

**University of Maine**

---

**From the Selected Works of Harlan J Onsrud**

---

2009

# Liability for Spatial Data Quality

Harlan J Onsrud, *University of Maine - Main*



Available at: [https://works.bepress.com/harlan\\_onsrud/3/](https://works.bepress.com/harlan_onsrud/3/)

# Liability for Spatial Data Quality<sup>1</sup>

Harlan J. Onsrud

## Abstract

*Liability in data, products, and services related to geographic information systems, spatial data infrastructure, location based services and web mapping services, is complicated by the complexities and uncertainties in liability for information system products and services generally, as well as by legal theory uncertainties surrounding liability for maps. Each application of geospatial technologies to a specific use may require integration of different types of data from multiple sources, assessment of attributes, adherence to accuracy and fitness-for-use requirements, and selection from among different analytical processing methods. All of these actions may be fraught with possible misjudgments and errors. A variety of software programs may be run against a single geographic database, while a wide range of users may have very different use objectives. The complexity of the legal questions surrounding liability for geospatial data, combined with the diversity of problems to which geospatial data and technologies may be applied and the continually changing technological environment, have created unsettling and often unclear concerns over liability for geospatial technology development and use. This article selects a single data quality issue to illustrate that liability exposure. In regard to that issue, it may have a substantial stifling effect on the widespread use of web-based geospatial technologies for such purposes as geographic data mining and interoperable web mapping services. The article concludes with a recommendation for a potential web-wide community solution for substantially reducing the liability exposure of geospatial technology and geographic data producers and users.*

## Introduction

Use of digital spatial data, unlike use of a music or journal article file, may often result in some action or decision. If errors or other shortcomings such as lack of data accuracy or completeness result in an inappropriate action or wrong decision, the possibility of liability arises for data suppliers as well as for all other parties in the chain of handling and processing the spatial data.

“As a general proposition, legal liability for damages is a harm-based concept. For instance, those who have been specifically hired to provide data for a database, or those who are offering data for sale to others, are responsible for some level of competence in the performance of the service or for some level of fitness in the product offered. If others are damaged by mistakes that a producer should not have made, or by inadequacies that should not have been allowed, the courts have reasoned that producers should bear some responsibility for the damages. But for their mistake or defective product, the damages would not have occurred. In commercial settings, liability exposure often may be reduced through appropriate communications, contracts, and business practices. However, liability exposure may never be eliminated completely. Nor as a society would we want it to be. Modern societies generally support the proposition that individuals and businesses should take responsibility for their actions if those actions have unjustifiably caused harm to others.

---

1



This work is in the public domain as of May 1, 2009  
<http://creativecommons.org/publicdomain/zero/1.0/>



However, the law does not require perfection. The law exists and responds to a realistic world. No general purpose dataset will ever be complete for all potential purposes that users might desire. Nor will the accuracy of data ever meet the needs of all conceivable uses. It is also inevitable that errors and blunders will be contained in any practical database. Thus, the law holds that those in the information chain should be liable only for those damages they had a duty to prevent. Establishing the nature and extent of rightful duties has traditionally been accomplished under theories of tort or contract law. Legislation may impose additional or alternative liability burdens. Legislation affecting liability for spatial datasets and software typically might be found in statutes addressing such issues as intellectual property rights, privacy rights, anti-trust issues, and access rights” (Onsrud, 1999).

Although the facts or legal issues may be complex in a specific dispute, the core legal concepts in imposing liability on map- or mapping system creators for inaccuracies or blunders that should not have occurred and that have caused harm have not changed substantially as societal uses have moved from paper to digital formats. For an overview discussion of liability in the use of geographic information systems and geographic data sets along with several examples that illustrate theories and limits of liability exposure within geographic information contexts, see Onsrud (1999). The core legal concepts change only slowly over time, so much of the discussion remains relevant. Another primer on liability and geographic information may be found in Cho (2005). While the core legal principles remain relatively stable, practical liability exposure has increased substantially due to the greatly expanded numbers of day-to-day users of digital mapping and guidance systems, the emergence of new use environments for spatial data, and the imposition by legislatures of much larger penalties being applied to cyber violations over comparable physical world violations.

