



LEGAL INTEROPERABILITY OF RESEARCH DATA: PRINCIPLES AND IMPLEMENTATION GUIDELINES

RDA-CODATA Legal Interoperability Interest Group

Version 1.0 | 20 October 2016



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

Legal Interoperability of Research Data: Principles and Implementation Guidelines

Acknowledgements

These Principles and Implementation Guidelines are the product of the entire RDA-CODATA Interest Group. The co-chairs of this Interest Group are Paul F. Uhler, Enrique Alonso Garcia, and Robert Chen. The full list of the Interest Group's members may be found at: <https://rd-alliance.org/groups/rdacodata-legal-interoperability-ig.html>. However, the following individuals were the principal authors of this document and they participated in frequent conference calls and in the RDA Plenaries to develop this between March 2015 and August 2016 (listed in alphabetical order): Donat Agosti, Enrique Alonso Garcia, Baden Appleyard, Christoph Bruch, Robert Chen, Gail Clement, Willi Egloff, Herbert Gruttemeier, Simon Hodson, Maria Lloset, J. Bernard Minster, and Paul F. Uhler.

This document went through two rounds of external reviews. The first round was conducted in December 2015-February 2016 by the RDA and CODATA, and those reviewers performed their services on the condition of anonymity. A second round of reviews was requested by the RDA-CODATA Interest Group itself in the spring of 2016, and the responses to those reviews led to this publication. The Interest Group is grateful for the review comments provided by all the anonymous reviewers and by the following experts who expressly consented to be named (listed in alphabetical order): George Alter, Amir Aryani, Adrian Burton, Jiang Dong, Giulia Dore, Elise Dunham, Katie Fortney, George Haringhuizen, Varsha K. Khodiyar, Michael Linksvayer, Wouter Los, Jesus Marco de Lucas, Thomas Margoni, Hollydawn Murray, Harlan Onsrud, Kevin Pomfret, Rina E. Pantalony, Fraser Taylor, Heiko Tjalsma, Kathryn Unsworth, Luis Villa, and J. Max Wilkinson. These experts provided their reviews in their personal capacities, rather than as representatives of their institutions of employment, and the views expressed here are solely those of the RDA-CODATA Interest Group.

Finally, we would like to thank the Research Data Alliance and CODATA for their support throughout the process of drafting this document, and in particular Simon Hodson, the Executive Director of CODATA, and Peter Fox, the Interest Group's liaison from the RDA Technical Advisory Panel, for their personal engagement.

Editorial Responsibility

Paul Uhler, Chief Editor

Gail Clement, Production Editor

Disclaimer

It is important to note as well that these Principles and their Implementation Guidelines are not comprehensive or "binding," nor should they be taken as legal advice. In developing the sound management of research data sharing, access, and reuse, there are other norms and perspectives --ethical, scientific, technical, cultural, and others--beyond just purely legal ones to consider.

TABLE OF CONTENTS

PART I. PRINCIPLES	1
EXECUTIVE SUMMARY OF THE IMPLEMENTATION GUIDELINES	3
PART II. IMPLEMENTATION GUIDELINES	7
Implementation Guidelines for Principle One	7
Implementation Guidelines for Principle Two	13
Implementation Guidelines for Principle Three	16
Implementation Guidelines for Principle Four	19
Implementation Guidelines for Principle Five	22
Implementation Guidelines for Principle Six	26
References	27
PART III: DEFINITION OF TERMS	34

PART I.

INTRODUCTION TO THE PRINCIPLES ON THE LEGAL INTEROPERABILITY OF RESEARCH DATA

The ability of the research community to share, access, and reuse data, as well as to integrate data from diverse sources for research, education, and other purposes requires effective technical, syntactic, semantic, and legal interoperability rules and practices. These Principles focus on legal interoperability because there tends to be misunderstanding and lack of knowledge and guidance about legal issues concerning research data generally.

Legal interoperability occurs among multiple datasets when¹:

- the legal use conditions are clearly and readily determinable for each of the datasets, typically through automated means;
- the legal use conditions imposed on each dataset allow creation and use of combined or derivative products; and
- users may legally access and use each dataset without seeking authorization from data rights holders on a case-by-case basis, assuming that the accumulated conditions of use for each and all of the datasets are met.

When data are combined from multiple sources, the resulting dataset will incorporate the accumulated restrictions imposed by each source. 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 restrictions².

The scope of this document is limited to laws and policies pertaining to access and reuse of data, either produced by researchers themselves or used in research and other activities. It is specifically focused on the ability to integrate those data for research purposes in the context of intellectual property law. Because the focus is on facilitating the open sharing, access, and reuse of data, these *Legal Interoperability Principles* are oriented primarily to data that are either produced in, or with funding by, the public sector. Many of the tensions, however, arise at the interface of different disciplines, institutions, sectors, or countries. We do not cover the protection of research data for human privacy, national security, public safety, liability and personal injury, or other unrelated legal matters, except in the context of balancing them with public access and reuse interests.

These *Legal Interoperability Principles* are offered as high-level guidance to all members of the research community—the funders, managers of data centers, librarians, archivists, publishers, policymakers, university administrators, individual researchers, and their legal counsel—who are engaged in activities that involve the access to and reuse of research data from diverse sources.³ The *Principles* are synergistic, so their greatest benefit is realized when they are considered together.

¹ This definition was derived from the pending publication by Doldirina, Catherine, Eisenstadt, Anita, Onsrud, Harlan, and Uhler, Paul. F., “Legal Approaches for Open Access to Research Data”.

² *Ibid.*

³ It is important to note as well that these Principles and their Implementation Guidelines are not comprehensive or “binding,” nor should they be taken as legal advice. In developing the sound management of research data sharing, access, and reuse, there are other norms and perspectives –ethical, scientific, technical, cultural, and others—beyond just purely legal ones to consider.

The *Principles* were developed by the RDA-CODATA Interest Group (IG) on the Legal Interoperability of Research Data, and they were subsequently reviewed in two rounds by independent experts and also approved by the RDA and CODATA. A set of Implementation Guidelines, also reviewed and approved by the RDA and CODATA, provides more detailed information that members of the research community may find helpful in implementing each Principle and in understanding different approaches in various countries. The Guidelines are presented below.

In drafting these Legal Interoperability Principles and related Implementation Guidelines, the IG was mindful that other organizations and groups have addressed similar issues pertaining to research data. Other relevant Declarations, Statements, and Principles, as well as additional key sources, are provided in the References section at the end of Part II: The Implementation Guidelines. However, almost all of these documents have been focused on promoting the value of “openness”, rather than on the legal conditions required for the interoperability of data.

The following *Principles on the Legal Interoperability of Research Data* focus on all types of data that are used primarily in publicly funded research in government and academia.

Principles on the Legal Interoperability of Research 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.
- Four. State the rights transparently and clearly.
- Five. Promote the harmonization of rights in research data.
- Six. Provide proper attribution and credit for research data.

EXECUTIVE SUMMARY OF THE IMPLEMENTATION GUIDELINES

Principle One: Facilitate the lawful access to and reuse of research data

Definition and Discussion of Terms

| [Copyright](#) | [Collection of data](#) | [Data](#) | [Database](#) | [Equity](#) | [Interoperability of data](#) |
| [Legal interoperability of data](#) | [Research data](#) | [Public domain](#) | [Open data](#) | [Rights holder](#) |
| [Sui generis database rights](#) |

Guideline 1A. Access to and reuse of research data should be open and unrestricted as a default rule, or otherwise be granted to users with the fewest limitations possible.

Guideline 1B. Governments, institutions, or researchers can apply one of an array of legal instruments to place collections of research data in the public domain, with no restrictions on reuse.

Table 1 on the following page summarizes what governments can do from the top down, and what individual researchers and their institutions—whether nongovernmental or governmental—can do from the bottom up, to promote open data free from reuse restrictions, and thus ensure their legal interoperability.

Guideline 1C. The CC0 or PDDL waivers of rights are the preferred voluntary, nongovernmental approaches to facilitate the legal interoperability of research data. The non-restrictive (“common-use, attribution-only, CC-BY 4.0”) license may also be considered when disseminating the data for broad use.

Guideline 1D. In asserting any rights and in applying access and reuse terms and conditions to any research data made publicly available, all members of the research community should make such data available equitably to all users, including the most disadvantaged ones.

Principle Two: Determine the rights to and responsibilities for the data.

Guideline 2A. Research data disseminators need to establish who or what entity has the rights to any given collection of data before the data are disseminated to others.

Guideline 2B. It is also the responsibility of research data users to abide by the rights applicable to the collection of research data, as well as the specific user rights in the jurisdiction that the data are being used.

Guideline 2C. Representatives of research communities who are also experts in policy and legal issues are encouraged to participate in fora that develop and implement laws and other norms governing access to, and the reuse and legal interoperability of, research data.

Guideline 2D. A well-conceived educational process for researchers regarding rights and responsibilities in research data should be developed and adopted by relevant institutions.

Table 1: Summary of Legal Mechanisms that Promote Open Data and Legal Interoperability⁴

Type of Legal Mechanism	Summary Description
Governmental Mechanisms	
Intergovernmental 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 intellectual property rights protection, and should be made publicly available if no other restrictions apply.
Intergovernmental Policies	International or 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 at the ministerial level, or adopt a broad range of policies to place publicly funded research data in the public domain. Government research funding agencies and government data centers or services can include requirements in grants or contracts to make resulting research data publicly available without restrictions on use or reuse.
Nongovernmental and Governmental Mechanisms	
Creative Commons Public Domain Mark (CC PDM, http://creativecommons.org/choose/mark/)	The CC Public Domain Mark is used to mark collections of data over which any potential copyright has expired or never existed, 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 data in historic documents that have been digitized, few collections of data should likely have this mark applied.
Creative Commons No Rights Reserved Instrument (CC0, http://creativecommons.org/choose/zero/)	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, such as the E.U. <i>sui generis</i> database right. It is has been declared compatible with the Open Definition by the Open Definition Advisory Council http://opendefinition.org/licenses/
Open Data Commons Public Domain Dedication and License (PDDL, http://www.opendatacommons.org/licenses/pddl/1-0/)	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.” It is has been declared compatible with the Open Definition by the Open Definition Advisory Council http://opendefinition.org/licenses/
Creative Commons Attribution 4.0 License (CC-BY 4.0, https://creativecommons.org/licenses/by/4.0/legalcode)	The CC-BY 4.0, is the least restrictive of the Creative Commons licenses. It permits the user to share and adapt material with minimal attribution requirements. It is has been declared compatible with the Open Definition by the Open Definition Advisory Council http://opendefinition.org/licenses/
Nongovernmental Policies and Agreements	Nongovernmental organizations 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.

⁴ This table was derived from the pending publication by Doldirina, Catherine, Eisenstadt, Anita, Onsrud, Harlan, and Uhler, Paul F., “Legal Approaches for Open Access to Research Data.”

Principle Three: Balance the legal interests.

Definition and Discussion of Terms

| [Confidentiality laws and policies](#) | [Copyright](#) | [Copyright infringement](#) | [Cultural resources protection](#) | [Endangered species protection](#) | [Intellectual property rights](#) | [National security or public safety laws](#) | [Patents](#) | [Periods of exclusive use](#) | [Personal privacy protection](#) | [Sovereign genetic resources protection](#) | [Sui generis database rights](#) | [Traditional Knowledge protection](#) |

Guideline 3A. As a default rule, intellectual property rights in government or publicly-funded collections of research data should be waived as far as possible and distributed in the least restrictive manner.

Guideline 3B. Governments and public research institutions need to justify any restrictions on research data that may be imposed over and above any restrictions on the access and reuse of data that are defined by existing legislation.

Guideline 3C. Policymakers should consider public interests in developing rules for access to and use of publicly generated research data.

Guideline 3D. Public research funding organizations and the rights holders of public research data sources should reduce time embargoes for exclusive personal periods of research use to the minimum necessary.

Guideline 3E. As a default rule, all rights holders of research data that are partly or fully funded by the public sector need to avoid the use of individual contracts or agreements that restrict access to and reuse of the data.

Principle Four: State the rights transparently and clearly.

Definition and Discussion of Terms

| [Certainty](#) | [Metadata](#) | [Research data](#) | [Rights holder](#) | [Rights statement](#) | [Transparency](#) |

Guideline 4A. The use of standardized electronic statements regarding the legal rights retained (if any) by the rights holders and providers of research data can greatly assist in their comprehensibility by a wide audience--including by machines.

Guideline 4B. The rights holder(s) of any given data used in research should engage competent legal counsel when it is necessary to determine the applicable law(s) and to clarify the differences among jurisdictions.

Guideline 4C. Rights holders should inform users about any special terms and conditions of use.

Principle Five: Promote the harmonization of rights in research data.

Definition and Discussion of Terms

| [Norms](#) |

Guideline 5A. Research policymakers and practitioners can use both top-down and bottom-up approaches, and mixes of both, to harmonize rights concerning research data.

