Legal Options for the Exchange of Data through the GEOSS Data-CORE

Draft White Paper—For Comment Only

Legal Interoperability Sub-Group[1]

Data Sharing Task Force

Group on Earth Observations

A. INTRODUCTION

1. Organizational Aspects and Definitions of Key Concepts

The Group on Earth Observations (GEO) is a voluntary, legally nonbinding consortium of Member countries and Participating Organizations (mostly not-for-profit and some for-profit) that seeks to promote human welfare in nine "societal benefit areas"[2] through the Global Earth Observation System of Systems (GEOSS) Common Infrastructure (GCI). As a "system of systems," GEOSS will make available through its portal data contributed from a variety of

[1] The information contained in this document does not constitute legal representation by the GEO Data Sharing Task Force (DSTF) or its Legal Liability Subgroup. Before using any information in this publication, it is recommended that an attorney licensed in the jurisdiction of interest be consulted for specific legal advice. The DSTF is grateful to its Legal Interoperability Sub-Group members for providing this background white paper. The Sub-Group members are: Paul F. Uhlir, Miles Gabriel, Joanne Irene Gabrynowicz, Jeff Heninger, Puneet Kishor, Harlan Onsrud, Kevin Pomfret, Daniel Quintart, and Glenn E. Tallia. We also wish to express our gratitude to Sarah Pearson, Senior Counsel at Creative Commons, for her comments on drafts of this paper. The views expressed here are those of the authors and not necessarily those of their employing institutions.

[2] According to the GEO document, "The Global Earth Observation System of Systems (GEOSS): 10-Year Implementation Plan" (as adopted 16 February 2005), the nine agreed societal benefit areas are:

- Reduction and Prevention of Disasters
- Human Health and Epidemiology
- Energy Management
- Climate Change
- Water Management
- Weather Forecasting
- Ecosystems
- Agriculture
- Biodiversity

existing earth observation systems, both space and air based and *in situ*, ranging from primary data collection systems to higher level processed data products and associated descriptive metadata. Together, the members of the consortium produce and hold the largest amount of geospatial data resources in the world.

While all GEOSS data systems are owned and operated by the Members, Participating Organizations and others registering resources, the participants can leverage each other so that the overall GEOSS becomes much greater than the sum of its many parts. Such synergy can be achieved and enhanced as each GEO participant supports common approaches designed to make shared observations and products more accessible, comparable, and understandable.[3]

According to the GEOSS 10-Year Implementation Plan (2005), to achieve the consortium's broad goals GEOSS will collectively:

- Address identified common user requirements;
- Acquire observational data;
- Process data into useful products;
- Exchange, disseminate, and archive shared data, metadata, and products; and
- Monitor performance against the defined requirements and intended benefits.

The GEO Members and Participating Organizations are developing technological, semantic, and legal approaches that will promote the major objectives of GEOSS to facilitate access to, use of, and interoperability of their data that are relevant in the nine agreed societal benefit areas. The 2005 GEOSS 10-Year Implementation Plan explicitly acknowledges the importance of data sharing in achieving the GEOSS vision and benefits when it states that: "*The societal benefits of Earth observations cannot be achieved without data sharing*". The GEOSS Data Sharing Principles, also adopted by consensus in 2005, state:

1. There will be full and open exchange of data, metadata and products shared within GEOSS, recognizing relevant international instruments and national policies and legislation.

[3] See the GEO "Strategic Guidance for Current and Potential Contributors to GEOSS" (October 2007).

2. All shared data, metadata and products will be made available with minimum time delay and at minimum cost;

3. All shared data, metadata and products being free of charge or no more than cost of reproduction will be encouraged for research and education.

2. Statement of the Problem in the Context of GEOSS Objectives and Principles

A fundamental feature of GEO is that it is organized as a voluntary, federated system of individually held, but linked, components. GEO itself therefore does not operate any of the GEOSS components nor does it own, possess, or control any of the data. Indeed GEO is not even a legal entity so it is unlikely that it could assert ownership, possession, or control of any data in its own right. The organization therefore also cannot license the data made available through the GCI.[4] Instead, GEOSS will enable data providers (the collectors or generators of data, or the rights holders[5]) to contribute their data sets by registering them through a Components and Services Registry enabling their access through the GEOSS Portal.

Principle 1 of the GEOSS Data Sharing Principles is the most relevant in the context of this white paper and in the goal of achieving legal interoperability, along with technical and semantic interoperability. On the one hand, the principle promotes the "full and open exchange of data" defined in the GEOSS Data Sharing Implementation Guidelines as "data, metadata and products made available through the GEOSS are made accessible with minimal time delay and with as few restrictions as possible, on a non-discriminatory basis, at minimum cost for no more than the cost of reproduction and distribution." On the other hand, the principle recognizes the impact that international agreements, national and sub-national laws and various policies and procedures pertaining to those data that may have on sharing of data, through the GCI or any other mechanism. This inherent tension between the data sharing purpose and goals of GEOSS and such laws and policies that may inhibit data sharing needs to be addressed and resolved through

[4] For clarity purposes, it should be noted that conditions of use posted on the GCI website may very well be enforceable, but the benefits and limitations would accrue to and be enforceable by those specific parties using the portal either as users or contributors. The same would hold true for those agreeing to terms, such as through a click agreement.

legally valid and defensible means that all GEO Members and Participating Organizations can accept.

