Mashing-up Culture

Proceedings from the COUNTER workshop Mashing-up Culture Uppsala University, May 13-14, 2009

Title: Mashing-up Culture: The Rise of User-generated Content

Eds.: Eva Hemmungs Wirtén and Maria Ryman

Abstract: These are the proceedings of the first COUNTER workshop "Mashing-up Culture: The Rise of User-generated Content", Uppsala University, May 13-14, 2009. COUNTER is a European research project exploring the consumption of counterfeit and pirated goods. Sampling, remixing, mash-ups and appropriation are part of the digital creative milieu of the twenty-first century. Sites such as YouTube and deviantART have offered new outlets for creativity and become hubs for innovative forms of collaboration, thus playing their part in challenging modernist notions of what it means to be a creator as well as a consumer. Drawing on this general background, the ten papers presented in these proceedings are examining areas such as: Sampling, mash-ups, and appropriation; Creativity and collaborative practices; Creative Industries and intellectual property; Copyright, Cultural Heritage and Cultural Policy; and Formal and informal regulations of intellectual property. The ten authors are all contributing to an international and interdisciplinary scientific discussion on the *mash-up*. Mash-ups and user-generated content are social and cultural phenomena which in these papers are put into various contexts, from legal ones, over technology, to the nation as a framework.

Keywords: Proceedings (COUNTER Counterfeiting and Piracy Research), Copyright—Adaptions, User-created content

COUNTER is funded by the European Commission's Framework 7 programme.

UPPSALA UNIVERSITET

Counter© Counterfeiting & Piracy Research

COOPERATION

Mashing-up Culture: The Rise of User-generated Content

Proceedings from the COUNTER workshop Mashing-up Culture
Uppsala University, May 13-14, 2009

Eds. Eva Hemmungs Wirtén & Maria Ryman

UPPSALA
UNIVERSITET

Meddelanden från Institutionen för ABM vid Uppsala universitet
Nr 2

Meddelanden från Institutionen för ABM vid Uppsala universitet
Nr 2

General eds.: Kerstin Rydbeck, Eva Hemmungs Wirtén,
Björn Asker and Inga-Lill Aronsson
Eds. for No. 2: Eva Hemmungs Wirtén and Maria Ryman

Distribution:
Department of ALM
Box 625, 751 26 Uppsala
publikationer@abm.uu.se

Cover: Graphic Services, Uppsala University
Cover photo: Mash-up of COUNTER logotype

© Department of ALM, COUNTER and the authors
ISBN 978-91-506-2081-8
Print: Universitetstryckeriet, Uppsala 2009

Public Policy: Mashing-up Technology and Law

Puneet Kishor, Oshani Seneviratne

Contact Information
Puneet Kishor
University of Wisconsin-Madison
1630 Linden Drive
Madison, WI 53706
Ph: +1 (608) 729-5832
Email: punkish@eidesis.org
Web: http://www.punkish.org

Oshani Seneviratne
Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology, Cambridge, MA 02139
Ph: +1 (857) 998-7451
Email: oshani@csail.mit.edu
Web: http://people.csail.mit.edu/oshani

Abstract

Online communities are changing the way we think about restricting or opening access to information. This is no more truer than in online mapping which is fueling community interest in collaborative applications. The open source communities are much like a commons of information where the non-subtractable resource actually grows with use. Restricting access to geospatial information will cut off the oxygen for both scientific as well as community activities that depend upon it. Information is restricted for many different reasons, but clearly, free and open access to new information is not guaranteed. The solution lies in not one but in multiple approaches that satisfy public policy, scientific, community and economic interests. In this paper we propose an approach for using technology to aid public policy by enabling the user to do the right thing instead of preventing the user from doing the wrong thing.