### **An Illustrative Data Quality Liability Problem**

There are obviously many aspects to spatial data quality. The most frequently cited components include accuracy, precision, consistency and completeness, with each of these components typically assessed in terms of space, time and theme (e.g., spatial accuracy, temporal accuracy, and thematic accuracy). One important aspect of data quality relative to completeness is knowing the legal status of a digital dataset that one may desire to copy, or a digital database from which one may wish to extract. That copying or extraction may be by action of a human or software (e.g., automated data mining). If a human or human-initiated software extracts from a geographic dataset or database from which no legal authority exists to extract, substantial liability exposure is incurred.

Just because one finds a music file readily available on the web, does not mean that legal authority has been granted to download it or incorporate all or part of the file content into a derivative work. The same rule applies to geographic data files.

Further, keep in mind that if you were to steal a CD from a music store, the maximum typical fine might be about \$1000. However, if you were to download the same ten songs from the Internet, your liability exposure could be as high as \$1.5 million<sup>2</sup> (Lessig, 2004: ch. 12). Thus, in a similar manner, if you use a web mapping service or an automated data mining software program to draw together data from ten sources distributed across the web, the same potential \$1.5 million liability exposure arises.

What if one only hosts a web mapping service, but it is users that use the technology to draw their own maps extracting data from those other sites? Is the web mapping host site relieved of liability? Among the relevant legal precedents include the Napster

---

<sup>2</sup> Under the 1999 Digital Millennium Copyright Act in the United States, statutory damages for behavior unproven as “willful” are set at “not less than \$750 or more than \$30,000” per infringement while statutory damages for “willful” behavior are not more than \$150,000 per infringement. The Recording Industry Association of America (RIAA) has been very effective in using these very high damage limits in extracting large settlements from individual students and others accused of illegal downloading of music files.

(*A&M Records, Inc. v. Napster, Inc.*, 2000) and the recent Pirate Bay cases. In both instances, the operators of the web sites were found guilty even though they only facilitated the sharing of files by others, and even though many instances of sharing were legal. In the Pirate Bay case, decided on 17 April 2009, each of the four defendants was sentenced to a year in jail and fined roughly \$900,000 in damages (Carrier, 2009). So in a legal context, in providing a map service, the question arises not whether the operation of the service provides a valuable public benefit or whether legal uses are being made of the site, but whether this is also an operation for facilitating illegal sharing and use of geographic data files without the explicit permission of the owners of those files.

Quality of data requires completeness, and completeness in the global legal environment requires that you know the legal status of the data that you draw from others and have strong confidence that you have legal authority to use the data in your specific context. By far the greatest legal liability exposure for those pursuing exchange and integration of geographic data in a networked world of interoperable web services and data mining is incurred through the violation, whether intended or unintended, of copyright, database legislation, and similar intellectual property protections. Such laws are in place and being continually strengthened by governments across the world.

### **The Existing Legal Problem with Internet Wide Geographic Data Sharing**

The core problem we are confronting is the acquisition by database and dataset developers of automatic copyright upon creation, whether creators want it or not for their data offerings. In most jurisdictions of the world, if one creates an original doodle, copyright occurs instantaneously in that doodle with no need to register the right and regardless of whether most people would find this instantaneous copyright to be reasonable. Therefore, if someone creates a story, song, image or dataset and places it openly on the web, is it free for anyone to copy without permission? Keep in mind that *common practice*, *not getting caught*, or *small likelihood of being sued* are not equivalent to having a clear legal right to copy. The lawyerly response to the question is that the answer will depend on answers to several additional questions. As a general proposition, however, there will be some minimal creativity in the vast majority of digital works made accessible through the Internet. If minimal creativity exists, the law assumes one must acquire permission from the copyright holder to copy, distribute, or display the work or generate derivative products from it.

One might argue that ‘data’ or ‘empirical values’ drawn from a database are all legally equivalent to ‘facts’, and therefore are not protected by copyright. Even if true in a specific jurisdiction in a specific instance, the selection, coordination and arrangement of facts may be protected by copyright. Further, the explicit legal tests for qualifying for, or determining what is protected by copyright vary from jurisdiction to jurisdiction (e.g., protection of sweat of the brow, industriousness, etc.) and many jurisdictions supply protections for datasets and databases that extend well beyond those granted by copyright (e.g., database protection legislation, unfair competition regulations, catalog rules, etc.). In truth, one cannot know for certain whether there exists some minimal creativity in a posted geographic data set (and neither can any lawyer) until the gavel falls in a court of law on a case-by-case basis. Thus, most lawyers across the globe, when asked, will advise their clients that they should always assume that a party will emerge to sue unless explicit permission to use is acquired in every instance of drawing from the materials of others.

