

The Legal Interoperability of Data

Paul F. Uhler, J.D.

Director, Board on Research Data & Information

National Academy of Sciences

Washington, DC, USA

puhler@nas.edu

The Legal Interoperability of Data

Acknowledgements

Group on Earth Observations Legal Interoperability Subgroup Members:

Paul Uhler, CODATA, Subgroup Co-Leader
Catherine Doldirina, Subgroup Co-Leader
Miles Gabriel, United Kingdom
Joanne Irene Gabrynowicz, Int'l Institute of Space Law
Jeff Heninger, NASA, USA
Puneet Kishor, Creative Commons
Harlan Onsrud, Global Spatial Data Infrastructure Assoc.
Kevin Pomfret, Open Geospatial Consortium
Daniel Quintart, European Commission
Glenn Tallia, NOAA, USA
Sarah Pearson, Creative Commons

Disclaimer: Views are my own and not necessarily those of NRC/NAS

The Legal Interoperability of Data

Imperatives of data interoperability:

- Technical
- Semantic
- Legal

The Legal Interoperability of Data

"Legal interoperability" for data may be defined as follows:

"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 compromising the legal rights of any of the data sources used."

From: Data Sharing Task Force (publication pending). *Draft White Paper on Legal Options for the Exchange of Data through the GEOSS Data-CORE*. Group on Earth Observations.

Legal interoperability is especially important in public research, where many sources are used and re-used, combined, and re-disseminated. In a derivative data product or database, the strongest restriction(s) will control.

The Legal Interoperability of Data

Hierarchy of Laws

Public law

- Treaties
- Legislation
- Regulations
- Government policies

Private law

- Contacts
- Licenses
- Waivers

The Legal Interoperability of Data

Public law default rule—data (and all other forms of information) are automatically subject to all existing legislative and regulatory requirements and restrictions, including:

- National security
- Law enforcement
- Personal Privacy
- **Intellectual property (IP)—copyright, database rights** (and patents, trade secrets, commercial misappropriation)
- And other laws specific to a certain sector or application (e.g., endangered species protection, indigenous rights, etc)

The Legal Interoperability of Data

Copyright

- Applies to original and creative works
- Databases generally protected by “thin” copyright—original and creative *selection, coordination, or arrangement* of the data, not facts themselves.
- Differs by type of data in database, portions of database, and national law (some may still apply “sweat of the brow” protection based on work and expense in compiling the database)
- Exclusions or limitations (e.g., subject matter, like facts) and exceptions (e.g., fair use, fair dealing) also differ
- Bottom line: unclear and uncertain application of copyright to databases, and varies significantly.

The Legal Interoperability of Data

Database Protection Law (EU and similar legislation)

- In addition to copyright
- Provides exclusive property protection to any compiler of information based on substantial investment
- Protects any portion of the database that is more than insubstantial, measured quantitatively or qualitatively
- 15-year period of protection, automatically extended with each new substantial investment (potentially in perpetuity)
- Few exceptions, none mandatory
- Government compilations *not* excluded
- Bottom line: provides much greater IP protection, but with uncertainty, and contravenes prior IP law by placing a strong exclusive property right (no actual harm required) on investment, rather than creativity.

The Legal Interoperability of Data

Summary of default *public law* (legislation/regulation) IP regime:

- Unsatisfactory to many producers and users (too strong for some, too weak for others)
- Uncertain application in scope of coverage, even within one jurisdiction
- Varies greatly across jurisdictions and types of databases, but applies automatically
- Encourages non-compliance with the law by many users

The shortcomings of public law:

- Stimulates producers to turn to more flexible and responsive *private law* common-use solutions (waivers, licenses, contracts)
- Digital networks provide means to implement private law options easily, cheaply, and with greater certainty (power of the two-party deal)

The Legal Interoperability of Data

Focus here on “public domain” and “attribution only” conditions. Not on other conditions of common-use (e.g., non-commercial, copyleft) or restrictive licenses (greater restrictions on data users than those allowed by statute).

Public domain may be defined as information that is “(1) not subject to copyright or related rights (including database rights), and (2) not subject to conditions on reuse imposed by other means.”

- Sarah Pearson, Creative Commons, private communication (2011).

It is the *yin* to the proprietary *yang*.

