RDA-CODATA Interest Group on the Legal Interoperability of Research Data

**Implementation Guidelines**

**for the Principles on the Legal Interoperability of Research Data**

March 28, 2016

 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 “Guidelines for Implementation”. The authoring Interest Group[[1]](#footnote-1) also has provided cross-references to other germane Principles and to key supporting documents in a Bibliography.

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

Definition and Discussion of Terms

| [Interoperability of data](#InteropDataDef) | Legal interoperability of data | Public domain | [Open Access](#OpenAccessDef) | | Equity |

**Guidelines for Implementation**

 It is widely recognized that much of the value of public research data lies in their broad dissemination and reuse, particularly on digital networks. There are many justifications for this, whether to improve research and education outcomes, enhance economic returns, promote social welfare goals, or support good governance. Public research data have public good characteristics, and are often global public goods. There is a body of literature that documents and supports these conclusions (Uhlir 2015).

The ability to access and reuse data is impaired, however, when there is an absence of clarity about the legal conditions under which the data may be reused and when restrictions are placed on the reuse of datasets. In most circumstances, legal restrictions on reuse run counter to the obligation to make research data publicly available.

Restrictions can inhibit reuse to a greater extent than is sometimes realized. This can be illustrated by analogy to the idea of a ”lowest common denominator”.  When considering the legal restrictions on reuse of datasets, 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 the underlying datasets will be transferred to the entire derivative dataset. 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.

In order to enable the maximum degree of interoperability, and to promote access and reuse (including data sharing), the following guidelines for implementation can be used by an original producer or subsequent rights holder 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. The designation of the research dataset in the public domain by the rights holder(s) is the best way to achieve legal interoperability.

From a legal perspective, placing data or information in the public domain means that there are no restrictions on their reuse. Research data and other types of information are in the public domain if:

1) The term of copyright protection has expired. By international agreement, this is the life of the author, plus a minimum of 50 years. In the national legislation of many countries (including all OEDC countries), 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.

2) The material is not copyrightable to begin with, such as a fact of nature, which makes up much of the content of many datasets.

3) The rights holder of the dataset waives all rights (copyright and other intellectual property rights) in it in perpetuity.

4) The exclusion of the type of information from copyright and other intellectual property rights (especially “neighboring rights”) by a government, either through a treaty, an executive agreement, a policy requirement, or by legislation.

**[ ] Guideline 1B. There is an array of legal instruments that can be used by governments, institutions, or researchers to place research datasets in the public domain.**

Table 1 summarizes what governments can do from the top down and what nongovernmental individual researchers and their institutions can do from the bottom up to promote open data and their legal interoperability.

**Table 1: Summary of Legal Mechanisms that Promote Open Data and Legal Interoperability[[2]](#footnote-2)**

|  |  |
| --- | --- |
| **Type of Legal Mechanism** | **Summary Description** |
| ***A. 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 copyright protection, place research data arising from joint research publicly in the public domain or place data collected in a particular location (i.e., Antarctica) in the public domain. |
| Intergovernmental Policies | 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 can include requirements in grants or contracts to make resulting research data publicly available without restrictions on use or reuse. |
| ***B. Nongovernmental Mechanisms*** |
| Creative Commons Public Domain Mark (http://creativecommons.org/choose/mark/) | The CC Public Domain Mark is used to mark datasets over which copyright has expired, and thus are already in the public domain, enabling their more ready identification in global web searches. Except for data arising in the public domain (e.g., facts) and datasets in historic documents that have been digitized, few datasets 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 EU database right. |
| 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.” |
| Nongovernmental Policies and Agreements | Nongovernmental organizations can implement policies to retain no rights in research data or implement agreements among institutional members of a consortium to retain no rights in research data. |

 **[ ] Guideline 1C. The use of non-restrictive (“common use”) licenses, with only “some rights reserved” is not recommended for use by the originator(s) or the rights holder(s) of the research data when disseminating the data for broad use.**

Many public and private research organizations and researchers now use non-restrictive licenses for both research datasets and more copyrightable information, such as research articles, books, pictures, software, and other types of information products. 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 legally validated in 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 online.

Creative Commons licenses, as well as other common-use licenses, are intended for use only with creative works, because they rely on copyright or a neighboring intellectual property right for enforcement. Because many datasets have few or no creative elements we do not recommend the use of the Creative Commons Attribution 4.0 International License, or more restrictive common-use licenses for general use with datasets even if the data content, its selection, or its arrangement includes some minimal creativity. Even minimal restrictions, such as an attribution requirement, raise legal interoperability challenges due to accumulating restrictions caused by the common day-to-day extraction of data from multiple databases and recombination for scientific purposes. Not knowing or resolving the legal status of the resulting data compilations results in substantial impediments to the sharing and use of research data. Tracking the legal status of millions of datasets on an internet-wide basis would create substantial technological and bureaucratic overhead with very little return in value, among other problems (Doldirina, Eisenstadt, Onsrud, and Uhlir, publication pending).

