



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

Implementation Guidelines for the Principles on the Legal Interoperability of Research Data

December 6, 2015

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 and in understanding different approaches in various countries.

The Implementation Guidelines build upon extensive previous work by scientific bodies and research data communities, reflected in Principles, Declarations and Statements issued over the past 15 years, focusing on open access and reuse of research data from public funding. Many of these are policy or normative documents developed by community initiatives, and we cite and link to them in [Appendix 1](#) as providing the relevant background for and evidence of our Principles and Implementation Guidelines.

The Implementation Guidelines will be useful for practitioners to achieve greater legal interoperability of research data to reference from a national, discipline-based, or specific sector perspective. As explained by the Group on Earth Observations (GEO) Data Sharing Working Group in their white paper *Legal Mechanisms to Share Data as Part of GEOSS Data-CORE* ([GEO, 2014a](#)), legal interoperability occurs among multiple datasets when:

- use conditions are clearly and readily determinable for each of the datasets,
- 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.

Legal interoperability also implies that the search for, or tracking of, licenses or other legal instruments and their compatibility with other legal conditions will occur online. When data are combined from multiple sources the resulting dataset will incorporate the accumulated restrictions imposed by each source. Therefore, 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. It is essential for practitioners to be cognizant of the challenges and opportunities inherent in intellectual property laws ([GEO, 2014a](#)).

It is important to note as well that neither these Principles nor the Implementation Guidelines are 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 many other norms and perspectives—ethical, technical, cultural, and others—beyond just purely legal ones to consider.

The Legal Interoperability Interest Group is also working with other RDA groups to develop use case examples and definitions of key terms pertaining to the intellectual property and proprietary law topics in research data, [and there will be links to those provided here as well].

The following Implementation Guidelines for the International 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. They are organized according to the ten Principles and each contain sections on “Definition and Discussion of Terms” and “Guidelines for Implementation”. The Interest Group endeavored to provide cross-references to other germane Principles and reference key supporting documents. [TO BE ADDED AFTER THE REVIEW: The Implementation Guidelines were reviewed by the Research Data Alliance and CODATA according to their review processes, as well as by selected independent experts, and this document has been approved by both organizations.]

Acknowledgements

These Implementation Guidelines are the product of the entire RDA-CODATA Interest Group. However, the following individuals were their principal authors and participated in the weekly conference calls to develop them (in alphabetical order): Donat Agosti, Enrique Alonso Garcia, Christoph Bruch, Robert Chen, Gail Clement, Willi Egloff, Herbert Gruttmeier, Simon Hodson, Maria Lloset, J. Bernard Minster, and Paul F. Uhlir.

GUIDELINES FOR IMPLEMENTATION

1. ACCESS AND REUSE: GUIDELINES FOR IMPLEMENTATION

Principle One: Access and Reuse

Facilitate access to and reuse of research data. In order to enable the maximum degree of interoperability, access to and reuse of research data should be open and unrestricted by default or otherwise be granted to users with the fewest limitations possible. Data in the public domain ensures that there are no restrictions

Definition and Discussion of Terms

| [Interoperability of data](#) | [Open Access](#) | [Unrestricted reuse](#) | [Rights waivers & non-restrictive licenses](#) | [Public domain](#) |

In order to enable the maximum degree of interoperability, and to promote access and reuse, we provide the following guidelines for implementation.

[] Guideline 1A. Rights holders should state the conditions of data reuse clearly.

The data should be clearly labelled with the conditions for reuse. Ideally, this should be by means of one of the following:

- A clear label or statement that the data is already in the public domain;
- A waiver of rights that dedicates the data to the public domain;
- A non-restrictive, common-use license that imposes no condition greater than the requirement to give attribution to the data producer.

The waiver of rights should be clearly displayed and explained. The information contained should be both human and machine readable. For humans, the explanation of the waiver of rights and its implications should be clear and consistent. Where some conditions or restrictions have been imposed, the implications of these also should be made clear to the potential user. Where the waiver of rights imposes no conditions, this should be transparent and should not be taken for granted. See [Principle 3](#) and [Principle 5](#), and the accompanying Implementation Guidelines, below.

[] Guideline 1B. A voluntary assignment of a waiver of rights by the rights holder(s) in any given dataset is the best way to achieve legal interoperability.

Some data by definition may already be in the public domain, with no restrictions on reuse. Where this is the case, the data should be clearly labelled as such.

Conditions under which datasets may be dedicated to the public domain: the broad international definition of copyright and the (mostly) European *sui generis* database right means that in most cases

the rights holder can assign a public domain rights waiver (by using a CC0 waiver) to a database that has required reasonable effort or originality in its making.

Where the dataset has been created as a result of public funding, the data should be made available and easily accessible to the greatest extent possible, and should be dedicated to the public domain using an appropriate waiver of rights.

[] Guideline 1C. Standard rights waivers and unrestrictive licenses should be applied by the rights holders as much as possible.

The GEO Data Sharing Working Group recommends “voluntary waivers or standard common-use licenses” that meet the objectives of legal interoperability. ([GEO, 2014a](#), p. 5)

- Creative Commons Public Domain Mark
- Statutory waiver of copyright
- Creative Commons Public Domain Waiver (CC0)
- Open Data Commons Public Domain Dedication and License (PDDL)
- Creative Commons Attribution License (CC BY 4.0)

[] Guideline 1D. Governmental bodies and other public or academic research institutions should develop a clear legal and administrative framework for public research data.

The GEO Data Sharing Working Group has further recommended “legislative, regulatory or administrative and other government measures placing all data and information produced by government entities in the public domain” ([GEO 2014a](#), p. 4) In the meantime, it is the responsibility of data rights holders to apply public domain waivers of rights or use a CC-BY license. There is a need to promote this practice through education and training, to clarify the normative language and understanding about the waiving of rights and other legal aspects of research data.

[] Guideline 1E. Governmental bodies and public or academic research institutions should also implement other basic prerequisites for open access and unrestricted reuse.

It should be observed that in order to be “open” and “unrestricted”, access and reuse must be facilitated. This requires online visibility, facilitated by data publication and citation good practices, as well as data preservation and curation standards (e.g., encoding, formats, protocols, persistent identifiers, or PIDS, and metadata schemas). Some of these aspects are taken up in other parts of these Implementation Guidelines.

Principle Two: Balance the Legal Interests

Freedom of access and reuse of research data that enables legal interoperability must be balanced against legitimate interests to withhold or restrict such data, or control the use of data by the rights holders. 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, and defined periods of exclusive use prior to the publication of research results.

Definition and Discussion of Terms

| [Intellectual property rights](#) | [Database rights](#) | [Patents](#) | [Trademarks](#) | [Infringement of copyright](#) |
| [National security or public order laws](#) | [Protection of endangered species](#) | [Protection of cultural resources](#) | [Protection of indigenous peoples' rights](#) | [Protection of genetic resources](#) |
| [Confidentiality laws and policies](#) |

2. BALANCE THE LEGAL INTERESTS: GUIDELINES FOR IMPLEMENTATION

Freedom of access and reuse of research data, especially data obtained from publicly funded research, is a public interest. It is of general interest for knowledge to be diffused as broadly as possible for the results of public funded research serve the whole society and for these results to be reused for any further research. This public interest, however, may conflict with other public interests that make it necessary to withhold certain data and to restrict access to them. It may also be contrary to the personal interests of researchers, publishers, and other participants in the research process who would prefer to withhold their data.

As a consequence and despite Principle One above, which advances open access and unrestricted reuse of research data as the default rule, there are numerous legitimate countervailing laws, policies, and norms that sometimes must supersede this principle. Not every restriction is valid, however: there are numerous restrictive declarations accompanying data and information that have either an invalid legal basis or are not justified.

