Remix Culture on the Web: A Survey of Content Reuse on Different User-Generated Content Websites

Oshani Seneviratne
CSAIL, MIT
oshani@csail.mit.edu

Andrés Monroy-Hernández
Media Lab, MIT
andresmh@media.mit.edu

ABSTRACT

In this paper we examine how remixing or content reuse is supported in a wide range of user-generated content websites. In particular, we analyzed the ways in which microblogging, social network, video, photo, music and scientific data sharing websites support some the following forms of content reuse: (a) Licensed content brought in from external sources. (b) Content uploaded to the website and given a license of the user's liking. (c) Licensed content from the website reused elsewhere. (d) Licensed content reused within the website.

We found that the majority of the websites we analyzed[1] had insufficient support for these different types of content reuse scenarios, acting merely as content distribution portals rather than platforms that support a culture of remixing. Some of the websites were found to have support for different licenses but most websites lack adequate functionality for finding licensed content, remixing and attributing sources for remixes.

Keywords

Social Media, Content Reuse, Intellectual Property, Commons, Licensing

1. INTRODUCTION

Remixing is one of the most important ways in which the web is empowering the creation of knowledge and culture nowadays. For example, scholars like Manovich [16], Benkler [8] and Lessig [15] argue that remixing plays an essential role in the creation and development of today's culture due to the affordances that new technologies provide. Manovich points out that remixing is "practically a built-in feature of digital networked media universe", while Benkler positions remixing as a fundamental part of the way culture is produced and argues that if "we are to make this culture our own, render it legible, and make it into a new platform for our needs and conversations today, we must find a way to cut, paste, and remix present culture." Based on these ideas, we look at how well some of the major content sharing websites support a healthy remix culture by (1) empowering people to build on the work of those creators who want their work to be reused; and (2) making it easier to understand what are the rights of authors and the licenses associated with their work.

In analyzing the websites for this survey, we tried to answer the following questions:

1. Can licensed content be brought from an external source?
2. What are the licenses supported by these websites?
3. How easy is it to find licensed content in the website?
4. How easy is it to legally remix content using the features available in the website?
5. How effective are the mechanisms for credit giving? Are they automatic? Or is the user required to give credit manually?

We focus on the social and technical infrastructure that supports remixing practices and the reuse of digital content, such as audio, images, videos and data, created by others in the websites analyzed for the study.

2. ANALYSIS OF REMIX CULTURE

We have chosen to study six major categories of user-generated content websites. Namely: video sharing, photo sharing, audio sharing, micro-blogging, social network, and scientific data sharing websites.

2.1 Video Sharing Websites

YouTube is the largest video sharing website [10], hosting a combination of amateur and professional content since 2005. It has become the quintessential media sharing website, and therefore, it has also become a common target of copyright and licensing disagreements related to content reuse. For example, Lessig, when advocating for changes to copyright laws that stifle amateur creativity, cites the case of a YouTube video uploaded by the mother of a baby showing her son dancing to one of the songs of a musician called Prince, and the subsequent deletion of the video by YouTube due to a copyright complaint by the company who owns the Prince song [15].

Since early 2009, YouTube started to allow special "partners" to select the type of license for their videos [7]. The available licensing options are: personal, public domain and the different Creative Commons (CC) licenses. In order to become a partner, an application has to be submitted to YouTube, which is then evaluated based on the size of the audience of the existing videos among other metrics [1].

[1]The observations were conducted in March 2010

Currently, partners are primarily universities and other non-profit organizations, hence, for a common user just starting on YouTube, it is difficult to explicitly state the kind of license for her videos. Additionally, not even the partner videos with specific licenses have machine-readable information about it and the only way a user can find what the type of license of some video is by clicking on the download button (see Figure 1).

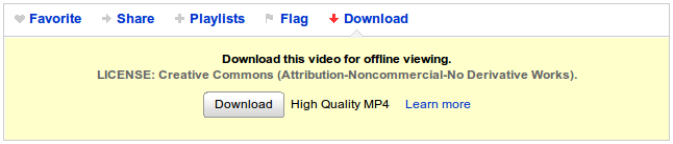

Figure 1: Screenshot of section of a YouTube page displaying the license under a partner video.