Assume that you have just extracted data elements from 42 other geographic datasets in an automated data mining or web mapping exercise. Many of us assume that the vast majority of those other sites probably placed their datasets on the web and are adhering to data format and other interoperability standards so that others might freely benefit from their postings. Yet, the law generally holds that we must *not* make this assumption. By example, one or more of those 42 sites has inevitably posted a license provision that your specific use breaches and the posted contract or license provisions of many of the sites are highly likely to be in conflict with each other. Some of the sites

will have no license language or use restrictions posted, but in those instances the intellectual property laws of some nation will apply by default. You are required to meet the requirements of all of the involved sites unless you have explicit permission stating otherwise. Shipping a request to the addresses of those 42 sites asking them for explicit permission to use the data extracted from their web sites is burdensome and nonsensical to the typical user. Most of the recipients of your message might also view the request as silly since they would not have posted the data if they did not want it to be used. In many instances you won't receive responses. One should assume however that there are one or more legal claimants lurking out there if you do copy or extract data without explicit permission from each and every source.

Copyright liability is a strict liability concept and thus no intent to break the law or even knowledge of breaking the law is necessary to be found guilty. Even innocent or accidental infringement may produce liability (*DeAcosta v. Brown*, 1944). For each violating extraction or copying, the potential damages are huge and are having a chilling effect on using the geographic data of others, except perhaps for the largest companies with strong legal staffs willing to fight wars of attrition.

Common sense argues that we need strong copyright protections in order to prevent digital thieves from stealing our geographic data offerings against our will. In the world of paper, the law had established a nuanced balance between the rights of creators to benefit from their original works and the rights of new creators to build from the past contributions of those who came before. This balance no longer exists. There are no longer legal exceptions for uses that do not involve a copy such as loaning a book many times or selling a used CD because virtually every use on the Internet involves a new copy. Fair uses and the right to extract such things as facts are readily negated on the Internet by requiring users to click a license before allowing the downloading of data or software. Laws across the globe are skewing even further the legal landscape that has been disturbed by changes in technology rather than restoring the balance. Changed technological conditions and our legislatures in the cause of protecting us from thieves have shifted us to what Lawrence Lessig refers to as a "permissions culture" in the world of the Internet (Lessig, 2004, ch 10). You now almost always need permission to extend from the work of others and this is a world dominated by lawyers.

As a result of the current lack of ability to reliably and efficiently know whether a data set found on the web is available for use without further permission, we have hundreds of millions of orphaned digital works on the web that are not generating any economic income for creators, and in many instances creators placed these data sets on the web for the purpose of being freely used by others. Yet the law states we must *not* assume that we may use them. As a result we do not use the geographic data and products of others in generating our own new products because we are afraid of the legal liability exposure. Either that, or many of us simply use what we find freely available and become law breakers whether we know it or not. This situation is not conducive to supporting democratic societies. When large segments of society become lawbreakers on a daily basis – whether downloading music or geographic data files, breaking the speed limit, or cheating on tax forms – a culture of general disrespect for the law is created. This is bad for society since we generally want our laws respected and enforced. Individuals should also be aware that if evidence exists of commission of a crime such as illegal downloading, certain civil liberties such as the right to privacy, the right to be free of search and seizure of your computer or the right to Internet access are likely to disappear (Lessig, 2004: ch. 12).

To avoid widespread breaking of the law and the creation of a growing culture of disrespect of the law generally, our societal choices are typically to change the law, enforce the law more aggressively, find ways for our community to live within the bounds of the law, or a combination of these approaches.

### **A Suggested License Solution**

Numerous legislative proposals have been submitted in jurisdictions across the globe

to restore the balance between the rights of creators and users under copyright law. To date they have consistently failed to gain passage. As mentioned previously, the trend is actually in the opposite direction as legislatures view the need to catch digital thieves as far more pressing than the nebulous need to support freedom to build on the works of others. Rather than fight the law or in addition to advocating changes in the law, another option is to use the law and technology to create an electronic commons in geographic data that all could openly use.