This background white paper addresses some legal approaches to sharing of data through the GEOSS Data Collection of Open Resources for Everyone (Data-CORE). The GEOSS Data-CORE is a distributed pool of documented datasets[6], contributed by the GEO community under the following principles, as set forth in the 2010 GEOSS Action Plan:

1. The data are free of restrictions on re-use;

2. User registration or login to access or use the data is permitted;

3. Attribution of the data provider is permitted as a condition of use; and

4. Marginal cost recovery charges (i.e., not greater than the cost of reproduction and distribution) are permitted.

It is important to note that (i) user registration, (ii) attribution of provider, and (iii) marginal cost recovery charges for access to the data are not considered restrictions in the context of the GEOSS Data-CORE. Under plain language and in a formal legal sense, however, they would be viewed as restrictions.

The paper focuses on the "legal interoperability" aspects of data made available through the GEOSS Data-CORE because it is essential for the effective sharing of data in GEOSS, which is a priority of the GEO Members. One may define legal interoperability for data as the compatibility of legal rights, terms, and conditions of databases from two or more sources so that the data may be combined and integrated by any user without further permission and without compromising the legal rights of any of the data sources used. Note that the concept of legal interoperability may be applied to the full range of openly available governmental, non-governmental, academic, and commercial data sets. However, we consider the concept here only in the context of databases that also meet the GEOSS Data-CORE Principles.

---

[5] The original collectors or generators of a particular data set may or may not be the rights holders or providers of that data set through GEOSS. For simplicity this paper refers to all of these parties collectively as "data providers".
[6] The term "database" in this paper refers to collections or compilations of data and information. The term encompasses metadata that document and explain the data contained in a database, and also include more highly processed data products.

Many GEOSS Members and Participating Organizations also may be expected to make other data available through GEOSS, but with restrictions on access and re-use that are greater than those allowed in the GEOSS Data-CORE. These legal conditions and approaches of data exchange that are beyond the GEOSS Data-CORE will be explored in a subsequent and separate white paper.

In order to explain the legal basis for any proposed approaches to data sharing in the GEOSS Data-CORE, we begin by providing some background on the legal status of data in the public statutory intellectual property laws that pertain to data and collections of data. The use of different private law instruments (waivers, licenses, and contracts) to either increase or decrease the statutory protections pertaining to any given data set is also explored. We then propose and assess the various legal options for GEO and the GEO Members and Participating Organizations for providing access to their data in the GEOSS Data-CORE through the GEOSS Portal. The paper ends with a set of conclusions and recommendations for broad consideration and consensus adoption of the GEO Members.[7]

B. DATA IN THE STATUTORY LAW CONTEXT

As noted in the Introduction, the GEOSS Data Sharing Principles and their Implementation Guidelines encourage "the full and open exchange of data, metadata and products shared within GEOSS," but subject to "recognizing the relevant international instruments and national policies and legislation." Various laws limit or restrict access, use and re-use of data and information based on a number of countervailing rationales and policies, including the protection of national security, privacy, confidentiality, and intellectual property. It is important to emphasize that when substantial amounts of statutorily protected data are combined from two or more data sources, the new resulting database often will acquire the accumulation of restrictive rights from the sources used.

This white paper is concerned only with the data and databases that will be made accessible through the GEOSS Data-CORE in the GEOSS portal and the legal mechanisms that should be

[7] A Summary of this white paper was submitted for review and consensus adoption by the GEO Members in the 2011 GEO Plenary.

considered and may be used to make those data and databases available globally on terms that are consistent with the GEOSS Data-CORE. The presumption is that the data providers will themselves take appropriate measures to restrict access and use of data that may be protected under other laws and policies.

1. Statutory Intellectual Property Laws that Protect the Rights Holder and Restrict the User of Information

There are two main types of intellectual property legislation, copyright and database protection rights, that are especially pertinent in the context of this paper. Other statutory protections that may have some applicability in certain circumstances in some jurisdictions—such as patent law, trade secret law, commercial misappropriation, and trespass—are not considered here.

a. Copyright

At the outset, it is important to understand that there is no such thing as an "international copyright" that automatically protects rights in creative content on exactly the same basis throughout the world. Such protection depends on the national laws of each country and their interpretation in the courts and other mechanisms for dispute resolution. [3] [expand] **[Need to discuss Berne and exemption of facts from copyright protection, and revise.]**

Data range from individual facts or uncorrected "raw" observations, such as the kind that are streamed from automated sensors, to various levels of interpreted data that have resulted from analysis, including visualized depictions in graphs, images, maps or computer simulations. Under traditional copyright law, a specific datum, such as an observation or description of a nucleotide sequence, is a fact not subject to copyright. Therefore, absent any other protection, it may be used, re-used, or re-disseminated by anyone for any (otherwise legal) purpose, once legally accessed.

However, data sets, databases, and other collections of facts may be subject to automatic copyright protection (i.e., the protection does not need to be expressly claimed or requested) in

whole or in discrete parts as "compilations" of information, even if they consist entirely of individually non-copyrightable facts, if their "selection, coordination, or arrangement" is achieved through some human creativity or originality. Thus, the classification, coding, formats, and interpretations of data in a compilation may be presumed to be covered by copyright. Compilations of facts and their ancillary information in this category are generally interpreted to have "thin" copyright that protects only against wholesale, verbatim copying. Compilations, particularly of factual material, that are arranged for ease of use, or to comply with standards in some disciplinary or business context, or in some obvious, routine, or mechanical ways, generally are not protected by copyright.