Guideline 5B. Bottom-up actions based on voluntary, private law mechanisms, can be an effective approach for research individuals or institutions to take in the absence of government action, but are relatively fragmented and less harmonized.

Guideline 5C. Top-down harmonization through “hard” law, such as multilateral treaties or executive agreements, or national legislation or administrative regulation, can work in some contexts and can be extremely useful as a broad harmonization tool.

Guideline 5D. Process-based approaches such as workflows decision-making charts, decision making apps and tools, or scoreboards might be considered useful tools to promote harmonization.

Principle Six: Provide proper attribution and credit for research data.

Definition and Discussion of Terms

| [Acknowledgement](#) | [Attribution](#) | [Citation](#) | [Credit](#) | [Plagiarism](#) |

Guideline 6A. Attribution of research data used in any scholarly output should be a normative convention established by good research policy and practice, and preferably not by a legal mandate or a license requirement.

PART II.

IMPLEMENTATION GUIDELINES FOR THE PRINCIPLES ON THE LEGAL INTEROPERABILITY OF RESEARCH DATA

Introduction

These Implementation Guidelines for the Principles on the Legal Interoperability of Research Data (Implementation Guidelines) provide more detailed information that members of the research community may find helpful in implementing each Principle. They are expected to be useful for practitioners in achieving greater legal interoperability of research data, but they are not a substitute for legal advice.

The following Implementation Guidelines focus on all types of data that are used primarily in publicly funded research in government and academia. They are organized according to the six Principles and contain sections on “Definitions and Discussion of Terms” and a summary of the “Guidelines for Implementation”. The authoring RDA-CODATA Interest Group also has provided cross-references to other germane Principles and to key supporting documents in the References.

Principle One: Facilitate the lawful access to and reuse of research data

Definition and Discussion of Terms

| [Copyright](#) | [Collection of data](#) | [Data](#) | [Database](#) | [Equity](#) | [Interoperability of data](#) |
| [Legal interoperability of data](#) | [Research data](#) | [Public domain](#) | [Open data](#) | [Rights holder](#) |
| [Sui generis database rights](#) |

Guidelines for Implementation

In an era of big data and data science, research data have become a first class scholarly object. Much of the academic value of these research resources can be realized through their broad dissemination and reuse, particularly on digital networks (Royal Society, 2012; Hey, Tansley, and Tolle, 2009; Uhlir and Schröder 2007; Benkler, 2006; National Research Council, 2004; National Research Council, 2003).

There are numerous justifications for this beyond research, as well, and there is a growing body of literature that supports such benefits (Uhlir, 2015). It includes many studies that show enhanced economic returns, the promotion of societal welfare goals, the improvement of educational outcomes, and the support of good governance practices. Research data whose production was at least partly publicly funded have public good characteristics, and are often global public goods (Stiglitz, 1999).

Moreover, there are numerous statements, declarations, and principles have been issued over the past two decades in support of open data and unrestricted reuse of data by various research organizations and disciplines, including: the broader research community (Science International, 2015; Tsoulaka, et al., 2015; LIBER, 2014; CODATA PASTD 2014; Denton Declaration, 2012; Bethesda Statement, 2003; Berlin Declaration, 2003; Budapest Open Access Initiative, 2002); international governmental research-related organizations (Data Sharing Working Group - GEOSS, 2014; Cabinet Office UK, 2013; OECD 2007) and

many national governments and their agencies, too numerous to cite here; and representatives of specific disciplines such as genomics (e.g., Human Genome Organization, 1996), systematic biology (Bouchout Declaration, 2014), medical research information in developing countries, and global change environmental data (Bromley, 1991). Taken together, these and other sources form a growing consensus of experts and evidence of customary practice of states and practitioners in publicly funded research.

At the same time, research evaluation in many cases still reward competition more than cooperation. While there may be a broad consensus to make data underlying publications available together with the publication, many researchers are hesitant when asked to make such data available that, from their point of view, may still hold a competitive advantage for them.

Indeed, the rights holder(s) of many research datasets may have legitimate reasons to restrict access and reuse (see Principle Three and its Implementation Guidelines). In the context of these Guidelines, the term “rights holder” refers to anyone who holds the copyright or database rights in the collection of data, assuming that the data within it are eligible for such rights (see Implementation Guideline 1A). There also may be more immediate economic concerns for restricting free dissemination of research data or datasets that are situation dependent (Ember and Hanisch, 2013). Moreover, there are many reasons—both legitimate and not—why researchers themselves do not wish to freely share the data generated in their research (Borgman, 2014). These Principles and their Implementation Guidelines take note of the heterogeneity of research data and sources, their management practices, and business models, but focus only on the legal conditions for best achieving their legal interoperability.

In order to enable the maximum degree of interoperability, and to promote access to and reuse of research data (including data sharing), the following guidelines for implementation can be used by an original producer or subsequent rights holder of collections of research data.

Guideline 1A. Access to and reuse of research data should be open and unrestricted as a default rule, or otherwise be granted to users with the fewest limitations possible.

At the outset, it is essential to characterize the legal differences among databases, collection of data (or datasets), and data: each of these terms are defined in the Definitions section (Part III at the end of the document).

Intellectual property rights, such as copyright and the *sui generis* database right (mostly in the European Union), provide exclusive property protection to protectable subject matter (see Principle Three and its Implementation Guidelines, as well as the Definitions of Terms section). These laws apply automatically by statute in the jurisdictions in which they are enacted.

In those cases in which the rights holder(s) of any given data want them to be broadly disseminated and used, the ability of researchers to access and reuse the data is impaired when there is an absence of clarity about the restrictions or legal conditions under which the data may be reused. Legal restrictions on reuse limit the reproduction and distribution of public research data.

Restrictions can inhibit reuse to a greater extent than is sometimes realized. This can be illustrated by analogy to the idea of the “lowest common denominator.” The lowest common denominator means that for a derivative dataset that is the result of the combination of parts of two or more other datasets, the most restrictive terms and conditions of any dataset used will be transferred to the entire derivative dataset (DSWG, 2014). In this way, the legal restrictions, perhaps unnecessarily imposed, can have broader, unwanted effects limiting the reuse of derived datasets in which most of the components may

otherwise be in the public domain. It should be noted, however, that it is possible to separate data files, even when merging two or more datasets.

Reuse of data (or other types of information) that are in the public domain are not restricted by intellectual property rights.

Guideline 1B. Governments, institutions, or researchers can apply one of an array of legal instruments to place collections of research data in the public domain, with no restrictions on reuse.

Table 1 summarizes what governments can do from the top down, and what individual researchers and their institutions—whether nongovernmental or governmental—can do from the bottom up, to promote open data free from reuse restrictions, and thus their legal interoperability.

Table 1: Summary of Legal Mechanisms that Promote Open Data and Legal Interoperability⁵

Type of Legal Mechanism	Summary Description
Governmental Mechanisms	
Intergovernmental 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 intellectual property rights protection, and should be made publicly available if no other restrictions apply.
Intergovernmental Policies	International or 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 at the ministerial level, or adopt a broad range of policies to place publicly funded research data in the public domain. Government research funding agencies and government data centers or services can include requirements in grants or contracts to make resulting research data publicly available without restrictions on use or reuse.
Nongovernmental and Governmental Mechanisms	
Creative Commons Public Domain Mark (CC PDM, http://creativecommons.org/choose/mark/)	The CC Public Domain Mark is used to mark collections of data over which any potential copyright has expired or never existed, 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 data in historic documents that have been digitized, few collections of data should likely have this mark applied.

⁵ This table was derived from the pending publication by Doldirina, Catherine, Eisenstadt, Anita, Onsrud, Harlan, and Uhlir, Paul F., "Legal Approaches for Open Access to Research Data."

<p>Creative Commons No Rights Reserved Instrument (CC0, http://creativecommons.org/licenses/by/4.0/legalcode)</p>	<p>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, such as the E.U. <i>sui generis</i> database right. It is has been declared compatible with the Open Definition by the Open Definition Advisory Council http://opendefinition.org/licenses/</p>
<p>Open Data Commons Public Domain Dedication and License (PDDL, http://www.opendatacommons.org/licenses/pddl/1-0/)</p>	<p>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.” It is has been declared compatible with the Open Definition by the Open Definition Advisory Council http://opendefinition.org/licenses/</p>
<p>Creative Commons Attribution 4.0 License (CC-BY 4.0, https://creativecommons.org/licenses/by/4.0/legalcode)</p>	<p>The CC-BY 4.0, is the least restrictive of the Creative Commons licenses. It permits the user to share and adapt material with minimal attribution requirements. It is has been declared compatible with the Open Definition by the Open Definition Advisory Council http://opendefinition.org/licenses/</p>
<p>Nongovernmental Policies and Agreements</p>	<p>Nongovernmental organizations 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.</p>

Guideline 1C. The CC0 or PDDL waivers of rights are the preferred voluntary, nongovernmental approaches to facilitate the legal interoperability of research data. The non-restrictive (“common-use, attribution-only, CC-BY 4.0”) license may also be considered when disseminating the data for broad use.

Many public and private research organizations and individual researchers now use non-restrictive licenses for both research data and more copyrightable information, such as research articles, books, pictures, software, and other types of information. One non-profit organization, the Creative Commons (CC), has developed a suite of such standard licenses with “some rights reserved”.

According to the CC website (<http://creativecommons.org>) 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, which 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. The CC licenses have been reviewed and ported by over 70 countries and are in use throughout the world. CC has estimated that there are over one billion documents with a CC license or waiver of rights online.

Creative Commons, and other common-use licenses, are predominately intended for use with creative works. They rely on copyright or other intellectual property rights for enforcement. In some cases and in some countries, the question of whether a dataset is copyrightable may be difficult to determine. For material published under the Creative Commons Attribution License or common-use licenses with

similar characteristics, enforcement can be largely a moot consideration, because of the time and expense of going to court.

Similarly, the point is often raised that these licenses have onerous attribution requirements, which compound when multiple data or other material are combined. Although that may be the case, the Creative Commons licenses also provide rights holders with options to negate such “attribution stacking.” For example, a research project group combining data from multiple datasets could agree on a single, brief attribution, recognizing the group, rather than a long list of individual project members. Alternatively, a rights holder may not require being attributed at all, although this would be uncommon.

Another important consideration in the selection of an appropriate license or waiver is the quality of the disclaimer of warranties and limitation of liability clauses. Generally, these clauses are robust, but various licenses contain subtly different disclaimers, which may be important. For example, where the CC-BY 4.0 (in fact, all of the Creative Commons 4.0 licenses) expressly disclaims liability for “negligence”, that feature is presently absent from the CC0 No Rights Reserved instrument and PDDL disclaimer clauses. The most recent versions of Creative Commons licenses also permit supplementary notices that, if properly drafted by the licensor (i.e., do not otherwise contradict the license or remove obligations or rights imposed by the license), may further disclaim liability in specific circumstances. These attributes may or may not be important, depending on, among other things, the characteristics of the collection of data, or the law of the jurisdiction in which the user or rights holder, or both, are located.

Another feature often noted about CC0 is its potential incompatibility with moral rights law⁶. That is because, as noted in the CC0 FAQ, “the laws of some jurisdictions don’t allow authors and copyright owners to waive all of their own rights, such as moral rights.” (Creative Commons, 2015). Moral rights are related to copyright and are used to protect the non-economic interests of the author of a work. Such rights may include the right to protect the integrity of the work, to honor and reputation, or to have the work published anonymously (Lerman, 2012). It is not possible in many jurisdictions to waive the inherent “author’s moral rights”. However, the CC0 only operates to the extent permitted by relevant law. If moral rights cannot be waived in a particular jurisdiction, then generally the interpretation and effect of CC0 will not extend to the waiver of moral rights in databases, to the extent that they do apply. Because CC0 only makes grants “[t]o the greatest extent permitted by, but not in contravention of, applicable law”, CC0 may not be fully equivalent to material that has actually fallen into the public domain by virtue of term expiration or non-protectability.

