

3-1-2010

The Semantic Web and Information Flow: A Legal Framework

Julien Mailland

Follow this and additional works at: <http://scholarship.law.unc.edu/ncjolt>



Part of the [Law Commons](#)

Recommended Citation

Julien Mailland, *The Semantic Web and Information Flow: A Legal Framework*, 11 N.C. J.L. & TECH. 269 (2010).
Available at: <http://scholarship.law.unc.edu/ncjolt/vol11/iss2/3>

This Article is brought to you for free and open access by Carolina Law Scholarship Repository. It has been accepted for inclusion in North Carolina Journal of Law & Technology by an authorized administrator of Carolina Law Scholarship Repository. For more information, please contact law_repository@unc.edu.

**THE SEMANTIC WEB AND INFORMATION FLOW: A LEGAL
FRAMEWORK**

Julien Mailland^{*}

Considered by many as the next key development in Internet technology, the semantic web has generated a lot of buzz recently. A creation of Tim Berners-Lee, inventor of the World Wide Web, the semantic web is a "web of data," a technology that promises to provide the world with enhanced access to information through more efficient and sophisticated content indexing. There are, however, dangerous side-effects to the system. As a labeling system, the semantic web bears the potential to provide sophisticated new tools for enhanced control of information distribution by governments. Ironically, however, perhaps because Dr. Berners-Lee's previous invention served as a catalyst of unrestrained information flow worldwide, there has been little if any analysis of the potential to use the semantic web as a tool of enhanced control. This paper aims at filling the gap, by unveiling unintentional externalities of content-tagging systems such as the semantic web from the standpoint of the free flow of and access to information and by alerting the computer science community that further work needs to be done as the semantic web is being developed, in order to minimize these externalities and further embed freedom in the technology.

I. INTRODUCTION

Considered by many as the next key development in Internet technology, the semantic web has generated a lot of buzz recently. A creation of Tim Berners-Lee, inventor of the World Wide Web,

^{*} Annenberg Fellow, University of Southern California, Annenberg School for Communication & Journalism. LL.M., 2000, New York University School of Law. M.C.L., 1999, LL.B., 1997, University of Paris II School of Law (Panthéon-Assas). Many thanks to Kevin Driscoll, Tom Goodnight, Martin Hilbert, Russ Korins, Bill McClain, James Polk, Monroe Price, Philip Spencer, and Matt Ward. The author can be reached at his last name at usc.edu.

the semantic web is a “web of data,” a technology that promises to provide the world with enhanced access to information through more efficient and sophisticated content indexing. Recognizing the current limitations of computer software, Dr. Berners-Lee pointed out that “[m]ost of the Web’s content today is designed for humans to read, not for computer programs to manipulate meaningfully [I]n general, computers have no reliable way to process the semantics.”¹ As a result, he suggested a new approach for encoding data in which:

[t]he Semantic Web will bring structure to the meaningful content of Web pages, creating an environment where software agents roaming from page to page can readily carry out sophisticated tasks for users. . . . The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation.²

As he put it, “[t]he Semantic Web will enable machines to *comprehend* semantic documents and data.”³ The semantic web is an offspring of an earlier project developed by Berners-Lee’s World Wide Web Consortium (“W3C”), called the *Platform for Internet Content Selection* (“PICS”).⁴ As described by the W3C, “[t]he PICS specification enables labels (metadata) to be associated with Internet content. It was originally designed to help parents and teachers control what children access on the Internet.”⁵ Both the semantic web and PICS revolve around the tagging of content, which is a critical tool in the control of information distribution through censorship. As such, both technologies bear the potential to provide sophisticated new tools for enhanced control of information distribution in cases where a given government would like to pursue censorship.⁶ Ironically, there has been little or no

¹ Tim Berners-Lee, James Hendler & Ora Lassila, *The Semantic Web*, 284 SCI. AM. 34, 36 (2001).

² *Id.* at 36–37.

³ *Id.* at 40 (emphasis in original).

⁴ See PICS, <http://www.w3.org/PICS/> (last visited Mar. 15, 2010) (on file with the North Carolina Journal of Law & Technology).

⁵ *Id.*

⁶ As put by Lawrence Lessig, “PICS is the devil. . . . PICS will be an extremely versatile and robust censorship tool.” Lawrence Lessig, *Tyranny in the Infrastructure: The CDA Was Bad—But PICS May Be Worse*, WIRED, July

analysis of the potential of the semantic web as a tool of enhanced control of ideas, perhaps because Dr. Berners-Lee's previous invention served as a catalyst of unrestrained information flow worldwide. This paper aims to fill that gap, by unveiling unwanted externalities of content-tagging systems such as the semantic web that may negatively impact the free flow of and access to information. It also aims to alert the computer science community that further work needs to be done as the semantic web is being developed in order to minimize these externalities and further embed freedom in the technology.

In this paper, I first contextualize the brief history of Internet regulation and censorship under the lens of international communication theory, the international relations theory that deals with "communication that occurs across international borders,"⁷ specifically with reference to free flow, dependence theory, and hegemony theory. Second, I examine the idea of a "semantic web" with reference to the theory from which it evolved: semantics. I point out that allowing meta-tags of information to be a crucial tool of retrieval of information in the Internet age is equivalent to allowing a meta-authority to classify—and to exclude—information on our behalf, in the same way that we rely on a librarian's tagging system to access information in a physical library. While the Internet allowed the populace to break away from such shackles, the semantic web could lead us backwards if unwanted externalities are not neutralized. Third, using specific case studies of recent semantic web developments, I demonstrate that by allowing for more sophisticated meta-classification of content and by promoting cooperation mechanisms between content producers and gatekeepers, the semantic web might instead allow for more refined and effective controls of information distribution. Building on Jürgen Habermas' construct of the

1997, http://www.wired.com/wired/archive/5.07/cyber_rights_pr.html (on file with the North Carolina Journal of Law & Technology).

⁷ ROBERT FORTNER, INTERNATIONAL COMMUNICATION: HISTORY, CONFLICT AND CONTROL OF THE GLOBAL METROPOLIS 6 (1993), *quoted in* DAYA KISHAN THUSSU, INTERNATIONAL COMMUNICATION xiv (2006).

“public sphere”⁸ and on the traditional First Amendment concept of the marketplace of ideas, I argue that the semantic web risks creating a “digital public sphere,” a global marketplace of ideas where publishers will offer their content and present it with labels, and where governments worldwide will select content for redistribution in their respective jurisdictions, all by automated means made possible by semantic web tools. The result would be an optimized, automated global censorship system, where it would not be the best idea that prevails but the one most suited to the receiver’s regulatory environment. This is the externality this paper aims at raising awareness towards, so the computer science community can better embed civil libertarian values into the technology as it develops it, preventing the dream of Internet visionaries from turning into their worst nightmare.

