Copyright Terms in Online Creative Communities

Casey Fiesler

Georgia Institute of Technology

Atlanta, GA 30332

casey.fiesler@gatech.edu

Amy Bruckman

Georgia Institute of Technology

Atlanta, GA 30332

asb@cc.gatech.edu

Abstract

A key usability problem for websites is the complexity of their terms and conditions. Within the HCI community, attention to this issue to date has primarily focused on privacy policies. We begin to build on this work, extending it to copyright terms. With so many people posting everything from status updates to digital art online, intellectual property rights are increasingly important to the end user. We conducted a content analysis of 30 different websites where users can share creative work, focusing on the licenses and usage rights that users grant to those websites. Due to difficult readability, legalese, and a lack of plain language explanations, it is likely that users may not know what rights they are granting. Next steps include a user survey to determine whether this is the case, and further exploration of the impact on usability.

Author Keywords

copyright; creativity; intellectual property; online communities; readability; terms of service

ACM Classification Keywords

K.4.1 **[Computers and Society]:** Public Policy Issues
--- *Intellectual property rights*

Introduction

Much of the content that we consume every day comes from amateurs on the web. From Facebook statuses to YouTube videos, more and more people are becoming content producers. They are sharing their work in a number of different online spaces, but when users put their content online, what rights to that content do they grant? Can websites do anything they wish with user content? What terms are users agreeing to when they post or upload?

Many users of the photo sharing site Instagram were seemingly unaware of the answers to these questions. In December 2012, a provision in their Terms of Service (TOS) allowing use of photographs "in connection with paid or sponsored content" gained some media attention. After users complained with fears that their content might be used in advertising without explicit consent, Instagram altered these licensing terms. This situation served as a reminder that those sharing creative work online may not always be aware of how their work can be used.

Considering that reading only the privacy policy of every site visited would take the average Internet user over 200 hours per year, it is not surprising that many do not take the time to read often complicated terms, conditions, and policies [4,9]. Though the readability of online privacy terms has been identified as a usability problem within the HCI community [4,7], intellectual property rights are also increasingly relevant to end users. Whether posting a social network status update or uploading a piece of art to an online community, creators are required to provide the websites with permission to display or use their work. These permissions are typically covered in the websites' terms, but we have no reason to think that these are read any more frequently than privacy policies. What rights in their creative work are users actually giving to websites? How could the work be used?

To begin exploring this usability problem, we examined the copyright license and rights related provisions for 30 different websites focusing on user-generated creative work. We wanted to find out what common provisions exist and in what frequencies. Based on these initial findings and past research, we also

hypothesize that users are not aware of these terms, and are preparing to test this hypothesis as a first step in determining how TOS might impact user experience in online creative communities.

Related Work
Much work within the HCI community has been devoted to issues of readability and web accessibility. Luger et al. pointed out that though these are major themes within HCI, the established tenants of good design are glaringly absent from consideration in crafting terms and conditions on websites [7]. Past work has shown a surprisingly high level of complexity within TOS, privacy policies, and other click-through conditions found on the web [4,7,8,9].

A number of researchers have examined privacy policies in detail, often analyzing their relationship to actual government regulations. Content analyses of the policies of library vendors, healthcare providers, and universities have shown that provisions are not always in line with regulations or expectations [1,2,8]. Some researchers have also taken steps toward solving the usability issue by making privacy policies more accessible—for example, Kelley's privacy nutrition labels [5]. Luger et al. have suggested that more transparency might be beneficial, providing a visualization for the readability of terms that users come across on the web [7]. Our preliminary results present copyright licenses as part of the overarching usability problem of website terms.

Terms of Service Analysis
In our prior work focusing on online creative communities in which issues of copyright are complicated by appropriation and remix [3], a question

that often arose was the relevance of the websites' TOS. This work is part of a further exploration of this space. Therefore, in creating a sample of sites to analyze, we began with a set of these same types of remix communities and added other user-generated content and social networking sites.

To ensure that the websites we examined would cover a variety of different types of creative communities, we sampled from four different media types: writing, music, art, and video. We used Alexa search engine rankings and keywords that specified remix or appropriation ("fan fiction" for writing, "music remix" for music, "fan art" for art, and "remix video" for video), and pulled out the highest ranked six websites for each category, manually checking that each was indeed a place for users to post work of that type. We limited ourselves to sites that are primarily in English.

In addition to these remix-based communities, we also included the most popular social networking and user-generated content websites as provided by a website that also bases its algorithm on Alexa rankings.[1] With 6 categories, this resulted in a total of 36 websites, 30 of which would be used in our final dataset. For each of these 36 websites, we retrieved the TOS as well as any supplemental copyright terms included on the site. In addition to analyzing the text of the TOS, we also gathered word count and reading level information.

For this first study, we focused just on the part of the terms that covered copyright licenses and rights—i.e.,

---

[1] http://www.ebizmba.com/articles/social-networking-websites, http://www.ebizmba.com/articles/web-2.0-websites

what licenses the user provides in their work by posting there and what rights of use the website claims in that work. The typical format of these terms is, for example, "You grant this website an A, B, and C license to X, Y, and Z." The first author is a law school graduate and copyright expert, and determined which sections of the terms were relevant.