Accordingly, for the reasons above, we consider that there are a few options open to the rights holder(s) of research datasets that encourage broad reuse of their data. For datasets that have no original and creative elements at all, or are not subject to copyright or *sui generis* database rights restrictions (e.g., because the term of protection has expired), we recommend the application of either the Creative

⁶ The Creative Commons wiki provides an explanation of the CC0 waiver and moral rights as follows:

(https://wiki.creativecommons.org/wiki/CC0_FAQ#Does_CC0_really_eliminate_all_copyright_and_related_rights.2C_everywhere.3F)

“The laws of some jurisdictions don’t allow authors and copyright owners to waive all of their own rights, such as moral rights. When the waiver doesn’t work for any reason CC0 acts as a free public license replicating much of intended effect of the waiver, although sometimes even licensing those rights isn’t effective. It varies jurisdiction by jurisdiction. While we can’t be certain that all copyright and related rights will indeed be surrendered everywhere, we are confident that CC0 lets you sever the legal ties between you and your work to the greatest extent legally permissible

Commons Public Domain Mark⁷ or a similar statement of no known rights. For datasets that contain original and creative elements and are subject to copyright or database restrictions, or if the rights holder is uncertain whether copyright or *sui generis* database rights apply (see Guideline 2B), we recommend the use of either the CC0 or the PDDL, or alternatively, the CC-BY 4.0, having due regard to the options for attribution expressed above. We include the latter because it contains arguably a superior disclaimer of warranties and limitation of liability clauses (with an option to further specify and limit liability by way of supplementary notice), which may be important to some dataset rights holders. It may also provide an alternative if CC0 is considered unacceptable for moral rights purposes, despite its self-limiting application to the extent that relevant law permits. Finally, we do not endorse the use of any other standard Creative Commons licenses for research datasets.

Guideline 1D. In asserting any rights and in applying access and reuse terms and conditions to any research data made publicly available, all members of the research community should make such data available equitably to all users, including the most disadvantaged ones.

Equity and fairness are important qualities for research conducted with public support. Because such these types of values support openness to public data as a default rule, they are important as well for justifying legal interoperability. Such values are generally not legal requirements, but are ethical and normative ones.

If the research data were collected using public funds, the taxpayer has already paid for generating them. The entire activity should be presumed in the public interest and therefore open to the public, absent some legitimate countervailing and overriding purpose, as discussed below in Principle Three and its Implementation Guidelines. The data may be considered as belonging to the public that paid for it, with the research data collector and disseminator acting only as an agent on behalf of the public. The benefits from the access to and use of public data can then be equitably available to all users, with the least number of restrictions, in many cases as a global public good.

The open provision of various types of research data makes them available for use by those least able to afford any access fees, especially in developing countries. They provide a more level playing field and opportunities for greater interactions with the global research community, among many other benefits (CODATA PASTD, 2014; National Research Council, 2012b). Such policies also enable the supply of data to data-poor areas, where such information can be most needed, even in more economically developed countries. Excluding the other researchers and the public in general from publicly supported data may be seen as unethical and inequitable. Such considerations are important for data rights holders to determine whether to make research data that are collected by the public sector and with public money freely available, and whether to refrain from claiming intellectual property rights in those collections of data (if they are even allowed to do so).

At the same time, it is important to recognize that most research funders have not committed to paying for research data access, which, absent a sufficiently robust financial sustainability plan, can lead to suboptimal curation or even a complete loss of useful research data. This is also particularly true in less

⁷ It should be noted that 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 all other jurisdictions.

economically developed countries, so such values and ethics need to be taken into consideration as well.

The users of research data, whether publicly or privately generated, also have ethical obligations. They must use the data in accordance with the terms and conditions imposed by the provider of the data, as well as any laws of the jurisdiction in which they are located and the norms of the community within which they work (see also Principle Two and Implementation Guideline 2B). Research data users need to appreciate that for many institutional and individual sources of data, reputational risks can be as important to protect as legal risks. At the same time, the fewer constraints that are imposed by the provider, the more creative the users can be with the data, leading perhaps to greater opportunities for serendipitous results. Also, the fewer restrictions there are on the users, the fewer chances there are to contravene them.

Principle Two: Determine the rights to and responsibilities for the data.

Guidelines for Implementation

All members of the research community need to know who has the rights to the research data that they disseminate or use. More specifically, both research data producers and users should know their rights and responsibilities under pertinent laws and policies, and participate through their proxy institutions in the legislative or policy processes related to data sharing principles and legal interoperability. This is especially true in the emerging public data environment that is often referred to as “open data.” This online environment can only be made to foster scientific progress effectively if data providers make their data available and usable to others within the rules, and if data users take advantage of the data that are made lawfully accessible and usable.

This Principle is therefore directed at every person in the research process who has some role and responsibility to produce, access, share, use, or combine research data.

Guideline 2A. Research data disseminators need to establish who or what entity has the rights to any given collection of data before the data are disseminated to others.

A disseminator of research data can be anyone who makes the data publicly available. This can be the original producer(s) of a dataset, the organization under whose auspices the dataset was produced, or a data repository in which the data have been deposited. Prior to making the data publicly available, the disseminator must ascertain who has the right to do so and what rights attach to those data.

Establishing the holder of rights (assuming there is any) in a research dataset is important because all downstream decisions on making the data available for reuse are dependent on knowing who has the right to decide what the terms and conditions are. Research data that are to be made available to third parties should therefore always include a statement on the rights.

Moreover, a disseminator of research data in many cases may not be the same entity as the holder of rights in those data. The rights holder of a collection of data also is not necessarily synonymous with its original producer, since the rights may be transferred—and made either less or more restricted—by agreement of the parties.

For example, an individual researcher may produce original data, but the rights are controlled by the employing institution. A data repository that a researcher may deposit data into, such as GenBank, Dryad, Figshare, and the like, should have some transfer of rights request as part of the deposit.

Likewise, a researcher or a data center may combine many data sources, but one or more of those data sources have restrictions placed on them that may “infect” the other, less restrictive data sources (unless fully separable) and make the derivative dataset subject to the restrictions of the most restrictive data source. Or, the person(s) who ought to be recognized or credited if data are to be reused may well be different from the rights holder(s). As a result, any provider or disseminator of research data must know what entities hold any rights to the data and abide by those rights.

In particular, research organizations should have clear rules that allow determining what rights are in data controlled by that organization and to what extent the rights in those data may be controlled by others, including their employees. The clarification of this matter should be incorporated into contracts of employment, and into research cooperation and grant agreements. The exact status of rights in any data should be transparent in two respects, stating clearly (i) who is or are the rights holder(s) of any subset of a dataset, and (ii) who is or are the rights holder(s) of the dataset(s) containing these data. The rights in the data that are part of a dataset do not have to be controlled by just one natural or legal person.

These aspects and any other relevant ones should be addressed in the data management plans that ideally precede the compilation of the data. The bureaucratic burden connected to clearly stating rights pertaining to any data and datasets can be greatly reduced by research funding and research performing organizations through the provision of clear guidance. Further simplification of these terms and conditions can be achieved if research funding and research performing organizations harmonize their policies with respect to research data rights and responsibilities (see also Principle Five and the accompanying Implementation Guidelines on Harmonization, below).

Guideline 2B. It is also the responsibility of research data users to abide by the rights applicable to the collection of research data, as well as the specific user rights in the jurisdiction that the data are being used.

It is well-known that “ignorance of the law is no excuse.” It is a legal duty for the user of any research data to ascertain the rights in those data and abide by them. Most data sources are not in the public domain and have various restrictions based on intellectual property rights or private contract placed on them (see Principle Three and its Implementation Guidelines). Data users thus need to be familiar with and generally understand the relevant laws, policies, and norms relating to the legal status of the research data that they are using. At the same time, it can be a lot to ask of every user of research data, and the recommendations provided elsewhere in these Implementation Guidelines can help obviate these burdens.

Although these Principles and Implementation Guidelines are focused primarily on providing guidance to the data producers and the rights holders in sharing their data from a legal perspective, users too, however, have some limitations and exceptions to copyright, despite the fact that the laws are written primarily for the rights holders. Researchers who receive data may have to deal with various restrictions on them if they wish to reuse, combine, or share those data. This is real-world problem that faces all researchers. Although a treatise on limitations and exceptions or defenses to copyright for research data (see, e.g., Reichman and Okediji, 2014) is beyond the scope of these Guidelines, the following points may be useful for research data users to consider:

- Assess whether a statutory limitation or exception permits an applicable research or other public-interest use. Limitations and exceptions to copyright law may include: quotation, news reporting, judicial and legislative documents, teaching, research, parody, rights of the visually impaired, and other public interest uses or users.
- Researchers in the United States also may be able to rely on the fair use provision contained in the Copyright Act, noting in particular the Codes of Best Practice in Fair Use established for various scholarly communities (Association of Research Libraries et al., 2012), which also promotes harmonization of practice within the communities adhering to the Code (see Principle 5). In the United Kingdom and other Commonwealth countries, researchers may be able to use the fair dealing defense. In other jurisdictions, such as in some of the countries that are members of the European Union, there may be some statutory exceptions available as well, but in all cases, the researcher needs to make sure that the use is lawful.
- If some statutory limitation or exception, or a defense option, does not apply, the researcher can go back to the source of the data and try to negotiate more reasonable terms for the intended use(s). This can entail a substantial transaction cost of time and money, however.
- If there is no statutory exception or limitation, or a defense, or the researcher is turned down by the rights holder (rather than being offered some kind of use fee or other conditional permission), or the rights holder is impossible to determine, then there may be the possibility of finding a substitute dataset with similar data.
- It also may be possible to abide by the restrictions and keep the restricted data separate from the other unrestricted data, so that the entire derivative dataset is not “infected” with the strongest restriction(s).

Various resources, including these Implementation Guidelines, describe these legal concepts in layman’s terms and explain the differences between legal jurisdictions (e.g., the application of copyright laws or the EU Database Directive). It also is important to emphasize that just like the laws, policies, and agreements used for making research data open may change over time, so can the restrictions that are placed on them, so the user needs to have access to the most recent and accurate information about any restrictions that might exist. In the event that clear explanatory resources are insufficient or nonexistent, the advice of legal counsel is encouraged by either the providers or users of research data (see also Principle Four and Implementation Guideline 4C).

Guideline 2C. Representatives of research communities who are also experts in policy and legal issues are encouraged to participate in fora that develop and implement laws and other norms governing access to, and the reuse and legal interoperability of, research data.

Research representatives with knowledge of the legal and policy intricacies of the research process need to engage in political and legislative processes—whether formal or informal—that relate to the legal interoperability of research data, and advocate for legal frameworks that facilitate scientific progress. For example, the National Academy of Sciences, and research library and university organizations in the United States, were actively involved in providing testimony, meeting with legislative staff, and drafting legislative options for the congressional attempts to enact U.S. database rights protection legislation over five years, beginning in 2007 (National Research Council, 1999; Reichman and Uhler, 1999).

It is also a responsibility of those developing laws to explicitly seek out expert opinions.

Guideline 2D. A well-conceived educational process for researchers regarding rights and responsibilities in research data should be developed and adopted by relevant institutions.

Institutions that educate and train researchers in all disciplines should include the legal interoperability of research data and the policies and procedures for managing research information, generally, as part of the responsible conduct of research. The overall goal should be to instruct future generations of researchers in the proper handling of research data. Useful educational processes might include: formal higher education courses, university training and continuing professional education from well-established research data organizations, seminars and webinars targeted at supervisors and mentors, participants in the publishing process (e.g., authors, editors, reviewers, publishers, librarians, archivists, data repository managers, and researchers generally), and the design and funding of targeted research programs.

Principle Three: Balance the legal interests.

Definition and Discussion of Terms

| [Confidentiality laws and policies](#) | [Copyright](#) | [Copyright infringement](#) | [Cultural resources protection](#) | [Endangered species protection](#) | [Intellectual property rights](#) | [National security or public safety laws](#) | [Patents](#) | [Periods of exclusive use](#) | [Personal privacy protection](#) | [Sovereign genetic resources protection](#) | [Sui generis database rights](#) | [Traditional Knowledge protection](#) |

Guidelines for Implementation

Freedom of access and reuse of research data that enables legal interoperability must be balanced against legitimate interests to withhold, restrict, or control the use of such data by the rights holders, or in the interest of the larger society, as determined by lawmakers. Legitimate interests vary by jurisdiction and situation, but generally reflect laws or established norms regarding protection of intellectual property rights, national security and public safety, personal privacy, confidentiality, endangered species or cultural resources, sovereign genetic resources, Traditional Knowledge, and defined periods of exclusive use prior to the publication of research results.

At the same time, not every restriction is valid. There can be restrictive declarations accompanying data and information that have either an invalid legal basis or are not justified. The following brief definitions are provided to help guide the reader in understanding these other restrictions and protections that may exist.

Intellectual property rights: As used in these Implementation Guidelines, “intellectual property rights” include copyright, *sui generis* database rights, patents, and other similar laws that establish a regime for access, use, or reuse of information, including research data or metadata. It does not include other restrictive information laws that are also discussed in these Guidelines. See the individual entries for Copyright, *Sui generis* database rights, and Patents.

National security and public safety laws: National security and public safety (or public order) laws can declare certain data as secret or confidential for a specified period of time. That means that such data are only held within certain sectors of the government by those with a need to know. It is forbidden to make use of the protected data and to disseminate them or the information based on these data to the

public. The main applications of such restrictions can be found in the field of military defense, intelligence agencies, or ministries and organizations that protect the public safety.