Rise of Online Communities

Online communities are turning many traditions on their heads – the Drudge Report is first-to-wire, Flickr and Twitter are quicker, and Reddit, Digg, Slashdot, Technorati and Del.icio.us aggregate diverse sources of information perhaps better than Associated Press or Reuters can. They are investigative and sensational, they break or even make careers, and sometimes they are also wrong, but, it cannot be denied that they are permanently game-changing. They are also changing the way we think about restricting or opening access to information, and this is no more truer than in online mapping. Web-based mapping is fueling community interest in collaborative applications inspiring astonishingly creative civic, environmental, social justice and journalistic maps while also mapping places to visit, sightsee or get a lunch.[56]

We believe restrictions in the form of use permissions and contractual obligations placed on access to information, both data and tools, will impede creation of legally conforming mash-ups. In this paper we offer some reasons for this impedance, and suggest that one

56. National Center for Geographic Information and Analysis organized a Specialist Meeting on Volunteered Geographic Information in 2007 in Santa Barbara, CA, to address the implications of this phenomenon. (NCGIA 2007).

way to lower this barrier is by using a technological solution that aids in implementing the permitted uses of the components of the mash-up.

Restricting Online Information

While public agencies might restrict access to information because of security concerns, private agencies do so for economic reasons as well. Public interests are served by publicly available information, but as corporations take over its collection and management, continued access to new information is not guaranteed for the various stakeholders. Uhlir and Schröder identify governments, research funding agencies, universities and not-for-profit research institutes, learned and professional societies, international scientific organizations, industry research institutions, individual researchers and the general public as the stakeholders (2007). They also identify the hidden costs of closed data systems as higher research costs, lost opportunity costs, barriers to innovation, less effective cooperation, education and training, sub-optimal quality of data and a widening gap between OECD nations and developing countries.[57]

Open Access to Information

There is much interest in free and open access to information nowadays.[58] What started off as a movement for openness in computer program source code has developed into multiple movements, from

57. Organization for Economic Cooperation and Development (OECD) is an international organization of 30 high-income economies helping governments tackle the economic, social and governance challenges of a global economy.

58. See the timeline of the open access movement maintained by the Open Access Directory for the rise in popularity of the movement (Directory 2009).

scientific data and research journals to governance to films, music, literature and other cultural products.[59] Both open access and the interest in open access are fueled by the Internet and the web.

As it becomes easier than ever to share virtually identical copies of the original virtually limitlessly, there is also much concurrent interest in enforcing intellectual property rights. This is being accomplished by both technological means as well as by tightening up intellectual property laws.

Tension of Open Access

While technology is making it easier to share information, technology is also making it easier to block information, to build walls that can keep people out. And, while technology is changing very fast, law is also changing, but very slowly. The current copyright law in the United States was passed in 1976, which, while only 33 years ago, was almost 15 years before the web came into existence. So there is arguably a need for copyright reform to keep pace with the technological changes (Samuelson 2007).

Culture is changing very rapidly in one segment, and not so rapidly in another segment. The open knowledge segment is growing fast, and changing consumption patterns, expectations and products. For example, changing readership of newspapers is attributed to the loss in advertising revenue as the money shifts to online outlets leading to a loss of paper-based newspapers and magazines. Concurrently, the large media firms are exerting a strong influence on the interpretation, application and execution of intellectual property laws in protecting their content.

Conventional wisdom has held the belief that given something of utility, its scarcity creates value. That belief drives our economic system, and everything from entrepreneurial startups to global trade.

59 See the work of The Open Knowledge Foundation that bills itself as "promoting open knowledge in a digital age." (See the work of The Open Knowledge Foundation that bills itself as "promoting open knowledge in a digital age."OKFN 2009).

Open source wisdom, on the other hand, maintains that sharing is good. Creating and maintaining scarcity is expensive, particularly in the context of information, and does not guarantee continued competitive advantage anyway.

This tension between law and technology and the gap between the cultural norms and expectations of different segments of content producers means that many laws are being broken. In fact, never before in our history has it been so easy to break so many laws with just one mouse click.