Without knowing exactly what licenses and rights would be present, we used a quasi-grounded theory approach in which we updated our coding scheme as we went through the set of documents. As we added new codes, we re-coded previous documents. Similar grounded theory approaches have been used in content analyses for privacy policies [2]. In addition to coding for each license and right mentioned, we also coded for whether the site included plain language explanations of copyright terms, whether there were any explicit waivers of rights, and provisions for destruction of content. Some of the websites did not have any copyright terms, and this was also coded.

Our copyright expert coded the first 5 websites in each category, for a total of 30 documents in our data set. The sixth document in each set was used for training a second coder, and this data was thrown out. The second coder then coded 5 documents chosen randomly from the data set, in order to determine inter-rater reliability. A comparison of this independent coding achieved an 89% inter-rater reliability using Cohen's Kappa [6].

Preliminary Results
Readability
For each of the 30 websites in our dataset, we determined word count and Flesch-Kincaid Grade Level

Website	Word	F-K
Asianfanfics.com	5823	18
Fanfiction.net	3661	18.3
Harrypotterfanfiction	985	11.9
Archive of Our Own	9142	12.1
Twisting the Hellmouth	3005	9
Club Create	5811	16.3
Overclocked Remix	1170	15.4
ccMixter	2693	17.3
Sound Cloud	7961	19.8
Remix 64	N/A	N/A
The Otaku	1189	8.4
DeviantArt	4031	14.7
Y-Gallery	2368	10.6
Elfwood	2300	13
Fanart Central	2591	13.2
Vidders.net	1677	14.2
Ebaum's World	4659	15.1
Warcraft Movies	1837	19
YouTube	3764	16.6
Daily Motion	3223	17.6
Wikipedia	5773	15.2
Twitter	3486	15.3
Craigslist	5006	17.6
Flickr	5763	15.2
IMDB	2613	14.2
Facebook	4477	12.9
LinkedIn	7294	15
Pinterest	2219	12.5
MySpace	5486	16.9
Google+	1691	11.5

Score using statistics in Microsoft Word. This is a common readability measure that has been employed in similar studies for privacy policies [1, 4, 8]. These measures were based solely on the actual TOS (or similar document) rather than supplemental material. Only one site on our list (Remix64, a music remix community) did not have any TOS. The table to the left provides a summary of these measures for the websites in our data set.

The average Flesch-Kincaid Grade Level Score (representing a U.S. educational grade level) is a college sophomore reading level of 14.8, ranging from 8.4 to 19.8. This puts the readability of these documents roughly on par with those of privacy policies, which one study has shown to have an average score of 14.21 [4]. The average scores for video sites (17.7) and for music sites (17.2) does skew higher than the other media types, which could possibly be accounted for by the additional legal complexities associated with sound. Just as reading privacy policies for all the websites one visits might take years, at an average adult reading speed of 250 words per minute, these would take almost 8 hours to read.

Copyright Terms
When a website user submits content to one of these websites, they are typically licensing that work for use by the site—at the very least, the site must be permitted to display the work, or posting it there would be pointless. For every website except for one (Club Create, a music remix platform in which users remix provided samples live on the site), the site only licenses rather than requiring a transfer of copyright. Sixteen of the sites specifically state that the user retains copyright (or ownership) in the work.

Usually the *type* of license required by the site is stated in a string of legalese in the TOS. For example, the license provision for Facebook states: "You grant us a non-exclusive, transferable, sub-licensable, royalty-free, worldwide license to use any IP content that you post on or in connection with Facebook."

In our dataset, we saw the following types of licenses: revocable, irrevocable, assignable, limited, nonexclusive, paid, perpetual, royalty-free, sub-licensable, transferable, unrestricted, and worldwide. The most common are also the most unsurprising and necessary. Nonexclusive (19 instances) means that the user is free to use the same content however else they like, including licensing it to others. Royalty-free (18 instances) means that the website is not required to pay the user for their work. Worldwide (18 instances) means the license is effective all over the world.

However, we speculate that some of these might be more surprising to users, based upon instances like the alarm of Instagram's TOS change and other observations of norms about content use for Internet users, including in our prior work [3]. Six of the sites require "irrevocable" licenses, meaning that once they give the website the right to use their content, they cannot take it back. Eleven sites also require licenses that are sub-licensable, meaning that they can license the content to a third party.

In addition to these lists of license type, the license explanations were typically followed by an enumeration of specific uses the website could make of the work. For example, when posting on Craiglist the user is giving the website a license "to copy, perform, display, distribute, prepare derivative works from (including, without limitation, incorporating into other works) and

otherwise use any content that you post." Based on the provisions in our data set, we developed the following codes to cover these different usage rights: transmit, translate, enforce, reproduce, perform, modify, adapt, transform, index, improve, edit, distribute, display, compile, backup, analyze, advertising/promotion, commentary, commercial use, in connection with site business, use by other parties, use of name/likeness., and unspecified use.