**[ ] Guideline 1D. In asserting any rights and in applying access and reuse terms and conditions to any dataset 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. They are generally not legal requirements, but are ethical or normative ones.

Such considerations are important for prospective data providers to determine whether to make research data collected by the public sector and with public money freely available, as well as what the users of those data may do with the data. If the research data were collected using public funds, the taxpayer has already paid for generating the data and 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 and Three and its Implementation Guidelines. The outputs 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. Moreover, for government employees at least, there is no need to apply intellectual property laws to incentivize and protect those public outputs. The benefits from the access and use of such data are equitably available to all users. The originators or rights holder(s) of research data therefore should refrain from claiming intellectual property rights for public or publicly funded research data (and other public information).

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 level playing field and opportunities for a leapfrogging effect. 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 is unethical and inequitable.

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 data provider, as well as any laws of the jurisdiction in which they are located and the norms of the community within which they work. At the same time, the fewer constraints that are imposed by the provider, the more creative the users can be with the data and the greater are the 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 access, use, share, produce, or combine research data.

[ ] Guideline 2A. Research data *providers* need to know who or what institution has the rights to the data before they are disseminated to others.

Establishing the holder of rights in a research dataset is important because all downstream decisions on making the dataset 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, if any.

Moreover, a provider or 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 dataset also is not necessarily synonymous with its original producer.

For example, an individual researcher may produce original data, but the rights are controlled by the employing institution. 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 entity holds the rights to the dataset 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 a mandatory part of contracts of employment, and of research cooperation and grant agreements. The exact status of rights in any data should be transparent in two respects, stating clearly i) who is the rights holder of any subset of a dataset and ii) who is the rights holder 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, and controlling rights do not necessarily imply any intellectual property rights protection for those data.

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 ownership and rights (see also [Principle Five and the accompanying Implementation Guidelines on Harmonization](#PrincipleFour), below).

 [ ] Guideline 2B. It is also the responsibility of research data *users* to know the rights of the 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 other policies 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.

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 law heavily favors the rights holder. 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 we do not intend to provide a treatise on limitations and exceptions to copyright (see, e.g., Reichman and Okediji 2014), we do offer the following general guidance to data users:

* Assess whether a statutory limitation or exception permits a research or other public-interest use. Researchers in the United States may be able to rely on the Fair Use exception to the Copyright Act. In the United Kingdom and other Commonwealth countries, researchers may be able to use the Fair Dealing exception. In other jurisdictions, 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 the statutory exception does not apply, the researcher can go back to the source of the database and try to negotiate more reasonable terms for their intended use(s). This can entail a substantial transaction cost of time and money, however.
* If there is no statutory exception or limitation, or the researcher is turned down (rather than being offered some kind of use fee), or the rights owner is impossible to determine, then there is the suboptimal 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, exist that 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). In the event that 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. Expert representatives of research communities 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 that relate to the legal interoperability of research data, and advocate for legal frameworks that facilitate scientific progress. The development of such rules is a mix of overarching and specialized legal, policy, or normative actions within each data community.

**[ ] Guideline 2D. A well-conceived educational process should be developed and adopted by relevant institutions, in order that future generations of researchers might be better prepared.**

With the aim of achieving a sustainable set of practices in the long term, institutions that educate and train researchers in all disciplines should include the legal interoperability of research data and the policies and procedures for research information, generally, as part of the responsible conduct of research. 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., editors, reviewers, librarians, data managers, and researchers generally), and the design and funding of targeted research programs.

Principle Three: Balance the legal interests.

Definition and Discussion of Terms

| [Intellectual property rights](#IPRsDef) | [Database protection rights](#DatabaseRightsDef) | [Patents](#PatentsDef) | [Trademarks](#TrademarksDef) | Copyright [infringement](#INfringementDef) | | [National security or public order laws](#NatSecurityDef) | Personal privacy protection | [Endangered species](#EndangeredDef) protection | [Cultural resources](#CulturalDef) protection | Traditional Knowledge protection | [Genetic resources](#GeneticDef) protection | [Confidentiality laws and policies](#ConfidentialityDef) |

**Guidelines for Implementation**

 As declared in The Hague Declaration on Knowledge Discovery in the Digital Age:

“the free flow of information and ideas is an essential human right. It is a catalyst for the production of human knowledge, which underpins welfare and prosperity. Societies around the world have chosen to protect certain limited rights in intellectual property as incentives both to innovation and the dissemination of knowledge. Intellectual property law was never intended to cover facts, ideas and pure data” ([LIBER, 2015, p.2](#LIBER)).