II. FREE FLOW, HEGEMONY, FILTERING, AND THE BATTLE FOR THE NODES

The concept of the “free flow of information,” as summarized by Daya Thussu, expressed the idea that the primary function of international communication was to promote democracy, freedom of expression, and markets⁹ and reflected Western, and specifically U.S., antipathy to state regulation and censorship of the media.¹⁰ Indeed, one can consider the free flow of information a necessary tool to give substance to John Winthrop’s vision of a “city upon a hill”¹¹ in an age of increased international relations and as a tool to globalize Madisonian ideals. For First Amendment scholar Alexander Meiklejohn, “in a society pledged to self-government, it is never true that, in the long run, the security of the nation is endangered by the people. . . . Freedom is always wise. That is the faith, the experimental faith, by which we Americans have

⁸ See JÜRGEN HABERMAS, *THE STRUCTURAL TRANSFORMATION OF THE PUBLIC SPHERE: AN INQUIRY INTO A CATEGORY OF BOURGEOIS SOCIETY* (Thomas Burger & Frederick Lawrence trans., MIT Press 1989) (1962).

⁹ THUSSU, *supra* note 7, at 55.

¹⁰ *Id.*

¹¹ JOHN WINTHROP, *A MODEL OF CHRISTIAN CHARITY* (1630).

undertaken to live.”¹² Indeed, under American constitutional law and political science theory:

[t]he very purpose of the First Amendment is to foreclose public authority from assuming a guardianship of the public mind through regulating the press, speech, and religion. In this field every person must be his own watchman for truth, because the forefathers did not trust any government to separate the true from the false for us.¹³

Free flow of information is therefore a prerequisite for those ideals to prevail in the global village.

Other theories of international communication compete with the American ideal. For instance, on the opposite end of the international communication theory spectrum lie proponents of the dependence theory (also known as the cultural imperialism thesis) and the hegemony theory. According to dependence theory, transnational corporations based in the West exercise control, with the support of their respective governments, over the developing countries by setting the terms for global trade.¹⁴ Further, under hegemony theory, as developed by Antonio Gramsci, mass media is a key player in propagating and maintaining the dominant ideology.¹⁵ The goal of both dependence theorists and hegemony theorists, therefore, is to cut free flow communications emerging from certain countries—namely, from the West, and, chiefly, from the country where speech likely enjoys the most freedom in the world, the United States. The discussion related to the free flow of information became part of the Cold War discourse. While for the supporters of capitalism, the primary function of international communication was to promote democracy, freedom of expression, and markets; Marxists argued for greater state regulation of communication and media outlets.¹⁶

¹² *The Meaning of the First Amendment: Hearing Before the S. Subcomm. on Constitutional Rights*, 84th Cong. (1955) (statement of Alexander Meiklejohn).

¹³ *Thomas v. Collins*, 323 U.S. 516, 545 (1945) (Jackson, J., concurring).

¹⁴ See HABERMAS, *supra* note 8, at 47.

¹⁵ See generally ANTONIO GRAMSCI, *SELECTION FROM THE PRISON NOTEBOOKS* (1971); see also THUSSU, *supra* note 7, at 53 (discussing the notion of hegemony in international communication).

¹⁶ THUSSU, *supra* note 7, at 55.

In terms of Internet infrastructure policy, dependence theorists and hegemony theorists have been hard at work to take control of Internet nodes—the hardware that creates bottlenecks—and Internet protocols—the software that allows for traffic-management at the node level. Expressions of such efforts have materialized in the *governance movement*, a coalition of governments, private sector, and civil society actors claiming a right to participate in the decision-making process related to the development of the Internet¹⁷ in an unprecedented partnership with censors of all kinds, including Western censors. However, merely taking control of the pipes, bottlenecks, and protocols does not suffice to enable control-oriented governments to *selectively* censor. Unless a government intends to prevent *all* Internet traffic from entering its borders, it must develop mechanisms of selective censorship to ensure that its values are asserted within the physical bounds of its territory, while still allowing the benefit of Internet technologies. Until now, however, mechanisms were so crude that they led to heavy spillover effects—externalities—causing much more content to be banned than was originally targeted. In turn, some Western governments abandoned or minimized their efforts to control content flow, as will be illustrated below. A look at two flow-control techniques, filtering and labeling, helps explain the reason behind the spillover effects.

Filters are pieces of software which are implemented along various steps of the network (routers, Internet service providers, the local library's server, or one's own browser).¹⁸ When triggered

¹⁷ See REPORT OF THE WORKING GROUP ON INTERNET GOVERNANCE 4 (June 2005), <http://www.wgig.org/docs/WGIGREPORT.pdf> (last visited Mar. 2, 2010) (on file with the North Carolina Journal of Law & Technology) ("Internet governance is the development and application by Governments, the private sector and civil society, in their respective roles, of shared principles, norms, rules, decision-making procedures, and programmes that shape the evolution and use of the Internet.").

¹⁸ See, e.g., SearchSecurity.com, Content Filtering, http://searchsecurity.techtarget.com/sDefinition/0,sid14_gci863125,00.html (last visited Mar. 15, 2010) (on file with the North Carolina Journal of Law & Technology) ("Content filtering is used by corporations as part of Internet firewall computers and also by home computer owners, especially by parents to screen the content their children have access to from a computer. Content filtering usually works by

by keywords, these filters block certain content from passing through to the user.¹⁹ The problem with filtering is that it blocks too much, because it is not intelligent. It does not understand the *semantic* implications of language. Consider the following explanation by Alan Brown, of the Digital Freedom Network:

Filters are computer algorithms, rules which seek patterns and follow instructions based on what they find. Filters do *not* make judgments, act *in loco parentis*, evaluate context, appeal to morality or understand you or your child's needs. They simply scan for letter combinations. This constitutes neither the exercise of judgment nor the enforcement of decency. . . . We do know, though, that whatever 'decency' means, it does not include blocking medical information on breast cancer, preventing students from botany information because the word 'cucumber' contains the letter sequence 'cum,' blocking people from sites on 'Mars exploration' because the phrase contains the sequence 'sex,' or preventing a woman from registering her name 'HillaryAnne' because it contains the sequence 'aryan.' Yet all of these are documented errors based on the way filtering software is designed.²⁰

Perhaps the most amusing example of a counterproductive effect of filtering was provided to us involuntarily by Dick Armey, former Republican Representative and House Majority leader. In 2000, the Digital Freedom Network released the results of a "Foil the Filters" contest, which it had organized to challenge the "people around the world to come up with the most absurd

specifying character strings that, if matched, indicate undesirable content that is to be screened out." See also the description of its service by Net Nanny, a leading filtering software provider: "Keyword filtering lets parents decide if kids should see pages that contain specific words. You can even customize this list for each individual user, giving you the flexibility and control over the internet content your family can view." NetNanny.com, <http://www.netnanny.com/products/netnanny> (last visited Mar. 15, 2010) (on file with the North Carolina Journal of Law & Technology).

¹⁹ See NetNanny.com, <http://www.netnanny.com/products/netnanny> (last visited Mar. 15, 2010) (on file with the North Carolina Journal of Law & Technology).