Benefits of Open Access

Research confirms that open access is good for research, particularly in the developing countries that are hungry for scientific information (Engelward and Roberts 2007; Evans and Reimer 2009). But, how do we reconcile the desire to access information freely with the desire to protect its inherent value for the benefit of its creator? How do we encourage innovation and entrepreneurism that is enabled by open access and open source while ensuring that those who come up with new ideas are able to benefit from them?

Commons as an Alternative

Commons is one alternative that is proposed as a model. A commons is traditionally a place in which the common public holds rights to its use, and no one is able to exercise a monopoly on its use. The idea of commons is a fairly old one, popular in reference to the English commons that were held open for enjoyment by all for various outdoor activities, from grazing cattle to walks and hunting. Of course,

the old English commons were not without their battle for public's rights (Shaw-Lefevre 1894).

"Freedom in a commons brings ruin to all," lamented Garrett Hardin who believed that the negative component of the utility of partaking in the commons inexorably eroded it for everyone (1968). The commons of information, however, is different – not only is it not eroded upon use, its utility, like goodwill, actually grows with its use.

The open source communities are much like a public commons of information (Hess and Ostrom 2003), but need innovative approaches to governance just as more traditional commons have needed new institutions (Ostrom 1990).

Commons of Information

The idea of a commons has been applied to information, and has been received with a surprising amount of enthusiasm. Since the original commons were physical spaces providing access to physical resources, they were naturally subtractable, that is, they would decrease in amount upon use. Information, especially digital, is inherently non-subtractable. Once produced, it costs relatively little to store and distribute. Unless access to it is restricted, scarcity does not apply because information never decreases upon use. And, because information at one level is merely data at another level, it serves as raw material for more information. The downstream benefits of public sector information can be tremendous to society, especially when put to possibly unforeseen uses because of widespread access afforded by digital information and computer networks (Uhlir and Schröder 2007).

Not surprisingly, commons has been proposed as a model for geospatial data as well. One proposed model depends upon an Internet based repository into which participants would deposit their data,

create a waiver of all their rights save an obligation for attribution, and thus, share in each others' data (Onsrud et al. 2004).

Making Commons Possible

A commons is a good idea, but it depends on the availability of one or more of such central repositories that could be trusted by its potential users not just for their security and integrity but also longevity. No one is going to deposit their intellectual output in a repository that will not instill a sufficient amount of confidence in its being around far into the future. In any case, for other reasons such as privacy and control, people are not going to be very comfortable giving up their information to an entity that may or may not be around in the future, that may or may not sell or in some other way give away or benefit from their data. Who can one trust? The private sector? The government? Themselves? And, a commons of geospatial information does not do away with legal implications but in fact, it creates many new and unique legal considerations (Mccurry et al. 2006).

Whatever the nature of the commons, be it a centralized institutional, governmental or community owned repository or a peer-to-peer based network, we require innovations in technology, the legal regime and the culture of data ownership to make it successful. Technology makes sharing possible, law makes it permissible, and culture makes it acceptable.

Most of us want things for free, but do not want to give them away for free. Culturally, we have to learn and understand the value that can come from "giving it away." This change in mindset is already happening in communities and common public, but is slower to materialize in institutions. More change in mindset will happen as a result of both activism as well as generational change. Already, all 20 years old were born after the web was invented. In another five years, every 25 year old leaving the university and entering the job

market will not have known a world without the web – cultural change is only a matter of time.

Legally, we have to make it easy to implement desired use restrictions on our own output as well as understand and adhere to the use restrictions placed on the output of others. Those who understand the value of free and open access to information in their work are willing to provide the same to their own intellectual output and data to others. Doing so, however, is onerous as it requires interpreting the intent of the creator of the product. An interoperable, interpretable, easily implementable mechanism for conveying the intent would greatly facilitate the sharing of data and information:

- Imagine a world in which everything we create automatically gets imbued with what we desire for its use and becomes available to everyone else;

- Imagine that anything we make is locatable by everyone else, is accessible to them, and they immediately know what they can do with it;

- Imagine that every time someone uses our output, somehow we get some kind of payment, be it in monetary format or in the form of attribution and credit.