To balance the legal interests means to weigh the public interest on free information exchange against these conflicting interests following objective criteria. We enumerate the terms of the justified countervailing interests used in Principle Two and provide some guidance to navigating those restrictions.

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](#)).

Restrictions on access to and reuse of research data, especially those generated through public funding, should therefore 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 will help achieve this balance.

[] Guideline 2A. Rights holders should refrain from reserving all intellectual property rights on publicly-funded research data.

Ethical rules impel researchers, institutions, publishers to refrain from reserving intellectual property rights in research data, especially to data obtained partly or fully from publicly funded research. If data are kept in texts, collections, or databases that qualify as copyrightable works, or are protected by a database right or other intellectual property rights, the holders of such rights are urged to allow access to and reuse of those data in the least restrictive manner ([Principle One: Access and Reuse](#))

[] Guideline 2B. 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 above should be applied as narrowly as possible. In order to keep data and other forms of knowledge secret, restrictions need to be justified by an explicitly stated and overriding public or private interest.

[] Guideline 2C. Public 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, by default rule, hamper access to research data or reuse of scientific data and information. Public scientific 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 2D. Public research funding organizations and the rights holders of public research data sources should reduce time embargoes for exclusive periods of research 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 specified by the funder of the research. The release of public domain data 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 2E. All rights holders of public research data need to avoid individual contracts or agreements that restrict access to and reuse of the data.

Individual contracts or use agreements 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 renegotiation on a case-by-case basis, if it is even possible. It is therefore good practice not to enter into

individual contracts that unnecessarily restrict access and reuse of research data, especially data obtained from publicly funded research.

3. TRANSPARENCY AND CERTAINTY: GUIDELINES FOR IMPLEMENTATION

Principle Three: Transparency and Certainty

State legal rights clearly. An unambiguous articulation or labeling of rights and policies governing 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 legal exposure to risk for all parties, can remove barriers to research, and facilitates the long-term access and use of data resources, including for those data in the public domain.

Definition and Discussion of Terms

| [Transparency](#) | [Certainty](#) | [Rights status](#) | [Research Data](#) |

The first two Principles and their Implementation Guidelines describe laws, policies, and practices that are relevant to the user of any data source in understanding the rights and responsibilities in its potential for reuse, including its legal interoperability. Many of the relevant laws and policies are not 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 and re-use of those data.

By the same token, in order to maximize 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 3A. Standardized statements regarding the status of rights holder(s) can greatly assist their comprehensibility by a wide audience—including machines.

Statements concerning the rights situation of data form part of the metadata of any dataset (see [Principle and Implementation Guideline Five on Metadata](#), below). Ideally, the metadata should be comprehensible to humans and machines. Accessibility and reuse of research data is not only dependent on their rights status, but also on the ability to communicate this status in an effective way. In many cases, reuse of research data will occur as a result of electronic searches (including text and data mining) and subsequent machine manipulation (such as merging and integration with other datasets, visualization, and so on) and dissemination. One criterion of these functions will likely concern the rights status of the data. If this information cannot be communicated electronically, reuse will not occur in many cases in which reuse would be possible on the basis of their rights status, because the user does not wish to assume any legal risk.

As noted in Principle and Implementation Guideline One: Access and Reuse, rights holders of datasets are encouraged to use the Creative Commons (CC) “Attribution Only” (CC-BY) license or the CC Waiver of Rights. These legal instruments have “human” (i.e., lay person), “machine”, and formal “legal” deeds that use standard language and are recognized worldwide.

[] Guideline 3B. The rights holder(s) of any research dataset should make the legal status of that dataset clear to all users.

Establishing the control or the holder of rights in a research dataset is of great importance because all downstream decisions concerning 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, in the data and of the dataset, if the data are part of a dataset.

Research organizations should have clear rules that allow determining what rights are in data controlled by that organization and what the rights in those data may be controlled by others, including especially 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 individual data 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 reside with just one natural or legal person, and controlling rights do not necessarily imply copyright protection. All 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 in respect to research data ownership and rights (see also [Implementation Guideline 4 on Harmonization](#), below).

[] Guideline 3C. Both the rights holder(s) and the originator(s) of any given research dataset should be accurately identified.

The rights holder of a dataset is not necessarily synonymous with its originator, in the sense that the person who ought to be recognized or credited if data are to be reused is in addition to the legal rights holder. This is important because it means that researchers do not need to control or hold the rights in data they produce as part of their research in order to be credited.

[] Guideline 3D. The rights holder(s) of any given research dataset used in international research should try to anticipate and state the differences among jurisdictions.

Statements concerning the rights situation of research data have to be interpreted in the context of applicable statutory law. In many cases, 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 that is applicable depending on which jurisdiction applies to the collection of the data. The rights statement should therefore distinguish between general statements with "no" special relation to a certain jurisdiction and those with such a relation. In the latter case, this jurisdiction and possibly the law should be mentioned.

Complications based on variations of statutory law in different jurisdiction could be limited considerably if the same set of data would be made available in several jurisdictions.

If legal statements are not available in a standardized form (as in the case of CC licenses for copyright and database rights) they will be very difficult to translate. If a translation is attempted, the original text should also be included in the statement.

[] Guideline 3E. Rights holders should inform users about any terms and conditions of use.

Even with the best efforts by the authors, statements concerning the rights situation will in many cases not be sufficiently comprehensible for potential users because of the complexity of the issue. We suggest therefore additional efforts to help inform potential users of research data about the terms of use. In the interim, rights holders and disseminators of data can use two approaches:

- Collections of data with common usage options.
Research data with a comparable rights situation 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 (in the sense of infrastructure) could provide special environments that enable user to meet above mentioned conditions, e.g., protected environments assuring anonymity.
- Accessibility to data governed by committees.
Another approach is to govern access and reuse rights on applications reviewed by committees. Often such bodies already exist. 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 Geospatial data confidentiality committee [\[add link\]](#). If they do not exist, the criteria used by these bodies 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 such a body.

Principle Four: Harmonization

Researchers should facilitate legal interoperability using compatible and consistent terms and conditions for research data 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 must be balanced against the need for flexibility.

Definition and Discussion of Terms

| [Total harmonization](#) | [Mutual recognition](#) | [Partial harmonization](#) | [Alternative harmonization](#) | [Optional harmonization](#) | [Fair use and fair dealing](#) | [Norms](#) | [Berne three-step test](#) |

4. HARMONIZATION: GUIDELINES FOR IMPLEMENTATION

The implementation of most of the other principles, in particular, Principle Three on Transparency and Certainty, can be facilitated through the harmonization of the law or policies of funding and research organizations with respect to rights concerning research data. This harmonization might be achieved by different means. Intellectual property rights (IPRs) are governed by laws or policies that are mandatory and apply automatically under each national jurisdiction (top-down approach). Such laws or rules, however, also leave a lot of room for individual or collective decisions based on the autonomy of the rights holder. Consequently, different stakeholder communities accessing, using, and re-using data (e.g., libraries, museums, or e-science infrastructures), as well as the distinct scientific communities themselves, can and should harmonize their practices whenever the legal system provides opportunities for developing and implementing common understandings (bottom-up approach). It is worthwhile examining several illustrative examples of different strategies through which harmonization have been achieved, focusing first in the bottom-up approaches [add link to the subsequent examples here].

Further discussion of mechanisms for Harmonization is included in Appendix One ([See below](#)).