Freedom of access and reuse of research data that enables legal interoperability, however, 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, personal privacy, confidentiality, endangered species or cultural resources, Traditional Knowledge, genetic resources, 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.

Restrictions on access to and reuse of research datasets and the data within them, especially those generated through public funding, therefore generally ought to be applied as sparingly as possible and only pursuant to legitimate reasons. They should fully respect the public interest for the diffusion of knowledge. The following Implementation Guidelines can help achieve this balance.

[ ] Guideline 3A. Rights holders should not assert intellectual property rights in government or publicly-funded research datasets.

Ethical norms compel all participants in the research process to refrain from reserving intellectual property rights in research datasets, especially to those obtained partly or fully from publicly funded research (see Principle One and Guideline 1.C). A statutory intellectual property 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 or other intellectual property rights, 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](#PrincipleOne) and Implementation Guideline 1.A and 1.B).

[ ] Guideline 3B. Governments and public research institutions need to justify any legally-imposed restrictions on research data.

The restrictions on open access and reuse of research data described in the definitions section below cannot be waived. 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 light of open access to knowledge.

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 datasets available in anonymized or aggregated form.

[ ] 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 or complete provisional results or to publish the findings. The duration may vary by discipline. However, such restrictions should be narrowly limited in time and be specified by the funder of the research. The release of research data partly or fully funded by the public sector should not be blocked when the data are delivered to the public funder, because this contradicts the overarching principle of open access to and unrestricted reuse of research data, as specified in Principle One.

[ ] Guideline 3E. All rights holders of research datasets that are partly or fully funded by the public sector need to avoid individual contracts or agreements that restrict access to and reuse of the data.

Individual, commercial contracts or “end-user licensing agreements” (EULAs) 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 pose significant obstacles to the interoperability of datasets and require costly renegotiation on a case-by-case basis, if that is even possible.

Principle Four: State the rights transparently and clearly

Definition and Discussion of Terms

| [Transparency](#TransparencyDef) | [Certainty](#Certainty) | [Rights status](#RightsStatusDef) | [Research Data](#ResearchDataDef) | Metadata

**Guidelines for Implementation**

An unambiguous statement of legal rights and policies pertaining to all datasets 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 data in the public domain. When a dataset 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 any data in understanding the rights and responsibilities in their potential for reuse, including their legal interoperability. Many of the relevant laws and policies are not always restated in conveying the data to the user. Despite the old legal principle that ignorance of the law is no excuse, users cannot be expected to know the intricacies of all the relevant laws and policies. It is therefore incumbent on the provider of the data to state clearly any restrictions that are imposed on the access to and reuse of those data.

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.

[ ] Guideline 4A. Standardized electronic statements regarding the legal rights retained (if any) can greatly assist in their comprehensibility by a wide audience--including by machines.

Statements concerning the legal rights pertaining to research data form part of the metadata (documentation for the user) of any dataset. Metadata are the principal mechanism through which transparency and certainty can be achieved. 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 datasets, 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 data appropriately. Rather, information about ownership rights and usage terms and conditions are only loosely coupled to the dataset files in the form of copyright notices or as open licenses posted on the dataset landing page. Legal information governing data reuse is therefore easily disconnected from the data and lost, leaving potential users concerned about violating the law, assuming legal risk, or actually infringing on intellectual property rights. Removing legal uncertainty regarding data reuse requires consistent and predictable rights information that remains associated with the data assets being used.

The metadata for any publicly available dataset 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 (e.g., subject to copyright or database protection legislation; in the public domain; status unknown/orphan work, and the like). Additionally, this field may indicate the presence of any terms and conditions governing use of the dataset, including whether it is freely accessible or only with 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 thus can be significantly enhanced by means of a common taxonomy of rights statements that are both human and machine readable: standard ways of expressing the open access or restrictions governing a given research output.

In light of this guidance, the rights holder(s) of any research-related datasets are encouraged, when appropriate, to use the Creative Commons Public Domain Mark or the Creative Commons No Rights Reserved Instrument (CC0)(see Principle One and Guideline 1B). These legal instruments have “human” (i.e., lay person), “machine”, and formal “legal” meanings that use standard language and are recognized worldwide.