Finally, some jurisdictions, such as Australia, have so-called "sweat-of-the-brow" laws that apply copyright based on the effort and investment in compiling the database, while still others have no such laws or have expressly rejected such a basis for protection of unoriginal and uncreative factual contents.

b. Database Protection Laws

In addition to copyright, a major statutory form of exclusive property rights protection of databases or "collections of information" is the 1996 *Directive on the legal protection of databases*, which has been enacted in the national legislation of all EU Member States and Participating States.[8] Several other countries (e.g., Mexico, South Korea) have adopted similar legislation. Such laws protect the information compiled in databases, even mere facts that form more than an "insubstantial part" of the database, defined either quantitatively or qualitatively, as long as the database is the result of a "substantial investment".[9]

We do not analyze here the legal merits of an exclusive property right that protects mere investment in factual compilations.[10] What is important to understand in the context of this paper is that such database protection legislation confers additional statutory rights to data providers,

[8] Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases. Available at: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31996L0009:EN:HTML

[9] Id., Section ____.

[10] For an analysis of the problems posed by exclusive rights protection of factual compilations, particularly in the context of public-sector and publicly funded research data, *see* Reichman, J.H., and Paul F. Uhlir, *Database Protection at the Crossroads*, Berkeley Tech. L. J., 1999; and National Research Council, A QUESTION OF BALANCE: PRIVATE RIGHTS AND THE PUBLIC INTEREST IN SCIENTIFIC AND TECHNICAL DATABASES, National Academy Press, 1999.

which they can use to enforce their license provisions (as discussed further below) in those jurisdictions that have enacted such legislation.

2. Limits to Statutory Intellectual Property Laws

The extent of protection of databases by treaties and legislation is also defined by what is *not* protected—the public domain *yin* to the proprietary *yang*. There are also various statutory limitations and exceptions that further limit the scope or application of protection in favor of different user groups or types of uses in order to promote diverse competing public interests.

a. Public Domain Status

The public domain may be defined as encompassing content that is (1) not subject to copyright or related rights (including database protection rights), and (2) not subject to conditions on reuse imposed by other means.[11] The public domain may be created formally by public laws through national legislation or regulation that expressly excludes certain categories of data and information from copyright or from other exclusive property protection, or by private-law waivers of rights. Public domain status may also be attained when the protection of the databases has exceeded the statutory term of protection (which is unlikely for almost all data made available through GEOSS), or by exclusions of certain subject matter from protection, such as facts. As noted above, rights under copyright or database protection laws arise automatically (i.e., they do not have to be claimed by a copyright filing or statement), unless expressly excluded or waived. Hence, express legislative or regulatory action is needed, or a waiver of all rights through a private law alternative (see Section C below) to make the data excluded or waived from protection, or to make the re-use and re-dissemination of data unrestricted.

As a matter of public policy, the period of protection conferred by intellectual property laws on rights holders is limited in time. Once the information exceeds the statutory time period of protection it enters the public domain and is no longer protected by that statute. Information that

is in the public domain and is legally accessed can be used without restriction and without attribution of the rights holder.[12]

In addition to the expiration of the term of statutory protection, public domain status may be achieved by several means. One is the statutory exclusion of a class of producers of creative works. A notable example of this is the placement of all works by the U.S. federal government and its employees in the scope of their employment in the public domain.[13] The public domain may also be created through a class of information (such as non-copyrightable facts in databases, discussed above), the explicit transfer of the information from the owner to public domain status by a waiver of all rights (as discussed further below), or by the failure of a government to enact copyright legislation. With regard to the latter instance, while copyright laws exist in most countries, there are some jurisdictions where such protection does not exist. None of these countries is currently a GEO Member, however.

In reality, databases that are not original or compiled by a sole source typically contain data aggregated from a mixture of data sets from different providers, some perhaps partially protected by statute or license and others perhaps unprotected, which is discussed in more detail below.

b. Limitations and Exceptions

All copyright protection statutes also allow for some limitations and exceptions for the users of copyrighted material. Limitations and exceptions can be based on the status of the user, the type of use, its extent, the type of content, or other factors. In the United States, the main set of limitations is referred to as "fair use," and in many other countries they are known as "fair dealing". [expand and reference]

Because limitations and exceptions to either copyright or the database protection right are narrowly drawn, situation-dependent, and inherently uncertain in their application, we do not

[11] Private communication from Sarah Pearson to Paul Uhlir, 1 September 2011.
[12] It should be noted that in many jurisdictions, however, the "moral rights" of the author, or *droit d'auteur*, applies indefinitely and attribution is required, although this is unlikely to be the case with factual compilations or databases that were protected originally by "thin" copyright, or not at all.
[13] United States Copyright Act (1976), 17 U.S.C. section 105.

find them suitable for providing a legally suitable solution for meeting the GEOSS Data-CORE requirements.

3. Statutory Intellectual Property Law in the Context of the GEOSS Data CORE

As explained at the outset, the GEOSS Data-CORE seeks to provide easy and open availability of data held by GEO Members and Participating Organizations and made available by them through the GEOSS portal, with no restrictions at all on re-use. Collections of data in the public domain fully meet these conditions.

Collections of data in databases that are protected to varying degrees by copyright statutes have a less certain status, unless their legal terms and conditions are specifically explained (e.g., in an accompanying license or metadata). As noted above, facts, such as those observed and collected by automated sensors in databases, are not copyrightable, so they may be extracted, re-used, and re-disseminated by users who lawfully access them, unless further protected by a restrictive license or contract. However, if the databases made available through the GEOSS Data-CORE have some original or creative selection and arrangement, or other information in them is copyrightable, their re-use and re-dissemination may constitute an infringement, absent a specific authorization of the user by the data provider to do so, or an express waiver of the providers' rights.