Personal privacy protection: In most countries, data referring to individual persons (e.g., medical information, data on social status, penal record, or financial information, among others) are treated differently from all other data and restricted. Such data may affect privacy insofar as they can contain information about a defined person that the person may want to keep private. In order to solve this conflict of interest, many jurisdictions have enacted personal data protection laws. As a common element, they state that data referring to an individual person may only be released to the public if certain conditions are fulfilled (e.g., agreement by the concerned person, legal authorization, or public interest).

Protection of confidentiality: Confidentiality laws can protect public and private interests by withholding certain data and information from the public. This may refer to administrative consultations in governments or to secrets in the private sector, such as technical know-how, customer information, trade secrets, non-competition, and the protection of other commercial interests. Such presumptive or real interests often form the background to information policies or agreements with employees by institutions and enterprises that try to withhold data and information they produce, or control the use of them.

Protection of endangered species: Specific data and information referring to endangered species must, in certain circumstances, be withheld in the interest of their protection. Open access to data on the precise location of an endangered species, for example, can contribute to the misuse or extinction of the organism. Open access to relevant data could therefore jeopardize its protection and even existence. (See, e.g., the Convention on International Trade in Endangered Species of Wild Fauna and Flora, IUCN, 1973.)

Protection of cultural resources: Secrecy may apply to data and information about cultural resources. The open accessibility of such data may, under certain circumstances, provoke the illegal exploitation of cultural resources and put those resources at risk of theft or destruction. Open access to research data could, under certain circumstances, compromise the protection of such resources. (See, e.g., the Convention on the Means of Prohibiting and Preventing the Illicit Import, Export and Transfer of Ownership of Cultural Property from UNESCO, 1970; the Convention on the Protection of World Cultural and Natural Heritage, UNESCO, 1972; and the Convention of the Protection and Promotion of the Diversity of Cultural Expressions, UNESCO, 2005.)

Protection of sovereign genetic resources: National authorities have the right to legislate on the access to genetic resources, based on the legal requirements of the Convention of Biological Diversity of 1992 and its Nagoya Protocol of 2010. This legislation should create conditions - via prior informed consent (PIC), mutually agreed terms (MAT), and access and benefit sharing (ABS) - that facilitate access for environmentally sound uses and not impose restrictions that run counter to the conservation and sustainable use of biodiversity. The laws apply equally to the physical genetic resources and to the genetic data about them.

Protection of Traditional Knowledge: The World Intellectual Property Organization defines Traditional Knowledge (TK) as “the intellectual and intangible cultural heritage, practices and knowledge systems of traditional communities, including indigenous and local communities” (WIPO date unspecified). The laws in some countries prohibit the misappropriation of various data and information that have cultural significance and protect indigenous peoples’ rights. Examples may include various designs and

traditional medicines. Some cultures also prohibit taking pictures of designated persons and disseminating them. (See, e.g., Recommendation on the Safeguarding of Traditional Culture and Folklore, UNESCO, 1989). However, many countries do not protect such rights at all. In those cases, “soft” law, such as governmental policies, and ethics can be used to protect the holders of Traditional Knowledge and help ensure that indigenous people’s rights are respected (Lead et al., 2016).

Periods of exclusive use of research data: Many research funding institutions, universities and research institutes, and academic publishers impose time embargos for the access to and the reuse of scientific data produced during the course of the research. These restrictions can be based on either a formal contractual basis or a less formal policy one, and are expressions of a self-defined institutional policy ideally incorporating on the norms of a defined research community. (National Research Council, 1997).

Restrictions on access to and reuse of research data, especially those generated through public funding, therefore generally ought to be applied only pursuant to legitimate reasons, such as those defined above. The following Implementation Guidelines can help achieve this balance.

Guideline 3A. As a default rule, intellectual property rights in government or publicly-funded collections of research data should be waived as far as possible and distributed in the least restrictive manner.

Ethical norms compel all participants in the research process to refrain from reserving intellectual property rights in collections of research data as a default rule, especially to those obtained partly or fully from publicly funded research (e.g., European Commission, 2016 and OECD, 2007) (see Principle One and Guideline 1.D). Copyright or *sui generis* database right applies automatically to any work that falls within its ambit. If a dataset qualifies as a copyrightable work, or is protected by a database right, the holder(s) of such rights are urged to place their data in the public domain or allow access to and reuse of those data in the least restrictive manner (see Principle One and Implementation Guidelines 1.A, 1.B, and 1.C).

Guideline 3B. Governments and public research institutions need to justify any restrictions on research data that may be imposed over and above any restrictions on the access and reuse of data that are defined by existing legislation.

Many of the restrictions on open access and reuse of research data described in the definitions section below cannot be waived or modified by the rights holder. Nevertheless, in order for the rights holder(s) to restrict the reuse of research data or to keep them secret, the restrictions should be justified by an explicitly stated and overriding public or private interest, and be limited in time.

Guideline 3C. Policymakers should consider public interests in developing rules for access to and use of publicly generated data.

Where scientific research is partly or fully funded by public institutions, private interests -- especially commercial interests -- should not, as a default rule, hamper access to or reuse of research data. Public research should serve the entire society.

Where personal privacy protection or confidentiality interests require withholding of certain data, an evaluation should determine whether sharing can be assured by making such data available in

anonymized or aggregated form. The privacy laws in many countries contain exceptions for using or processing personal data for research purposes, leading to access to these data for researchers only and to reporting based on the data only in a fully anonymized way.

Guideline 3D. Public research funding organizations and the rights holders of public research data sources should reduce time embargoes for exclusive personal periods of research use to the minimum necessary.

Time embargos on the release of research data may be justified by scientific needs, especially in order to verify the results. The duration may vary by discipline and is already frequently prescribed by research funding agencies or scientific journals. However, such restrictions should be narrowly limited in time and be specified by the funder of the research. Researchers also should be aware and abide by any defined embargo periods for research data.

Guideline 3E. As a default rule, all rights holders of research data that are partly or fully funded by the public sector need to avoid the use of individual contracts or agreements that restrict access to and reuse of the data.

Individual contracts or “end-user licensing agreements” (EULAs), whether commercial or not, may impose restrictions to access and reuse of research data and information that go far beyond legally justified interests. Even if justified, they are non-standard, expressing the individual interests of the parties for a particular situation or subject matter. As non-standardized rules, they can invalidate the general interests in information exchange and freedom of access and reuse of research data that are partly or fully funded by the public sector. They also can pose significant obstacles to the interoperability of data and require time-consuming and costly solutions on a case-by-case basis, if that is even possible. We therefore recommend not to enter into such agreements regarding publicly funded research data, if at all possible.

Principle Four: State the rights transparently and clearly.

Definition and Discussion of Terms

| [Certainty](#) | [Metadata](#) | [Research data](#) | [Rights holder](#) | [Rights statement](#) | [Transparency](#) |

Guidelines for Implementation

An unambiguous statement of legal rights and policies pertaining to all data used in research is essential to provide sufficient notification of the legal rights (if any) retained by the rights holder(s). The legal terms and conditions of those data must be clearly stated to enable legal interoperability. Legal transparency and certainty minimizes exposure to legal risk for all parties, can remove barriers to research, and facilitates the long-term access and reuse of data resources, including for those in the public domain. When a collection of data is in the public domain, it should be labeled clearly as such, consistent with Implementation Guideline 1.B.

The first three Principles and their Implementation Guidelines describe laws, policies, and practices that are relevant to the producer, disseminator, or user of research data. They assist the stakeholders in the research process to understand the rights and responsibilities concerning collections of research data and the data contained in them, including an array of mechanisms for promoting their legal interoperability. Many of the relevant laws and policies are not always restated in conveying the data to the user. Providers or distributors are in the best position to state clearly the restrictions imposed. By the same token, in order to increase the amount of data being made available for reuse, the demands on transparency and certainty should not be excessive and should not involve any risk of liability, to the extent possible and foreseen. Towards this end, members of the research community, especially committees of legal experts and data professionals, should develop the appropriate language to use in an ongoing process.

Guideline 4A. The use of standardized electronic statements regarding the legal rights retained (if any) by the rights holders and providers of research data can greatly assist in their comprehensibility by a wide audience—including by machines.

Statements concerning the legal rights pertaining to research form part of the metadata (documentation for the user). Metadata are the principal mechanism through which transparency and certainty can be achieved. The accessibility and reuse of research data is not only dependent on the rights in them, but also on the ability to communicate this status effectively. In many cases, reuse of research data will occur as a result of electronic searches (including text and data mining), in subsequent machine manipulation (such as merging and integration with other data, visualization, and so on), and further dissemination of the original or derivative dataset(s).

In practice, however, the majority of established metadata standards in use today for research data do not include rights elements that convey the legal information needed by users to clearly understand their rights and responsibilities in reusing the data appropriately. Sometimes, information about ownership rights and usage terms and conditions are loosely coupled to the dataset files in the form of copyright notices or as open licenses posted on the online dataset “landing page.” In such cases, legal information governing the reuse of the data can be easily disconnected from the data and lost, leaving potential users unaware about violating the law, assuming legal risk, or actually infringing potential intellectual property rights. Removing legal uncertainty regarding reuse requires consistent and predictable rights information that remains associated with the data being used.

The metadata for any publicly available collection of data therefore should include all information—a rights statement—necessary to understand the legal control of the data and any terms and conditions governing their access and reuse. Specifically, the rights statement should include a set of standard expressions that, at a minimum, communicate the intellectual property status of the research dataset (see, e.g., Principles and Implementation Guidelines One and Three). Additionally, this field may indicate the presence of any terms and conditions governing use of the data, including whether it is freely accessible or only with a login requirement or perhaps a defined payment. Finally, the rights field may be used to convey the name and version of any mechanism used to retain rights in the dataset, and the legal jurisdiction where it applies. Legal interoperability—even for data that may have some additional restrictions—thus can be significantly enhanced by means of a common taxonomy in rights statements that are both human and machine readable, that is, standard ways of expressing the open access or restrictions governing a given research output.

There are several models for established rights statements that are being applied to research materials. They include the RIOXX metadata profile, developed for UK repositories to track compliance with funder open access mandates (<http://riox.net/>), and the RightsStatements.org taxonomy, from the Digital Public Library of America and the Europeana Digital Library, which provides 12 standardized rights statements for online cultural heritage resources (<http://rightsstatements.org/en/>).

Guideline 4B. The rights holder(s) of any given data used in research should engage competent legal counsel when it is necessary to determine the applicable law(s) and to clarify the differences among jurisdictions.

Statements concerning the rights inherent in research data need to be interpreted in the context of applicable statutory law. In many cases, however, a reasonable and transparent statement of the rights concerning research data will not answer all of the questions potential users may have about the terms of reuse, because of the great variations in statutory law and institutional policies that may be applicable, depending on which jurisdiction applies to the dataset. The advice of well-informed legal counsel is important to resolve such questions and others concerning the legal status of research data, especially if the data are used in other jurisdictions or sectors.

Guideline 4C. Rights holders should inform users about any special terms and conditions of use.

Even with the best efforts by the rights holders, statements concerning the rights situation may not be sufficiently comprehensible for potential users in some cases because of the complexity of the issues. We suggest therefore additional efforts to help inform potential users of research data about the terms of use. Rights holder(s) and those authorized by the rights holder(s) of research data can use two approaches:

Collection of data with common usage options.

Research data with a comparable set of specialized rights could be gathered in collections that are made available by public data repositories and other institutional providers, with special support concerning legal issues of reuse. Data within the collection could be marked to be available for certain uses, which could include conditions that ought to be met by the users. The collections of such research data may provide special environments that enable qualifying users to meet designated access conditions. For example, restricted data access collections can be used to protect endangered species or sensitive archeological sites, or to protect against breach of anonymity protections through combinations with externally available data sources.

Accessibility to data governed by specialized subject-matter organizations or committees.

The approach used here is for specialized bodies with expertise in the subject-matter of the data, such as managers of community data repositories, to develop access and reuse rights. Existing examples include the Data Access Committees of the European Genome-Phenome Archive (<https://www.ebi.ac.uk/ega/dacs>); the NIH Data Access Committees (https://gds.nih.gov/pdf/NIH_DACs_Chairs.pdf); the policy approved by the Steering Committee of the Federal Geographic Data Committee (FGDC) in 1998 (<https://www.fgdc.gov/policyandplanning/privacy%20policy/?searchterm=confidentiality>); and in the social sciences (<http://www.dwbproject.org/>). If they do not exist, the criteria used by such entities

for checking research projects should be aligned in order to increase the number of collections that can be made available on the basis of the decision of the entity.

Principle Five: Promote the harmonization of rights in research data.

Definition and Discussion of Terms

| [Norms](#) |

The implementation of the other principles can be facilitated through the harmonization of the data rights laws, policies, or norms across funding and research organizations. This harmonization might be achieved by different means. Intellectual property rights and their waivers are governed by laws or policies that are mandatory and apply automatically under each national jurisdiction. This can be called the top-down approach.