 [ ] Guideline 4B. The rights holder(s) of any given research dataset used in research should have access to competent legal counsel 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:

* *Collections of data with common usage options.*

Research data with a comparable set of specialized rights could be gathered in collections that are made available with special support concerning legal issues of reuse. Datasets within the collection could be earmarked 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.

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

The approach used here is to govern access and reuse rights on applications reviewed by specialized bodies with expertise in the subject-matter of the data. 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); and the policy approved by the Steering Committee of the Federal Geographic Data Committee (FGDC) in April 1998. See [https://www.fgdc.gov/policyandplanning/privacy%20policy/?searchterm=confidentiality](https://www.fgdc.gov/policyandplanning/privacy%20policy/?searchterm=confidentiality" \t "_blank). 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

| [Total harmonization](#TotalHarmDef) | National treatment | [Partial harmonization](#PartialHarmDef) | [Alternative harmonization](#AltHarmDef) | [Optional harmonization](#OptionalHarmDef) | [Fair use and fair dealing](#FairUseDef) | [Norms](#NormsDef) |

**Guidelines for Implementation**

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, museums, publishers, or e-science 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. Both top-down and bottom-up approaches, and mixes of both, can be used to harmonize rights concerning research data.

There are different levels of the harmonization of rights in research data, but the goal should be achieve total harmonization to facilitate legal interoperability of the data. Levels of harmonization 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.

[See, e.g., at the global level, 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.]

*National treatment of laws.* National treatment (or “mutual recognition”) of laws is a type of loose harmonization achieved through the recognition of full value given to each other´s laws (nobody contests the open or limited scheme of the others and fully agree to respect it as if it were its own).

***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). [See, e.g., 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 dataset 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.]

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, are an easy approach for individuals or groups of researchers to take, absent government action, but are relatively fragmented, with harmonization more elusive.

A rights holder can voluntarily use a Creative Commons (CC0) waiver of rights, with “no rights reserved”, or the Public Domain Dedication License (PDDL) of the Open Data Consortium, which are described in Table 1 of Principle One Implementation Guideline 1B. The CC0 waiver of rights provides 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 <http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf>.

Also, for example, 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 that opted for it instead of negotiating an amendment to the Convention that established CERN: data are released under the emerging standard Creative Commons CC0 waiver; even though it has 4 levels of data releases based on a policy under which CMS “will provide open access to its data at different points in time with appropriate delays, which will allow CMS collaborators to fully exploit the scientific potential of the data before open access is triggered (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 typical and undisputed research data, the easier it is to achieve total harmonization of the rules applicable to public access and reuse of those data (see, e.g., for astronomical datasets, 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 domestic level of each jurisdiction, the professional codes “of conduct” or “of practice” on “fair use” (in the United States) are a good example of some degree of partial harmonization. This is the case, in particular, since U.S. Courts interpret §§ 106 and 106A of 17 U.S.C. § 106A, by looking not only at context, amount, and value of the use, but also to the standards and practices of the professional communities where the case comes from, although the real depth of the level of harmonization might not be sufficiently satisfactory since the professional codes fail to provide one-size-fits-all solutions to legal disputes due to the willingness of courts to examine their scope and extension in a case-by-case basis.

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 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 dataset(s) 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 Principle One Implementation Guideline 1.B). It 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 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), is another example of top-down harmonization using a soft-law approach for a certain type of public data.

Another example of harmonization through treaties is the agreed scope of “limitations and exceptions to copyright” based on antitrust competition law under the so-called “Berne three-step test” (the clause that is included in several international treaties on intellectual property).

Sometimes harmonization is achieved through the addition of unilateral actions of different States 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.

For an example of judicial interpretations of statutory law on exceptions and limitations to copyright, see, e.g., the case law on the scope and meaning of “fair use” under US §§ 106 and 106A, 17 U.S. Code, as interpreted by courts in the United States, when compared to “fair dealing” as applied in Canada, Australia, Canada, New Zealand, Singapore, India, South Africa and the United Kingdom, among others, or to the interpretation of exceptions and limitations to copyright in other jurisdictions.

[ ] Guideline 5D. Process-based approaches such as workflows decision-making charts, decision making apps and tools, or scoreboards (versus substantive harmonization by negotiating common standards) might be considered useful tools to promote harmonization.

Some 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 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>.

Even the level or degree of adherence to the present RDA-CODATA Principles and their Implementation Guidelines could be subject to self-evaluation by actors that may commit to them in the future, although producing such charts, apps, and tools is only envisioned as potential future work for the RDA-CODATA IG at this stage.

