

# “I Am Not a Lawyer”: Copyright Q&A in Online Creative Communities

Casey Fiesler  
GVU Center  
Georgia Institute of Technology  
85 5th St. NW Atlanta, GA 30332  
casey.fiesler@gatech.edu

Jessica Feuston  
GVU Center  
Georgia Institute of Technology  
85 5th St. NW Atlanta, GA 30332  
jfeuston@gatech.edu

Amy S. Bruckman  
GVU Center  
Georgia Institute of Technology  
85 5th St. NW Atlanta, GA 30332  
asb@cc.gatech.edu

## ABSTRACT

Once referred to by the Supreme Court as the “metaphysics” of law, many parts of copyright policy are historically confusing. Therefore, it isn’t surprising that in communities where amateur content creators work within a legal gray area, copyright is a frequent topic of conversation. Here, people with often little knowledge of the letter of the law are asking and answering complex legal questions in the context of their creative activities. Working from a content analysis of public forum conversations in eight different online communities, we have examined these questions and answers more closely. By studying these interactions, what can we learn about how people engage with the law and how non-expert advice affects behavior and knowledge?

## Categories and Subject Descriptors

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

## General Terms

Legal Aspects

## Keywords

Copyright; Creativity; Expertise; Intellectual property; Law; Online communities; Policy; Social Q&A; User-generated content

## 1. INTRODUCTION

The phrase “I am not a lawyer” is so common in online discussions that it comes with a commonly used acronym: IANAL. Typically followed by a “but...” this disclaimer goes hand-in-hand with frequently offered amateur legal advice in a variety of contexts all over the Internet. One common topic is that of copyright, which though once mostly relevant to professional artists and big corporations now touches ordinary Internet users on a daily basis. Thanks to the ease of digitization and wide dissemination of content, anyone with an Internet connection and a “share” button is engaging with copyright.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [Permissions@acm.org](mailto:Permissions@acm.org).

GROUP '14, November 09 - 12 2014, Sanibel Island, FL, USA  
Copyright is held by the owner/author(s). Publication rights licensed to ACM.  
ACM 978-1-4503-3043-5/14/11...\$15.00  
<http://dx.doi.org/10.1145/2660398.2663774>

However, more people engaging with this aspect of the law does not make it less complex. On the contrary, technological advances only exacerbate existing confusions in the law. Consider, for example, how easy and common it is now to appropriate existing copyrighted work—from popular YouTube remix videos to viral image memes to *Star Trek* fan fiction. Whereas the average Internet user might know that it is illegal to download a copyrighted song, whether that song can be used as part of a remix is a more difficult question. In a seminal case about remix prior to widespread use of the Internet, the Supreme Court in *Campbell vs. Acuff-Rose* (1992) referred to issues of appropriation as the “most troublesome” part of all of copyright law.

Most online sites where people share creative work include some way for users to interact beyond simply commenting on one another’s work. In these spaces where intellectual property matters come up frequently, much of this discussion focuses on questions and answers. In the same online communities where users are posting digital art, fan fiction, remix videos, and music remix, they are asking each other questions like “How does the law apply to my work?”, “Is this illegal?”, and “Why did I get in trouble?” This is a rich example of an online question-and-answer space in which non-experts are relying on one another’s (often imperfect) knowledge and advice in order to make decisions.

As part of a broader study on copyright knowledge and norms in online creative communities [6,7], we conducted a content analysis of public forum conversations about copyright in eight different online communities. In looking specifically at how users are asking and answering questions, we have begun to see some patterns in social question asking (Q&A) taking place among strangers with a shared creative interest.

## 2. BACKGROUND

Our prior work began with the problem of gray areas in the parts of copyright law relevant to online content creators. Judging from previous research, including studies of documentary filmmakers [10], remix video creators [2], and knitters [9], we expected to see misunderstandings of the law among remixers. However, a more surprising finding was how similar their ideas were. Focusing on fair use,<sup>1</sup> the legal doctrine that allows for some uses of copyrighted content, we found both nuances of understanding and patterns of misconceptions [6]. Many of these patterns seemed tied to social norms, which suggests that they form within the

<sup>1</sup> Fair use in U.S. law (though there are similar concepts for other countries) is a codified exception to copyright law. It covers, for example, parody, news reporting, and creative re-uses such as remix. Determinations are made by a judge on a case-by-case basis.

community itself. This includes not only misconceptions but also social constructions of the law. One hypothesis is that these constructions could be formed by learning within the community, including information seeking behavior.