Even more problematic is the statutory *sui generis* database law that goes beyond copyright to provide an exclusive right in more than "insubstantial parts" of compilations of information, even of otherwise non-copyrightable factual data that are the result of a "substantial investment." Since the user may not know if there was a "substantial investment", what is deemed to be a substantial investment, or what parts of the database the provider deems "substantial", "either quantitatively or qualitatively," there is legal uncertainty and the potential for infringement with the extraction and re-use of more than a small amount of facts from a database that is covered by such a statutory right.[14] As in the case of copyrightable portions of any given database, the provider needs to either expressly authorize the user to re-use and re-disseminate the data

[14] Reichman and Uhlir, *op. cit.*, note 9.

consistent with the operating principles of the GEOSS Data-CORE, or waive the provider's rights under the law.

Thus, the questionable applicability of the statutory law protections to databases and their subsets accentuates the uncertainty of the actual scope of protection or the possibility of infringement by the user. Sometimes, even lawyers who are expert in this field will disagree on the scope of the application of the law, so non-experts are much less likely to understand this or even to be aware of their rights and responsibilities. Moreover, the institutions in which the data users work frequently take a risk-averse position to the use of databases, which assumes that all the contents of the database are protected, even if they are not, leading to high barriers and associated transaction costs for socially beneficial re-use and re-dissemination of the data resources.[15]

This uncertainty and risk of legal dispute is compounded by the global nature of GEO and GEOSS, and the breadth of the relevant data and information types. The inter-jurisdictional transfers and the complexities of the data and their many different uses make the untangling of the legal rights and responsibilities especially vexing for the legally responsible user. This is why it is important to make the data available through the GEOSS Data-CORE with simple, known, and described terms and conditions that enable and encourage the socially beneficial data access and re-use that are the key drivers of GEO and GEOSS.

C. THE USE OF PRIVATE LAW WAIVERS, LICENSES, AND CONTRACTS FOR DATABASES

As we outlined in the preceding section, data and all other forms of information are automatically subject to existing legislative and regulatory requirements and restrictions, including intellectual property rights conferred by copyright and database rights. The application of IP protection, however, is unsatisfactory to many producers and users for a number of reasons.

[15] For a discussion of the effects of such artificial legal barriers, *see* Uhlir, Paul F., and Peter Schröder, *Open Data for Global Science*, Data Science Journal, CODATA, Paris, 2007.

The laws provide a one-size-fits-all protection that is too strong for some and too weak for others. There is uncertain application in scope of coverage for factual compilations (databases), even within one jurisdiction. The public laws vary significantly across jurisdictions and types of databases. Because of these deficiencies, the laws encourage non-compliance by many users and encourage producers to turn to more flexible and responsive *private law* solutions in the form of waivers, licenses, and contracts. Digital networks provide the means to implement private law options easily, cheaply, and with greater certainty.[16]

Moreover, although public-domain status is the best legal option for promoting the various social benefits and goals intended by GEO through the GEOSS Data-CORE because it enables the unrestricted re-use, re-dissemination, and legal interoperability of data, a statutorily created public domain is limited as well. It is not broadly implemented for public sector data and waiting for expiration of statutory IP protection is not a good option.

The focus in this section therefore is on "public domain" and "attribution only" conditions created through private law instruments—waivers, licenses, and contracts—consistent with the terms and conditions of the GEOSS Data-CORE. Because the discussion here is limited to the GEOSS Data-CORE, we do not examine other conditions of common-use (e.g., non-commercial, share-alike, or copyleft uses) or restrictive licenses and contracts that have restrictions on data users greater than those allowed by statute.

1. Waivers, Licenses, and Contracts Explained

Waivers are an express written statement by the rights holder that no statutory or other rights are retained by that rights holder in the database or other information product. A waiver is a private law dedication of the database to the public domain, with no rights reserved. This is the most favorable condition for the user of the database, since it provides equivalent status to the statutory public domain and allows complete freedom for any user to integrate, re-use, re-disseminate all or a portion of the database. It provides full interoperability with no restrictions whatsoever. It retains no protections for the database provider, however, including no legally enforceable attribution or any other requirement. The lack of a legally enforceable attribution requirement may not have much practical effect in most cases, since attribution and citation are

[16] (Cite: power of the two-party deal) [to be added]

normative and ethical practices anyway. Also, many jurisdictions do not allow the waiver of all rights, since the author's moral rights, if applicable, cannot be waived.

Licenses and contracts are used if the database provider wishes to retain some rights and control the use(s) of the data in some way. There is a popular misconception, however, that licenses and contracts are the same thing. They are not.[17]

Licenses are based on existing statutory rights for enforcement. They are applied automatically and do not depend on "agreement" between the rights holder and the user(s). They do not extend to facts or materials already in the public domain, because there is no underlying statutory protection for that material, but can extend to databases or protectable portions of databases, although the uncertainty of enforcement remains. Finally, licenses can be used to decrease or increase level of protection, based on what the database rights holder wants. Decreased protection creates what may be referred to as "common use" conditions, while increased protection confers added protection to the database rights holder through user restrictions over and above the level of statutory intellectual property or exclusive rights protection.

Unlike licenses, contracts are based on the express agreement of the parties. Contracts require formal offer, acceptance, consideration, and (usually) written terms. Formal offer and acceptance for databases and other digital information products are made with "click through" agreements online or "shrink wrap" agreements on CDs and other physical media. Unlike licenses, contracts are not dependent on their enforcement for an underlying statute, although of course they must not be made for an illegal purpose. Also unlike licenses, they can apply to data otherwise unprotected by statute (e.g., factual material in the public domain). Contracts are only valid for the agreeing parties, so others who may obtain the data(base) are not bound by the terms of the original agreement. This makes contracts susceptible to leakage and they can therefore be an uncertain mechanism for rights holders. Finally, contracts and agreements are not standard, unlike licenses, and frequently are long, confusing, and ignored by the user. An example of a restrictive contract is the familiar End User Licensing Agreement (EULA) that accompanies most commercial software or databases.