Governments and organizations engaged in the research process can facilitate legal interoperability using compatible and consistent terms and conditions for research data rights across as many jurisdictions as possible. Possible mechanisms for achieving such harmonization include treaties, legislation, public policy, common-use licenses, and waivers of rights. In harmonizing laws and policies, including statutory limitations and exceptions, stakeholders need to take into account prevailing practices, norms, and other factors relevant to the research community. The goal of standardization, however, must be balanced against the need for flexibility.

Nevertheless, such laws or rules also leave a lot of room for individual decisions based on the autonomy of the rights holder. Consequently, different stakeholder communities accessing and reusing data (e.g., libraries, archives, museums, academic publishers, or various digital infrastructures), as well as the distinct, non-governmental research discipline communities themselves, can and should harmonize their practices whenever the legal system provides opportunities for developing and implementing common understandings. These kinds of actions by the practitioners themselves are known as the bottom-up approach.

In this section, we propose Implementation Guidelines to assist in the harmonization of laws, policies, and norms pertaining to the legal interoperability of research data.

Guideline 5A. Research policymakers and practitioners can use both top-down and bottom-up approaches, and mixes of both, to harmonize rights concerning research data.

There is a tendency to understand that harmonization implies an adjustment of differences and inconsistencies among different rights in research data to make them uniform. But while uniformity is the most expected outcome when parties engage in harmonization processes, it is only one of the potential results. Many other results can also imply trends towards mutually compatible results, and not necessarily uniformity. The idea that harmonization is a black or white scenario (harmonization equals uniformity and uniformity is either achieved or not, hence harmonization either is total harmonization or no interoperability is possible) is wrong and moves scientists and policy makers away from other types of harmonization that paradoxically might lead to more easily achievable mutual compatibility of data. Scientists and policy makers should be aware of these other results, different from uniformity, in

order not to reject harmonization processes right away when the ideal goal of total harmonization appears to be unrealistic.

Thus, there are different levels of harmonization of rights in research data, which include:

Total harmonization. When several or all jurisdictions or members of a distinctive scientific community use a common standard –whether legally binding or normative--it can be:

- the result of State action via an international treaty, executive agreement or similar legal instrument broadly adopted by governments, or consistent State practice developed as customary law over time; or
- the result of an agreement within a scientific community, which can include the legal approaches recommended in the Implementation Guidelines in Principle One.

[At the global level, see e.g., the Data and Information Management Strategy (DIMS), developed by the Standing Committee on Antarctic Data Management of the Scientific Committee on Antarctic Research (SCAR) as an implementation of the Antarctic Treaty, available at: <http://www.scar.org/data-products/scadm>; or the 1996 Bermuda Principles of the Human Genome Project requiring that all DNA sequence data be released in publicly accessible databases within twenty-four hours after their generation, available at: http://web.ornl.gov/sci/techresources/Human_Genome/research/bermuda.shtml.]

Partial harmonization. Partial harmonization can occur when a common standard that is only applicable to some aspects of any given legal issue is adopted by all jurisdictions or members of a distinct scientific community (e.g., only for access, but not for use or reuse of data). For example, see the Electronic Chart Display and Information System (ECDIS) for nautical charts, under which governments allow only visualization/display, but not reuse most of the data, based on safety reasons <http://www.ecdis-info.com/>.

Alternative harmonization. This occurs when several standards are agreed and available, and fully harmonize access in two or more different ways, leaving the choice of the standard to the data provider. [See, e.g., the potential use of DiGIR, BioCAsE, or TAPIR metadata standards in the Global Biodiversity Information Facility, GBIF, <http://www.gbif.org/resource/80635>.]

Optional harmonization. A standard is available to everybody, but is not mandatory; it is up to the rights holder of the collection of data to determine if it wants to opt-in and adopt the standard, or to opt-out and not use the agreed standard; but if the agreed standard is used, its terms and conditions are the same for everybody. Of course, it can also be partial: opting-in, or more frequently, opting-out can be limited to some aspects or issues. [See, e.g., the Open Research Data Pilot option provided for projects that decide to opt for open access and the sharing of data in the EU Horizon2020 research program at: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf. This rule was subsequently changed to an opt-out approach, where open research data becomes the default setting.]

In all cases, harmonization approaches for open data should be based on three key factors: a) legal predictability and certainty; b) ease of use and understanding; and c) low costs to users.

Guideline 5B. Bottom-up actions based on voluntary, private law mechanisms, can be an effective approach for research individuals or institutions to take in the absence of government action, but are relatively fragmented and less harmonized.

A rights holder can voluntarily use a Creative Commons CC0 No Rights Reserved instrument, the Public Domain Dedication License (PDDL) of the Open Data Consortium, or the Creative Commons Attribution (CC-BY) 4.0 license, which are described in Table 1 of Principle One Implementation Guideline 1B. These legal instruments provide bottom-up harmonization, since the standards are interpreted in the same way in almost all jurisdictions (see, e.g., the “encouragement” of the European Commission to use CC0 (or CC-BY) in all Horizon2020 research projects (European Commission, 2016).

The CC0 waiver can provide a workable scheme for self-governing by scientific communities that harmonizes even better than the top-down statutory law or national policy schemes. For example, the CERN Large Hadron Collider (LHC) Computing Grid Compact Muon Solenoid (CMS) data management system opted for a CC0 approach instead of negotiating an amendment to the Convention that established CERN (available at: <https://cms-docdb.cern.ch/cgi-bin/PublicDocDB/RetrieveFile?docid=6032&version=1&filename=CMSDataPolicy.pdf>).

The more that all actors in a distinct scientific community agree that they deal with a certain type of research data, the easier it is to achieve total harmonization of the rules applicable to public access and reuse of those data. In astronomy, for instance, see the work of the International Virtual Observatory Alliance-IVOA (<http://www.ivoa.net/>); or for data on metabolic phenotyping, see MetaboLights (<http://www.ebi.ac.uk/metabolights/>). At the same time, other discipline areas, such as the humanities generally, encounter greater difficulties in the legal interoperability of research data, since many of the collections of data they use or extract data from are copyrighted or restricted in some way.

Harmonization can be also built on norms instead of legal rules/agreements/licenses/waivers (see, e.g., the Human Genome Project agreement on DNA digital sequences flowing into the public domain within 24 hours, cited in Implementation Guideline 5A, above). The adoption of norms signifies consensus within research communities, but takes time to develop; individual and voluntary waivers of rights, or institutional agreements are much faster to adopt, but are specifically limited to the collection of data to which they pertain. Furthermore, norms require additional incentives and peer acceptance, because their logic can be less evident than the use of more formal agreements. They have proven to be effective in promoting research data sharing, however, since they maximize the welfare of common advancement of research while minimizing the transaction costs of having to negotiate a legally binding agreement.

Guideline 5C. Top-down harmonization through “hard” law, such as multilateral treaties or executive agreements, or national legislation or administrative regulation, can work in some contexts and can be extremely useful as a broad harmonization tool.

Harmonization through government action (top-down approaches) is efficient, although it is likely to be more controversial and difficult to achieve consensus, and therefore more time consuming (see “Government Mechanisms” in Table 1 under Implementation Guideline 1.B). Such harmonization also is dependent on the willfulness or the capacity of the scientific community to actively engage in the political process.

Harmonization through treaties is certainly possible although seldom achieved. For example, the total harmonization of data policies under the Antarctic Treaty system [Section III.1.c) provides that:

"Scientific observations and results from Antarctica shall be exchanged and made freely available" (see <http://www.scar.org/>). This represents an approach that should not be regarded as a mere curiosity, but as an achievable—though difficult—goal. The agreed policies of major international organizations, such as the Group on Earth Observations with its 3 Data Sharing Principles (GEO, 2014b), provides another example of top-down harmonization using a soft-law approach for a certain type of public research data.

Sometimes harmonization is achieved through the addition of unilateral actions of different jurisdictions by passing or amending legislation or policies that approach national requirements to those of other countries, or even through more internationally oriented judicial interpretation of national statutes and regulations. For example, see, the Australian Research Council Open Access Policy, established in January 2013 (<http://www.arc.gov.au/arc-open-access-policy>); the U.S. White House Executive Memorandum on Public Access to Research Results of 22 February 2013 (<https://www.whitehouse.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research>); and the E.U. Open Access policy established in July 2012 (<http://ec.europa.eu/research/swafs/index.cfm?pg=policy&lib=science>). These three policy statements, although of somewhat different legal effect, seem to have opened a trend toward open access as a general rule for data derived from publicly funded research. While the objectives of these individual statements of policy are broadly applicable at the national level, such an approach to actual harmonization remains underdeveloped.

Guideline 5D. Process-based approaches such as workflows decision-making charts, decision making apps and tools, or scoreboards might be considered useful tools to promote harmonization.

Some research communities might have totally different approaches and differing views about the exact scope of the openness of the data provided or reused by their members, making it impossible to harmonize to any degree rules or norms on access and reuse of such data. An alternative approach could be the adoption of common decision-making strategies that clarify, under a common understanding, the implications and legal consequences of the different choices available. This may be the case, in particular, when agreement to common substantive protocols, standards, or rules or norms are not foreseen as achievable in the short or mid-term. Transparency of decisions taken under common decision-making processes may ultimately lead to common understandings on the openness model to be adopted by research communities; scoreboards might even help through indicators, to self-assess the level of openness that the tentative or taken decisions entail.

See, e.g., as an example of a chart, the Licensing Decision Flowchart developed by the Web2Rights OER IPR Support project <http://www.web2rights.com/OERIPRSupport/diagnostics.html>, and, as an example of an app, Camden, the open-source, cross-platform tool designed to provide legal guidance on issues concerning copyright, in Richard Hosking et al., "An eScience tool for understanding Copyright in Data Driven Sciences." <https://rd-alliance.org/sites/default/files/eScience-Camden-final.pdf>.

Principle Six: Provide proper attribution and credit for research data.

Definition and Discussion of Terms

| [Acknowledgement](#) | [Attribution](#) | [Citation](#) | [Credit](#) | [Plagiarism](#) |

Guidelines for Implementation

Attribution and citation of a well-prepared, quality collection of data are an essential part of the research enterprise, but constitute mostly a scholarly rather than legal requirement, as explained below. In research communities worldwide, the right to attribution for one's contributions to research is considered a core value and obligatory practice that underpins the free exchange of knowledge, and provides an important means to trace the data and establish the correct provenance. Researchers often are paid for their ingenuity and efforts not by money (at least not directly) but, rather, in the currency of credit that enhances their reputation.

Despite some substantial progress in developing standards for data citation, however, there remain challenges to the practice of giving proper attribution and credit to the correct parties. Good practice is not widespread and has not been established for giving proper attribution and credit where the collection of data may be the product of activity involving many individual contributors, or where very large numbers of collections of data have been used to create a further data product. Providing attribution information in an external metadata record may be a means of acknowledging contributions, while conforming to the limits imposed by citation styles in scholarly communications.

Guideline 6A. Attribution of research data used in any scholarly output should be a normative convention established by good research policy and practice, and not by a legal mandate or a license requirement.

There has been a difference of opinion about whether to make attribution a specifically legal requirement or simply leave it to the norms of the research community. Despite its essential place within research communities, attribution for the use of data is not universally regulated by law, especially for information that is not copyrightable. European jurisdictions do protect attribution through various legal measures, including moral rights provisions in national copyright statutes. Copying with intentional lack of attribution or, in some cases, with copyright infringement, arises to the level of plagiarism (Harris date unspecified). It constitutes an ethical infringement punishable under community norms as well as possibly an infringement of copyright punishable under statutory law.

The one instance in which attribution has been a matter of broad legal compliance is for broad distribution of research publications under a Creative Commons (CC) Attribution 4.0 license (Creative Commons, 2015). This license requires attribution to the "author" (the collection of data producer) and the source (provider) in a manner specified by the rights holder (through the license), but its legal enforceability is uncertain. Although a failure to properly attribute CC licensed works results in termination of the license, exposing the user to the risk of legal consequences if the copyrightable portion is not attributed, the instances of such legal enforcement in practice—especially for research data—have been rare because of the expense and legal uncertainty.

The disparity in legal treatment of attribution across jurisdictions leaves the research data community to rely on normative values and conventions adopted by communities of practice to assure that the

producers and rights holders of collections of data receive the credit they expect and deserve. Norms of attribution in scientific publishing have been formalized to increase transparency and equality and they are promulgated in guidelines issued by well-established editorial and publishing groups such as COPE (the Committee on Publication Ethics), the Council of Scientific Editors, and the International Council of Medical Journal Editors, among others (Albert and Wager, 2003; Council of Scientific Editors 2012, 2016; International Council of Medical Journal Editors, date unspecified). The Digital Public Library of America's "Data Use Best Practices" is particularly instructive (DPLA, 2013). Applying these established guidelines to attribution practices for research data is a helpful starting point for the research community.