²⁰ Alan Brown, *Four Myths and Facts for Parents and Their New Computers*, DIGITAL FREEDOM NETWORK, Dec. 22, 2000, <http://web.archive.org/web/20010301231819/dfn.org/focus/intl/netmyth.htm> (on file with the North Carolina Journal of Law & Technology).

examples of problems with filtering.”²¹ The “Poetic Justice Award” went to “[a]n anonymous submitter [who] noticed that the Web site of . . . Richard ‘Dick’ Armey, Majority Leader of the U.S. House of Representatives and a staunch defender of censorware and strict Internet regulation, is himself a victim of censorware. Netnanny, Surfwatch, Cybersitter, N2H2, and Wisechoice are among the ‘software solutions’ which Armey advocates. All of them filter his site because it contains the word ‘dick.’ ”²²

In other words, because of its semantic incompetence, filtering creates spillover effects: it blocks content which is not only harmless but also fully lawful and useful from the standpoint of education and of democracy in general. I am not suggesting that Dick Armey made a significant contribution to education or democracy, but materials on topics such as breast cancer and World War II history must be widely distributed in order to further important educational purposes. Under a filter-based information-distribution control system, however, materials on World War II history that refer to words such as “Hitler” and materials on breast cancer that refer to “breasts” tend to be blocked and unavailable to those who seek valuable educational information.

Fortunately for Americans, the Supreme Court has recognized the danger of techniques that involve such broad externalities and has struck down statutes because of their chilling effects on speech. In the landmark case of *Reno v. ACLU*,²³ the Court examined the constitutionality of the Communication Decency Act of 1996 (“CDA”). The CDA criminalized the knowing transmission of obscene or indecent messages to any recipient under eighteen years of age.²⁴ It further prohibited the knowing, sending, or displaying to a person under eighteen of any message that, in context, depicts or describes, “in terms patently offensive as measured by contemporary community standards, sexual or

²¹ *Winners of the Foil the Filters Contest*, DIGITAL FREEDOM NETWORK, Sept. 28, 2000, http://attrition.org/misc/ee/why_censorware_sucks.txt (on file with the North Carolina Journal of Law & Technology).

²² *Id.*

²³ 521 U.S. 844 (1997).

²⁴ *Id.* at 846.

excretory activities or organs.”²⁵ Affirmative defenses were provided for those who take good faith, effective actions to restrict access by minors to the prohibited communications and those who restrict such access by requiring certain designated forms of age proof, such as a verified credit card or an adult identification number. In its analysis of the CDA, the Court focused on the “special First Amendment concerns [created by the Act] because of its obvious chilling effect on free speech.”²⁶ First, the Court pointed out that age verification systems were “not economically feasible for most noncommercial speakers.”²⁷ As a result, the CDA pursued its goals of protecting children from potentially harmful materials “by suppressing a large amount of speech that adults have a constitutional right to send and receive”—that is, indecent materials and materials that “describe, in terms patently offensive as measured by contemporary community standard, sexual or excretory activities or organs.”²⁸ This effect doomed the CDA. It is worth pointing out that the effect of filtering techniques on speech that adults have a constitutional right to send and receive is similar to that of age verification systems, because filtering techniques block materials based on keywords, indiscriminate of the age of the recipient. The chilling effects of such techniques, therefore, are at least as strong as those that led the Court to strike down the CDA. Further, the Court, in its detailed analysis of the dooming chilling effects, pointed out that even where “technologically feasible to block minors’ access to newsgroups and chat rooms containing discussions of art, politics, or other subjects that potentially elicit ‘indecent’ or ‘patently offensive’ contributions, it would not be possible to block their access to that material and ‘still allow them access to the remaining content, even if the overwhelming majority of that content was not indecent.’”²⁹ Here again, the similarity with the chilling effects created by

²⁵ *Id.*

²⁶ *Id.* at 845.

²⁷ *Id.* at 847.

²⁸ *Id.* at 846. The Court noted that the definition of the speech was too vague, as opposed to the three-prong obscenity test set forth in *Miller v. California*, 413 U.S. 15, 24 (1973). *Id.* at 845–46.

²⁹ *Reno*, 521 U.S. at 856.

filtering techniques is obvious. As has been shown above, because of its semantic incompetence, filters indiscriminately block both materials which are not constitutionally protected—obscene materials—as well as materials that are afforded full First Amendment protection. This is the very externality that doomed the CDA: “Given the vague contours of the coverage of the statute, it unquestionably silences some speakers whose messages would be entitled to constitutional protection.”³⁰ And the Court concluded that:

[a]s a matter of constitutional tradition, in the absence of evidence to the contrary, we presume that governmental regulation of the content of speech is more likely to interfere with the free exchange of ideas than to encourage it. The interest in encouraging freedom of expression in a democratic society outweighs any theoretical but unproven benefit of censorship.³¹

Labeling is another technique which the censoring world has and continues to experiment with in order to curb free flow of information. The best explanation of a labeling system is, perhaps, provided by the French government, in a 1999 policy paper:

The various segments of the general public must be allowed to protect themselves against certain types of site content. In the case of children, for example, web filter software is available today which makes it possible to select the types of content that they are allowed to access. To accomplish this, sites must first define the target audience for their content, using a standardized ‘label.’ Such tools are particularly important where Internet access in schools is concerned, and more generally provide one answer to the problem raised by the Internet’s ability to provide immediate access to any kind of site content. Site labelling is essential if self-regulation by private-sector players is to succeed. It will also make it possible to adopt a shared approach at international level to cover different legal systems. In order to ensure that the most vulnerable cybernauts are protected, the authorities will continue to support, as they have in the past, the development of French Web filter and site labeling software, as well as the testing of such systems. Along side this, work will be done with industry players in order to encourage them to adopt a satisfactory site labelling policy. In

³⁰ *Id.* at 874.

³¹ *Id.* at 885.

particular, the authorities will seek to ensure that the various labels are allocated with due regard for the rules of fair competition.³²

Two phrases in the above statement are crucial if one is to understand the implications of a labeling system: “sites must first define the target audience for their content, using a standardized ‘label,’ ” and “the authorities will seek to ensure that the various labels are allocated with due regard for the rules of fair competition.”

In other words, under such a set-up, sites must register with a local authority in order to obtain a label. And such a system can only be effective if all Internet service providers in a particular jurisdiction block by default all content that has not been registered. This leads to an implementation of the very strict definition of censorship: one must obtain *prior* governmental approval in order to be published. The implication of such a model is drastic as far as access to information is concerned. Today, one located in a country that does not censor Internet content can access material produced and published anywhere in the world, *unless the sender herself* prevents the free dissemination of its content using, for example, IP-based geolocalization techniques.³³ Under a labeling system however, residents of a country could only receive material published by speakers that actively target

³² Consultation Publique sur l’Adaptation du Cadre Législatif de la Société de l’Information [Policy Paper on the Adaptation of the Legal Framework to the Information Society], pt. 1.3 (Oct. 1999) (on file with the North Carolina Journal of Law & Technology) [hereinafter Policy Paper].