[17] The discussion of the distinctions between licenses and contracts is based on a presentation by Sarah Pearson at the National Research Council symposium on Developing Data Attribution and Citation Practices and Standards, August 23, 2011, Berkeley, CA; available at:[to be completed],

Examples of waivers and licenses are provided in section C.3 below.

2. The Use of Waivers and Licenses for Data Compilations in the GEOSS Data-CORE

From the perspective of meeting the requirements of access and re-use in the GEOSS Data-CORE, the most compatible legal status is the public domain. In public law this can be accomplished either with formally excluding the databases from copyright or exclusive property protection of other legislation, or, in the much less likely situation for data in GEOSS, the protection of the databases has exceeded the statutory term of protection. In private law, this can be accomplished by an express waiver of rights by the rights holder.

As pointed out by Thinh Nguyen, former counsel for the Science Commons, public domain status is the best option to implement the following goals.[18] The data are not restricted in their re-use, or re-disseminated to anyone. The data are fully legally interoperable, in that they can be combined without any restrictions from all public-domain sources. There are low transaction costs and administrative burdens. There is legal certainty in the use of the data without fear of infringement by the user. And data in the legal public domain is consistent with the community expectation and use, in this case, in the context of the GEOSS Data-CORE.

The downside, however, is that database producers, even in the public sector, will not have sufficient incentives to release their data with no protection, unless this is part of their mission in the public sector or part of their business plan in the private sector. Database producers may make only their least valuable data available under pure public domain conditions or withhold data completely. The balance of producer and user rights is a policy decision for GEOSS Data-CORE participation, as with other data release decisions.

In general, the simplest case of legal interoperability is if many producers in the world distributing data impose the fewest restrictions possible by using the same waiver or license. By having minimal restrictions, conflicting interpretations of those restrictions in different jurisdictions are minimized. A less simple case is where a small subset of open access data licenses might be used, yet still be potentially interoperable where the most stringent conditions

[18] Nguyen, Thinh [forthcoming]. The Web Enabled Research Commons: Applications, Goals, and Trends, in *Designing the Microbial Research Commons*, Paul F. Uhlir, ed. National Academies Press. Washington, D.C.

in each license may control the use conditions of the resulting derivative data set or product. The least favorable condition is the use of non-standard custom licenses or contracts that make the resolution of rights and legal interoperability most difficult.

More specifically, in the organizational context of GEOSS, many users of geospatial data work with more than one data set, typically mixing one or more data sets with their own data. Moreover, many potential users of GEOSS data will not be end users, but re-users or re-disseminators of the data they obtain from other sources. When data from databases with different licenses are mixed or integrated, a new database is created, but the legal terms and conditions, to the extent they are applicable, are transferred with the data that are used from each database. The use and re-use conditions of the resulting database become as restrictive as the most restrictive license of the component data.[19] The restrictions of the component data sets also accumulate, which means that they all apply. In many instances these multiplying restrictions may conflict with each other, creating a non-viable legal status for the resulting dataset. Under certain conditions, while it may be possible to legally acquire certain data, re-using them or mixing them together might be a violation of the terms of one or more licenses, thereby restricting the value of those data in promoting the nine societal benefit areas of GEOSS, and other social benefits more generally.

There are many kinds of standard licenses, ranging from all rights reserved under any applicable statutory law plus other restrictions by the provider, to no rights reserved, or with just some rights reserved between the two extremes. Moreover new, custom licenses can be created by any provider with any mix of terms and conditions.

It also is important to note that transferring data under a license or other data sharing agreement involves more than a transfer of intellectual property rights. It is also a means by which parties allocate the risk associated with such matters as liability compliance with laws, privacy and national security, liability. Therefore, failing to specifically address these issues in a license or data sharing agreement does not make the issues go away. Rather, it simply means that the parties have chosen to let others (courts, legislatures, regulators) decide how the risk is allocated.

[19] Hanson, Chris, Lalana Kagal, Tim Berners-Lee, Gerald Jay Sussman, and Daniel J. Weitzner (2007). Data-Purpose Algebra: Modeling Data Usage Policies, *IEEE Policy.* Available at: http://dig.csail.mit.edu/2006/Papers/Policy07/data-purpose-algebra.pdf

It is easy to see how these facts work together to hamper legal interoperability, and the ability of others to use or re-use data. One way to prevent this from happening would be to agree on a set of specific, restriction-free waivers or licenses for all the databases contributed to the GEOSS Data-CORE. That would ensure that different data could be integrated, re-used, and re-disseminated without any potential infringement problem. The voluntary association of the GEO Members and Participating Organizations, however, does not allow for on the imposition of a mandatory waiver or license for use by all GEO participants. Nevertheless, if GEO does not encourage the use of any such standard instruments, there is a danger that data providers will use any license they want, including their own custom licenses, without completely realizing the detrimental impact of their choice for GEO societal benefit areas.

An intermediate option, that we believe would also be strategically acceptable, is to encourage, but not mandate, adoption of a waiver or license, or terms and conditions from a small set of carefully vetted waivers or licenses. Such private law instruments should enable the legally unfettered interoperability of data, consistent with the principles in the GEOSS Data-CORE.