Though there is little work in the area of Q&A specifically related to intellectual property, Humphries' study of Ravelry, an online community of knitters, examined copyright discussions in the site's online forum [9]. The conclusion was that the community seemed to have very little consensus over what constituted legal or ethical behavior. As the author pointed out, if a simple request for information resulted in a thread with over 80 posts that culminates in a suggestion that the poster consult an attorney, there is likely a problem with both legal literacy and uncertainty. Humphries also suggested that these Q&A sessions result only in frustration rather than encouraging people to learn more.

Of course, it is not unusual for people to seek answers to their questions online, and from a number of different sources. Research in this area has explored formal Q&A sites such as Quora and Yahoo! Answers [1,13] where most of the interaction is between strangers, as well as the use of personal social networks [8,12] (such as Twitter and Facebook) for information-seeking. We know some reasons why someone might choose to ask a question on Facebook, for example, as opposed to seeking out a specific expert, such as personal context and trust [12]. Choi et al. have proposed four models of Q&A sites: community, collaborative, expert-based, and social [4]. One might consider an online affinity space, in which people come together because of a shared interest or common activity (rather than simply Q&A) to be somewhere in between a social and community model.

In these spaces, when someone asks a question and the answer comes from a stranger, they know something about them—that they are likely also an artist, or a writer, or a knitter. They share a common experience, and common problems related to the question—in this case, the struggle to understand the boundaries of copyright law in the context of their creative work. Trust and context are still benefits, but information seekers also have to make decisions about authority and expertise—which we know from credibility research can be difficult to judge [5,13]. Additionally, the legal context makes the questions potentially high risk—could trusting a wrong answer lead to legal trouble? In this way, the environment is similar to health information seeking, where source credibility is an important factor in Q&A [5].

### 3. METHODS

As part of a broader research project [7], we are focusing on creative activities often burdened with unclear copyright issues—i.e., remix and appropriation. Therefore, our data set comes from online communities for which creative appropriation is common. We identified eight popular online communities representing four common remix media types: art, music, writing, and video (see Table 1). Each site features user-generated content and has a public forum with posts primarily in English.

Our data set comes from public (publicly viewable to anyone on the web without account creation) forum posts scraped from these websites in the spring of 2013 [7]. We used an inter-rater reliability measure to validate a keywords used to identify posts about copyright from this set of posts, and then created a random sample with a maximum of 50 posts from each site. In sum, we began with a set of millions of forum posts across these eight different online communities, narrowed down this data set to

posts that were likely about copyright, and then narrowed this down to a tractable number of posts about copyright based on a random sample (see Table 1). The final data set consists of 339 posts. Table 1 also shows an estimate of the percentage of copyright-related posts in each forum generally, based on analysis of an additional random sample. Though these numbers might seem small, for DeviantArt's 15 million posts, for example, three percent still represents hundreds of thousands of conversations.

For data analysis, we looked to grounded theory, which provided us with a systematic process for analysis while maintaining flexibility to fit this unique data [2]. We began with inductive, open coding, and coded the data for emergent phenomena. Two independent coders met to periodically discuss codes and then synthesized them into a final set. One coder was the first author, who is a law school graduate and copyright expert. Once the categories were finalized we coded the data again and used an overlapping set of 10% to calculate inter-rater reliability with a percent agreement of 94% and Cohen's Kappa of .77 [6].

### 4. FINDINGS

In our analysis of the resulting data set, we grouped our codes into eight higher level categories: type of post, legal concepts, policies and enforcement, ethics, attitudes, media, problems, and Q&A. For more detail on methods and additional findings, see [7]. For the current discussion, we are focusing on our initial findings about Q&A based on this analysis.

The general types of copyright-related conversations we identified in this data set were: Q&A, requests for action, policy commentary, complaints, discussion of a specific case, and other (veering off-topic to copyright). In categorizing the posts by type, we found that the most common by far was Q&A. In the two largest sites that we studied, DeviantArt and YouTube, Q&A posts made up more than half of those posts in our data set.

Essentially, the majority of the copyright-related conversation taking places in these communities stems from someone asking a question. We also categorized the most common questions that posters asked:

- Is this okay?
- How can I avoid getting into trouble?
- Why did I get into trouble?
- How can I protect my work?
- I don't understand this rule; can someone explain it to me?

A great many of these questions boil down to requests for amateur legal advice. "Is this okay?" is the most common, with a poster expressing concern about some creative activity and whether it might be illegal or against site policies. What complicates matters is that rarely are there black-and-white answers to these questions, but instead they can be legally complex. For example, one poster on Overclocked, a music remix site, asked whether it is legal to use extracted vocals in a remix of a commercial song. Another asked if it was legal to use chiptune effects on an album intended to generate revenue for a religious entity. The answers to both of these questions depend upon a fairly complex fair use analysis.