We believe that one way this can be achieved is with the help of technology, and in a manner so that our existing workflows require as little added impedance or modification. This is how technology can assist in implementing our desires.

Balancing Freedom and Rights

We all talk of policy, but what do we mean by "policy"? Policy is a stance, a disposition toward a favored state. To realize a viable commons of information, we have to simultaneously provide users with the freedom to access data while provide the data providers with an avenue to benefit from their work if they so desire. This is our policy, and we are interested in the interface of technology and law as offering possible solutions to meet our policy goals.

Licenses and Contracts

Our desire for how we want our output to be used is conveyed through legal instruments called licenses and contracts.

A **license** is an instrument that conveys a right, accompanied by a promise, by the grantor to not sue the grantee if that right is exercised. In the context of property law, a license is a unilateral permission to use someone else's property. In the context of *digital files*, a license describes the conditions of usage under which those files may be used. A license on a digital file can exist whether or not there are any corresponding users of that file. A user would have to abide by the license that covers the usage of that file, and if any of the conditions of usage described in that license are violated, then the user would have to cease using that file. Licenses are covered by federal copyright laws.

A **contract** is like a license, but requires at least two parties agreeing to it. Without at least two parties, a contract cannot exist. A contract specifically describes the obligations of both parties to the contract. For example, "If we give you this data file with experiment readings, you will give me a chart showing a scattergram" is a contract, provided both you and we agree to it. Contracts fall under the purview of state law.

So, one could have a digital file that is released under a particular license, but could also be given to someone under a contract to actually deliver something in return. For example, we could license our music mp3 under a *Creative Commons 3.0 NC* license and give it to you under a contract that obligates you to add a strings soundtrack to the file and give the file back to us. If you agree to the contract and take our mp3, then you will have to deliver to us that mp3 with a strings soundtrack added to it. If you fail to give us the specified mp3 with the promised strings soundtrack incorporated in it, you will be in breach of your contract. Of course, the license itself would allow you to do anything else with the mp3 as well, as long as it was not used in a commercial project (that condition comes from the **NC** clause in the CC 3.0 NC license). This example should make the difference between a license and a contract very clear.

It should be noted that a CC license can also take on the nature of a contract. For example, a CC 3.0 BY license obligates the user of the licensed item to give attribution to the creator of the item (that obligation stems from the **BY** clause in the CC 3.0 BY license). This can be problematic in the case of data set because of attribution-stacking whereby a user can get legally obligated to attribute **all** the contributors to a crowd-sourced data set.

In the world of digital data, licenses are utilized to grant certain rights to the user by actually reducing the rights held by the user. For example, data created by me are protected for my use completely and fully by the copyright law. However, we can grant you a license to use the data for your work provided you use it for noncommercial purpose. If you start using our data for commercial purpose, we can revoke your license and ask you to stop using our data.

We can further obligate you, the user, to do something if you use our data – for example, we can ask you to pay us money, or we can ask you to give us credit. If you agree to this condition, then you are entering into a contract with us.

It is to be noted that while factual data such as pure geographic data cannot be protected by licenses, they can very well be protected by contracts. In the case of databases, this can lead to a unique problem, especially if a database contains within it data from many contributors. The user, the contractee, ends up having agreed to not one

but many contracts, and fulfilling all those contracts can become exceedingly onerous or even impractical as in the case of mash-ups.

Mash-ups Defined

Mash-ups have existed for as long as content has existed. Jazz musicians routinely quote other songs and tunes, and borrowing and mixing are considered an integral part of jazz music. Robert Rauschenberg is famous for having used Willem de Kooning's painting to create a new painting called *Erased de Kooning* (Rauschenberg 1953). Scientists routinely utilize data from different sources to conduct their own experiments.