[] **Guideline 4A. Harmonization tools 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 4B. Both top-down and bottom-up approaches, and mixes of both, can be used to harmonize rights concerning research data.**

[] **Guideline 4C. Non-legally binding “soft” instruments are available in both top-down and bottom-up approaches, and offer more opportunities for scientific communities to reach harmonization based on their own decisions and strategies**

- In the case of top-down approaches: codes of practice or recommendations passed by the communities of stakeholders provide a good tool to harmonize open-ended statutory clauses (versus official amendment of statutes and regulations, or the negotiation of treaties). Sometimes systems organized by communities in line with official policies provide a very

productive environment for harmonization. See for example, the Group on Earth Observations process, with its 3 Data Sharing Principles and 10 Data Management Principles, based on 5 foundational elements ([GEO, 2014b](#); [still need references for the 10 data mgmt. principles and 5 foundational elements](#)).

In the case of bottom-up approaches, the adoption of norms signify consensus within scientific communities, but take time to develop; individual and voluntary waivers of rights, or non-restrictive licenses and agreements are much faster to adopt, but are specifically limited to the dataset(s) to which they pertain. Furthermore, norms need additional adequate 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 scientific research data sharing, however, since they maximize the welfare of common advancement of science while minimizing the transaction costs of having to negotiate a legally binding agreement.

[] Guideline 4D. 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.

At the same time, they are much more difficult and slow to develop than “soft law” alternatives. Nevertheless, scientific communities should not regard these processes as “political” or “legalistic” efforts alien to their agendas (see [Principle Eight on Responsibility](#)).

[] Guideline 4E. 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.

Principle Five: Metadata

Improve metadata to enhance legal interoperability. The metadata for any publicly available dataset should include all information necessary to understand the legal ownership of the data and any terms and conditions governing its access and reuse. To ensure interoperability across disparate online systems, rights metadata accompanying datasets should be made available, to the maximum extent possible, in machine-readable form using available standards. All metadata describing a dataset should be freely accessible with no legal restrictions imposed on its reuse.

Definition and Discussion of Terms

| [Metadata](#) | [Rights statement](#) | [Namespace](#) |

5. METADATA: GUIDELINES FOR IMPLEMENTATION

Metadata are the principal mechanism through which transparency and certainty can be achieved (see [Principle Three on Transparency and Certainty](#)). They are also a means to harmonize information about

rights statements (see [Principle Four on Harmonization](#)). Many metadata standards are developed by particular communities of practice and apply to discipline- or genre-specific documents or datasets. For examples of disciplinary metadata standards, see the directory of metadata standards provided by the Digital Curation Centre and maintained by the RDA Metadata Standards Directory Working Group ([DCC/RDA, 2015](#)).

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. Removing legal uncertainty surrounding data reuse requires consistent and predictable rights information that remains associated with the data assets being used. Rights metadata helps both humans and machines confidently reuse data assets without concerns of infringement or breach of license.

The specific metadata element that most effectively promotes legal interoperability of research data is the rights field, used minimally to communicate the copyright status of the resource (Copyrighted; Public Domain; Status Unknown/Orphan Work). Additionally, this field may indicate the presence of any license terms and conditions governing use of the object. Finally, the rights field may be used to convey the name and version of the license, and the legal jurisdiction where it applies.

Thus, a number of widely-used metadata standards do contain one or more “rights” elements and their sponsoring communities provide guidelines on how to populate rights fields most effectively. Each of these models represents positive steps towards recognition of the importance of rights metadata to facilitate online resource sharing at global scale. For example, the XML schema of the DataCite organization includes a <rightsList> element where the <rights rightsURI> can be expressed, as well as the identification of the rights holder among the list of contributors (DataCite Metadata Working Group, 2015). The recently announced rights statement standard under development by the International Rights Statement Working Group of the Digital Public Library of America (DPLA) – Europeana digital library consortium may offer an even more promising model for the research data community because it reflects the particular challenges of harmonizing rights statements across national jurisdictions ([International Rights Statements Working Group, DPLA-Europeana, 2015](#)). Additional examples from other initiatives reflecting potential of rights statements metadata are provided in Appendix Two ([See below](#)).

To standardize the content of rights statements associated with digital resources available over the global Web, some initiatives have made calls for a universal common namespace and URI design in the “rights” core elements of metadata that would allow further specifications, both machine- and human-readable, for the different scientific communities. Examples include the DPLA-Europeana efforts toward establishing international & interoperable rights statements ([ibid](#)) and the National Information Standards Organization (NISO) Access and License Indicators project “aimed to determine the optimal mechanisms to describe and transmit the right, if any, an arbitrary user has to access a specific article from any internet connection point” ([NISO, 2014](#)). This emerging area is likely to be fertile for further work in the research data community.

In addition to the minimal requirement of using appropriate metadata to articulate the rights status of a given research dataset, data providers concerned with machine readability of rights information may also wish to consult the various Rights Expression Languages (REL) emerging for use over the World Wide Web. Rights Expression Languages aim to encode restrictions on the use of content and provide formal, machine-readable expressions of copyright usually through creating a controlled vocabulary of verbs standing in as restricted actions. One example is the World Wide Web Consortium's [Open Digital Rights Language, ODRL](http://www.w3.org/TR/odrl/) (<http://www.w3.org/TR/odrl/>).

[] Guideline 5A. Any disseminator of research data set(s) should incorporate metadata containing a rights statement necessary to assure legal interoperability of those data.

The rights statement should include:

- a set of standard expressions that, at a minimum, distinguish between public domain and copyrighted content, and indicate whether the data are freely accessible or only with payment; and
- a URI with a standard code to address international use and translation issues (for example, ISO's [Reference Codes for the Representation of Names of Languages](#) and their [Country Codes](#))

[] Guideline 5B. Legal interoperability is enhanced with 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.

The research data community should devise a set of rights statements, both human and machine readable, and for this task should draw on the prior work of Europeana, DPLA and NISO.

Principle Six: Attribution and Credit

Reaffirm the value of giving credit. Attribution and citation are an essential part of the research process for providing appropriate credit, especially when data from different sources are incorporated into a dataset or database. They should be encouraged through community norms rather than through legal requirements.

Definition and Discussion of Terms

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

6. ATTRIBUTION AND CREDIT: GUIDELINES FOR IMPLEMENTATION

Appropriate attribution and credit for creating or contributing to a well-prepared, quality dataset is essential to the research enterprise. There is a difference of opinion though about whether to make attribution a specifically legal requirement or simply leave it to the norms of the research community. Ethical rules and practices vary from community to community and may often refer to particularities of a defined scientific domain. For these reasons, attribution requirements should be defined by the communities, not by lawmakers.

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. Philosopher of Science George Franck, writing in *Science* (1999), observed:

“Success in science is rewarded with attention. You gain full membership in the scientific community only by receiving the attention of your fellow scientists. Earning this attention “income” is a prime motive for becoming a scientist and for practicing science. “

Researchers 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. “The commodity which scientists traditionally exchange is knowledge or information, and in drawing on the intellectual property of their peers scientists have to enter the exchange system and “pay the going rate”, so to speak. The currency, to maintain the economic metaphor, is the “coin of recognition” (Merton, 1968, p. 56).

Despite its essential place within research communities, however, the right of attribution is not universally regulated by law (Fisk, 2006), especially for information that is not copyrightable. European jurisdictions do protect attribution through various legal measures including moral rights provisions of national copyright statutes (Lastowka, 2007). United States copyright law limits the right of attribution to the narrow case of visual artists who produce creative works as unique or highly limited editions (US Code, Title 17, Section 106A)

The one case where attribution is a matter of legal compliance is with research outputs distributed under a Creative Commons license. All such licenses require attribution to the data producer or provider in a manner s/he specifies. Failure to properly attribute CC licensed works results in termination of the license, exposing the user to the risk of legal consequences (Creative Commons, 2015).

