publications/open-data-charter>. The G-8 is a group of nations that meet to advance political and economic objectives, but is not itself an international organization.

¹⁹³ See http://www.earthobservations.org/geoss_dsp.shtml and GEOSS DATA-CORE WHITE PAPER 2014 in note 24.

¹⁹⁴ See <http://www.gbif.org/terms/licences/data-sharing>.

Four: State the rights transparently and clearly.

Five: Promote the harmonization of rights in research data.

Six: Provide proper attribution to the data originator(s) and rights holder(s).¹⁹⁵

- Gates Foundation Policy on Open Access Research, November 2014. The Foundation's open access policy requires publications resulting from its funding to be published on open access terms to enable users of the publication to copy and reuse the material and build upon it without further permission or fees.¹⁹⁶
- International Council for Science (ICSU) is a non-governmental organization with members from national scientific bodies and international scientific unions representing 143 countries. ICSU was founded in 1931 to promote international scientific activity. ICSU published a Statement on Open Access to Data in 2013.¹⁹⁷
- Science International Accord on Open Data for a Big Data World, December 2015. [fill in]
- Sunlight Foundation 2010: Ten Principles for Opening up Government Data (August 11, 2010) Sunlight Foundation identified ten principles for government datasets to make them publicly available, including placement in the public domain and use of commonly owned standards.¹⁹⁸

International, interdisciplinary statements on open access to research publications and databases arising from the research community¹⁹⁹

- The Australian Open Access Support Group aims to achieve greater access to Australian scholarship and research. The underlying principle of the Group is support for free and open access to research outputs to promote discoverability. The Group's website lists supporting statements, mandates, and policies around the world that promote open access at government and large funder level since mid-2012.²⁰⁰
- The Budapest Open Access Initiative (2002). The statement is one of the first articulations for broadly promoting open access to publicly funded peer-reviewed research.²⁰¹
- The Bethesda Statement on Open Access Publishing (2003). This statement of principles, drafted at a meeting held at the Howard Hughes Medical Institute in Maryland, promotes open access models to the primary biomedical scientific literature.²⁰²
- The Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (2003), whose goal is to promote the internet as an instrument for distributing global scientific knowledge and human reflection and to specify measures which research policy makers, research institutions, funding agencies, libraries, archives, and museums need to

¹⁹⁵ CODATA is an interdisciplinary Scientific Committee of the International Council for Science (ICSU), which works to improve the quality, reliability, management and accessibility of scientific data. Its members include scientific academies, scientific unions, and international organizations.

¹⁹⁶ See <http://www.gatesfoundation.org/How-We-Work/General-Information/Open-Access-Policy>.

¹⁹⁷ See <http://www.icsu.org/general-assembly/news/ICSU%20Report%20on%20Open%20Access.pdf>.

¹⁹⁸ See <http://sunlightfoundation.com/policy/documents/ten-open-data-principles/>.

¹⁹⁹ These statements have been adopted by ad hoc research communities rather than established organizations.

²⁰⁰ See <http://aoasg.org.au/statements-on-oa-in-australia-the-world/>.

²⁰¹ See <http://budapestopenaccessinitiative.org/>.

²⁰² See <http://legacy.earlham.edu/~peters/fos/bethesda.htm>.

consider to achieve the goal of open global access.²⁰³ They hold annual conferences on this topic in different places around the world.²⁰⁴

- The Hague Declaration (2014). The declaration emphasizes the importance of content mining from big data to harness knowledge, address global challenges, and revolutionize research. The declaration endorses the following principles: (1) intellectual property laws should not be designed to regulate the flow of facts, data, and ideas; 2) licenses and contract terms should not restrict individuals from using facts, data, and ideas; and 3) ethics around the use of content mining techniques will need to continue to evolve in response to changing technology. The declaration also encourages universities, research organizations, library associations, medical communities, and businesses to advocate the benefits of content mining.²⁰⁵
- The Ghent Declaration (2011) is an initiative of the reviewers of the European Commission's OpenAIRE project, OpenAIRE was funded by the European Commission's Seventh Framework Programme to implement the EC Open Access policy by establishing an infrastructure for EC-funded researchers to make their published work open access. The Ghent Declaration encourages the EC to move from open access to research and scholarship towards the creation and use of open data, open source software, and open educational resources.²⁰⁶
- RECODE Policy Recommendations for Open Access to Research Data in Europe Project (RECODE) (2015). This EU-funded research project leveraged existing networks, communities and projects to address challenges to open. Legal issues focused on intellectual property rights, privacy and data protection, and open access mandates. RECODE produced policy recommendations for research funders, institutions, data managers and publishers. Recommendations include requiring the use of harmonized open licensing frameworks, and systematically addressing legal and ethical issues arising from open access to research data. RECODE recommends the extensive use of open licensing and implementing technical solutions for legal and ethical issues, systematically turning institutional attention to developing solutions for legal and ethical problems arising from open access to research data, including internal review processes.²⁰⁷
- SPARC Europe's Statement on Open Access, 2011. The statement promotes access to scholarly publications and data.²⁰⁸