However, mash-ups, as we know them now, are a peculiarly digital phenomenon of the internet age. They are entirely a product made possible by the portable, mixable and immediate nature of digital technology. We define a mash-up as content formed by an inline combination of two or more different content streams in real time. So, A and B data streams can be utilized to create a mashed-up content C by programmatically merging and modifying A and B from source instead of copying A and B to a different location and then merging and modifying them. That way, as A and B continue to change, the resultant changes are reflected in C in real time.

There is no doubt that mash-ups are very popular, and within mash-ups, map mash-ups are by far the most popular. One API directory, at the time of writing this paper, lists 1,254 documented mash-up APIs with 6,852 mash-ups out of which a highly disproportionate 2,281 or 33 percent are mapping mash-ups, with Google Maps API being the source of 1,658 mash-ups itself.[60]

60. The API Directory by the Programmable Web provides one measure of the popularity of mash-ups. (Web 2009).

The Problem of Mash-ups

Mash-ups create a potential legal problem by merging more than one legally encumbered data streams. By adding an additional input of creativity, they create a derivative work. This can be problematic on two fronts – the creator of the mash-ups has to worry about remaining within the bounds of the permitted uses of the components of the mash-ups, and the users of the mash-ups have to deal with not one but multiple sets of permissions should they decide to use the mash-up in their own work.

Unless the mash-up involves using two or more pure data-streams, whereby data are made up of facts that cannot be copyrighted, there will always be a finite set of permitted uses that both the mash-up creator and the mash-up users will have to contend with.

Creators and users of mash-ups can either ignore these permissions, or follow them. Either way, this creates a burden on them – ignoring them puts the mash-ups makers and users at the peril of breaking the law, and following them slows the creative process. Much creativity visible on the web is arguably because of the instantaneousness and spontaneity with which content can be accessed and used. The lack of impedance is what greases the creative process.

So, once again, we are back to a familiar problem – balancing the freedom of the users with the rights of the creators. And, there are really no two parties here since creators of content are also users of content and vice versa.

The Problem Domain

This is a multifaceted problem, and has stemmed from multiple domains intersecting at rapid speed – the cultural need to express oneself, the technological capability to do so rapidly and in innovative ways, and the legal boundaries within which it all can be done. Technological innovations makes mash-ups possible, legal innova-

tions are required to make them permissible, and cultural change makes them acceptable.

The solution to the problem is also going to be multifaceted, otherwise tackling any one aspect in isolation is only going to lead to frustration. In the rest of the paper we describe one approach that mashes up technology and law to create a solution that makes the task of abiding by the permissions on a dataset imposed by its producer easier for the consumer of the dataset.

Cleaning up the Legal Minefield

With almost a hundred different licenses available for a creator of digital data, what we have is a babel of licenses.[61] Creative Commons (CC) makes our task easier by clearing up the minefield of licenses. Through an innovative license chooser that makes the job of choosing the appropriate license easy even for a non-initiate, and through a choice of less than half a dozen possible licenses, CC presents an innovative solution to an otherwise confusing problem.[62]

A Data Purpose Algebra

One of the hazards of combining multiple data sources is that incompatible licenses can get mixed up creating a license that basically freezes the creative process. Take for example a Non-Commercial (NC) license that gets mixed with a Share-Alike (SA) license. The resulting license becomes NC-SA, but while the result satisfies the first license by also being NC, it fails the second license by not being only SA. Remember, SA requires that the resulting product be shared under exactly the same conditions as the component product under SA. So, if the component product under SA was unencumbered by

61. There are more than 90 different digital licenses that create a confusing choice for potential licensors. (Kishor 2006).

62. Creative Commons provides fewer than half a dozen licenses and an easy to use license chooser on the web. (Commons 2009).

the NC clause, adding it to an NC restricted component violates the SA requirement.