³³ These techniques are widely used by the entertainment industry, which uses them to enforce geographical windowing in order to maximize the return of their digital information properties. *See generally* CARL SHAPIRO & HAL R. VARIAN, INFORMATION RULES: A STRATEGIC GUIDE TO THE NETWORK ECONOMY (1999). The same techniques are also used by portals and content providers to prevent access by residents of certain jurisdictions to material deemed illegal in the jurisdiction at stake, in order to minimize the portals and content providers’ liability risks. For example, it was reported that Yahoo!, Flickr, and Microsoft’s Bing have introduced filters to prevent Indian users from accessing sexual content, as a response to India’s Information Technology Act of 2000, which bans the publication of pornographic material. *See* Gethin Chamberlain, *No Internet Sex Please, We’re Indian. Web Firms Observe New Law*, THE GUARDIAN, Dec. 28, 2009, available at <http://www.guardian.co.uk/technology/2009/dec/28/sex-internet-india-law>.

recipients in the country at stake and that have passed administrative muster. This means that the resources in the country at stake would be depleted from the greater part of the Internet, since most Internet speakers do not have the resources to register with the labeling authorities of the other countries in the world.³⁴

While France eventually abandoned such plans amid protests and a possible understanding that such a scheme would be unconstitutional,³⁵ such plans come back as cyclically as an ocean tide. For example, under a law passed in May and September 2009, the French government has created new flow-breaking gates by creating incentives, in the form of criminal penalties, for individuals to password-protect their personal wifi hubs, as I have detailed at length elsewhere.³⁶ And while the final text did not include a labeling provision, the government originally discussed

³⁴ Including the 192 United Nations members and miscellaneous territories which have claimed independence but have yet to be recognized. See United Nations, Member States, <http://www.un.org/en/members/growth.shtml#2000> (last visited Mar. 2, 2010) (on file with the North Carolina Journal of Law & Technology).

³⁵ The French Constitutional Council had previously hinted that for the executive to set up standard-setting bodies in the area of speech was unconstitutional under separation of powers principles. CC decision no. 96-378DC, July 23, 1996.

³⁶ Julien Mailland, France's Rhetoric of Control and the New "Anti-Piracy" Apparatus (May 28, 2009), <http://www.globalmedialaw.com/blog/?p=430>, (on file with the North Carolina Journal of Law & Technology). The law was eventually struck down by the French constitutional court. CC decision no. 2009-580, June 10, 2009, *available at* <http://www.conseil-constitutionnel.fr/conseil-constitutionnel/francais/les-decisions/2009/decisions-par-date/2009/2009-580-dc/decision-n-2009-580-dc-du-10-juin-2009.42666.html>. The law was passed again by parliament with minor modifications and eventually passed constitutional muster. The law has been dubbed a "three-strike" law. An administrative agency, named "HADOPI," is responsible for spying on internet communications and detecting "suspected" acts of illegal downloading of copyrighted materials. It is responsible for notifying the suspected illegal downloader of its suspicion. After three notifications, the suspected culprit's Internet connection is shut down for up to one year. Further, failure to "secure" one's internet access (including wifi access), can lead to a one-month access suspension and a fine of up to EUR 1,500. See Law of Sept. 22, 2009, *available at* <http://www.assemblee-nationale.fr/13/ta/ta0337.asp>.

requiring operators of public wifi hubs to block the entire Internet except for a list of “white sites” selected by the executive.³⁷

It is therefore clear that filtering and labeling systems are damaging to freedom of access to information, not simply because they form a basis for effective Internet censorship, but also because they create massive unwanted externalities. They prevent the reception of material that is not itself illegal or itself targeted. In the case of filtering, the spillover effects come from the fact that software, until now, has not been able to understand semantics. In the case of labeling, it comes from the fact that a speaker who has not taken the time to register with the politbureau cannot enter the public discursive space.

In the case of the W3C’s work, externalities have long existed that would provide censors with tools to give substance to their efforts to “embed or hardwire the legal regime in the technology itself,”³⁸ to quote James Boyle—in this case, a legal regime of prior governmental approval for publication (censorship). Starting in 1995, the W3C developed what it calls a *Platform for Internet Content Selection* (“PICS”), a “specification [that] enables labels (metadata) to be associated with Internet content.”³⁹ In practice:

[a] browser or stand-alone software filter can be set to check labels supplied by an independent rating agency before connecting to a chosen site. . . . When an end-user asks to see a particular URL, the software filter fetches the document but also makes an inquiry to the label bureau to ask for labels that describe that URL. Depending on what the labels say, the filter may block access to that URL.⁴⁰

While these technologies have long existed, they have been prevented from being used on a large scale by their inherent limitations. What has so far protected citizens eager to access information freely from filtering and censorship schemes was in

³⁷ Mailland, *supra* note 36.

³⁸ James Boyle, *Foucault in Cyberspace: Surveillance, Sovereignty, and Hardwired Censors*, 66 U. CIN. L. REV. 177, 188 (1997–1998).

³⁹ W3C, Platform for Internet Content Selection, <http://www.w3.org/PICS/> (last visited Mar. 2, 2010) (on file with the North Carolina Journal of Law & Technology).

⁴⁰ W3C, PICS Frequently Asked Questions, <http://www.w3.org/2000/03/PICS-FAQ> (last visited Mar. 3, 2010) (on file with the North Carolina Journal of Law & Technology).

part the fact that such systems involved so many spillover effects that censors were, in certain countries and to a certain extent, reluctant to adopt them on a large scale. Ironically though, the creation of more subtle filtering and labeling tools with less spillover effect might lead to more censorship. The Supreme Court itself left the door open for the implementation of such techniques when it suggested, in reference to the CDA, that the Act's "burden on adult speech is unacceptable if less restrictive alternatives would be at least as effective in achieving the Act's legitimate purposes."⁴¹ The burden to prove that no less restrictive alternatives exist is on the government, and, in the CDA's case, the Court noted that "[t]he Government has not proved otherwise."⁴² Therefore, the Court partially struck down the Act. *A contrario*, a more refined tool of content control—one that would achieve a legitimate government interest without otherwise creating externalities, chilling speech—might pass constitutional muster. Enter the semantic web.

III. OF BOXES AND PUSSYCATS: SNATCHING THE MEANING OF THE SEMANTIC WEB

Perhaps a quote from the late George Carlin's *Filthy Words* could best explain the concept behind the semantic web. The satiric monologue is best known for providing the factual basis upon which the Supreme Court of the United States delimited the rules of indecent broadcasting in the 1978 case *FCC v. Pacifica Foundation*.⁴³ In the speech, Carlin recites the following, amongst other things:

[n]ow the word twat is an interesting word. Twat! Yeh, right in the twat. Twat is an interesting word because it's the only one I know of, the only slang word applying to the, a part of the sexual anatomy that doesn't have another meaning to it. Like, ah, snatch, box and pussy all have other meanings, man. Even in a Walt Disney movie, you can say, we're going to snatch that pussy and put him in a box and bring him on the airplane.⁴⁴

⁴¹ *Reno v. ACLU*, 521 U.S. 844, 846 (1997).