To deal with the data quality issue of completeness, metadata creation for geographic data needs to become very easy, efficient, and fast and become part and parcel of the ordinary course of doing business, science, and government. Those providing data with location components should not even recognize that they are helping to provide standards compliant metadata. Affirmative licensing to allow others to openly access and use geographic data needs to be embedded into the metadata creation process. Further, the ability to determine whether data accessible through the web is available legally for open access use needs to be determinable by browser software. How can these objectives be efficiently accomplished?

A legal commons has already been created for creative works on the web through the use of Creative Commons (CC) licenses (Creative Commons 1, 2009). With a few clicks, in less than a minute, one may create iron clad licenses for any of your creative work to make it practically and legally accessible to others. Well over 100 million such open access licenses have already been created and the advanced searches of most major web browsers allow one to restrict web searches to return hits to only those sites with the standard CC license you have specified in the search. The browsers automatically pick up embedded html code indicating that the returned sites contain CC licensed material.

There are several versions of CC licenses offered. Every license helps you “retain your copyright (and) announce that other people’s fair use, first sale and free expression rights are not affected by the license.” .... “Every license allows licensees (i.e., users), provided they live up to (the license provisions), to copy the work, to distribute the work, to display or perform it publicly, to make digital public performances (e.g., web casting), and to shift the work into another (medium). Every license applies worldwide, lasts for the duration of the work’s copyright and is not revocable.” ... “Every license requires licensees to obtain your permission to do any of the things you choose to restrict.”(Creative Commons 2, 2009) Among the optional restrictions or conditions you may choose to impose include requiring attribution, restricting uses to noncommercial purposes, not allowing derivative works, and allowing others to distribute derivative works but only under the condition that those works use a license identical to your license (Creative Commons 3, 2009).

The use of such licenses begs the question of what specific CC license would best fit the mores and traditions of science of the past several hundred years, and which would be best for the geographic information science community? For creative works such as research articles and reports I argue that the attribution license with no further restrictions provides the greatest freedom for all of us to extend from the scholarly advancements of each other. Obtaining credit for their contributions is the primary motivating factor and concern of scholars, researchers and students.

Creating licenses for data sets and databases is much more problematic. Creative Commons licenses apply only to creative works and the existence of *creative expression* is very minimal or questionable in many datasets that have been compiled in standard formats to ease machine processing. Thus, as a general proposition, CC licenses should not be applied to data files and databases. Legal scholars have been struggling with what to recommend for several years, and the Science Commons division of CC has recently recommended the use of provisions adhering to the Science Commons open access database protocol. To ensure that geographic data may be used legally across the web for the widest range of purposes, I recommend use of CC0 i.e. Creative Commons Zero or No Rights Reserved (Creative Commons 4, 2009). This language waives all copyright and database rights to the extent that one may have these rights in

any jurisdiction and is the best current option to ensure that data can be used legally for general web mapping and feature services, data mining, copying, and extraction. This approach completely avoids the license stacking problem and the need to resolve conflicting restrictions among licensed data sets and those that have been posted without licenses such that default laws of the jurisdiction apply. If contributors affirmatively waive all rights to the greatest extent possible there are no restrictions to conflict with each other. In order to protect attribution for contributors, other means than a copyright derived right may be utilized.

Although minimizing liability for data quality requires completeness of data that incorporates the legal status of the data, merely knowing the legal status of data or knowing the best license for support of web-wide sharing is insufficient. As stated previously, the open access data license generation process needs to be embedded in the metadata generation and recording process in the ordinary day-to-day course of doing business, science and government, or we will merely continue with the status quo of an inefficient system for contributing and finding geographic data and services.

### **A Suggested Web-Wide Solution**

Most people now gathering geographic data that they view as important for some purpose (or they would not have gathered it) are unable to archive these data sets so that the data (a) will be preserved over the long term, (b) can be found again readily, (c) will retain credit for the collector or creator, and (d) can be used legally by others without asking for further permission. Even if accessible through the web, geographic data appropriate for a specific purpose or germane to a specific geographic location is currently very difficult to find, use and share. Producers and users cannot find each other efficiently, nor agree on terms of use efficiently

Let us consider the example of volunteered geographic information by scientists, students, hobbyists, and many other average citizens. We argue that the typical person engaged in volunteering geographic information about their projects or communities would like to have a simple, comprehensive, practical, and universal solution for providing increased ‘findability’ for the data and products she produces, wants a solution that is legally, economically and ethically defensible, and does not want details but just a simple solution implemented through the web. We surmise that contributions to the commons would greatly increase and would be far more useful if (1) the ability to contribute geographic data sets was much easier to do, (2) creators could reliably retain credit and recognition for their contributions, and (3) creators would gain substantial benefits by contributing their geographic data to a commons.