Again, the most common codes were also the least surprising. To reproduce (17 instances), distribute (18 instances), and display (17 instances) are technically necessary in order to have the work appear on the website. However, perhaps more surprising to users would be the number of sites that require being able to *change* their work—we saw 18 instances of requiring a modify, adapt, or transform use. At times this may only mean something like formatting, and at others, as in the case of Craigslist noted above, the user gives the site the right to make and distribute derivative works. Additionally, two sites have the right to use the submitted work in marketing materials and are able to use the user's name and/or likeness. Two more sites required a waiver of moral rights and the right to publicity. Considering the social norms pertaining to commerciality within amateur creative communities, these terms are especially problematic [3].

Half of the sites we examined contain specific provisions for their right to destroy or delete the user's content. Eight specifically stated that this can only be done after a TOS violation or similar, and six can delete content for any reason.

Out of our 30 sites, only 5 included any kind of plain language explanation of these copyright provisions. For

example, Pinterest's TOS includes the statement: "More simply put, if you post your content on Pinterest, it still belongs to you but we can show it to people and others can re-pin it." Even this level of explanation was rare.

The majority of these sites included specific license and use rights provisions. Only one did not include a TOS, and another 5 did not have any information about copyright terms. These 5 were all smaller online remix communities. The number of codes (copyright provisions) generally did not fall into a pattern with respect to media type or popularity of the site. The site with the most number of codes (26), the small fan fiction website Asian Fan Fics, contains provisions for the site to essentially do whatever they like with whatever is posted there without any notice or attribution to the creator.

Continuing Work

Though our results suggest that unconscionable copyright terms are not lurking around every corner, we did identify terms that users might not expect. The potential for unusual uses of posted content is information that matters, and if this information is buried within difficult-to-read terms, then this is a usability problem.

Previous work suggests that online content creators can sometimes have surprising knowledge of copyright law, and that even in the context of confusions about the law there can be nuanced social norms about content use [3]. If Instagram users were upset about a single provision allowing advertising uses of content, imagine how the users of Asian Fan Fics might react to an explanation of their agreed-to copyright terms. However, given the high reading level score and the

lack of plain explanations in these terms, it is likely that many users would still not understand the agreement.

This work so far has provided us with an idea of the kinds of copyright provisions that creative sharing websites include. The next step is to determine how prevalent the usability problem really is—how misinformed are users, really? In order to answer this question, we are conducting a survey to find out if users are aware of these copyright terms in the websites they frequent. Using layman's explanations for the common licensing terms we found, we are asking users both whether they think that each website *can* use their work that way, and whether the website *should* be able to do so. This will provide us with information as well about whether the existing terms are surprising to users.

Part of the intended purpose of the terms and conditions on websites is informed consent. If we are correct and users are not aware of the rights in their creative work that they are giving away, this could be a serious usability problem. Finding ways to make the information more accessible would be a step towards solving this problem. Previous solutions have ranged from Kelley's privacy nutrition labels [5] to an open source browser plug-in (TOSDR.com) that rates websites' TOS from good to bad based on user input. Further explorations into this space should provide us with guidance as to what kinds of solutions might be appropriate for copyright terms as well.

Acknowledgments

References

[1] Culnan, M.J. and Carlin, T.J. Online Privacy Practices in Higher Education: Making the Grade? *Communications of the ACM 52* (2006), 126–130.

[2] Earp, J.B., Antón, A.I., Aiman-smith, L., and Stufflebeam, W.H. Examining Internet Privacy Policies Within the Context of User Privacy Values. *IEEE Transactions on Engineering Management 52*, 2 (2005), 227–237.

[3] Fiesler, C. and Bruckman, A. Remixers' Understandings of Fair Use Online. *Proc. CSCW 2014* (2014).

[4] Jensen, C. and Potts, C. Privacy Policies as Decision-Making Tools: An Evaluation of Online Privacy Notices. *Proc. CHI 2004* (2004), 471–478.

[5] Kelley, P.G., Cesca, L., Bresee, J., and Cranor, L.F. Standardizing Privacy Notices: An Online Study of the Nutrition Label Approach. *Proc. CHI 2010* (2010), 1573–1582.

[6] Krippendorff, K. *Content Analysis: An Introduction to Its Methodology*. SAGE Publications, Thousand Oaks, CA, 2004.

[7] Luger, E., Moran, S., and Rodden, T. Consent for All: Revealing the Hidden Complexity of Terms and Conditions. *Proc. CHI 2013* (2013), 2687–2696.

[8] Magi, T.J. A Content Analysis of Library Vendor Privacy Policies: Do They Meet Our Standards? *College & Research Libraries 71* (2010), 254–272.

[9] Mcdonald, A.M. and Cranor, L.F. The Cost of Reading Privacy Policies. *I/S: A Journal of Law and Policy for the Information Society 4* (2008), 543–565.