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 (Albert and Wager, 2003; CSE, 2015; ICMJE, 2015). Applying these established guidelines to attribution practices for research data is a helpful starting point for the RDA community

Beyond the norms 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 2009 creation of DataCite has made it possible to formalize the concept of data citation (DataCite Metadata Working Group, 2015). Building largely on the work of DataCite, the Digital Curation Centre has developed a detailed guide on citing datasets and linking datasets to publications (Ball and Duke, 2015). Another notable effort has been CODATA's Data Citation Standards and Practices Task Group whose analyses and reports (CODATA 2013) 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. Citation styles vary, but the Declaration provides general principles and examples to ensure that data can be properly cited (see Force11 example at <https://www.force11.org/node/4771>). This includes providing the necessary elements to give proper credit and attribution for the use of others research outputs, as well as the use

of a unique identifier which allows the data to be located without ambiguity ([Data Citation Synthesis Group, 2014](#))

Good data citation practices are also promoted by most data repositories that include 'Please cite these data as...' statements with the information contained on the landing page associated with the permanent identifier (see examples in [Ball and Duke, 2015](#)).

Additionally, the major challenges to citing very large and dynamic datasets, or subsets produced by a specific query, are being addressed by the RDA Working Group on Data Citation. Their recent recommendations go some way to addressing this barrier ([RDA Working Group on Data Citation, WGDC, 2015](#)).

Despite the marked progress in developing standards for data citation, however, there remain challenges to the practice of giving proper credit and attribution to data creators. The practice is not as widespread as it should be which represents a failure to give proper credit to data creators. References to data are often made inter-textually or in other forms of acknowledgement and do not typically feature in the reference list, with the result that they are not included in citation metrics.

Good practice 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.

Another concern is that the need to establish normative practice for determining and crediting authorship, particularly in such complex cases, is not currently being addressed. Given that the international Committee on Publication Ethics indicates that authorship disputes and issues of questionable authorship account for the second largest group of cases they consult on, it may be important for the research data community to review this important area in the future (for examples of authorship disputes around research outputs including data, see the COPE Database of Cases, <http://publicationethics.org/cases>).

[] Guideline 6A. Creators of datasets, as authors of research outputs, should follow and adapt established guidelines for attributing authorship

Established criteria for assigning authorship in a research work are promulgated by scholarly publishing and editorial advisory groups such as the Committee on Publication Ethics (COPE, <http://publicationethics.org/>), the International Committee of Medical Journal Editors (ICMJE, <http://www.icmje.org/>), and equivalent groups in particular disciplinary domains (e.g., learned societies and professional associations). These groups have a long history of successfully applying these criteria when evaluating and resolving authorship disputes in the scientific research realm and provide a proven approach that may be adapted when attributing authorship of research data.

Guideline 6B. Authors using data in scholarly work should give attribution and credit to the data creators and unambiguously identify the source by using a citation that conforms to the Joint Declaration of Data Citation of Principles.

Guideline 6C. When very large numbers of datasets have been used to create a further data product, or a dataset has resulted from many individual contributions, best efforts should be made to provide full attribution in an external metadata record, while conforming to the limits imposed by citation styles in scholarly communications.

7. EQUITY: GUIDELINES FOR IMPLEMENTATION

Principle Seven: Equity

Promote equitable access and use. 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, and to share the benefits from the use of such data.

Definition and Discussion of Terms

| [Equity](#) | [Equitable access](#) | [Equitable use](#) | [Developing country](#) |

Equity and fairness are important qualities for research conducted with public support, especially in the public interest. There are not many expressions of such a principle in laws and policies at either an international or national level. These generally are not legal requirements, but ethical or normative ones.

There are important ethical considerations, not just legal ones, in determining 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 them. If the research data were collected using public funds, the taxpayer already paid for that 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 above in Implementation Guideline #2 on Balance. 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. Excluding the public is thus 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 within laws of the jurisdiction in which they are located and the norms of the community within which they work. At the same time, the less 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.

One of the most important advantages from open data practices, generally, and from the Legal Interoperability Principles, specifically, are for developing countries and other economically

disadvantaged users. This truism has been pointed out by many research data sharing initiatives, including the 'Nairobi Data Sharing Principles' developed by participants of the CODATA Workshop on Open Data for Science and Sustainability in Developing Countries ([CODATA, 2014](#)) and by the International Council For Science (ICSU)'s Ad-hoc Strategic Coordinating Committee on Information and Data (SCCID) in their 2011 report ([International Council for Science, 2011](#)). The group of less economically developed countries and those with economies in transition typically do not have many of their own means of research data production or data centers, but they particularly can benefit from the applications of those data. The same is true for those individuals or institutions that are economically disadvantaged within countries, even the economically more developed ones.

The open provision of various types of research data makes them available for use by those least able to afford any access fees. 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.

[] Guideline 7A. Researchers and research communities as data creators should refrain from claiming intellectual property rights for public research data and information.

[] Guideline 7B. Data repositories and other institutions that publish or provide access to data should not charge more than marginal cost for accessing them.

These recommendations follow the ethical norms discussed here and are consistent with Recommendation 1 in Implementation Guideline for Principle Two on Balance.

[] Guideline 7C. Waivers of rights that are applied to any research dataset by the rights holder should be made in perpetuity; application of any non-restrictive license, however, should be limited in time, consistent with the statutory term of the underlying copyright or database protection law.

This is an ethical obligation consistent with the discussion above and applies to waivers of rights. For non-restrictive licenses with "some rights reserved" (such as a CC-BY Attribution-Only license), it is an ethical obligation not to exceed the term of any underlying and enforcing statutory intellectual property law (such as copyright or database protection law), after the expiration of which, the licensed dataset will fall automatically into the public domain. This will ensure legal conformity and certainty for the rights of the user of those data.

[] Guideline 7D. Licenses should take into account specific conditions arising for users in less developed regions of the world, in order to achieve equitable access to data.

The specific conditions which create inequitable barriers to data include: excessive or unaffordable cost for access; lack of baseline technical standards, in line with common goals of capacity building; unwillingness to diversify access modes, in recognition of potential connectivity and other technological challenges.

Principle Eight: Responsibilities

Encourage responsibility. All members of the research community need to be fully informed about the laws, policies, and norms pertaining to the data they produce or use, and should engage directly in the applicable legal, policy, and community processes governing research data.

8. RESPONSIBILITIES: GUIDELINES FOR IMPLEMENTATION

It is worth reiterating that ignorance of the law is no excuse. This is true of all people, whether they are working in their own jurisdiction or in another one. In many ways, the familiarity with the norms and practices of one's own areas of research and the data used in that discipline is frequently better than with the laws and policies, even within one's own locale. Yet, obligations and responsibilities of all parties must be governed by perhaps poorly known laws and policies. This is especially true in the emerging environment that is often labeled as "open data." This 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 consumers take advantage of all data, thereby made accessible within the rules.

This Principle is therefore directed at all members of the research community. Every person in the research process has some responsibility to access, use, share, and combine research data, and must understand that responsibility.

[] Guideline 8A. Data providers and data users need to be familiar with and understand all relevant rules and policies relating to the legal status of research data.

At the very least data providers and data users must be aware of these rules and policies and know to access up-to-date versions thereof. The laws and policies in scope include:

- Prevailing laws in the jurisdiction where the work is performed, and where the data and results are published.
- Rules and policies imposed by funding agencies (from public, but also by private sources).
- Rules and policies imposed by their employer.
- Rules and policies imposed by publication channels (e.g., journals, audio-visual programs, and online media).
- Rules and policies of cognizant scientific organizations (e.g., professional societies, large international projects).