Although GEO cannot mandate the use of any particular waiver or license, it could choose to label and highlight in the Registry for Components and Services and in the GEO Portal those data registrations that are compatible with the terms and conditions of the GEOSS Data-CORE and that meet the basic requirements for legal interoperability. The waivers and licenses listed below are given as legally valid examples, but data providers in GEOSS may choose to use other similar alternatives. That is, they may still use their own waivers or licences (or none, as the U.S. government currently would do), as long as their approach and terms are compatible with the principles of the GEOSS Data-CORE. Forcing data providers to adopt a specific legal instrument is not the way to maximise the number of datasets within the GEOSS Data-CORE. Legal interoperability does not mean everybody has to use the same waiver or licence, although clearly that is the simplest approach.

Therefore, the presentation by GEO of a small set of universally accepted, well recognized waivers or licenses as choices can be strategically very useful as it can guide data providers toward adopting licenses that can promote interoperability, and thus be a positive move for GEO in achieving its goals for the GEOSS Data-CORE.

3. Examples of Standard Common-Use Waivers or Licenses Compatible with the GEOSS Data-CORE

There are only a few common-use waivers or licenses that have been developed for broad adoption that meet the requirements of the GEOSS Data-CORE. Waivers of rights are the least restrictive and most permissive legal instruments, as discussed above. Licenses intended to allow others to access creative and non-creative content without seeking permission from the owner are sometimes referred to as open content, commons, open access, or open data licenses. The most widely used and prevalent set of open access licenses for creative works is the suite of licenses offered by Creative Commons. Not all of these licenses are suitable for use with marginally creative works, such as databases, nor would all Creative Commons licenses qualify the data for the GEOSS Data-CORE.

Waivers and common-use licenses that would likely meet the requirements of the GEOSS Data-CORE include the licenses shown in Table 1, listed in order of least number of terms and conditions to the most.

Table 1. Waivers and Open Access Licenses that Fulfill the GEOSS Data-CORE Requirements

Name of Waiver or License	Summary Description and URL
Acknowledgement of Public-Domain Status: Creative Commons Public Domain Mark	The CC Public Domain Mark is used to mark data sets already in the public domain, enabling their more ready identification in global web searches. See http://creativecommons.org/choose/mark/ for a description.
Public-Domain Waiver: Creative Commons Public Domain Dedication (CC0)	To the extent possible under law across the world, the person or authority who associates CC0 with the work waives all copyright and related or neighboring rights to this work. For the text, see:

	http://creativecommons.org/choose/zero/
Public-Domain Waiver/License: Open Data Commons Public Domain Dedication and License (PDDL)	The PDDL allows the database user to "copy, distribute and use the database"; "produce works from the database"; and "modify, transfer and build upon the database." See: http://www.opendatacommons.org/licenses/pddl/1-0/ for the full text.
Attribution License: Creative Commons Attribution License (CC BY 3.0)	The CC BY 3.0 license allows the database user "to.Share – to copy, distribute and transmit the work", and "to Remix – to adapt the work", as long as the user "attribute[s] the work in the manner specified by the author or licensor" (plus some other conditions described below). See: http://creativecommons.org/licenses/by/3.0/legalcode for the full text of the license.
Attribution License: Open Data Commons Attribution License (ODC BY 1.0)	The ODC BY 1.0 license allows the database user "To Share: To copy, distribute and use the work", "To Create: To produce works from the database"; and "To Adapt: To modify, transform and build upon the database", as long as the user "attribute[s] any public use of the database, or works produced from the database, in the manner specified in the license." See http://www.opendatacommons.org/licenses/by/ for a full text of the license.

It should be noted that the "Attribution Only" licenses listed in Table 1 are not recommended typically for use with data. There are two main reasons for this. One is primarily philosophical and the other is practical.

(1) *Philosophical*. The open access licensing of data can potentially lead to overclaiming ownership or property rights in facts. That is, facts are in the public domain and yet by recommending a CC license this might lead people to claim ownership in data and impose an attribution condition in a database when it otherwise would not be required in a specific jurisdiction. To recommend a license that might actually facilitate conditions greater than the law would otherwise demand (albeit minimal) is cause for concern.

(2) *Practical*. It is very difficult to develop a license that applies across all legal jurisdictions and takes into account variations in law across the entire globe. For example, when does a compilation of facts reach a point in its coordination, selection, and arrangement so that it is deemed sufficiently "creative" or "original" to make it protectable under copyright? The law and the accurate response varies substantially from jurisdiction to jurisdiction.

In short, in both (1) and (2) the issues are far more complex than for creative works that are fully copyrightable.

It is important to note that further terms or conditions may not be added to the standard instruments in Table 1 or they become no longer "standard." That is, including additionally in a license that all users must pay a marginal cost recovery fee would make the license no longer standard. From a practical perspective however, an agency that charged marginal cost recovery fees to those downloading datasets directly from the agency would not violate the terms of any of the recommended licenses in Table 1 nor would this practice violate the GEOSS open exchange of data sharing principles.

It also should be noted that the licenses listed in Table 1 provide numerous terms in the license that the user of the licensed work is expected to accept as conditions of use. For example, the *Creative Commons Attribution License* imposes restrictions that require licensees to keep any copyright notice intact on all copies of the work, to link to the license from copies of the work, to not alter the terms of the license, not to use technology to restrict other licensees' lawful uses of the work, and to obtain the owner's permission to do any of the things restricted by the license

(e.g., remove attribution in a specific instance).[20] The licenses may also include conditions of use provisions addressing issues such as Representations, Warranties, and Disclaimers, Limitations on Liability and Termination.