In June 2007, YouTube implemented the first version of their content identification system that allows copyright holders to automatically identify videos that use their material and decide to either block, monetize (by sharing some of the ad revenue) or gather metrics about the video [3]. Shortly after this process was put in place, a study [10] found that only 5% of the videos were deleted due to copyright infringement. One of the negative side effects of the automated content identification process is that even those who have purchased the rights to include a piece of someone else's song or video in their own video, can be automatically blocked. Those affected by this can, if they chose to do so, file a counter-notification through YouTube's legal department.

Since downloading is a prerequisite to remixing, most of the remixing of YouTube videos happens offline and the remixes are often uploaded back to YouTube without any automated way of connecting the remix to the original video other than using the text in the video description where the remixer can manually reference the original works used in the video. Additionally, YouTube only displays a download for videos from partners that have allowed to do so, leaving non-partners unable to activate the download link on their videos. A simple web search, however, would list myriad of tools and websites that offer the service of downloading YouTube videos for offline view.

Some of the few mechanisms for video owners to control the use of their content on YouTube is through the features known as "video responses" and "embedding". When a user enables video responses, other people can respond to a video and the links to those responses will show up on the original video page. Embedding can be enabled or disabled as well, which gives video owners the ability to allow external websites (such as blogs or social network sites) to include their videos.

Another similar and popular video sharing website is Vimeo. This site does not let regular users set the license of their videos, and it even goes as far as prohibiting people from uploading videos that are in the public domain [6]. Vimeo iterates that "I have permission" does not mean that the user has created it, and thus the user does not have the right to upload to the website. However, Vimeo allows video creators to easily give credit to other members of the Vimeo community (see Figure 2). It would be very useful if this feature was extended so that it is possible to credit people who are not Vimeo users.

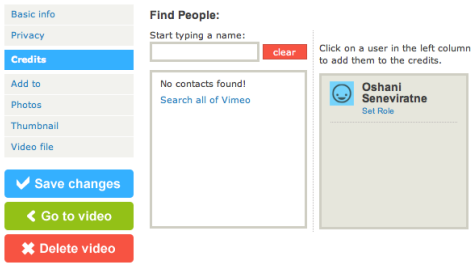

Figure 2: Screenshot of a Vimeo page that allows users to give credit to other Vimeo users.

Jumpcut was one of the few websites that provided not only hosting of videos like YouTube and Vimeo, but also web-based editing tools to mix different pieces of media. The website was closed in 2009. While the website encouraged remixing and provided automated tools to give credit to those whose content were reused, a qualitative study [12] found that people gave explicit attribution and notified the original creators due to a "moral obligation" people felt towards other people, but not towards companies in the case of professional content were being reused.

2.2 Photo sharing sites

Flickr, a very popular photo sharing website, was one of the early adopters of CC licenses. When photos are uploaded to the site, the default restriction given is "all rights reserved". However, users are given the option to choose from one of the six available CC licenses (see Figure 3). Once a user selects one of the CC licenses for all her photos, the statement "some rights reserved" will appear under each photo with a link to a page explaining what those rights are.

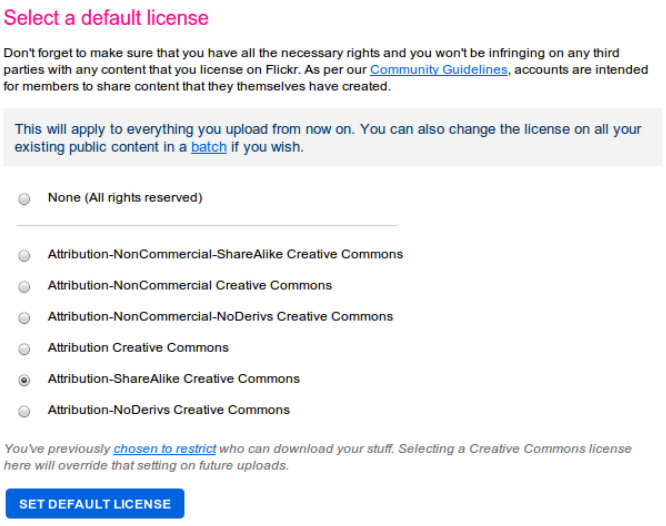

Figure 3: Screenshot of Flickr page where user selects the license for her photos