The ethical rules and practices vary from community to community and may often refer to particularities of a defined scientific domain. For all these reasons, we believe that attribution requirements (except actual copyright infringement) should be defined by the communities, not by lawmakers. Consistent with this view and the advice given by Creative Commons itself, we endorse the approach taken by the CC0 No Rights Reserved instrument, which suggests that the normative attribution to be given to the research data in question be consistent with the practice of the relevant research discipline (see Creative Commons, 2016 and Implementation Guideline 1B).

Beyond the norms and ethics inherited from the scientific publishing system, specific practices for attribution for the use of research data are being established and finding widespread support. Specifically, the development of persistent identifiers assigned to a collection of data, and the creation of the DataCite organization in 2009, has made it possible to formalize the concept of data citation (DataCite Metadata Working Group, 2016). Another notable effort has been CODATA's Data Citation Standards and Practices Task Group, whose analyses and reports (CODATA ICSTI, 2013; National Research Council, 2012a) spurred the development of the Force11 Joint Declaration of Data Citation Principles -- a widely endorsed statement covering the purpose, function and attributes of citations for data (Data Citation Synthesis Group, 2014). Other organizations have also been active in this area.

References

Albert, Tim and Wager, Elizabeth, 2003, "How to handle authorship disputes: a guide for new researchers", *COPE Report 2003*, p. 32-34. Available online: http://publicationethics.org/files/2003pdf12_0.pdf. Accessed September 5, 2016.

Association of Research Libraries, Center for Social Media-School of Communication at American University, and Program on Information Justice and Intellectual Property-Washington College of Law at American University (ARL et al.), 2012, *Code of Best Practices in Fair Use for Academic and Research Libraries*. Available online: http://www.cmsimpact.org/sites/default/files/documents/code_of_best_practices_in_fair_use_for_arl_final.pdf. Accessed September 5, 2016.

Ball, A. and Duke, M., 2015. "How to Cite Datasets and Link to Publications". *DCC How-to Guides* (web page). Edinburgh: Digital Curation Centre. Available online: <http://www.dcc.ac.uk/resources/how-guides/cite-datasets>. Accessed September 5, 2016.

Benkler, Yochai, 2006, *The Wealth of Networks—How Social Production Transforms Markets and Freedom*. Yale University Press, 515 pp. Available at: http://www.benkler.org/Benkler_Wealth_Of_Networks.pdf. Accessed September 5, 2016.

Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, 2003 (web page). Available at: <http://openaccess.mpg.de/Berlin-Declaration>. Accessed September 5, 2016.

Berne Convention for the Protection of Literary and Artistic Works of September 9, 1886, 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. Geneva: United International Bureaux for the Protection of Intellectual Property Available at: http://www.wipo.int/treaties/en/text.jsp?file_id=283698. Accessed September 5, 2016.

Bethesda Statement on Open Access Publishing, (Bethesda Statement), 2003 (web page), http://dash.harvard.edu/bitstream/handle/1/4725199/suber_bethesda.htm?sequence=1. Accessed September 5, 2016.

Borgman, Christine L. , 2015. *Big Data, Little Data, No Data, Scholarship in the Networked World*, Cambridge, MA: MIT Press, 416 pp.

Bouchout Declaration for Open Biodiversity Knowledge Management (Bouchout Declaration) 2014, (web page), Plazi. Available at: <http://bouchoutdeclaration.org/>. Accessed September 5, 2016.

Bromley, Allan. 1991, *Policy Statements on Data Management for Global Change Research*. Washington D.C.: U.S. Global Change Research Program, Office of Science and Technology Policy. Available online from UNT Digital Library. <http://digital.library.unt.edu/ark:/67531/metadc11862/>. Accessed September 5, 2016.

Budapest Open Access Initiative (BOAI), 2002, *Budapest Open Access Initiative* (web page), <http://www.budapestopenaccessinitiative.org/>. Accessed September 5, 2016.

Buranen, Lise and Myers Roy, Alice, Eds., 1999. *Perspectives on Plagiarism and Intellectual Property in a Postmodern World*. SUNY Press, 324 pp.

Cabinet Office, UK, 2013, *Open Data Charter* (2013). Available at: <https://www.gov.uk/government/publications/open-data-charter>. Accessed September 5, 2016.

The Committee on Data for Science and Technology (CODATA) Task Group on Preservation and Access of Scientific and Technical Data in Developing Countries (CODATA PASTD), 2014, *Data Sharing Principles in Developing Countries* (The Nairobi Data Sharing Principles), Available online: https://zenodo.org/record/22117/files/Data_Sharing_Principles_in_Developing_Countries.pdf. Accessed September 5, 2016.

CODATA-ICSTI Task Group on Data Citation Standards and Practice, (2013). "Out of Cite, Out of Mind: The Current State of Practice, Policy, and Technology for the Citation of Data". *Data Science Journal*. 12. CIDCR1–CIDCR7. <http://doi.org/10.2481/dsj.OSOM13-043>

Convention on Biological Diversity (CBD), 1992. (web page) Available at: <https://www.cbd.int/convention/>. Accessed September 5, 2016.

Convention on Biological Diversity, United Nations, 2014. Nagoya Protocol on the Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization, <https://www.cbd.int/abs/doc/protocol/nagoya-protocol-en.pdf>. Accessed September 5, 2016.

Council of Science Editors, 2012, *White Paper on Publication Ethics: CSE's White Paper on Promoting Integrity in Scientific Journal Publications, 2012 Update*, (web page), <http://www.councilscienceeditors.org/resource-library/editorial-policies/white-paper-on-publication-ethics/>. Accessed September 5, 2016.

Council of Scientific Editors, “Scientific Style and Format Citation Quick Guide” (web page), <http://www.scientificstyleandformat.org/Tools/SSF-Citation-Quick-Guide.html>. Accessed September 5, 2016.

Creative Commons, 2015, “Attribution 4.0 International License” (web page), <http://creativecommons.org/licenses/by/4.0/legalcode#s3a>. Accessed September 5, 2016.

Creative Commons. 2016. *CCO FAQ* (web page). https://wiki.creativecommons.org/wiki/CCO_FAQ. Accessed September 6, 2016.

DataCite Metadata Working Group, 2016, *DataCite Metadata Schema for the Publication and Citation of Research Data*. Available online: <http://doi.org/10.5438/0012>. Accessed September 5, 2016.

Data Sharing Working Group (DSWG), 2014a, *White Paper: Mechanisms to Share Data as Part of the GEOSS Data-CORE, Group on Earth Observations*. Available at: https://www.earthobservations.org/documents/dswg/Annex%20VI%20-%20Mechanisms%20to%20share%20data%20as%20part%20of%20GEOSS%20Data_CORE.pdf.

Data Sharing Working Group, 2014b, *GEOSS Data Sharing Principles post-2015, Group on Earth Observations*. Available at: http://www.earthobservations.org/documents/dswg/10_GEOSS%20Data%20Sharing%20Principles%20post%202015.pdf.

Data Citation Synthesis Group, 2014. *Joint Declaration of Data Citation Principles* (web page), Martone M., Ed. San Diego CA: FORCE11. Available online: <https://www.force11.org/group/joint-declaration-data-citation-principles-final>. Accessed September 5, 2016.

The Denton Declaration: An Open Data Manifesto, 2012 (web page), University of North Texas Open Access Conference. Available at: <https://openaccess.unt.edu/denton-declaration>. Accessed September 5, 2016.

Digital Curation Center and Research Data Alliance (DCC and RDA), Metadata Standards Directory Working Group, 2015, “Metadata Directory” (web page), <http://rd-alliance.github.io/metadata-directory/>. Accessed September 5, 2016.

Digital Public Library of America, 2013, *DPLA Data Use Best Practices* (web page) <http://dp.la/info/wp-content/uploads/2013/04/DPLADataUseBestPractices.pdf>. Accessed September 5, 2016.

Doldirina, Catherine, Eisenstadt, Anita, Onsrud, Harlan, and Uhlir, Paul. F., (pending publication), “Legal Approaches for Open Access to Research Data”.

Ember, Carol and Hanisch, Robert, 2013, *Sustaining Domain Repositories for Digital Data: A White Paper* Prepared from a workshop organized by the Inter-university Consortium for Political and Social Research. Available at: http://datacommunity.icpsr.umich.edu/sites/default/files/WhitePaper_ICPSR_SDRDD_121113.pdf. Accessed September 5, 2016.

European Commission, 2016, *Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020*, Version 3.1. Available online: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf. Accessed September 5, 2016.

European Parliament, 1996, "Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases". *Official Journal of the European Communities*, No L 77/20, March 27, 1996 Available online: <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:31996L0009&from=EN>. Accessed September 5, 2016.

Europeana-DPLA International Rights Statements Working Group, 2016, *Recommendations for Standardized International Rights Statements*. Available online: http://rightsstatements.org/files/151002recommendations_for_standardized_international_rights_statements.pdf. Accessed September 5, 2016.

Fisk, Catherine L., 2006, "Credit Where It's Due: The Law and Norms of Attribution," *Georgetown Law Journal*, vol. 95, p. 49-117. Available online: http://scholarship.law.duke.edu/cgi/viewcontent.cgi?article=2094&context=faculty_scholarship. Accessed October 19, 2016.

Guibault, Lucie and Wiebe, Andreas (Eds.), 2013, *Safe to be open. Study on the protection of research data and recommendations for access and usage*. OpenAIRE, (Universitätsverlag Göttingen). Available at: <http://www.ivir.nl/publicaties/download/1129>. Accessed September 5, 2016.

Harris, Lesley Ellen, "What to Do If You're Accused of Copyright Infringement", date unspecified, World Intellectual Property Organization (WIPO) (web page). http://www.wipo.int/sme/en/documents/copyright_infringement_fulltext.html. Accessed September 5, 2016.

Health and Human Services, Office of Research Integrity, 2000, *Federal Research Misconduct Policy*, (web page) <http://ori.hhs.gov/federal-research-misconduct-policy>. Accessed September 6, 2016.

Hey, Tony, Tansley, Stewart, and Tolle, Kristin, Eds., 2009. *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research, 284 pp.

Human Genome Organization, *Bermuda Principles, 1996, Summary of Principles Agreed at the First International Strategy Meeting on Human Genome Sequencing*. Available at: http://www.casimir.org.uk/storyfiles/64.0.summary_of_bermuda_principles.pdf.

International Committee of Medical Journal Editors (ICMJE), "Scientific Misconduct, Expressions of Concern, and Retraction" (web page) <http://www.icmje.org/recommendations/browse/publishing-and-editorial-issues/scientific-misconduct-expressions-of-concern-and-retraction.html>.

International Committee of Medical Journal Editors (ICMJE), "Defining the Role of Authors and Contributors" (web page), <http://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html>. Accessed September 5, 2016

IUCN (The World Conservation Union), 1973. *Convention on International Trade in Endangered Species of Wild Fauna and Flora*.. Available at: <https://cites.org/eng/disc/text.php>. Accessed September 5, 2016.

Kubicek, Herbert, Cimander, Ralf, and Scholl, Hans Jochen, 2011. "Layers of Interoperability", in *Organizational Interoperability in E-Government: Lessons from 77 European Good-Practice Cases*. Heidelberg: Springer-Verlag, p. 85-96.

Lastowka, Greg, 2007, "Digital Attribution: Copyright and the Right to Credit", *Boston University Law Review*, vol. 87, p. 41-59. Available online: <https://www.bu.edu/law/central/jd/organizations/journals/bulr/volume87n1/documents/LASTOWKAv.2.pdf>. Accessed September 5, 2016.

Lead, David Fewer, Lauriault, Tracey P. , and Scassa, Teresa, Eds., 2016, *A Proposal: An Open Licensing Scheme for Traditional Knowledge*, Carleton University Geomatics and Cartographic Research Centre (GCRC) and the Canadian Internet Policy and Public Interest Clinic, https://cippic.ca/sites/default/files/file/CIPPIC_GCRC--TK_License_Proposal--July_2016.pdf. Accessed September 5, 2016.

Lerman, Celia, Unwaivable Moral Rights and the Need for New International Substantive Minima (Creative Commons, February 20, 2012). Available online at SSRN: <http://dx.doi.org/10.2139/ssrn.2159697>. Accessed September 5, 2016.

LIBER, Association of European Research Libraries, 2015, “The Hague Declaration on Knowledge Discovery in the Digital Age” (web page). Available online: <http://thehaguedeclaration.com/the-hague-declaration-on-knowledge-discovery-in-the-digital-age/>. Accessed September 5, 2016.

National Information Standards Organization (NISO), Open Access Metadata & Indicators Working Group, 2014, “Access and License Indicators” (web page), <http://www.niso.org/workrooms/ali/>. Accessed September 5, 2016.

National Research Council (NRC), 1997, *Bits of Power: Issues in Global Access to Scientific Data*. Washington DC: National Academy Press. Available online at: <http://www.nap.edu/catalog/5504/bits-of-power-issues-in-global-access-to-scientific-data>. Accessed September 5, 2016.