Hanson et al. show that it is possible to model data usage policies programmatically (2007). They describe each dataset as a bundle of its content, a source or agent that processes the data, the category of data, and its purpose. When another agent combines two or more datasets, a new dataset is created whose content, category and purpose are some function of the agent, content, category and purpose of each of the component datasets.

Inline Provenance

Extensible Metadata Platform (XMP) is another technology that allows one to transfer metadata along with the content by embedding the metadata in machine readable Resource Description Framework (RDF).[63] This technology is widely deployed in embedding licenses in free-floating multimedia content such as images, audio and video on the web. Jones developed a bookmarklet to copy and paste document fragments while preserving their provenance information that was stored inline using Extensible Hypertext Markup Language (XHTML) syntax (2007). Building up on that work, Seneviratne is creating a Semantic Clipboard, an extension to the open source Firefox browser to seamlessly integrate the metadata to content upon reuse, with the goal of capturing intent of the usage and making sure that the content will be reused in a policy aware manner (2009).

These projects have contributed a clearly defined document fragment ontology that can represent the information about the sources of the content, a method of excerpting from these sources, and a reasoning engine which reasons over the acceptable use of the source and the composite CC licenses.

63. Adobe, Inc., created the XMP specification (Adobe 2009).

Using Technology to Implement Licenses

How can technology be used to help implement a legal regime that makes a desired state possible? We believe that instead of using technology to make it difficult to do the wrong thing, it is better to use technology to make it easy to do the right thing.

Now that we have a potentially smaller menu of licenses to choose from, would it not be nice if we could embed the license into our work so that anyone using our work would also automatically get a copy of the license. CC allows doing just that utilizing a special kind of language called the Creative Commons Rights Expression Language (ccREL) (Commons 2009). The ccREL language allows using a world wide web consortium published specification called RDFa that allows attributes to be expressed as structured data in any markup language (W3C 2008).

Metadata

To convey our desired use restrictions with our output, we have to depend on metadata. Metadata, or literally, data about data, are used to describe mundane but useful facts such as provenance (whence the data came from?), currency (when it came into being?) and many other properties such as accuracy, author, etc. Metadata can also be used to describe use restrictions such as CC or other licenses.

To be useful, metadata need to have three important characteristics: they have to be easy to produce, be embedded within the data they describe, and be easily readable.

The easiest way to produce metadata is to have them be produced automatically. Any metadata that has to be produced manually by the user usually does not get produced at all.

The easiest way to ensure that the link between metadata and the data they describe is not broken is by embedding the former inside the latter. This way, the two travel together as a package, inseparable

always. Imagine if we all had name tags, but our name tags were kept in a different room from us. Every time we had to determine someone's name, we had to find some identifying code about them, then run into the other room and find the corresponding name tag to find out their name, and then come back and continue our interaction. That would not be very productive.

Finally, the metadata have to be accessible easily, readable both manually as well as programmatically. At best, the metadata should be readable by crawlers of the various search engines. Since metadata and data are traveling together, if popular search engines such as Google and Yahoo can read the metadata, by default the data become available to anyone who searches for them.

License Violations Validator

We find inspiration in digital photos and their Exchangeable Image File Format (Exif) data to implement a technological solution to help users stay in conformance with the licensed uses of photos. The Exif information describes the photo, is embedded inside the photo itself, and is readable using simple tools. This allows embedding within the photos licensing information about the photos.

This is essentially a tool to help an honest person remain honest when reusing content on the web. When someone aggregates content from many different sources, it is inevitable that some attribution details may be accidentally forgotten. An important consideration on the design of this tool was not to focus on the digital rights management (DRM) approach. It was also not meant to be entirely focused on Creative Commons (CC) either. But in the current implementation it is based on CC licenses with possible extensions to scenarios modeled in Policy Languages such as "Accountability In RDF" (AIR) (Kagal, Hanson, and Weitzner 2008).