Finally, combining data from ODC-BY and CC BY could be uncertain when it comes to figuring out when attribution is triggered when developing a derivative data product, because ODC-BY only applies to the database, whereas CC BY applies to any data that is subject to copyright. It is also worth mentioning that CC BY and ODC-BY do not have parallel attribution requirements, which could further complicate matters. One potential solution to that problem is to suggest that GEO participants contributing data through GEOSS and are using CC BY, customize the attribution requirements for their material, which is possible using the Creative Commons technical infrastructure, in order to match with the requirements set forth in ODC-BY.[21]

4. Characteristics of Other Custom Waivers or Licenses that Would Allow Designation of Data Sets as Part of the GEOSS Data-CORE

As we have already noted, GEO should not mandate any single waiver or license, or even a menu of such instruments for use by data providers in the GEOSS Data-CORE. The preceding discussion was only intended to identify private-law instruments that have characteristics that are compatible with the GEOSS Data-CORE principles and that would make the available data legally interoperable. Any other waivers or common-use licenses that data providers to the GEOSS Data-CORE may use should have the following characteristics:

- They must be compatible with the GEOSS Data-CORE principles.
- They must be valid under the laws of different jurisdictions. GEOSS data currently are to be provided by over 80 Member nations and over 50 Participating Organizations in GEO, with users of the data potentially located in every country in the world. GEO thus should seek to promote the use of waivers or licenses with terms and conditions found to be valid internationally, preferably ones that have a proven track record of use in multiple jurisdictions.

[20] See: http://wiki.creativecommons.org/Baseline_Rights
[21] Private communication from Sarah Pearson to Paul Uhlir, 1 September 2011.

- They should be clear and simple enough not be confusing to the data provider or user. Many types of licenses, particularly restrictive and customized end-user license agreements, are very long and difficult for many users to understand. This value, however, needs to be balanced against the need to maintain the legal validity and integrity of the license, and that there is some risk in over-simplifying licenses. The licenses that are promoted by GEO therefore should not only be legally sound, but should be clear and simple enough so they can be understood even by those who are not lawyers.

- They should be easy to recognize and find. Related to the first two characteristics, the waivers or licenses themselves should be easy to access online by all potential users and not hidden or obscured. This will promote the goal of legal certainty and acceptance.

- They should be available in different languages. Although the common language used in GEO is English, many potential users of GEOSS data, as well as many data providers, speak English as a second language or not at all. The waivers or licenses, and the key metadata, should be available in as many other languages as is practicable, beginning with the language(s) of the country making the data available, plus English, followed by those languages that are the most widely spoken by the greatest number of GEOSS data users.

- They should be embeddable in the data as machine readable metadata. Just as the waivers or licenses should be easy for the human users to find and understand, they also should be machine readable, searchable, and trackable online. This will promote greater use and interoperability of the data, particularly since data are increasingly accessed and used on a machine-to-machine basis, without human intervention.

- Finally, and perhaps most important, the data and databases that are being made available through the GEOSS portal must be kept under the legal control of the data providers. By registering their data with GEOSS, data providers will benefit from greater potential discovery of their data. GEOSS itself, however, will not impose any access or use conditions on the data, which will continue to be held by or kept under the legal control of the providers themselves. Terms and conditions of access and (re)use, if any, will be set by the data providers, and the responsibility of ensuring compliance with those terms and conditions also will rest with the data providers.

5. Standard and Custom Licenses for Data Outside the GEOSS Data-CORE

There are many hundreds of licenses, and especially contracts, in use for data products (many thousands for other information products), with a variety of restrictions that are not compatible with the requirements of the GEOSS Data-CORE. Some of these licenses are intended to be standard or broadly adopted and have other common-use terms and conditions with some rights reserved, such as “non-commercial use only”, whereas many of these instruments were developed specifically by a single company or organization for use with their data products.

Many of the custom licenses are more restrictive on the user than the applicable statutory law, and are meant to protect the proprietary and commercial interests of the data or information provider, further limiting various user rights. Such restrictive licenses are used both for products intended for end-users (rather than re-users and re-disseminators, such as GEOSS data users) or for commercial re-sellers or distributors. This white paper, however, focuses on the legal interoperability of private-law waivers and licenses used in the GEOSS Data-CORE. A subsequent paper will address licenses with restrictions beyond that, such as those seeking to promote non-commercial uses only.

D. CONCLUSIONS AND RECOMMENDATIONS

The foregoing analysis leads to a number of conclusions and recommendations for consideration by the GEO Members and Participating Organizations.

1. Conclusions

“Legal interoperability” of data made available through the GEOSS Data-CORE is essential for the effective sharing of data in GEOSS, which is a priority of the GEO Members. Legal interoperability for data means that the legal rights, terms, and conditions of databases from two or more sources are compatible and the data may be combined by any user without further permission and without compromising the legal rights of any of the data sources used.

When substantial amounts of statutorily protected data are combined from two or more data sources, the new resulting database often will acquire the accumulation of restrictive rights from the sources used.

Public domain status is the best legal option for promoting the various social benefits and goals intended by GEO through the GEOSS Data-CORE by enabling the unrestricted re-use, re-dissemination, and legal interoperability of data, and. The public domain may be defined as encompassing content that is (1) not subject to copyright or related rights (including database protection rights), and (2) not subject to conditions on reuse imposed by other means.[22] The public domain may be created formally by public laws through national legislation or regulation that expressly excludes certain categories of data and information from copyright or from other exclusive property protection, or by private-law waivers of rights. Public domain status may also be attained when the protection of the databases has exceeded the statutory term of protection (which is unlikely for almost all data made available through GEOSS), or by exclusions of certain subject matter from protection, such as facts. Rights under copyright or database protection laws arise automatically (i.e., they do not have to be claimed by a copyright filing or statement), unless expressly excluded or waived. Hence, express legislative or regulatory action is needed, or a waiver of all rights through a private law alternative (see, e.g., the CC0 or PDDL waivers in section 3.2, below) to make the data excluded or waived from protection, or to make the re-use and re-dissemination of data unrestricted.