Statements focusing on open access in the context of developing countries

- Committee on Data for Science and Technology (CODATA) Data Sharing Principles in Developing Countries (Nairobi Principles, 2014). These data sharing principles were endorsed at the International Workshop on Data for Sciences and Sustainable

²⁰³ See <http://openaccess.mpg.de/Berlin-Declaration>.

²⁰⁴ *Id.*

²⁰⁵ See <http://thehaguedeclaration.com/the-hague-declaration-on-knowledge-discovery-in-the-digital-age/>

²⁰⁶ See <https://www.openaire.eu/did-you-know/highlights/seizing-the-opportunity-for-open-access-to-european-research-ghent-declaration-published>.

²⁰⁷ See <http://recodeproject.eu/>;

http://recodeproject.eu/wpcontent/uploads/2015/01/recode_guideline_en_web_version_full_FINAL.pdf.

²⁰⁸ See http://sparceurope.org/wp-content/uploads/2011/07/SPARC-Europe-Statement-on-Open-Access-Luxembourg-30-May-2011_FINAL1.pdf.

Development held in August 2014 at UNESCO in Nairobi, Kenya. They promote open and unrestricted data, discoverability, interoperability, equitable data access, and attribution. Data may be restricted, in exceptional cases, if adequately justified.²⁰⁹

- ICSU Committee on Freedom and Responsibility in the Conduct of Science (CFRS), Advisory Note on “Sharing Scientific Data with a Focus on Developing Countries. This Advisory Note is concerned with the barriers and challenges to international data access and sharing, with a special focus on developing countries. It cites Article 27 of the Universal Declaration of Human Rights, which affirms that: “Everyone has the right to (...) share in scientific advancement and its benefits,” in support of the Principle of the Universality of Science and the promotion of equitable access to scientific data. It provides that governments may restrict access in exceptional circumstances for reasons of security, privacy, or legal provisions. Such restrictions must be kept to the minimum necessary and justified explicitly. Data should consistently be shared openly unless the potential for harm to society is greater than the anticipated benefits.²¹⁰
- Salvador Declaration on Open Access (The Developing World Perspective, 2005). The Latin American and Caribbean Center on Health Sciences Information / Pan American Health Organization / World Health Organization issued a statement supporting unrestricted access to, and use of scientific information. Recommendations include requiring that publicly funded research be made available through open access (OA); considering the cost of publication as part of the cost of research; strengthening the local OA journals, repositories, and other relevant initiatives; and promoting integration of developing countries’ scientific information in the worldwide body of knowledge.²¹¹

Biological and biodiversity data

- The Bermuda Principles (1996). At a 1996 summit in Bermuda, leaders of the Human Genome Project agreed that all human genomic sequence information generated by centers funded for large-scale human sequencing should be made freely available and in the public domain within 24 hours after generation. The “Bermuda Principles” reshaped the practice of waiting until publication to release data and established a new norm in genetics and genomics to the rapid release of sequence assemblies by large-scale sequencing efforts.²¹²
- Sharing Data from Large-scale Biological Research Projects (Ft. Lauderdale Principles, 2003). The Wellcome Trust convened a meeting to discuss. Participants reaffirmed the Bermuda Principles for large-genome sequencing projects and defined responsibilities for researchers, publishers, and research funders.²¹³
- Prepublication of Data Sharing (the Toronto Statement, 2009). At a Toronto International Data Release Workshop, participants agreed to extend the practice established by the Bermuda principles of rapid release of prepublication data to other biological data sets

²⁰⁹ <http://www.codata.org/task-groups/preservation-of-and-access-to-scientific-and-technical-data-in-for-with-developing-countries-pastd/pastd-past-achievements>.

²¹⁰ See <http://www.icsu.org/publications/cfrs-statements/data-sharing>.

²¹¹ See <http://www.icml9.org/meetings/openaccess/public/documents/declaration.htm>.

²¹² See http://www.casimir.org.uk/storyfiles/64.0.summary_of_bermuda_principles.pdf.