It should be emphasized that sources of the vast majority of rules and policies listed above do not necessarily carry the force of law, except of course for the first bullet point above.

[] Guideline 8B. It is the responsibility of data providers and consumers to understand key concepts relating to the legal status of research data.

These concepts include:

- The nature and meaning of relevant legal terms such as “copyright”, “license”, “waiver”, and the like.
- The differences between “rights” and “permissions”.
- The semantic differences between “public records”, “open access data”, and “public domain data”.
- The semantic differences between “public domain accessibility”, “rights waivers”, and “fair-use”.
- The various levels of “plagiarism” and the processes that are in place to adjudicate complaints.

Resources exist that explain these terminologies in layman’s terms, and explain differences between legal environments (e.g. US copyright law vs. EU Database Directive.) As a common example, it is important for data users and providers alike to understand that, just because an item can be found through an internet search engine does not imply that this item belongs to the public domain.

As an example, the IEEE guidelines on evaluating plagiarism are referenced here (http://www.ieee.org/documents/Level_description.pdf). Other organizations (e.g. American Geophysical Union) participate in and rely on guidelines prepared and maintained by the Committee on Publication Ethics (COPE) (<http://publicationethics.org/resources/guidelines>)

[] Guideline 8C. It is the responsibility of data providers and users to take effective steps to make data discoverable, accessible, usable, and legally interoperable.

Usability may be achieved either by use a standard format or by providing tools and software to deal transparently with any non-standard format. In that example, the use of proprietary software to ingest and process data entails a risk of violating the principles of open data access

Making data discoverable and citable can be achieved through use of persistent identifiers (e.g. DOI). Details are of course dependent on the relevant community of practice, and in turn on the nature of the data (e.g., genomic data sets are not expected to be dealt with in a way as satellite remote sensing data, or population survey data.)

Making data legally interoperable can be done by adopting the recommendations above. Most important and easy to achieve in this regard is by assigning where possible rights waivers and open licenses, and clearly expressing the legal status of the research data in human and machine readable statements and metadata.

[] Guideline 8D. Researchers and research communities should engage in the political processes that lead to laws and other norms ruling the access to, the reuse of and the legal interoperability of research data.

It is important that members of the research community engage with the political and legislative processes that relate to the legal interoperability of research data, and that they explore legal frameworks—including enforcement of rules— that will facilitate scientific progress.

They should have a right to be consulted by the authorities before such laws are enacted.

Enforcement of rules can be a legal, policy, or normative action within each data community.

[] Guideline 8E. A well-conceived educational process should be constructed and adopted by relevant institutions, that leads to a better prepared generations of future researchers.

With the aim of achieving a sustainable set of practices in the long term, this educational process should stress responsibility to ensure legal interoperability of research data as part of the responsible conduct of research. As such the educational process might include:

- Formal courses in “Research Data Science” with a comprehensive syllabus, targeted at students at all levels (see the courses run at RPI <http://tw.rpi.edu/web/courses/datascience> and the CODATA-RDA Research Data Science Short Courses <http://www.codata.org/working-groups/research-data-science-summer-schools> and the definition offered of research data science)
- University training in the Responsible Conduct of Research
- Continuing professional education from well-established and widely supported curriculum of research data organizations such as the Data Carpentry Foundation (<http://www.datacarpentry.org/>), the Open Science Framework (<https://osf.io/>), as well as the Research Data Management programs offered by many research libraries and universities such as MANTRA offered by the Data Library at the University of Edinburgh (<http://datalib.edina.ac.uk/mantra/>)
- Seminars targeted at supervisors and mentors
- Seminars and webinars targeted at participants in the publishing process (e.g. editors, reviewers, librarians, data managers, other stakeholders) (For example webinars run by RDA, WDS and other organizations: RDA <http://europe.rd-alliance.org/rda-europe-webinars>; WDS <http://www.icsu-wds.org/community/webinars>; ANDS <http://ands.org.au/events/webinars/>)
- Design and funding of targeted research programs (e.g. NSF Research Traineeship Program, NRT <http://www.nsf.gov/pubs/2014/nsf14548/nsf14548.htm>; Belmont forum <http://www.belmontforum.org/>)

DEFINITIONS AND DISCUSSION OF TERMS

<p>Interoperability of data</p> <p>[Return to Principle One: Access and Reuse]</p>	<p>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. Semantic interoperability depends on shared and unambiguous properties to which data refer allowing comparison or integration at scale. Similarly, as defined in the Introduction above, legal interoperability facilitates the reuse and recombination of research data through waiving proprietary rights and providing clarity about any restrictions.</p> <p>It is widely recognized that the value of data lies in reuse. The ability to 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.</p> <p>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”. In the mathematics of fractions, operations must be performed with the lowest common denominator. Similarly, 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 be subject to unrestrictive (common-use) licenses or waivers of rights.</p>
<p>Open access</p> <p>[Return to Principle One: Access and Reuse]</p>	<p>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:</p> <p>“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)</p> <p>By “open access” to this literature, we mean its free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. As established in the Budapest Open Access Initiative (BOAI), “The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited” (BOAI, 2002).</p>

	<p>Open access to research data derives many principles from the open access movement and from such definitions. A key statement is found in the OECD Principles and Guidelines for Access to Research Data from Public Funding, which offers the following definition of openness and open access to research data:</p> <p>“Openness means access on equal terms for the international research community at the lowest possible cost, preferably at no more than the marginal cost of dissemination. Open access to research data from public funding should be easy, timely, user-friendly and preferably Internet-based”. (OECD, 2007).</p> <p>The <i>OECD Principles and Guidelines</i> have had widespread influence and are mentioned in many research funder policy documents. In recent years, there has been an effort to clarify that for publicly-funded research data, it is not enough simply to make the data available, but that usability must be facilitated. Influenced by the definition of “intelligent openness” advanced in the Royal Society’s Science as an Open Enterprise report, the G8 Science Ministers Statement declared that: Open scientific research data should be easily discoverable, accessible, assessable, intelligible, useable, and wherever possible interoperable to specific quality standards. (G8 Science Ministers 2013). The G8 definition, in turn, is used in the documents presenting and supporting the European Commission’s Guidelines on Open Access to Scientific Publications and Research Data (European Commission 2015). See also Guidelines on Data Management in Horizon 2020, p.6; http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf</p> <p>From each of these definitions it can be seen that “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. The Budapest initiative is more specific about that the types of reuse includes computer assisted processing and analysis at scale</p>
<p>Unrestricted reuse</p> <p>[Return to Principle One: Access and Reuse]</p>	<p>There are widely acknowledged and necessary restrictions which may – and sometimes by definition must – be placed on the reuse of research data. These include, among others, the need to protect personal privacy, issues of national security, and in some instances commercially confidential information. (See Principle 2 and the accompanying Implementation Guidelines, below, for additional information.) All statements and policies on open access to research data acknowledge these limitations.</p> <p>Unrestricted reuse means, therefore, the absence of any restrictions over and above these and similar principles. For example, the Budapest Open Access Initiative statement observes: “The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited.” [http://www.budapestopenaccessinitiative.org/] An analogous principle holds for the assertion of intellectual property rights over data produced by publicly funded research.</p> <p>The OECD Principles and Guidelines argue that where data produced by publicly funded research are protected by intellectual property rights “the holders of these rights should nevertheless facilitate access to such data particularly for public research or other public-interest purposes.” (OECD 2007, p.17).</p>
<p>Rights waivers and non-</p>	<p>The principle of legal interoperability means that data should be clearly labeled with</p>