⁴² *Id.*

⁴³ 438 U.S. 726 (1978).

⁴⁴ *Id.* at 755 (quoting transcript of George Carlin's *Filthy Words*).

If all words, like the word *twat*, had only one meaning, filtering systems would not have spillover effects. And if all countries in the world agreed that a given word should not be pronounced, that would also minimize the spillover effects of labeling systems, because one label granted by one global label agency would serve as a passport to all countries in the world.⁴⁵ The burden would therefore be minimal, if one were to morally accept the premise of censorship. But then there are words like *snatch*, which, like the word *breasts* or the name *Hitler*, have different meanings and implications in different contexts. If software was intelligent and was able to tell differences by contextualizing language, then the spillover effects would be minimized. The semantic web aims at doing just that. Behind the semantic web is the goal to allow software to understand that the expression “we’re going to snatch that pussy and put him in a box and bring him on the airplane” belongs to a Disney movie, and to perceive the situations in which, when the words are reshuffled, they actually originate from *Hustler* or *Playboy* magazines.

Tim Berners-Lee recognized the limitations of computer software as it then existed when he wrote that “[m]ost of the Web’s content today is designed for humans to read, not for computer programs to manipulate meaningfully. . . . [I]n general, computers have no reliable way to process the semantics.”⁴⁶ As a result, he suggested a new approach for encoding data in which “[t]he Semantic Web will bring structure to the meaningful content of Web pages, creating an environment where software agents roaming from page to page can readily carry out sophisticated tasks for users. . . . The Semantic Web will enable machines to COMPREHEND semantic documents and data.”⁴⁷

What this implies is that the semantic web can minimize spillover effects of filtering, because it can contextualize words. The machine will know when the word *breast* is used in the context

⁴⁵ In fact, such systems exist. For example, in the world of finance, where a financial institution registered in one European Union country can carry operations in all EU countries under what is known as the “European Passport Procedure.”

⁴⁶ Berners-Lee, Hendler & Lassila, *supra* note 1, at 36.

⁴⁷ *Id.* (emphasis in original).

of medical information and when it is used to appeal to the prurient interest. This will allow for more targeted and efficient filtering. Ironically, it might also lead to more filtering than before, since in a number of countries such as the United States, excessive spillover effects were the barriers that kept filtering from becoming the norm.⁴⁸

The semantic web also remains, just like PICS, a labeling system based on meta-tagging, albeit a more sophisticated version than PICS. As described by Martin Hilbert:

[t]he applied logic is basically similar to the functionality of search engines in libraries. Any kind of information and text can be classified into different categories of meta-information, adding "information about information." . . . Of course, the machine does not automatically "know" that a book with the title "democratic theory" belongs to the category "political science." The machine does not "understand" the title of the book in a traditional sense. It is the additional tag of meta-information that allows the machine to "understand" where this specific book belongs.⁴⁹

The system is based on a metadata model called the Resource Description Framework ("RDF").⁵⁰ It is a metadata model that provides the structure to make statements about Web resources in the form of "subject-predicate-object" expressions called "triples."⁵¹ The subject identifies the resource about which the statement is made. The predicate expresses a relationship between the subject and the "target" object. Taking public budget priorities as an example, one way to represent the notion "Citizen B prioritizes education" in RDF is as the triple: subject = "Citizen B"; predicate = "priority"; object = "education."⁵² Further, "[a]dditional ontology languages have been built upon the RDF syntax. W3C's Web Ontology Language (OWL), for example, enables users to describe the types of relationships ('property assertions') between a set of individuals ('classes'). The

⁴⁸ See *Reno v. ACLU*, 521 U.S. 844 (1997).

⁴⁹ Martin Hilbert, *The Maturing Concept of E-Democracy: From E-Voting and Online Consultations to Democratic Value Out of Jumbled Online Chatter*, 6 (2) J. INFO. TECH. & POL. 87, 99 (2009).

⁵⁰ *Id.*

⁵¹ *Id.*

⁵² *Id.*

relationships are axioms that provide semantics by allowing systems to infer additional information based on the data explicitly provided.”⁵³

It is worth noting that the basic logic behind the semantic web relies on “users” to tag information. This will raise issues of cultural hegemony because one can wonder who exactly the users in question will be. If the “users” are predominantly from Western developed countries, it is likely that the semantic web will replicate the linguistic, cognitive, and ideological frames of these Western, developed countries, which in turn would support an Internet-age Gramscist argument that digital media is a key player in propagating and maintaining the dominant ideology.⁵⁴ While these questions are outside the scope of this paper, they are important to point out.

For the purpose of understanding the effects of the semantic web, two points are essential to recognize. First, the semantic web promises to dramatically reduce the spillover effects of filtering, because, if it is widely implemented, the system will allow for filters to contextualize information and to let through information which, literally speaking, could be unwanted—“dick,” “breast,” “snatch,” “Hitler”—but which is now deemed to be part of a “safe” context. Second, the semantic web relies on meta-tagging to organize information. The use of meta-tagging in this context will by no means negate the anti-democratic character of labeling. One will still rely on third parties to categorize speech. This raises the issue of due process as, in a modern democracy, only an independent judiciary should be able to make a final determination as to what behavior is legal and what is illegal.⁵⁵ Under a semantic

⁵³ *Id.* at 100.

⁵⁴ See *supra* note 15 and accompanying text.

⁵⁵ See e.g., in the U.S. context, the statement of administrative law expert Bernard Schwartz: “The requirement of due process has been interpreted as requiring a formal adversarial hearing—what has come to be called an evidentiary hearing—before administrative decisions which adversely affect private individuals may be made. . . . This means that, before an administrative decision which adversely affects an individual may be made, that person has a right to an evidentiary hearing, which means ‘a hearing closely approximating a judicial trial.’” Bernard Schwartz, *Administrative Procedure*, in

web model, the meta-tagging function will most likely end up in the hands of the executive branch, more specifically in the hands of what could be dubbed in democratic countries an “independent administrative agency”—a friendlier word than “politbureau,” but a politbureau nonetheless. Further, the system still implies, as an inevitable consequence, that un-tagged or un-labeled speech will by default be blocked—otherwise, what good are labels for? This necessarily implies that the semantic web, as it is envisioned today, can only live in a system of strict censorship. At the risk of appearing sanguine, censorship is the semantic web’s lifeblood.

For a long time, the development of the Internet, in terms of free flow, was the result of the interplay between the efforts of civil society, led by civil libertarian thinkers, computer scientists, and governments, who accurately perceived the Internet as a threat to their sovereignty. But more and more, the business sector played a part in shaping the technology, in shaping the code that is the law of the Internet to paraphrase Lawrence Lessig.⁵⁶ AOL took pride in its “gated community.”⁵⁷ Hollywood realized that IP-based geolocation techniques could support enforcement of geographical windowing for the release of its communication goods.⁵⁸ Google, whose informal motto is “don’t be evil,” provides Chinese residents with a culturally-clean version of its search results using filters, as has been well documented by the Harvard Berkman Center.⁵⁹ And Yahoo!, Flickr, and Microsoft’s

FUNDAMENTALS OF AMERICAN LAW 139 (Alan B. Morrison ed., 1996) (quoting *Mathews v. Eldridge*, 424 U.S. 319, 325 (1976)).