Nearly every question in our data set was answered by at least one

	Description	~ Total Posts	~ Posts Scraped	Posts in Data Set	Estimated Copyright Percentage
<b>DeviantArt</b> Art c. 2000 forum.deviantart.com	One of the largest social networks on the web, a popular space for artists both amateur and professional to showcase their work	15,800,000	49,464	50	3%
<b>Fanart Central</b> Art c. 2004 forums.fanart-central.net	An online art gallery that hosts primarily fandom-based art and fiction, but also allows original submissions	278,000	20,875	50	11%
<b>Remix64</b> Music c. 2002 remix64.com/board	A Commodore 64 and Amiga music remix community, containing news and reviews as well as a place for users to upload their work	73,000	1,099	50	5%
<b>OverClocked Remix</b> Music c. 2003 ocremix.org/forums	A video game music community featuring fan-made remixes and information	636,000	7,642	50	5%
<b>YouTube</b> Video c. 2005 productforums.google.com/forum/#!forum/youtube	The largest user-generated content video site on the web; though it does not have a general forum for discussion, it does have a very large help forum in a Google Group	500 per day	17,546	50	13%
<b>MMORPG Forum</b> Video c. 1999 mmorpgforum.com	A site for discussion of massively multiplayer online roleplaying games; the largest sub-forums are dedicated to machinima, Warcraft movies and Warhammer movies	113,000	364	23	3%
<b>HarryPotterFanFiction.com</b> Writing c. 2009 harrypotterfanfiction.com/forums	The oldest fan fiction site on the web dedicated to the <i>Harry Potter</i> novels, housing over 78,000 stories	70,000	1,211	16	2%
<b>Twisting the Hellmouth</b> Writing c. 2008 forum.tthfanfic.org	A fan fiction archive with over 15,000 stories based on the <i>Buffy the Vampire Slayer</i> and <i>Angel</i> television shows	54,000	1,590	50	7%

**Table 1: Website information and statistics**

person—sometimes by several people, with conflicting answers. The question above about chiptunes and religious entities generated four pages of discussion, including one complex (and fairly accurate) fair use analysis, some blatant misinformation, and advice to consult a lawyer. One later commentator pointed out the humor in the conflicting advice, writing that the comments either boiled down to “Don’t do it or you could go to jail” or “Come on, just do it, nobody will notice.” Another poster pointed out that the Internet is a poor substitute for legal advice, and yet another praised the thread for helping them to learn about copyright law.

Though that particular thread benefited from participants who had some understanding of law, many do not. Some answers to questions express basic misunderstandings about copyright, such as telling a poster that they have to register with the copyright office in order to have a copyright in their work, that any image found online is public domain and thus available to use, or that writing “no copyright infringement” as a disclaimer carries legal weight (all of which are incorrect). Misinformation also comes in the form of too strictly construing copyright law. One respondent incorrectly stated that remixing anything without a license in writing constitutes criminal activity. When one YouTube poster questioned whether their video taken down for a copyright violation may have been fair use, one commenter answered that there is no such thing as fair use. In our data set, nearly a quarter of the posts contained some incorrect legal information.

Even when incorrect, more often than not, answers are presented as fact rather than opinion. We coded for degree of confidence,

and absolute confidence was far more common than “I’m not sure, but...” or IANAL disclaimers. However, it was also rare for answers to provide either citations or indications of expertise (occurring in less than 5% of the posts). Occasionally a poster would link to an external resource, such as a Wikipedia page or even to the actual legal code, but this was unusual. Even more unusual—seen only once in our entire data set—did someone express personal expertise. Here, we saw the opposite of IANAL: “I *am* a lawyer, but this is not legal advice” (a necessary disclaimer for attorneys to avoid creating attorney-client relationships strangers on the Internet).

Based on this data alone it is difficult to speculate about the outcomes of questions being answered. However, we can infer that posters take the advice offered at least some of the time. In cases of incorrect information, this could be potentially troubling. Though interestingly, what we see *less* often is advice that might get the original poster into legal trouble (telling them that something is okay when it isn’t). Instead, the more common problem would be advice that might stifle the poster’s creativity (telling them that something *isn’t* okay when it might be). In legal terms, not doing something that you should be able to do because you fear legal consequences is a *chilling effect*. Erroneously telling a creator that their work is illegal could result in that creator choosing not to share their work.

Throughout the entirety of our data set, an overarching theme was that of incomplete information: information seeking, expressions of confusion about the law, misleading or contradictory answers.