***Principle Six: Provide proper attribution to the correct parties.***

***Definition and Discussion of Terms***

| [Acknowledgement](#AcknowledgementDef) | [Attribution](#AttributionDef) | [Citation](#CitationDef) | [Credit](#CreditDef) | [Plagiarism](#PlagiarismDef) |

**Guidelines for Implementation**

Attribution and citation of a well-prepared, quality dataset is an essential part of the research enterprise. 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. 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 dataset may be the product of activity involving many individual contributors, or where very large numbers of datasets 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 of national copyright statutes. Copying with intentional lack of attribution or, in some cases, with copyright infringement, arises to the level of plagiarism. 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 International license. This license requires attribution to the “author” (the data producer) and the source (provider) in a manner specified by the rights holder (through the license), but its legal enforceability is limited to the copyrightable portions of the dataset. The copyrightability of most datasets is “thin”; that is, it often does not apply to the full contents of the dataset and is uncertain in its application. 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—are rare because of the expense and legal uncertainty. Consistent with the advice given by Creative Commons and for the reasons provided in Principle One and Implementation Guideline 1C, we do not endorse the use of the CC Attribution 4.0 International license for research data.

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 data producers and providers 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 (see the Bibliography). 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, we do endorse the CC0 waiver of rights that suggests the normative attribution be given to the research data in question consistent with research discipline practice (see https://creativecommons.org/about/cc0/ and Principle One and Implementation Guideline 1B).

Beyond the norms and ethics inherited from the scientific publishing system, specific practices for attribution for research data are being established and finding widespread support. Specifically, the development of persistent identifiers assigned to data sets, and the creation of the DataCite organization in 2009, has made it possible to formalize the concept of data citation. Another notable effort has been CODATA’s Data Citation Standards and Practices Task Group whose analyses and reports 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. Other organizations have also been active in this area.

*

Definitions and Discussion of Terms

 Note: These definitions are organized by principle and Guideline, in the order they appear in the text.