⁵⁶ LAWRENCE LESSIG, CODE AND OTHER LAWS OF CYBERSPACE (1999).

⁵⁷ Consider, e.g., David Carr, *AOL Blossoms as Print Retreats*, N.Y. TIMES, Aug. 17, 2009, at B1 (noting that AOL was “often derided as the original gated community”).

⁵⁸ See *supra* note 33 and accompanying text.

⁵⁹ Consider for example the Berkman Center’s OpenNet Initiative report, which states: “While Google and Microsoft, which are hosted outside China, actually de-listed certain search results, the two search engines hosted inside China, Yahoo! and Baidu, ran their Web crawlers behind the China’s filtering system, and therefore did not index Web sites already blocked by the Chinese government. Although Google censored considerably less than [sic] the other search engines, it also has a practice of prioritizing authorized local content, which researcher Nart Villeneuve found amplified the significance of the censored Web sites as they were the only ones to offer differing viewpoints.”

Bing seem to have introduced filters to prevent Indian users from accessing sexual content, as a response to India's Information Technology Act of 2000, which bans the publication of pornographic material.⁶⁰ PICS, as a labeling system, failed to be adopted on a wide scale in part because of its imperfections and crude character.⁶¹ The semantic web, however, as a more refined classification tool, is very appealing to the private sector.

Unfortunately, the private sector, in engaging the semantic web, is in effect furthering the development of censorship systems.

OpenNet Initiative, *Internet Filtering in China*, June 15, 2009, <http://opennet.net/research/profiles/china> (on file with the North Carolina Journal of Law & Technology). According to recent reports, however, following attacks suffered by Google that originated in China and were widely attributed to the Chinese government, "[a]fter revealing that the attacks targeted not only its core intellectual property but the e-mail accounts of Chinese human-rights activists, Google announced that it would stop censoring search results on Google.cn, its Chinese-language search engine." Michael Moyer, *Internet Ideology War: Google's Spat with China Could Reshape Traditional Online Freedoms*, SCI. AM., Feb. 25, 2010, <http://www.scientificamerican.com/article.cfm?id=internet-ideology-war> (on file with the North Carolina Journal of Law & Technology).

⁶⁰ See *supra* note 33.

⁶¹ See, e.g., Cyber-Rights & Cyber-Liberties (UK) Report, *Who Watches the Watchmen: Internet Content Rating Systems, and Privatised Censorship*, Nov. 1997, <http://www.cyber-rights.org/watchmen.htm> (last visited Mar. 2, 2010) (on file with the North Carolina Journal of Law & Technology) ("The second stage in content control thus began with the introduction of rating and filtering products that claim to permit users to block unwanted material from their personal systems. The most sophisticated and widely recognised of these systems is the Platform for Internet Content Selection ('PICS'), introduced by the World Wide Web Consortium. European governments were especially interested in this hoped-for solution. They backed away quickly from incidents in the first stage of direct suppression and put forward PICS and rating systems as a proposed standard, both through national governments and the European Union as a self-regulatory solution to Internet content. There are many problems, however, in rating and filtering systems as will be explained in this report. They are crude and tend to block too many sites. Most focus on the World Wide Web, offering no way to block objectionable content on other distribution mechanisms of the Internet such as newsgroups and ftp sites. Each system is extremely subjective and affected by cultural assumptions, so international exchanges of systems will not satisfy users. Finally, the systems were designed for individual users and do not scale well to use by entire countries and third parties.").

In the following part, I will argue that the semantic web will create a “digital public sphere,” a global marketplace of ideas where publishers will offer their content and present it with labels, enabling governments worldwide to easily select content for redistribution in their respective jurisdictions through automated means. The result would be an optimized, global censorship system, where it would not be the best idea that prevails but the one most suited to the receiver’s regulatory environment.

IV. THE SEMANTIC WEB AT WORK

In a capitalist society, the business sector’s main goal, by definition, is to create wealth. While some businesses may have other considerations in mind—the craftsman, the so-called “green bank,” the so-called “ethical investor”—in reality the aggregation of value comes before the development of democracy for the overwhelming majority of businesses. A close examination of case studies published by the W3C indicates both the ways in which businesses are refining the semantic web and pushing for its development for capitalistic interests, as well as the ways in which the business sector shows a complete disregard for the implication of its actions on democracy in this case. This is demonstrated through two case studies discussed below, which, in turn call for a greater involvement of the computer science community to neutralize the semantic web’s externalities.

A. *Search Thresher: A Censor’s Dream, or How the Dream of Internet Visionaries Could Turn into their Worst Nightmare*

Segala, a corporation which defines itself as “one of Europe’s most trusted suppliers of testing and certification services,” promotes a product they call Search Thresher.⁶² Segala sees a problem:

Search results are not as reliable as they can be. They don’t provide any form of trust to end users to help them make informed decisions before entering Web sites. Some users may not wish to wade through search results before discovering a site that “looks” like it ‘can’ be trusted, or until they stumble upon a site with a Trustmark. Some users

⁶² Segala, <http://segala.com> (last visited Mar. 2, 2010) (on file with the North Carolina Journal of Law & Technology).

may only trust Web sites that have been vetted by an independent authority to guarantee that the information provided is trustworthy enough to rely on it. For example, if you conducted a search on treatment for a particular illness, how would you know which Web sites to trust?⁶³

Segala also sees a solution:

Based on the principles of the Semantic Web, Content Labels are files that contain Metadata that enable search engines and browsers to provide more information about trust in search results. Like the title and description tags, Content Labels can be read and “utilised” by search engines and browsers to display more information about a Web site in search results.

“In its earliest days, W3C recognized a need to be able to describe content according to a defined vocabulary. This could be done for a variety of reasons including, but not limited to, child protection. The result was the PICS system which, despite early promise, has achieved limited support.” Content Labels will be proposed as a replacement of PICS now that it has made it onto a W3C Full Recommendation Track.⁶⁴

What is interesting is that the rhetoric used by Segala is the very same rhetoric used by Western censors, such as the authors of the aforementioned 1999 French policy paper.⁶⁵ First, it implies that recipients of information need outside help in order to make informed decisions, saying “[sites] don’t provide any form of trust to end users to help them make informed decisions.” Second, it also implies that absent the outside help, terrible consequences may occur, warning “if you conducted a search on treatment for a particular illness, how would you know which Web sites to trust?” The implication is that you could remain gravely ill but for that third party’s assistance, or you could become even more ill if you were to follow untrustworthy advice you found on the Internet. Finally, the rhetorical argument focuses on the role of an “independent authority” by saying “[s]ome users may only trust Web sites that have been vetted by an independent authority.” The

⁶³ DAVID ROOKS, CASE STUDY: IMPROVING THE RELIABILITY OF INTERNET SEARCH RESULTS USING SEARCH THRESHER (2007), <http://www.w3.org/2001/sw/sweo/public/UseCases/Segala/> (last visited Mar. 2, 2010) (on file with the North Carolina Journal of Law & Technology).