Ideally, databases already having public domain status should include a notice in their metadata or on the database owner's server informing potential users of their public domain status. The Creative Commons Public Domain Mark, listed in section 3.2, serves this purpose. Such a notice could help to overcome the incorrect assumption by some potential users that the data are subject to protection and have attendant restrictions on reuse. Such a notice would thereby promote the further use of the data and legal interoperability through the GEOSS Data-CORE.

Most databases, however, do not have public domain status and are protected in whole or in part under statutory intellectual property laws. In those cases, a legally valid waiver of rights can achieve a private-law equivalent of public domain status, or a common-use license can

incorporate the attribution conditions allowed by the GEOSS Data-CORE (see the CC BY 3.0 and ODC BY 1.0 licenses in section 3.2).

The endorsement by the GEO Plenary of either standard, accepted waivers or licenses, or other customized common-use licenses that meet all of the GEOSS Data-CORE conditions of access and unrestricted re-use of data, would help ensure certainty and legal interoperability of the data, and thus support the important GEO societal benefit goals. Common-use licenses and waivers also would help promote the contribution of databases through the GEOSS Data-CORE, because most jurisdictions do not have public domain status created by statute for the data compilations relevant to GEOSS.

It is important to note that the attribution term may not be legally enforceable for all data used in all jurisdictions. In those cases that it is not, attribution may be seen as a standard community practice or norm, or a moral or ethical imperative that is not to exactly the same as the legally enforceable attribution condition.

Data policies that promote full and open data exchange, but that are not formally codified through public laws, or through licenses and agreements, do not have the force of law.

2. Recommendations for the 2011 GEO Plenary

The GEOSS Data-CORE's terms and conditions can best be achieved through any of the following mechanisms: statutory public domain, a private-law waiver of rights, or a common-use license.

If the database is not in the public domain as a result of a statutory or private-law waiver of rights, or by the expiration of the term of protection of any rights, the GEO Members and Affiliated Organizations should consider adopting a waiver or common-use data license with the following characteristics:

a. *The waiver or license must be compatible with the GEOSS Data-CORE principles as established in the 2010 GEOSS Action Plan; specifically:*

   - The data are free of restrictions on re-use;

606 *- User registration or login to access or use the data is permitted;*

607 *- Attribution of the data provider is permitted as a condition of use; and*

608 *- Marginal cost recovery charges (i.e., not greater than the cost of reproduction and*
609 *distribution) are permitted.*

610 *b. They should be valid under the laws of as many different jurisdictions as possible.*

611 *c. They should be clear and simple enough not be confusing to the data provider or user.*

612 *d. They should be easy to recognize and find.*

613 *e. They should be embeddable in the data as machine readable metadata whenever*
614 *possible.*

615 *f. They should be available in different languages, at a minimum in the language(s) of the*
616 *country making the data available, as well as in English.*

617 *g. They may have any other terms and conditions, such as a disclaimer of warranty and*
618 *liability, that do not restrict the user or conflict with any of the terms and conditions*
619 *summarized in a-f above.*

620 *h. Finally, and perhaps most important, the data and the applicable license must be kept*
621 *under the legal control of the data providers, and not GEO or GEOSS.*

622

623 Based on these characteristics, the GEO Members and Participating Organizations should
624 consider adopting one of the following existing private-law waivers or standard common-use
625 licenses, which are presented below from pure public domain to the adoption of the legal
626 attribution condition by license[23]:

627

[23] Examples of standard, common-use licenses that meet the GEOSS Data-CORE terms and conditions, but that are geographically limited or constrained to a particular type of data and information (e.g., information generated by a government agency) include: the U.K. Open Government Licence for Public Sector Information (OGL), available at http://www.nationalarchives.gov.uk/doc/open-government-licence/, and the Norwegian Open Data License for Public Sector Information (NLOD), available at http://data.norge.no/nlod.

a. Creative Commons Public Domain Mark. The CC Public Domain Mark is used to mark and identify data sets already in the public domain, enabling their more ready identification in global web searches. For a full description, see http://creativecommons.org/choose/mark/.

b. Creative Commons Public Domain Dedication (CC0). To the extent possible under law across the world, the person or authority who associates CC0 with the work waives all copyright and related or neighboring rights to this work. For the text of this waiver, see: http://creativecommons.org/choose/zero/.

c. Open Data Commons Public Domain Dedication and License (PDDL). The PDDL allows the database user to "copy, distribute and use the database"; "produce works from the database"; and "modify, transfer and build upon the database." See: http://www.opendatacommons.org/licenses/pddl/1-0/ for the full text of the license and waiver.

d. Creative Commons Attribution License (CC BY 3.0). The CC BY 3.0 license allows the database user "to Share – to copy, distribute and transmit the work", and "to Remix – to adapt the work", as long as the user "attribute[s] the work in the manner specified by the author or licensor" (plus some other conditions described in the license). See: http://creativecommons.org/licenses/by/3.0/legalcode for the full text.

e. Open Data Commons Attribution License (ODC BY 1.0). The ODC BY 1.0 license allows the database user "To Share: To copy, distribute and use the work", "To Create: To produce works from the database"; and "To Adapt: To modify, transform and build upon the database", as long as the user "attribute[s] any public use of the database, or works produced from the database, in the manner specified in the license." See http://www.opendatacommons.org/licenses/by/ for the full text.

Custom licenses that have the same terms and conditions as the characteristics listed above can also be used to provide data through the GEOSS Data-CORE, although such custom licenses will not be vetted and approved by the GEO Members in advance.