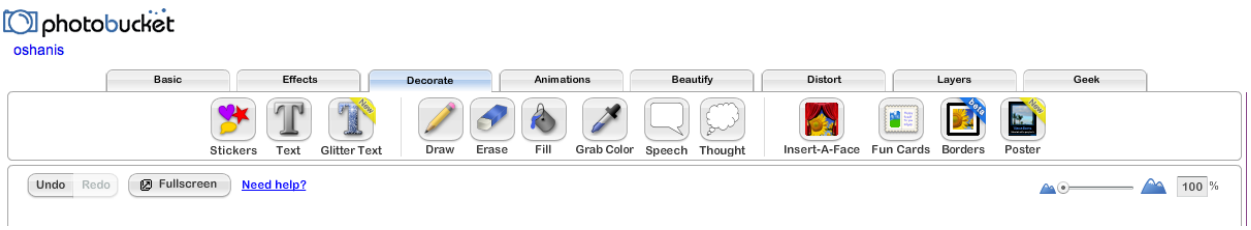

Figure 4: Screenshot of the "Remix" interface of photobucket

The information exposed by Flickr through their embedded metadata seems to assume that all the photos uploaded are owned by the uploader. If the user wishes to let other people reuse her photos, she can display an appropriate CC license that grants the rights to her. However, if she used a CC licensed photo from somebody else in an image that she is uploading to Flickr, there is no in-built support to display the proper attribution to the original owner of the component photo. The same situation applies when a user wishes to reuse a Flickr image from somebody else within the website. Although the Flickr API can be queried for the license for a given photo, Flickr does not facilitate transferring the rights to another person. It becomes the remixer's duty to give the proper attribution when reusing other people's photos, which is confirmed by a study that revealed a very high rate of attribution license violation of Flickr images on the web [18]. This seems to suggest that websites like Flickr could provide the technical affordances to make it easier for people to automatically give the proper attribution when remixing images.

Photobucket.com is another website that allows its users to upload photos and even videos through a variety of methods. The website is primarily used for hosting photos, and has the functionality to reuse images in the website and build scrapbooks, slideshows and even remix images from other users through a very easy to use interface (see Figure 4). The terms of use of the website allows Photobucket and other users to reuse such content under a limited license [5], but does not specify whether it allows CC licenses.

DeviantArt is an online artist community that acts as a creative outlet for over 7 million users. The website has CC licensing built into their UI (see Figure 5). The website also has a mechanism to share works of art within the website, and give automatic attribution to the original source (see Figure 6)

Figure 5: Screenshot of selecting a CC license in the DeviantArt website

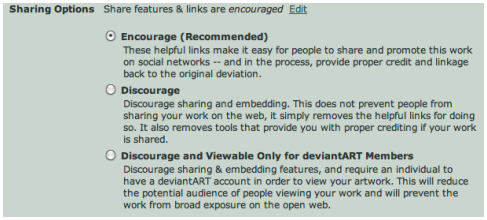

Figure 6: Screenshot of different "sharing" options in the DeviantArt website

2.3 Audio sharing sites

CCMixter is a music website that accepts original music samples and remixes. All the content featured in the site has a CC license. When an original music sample is uploaded, the user can specify any CC license. When a remixed composition is being uploaded, the remixer is presented with a simple interface that helps her identify which samples or any other remixes she used in her composition. Based on the components used in the remix, the remix will "inherit" the most restrictive license from the samples used. Identifying the samples used in the remix allows all individual components of the remix to be linked together, essentially creating an attribution tree [21]. This makes attribution to all those who are involved very easy (see Figure 7). If someone wishes to use content from the ccMixter in an external site, the reuser will have to honor the CC license associated with the music sample [11].

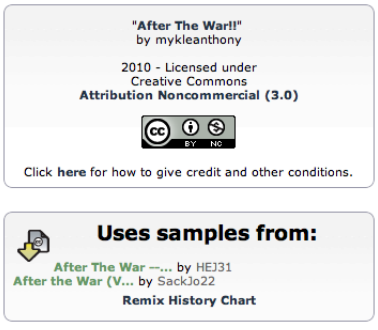

Figure 7: Screenshot from CCMixter showing the CC license for a particular remix and the derivation tree of that remix

IndabaMusic.com is a website that lets people, either in groups or individually, to create songs by putting together different tracks. The website lets people upload audio files for which people have permissions to do so and select from one of three licensing options for those uploads: (1) All rights reserved (this basically tells the community that you own the file and are not granting anyone any special permissions (2) Creative Commons Attribution, and (3) Creative Commons Attribution, Non-Commercial. If a user wishes to specify that a certain music file was used in a remix it can be easily done, and the remixes of the current music file will be shown (see Figure 8). However, it does not have as clearly stated rules or options for songs that are published on the site to be remixed within indaba nor for use outside indaba. This lack of licensing options for finished mixes could be explained in part due to the lack of explicit ways of download finished mixes.