In light of the previous discussion about the need for completeness of geographic data to ensure quality and usefulness, the minimum criteria for a widely used and effective web-based contribution facility we hypothesize should include *all* of the following capabilities:

- The ability of a contributor to waive their rights to the greatest extent possible in all jurisdictions using iron clad open access legal language (e.g., use the CC0 language) for each contributed data file in less than a minute.
- Accomplish a permissions request to any other potential current or prior rights holder in the data set in less than five minutes with responses automatically fed to the system without human intervention.
- The ability to create standards-based metadata in less than 10 minutes, situating the data in space and time, and tied through an ontology covering all fields of science. We suggest a user interface process with a sophisticated back end by which the system automatically guesses at the field or specialty area of the contribution based on terms supplied by the contributor when describing the file and terms that may be contained in the contributed file. The system would tie the meanings of the terms used to the most appropriate thesauri but as well allow the user to readily correct the guessed at meanings of the words.
- Automatic long-term archiving and backups.

- Automatic conversion of files to interchange formats to enhance survivability of the information over time.
- Attribution acknowledgement. We suggest that the data contributor be credited in the repository for contributing to the shared resource and be granted the right to use the trademark of the facility as acknowledgement of their contributions. This could link to their contributions in the facility (preferred), or attribution could be provided by automatic lineage tracking in all the interchange formats as portions of the contributed data sets are sliced and diced over time. The second option requires more computational and storage overhead, but would allow contributors to search the web to discover how many times their contribution or significant portions of it had been used by others.
- An efficient peer recommendation system.

All of these would be new benefits not typically provided to the typical person volunteering geographic information to web projects today. The contributor would gain easy off-site management and backups of their contributions, enhanced ‘findability’ of their data sets and data products and that of others, increased recognition and credibility and enhanced peer review of their contributions.

## Conclusion

The world is unlikely to see efficient widespread notice of the rights in geographic data sets that would allow legally defensible sharing of geographic data until appropriate rights transfer language is embedded in widely used metadata creation processes. Further, it is doubtful that widespread metadata documentation of geographic data will be universally achieved in doing the day-to-day tasks of business, government and science until all of the above criteria at a minimum are met by operational facilities and distributed across the web.

Similar to the web itself, not all geographic data creators and collectors will contribute their data to a legal commons environment, but many would. Such a commons would constitute a valuable web-wide resource providing assured legal authorization to copy datasets, extract from databases, provide web mapping and web feature services, and engage in data mining. The legal right to carry out such activities under the current status of web and geospatial technology development and in the current global legal environment, however, is very much in question.

## References

- Carrier, M.A. (2009), “BitTorrent: Legal Nightmare or Future Business Model?” *OUPBlog* (April 17<sup>th</sup> 2009) Oxford University Press USA. <http://blog.oup.com/2009/04/bittorrent/>.
- Cho, G. (2005), *Geographic Information Science: Mastering the Legal Issues*. John Wiley and Sons, Inc., Chichester, 440p.
- Creative Commons 1 (2009) Creative Commons <http://creativecommons.org>
- Creative Commons 2 (2009) Baseline Rights [http://wiki.creativecommons.org/Baseline\\_Rights](http://wiki.creativecommons.org/Baseline_Rights)
- Creative Commons 3 (2009) Creative Commons Licenses <http://creativecommons.org/about/licenses/meet-the-licenses>
- Creative Commons 4 (2009) CC0 <http://creativecommons.org/license/zero/>
- Lessig, L. (2004), *Free Culture: The Nature and Future of Creativity*. Penguin Press, New York, 345p.
- Onsrud, H. (1999), “Liability in the Use of GIS and Geographical Datasets”. In: P. Longley, M. Goodchild, D. Maguire and D. Rhind (eds.), *Geographical Information Systems: Vol.2 Management Issues and Applications*, John Wiley and Sons, Inc., New York, pp. 643-652.

## Legal cases cited



A&M Records, Inc. v. Napster, Inc., 239 F.3d 1004 (9th Cir. 2001), affirming, 114 F.Supp.2d 896 (N.D. Cal. 2000).  
DeAcosta v. Brown, 146 F.2d 408 (1944), cert. denied, 325 U.S. 862 (1945).