|  |  |
| --- | --- |
| Interoperability of data[[Return to Principle One](#PrincipleOne): Access and Reuse] | Interoperability of data at the technical level may be defined as the “property of a product or system … to work with other products or systems, present or future, without any restricted access or implementation” [http://interoperability-definition.info/en]. Interoperability is an attribute that greatly facilitates usability of research data.  [NOTE: we also need to add the definition of “legal interoperability”.] |
| Open access[[Return to Principle One](#PrincipleOne): Access and Reuse] | Definitions of “open access” started in debates to promote the wider availability of scientific literature.  Open access to scholarly literature is defined by the Budapest (Feb. 2002), Bethesda (June 2003), and Berlin (Oct 2003) statements or declarations. The following text is shared by the Bethesda and Berlin definitions:“The author(s) and copyright holder(s) grant(s) to all users a free, irrevocable, worldwide, perpetual right of access to, and a license to copy, use, distribute, transmit and display the work publicly and to make and distribute derivative works, in any digital medium for any responsible purpose, subject to proper attribution of authorship.” ([Suber, 2003](#Suber2003)) Thus, ”open access” means unrestricted access to and use of scientific information and data. Open access exists to facilitate reuse and legal interoperability is an important component of this process. |
|  |  |
|  |  |
|  |  |
| Public domain[[Return to Principle One](#PrincipleOne): Access and Reuse] |  |
| Intellectual property rights (IPRs) | As used in these Implementation Guidelines, “intellectual property rights” include copyright, database protection rights, trade secret, patents and other similar laws that establish a regime for access, use or reuse of data or metadata. It does not include other restrictive information laws that are also discussed in these Implementation Guidelines. See the individual entries for [Copyright](#CopyrightDef), [Database protection rights](#DatabaseRightsDef), [Patents](#PatentsDef), and Confidentiality. |
| Copyright[Return to [Principle Three](#PrincipleTwo): Balance Legal Interests] | Copyright is the right to monopolize a work of art and literature. 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. 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 of this content.  |
| Database protection rights[Return to [Principle Three](#PrincipleTwo): Balance Legal Interests] | Database protection in a legal form of a *sui generis* (“of its own kind,” or unique) right exists mostly in the E.U. (with a few similar applications in other countries) ([European Parliament, 1996](#EUPARLIAMENT)). It applies to databases that show a considerable private investment in the generation, verification, and presentation of the contents. Database protection refers to the entire 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. 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. |
| Patents [Return to [Principle Three](#PrincipleTwo): Balance Legal Interests] | Patents are granted for by a national 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, but may impede certain commercial reuses of these data for a given time period. |
|  |  |
| Copyright infringement[Return to [Principle Three](#PrincipleTwo): Balance Legal Interests] | 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 US 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 at: <http://www.wipo.int/sme/en/documents/copyright_infringement_fulltext.html>.) |
| National security or public order laws[Return to Principle Three: Balance Legal Interests] | National security laws, or more broadly, 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. |
| Endangered species protection[Return to Principle Three: Balance Legal Interests] | 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, 1973](#CONVENTION). |
| Cultural resources protection[Return to Principle Three: Balance Legal Interests] | 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, 1970, [http://portal.unesco.org/en/ev.php-URL\_ID=13039&URL\_DO=DO\_TOPIC&URL\_SECTION=201.html](http://portal.unesco.org/en/ev.php-URL_ID%3D13039%26URL_DO%3DDO_TOPIC%26URL_SECTION%3D201.html); the Convention on the Protection of World Cultural and Natural heritage, 1972, [http://portal.unesco.org/en/ev.php-URL\_ID=13055&URL\_DO=DO\_TOPIC&URL\_SECTION=201.html](http://portal.unesco.org/en/ev.php-URL_ID%3D13055%26URL_DO%3DDO_TOPIC%26URL_SECTION%3D201.html); and the Convention of the Protection and Promotion of the Diversity of Cultural Expressions, 2005, http://portal.unesco.org/en/ev.php-URL\_ID=13055&URL\_DO=DO\_TOPIC&URL\_SECTION=201.html. |
| Traditional Knowledge laws[Return to Principle Three: Balance Legal Interests] | There are Traditional Knowledge (TK) laws in many countries that prohibit the misappropriation of various data and information that have cultural significance. Such laws typically 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, 1989, http://portal.unesco.org/en/ev.php-URL\_ID=13141&URL\_DO=DO\_TOPIC&URL\_SECTION=201.html |
| Genetic resourcesa protection[Return to Principle Three: Balance Legal Interests] | 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 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, <https://www.cbd.int/convention/>; Nagoya Protocol on the Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization, 2014, https://www.cbd.int/abs/doc/protocol/nagoya-protocol-en.pdf |
| Confidentiality laws[Return to Principle Three: Balance Legal Interests] | 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, and the protection of other commercial interests. Such presumptive or real interests often form the background to information policies of institutions and enterprises that try to withhold data and information they produce, or at least control the use of them.  |
| Period of exclusive use of research data[Return to Principle Three: Balance Legal Interests] | Many scientific funding institutions, universities and research institutes, and scientific 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. |
| Personal privacy protection[Return to Principle Three: Balance Legal Interests] | In many 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).  |
| Individual contracts or use agreements[Return to Principle Three: Balance Legal Interests] | Through the use of individual contracts, often referred to as end-user licenses 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. |
| Transparency[Return to Principle Four: Transparency and Certainty] | Provision of all pertinent information concerning the holder of rights and the status of the rights, if any, in a dataset to the extent that is feasible, provided with reasonable effort and cost by the person or organization making the data available. |
| Certainty[Return to Principle Four: Transparency and Certainty] | Certainty is based on the completeness and accuracy of the information given about the rights in a dataset. Such information reduces ambiguities and the exposure to legal risks for the user of that dataset.  |
|  |  |
| Research Data[Return to Principle Four: Transparency and Certainty] | 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.” <http://dictionary.casrai.org/Research_data>. |
|  |   |
|  |  |
|  |  |
|  |  |
|  |  |
| Fair use and fair dealing[[Return to Principle Five](#PrincipleFour): Harmonization] | Both of these legal exceptions and limitations are based on a statutory clause (or a judicial doctrine in some national jurisdictions) that permit limited use of copyrighted material without acquiring or requiring permission from the rights holders. Examples of fair use (under U.S. law) include commentary, search engines, criticism, parody, news reporting, research, teaching, library archiving, or scholarship. Although not exactly similar, in many other national (or supranational, as it is the case of the European Union) jurisdictions, open-ended provisions in statutes listing limitations and exceptions to copyright are known as “fair dealing,” which allow for copyrighted works to be used without a license from the copyright owner, and that include a number of important considerations, such as: competition law, freedom of speech, education, or equality of access (e.g., by the visually impaired). When judicially interpreted, fair dealing accomplishes a similar function as fair use, but is more usual in common law countries. |
| Norms[[Return to Principle Five](#PrincipleFour): Harmonization] | 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. By example, The Code of Best Practices in Fair Use for Academic and Research Libraries reflects librarians’ application of Fair Use, drawn from the actual practices and experience of the library community itself. (Association of Research Libraries et al., 2012).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). |
|  |  |
| Metadata[[Return to Principle Four](#PrincipleFive): Metadata] | Metadata are the structured descriptions of data sets 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. |
| Rights statement[[Return to Principle Four](#PrincipleFive): Metadata] | 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 legally-binding rights attached to the licenses, norms, agreements, or public domain status of the content, of the means of expression, or of the data itself. This definition is based on Europeana-DPLA, May 2015 “Recommendations for the Technical Infrastructure for Standardized International Rights Statements”, International Rights Statements Working Group. |
|  |  |
| Acknowledgement[[Return to Principle Six](#PrincipleSix): Attribution and Credit] | 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.  |
| Attribution[[Return to Principle Six](#PrincipleSix): Attribution and Credit] | 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. |
|  |  |
| Citation[[Return to Principle Six](#PrincipleSix): Attribution and Credit] | 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. |
| Credit[[Return to Principle Six](#PrincipleSix): Attribution and Credit] | Formal recognition of the contribution made by an individual or group to a research output. |
| Plagiarism[[Return to Principle Six](#PrincipleSix): Attribution and Credit] | 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, Federal Research Misconduct Policy, http://ori.hhs.gov/federal-research-misconduct-policy) |
| Equity[[Return to Principle One](#PrincipleSeven): Equity] | Equity may be defined as "the quality of being fair and impartial." Oxford English Dictionary. |
|  |  |
|  |  |
|  |  |