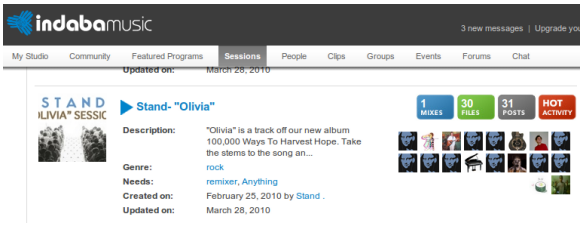

Figure 8: Screenshot from IndabaMusic showing how many remixes of the current music file are there

2.4 Micro-blogging sites

Twitter, a very popular micro-blogging site, does not enforce any form of licensing for status updates. However according to the site's terms of service [20] it advocates users to contribute their creations to the public domain or consider progressive licensing terms. As more and more people use twitter, collective opinion of ordinary citizens becomes very important. For example, if someone wishes to reuse some twitter users' quotes and opinions in a publication, it would be very cumbersome, and very much time consuming for that person to contact each and every person that issued the status updates. Therefore, the need for a proper licensing mechanism becomes a necessity. There are third party applications that address the lack of license support for reuse of the status updates outside of the website. Tweetcc (www.tweetcc.com) is one such example. Tweetcc facilitates people to make their licensing terms more explicit. To use the this service, a twitter user has to send the status update: "@tweetcc: I license tweets under CC Attribution http://creativecommons.org/licenses/by/3.0/". The user can select any other form of CC licensing option, and can change the license at any time by sending another status update that specify the new license. If Alice wishes to reuse a status message from Bob, all she needs to do is to check the twitter username of Bob on the tweetcc service to see if Bob is using any license for his tweets.

When reusing (or rebroadcasting) tweets within twitter, there are certain social norms to let a user's followers know that the tweets are from someone else. This used to be in the form of "RT @Bob" or "retweeting @Bob" or "via @Bob". This was very much similar to email forwarding, where you would retain the original email headers in the body of the message and optionally add a bit of your own commentary. However, as pointed out in [9], this leads to many inconsistencies in attribution. In November 2009, twitter introduced the automatic retweet button (see Figure 9): now if Alice wants to retweet something that Bob had said, the original tweet from Bob will appear in Alice's followers' timeline. This is accompanied by some text indicating that it was re-tweeted by Alice. This particular feature was met with mixed reactions, where some twitter users did not like the fact that they are not able to add their own commentary, and some others complained that it is strange to see someone who you do not follow appear in your timeline. However this new feature does not in fact prevent people from retweeting the old fashioned way. It is merely a measure to reduce the noise in the twitter timeline, and to give prominence to the original tweeter [14].

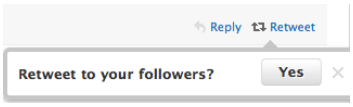

Figure 9: Screenshot of the "retweet" button in Twitter

Twitter terms of service gives Twitter an unrestricted and undefined license for all the content posted on their website while still having the person who posted as the copyright holder [20]. Identi.ca, an alternative microblogging service, is more explicit about the type of license the updates are issued under. Identi.ca has chosen CC-BY (attribution) license by default and it does not let users to select any other type of license. In terms of rebroadcasting status updates, identi.ca supports a mechanism very similar to Twitter. However, the terminology they use is a bit different: instead of "Retweet to your followers?", it uses "Repeat this notice?".

2.5 Social Network sites

Facebook allows users to publish content in their activity streams, but it does not support any licensing of content. However, Facebook users can install the CC application and choose a license for photos, profile text, and status updates, as well as any video content uploaded. This application will display a "badge" in the user's profile (see Figure 10). Each media, such as every photo posted on a user's profile, must fall under one of the six licenses supported by the application. Once installed, this application will show the CC license chosen by the user on her profile. The CC license merely acts as a notice, and it does not enforce anything, nor will provide automatic attribution notices for shared items.