<p>restrictive licenses</p> <p>[Return to Principle One: Access and Reuse]</p>	<p>conditions of reuse. (See Principles 3 and 5, and the accompanying Implementation Guidelines for additional information.) When data are already in the public domain, this should be clearly stated. Otherwise, a waiver of rights or a non-restrictive (common-use) license should be used.</p> <p>Specific Creative Commons common-use licenses were established to formally allow reuse, subject to certain restrictions (see: http://creativecommons.org).</p> <ul style="list-style-type: none"> • CC-BY (attribution), allows unrestricted use of a work so long as the creator is attributed. • CC-NC (non-commercial) allows uses that are non-commercial, defined as uses that are primarily intended for commercial advantage or monetary compensation. Attribution is also required. • CC-SA (share alike) compels any user to license a derivative product under the same terms of the original work. Attribution is also required. • CC-ND (no derivatives) prohibits the user from changing the licensed work or incorporating it in another work. Attribution is also required. <p>All CC licenses require attribution, reflecting the fundamental principle of giving credit where due in the trusted system of scholarly discourse. (See Principle 6 for further discussion). It is commonly deemed acceptable to impose a legal requirement for attribution in this way. However, it can be argued that attribution is something covered by the norms and practices of scholarly discourse and by an author's moral rights so that such a license is unnecessary. It can also result in the unwieldy accumulation of attributions, known as "attribution stacking".</p> <p>Further, each of the licenses beyond CC-BY imposes reuse restrictions that may create incompatibilities and licensing difficulties for the sharing of research data and for the legal interoperability in derivative datasets. We recommend not to use licenses that impose these restrictions. This advice is consistent with that of the Creative Commons organization itself (http://creativecommons.org)</p> <p>Ideally, then, to achieve the objectives of legal interoperability, data should be placed in the public domain with no restrictions. This can be accomplished by a voluntary waiver of rights by the rights holder of any given dataset. Creative Commons designates such waivers with CC0 (see http://creativecommons.org).</p>
<p>Public domain</p> <p>[Return to Principle One: Access and Reuse]</p>	<p>Public domain status of data is achieved in multiple ways: because the term of copyright protection and all other germane restrictions have expired; because copyright protection has been abandoned (e.g. the so-called orphan works European Union Directive 2012/28/EU http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2012:299:0005:0012:EN:PDF); through dedication of the data to the public domain by public or private data holders of the data via an exclusion in the copyright statute (e.g., section 105 of the 1976 Copyright Act of the United States (http://www.copyright.gov/title17/92chap1.html#105) or a simple declaration (always revocable) of the rights holder through a waiver of rights (e.g., CC0); or By eminent domain, through the statutory mandate of a government (or a regulation or policy of an agency, if that power is delegated) to declare the data are public domain, per se.</p> <p>There are subtle legal differences on what exactly public domain implies depending on those different origins (enforcement, conditioning, revocability...), but each of</p>

Simon Hodson 5/12/15 14:56

Commentaire: Yellow highlight designates need for an internal link.

	<p>these leads to the same common effect: i.e., that the data may be used by anyone, anywhere, anytime without permission, license or royalty payment, and without prejudice to its conditioning to requirements of user registration and/or proper attribution.</p> <p>Public release, disclosure, or dissemination are not synonymous to public domain and should not be used interchangeably. They describe the availability of a work and any relevant statutory restrictions apply automatically. For example, publicly released, disclosed, or disseminated information may be protected by copyright and therefore not be in the public domain.</p> <p>Finally, public domain data might have restricted distribution based on other public interests that could limit accessibility (see Principle 2 on Balance and accompanying Implementation Guidelines).</p>
Intellectual property rights (IPRs)	See individual entries for Copyright , Database rights , Patents , Trademarks
Copyright [Return to Principle Two: Balance Legal Interests]	<p>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/). Today, the field of application is much wider than art and literature. It refers also to technical illustrations, architecture, software and many other intellectual creations.</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 of this content. Research data refer mostly to facts and are not individual creative works. The datasets are often presented in a standardized, community-based form that shows no individuality or creativity, and therefore do not qualify for copyrighted protection (Patterson et al., 2014). Although there is a widespread opinion that copyright can be claimed for research data, this claim is often lacking any legal grounds. Copyright may be applicable, however, to a collection of data, by reason of the selection or arrangement of their contents, and that consequently rise to a status of intellectual creations with individual character.</p>
Database rights [Return to Principle Two: Balance Legal Interests]	<p>Database protection in a legal form of a <i>sui generis</i> ("of its own kind," or unique) right exists mostly in the E.U. (with a few similar applications in other countries) (European Parliament, 1996). 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 or "insubstantial" data point(s). 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 qualify as works in the meaning of copyright.</p>
Patents [Return to Principle Two: Balance Legal Interests]	<p>Patents are granted for inventions applicable in industry. They give the right to monopolize the commercialization of such inventions, 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.</p>
Trademarks, patents for design, and indications of sources	<p>Trademarks protect signs that are capable of distinguishing goods or services of one undertaking against goods or services of other undertakings. Patents for design are granted to new, original designs for products of manufacture. Protection of indications of source guarantees the correct use of direct or indirect references to the</p>

<p>[Return to Principle Two: Balance Legal Interests]</p>	<p>geographical origin of goods or services, including references to their properties or the quality associated with their origin. All these modes of protection do not refer to the access to data, but may exclude certain commercial reuses for a given time period.</p>
<p>Infringement of copyright</p> <p>[Return to Principle Two: Balance Legal Interests]</p>	<p>Violation of any of the exclusive rights of the copyright owner as provided by law represents infringement of copyright (For further information, see the copyright infringement section of US Copyright law (http://www.copyright.gov/title17/92chap5.html) and the "What to Do If You're Accused of Copyright Infringement" section of the WIPO web site (http://www.wipo.int/sme/en/documents/copyright_infringement_fulltext.html))</p>
<p>National security or public order laws</p> <p>[Return to Principle Two: Balance Legal Interests]</p>	<p>National security laws, or more broadly, public order (or enforcement) 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 and that 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 restriction can be found in the field of military defense, intelligence agencies, or ministries and organizations that protect the public safety.</p>
<p>Protection of endangered species</p> <p>[Return to Principle Two: Balance Legal Interests]</p>	<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. (Convention on International Trade in Endangered Species of Wild Fauna and Flora, 1973)</p>
<p>Protection of cultural resources</p> <p>[Return to Principle Two: Balance Legal Interests]</p>	<p>Secrecy may apply to data and information about cultural resources as well. 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. In this case again, open access to research data could, under certain circumstances, compromise the protection of such resources. [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]; 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]; 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.</p>
<p>Protection of indigenous peoples' rights</p> <p>[Return to Principle Two: Balance Legal Interests]</p>	<p>There are Traditional Knowledge (TK) laws in many countries that prohibit the misappropriation of various data and information that have cultural significance. Examples may include various designs and traditional medicines. Some cultures also prohibit taking pictures of designated persons and disseminating them. 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</p>
<p>Protection of genetic resources</p> <p>[Return to Principle Two: Balance Legal Interests]</p>	<p>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</p>

	run against 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 and policies [Return to Principle Two: Balance Legal Interests]	Confidentiality laws can also protect public and private interests by withholding certain data and information. This may refer to administrative consultations in governments or to the production of secrets in the private sector, such as technical know-how, customer information, trade secrets, and many 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.
Defined periods of exclusive use of research data [Return to Principle Two: 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 have a contractual or less formal policy basis, and are expressions of a self-defined institutional policy based on the norms of the community.
Personal privacy protection [Return to Principle Two: Balance Legal Interests]	In numerous countries, data referring to individual persons, such as medical information, data on social status, penal record, financial situation and many others, are treated differently from all other data. 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, public interest). Also, please consider the order of these restrictions. Personal privacy protection and perhaps others may need to be moved up in order of importance or to group certain types of restrictions.
Individual contracts or use agreements [Return to Principle Two: 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, as described above. 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 Three: Transparency and Certainty]	Provision of all 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 Three: Transparency and Certainty]	The term refers to the completeness of the information given on the holder and status of rights of a dataset, which results in legal certainty and the reduction of risk for the user of that dataset.
Rights Status [Return to Principle Three: Transparency and Certainty]	This term is to be understood broadly, including the terms and conditions derived from statutory law, private law, and customary law including ethics, such as codes of conduct in the arts and sciences.
Research Data	The term research data is defined deliberately loosely in this document as its primary