However, also among incorrect information is *correct* information, and the occasional nuanced discussion that goes beyond typical copyright knowledge. Within our data, we saw examples of posters going out of their way to answer questions in their communities. These individuals tended to have more legal knowledge than those providing quick, one-off answers. Another side effect of these discussions is that community members sometimes do legal research and learn about things that they wouldn't otherwise. Copyright Q&A sessions therefore are in a position to both contribute to knowledge and spread misinformation. As in other potentially high-stakes contexts such as health information, this position can be precarious.

## 5. DISCUSSION & FUTURE WORK

Social media Q&A research has been exploring important issues of how people seek information online from varying sources. Much of this research has focused on either answering questions from strangers (e.g., Yahoo! Answers) or from their social networks (e.g., Facebook). The domain of copyright conversations among online content creators offers some potentially unique questions and problems in this space. Here we see information-seeking behavior towards people who are neither trusted friends nor strangers nor experts—but rather, people with a shared experience. However, similar to health information seeking behavior, legal advice could have potentially high stakes.

Our work so far begins to describe these communities as a Q&A space, examining the type of questions being asked, and the answers provided. We see that questions are almost always answered with a high degree of confidence and with few citations. This coupled with a high degree of incorrect information is potentially problematic.

Future work includes an examination of how judgments of expertise are actually happening within these communities. How are question askers deciding whom to trust? What are the outcomes of these interactions? Are there times when Q&A is more harmful than helpful if misinformation is disseminated? Are the interactions more similar to those with expert-strangers or non-expert-friends?

Moreover, there is a potential design space here. How can the designers of these communities better support more productive and useful conversations about copyright? All of this conversation is taking place despite none of these communities providing dedicated spaces for legal discussions or Q&A. Encouraging better understandings of copyright could both discourage chilling effects on creativity and help mitigate the air of legal uncertainty in these communities.

## 6. ACKNOWLEDGMENTS

This work is supported by NSF IIS-1216347.

## 7. REFERENCES

1. Adamic, L.A., Zhang, J., Bakshy, E., Ackerman, M.S. and Arbor, A. Knowledge Sharing and Yahoo Answers: Everyone Knows Something. In *Proceedings of the 17<sup>th</sup> International Conference on World Wide Web*, (2008).
2. Aufderheide, P., Jaszi, P., and Brown, E.N. The Good, The Bad, and The Confusing: User-Generated Video Creators on Copyright. *American University Center for Social Media*, 2007.
3. Charmaz, K. *Constructing Grounded Theory: A Practical Guide Through Qualitative Analysis*. SAGE Publications, London, UK, 2006.
4. Choi, E., Kitzie, V. and Shah, C. Developing a Typology of Online Q&A Models and Recommending the Right Model for Each Question Type. *Proceedings of the American Society for Information Science and Technology*, (2012).
5. Eastin, M.S. Credibility Assessments of Online Health Information: The Effects of Source Expertise and Knowledge Content. *Journal of Computer Mediated Communication*, 6(4), (2001).
6. Fiesler, C. and Bruckman, A.S. Remixers' Understandings of Fair Use Online. *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, (2014).
7. Fiesler, C., Feuston, J. and Bruckman, A.S. Understandings of Copyright in Online Creative Communities. *Proceedings of the 18<sup>th</sup> ACM Conference on Computer Supported Cooperative Work & Social Computing*, (2015, in press).
8. Gray, R., Ellison, N.B., Vitak, J. and Lampe, C. Who Wants to Know? Question-Asking and Answering Practices Among Facebook Users. *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, (2013).
9. Humphreys, S. The Challenges of Intellectual Property for Users of Social Networking Sites: A Case Study of Ravelry. *MindTrek '08: Proceedings of the 12th International Conference on Entertainment and Media in the Ubiquitous Era*, (2008).
10. Larsen, L.O. and Nærland, T.U. Documentary in a Culture of Clearance: A Study of Knowledge of and Attitudes Toward Copyright and Fair Use Among Norwegian Documentary Makers. *Popular Communication* 8, 1 (2010), 46–60.
11. Lombard, M., Snyder-Duch, J., and Bracken, C.C. Content Analysis in Mass Communication: Assessment and Reporting of Intercoder Reliability. *Human Communication Research* 28, 4 (2002), 587–604.
12. Morris, M.R., Teevan, J., and Panovich, K. What Do People Ask Their Social Networks, and Why? A Survey Study of Status Message Q&A Behavior. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (2010).
13. Paul, S.A., Hong, L., and Chi, E.H. Who is Authoritative? Understanding Reputation Mechanisms in Quora. In *Proceedings of Collective Intelligence*, (2012).