While tools such as http://validator.creativecommons.org provided by the CC detects the embedded licenses in a page and gives

information about the license, it cannot be used to detect a license "violation" and obtain helpful hints as to how to make the user's composite document compliant with the original source licenses. Also, not all the CC licenses link back to the CC deed referrer Uniform Resource Identifier (URI), which, by clicking on it composes the attribution XHTML from the license deed if it is available.

As a proof of concept, we have implemented the Flickr CC license violations validator, which pinpoints any Attribution license violations on Flickr images used in composite works.In order to make sure that no CC license terms are violated, the author can run the CC License Violations Validator hosted at http://oshani.mit.edu/cc_validator.py and see if some sources have been left out or whether some have been misattributed. Once the user gives the URI where the composite work could be found at a particular URI, a site crawler will search for all the links embedded in the given site and filter out any embedded Flickr photos. From these Flickr photo URIs, it is possible to glean the Flickr photo id. Using this photo id, all the information related to the photo is obtained by calling several methods in the Flickr API. This information also includes the original creator's Flickr user account id, name and CC license information pertaining to the photo. If a Flickr photo has a CC license attached, regardless of the purpose for what it is used for, the photo should be given proper attribution as Flickr is still using the older CC2.5 recommendation. Therefore, if it was determined that a Flickr photo on a particular page has a CC License, the tool checks for the attribution information which can be either the attribution name, attribution URI, source URI or any combination of those within a reasonable scoping in the containing DOM element where the image was embedded in. If such information is missing, the user is presented with the details of the original content creator's name, the URI and the license it is under to compose the proper XHTML to properly attribute the sources used.

But, back to our problem of data – it is very difficult to determine where raw data, or facts, end and interpreted data, or creative expression of facts, begin. For example, sensors readings on a charge-coupled device (CCD) are certainly raw data, however, when such readings are acquired through a camera and stored on a reproducible medium, they become photographs that are protected by copyright. In the geospatial realm, raw data collected by remote sensing satellites are reflectance values of geographic features, that is, pure facts. These sensor readings are processed in image-processing programs to create colored photographs that can be protected by copyright.

Further complications arise vis à vis data in that while factual data themselves cannot be copyrighted, the manner in which the data are organized can indeed be copyrighted. Data, typically held in a database, are organized to optimize any one or more of atomicity, consistency, isolation, durability as well as security and speed of access. This organization reflects creativity, and hence, it is protected by copyright. Insofar as the organization of data assists in the execution of a function or a process, it can be patented. And, to make matters even more complicated, data can even be protected by trade secret – for example, names and addresses of customers can be of strategic value to a business, and the business can protect those data as a trade secret.

Since determining what portion of our data is factual and what portion is interpreted, hence, protected by copyright, is so difficult, one school of thought believes that preemptively negotiating a contractually obligatory scientific commons where all information is unencumbered would be a way toward lasting free and open access (Reichman and Uhlir 2003). The Creative Commons Zero (CC0) Protocol goes even further by suggesting that one waive all rights that one might have in one's data, the CC0 data mark serving as a mark of quality that a potential user can depend upon to represent no encumbrances on the dataset (Commons 2007).

Whatever the nature of the commons, we are hoping to apply the ideas from data purpose algebra and learned from experiments such

as the license validator to geographic datasets. We hope such effort will make it easy to do the right thing without burdening the user, and thus, greatly aid in spontaneity and creativity on the web. Our hope stems from our firm belief in using technology to aid the user in doing the right thing rather than preventing the user from doing the wrong thing.