<p>[Return to Principle Three: Transparency and Certainty]</p>	<p>purpose is not to define research data, but to identify conditions that govern the legal interoperability of information/works that are considered research data by researchers. It is important to note also that the conditions and practices that govern legal interoperability of research data are not necessarily specific to information considered to be research data.</p> <p>More generally, the term research data is not clearly defined or very well agreed upon: some would consider all sources and even physical samples research data which is out of scope here. In the context of these Implementation Guidelines, the meaning is construed more narrowly, referring only to information considered by researchers as research data and available in a digital format. If these two conditions are met, the source data and information as well as processed data, individual data as well as data sets organized in databases independent of format, and whether it falls under copyright/database protection or not is considered research data. [See further: The CASRAI Research Data Domain dictionary gives the following broad definition: ‘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; meanwhile, Christine Borgman offers this useful and broad definition “representations of observations, objects, or other entities used as evidence of phenomena for the purposes of research or scholarship”. C.L. Borgman, 2015. Big Data, Little Data, No Data: Scholarship in the Networked World. The MIT Press, p.28.]</p>
<p>Total harmonization</p> <p>[Return to Principle Four: Harmonization]</p>	<p>When several or all jurisdictions or members of a distinctive scientific community use a common standard—whether legally binding or normative—it can be:</p> <p>(a) 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</p> <p>(b) the result of an agreement within a scientific community.</p> <p>[Examples to endnotes: 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 http://www.scar.org/data-products/scadm; for harmonization between the 28 EU countries, the EU Inspire Directive http://inspire.ec.europa.eu/; 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 generation http://web.ornl.gov/sci/techresources/Human_Genome/research/bermuda.shtml.]</p>
<p>Mutual recognition</p> <p>[Return to Principle Four: Harmonization]</p>	<p>Mutual recognition (or “national treatment”) 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). [Examples to endnotes: See, e.g., the so-called "Cross-Strait Intellectual Property Rights Protection Cooperation Agreement" (the IPR Agreement) on June 26, 2010, http://www.mac.gov.tw/ct.asp?xItem=85847&ctNode=6256&mp=3. Mutual</p>

	recognition was the first tool used to harmonize IPR rules until the Berne Convention of 1886 was signed, and is still used mainly in bilateral agreements to harmonize enforcement techniques.
Partial harmonization [Return to Principle Four: Harmonization]	<p>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 re-use of data). [Examples to endnotes: For example, see, e.g., the Electronic Chart Display and Information System (ECDIS) for nautical charts, under which governments allow only visualization/display but not reuse of most of the data –based on safety reasons http://www.ecdis-info.com/.]</p>
Alternative harmonization [Return to Principle Four: Harmonization]	<p>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. [Examples to endnotes: See, e.g., the potential use of DiGIR, BioCASE or TAPIR in the Global Biodiversity Information facility, GBIF, http://www.gbif.org/resource/80635; or the alternative CC licenses.]]</p>
Optional harmonization [Return to Principle Four: Harmonization]	<p>A standard is available to everybody, but is not mandatory; it is up to the data provider 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. [Examples to endnotes: 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.]</p>
Fair use and fair dealing [Return to Principle Four: Harmonization]	<p>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 EU) jurisdictions, open-ended provisions in statutes listing limitations and exceptions to copyright, 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</p>
Norms [Return to Principle Four: Harmonization]	<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. 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, p. http://www.cmsimpact.org/sites/default/files/documents/code_of_best_practices_in_fair_use_for_arl_final.pdf)</p> <p>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</p>

	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)
Berne three-step test [Return to Principle Four: Harmonization]	As first enacted by article 9(2) of the Berne Convention for the Protection of Literary and Artistic Works in 1967, member countries may include in their national laws a provision for limited copying of protected works without permission, subject to the following three step test as follows : "It shall be a matter for legislation in the [different] countries (...) to permit the reproduction of such works [a] in certain special cases, provided that [b] such reproduction does not conflict with a normal exploitation of the work and [c] does not unreasonably prejudice the legitimate interests of the author." http://www.wipo.int/treaties/en/text.jsp?file_id=283698
Metadata [Return to Principle Five: Metadata]	Metadata are the structured descriptions of data sets and data services that facilitate their discovery, assessment, inventory, and use. In the case of these Implementation Guidelines, the focus of the metadata discussed here concerns the legal status of the dataset being described.
Rights statement [Return to Principle Five: 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. [Link to endnote: Based on Europeana-DPLA, May 2015 "Recommendations for the Technical Infrastructure for Standardized International Rights Statements", International Rights Statements Working Group.]
Namespace [Return to Principle Five: Metadata]	A prefix added to XML element names to prevent conflicts when XML documents from different XML applications are combined.
Acknowledgement [Return to Principle Six: 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: 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.
Authorship [Return to Principle Six: Attribution and Credit]	The International Committee on Publication Ethics (COPE) recognizes that there is no universally accepted definition of Authorship; this group defines authors as "the creator or originator of an idea (e.g., the author of the theory of relativity) or the individual or individuals who develop and bring to fruition the product that disseminates intellectual or creative works (e.g., the author of a poem or a scholarly article)" (What constitutes authorship? COPE Discussion Document, COPE Council 9 June 2014, http://publicationethics.org/files/Authorship_DiscussionDocument.pdf). Building on that core definition, the Council of Science Editors has further elaborated that "Authors are individuals identified by the research group to have made substantial contributions to the reported work and agree to be accountable for these contributions. In addition to being accountable for the parts of the work he or she has done, an author should be able to identify which of their coauthors are responsible