²¹³ See <http://www.wellcome.ac.uk/About-us/Publications/Reports/Biomedical-science/WTD003208.htm>; <https://www.genome.gov/Pages/Research/WellcomeReport0303.pdf>.

that have broad utility. The statement establishes best practices for data producers, publishers, data funders, and data users.²¹⁴

- The Declaration on Open Biodiversity Knowledge Management (Bouchout Declaration, 2014). The Bouchout Declaration's goal is to establish an Open Biodiversity Knowledge Management System. Its principles include the open and free use of digital resources about biodiversity, and licenses or waivers that grant or allow all users a free, irrevocable, world-wide, right to copy, use, distribute, transmit, and display the work publicly as well as to build on the work and to make derivative works, subject to proper attribution consistent with community practices, while recognizing that providers may develop commercial products with more restrictive licensing.²¹⁵

Agricultural research data

- GODAN Statement of Purpose (2013) The GODAN initiative was announced at the Open Government Partnership Conference in October 2013. GODAN supports the sharing of open data to make information about agriculture and nutrition available, accessible, and usable. It advocates for open data and open access policies by default, in both public and private sectors, while respecting and working to balance openness with legitimate concerns in relation to privacy, security, community rights, and commercial interests.²¹⁶

²¹⁴ See http://www.nature.com/nature/journal/v461/n7261/box/461168a_BX1.html.

²¹⁵ See <http://www.bouchoutdeclaration.org/declaration/>.

²¹⁶ See <http://www.godan.info/about/statement-of-purpose/>.

Appendix B

Selected Government Instruments and Policies Promoting Open Access to Data in the United States and European Union

United States

- OMB Circular A-130, The White House Office of Management and Budget (OMB) Memorandum on Management of Federal Information Resources, November 28, 2000. “Government information is a valuable national resource. As a public good, it furthers transparency in government, democracy, and fuels economic benefits.” “The open and efficient exchange of scientific and technical government information, subject to applicable national security controls and the proprietary rights of others, fosters excellence in scientific research and effective use of Federal research and development funds.”²¹⁷
- Executive Order 13642, Making Open and Machine Reusable the New Default for Government Information, May 9, 2013. This Executive Order makes machine-readable and open the new default for government information.²¹⁸
- OMB Memorandum to Federal Agencies, “Open Data Policy-Management Information as an Asset” (M-13-13, May 9, 2013). This OMB Memorandum establishes a framework to institutionalize the principles of effective information management to promote interoperability and openness.²¹⁹
- White House Office of Science and Technology Policy’s Memo” Increasing Access to the Results of Federally Funded Scientific Research (February 22, 2013). Federal agencies with over \$100 million in the annual conduct of research and development expenditures are required to develop plans and policies to support increased public access to the unclassified results of research funded by the Federal Government with such results including peer-reviewed publications, metadata, and data.²²⁰
- Executive Office of the President, U.S. Open Data Action Plan, May 9, 2014. The plan outlines plans to enhance and release data assets set forth in Open Data Charter.²²¹
- White House Office of Science and Technology’s Bromley Principles Regarding Full and Open Access to Global Change Data (1991) set forth data management policies that promoted the sharing of global change research.

European Union

- EU Press Release: All European Scientific Articles to be Freely Accessible by 2020, May 27, 2016. EU Ministers for Research and Innovation announced that scientific publications reporting on the results of research supported by public funds must be freely

²¹⁷ See https://www.whitehouse.gov/omb/circulars_a130_a130trans4/.

²¹⁸ See <https://www.whitehouse.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government->.

²¹⁹ See <https://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf>,

²²⁰ See https://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf.

²²¹ See https://www.whitehouse.gov/sites/default/files/microsites/ostp/us_open_data_action_plan.pdf.

available by 2020. The data underlying the research publication should be made accessible, unless there are well-founded reasons for not doing so, for example intellectual property rights or security or privacy issues.

- Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Region, “Open data: An engine for innovation, growth and transparent governance” (12.12.2011). This Communication adopts a legal Framework for reuse of public data, among other things.²²²
- Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020, December 11, 2013. The European Commission’s guidelines provide a context and explanation for the open access rules for scientific data and publications arising from funding under Horizon 2020, the EU Framework Programme for research and innovation for 2014-2020. The European Commission's vision is that information already paid for by the public should not be paid for again each time it is accessed or used, and that it should benefit European companies and citizens to the full. Provisions for open access to the data and publications are contained in the Model Grant Agreements. The default is open access but grantees may restrict access to data to comply with data protection requirements, security rules, or protect intellectual property rights.²²³
- EU Regulation on Copernicus Data, 2014. The European Commission adopted an open data policy for remote sensing data from Copernicus, the EU’s flagship Earth observation programme.²²⁴

²²² See http://ehron.jrc.ec.europa.eu/sites/ehron/files/documents/public/2011.ec_communication_open_data_-_engine_for_innovation_growth_and_transparent_governance.pdf.

²²³ See http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf.

²²⁴ See http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2014.122.01.0044.01.ENG.