References

Adobe, Inc. 2009. *Adobe XMP: Adding Intelligence to Media* 2009 [cited April 4 2009]. Available from http://www.adobe.com/products/xmp/

Commons, Creative. 2008. *CC0 Waiver 1.0 Universal*. Creative Commons 2007 [cited December 12, 2008]. Available from http://labs.creativecommons.org/licenses/zero/1.0/

---. *ccREL - CC Wiki*. Creative Commons, April 6, 2009. Available from http://wiki.creativecommons.org/CcREL

---. 2009. *Licenses* 2009 [cited April 4, 2009]. Available from http://creativecommons.org/about/licenses

Directory, Open Access. 2009. *Timeline of the Open Access Movement*, March 20, 2009 2009 [cited March, 31 2009]. Available from http://oad.simmons.edu/oadwiki/Timeline

Engelward, B.P., and R.J. Roberts. 2007 "Open Access to Research Is in the Public Interest." *PLoS Biology* 5(2):e48.

Evans, J.A., and J. Reimer. 2009 "Open Access and Global Participation in Science." *Science* 323: 5917.

Hanson, C., L. Kagal, T. Berners-Lee, G.J. Sussman, and D.J. Weitzner. 2007 "Data-Purpose Algebra: Modeling Data Usage Policies." *IEEE Policy*.

Hardin, G. 1968 "The Tragedy of the Commons." *Science* 162:1243–1248.

Hess, C., and E. Ostrom. 2003. "Ideas, Artifacts, And Facilities: Information As A Common-Pool Resource." *Law and Contemporary Problems* 66 (1 & 2).

Jones, H.C. 2007 "XHTML Documents with Inline, Policy-Aware Provenance. "Department of Electrical Engineering and Computer Science, MIT.

Kagal, L., C. Hanson, and D. Weitzner. 2008. Using Dependency Tracking to Provide Explanations for Policy Management. In IEEE International Workshop on Policies for Distributed Systems and Networks.

Kishor, P. 2009. Babel of Licenses 2006 [cited March 31, 2009]. Available from http://www.punkish.org/Babel-of-Licenses

Mccurry, D., J. Campbell, M. Lutz, H. Onsrud, and K. Williams. 2006 "Managing Intellectual Property Issues in a Commons of Geographic

Data." *Digital Libraries Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries*: 347–347.
NCGIA. 2008. Specialist Meeting on Volunteered Geographic Information. NCGIA 2007 [cited March 15, 2008 2008]. Available from http://www.ncgia.ucsb.edu/projects/vgi
OKFN. 2009. The Open Knowledge Foundation 2009 [cited April 2, 2009 2009]. Available from http://www.okfn.org/
Onsrud, H., G. Camara, J. Campbell, and N. Sharad Chakravarthy. 2004. *Public Commons of Geographic Data: Research and Development Challenges*. M.J. Egenhofer, C. Freska and H.J. Miller, eds.: Berlin: Springer-Verlag.
Ostrom, E. 1990. *Governing the Commons: the Evolution of Institutions for Collective Action*. Cambridge Cambridge University Press.
Rauschenberg, R. 1953. *Erased de Kooning Drawing*. San Francisco, CA: San Francisco Museum of Modern Art.
Reichman, J.H., and P.F. Uhlir. 2003 "A Contractually Reconstructed Research Commons For Scientific Data In A Highly Protectionist Intellectual Property Environment." *Law and Contemporary Problems* 66 (1 & 2).
Samuelson, P. 2007 "Preliminary Thoughts on Copyright Reform." *Utah Law Review* 2007(3): 20.
Seneviratne, O. 2009. Semantic Clipboard, Feb 17, 2009 [cited March 31, 2009]. Available from http://dig.csail.mit.edu/2009/Clipboard/extension/
Shaw-Lefevre, G. 1894. *English Commons and Forests: The Story of the Battle During the Last Thirty Years for Public Rights Over the Commons and Forests of England and Wales*: Cassell and Company, Ltd.
Uhlir, P., and P. Schröder. 2007 "Open Data for Global Science." *Data Science Journal* 6: OD36–OD53.
W3C. 2008. *RDFa in XHTML: Syntax and Processing*. http://www.w3.org/TR/rdfa-syntax/
Web, Programmable. 2009. API Directory 2009 [cited April 1, 2009]. Available from http://www.programmableweb.com/apilist/bymashups#