	for specific other parts of the work. In addition, an author should have confidence in the integrity of the contributions of their co-authors. All authors should review and approve the final manuscript." (Council of Scientific Editors, 2012, "2.2. Authorship and Author Responsibilities", White Paper on Publication Ethics, http://www.councilscienceeditors.org/resource-library/editorial-policies/white-paper-on-publication-ethics/2-2-authorship-and-authorship-responsibilities/)
Citation [Return to Principle Six: 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) of enabling others to identify and locate those sources used in the research.
Credit [Return to Principle Six: Attribution and Credit]	Formal recognition of the contribution made by an individual or group to a research output.
Plagiarism [Return to Principle Six: Attribution and Credit]	As defined by the US 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 Seven: Equity]	The Oxford English Dictionary defines equity as "the quality of being fair and impartial."
Equitable access [Return to Principle Seven: Equity]	According to an International Council for Science (ICSU) Advisory Note by its Committee on Freedom and Responsibility in the conduct of Science (CFRS), on "Sharing Scientific Data, with a Focus on Developing Countries (2011): "Article 27 of the Universal Declaration of Human Rights affirms that: 'Everyone has the right to (...) share in scientific advancement and its benefits.' This includes universal and equitable access to scientific data, which international organisations should strive to ensure for scientists, notably in developing countries." It goes on to say that: "Research funders should provide for full and open access to data at the lowest possible cost, preferably free and online, so that scientists and other users in developing countries can afford to access it and share their own data." Sharing Scientific Data, with a Focus on Developing Countries (2011) http://www.icsu.org/publications/cfrs-statements/data-sharing/
Equitable use [Return to Principle Seven: Equity]	We define equitable use or reuse in the context of Principles 1 and 2, above. That is, we encourage "open access" and "unrestricted use" of public research data with the fewest limitations possible, but understand that there is a balance with other countervailing and sometimes superseding values that may be required by law.
Developing country [Return to Principle Seven: Equity]	According to the World Bank a developing country "is one in which the majority lives on far less money - with far fewer basic public services - than the population in highly industrialized countries". And, among other criteria, "A developing country may be one: (...) Where the populace is often hungry and sorely lacks education, where there is a large knowledge gap and technological innovation is scarce; Where health and education systems are poor and/or lacking and where transportation, potable water, power and communications infrastructure is also scarce".

REFERENCES CITED

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/>. Last accessed 12-6-2015.

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>. Last accessed 12-6-2015.

Creative Commons, 2015, "Attribution 4.0 International License" (web page), <http://creativecommons.org/licenses/by/4.0/legalcode#s3a>. Last accessed 12-6-2015.

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 Citation Synthesis Group, 2014. *Joint Declaration of Data Citation Principles*. Martone M. (ed.) San Diego CA: FORCE11. Available online: <https://www.force11.org/datacitation>.

Digital Curation Center and Research Data Alliance, Metadata Standards Directory Working Group, 2015, "Metadata Directory" (web page), <http://rd-alliance.github.io/metadata-directory/>. Last accessed 12-6-2015.

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, 2015, *Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020*, Version 2.0. Available online:

[rhttp://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf](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>. Last accessed 12-6-2015.

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. Last accessed: 12-6-2015.

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>. Last accessed 12-6-2015.

International 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. Last accessed 12-6-2015.

International Organization for Standardization (ISO), "Country Codes (ISO 3166)" (web page), http://www.iso.org/iso/country_codes. Last accessed 12-6-2015.

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/wp-content/uploads/sites/2/2015/04/Liber_DeclarationA4_2015.pdf

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/>. Last accessed 12-6-2015.

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>

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. Last accessed 12-6-2015.

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

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>. Last accessed 12-6-2015.

United States Code, Title 17, Section 501, "Infringement of Copyright" (web page), <http://www.copyright.gov/title17/92chap5.html#501>. Last accessed: 12-6-2015.

World Wide Web Consortium, 2015, "Open Digital Rights Language--ODRL" (web page), <http://www.w3.org/TR/odrl/>. Last accessed 12-6-2015.

APPENDIX ONE

Discussion of mechanisms for Harmonisation (Principle Four)

1. A rights holder can voluntarily use a Creative Commons (CC) waiver of rights (such as the “no rights reserved” CC0), or an “Attribution Only” (CC-BY) non-restrictive, common-use license, both of which are described in the Implementation Guidelines for Principle One [add link]. These legal instruments have “human” (i.e., lay person), “machine” (automatic, machine-readable tags), and formal “legal” deeds (in legal language used by lawyers) that use standard language and are recognized worldwide. These type of licenses, with their additional five variations (not recommended for datasets) of CC licenses (see <http://creativecommons.org/licenses/>) provide bottom-up harmonization, since the standards are interpreted in the same way almost in 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).

This CC0 waiver and CC-BY license can provide a workable scheme for self-governed by scientific communities that harmonizes even better than the top-down statutory law or national policy schemes (see, e.g., 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>).

2. 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 re-use 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/>).

3. 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.

4. Some scientific communities achieve harmonization by recommending or adhering to standardized clauses in agreements (of special value in the data distribution chain) as it is the case with the relatively

frequent liability clause used by databases such as, e.g., the clause use by data providers and users in the Biobanks data base (<http://www.biobank.org/>), which states:

Both parties acknowledge and agree that the data is being supplied with no warranties, express or implied, and XXX expressly disclaims any warranty of merchantability, fitness for a particular purpose, or non-infringement. Neither party makes any representation that the use of the data will not infringe the patent or proprietary rights of any third party.

In no event shall either party be liable for any indirect, incidental, special or consequential damages arising out of or in connection with this agreement whether or not that party has been advised of the possibility of or is otherwise on notice of such possibility.

5. Harmonization can be also built on norms instead of legal rules/agreements/licenses/waivers (see, e.g., the previously cited Human Genome Project agreement on DNA digital sequences flowing into the public domain within 24 hours)

6. Harmonization should be complemented to the maximum extent possible with metadata standards (see Principle Five) and dynamic on-going and more-or-less formal processes for discussion and decision-making as a supplementary tool to maintain the existing standards and promote new ones.

7. Harmonization through State action (top-down approaches) has also proven to be efficient, although it is likely to be more controversial and difficult to achieve consensus, and therefore more time consuming. It also is dependent on the willfulness or the capacity of the scientific community to actively engage in the political process. Where national jurisdictions openly diverge, the negotiation of consensus treaties is a possible recourse (see, e.g., the IPRs on databases, which are based on the level of creativity in the U.S., the level of skill needed in Canada, the levels of industriousness required in Australia, and mere economic investment in the E.U., respectively).

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 Group on Earth Observations process, with its 3 Data Sharing Principles (GEO, 2014b) and 10 Data Management Principles (based on 5 foundational elements).

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).

8. 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, the approach to actual harmonization in the real world is still very soft or does not actually exist.

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 U.S. Courts, 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.

9. Well-designed openness scorecards might lead to the adoption of best practices or collective self-help codes when the indicators are not disputed and the final scores create normative peer pressure (see, e.g., the “PSI [public sector information] Scoreboard”, <http://www.epsiplatform.eu/content/european-psi-scoreboard>, for the global analysis, the open data index, <http://www.programmableweb.com/news/open-data-scorecard-new-research-shows-governments-not-open-enough/2013/10/28>). Nevertheless, metrics systems like the scoreboards can be more subject to criticism than consensus.

POSSIBLE APPENDIX TWO: EXAMPLES OF METADATA STANDARDS WITH RIGHTS ELEMENTS

The Data Documentation Initiative used with social science data contains the field <copyright> 2.1.3.2 (<http://xml.coverpages.org/ddiTagLibraryAll.html>). The Dublin Core metadata scheme commonly used with digital library and Web resources also includes a rights field <dc.rights> (<http://dublincore.org/usage/decisions/2004/2004-01.Rights-terms.shtml>). The PREMIS metadata standard applied in digital preservation to convey administrative information necessary to steward digital resources over time includes the rights element, with recommendations to populate it (see <http://www.loc.gov/standards/premis/>, and point 3.5 of Priscilla Caplan, “Understanding Premise”, Library of Congress, 2009, available at <http://www.loc.gov/standards/premis/understanding-premis.pdf>). Additionally, the Data Catalog Vocabulary (DCAT) (<http://www.w3.org/TR/vocab-dcat/>), Description of a Project (DOAP) (<https://github.com/edumbill/doap/wiki>), or myExperiment base Ontology, (<http://rdf.myexperiment.org/ontologies/base/>), also incorporate a rights field. The Australian National Data Service (ANDS), for instance, has utilized this additional information to allow filtering of data by license (<http://guides.ands.org.au/rda-cpg/rights>).