⁶⁴ *Id.*

⁶⁵ See Policy Paper, *supra* note 32.

rhetoric is the very same that was used by the French government when it suggested that a labeling system must be put in place.⁶⁶ First, outside help is needed: “The various segments of the general public must be allowed to protect themselves against certain types of site content.”⁶⁷ Second, absent the outside help, terrible consequences may occur: “In the case of *children*, for example, web filter software is available today which makes it possible to select the types of content that they are allowed to access . . . [i]n order to ensure that the most *vulnerable* cybernauts are protected.”⁶⁸ Notice the use of children in order to evoke fear. Finally, it too focuses on the role of an “independent authority,” saying that “the authorities will seek to ensure that the various labels are allocated with due regard for the rules of fair competition.”⁶⁹

And since the French government “support[s], as [it] ha[s] in the past, the development of French Web filter and site labeling software, as well as the testing of such systems,”⁷⁰ Segala decided to capitalize on such a promising market and develop Content Labels, which are “files that contain Metadata that enable search engines and browsers to provide more information about trust in search results.”⁷¹

As discussed in Part I, however, labeling systems create very strong externalities in the form of chilling effects when speakers do not play along, because no data can pass through the governmental safety seal unless they have been properly labeled. Information that could have received a favorable label gets blocked when its publisher has not taken the steps necessary to receive the cyber-transit visa. As a result, the local Internet is deprived of valuable material; hence an interest exists for the government to “work . . . with industry players in order to encourage them to adopt a satisfactory site labeling policy.”⁷²

⁶⁶ *Id.*

⁶⁷ *Id.*

⁶⁸ *Id.* (emphasis added).

⁶⁹ *Id.*

⁷⁰ *Id.*

⁷¹ See ROOKS, *supra* note 63.

⁷² See Policy Paper, *supra* note 32.

B. *Digital Habermas*

If the foregoing spillover effects of labeling can be mitigated, then there is a risk that the system will actually be adopted by governments on a very large scale. The mitigation of this, as described above, requires not only the participation of the government as label grantor but also the participation of publishers, as label requesters, and even label developers. Successful development of labeling systems, therefore, requires a digital version of Habermas' public sphere: an area in social life where people can get together and freely discuss and identify societal problems, influencing political action through that discussion, "a discursive space in which individuals and groups congregate to discuss matters of mutual interest and, where possible, to reach a common judgment,"⁷³ a place that mediates between the "private sphere" and the "Sphere of Public Authority."⁷⁴

If the private sector is to participate in such discourse, it needs incentives. Private sector W3C participants, and, in particular, content management companies such as Saltlux,⁷⁵ have been hard at work demonstrating the financial benefits the private sector might derive from its solutions. In the field of mobile content delivery, for example, Saltlux recognizes the fact that "[t]o increase the user friendliness in terms of the content and services, the correct delivery system has to be fitted for each user and the appropriate content must be delivered to each user at the appropriate time and using appropriate means."⁷⁶ Here, we see a clear understanding by the private sector of the need for compatibility between the content and *each* user—or, if one were to take the logic further, between the content and each region or

⁷³ Gerard A. Hauser, *Vernacular Dialogue and the Rhetoricity of Public Opinion*, 65 COMM. MONOGRAPHS 83, 86 (1998).

⁷⁴ HABERMAS, *supra* note 8.

⁷⁵ Saltlux, <http://www.saltlux.com> (last visited Mar. 2, 2010) (on file with the North Carolina Journal of Law & Technology).

⁷⁶ TONY LEE ET AL., USE CASE: MOBILE CONTENT RECOMMENDATION SYSTEM (SEPT. 2008), <http://www.w3.org/2001/sw/sweo/public/UseCases/SaltLux-KTF> (last visited Mar. 2, 2010) (on file with the North Carolina Journal of Law & Technology).

each *jurisdiction*. Thus, the developer continues, “it is necessary to have a system that can provide identification between user preferences and user relationships.”⁷⁷

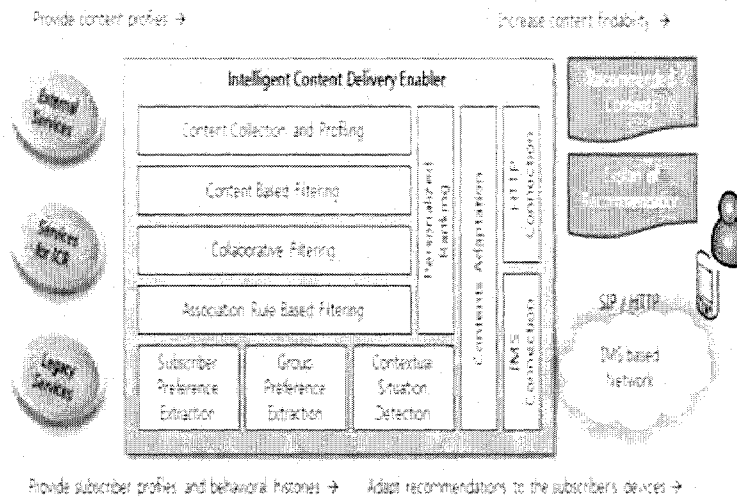
The needs addressed here are the same needs that transnational corporations would have, should different labeling systems be implemented on a large scale by censors in the various jurisdictions of the world. They also address the same issue which the same transnational corporations are faced with today—that is, the difficulty of complying with the laws of every country, each having different values and different content control laws. If technology can develop a digital sphere where various governments could present the private sector with their requirements, and where private content could be contextualized by intelligent software agents and routed only through the appropriate pipes but, crucially, through *all* the appropriate pipes, then the spillover effects of labeling might be mitigated to a point that would make it financially acceptable for the private sector to adopt and for governments worldwide to implement such controls without depleting the local Internet of too much useful information.

The Saltlux example indicates that such development of a “digital public sphere” is exactly what is happening at the moment. Saltlux suggests the following:

Individualized user preferences are identified through analysis of both wired and wireless use histories. Identified preferences are used to determine suitable content recommendations using techniques of content-based filtering and collaborative filtering. It enables content recommendations suitable to the user’s existing situation by application of recommendation policies according to the user’s current situational information such as user preference, time, place, weather, schedule, etc. Through reasoning of the social relationship between users, stereotypes of subscribers whose preference could not be retrieved can be assumed and recommended, or contents of other subscribers who are related to the current subscriber are recommended.⁷⁸

⁷⁷ *Id.*

⁷⁸ *Id.*



[Chart available at <http://www.w3.org/2001/sw/swco/public/UseCases/SaltLux-KTF/>]

This illustrates how my “digital public sphere” metaphor would work in practice. The digital public sphere is represented in the above chart by the “Intelligent Content Delivery Enabler.” On the left side of the sphere are the publishers (referred to here as External Services), which “provide content profiles”—that is, information about the content—using meta-tags. On the right side of the sphere appear censors and their recommended content. Using semantic web concepts, the software agent “enables content recommendations suitable to the user’s existing situation by application of recommendation policies according to the user’s current situational information.”⁷⁹ Saltlux refers to situational information “such as user preference, time, place, weather, schedule,”⁸⁰ but the preferences could of course be those of the ruler of the user’s jurisdiction and be based on considerations less trivial than weather.