BIBLIOGRAPHY

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, 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>.

Ball, A. and Duke, M., 2015. “How to Cite Datasets and Link to Publications”. *DCC How-to Guides*. Edinburgh: Digital Curation Centre. Available online: <http://www.dcc.ac.uk/resources/how-guides/cite-datasets>.

Budapest Open Access Initiative, 2002, “Read the Budapest Open Access Initiative” (web page), <http://www.budapestopenaccessinitiative.org/>.

Perspectives on Plagiarism and Intellectual Property in a Postmodern World(Lise Buranen and Alice Myers Roy, eds., 1999).

Albert, Tim and Wager, Elizabeth, 2003, “How to handle authorship disputes: a guide for new researchers”, *COPE Report 2003*. Available online: <http://publicationethics.org/files/2003pdf12_0.pdf>.

CODATA, 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>.

1973, *Convention on International Trade in Endangered Species of Wild Fauna and Flora*, Available online: <https://cites.org/eng/disc/text.php>.

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

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

DataCite Metadata Working Group, 2015, *DataCite Metadata Schema for the Publication and Citation of Research Data*. Available online: <http://schema.datacite.org/meta/kernel-3/doc/DataCite-MetadataKernel_v3.1.pdf>

Data Sharing Working Group, 2014, *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-%20%20Mechanisms%20to%20share%20data%20as%20part%20of%20GEOSS%20Data_CORE.pdf>.

Data Citation Synthesis Group, 2014. ***Joint Declaration of Data Citation Principles.*** Martone M. (ed.) San Diego CA: FORCE11. Available online: <https://www.force11.org/group/joint-declaration-data-citation-principles-final>.Digital Curation Center and Research Data Alliance, Metadata Standards Directory Working Group, 2015, “Metadata Directory” (web page), <http://rd-alliance.github.io/metadata-directory/>.

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

DPLA-Europeana International Rights Statements Working Group*, 2015, Rightsstatements.org White Paper: Recommendations for Standardized International Rights Statements.* Available online: <http://rightsstatements.org/files/151002recommendations_for_standardized_international_rights_statements.pdf>

European Commission, 2016, *Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020*, Version 2.1. Available online: http://ec.europa.eu/research/participants/data/ref/h2020/grants\_manual/hi/oa\_pilot/h2020-hi-oa-pilot-guide\_en.pdf.

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>.

Fisk, Catherine L., 2006, “Credit Where It’s Due: The Law and Norms of Attribution,” *Georgetown Law Journal*, vol. 95, p. 49-117.

Franck, George, 1999, “Scientific Communication--A Vanity Fair?, *Science*, 1 October 1999, p. 53-55.

“G8 Science Ministers Statement”, 13 June 2013 (web page), <https://www.gov.uk/government/news/g8-science-ministers-statement>.