National Research Council (NRC), 1999, *A Question of Balance: Private Rights and the Public Interest in Scientific and Technical Databases*. Washington DC: National Academy Press. Available online at: <http://www.nap.edu/catalog/9692/a-question-of-balance-private-rights-and-the-public-interest>. Accessed September 5, 2016.

National Research Council (NRC), 2003, *The Role of Scientific and Technical Data and Information in the Public Domain*. Washington DC: National Academies Press. Available at: <http://www.nap.edu/catalog/10785/the-role-of-scientific-and-technical-data-and-information-in-the-public-domain>. Accessed September 5, 2016.

National Research Council (NRC), 2004, *Open Access and the Public Domain in Digital Data and Information for Science*. Washington DC: National Academies Press. Available at: <http://www.nap.edu/catalog/11030/open-access-and-the-public-domain-in-digital-data-and-information-for-science>. Accessed September 5, 2016.

National Research Council (NRC), 2012a, *For Attribution—Developing Data Attribution and Citation Practices and Standards*, Washington DC: National Academies Press. Available at: <http://www.nap.edu/catalog/13564/for-attribution-developing-data-attribution-and-citation-practices-and-standards>. Accessed September 5, 2016.

National Research Council (NRC), 2012b, *The Case for International Sharing of Scientific Data: A Focus on Developing Countries*. Kathie Bailey Matthae and Paul F. Uhler, eds. Washington DC: National Academies Press. Available at: <http://www.nap.edu/catalog/17019/the-case-for-international-sharing-of-scientific-data-a-focus>. Accessed September 5, 2016.

NautaDutilh, 2001, *The Implementation and Application of Directive 96/9/EC on the Legal Protection of Databases*, report written under contract to the European Commission. Available at: http://ec.europa.eu/internal_market/copyright/docs/databases/etd2001b53001e72_en.pdf. Accessed September 5, 2016.

- “Declarations in support of OA,” Open Access Directory (web page), date unspecified. Available at: [http://oad.simmons.edu/oadwiki/Declarations in support of OA](http://oad.simmons.edu/oadwiki/Declarations_in_support_of_OA). Accessed September 5, 2016
- Organisation for Economic Co-operation and Development (OECD), 2007, *OECD Principles and Guidelines for Access to Research Data from Public Funding*. Available online: <http://www.oecd.org/sti/sci-tech/38500813.pdf>. Accessed September 5, 2016.
- Rasmusen, Eric, and Posner, Richard A., 2000 "Creating and Enforcing Norms, with Special Reference to Sanctions, University of Chicago Law School, John M. Olin Program in Law and Economics *Working Paper No. 96*. Available online: http://chicagounbound.uchicago.edu/cgi/viewcontent.cgi?article=1146&context=law_and_economics. Accessed September 5, 2016.
- Reichman, Jerome H. and Okediji, Ruth, 2012. “When Copyright Law and Science Collide: Empowering Digitally Integrated Research Methods on a Global Scale”, *Minnesota Law Review* 96: 1362-1480. Available online at: http://www.minnesotalawreview.org/wp-content/uploads/2012/08/ReichmanOkediji_MLR1362.pdf. Accessed October 19, 2016.
- Reichman, Jerome H., and Uhler, Paul F., 1999, “Database Protection at the Crossroads: Recent Developments and Their Impact on Science and Technology”, *Berkeley Technology Law Journal*, 793-838. Available online: http://scholarship.law.duke.edu/cgi/viewcontent.cgi?article=2230&context=faculty_scholarship. Accessed October 19, 2016.
- Royal Society, 2012. *Final Report: Science as an open enterprise*. Available online at: <https://royalsociety.org/~media/policy/projects/sape/2012-06-20-saoe.pdf> Accessed September 5, 2016.
- Science International, 2015, *Open Data in a Big Data World: An international accord* (Extended version). Available online at: <http://www.icsu.org/science-international/accord/open-data-in-a-big-data-world-long>. Accessed September 5, 2016.
- Stiglitz, Joseph E., 1999, “Knowledge as a Global Public Good” p. 308-325 in *Global Public Goods: International Cooperation in the 21st Century*, Inge Kaul, Isabelle Grunberg, and Marc Stern (Eds.) Available at: <http://web.undp.org/globalpublicgoods/TheBook/globalpublicgoods.pdf> . Accessed September 5, 2016.
- Tsoukala, Victoria, Angelaki, Marina, Kalaitzi, Vasso, et al., 2015, *Policy Guidelines for Open Access and Data Dissemination and Preservation*. European Commission. RECODE Project 2013-2015. Available at: <http://recodeproject.eu/wp-content/uploads/2015/02/RECODE-D5.1-POLICY-RECOMMENDATIONS-FINAL.pdf>. Accessed September 5, 2016.
- Uhler, Paul F., 2015. *The Value of Open Data Sharing, A CODATA Report for the Group on Earth Observations*. <http://doi.org/10.5281/zenodo.33830>. Accessed October 19, 2016.
- Uhler, Paul F, and Schröder, Peter, 2007. “Open Data for Global Science”. *Data Science Journal* 7. <http://doi.org/10.2481/dsj.6.OD36>. Accessed October 19, 2016.
- UNESCO, 1989. *Recommendation on the Safeguarding of Traditional Culture and Folklore* (web page) Available at: http://portal.unesco.org/en/ev.php-URL_ID=13141&URL_DO=DO_TOPIC&URL_SECTION=201.html. Accessed September 6, 2016.
-

UNESCO, 1970. *Convention on the Means of Prohibiting and Preventing the Illicit Import, Export and Transfer of Ownership of Cultural Property*. Available at: http://portal.unesco.org/en/ev.php-URL_ID=13039&URL_DO=DO_TOPIC&URL_SECTION=201.html. Accessed September 5, 2016.

UNESCO, 1972, *Convention on the Protection of World Cultural and Natural Heritage*. Available at: http://portal.unesco.org/en/ev.php-URL_ID=13055&URL_DO=DO_TOPIC&URL_SECTION=201.html. Accessed September 5, 2016.

UNESCO, 2005. *Convention of the Protection and Promotion of the Diversity of Cultural Expressions*. Available at: http://portal.unesco.org/en/ev.php-URL_ID=13055&URL_DO=DO_TOPIC&URL_SECTION=201.html. Accessed September 5, 2016.

Copyright Law of the United States, United States Code, Title 17. Available at: <http://www.copyright.gov/title17/circ92.pdf>. Accessed September 5, 2016.

Copyright Law of the United States, United States Code, Title 17, Section 106A. “Rights of certain authors to attribution and integrity.” Available online at: <http://www.copyright.gov/title17/92chap1.html#106a>. Accessed September 5, 2016.

Copyright Law of the United States, United States Code, Title 17, Section 501, “Infringement of Copyright.” Available at: <http://www.copyright.gov/title17/92chap5.html#501>. Accessed September 5, 2016.

World Intellectual Property Organization (WIPO), “Traditional Knowledge,” *WIPO Glossary* (date unspecified). Available online at: www.wipo.int/tk/en/resources/glossary.html. Accessed September 4, 2016.

WIPO Committee on Development and Intellectual Property, Document CDIP/13/INF/6, April 14, 2014 “Comparative Analysis of National Approaches on Voluntary Copyright Relinquishment,” prepared by Dr. Andres Guadamuz, University of Sussex, United Kingdom. Available online: http://www.wipo.int/edocs/mdocs/mdocs/en/cdip_13/cdip_13_inf_6.pdf Accessed September 5, 2016.

World Wide Web Consortium, 2015, “Open Digital Rights Language--ODRL” (web page), <http://www.w3.org/TR/odrl/>. Accessed September 6, 2016.

PART III. Definitions of Terms

<p>Acknowledgement</p>	<p>Formal recognition of contributions to a research output (e.g., intellectual, advisory, financial, facilities, resources, technical) by contributors who do not meet the criteria for authorship.</p> <p>[Principle Six: Attribution & Credit]</p>
<p>Attribution</p>	<p>The practice of indicating the source of a work, of a contribution to a work, or of an idea to one or more creators or authors. In academic and research communications, attribution may take the form of acknowledgement, citation, or authorship. (Fisk, 2006)</p> <p>[Principle Six: Attribution & Credit]</p>
<p>Certainty</p>	<p>Certainty is based on the completeness and accuracy of the information given about the rights in a collection of data. Such information reduces ambiguities and the exposure to legal risks for the user of those data.</p> <p>[Principle Four: Transparency & Certainty]</p>
<p>Citation</p>	<p>A reference in an academic or research communication that documents any sources used in a research output, for the two-fold purpose of: (a) giving credit to existing sources of ideas, data, and information, and (b) enabling others to identify and locate those sources used in the research. For more information about data citation, see Ball and Duke, 2015.</p> <p>[Principle Six: Attribution & Credit]</p>
<p>Collection of data</p>	<p>A collection of data (also referred to as a “dataset”) means a “collected, selected, coordinated, or arranged set of data elements in electronic form consisting often of observed, discovered, or derived values”. (Data Sharing Working Group, 2014a)</p> <p>[Principle One: Access & Reuse]</p>
<p>Confidentiality laws</p>	<p>Confidentiality laws can protect public and private interests by withholding certain data and information from the public. This may refer to administrative consultations in governments or to secrets in the private sector, such as technical know-how, customer information, trade secrets, non-competition, and the protection of other commercial interests. Such presumptive or real interests often form the background to information policies or agreements with employees by institutions and enterprises that try to withhold data and information they produce, or control the use of them.</p> <p>[Principle Three: Balance Legal Interests]</p>

<p>Copyright</p>	<p>The World Intellectual Property Organization defines copyright (or the author’s right) as “a legal term used to describe the rights that creators have over their literary and artistic works. Works covered by copyright range from books, music, paintings, sculpture, and films, to computer programs, databases, advertisements, maps, and technical drawings. There are two types of rights under copyright: economic rights, which allow the rights owner to derive financial reward from the use of his works by others; and moral rights, which protect the non-economic interests of the author.” (http://www.wipo.int/copyright/en/). Its most important legal basis at the international level is the Berne Convention for the Protection of Literary and Artistic Works (http://www.wipo.int/treaties/en/ip/berne/), although copyright is implemented through individual national legislation that is consistent with the treaty.</p> <p>Copyright applies to individual works, but not to facts, ideas, or concepts. It refers not to the content of a work, but to the form of presentation (the “expression”) of this content.</p> <p>[Principle One: Access & Reuse]</p>
<p>Copyright infringement</p>	<p>Infringement of copyright is a violation of any of the exclusive rights of the copyright owner, as provided by legislation. (See, e.g., the copyright infringement section of the 1976 U.S. Copyright Act at: http://www.copyright.gov/title17/92chap5.html, and the “What to Do If You’re Accused of Copyright Infringement” section of the UN’s World Intellectual Property Organization (WIPO) web site (Harris, date unspecified)</p> <p>[Principle Three: Balance Legal Interests]</p>
<p>Credit</p>	<p>Formal recognition of the contribution made by an individual or group to a research output. (Lastowka, 2007)</p> <p>[Principle Six: Attribution & Credit]</p>
<p>Cultural resources protection</p>	<p>Secrecy may apply to data and information about cultural resources. The open accessibility of such data may under certain circumstances provoke the illegal exploitation of cultural resources and put those resources at risk of theft or destruction. Open access to research data could, under certain circumstances, compromise the protection of such resources. See, e.g., the related Conventions referenced in UNESCO, 1970; UNESCO, 1972; UNESCO, 2005.</p> <p>[Principle Three: Balance Legal Interests]</p>
<p>Data</p>	<p>In accordance with the definition of the term “research data” in the CASRAI Research Data Domain dictionary (http://dictionary.casrai.org/Data) the term refers to any record - independent of which medium and format - containing a record of observations of the real world. This categorization does not deny that “observations of reality” can also be considered as subjective constructions. Data</p>