In other words, such a digital public sphere would serve as a marketplace of ideas, where publishers would offer their content and present it with labels, and governments would select such

⁷⁹ *Id.*

⁸⁰ *Id.*

content for redistribution in their respective jurisdictions—all of it in automated ways made possible by semantic web technology. The result is an optimized, global censorship system where receivers based in any given jurisdiction would only be able to receive content deemed suitable by the ruler of said jurisdiction. In the words of Saltlux, “[c]ontext-awareness-based content recommendations through consideration of physical, theoretical and social situations of preference of subscribers, groups and stereotypes enables the optimization of access costs and increases the usability and convenience of information discovery through providing customized contents easily and conveniently to users.”⁸¹ This calls for a greater involvement of the computer science community to neutralize the semantic web’s externalities.

V. CONCLUSION

The suggestion that engineers must refocus on free flow of information relies on two assumptions. First, it assumes that the semantic web is indeed the path on which influential Internet actors are walking. Second, it assumes that the development of the semantic web bears a very real potential to unintentionally provide new tools for enhanced Internet censorship and that we must explore how to neutralize this unwanted externality of an otherwise useful tool of information organization to avoid censorship. As has been discussed, I believe both assumptions are true.

First, the semantic web is based on meta-tagging, on labels. As we have seen, there is no point in having a labeling system unless the labels are actually used to block content at the door. Labels are used to block content before it is received. This matches the very definition of censorship, *a priori* suppression of speech. The semantic web, therefore, as is currently conceived, is, in its very nature, a censorship enabler as much as it is a tool of improved access to information. Censorship is the semantic web’s lifeblood.

Second, the semantic web is the future, as far as for-profit Internet actors are concerned, because by definition those actors are concerned with increasing profit, not with furthering the information flow for the purpose of developing democracy and

⁸¹ *Id.*

freedom. As explained by Saltlux in relation to its mobile recommendation system, because the service provider adjusts to the needs of the user, and provides intellectual content, “it *increases its competitiveness* and is able to not only maintain its customers but also *increases its market share*.”⁸² These business interests are identical to those that led Yahoo!, once the self-proclaimed defender of free speech, to turn over Chinese dissidents to the Chinese government and to block residents of certain jurisdictions from accessing certain content.⁸³ Facts have proven that large and influential Internet players prefer profits to the free flow of and access to information,⁸⁴ and the semantic web is the tool that will allow them to keep arbitrating content flow in even more efficient ways. Also, even though the debate on net neutrality is too wide to be addressed comprehensively in this paper, it is nonetheless worth noting that the semantic web is a tool that could also be used successfully in the context of undermining net neutrality. For example, if AOL Time Warner is able to realize, using semantic web tools, that the sentence “[w]e’re going to snatch that pussy and put him in a box and bring him on the

⁸² *Id.* (emphasis added).

⁸³ See *supra* note 59. Chinese cyber-dissidents and activists, such as the journalist Shi Tao, have been convicted in part because of some email service providers’ disclosure of their users’ personal information to the Chinese police. Press Release, Reporters Without Borders, Cyber-dissident Convicted on Yahoo! Information Is Freed After Four Years (Nov. 9, 2006), *available at* http://www.rsf.org/article.php3?id_article=8453. See also Human Rights in China, Case Highlight: Shi Tao and Yahoo, <http://hrchina.org/public/highlight/index.html> (last visited Mar. 15, 2010) (on file with the North Carolina Journal of Law & Technology).

⁸⁴ See *supra* note 59. It is also worth noting that Yahoo! insisted that its move had nothing to do with a 2000 French court case enjoining Yahoo!, Inc to get rid of links to neo-Nazi websites and to take down Nazi memorabilia from its auction service, a decision Yahoo! insisted was not enforceable in the United States. It did, however, “voluntarily” take down the Nazi memorabilia from its auction website. See Press Release, Yahoo!, Yahoo! Enhances Commerce Sites for Higher Quality Online Experience (Jan. 2, 2001), *available at* <http://docs.yahoo.com/docs/pr/release675.html>. For a detailed analysis of the Yahoo! case and its impact on global information flow, see Julien Mailland, *Freedom of Speech, the Internet, and the Costs of Control: The French Example*, 33 N.Y.U. J. INT’L L. & POL. 861 (2001), *available at* <http://www1.law.nyu.edu/journals/jilp/issues/33/pdf/33aa.pdf>.

airplane” that is trying to pass through its nodes is an excerpt from a Disney movie, it might block it and instead promote its own content. Conversely, if a Disney-owned Internet service provider made the determination, using semantic-web tools, that said sentence is not from a Disney movie but actually an excerpt from a non-children-friendly George Carlin monologue, it might also prevent it from flowing through its own nodes.

All in all, it seems unavoidable that the business-and-censor-friendly digital public sphere is the way of the future, unless alternative models, re-embedding the values of freedom into the technology, prevail.

Recent history shows that the Internet’s positive impact on promoting the free flow of and access to information resulted because developers were always one step ahead of censors. The Internet would not be what it is had it been built from the top and not from the edges.⁸⁵ Governments, until now, have only *reacted* to things like decentralized networks, open source software, and peer-to-peer file exchange. And while governments have been hard at work to “embed or hardwire the legal regime in the technology itself,”⁸⁶ they have been working on old technologies, unaware of what the next development would be and how civil liberties would be affected. They have therefore been unable to react quickly and efficiently. With the semantic web, things might be changing, because the interests of the governments coincide with the interests of business in the digital public sphere and because business has come to control so much of the Internet.

⁸⁵ Vint Cerf, “father of the Internet,” explained that the Internet: has had the advantage of rapid innovation by users at the ‘edge’ of the network, largely without much or any regulatory interference. Indeed, because much of the flexibility of the Internet is a consequence of its dependence on software running in devices at the edge of the network, rather than in systems embedded in the net, virtually anyone is free to invent new applications and to put them up for use.

Vint Cerf, *Does the Internet Need to be Governed?*, CIRCLEID, Nov. 4, 2004, http://www.circleid.com/posts/does_the_internet_need_to_be_governed (on file with the North Carolina Journal of Law & Technology).

⁸⁶ See Boyle, *supra* note 38, at 188.

This is therefore a call to W3C and the computer science community at large that the semantic web's externalities are currently so strong that there is a real risk that the keys to the machine will be involuntarily handed out to censors. Developers need to refocus their attention on ways to neutralize these externalities in order to further develop the web of data while at the same time preventing control-oriented governments from taking advantage of technology intended to further, rather than curb, knowledge and information flow.