Group on Earth Observations, Data Sharing Working Group, 2014a, *Legal Mechanisms to Share Data as Part of GEOSS Data-CORE*. Available online: <https://www.earthobservations.org/documents/dswg/Annex%20VI%20-%20%20Mechanisms%20to%20share%20data%20as%20part%20of%20GEOSS%20Data_CORE.pdf>

Group on Earth Observations, Data Sharing Working Group, 2014b, *GEOSS Data Sharing Principles Post-2015*. Available online: https://www.earthobservations.org/documents/dswg/Annex%20III%20-%20GEOSS%20Data%20Sharing%20Principles%20Post-2015.pdf.

Harris, Lesley Ellen, “What to Do If You're Accused of Copyright Infringement”, World Intellectual Property Organization (WIPO) (web page). <http://www.wipo.int/sme/en/documents/copyright_infringement_fulltext.html>.

International Council for Science (ICSU), Ad -hoc Strategic Coordinating Committee on Information and Data, 2011, Interim Report to the ICSU Committee on Scientific Planning and Review. Available online: <http://www.icsu.org/publications/reports-and-reviews/strategic-coordinating-committee-on-information-and-data-report/SCCID_Report_April_2011.pdf>.

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>.ternational Council for Science, Committee on Data for Science and Technology (CODATA), 2013, *Out of Cite Out Of Mind: The Current State of Practice, Policy, and Technology for the Citation of Data*. Available online: <http://datascience.codata.org/articles/abstract/10.2481/dsj.OSOM13-043/>.

International Organization for Standardization (ISO), “Reference Codes for the Representation of Names of Languages (ISO 639-2)” (web page), <https://www.loc.gov/standards/iso639-2/php/code_list.php>.

International Organization for Standardization (ISO), “Country Codes (ISO 3166)” (web page), <http://www.iso.org/iso/country_codes>.

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>.

LIBER, Association of European Research Libraries, 2015, *The Hague Declaration on Knowledge Discovery in the Digital Age*. Available online: http://thehaguedeclaration.com/the-hague-declaration-on-knowledge-discovery-in-the-digital-age/.

Merton, R. K., 1968, “The Matthew effect in science: the reward and communication systems of science are considered”. *Science*, 1 59, 56-63.

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

Organization 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>.

Patterson [complete]

Research Data Alliance Working, Group on Data Citation (WGDC), 2015. *Data Citation of Evolving Data*. Available Online: <https://rd-alliance.org/system/files/documents/RDA-DC-Recommendations_150924.pdf>.

Suber, Peter, 2003, “Bethesda Statement on Open Access Publishing” (web page), <http://dash.harvard.edu/bitstream/handle/1/4725199/suber_bethesda.htm?sequence=1>.

Rasmusen, Eric, and Posner, Richard A., 2000 "Creating and Enforc ing 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>.

Reichman, Jerome H., and Okediji, Ruth, 2014,

Reichman, Jerome H., and Uhlir, Paul F., 1999, “Database Protection at the Crossroads”, Berkeley Technology Law Journal, pages Available online:

Reichman, Jerome H., and Uhlir, Paul F., 2003, “A Contractually Reconstructed Public Domain in Research Data in a Highly Protectionist Intellectual Property Environment”, *Journal of Contemporary Problems, Duke Law School*, pages 315-459. Available online:

Uhlir, Paul F., 2015, “The Value of Open Data Sharing”, Group on Earth Observations, CODATA, 39 pages. Available online:

United States Code, Title 17, Section 106A, “Rights of certain authors to attribution and integrity” (web page), <http://www.copyright.gov/title17/92chap1.html#106a>.

United States Code, Title 17, Section 501, “Infringement of Copyright” (web page), <http://www.copyright.gov/title17/92chap5.html#501>.

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

1. Acknowledgements: These Implementation Guidelines are the product of the entire RDA-CODATA Interest Group. The co-chairs of this Interest Group are Paul F. Uhlir, Enrique Alonso Garcia, and Robert Chen. The list of all members may be found at: https://rd-alliance.org/groups/rdacodata-legal-interoperability-ig.html. However, the following individuals were the principal authors and participated in weekly conference calls and RDA Plenaries to develop them between March 2015 and March 2016 (in alphabetical order): Donat Agosti, Enrique Alonso Garcia, Christoph Bruch, Robert Chen, Willi Egloff, Herbert Gruttemeier, Simon Hodson, Maria Llosent, J. Bernard Minster, Gail Peretsman-Clement, and Paul F. Uhlir. [↑](#footnote-ref-1)
2. This Table is reproduced from the following article: Doldirina, Catherine, Anita Eisenstadt, Harlan Onsrud, Paul F. Uhlir (pending publication). Legal Approaches to the Open Sharing of Research Data. [↑](#footnote-ref-2)