A recent Facebook feature allows users to share content in their activity streams, and specify who it is from automatically. The share button has been tweaked to include "via [friend]" that provides automatic attribution. In addition to that, and unlike in twitter's automatic retweet button, it also allows the reposter to add her own comment about the posting [13] (see Figure 11). However, this feature does not retain the original broadcaster of the status update; users only get to see the immediate broadcaster. Therefore, the original source of the content will get lost in subsequent re-

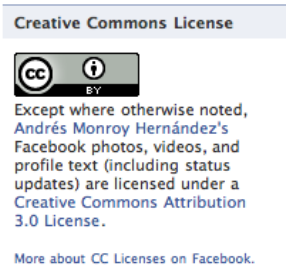

Figure 10: Screenshot of a Creative Commons license badge displayed on a Facebook profile page

postings. A study done on Facebook fan pages indicates that content can be propagated in long chains of up to about 82 levels [19]. Therefore, it would be useful to have a feature associated with Facebook's share button to preserve the attribution trail of any content that is being shared.

Figure 11: Screenshot of a section of a page showing automatic attribution of a status message on Facebook

2.6 Scientific Data Sharing Sites

The amount of research data that is being produced in laboratories around the world is growing rapidly. The large hadron collider will itself will produce 15 petabytes of data per year [2]. Socially the scholarly concept of citation is fairly well understood, but when several thousands of datasets are used in a single query to derive a result, the attribution requirement becomes a daunting task for the researcher.

Science Commons, an organization that crafts policy and legal tools to lower the barriers to scientific knowledge sharing, has identified the need for a sustainable licensing mechanism for reuse of this growing scientific data [4]. As a pilot initiative of the ScienceCommons, the Neurocommons project is aiming to facilitate an open source knowledge management platform for biological data sharing in a manner that fosters license compliance.

In terms of research publications the open access movement has been promoting free availability and unrestricted use of research work. Supporting this theme of open sharing of research results, there are sites such as arxiv.org that support open publication. In addition to that, sites such as openwetware.org allows scientists to record their finding in a wiki and license the text under a CC license of their choice.

Going even further, websites such as myexperiment.org allows scientists to share scientific workflows in a virtual research environment [17]. Workflows are a key aspect in science, but often there are multiple workflows tackling the same problem. By allowing users to upload workflows and share those amongst each other, a culture of open collaboration where scientists can build upon each others' knowledge will be fostered. Myexperiment.org allows users to upload workflows (these can be brand new or can be remixes of several other workflows) and specify to whom credit should be given for the workflows uploaded. This can be to the user, friends of the user, any other user or an entire group. It is also possible to share the workflows and let others update the workflow. This website supports variety of licenses including: CC, MIT, BSD, GPL, LGPL, Apache or Public Domain. Figure 12 shows how a scientist can upload a workflow and specify the sharing options, licenses and rights.

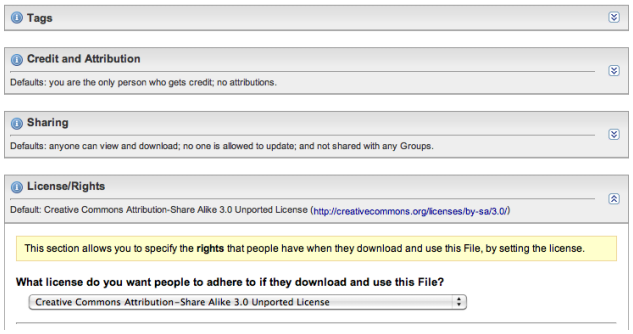

Figure 12: Myexperiment attribution, sharing and licensing options

3. CONCLUSION

We have analyzed a wide range of popular media sharing websites, from text to videos to photos, looking at how well they support the emerging remix culture. We found deficient support mechanisms for entering, presenting and linking licensing and attribution information in both human and machine readable formats. Our findings are summarized in Table 1. In order to address these issues, we propose simple design interventions that include:

1. Letting content creators choose the license for their work and display this license in human and machine readable form.

2. Allowing content uploaders to give credit to the sources of their work by providing hyperlinks and metadata of such sources.

3. Displaying provenance networks that display the tree of derivative work of some content as well as its antecedent work.

4. Giving people the tools to easily embed and remix content in a way that follows the license chosen by its creator.

Content sharing and the associated problems are not just limited to these websites outlined in this paper, there are many more websites and domains that we did not look at. Also, more work is needed to solve the legal complexities of the licensing in the context of remixing different pieces

of media that are licensed under different incompatible licenses. We hope however, that these sample of websites can provide some insight into what system designers have done to support or prevent the development of the remix culture. Future work should include other domains such as blog comment systems or Q&A systems, as well as empirical evidence on the efficacy of each one of the design decisions listed here.

4. REFERENCES

[1] Criteria for partnership : Partner program basics - youtube help.

[2] Large hadron collider: Worldwide lhc computing grid - http://public.web.cern.ch/public/en/lhc/computing-en.html.