The Legal Interoperability of Data

Public domain status provides greatest flexibility and freedom for data users. Advantages of public domain include:

- *Full interoperability*: data from many sources can be used and combined without restriction.
- *Reusable*: data can be repurposed into new and interesting contexts and disseminated to the world.
- *Administrative burden*: low transaction costs and administrative costs over time.
- *Legal certainty*: users can rely on legal usability of the data.

Source: Thinh Nguyen (2011). "The Web-Enabled Research Commons: Applications, Goals, and Trends." In *Designing the Microbial Research Commons*. National Academies Press. Washington, D.C.

Disadvantages of public domain mostly for producers:

- Loss of any control by the producer of the data use downstream, including commercial uses, misuse, etc.
- Cannot legally require attribution (but community practice may successfully substitute for that)
- Not broadly available

The Legal Interoperability of Data

Public domain status can be achieved through:

- Expiration of copyright or other legislative protections. This is very long and currently perpetual, unless using historical data in the research. It is passive and requires the user to wait. Access can still be subject to payment, although once accessed the database is fully open.

Or

- The database is composed fully of non-copyrightable subject matter (e.g., an alphabetical directory)—but beware the database protection legislation in the EU!

Or

- Express waiver by public statute or private law.

The Legal Interoperability of Data

Waiver of all IP rights by public statute is rare.

Example of a public statutory waiver:

Section 105 of U.S. Copyright Act (1976), waiving copyright protection for all federal government works produced in scope of employment.

The Legal Interoperability of Data

Examples of private law waiver include:

- Creative Commons (CC) Public Domain Mark (used to mark and identify databases already in the public domain)
- or
- CC0 (waiving all copyright and neighboring rights to the work, to the extent allowed by different jurisdictions)

Both waivers rely on community practices or norms for attribution.

The Legal Interoperability of Data

Licenses and contracts are not the same thing.

Licenses:

- Are based on existing statutory rights for enforcement
- Automatic (do not depend on “agreement” by user)
- Do not extend to facts or materials already in public domain (because there is no underlying statutory protection), but can extend to databases or protectable portions of databases (but uncertainty of enforcement remains)
- Can be used to decrease or increase level of protection, based on what the database producer/distributor wants. Decreased protection=common use conditions; increased protection=added user restrictions

The Legal Interoperability of Data

Example of common-use license for databases:

- Creative Commons Attribution Only—CC BY 4.0: 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”.
- However, CC does *not* recommend use of the CC BY 4.0 license for databases, because of fears of “attribution stacking” where 100s of data sources may be combined over time, and because the legal attribution requirement may not satisfy all norms or expectations.

The Legal Interoperability of Data

Contracts:

- Unlike licenses, based on express *agreement* of the parties.
- Requires formal offer, acceptance, consideration, and (usually) written terms. Formal offer and acceptance for databases (and other digital information products) made through click through agreements online or shrink wrap agreements on CDs and other physical media
- Unlike licenses, not dependent on enforcement for underlying statute (but must not be for an illegal purpose)
- Unlike licenses, can apply to data otherwise unprotected by statute
- Contracts are only valid for the agreeing parties (rest of the world not bound), so they can be an uncertain mechanism for rights holders.
- Contracts/agreements are not standard, unlike licenses, and frequently long, confusing, ignored by the user.

The Legal Interoperability of Data

Examples of common-use contracts requiring click-on assent:

- GBIF Data Use Agreement (requiring attribution)
- Long-term Ecological Data General Data Use Agreement (requiring attribution)

Examples of restrictive contracts:

- Most commercial End User Licensing Agreements (EULAs), that provide pages and pages of the owner's rights and the user's restrictions.

The Legal Interoperability of Data

Take-away points:

- Legal interoperability is important for data re-users (e.g., public researchers) rather than for end users who are just consumers.
- Public law status quo is uncertain and can be very restrictive for public research users of databases. Data users ignore the law at their own peril.
- Public domain status, whether created by lapse or exemption of protection, or by express waiver, provides greatest interoperability and freedom for users, but no control or protection for the producer/original rights holder.
- An attribution-only license may be used for databases, but is not recommended by Creative Commons because of potential for attribution stacking. Such licenses, however, result in only minor restrictions for data users but the scope is uncertain for the database rights holder.
- Contracts (attribution only) are more certain for the rights holder, but are not standard for the user and can be more burdensome.
- The more legal restrictions that are added, the less appropriate for science.