	<p>can be a single datum in a compilation of data. As a legal matter, data may be eligible for intellectual property protection, but the application of any law is subject to a case-by-case determination. Further details of the definition of data in the CASRAI Research Data Domain dictionary are not important in the context of legal interoperability.</p> <p>[Principle One: Access & Reuse]</p>
Dataset	<p>See "Collection of data"</p>
Database	<p>Here, the meaning of the term "database" is limited to its legal usage. In the legal sense, a database is an organized compilation of data that enjoys applicable statutory protection if it meets certain criteria that are stated in the statutes. Whether the database was created in the course of doing any form of research is not important in this context.</p> <p>[Principle One: Access & Reuse]</p>
Endangered species protection	<p>Specific data and information referring to endangered species must, in certain circumstances, be withheld in the interest of their protection. Open access to data on the precise location of an endangered species, for example, can contribute to the misuse or extinction of the organism. Open access to relevant data could therefore jeopardize its protection and even existence. See, e.g., IUCN 1973.</p> <p>[Principle Three: Balance Legal Interests]</p>
Equity	<p>Equity may be defined as "the quality of being fair and impartial." <i>Oxford English Dictionary</i>.</p> <p>[Principle One: Access & Reuse]</p>
Intellectual property	<p>As defined by the World Intellectual Property Organization, "[i]ntellectual property (IP) refers to creations of the mind, such as inventions; literary and artistic works; designs; and symbols, names and images used in commerce. IP is protected in law by, for example, patents, copyright and trademarks, which enable people to earn recognition or financial benefit from what they invent or create. By striking the right balance between the interests of innovators and the wider public interest, the IP system aims to foster an environment in which creativity and innovation can flourish. (http://www.wipo.int/about-ip/en/)</p> <p>[Principle Three: Balance Legal Interests]</p>
Individual contracts or end-user license agreements	<p>Through the use of individual contracts, often referred to as end-user license agreements (EULAs), contractors may be able to legally impose restrictions to the access and reuse of research data and information that go far beyond legally justified interests. Not all jurisdictions allow such overrides, however. Contracts are, by default, only binding to the parties who have signed them because they imply an equal bargain and knowing consent.</p>

	<p>[Principle Three: Balance Legal Interests]</p>
Interoperability of data	<p>“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.” (Doldirina et al., pending publication; Kubicek et al., 2011)</p> <p>Interoperability is an attribute that greatly facilitates usability of research data.</p> <p>[Principle One: Access & Reuse]</p>
Legal interoperability	<p>“Legal interoperability occurs among two or more datasets when: the legal use conditions are clearly and readily determinable for each of the datasets, typically through automated means; the legal use conditions imposed on each dataset allow creation and use of combined or derivative products; and users may legally access and use each dataset without seeking authorization from data rights holders on a case-by-case basis, assuming that the accumulated conditions of use for each and all of the datasets are met.</p> <p>When data are combined from multiple sources the resulting dataset will incorporate the accumulated restrictions imposed by each source.... 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....” (derived from Doldirina et al., pending publication).</p> <p>[Principle One: Access & Reuse]</p>
Metadata	<p>Metadata are data about data. Metadata provide structured descriptions of data collections and data services that facilitate their discovery, assessment, inventory, and use. The focus of the metadata discussed in these Implementation Guidelines concerns the legal status of the dataset being described (NISO 2014). For more general guidance on metadata, see, e.g., DCC and RDA Metadata Standards Directory WG, 2015.</p> <p>[Principle Four: Metadata]</p>

<p>Moral rights</p>	<p>“The Berne Convention (Article 6bis) requires Member countries to grant to authors: (i) the right to claim authorship of the work, (sometimes called the right of paternity); and (ii) the right to object to any distortion or modification of the work, or other derogatory action in relation to the work, which would be prejudicial to the author’s honor or reputation, (sometimes called the right of integrity). These rights are generally known as the moral rights of authors. The Convention requires them to be independent of the author’s economic rights, and to remain with the author even after he has transferred his economic rights. It is worth noting that moral rights are only accorded to individual authors. Thus even when, for example, a film producer or a publisher owns the economic rights in a work, it is only the individual creator who has moral interests at stake.” http://www.wipo.int/edocs/pubdocs/en/intproperty/909/wipo_pub_909.pdf</p> <p>[Principle One: Access & Reuse]</p>
<p>National security and public safety laws</p>	<p>National security and public safety (or public order) laws can declare certain data as secret or confidential for a specified period of time. That means that such datasets are only held within certain sectors of the government by those with a need to know. It is forbidden to make use of the protected data and to disseminate them or the information based on these data to the public. The main applications of such restrictions can be found in the field of military defense, intelligence agencies, or ministries and organizations that protect the public safety.</p> <p>[Principle Three: Balance Legal Interests]</p>
<p>Norms</p>	<p>Norms are behaviors in which communities of practice engage out of a sense of a peer obligation or as a means to articulate and demonstrate a consensus position about a particular area of practice. Norms are not legally binding private or public rules, or enforceable as a legal rule. Nonetheless, departure from norms could result, at a minimum, in disputes over what is “best” or “reasonable” practice for a given community. At worst, a breach in norms could lead to informal social sanctions, such as informal ostracism, professional censure, or even more formal exclusion of the community group governed by them (Rasmusen and Posner, 2000).</p> <p>[Principle Five: Balance Legal Interests]</p>
<p>Open data</p>	<p>“Open data is data that meets the criteria of intelligent openness. Data must be accessible, useable, assessable and intelligible.”</p> <ul style="list-style-type: none"> • “Accessible” – “Data must be located in such a manner that it can readily be found and in a form that can be used.” • “Useable” – “In a format where others can use the data or information. Data should be able to be reused, often for different purposes, and therefore will require proper background information and metadata. The usability of data will also depend on those who wish to use them.”

	<ul style="list-style-type: none"> • “Assessable” – “In a state in which judgments can be made as to the data or information’s reliability. Data must provide an account of the results of scientific work that is intelligible to those wishing to understand or scrutinize them. Data must therefore be differentiated for different audiences.” • “Intelligible” – “Comprehensive for those who wish to scrutinize something. Audiences need to be able to make some judgment or assessment of what is communicated. They will need to judge the nature of the claims made. They should be able to judge the competence and reliability of those making the claims. Assessability also includes the disclosure of attendant factors that might influence public trust.” (Royal Society 2012) <p>[Principle One: Access & Reuse]</p>
Patents	<p>The World Intellectual Property Organization defines patent as “an exclusive right granted for an invention, which is a product of a process that provides a new way of doing something, or offers a new technical solution to a problem.... A patent provides protection for the invention to the owner of a patent. The protection is granted for a limited period, generally 20 years.” (http://www.wipo.int/edocs/pubdocs/en/patents/450/wipo_pub_l450pa.pdf)</p> <p>Patents are granted by a national (or regional) authority as the right to monopolize the commercialization of an invention, but they do not prohibit the exchange or distribution of knowledge on which the invention is based. Patents therefore should not hamper the access to research data, although they may impede certain commercial reuses of these data for a given time period.</p> <p>[Principle Three: Balance Legal Interests]</p>
Periods of exclusive use of research data	<p>Many research funding institutions, universities and research institutes, and academic publishers impose time embargos for the access to and the reuse of scientific data and information that they have produced. These restrictions can be based on either a formal contractual basis or a less formal policy one, and are expressions of a self-defined institutional policy ideally incorporating on the norms of a defined research community.</p> <p>[Principle Three: Balance Legal Interests]</p>
Personal privacy protection	<p>In most countries, data referring to individual persons (e.g., medical information, data on social status, penal record, or financial information, among others) are treated differently from all other data and restricted. Such data may affect privacy insofar as they can contain information about a defined person that the person may want to keep private. In order to solve this conflict of interest, many jurisdictions have enacted personal data protection laws. As a common element, they state that data referring to an individual person may only be released to the</p>

	<p>public if certain conditions are fulfilled (e.g., agreement by the concerned person, legal authorization, or public interest).</p> <p>[Principle Three: Balance Legal Interests]</p>
Plagiarism	<p>As defined by the U.S. Department of Health and Human Services (HHS), “Plagiarism is the appropriation of another person's ideas, processes, results, or words without giving appropriate credit.” (HHS, Office of Research Integrity, 2000). For more information about plagiarism, see: Buranen and Myers Roy, 1999.</p> <p>[Principle Six: Attribution & Credit]</p>
Public domain	<p>Research datasets are in the public domain if any of these conditions is met:</p> <ol style="list-style-type: none"> 1) The term of intellectual property protection - if applicable - have expired. (For example, under international agreement in copyright law, this is the life of the author, plus a minimum of 50 years. In the national legislation of many countries (including all OECD countries, except Canada), the term has been extended to life of the author, plus 70 years. A special rule applies in the United States in the case of "works made for hire" [that is, through employment], the copyright term is 95 years from the date of first publication or 120 years from the date of its creation, whichever is earliest. Under the database protection legislation, the original term of exclusive property protection is 15 years, but is renewable for 15 years with any substantial investment in the maintenance of the database.) 2) The subject matter is not protectable to begin with. (In copyright law, for example, unprotectable subject matter includes facts of nature, which make up much of the content of many datasets.) 3) The rights holder of the dataset waives all rights (copyright and other intellectual property rights, but not an author’s moral rights—see Principle One Implementation Guideline 1C) in it in perpetuity. 4) The exclusion of the type of information from copyright and other intellectual property rights by a government, either through a treaty, an executive agreement, a policy requirement, or by legislation. <p>[Principle One: Access & Reuse]</p>
Research Data	<p>The CASRAI Research Data Domain dictionary defines research data as “Data that are used as primary sources to support technical or scientific enquiry, research, scholarship, or artistic activity, 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. Research data may be experimental data, observational data, operational data, third party data, public sector data, monitoring data, processed data, or repurposed data.”</p> <p>http://dictionary.casrai.org/Research_data.</p> <p>Generally speaking, the qualification of data as research data has no implications for intellectual property rights. This does not preclude conditions limiting the</p>

	<p>usage of certain data to research purposes. Also, the fact that data are to be used for research purposes may legitimize their production or collection.</p> <p>[Principle One: Access & Reuse / Principle Four: Metadata]</p>
Rights holder	<p>Techopedia defines ‘Rights holder’ as “a legal entity or person with exclusive rights to a protected copyright, trademark or patent, and the related rights of producers, performers, producers and broadcasters. A right holder may license a portion or all of a protected work through international legal and licensing provisions.... Right holders control the use of their exclusive rights, including reproduction and distribution. However, certain limitations and exceptions related to the interest of public policy override these rights.... If a protected work - such as a digital video, audio or electronic research content - is used without permission, the right holder may pursue legal infringement action against the unauthorized user.”</p> <p>(https://www.techopedia.com/definition/24896/right-holder)</p> <p>[Principle One: Access & Reuse / Principle Four: Metadata]</p>
Rights statement	<p>A rights statement is the assertion about the copyright and other legal status of the dataset in question, or the means of expression of the data, whether conveyed through a written text or other digital object. It is not a legal document per se, but a categorization that describes a diverse set of rights attached to the licenses, norms, agreements, or public domain status of the content, of the means of expression, or of the data itself. (Europeana-DPLA International Rights Statements Working Group, 2016).</p> <p>[Principle Four: Metadata]</p>
Sovereign genetic resources protection	<p>National authorities have the right to legislate on the access to a nation’s genetic resources, based on the legal requirements of the Convention of Biological Diversity of 1992 and its Nagoya Protocol of 2010. This legislation should create conditions - via prior informed consent (PIC), mutually agreed terms (MAT), and access and benefit sharing (ABS) - that facilitate access for environmentally sound uses and not impose restrictions that run counter to the conservation and sustainable use of biodiversity. The laws apply equally to the physical genetic resources and to the genetic data about them. (Convention on Biological Diversity, 1992, 2014).</p> <p>[Principle Three: Balance Legal Interests]</p>
<i>Sui generis</i> database rights	<p>Database protection in a legal form of <i>sui generis</i> (“of its own kind,” or unique) rights exists mostly in the European Union (with a few similar applications in other countries) (European Parliament, 1996). It applies to databases that show an investment in the verification and presentation of the contents. Database protection refers to the entire or “substantial part” of a database, not to the single datum or “insubstantial” part of a database. It prevents unauthorized persons from extracting and reusing substantial parts of the protected database, or even</p>

	<p>repeated extractions of insubstantial amounts of data. In most non-E.U. countries, databases are only protected if they (or certain portions or characteristics) qualify as “works” within the meaning of copyright. For an analysis of the effects of the E.U. Database Protection legislation, see, e.g., NautaDutilh, 2001; and on research, Guibault and Wiebe, 2013; Reichman and Uhlir, 1999.</p> <p>[Principle One: Access & Reuse / Principle Three: Balance Legal Interests]</p>
<p>Traditional Knowledge protection</p>	<p>The World Intellectual Property Organization in its <i>Glossary</i> (http://www.wipo.int/tk/en/resources/glossary.html) defines Traditional Knowledge (TK) as “the intellectual and intangible cultural heritage, practices and knowledge systems of traditional communities, including indigenous and local communities.” The laws in some countries prohibit the misappropriation of various data and information that have cultural significance and protect indigenous peoples’ rights. Examples may include various designs and traditional medicines. Some cultures also prohibit taking pictures of designated persons and disseminating them. (See, e.g., UNESCO, 1989.) However, many countries do not protect such rights at all. In those cases, “soft” law, such as governmental policies, and ethics can be used to protect the holders of Traditional Knowledge and help ensure that indigenous people’s rights are respected (Lead et al., 2016)</p> <p>[Principle Three: Balance Legal Interests]</p>
<p>Transparency</p>	<p>Provision of all pertinent information concerning the holder of rights and the status of the rights, if any, in a collection of data to the extent that is feasible, provided with reasonable effort and cost by the person or organization making the data available.</p> <p>[Principle Four: Metadata]</p>