[3] Official google blog: The state of our video id tools - http://googleblog.blogspot.com/2007/06/state-of-our-video-id-tools.html.

[4] Science commons - http://sciencecommons.org.

[5] Terms of service of dropbox.com (a photo sharing and remixing website) - http://photobucket.com/terms.

[6] Vimeo - http://vimeo.com/upload/video.

[7] Youtube blog: Youtube goes offline - http://youtube-global.blogspot.com/2009/02/youtube-goes-offline.html.

[8] Y. Benkler. *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. Yale University Press, 2006. Some chapters.

[9] boyd danah, S. Golder, and G. Lotan. Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In *HICSS-43. IEEE: Kauai, HI, January 6*, 2010.

[10] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. Moon. I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system. In *IMC '07: Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, pages 1–14, New York, NY, USA, 2007. ACM.

[11] G. Cheliotis and J. Yew. An analysis of the social structure of remix culture. In *C&T '09: Proceedings of the fourth international conference on Communities and technologies*, pages 165–174, New York, NY, USA, 2009. ACM.

[12] N. Diakopoulos, K. Luther, Y. E. Medynskiy, and I. Essa. *The evolution of authorship in a remix society*. ACM, 2007.

[13] Eric Eldon. Facebook gives resharing a boost with new via attribution feature - http://www.insidefacebook.com/2010/01/18/facebook-gives-resharing-a-boost-with-new-via-attribution-feature.
2010.

[14] Evan Williams. Why retweet works the way it does - http://evhead.com/2009/11/why-retweet-works-way-it-does.html.
2009.

[15] L. Lessig. *Remix: Making Art and Commerce Thrive in the Hybrid Economy*. Penguin Press HC, The, Oct 2008.

[16] L. Manovich. Remixability and remixability. 2005.

[17] D. D. Roure, C. Goble, and R. Stevens. The design and realisation of the virtual research environment for social sharing of workflows. *Future Generation Computer Systems*, 25(5):561 – 567, 2009.

[18] O. Seneviratne, L. Kagal, and T. Berners-Lee. Policy-aware content reuse on the web. In *International Semantic Web Conference*, pages 553–568, 2009.

[19] E. Sun, I. Rosenn, C. A. Marlow, and T. M. Lento. Gesundheit! modeling contagion through facebook news feed. In *Third International AAAI Conference on Weblogs and Social Media*, 2009.

[20] Twitter Team. Twitter Terms of Service - http://twitter.com/terms. 2006-2010.

[21] Victor Stone. ccMixter: A Memoir, How I Learned to Stop Worrying about the RIAA and Love the Unexpected Collaborations of Distributed Creativity During the First Four Years of Running ccMixter - http://fourstones.net/ccmixter-a-memoir. 2009.

Website	Licenses available in the website	Support for remixing	Mechanisms of credit giving
YouTube	Public Domain, Creative Commons (only for special partners)	Users have to download to remix.	The site also automatically recognizes when people use YouTube partners' content and it lets partners choose to either block, monetize or advertise.
Jumpcut (This site no longer exists.)	Creative Commons	Was available in their UI.	An automatic mechanisms was available.
Vimeo	No licenses are allowed.	Not available	Manual (it lets people give credit to other Vimeo users.)
Flickr	Creative Commons	Not available	Not available
PhotoBucket	A 'free' license which allows the website, and it's users to reuse content.	Not available	Not available
DeviantArt	Creative Commons	Available via their premium "Deviations" interface.	Available via their premium "Deviations" interface.
CCMixter	Creative Commons	Available (Although it does not provide an interface to perform the remix, the components of the remix can be attributed to other works.)	Manual (It lets people give credit to other CCMixter users.)
IndabaMusic	Creative Commons (Attribution and Non-commercial use only)	Not available	Manual
Twitter	None, Twitter can do anything with the content. User remains as the copyright holder.	Available (using 'RT' or 'via')	Automatic Retweeet gives automatic credit. Manual retweeters use the RT @originator or "via @originator" convention.
Identi.ca	Creative Commons	Available (same as twitter, slightly different terminology)	Can be rebroadcast to the user's followers (same as twitter, slightly different terminology.)
MyExperiment.org	CC, MIT, BSD, GPL, LGPL, Apache, Public Domain	Available	Available

Table 1: Comparison of several user generated content websites in terms of the licenses available, support for creating remixes and the availability of mechanisms of credit giving.