

eModeration

User generated content moderation

Contact information:

Tamara Littleton
CEO

T: 44 (0)20 7250 4736
M: 44 (0) 7776 138 642
E: tamara@emoderation.com

eModeration Limited
24 Greville Street
London EC1N 8SS
United Kingdom

Six techniques for building safer
User Generated Content (UGC) campaigns

Six techniques for building safer User Generated Content (UGC) campaigns

Executive Summary

In 2006, Time Magazine named "You" as its Person of the Year, bringing even more attention to the new social phenomenon of engaging consumers in the creation and submission of content of all types. From Super Bowl ads to personal profiles, consumers are now participating side by side with companies to create User Generated Content (UGC) sites.

The power of these sites can be significantly diminished by negative or divisive submissions from that small percentage of consumers who can ruin an experience for everyone.

This white paper will introduce six key concepts to help understand how to offset, and hopefully eliminate those negative submissions in order to create a fantastic experience for the community and increase brand protection.

Introduction

"Try to 'let your brand go' online whilst at the same time managing and monitoring the process carefully so that consumers don't undermine it."

- The Internet Advertising Bureau (IAB)

As agencies are called upon by their clients more and more to build marketing campaigns focused around User Generated Content ("UGC") concepts, both the client and the agency find themselves trying to create projects that are both fun and engaging for consumers, while simultaneously ensuring the client's brand isn't put at risk. Even clients who understand the value of an open exchange of ideas and opinions, both positive and negative, between a brand and its consumers have real reason to be worried about how far potential negativity might go.

Balance this against the need of a UGC campaign to allow for the level of freedom consumers require in order to get excited about a project and you find yourself wondering whether it's possible to find the middle ground between freedom and flexibility, and protection and direction.

Nearly every UGC project has built-in tools for moderating content that allow the agency, the client, or a trained team of outsourcers to remove offensive content, and perhaps even drive participation. But are those tools delivering as much as they could?

Moderation and the tools of that process are all too often an afterthought to development. Seemingly, moderation tools are just another hard cost, like hosting or server maintenance. Smaller projects tend to suffer the most from this thinking, where moderation tools and staffing are minimal, if they exist at all. This lack of tools and staff on projects of all sizes can result in a major backfire for the project goals.

With the right design and preparation, moderation tools and staff can deliver even greater success to UGC projects, while finding the right balance between consumer freedom and brand protection.

Technique one: Craft Your Guidelines

We've all become accustomed to seeing "terms and conditions" and other forms of legal documentation on websites. We typically find ourselves facing one of these menacing documents when trying to register for a new website, or upload content to a UGC campaign. Typically thick with complex language, traditional forms are hard to follow and nearly impossible to translate into an understanding of what is acceptable on any given website.

Imagine if guidelines for acceptable behaviour were easy to find and easy to follow! Even crazier, what if people actually enjoyed reading legal requirements?

Turning this vision into reality requires setting aside any ideas or legacy of "legal copywriting" and adopting a crafting mindset. This is similar to the creation of a site's logo or design style. Well-written guidelines can provide a clear picture to users and moderation staff alike about what's acceptable content and what's off-limits.

Look no further than Flickr.com for a great example. The Flickr version of "terms & conditions" is brilliant on several fronts:

- **Choosing the right title** – Flickr's terms are actually called the "Community Guidelines". This is much more understandable to a user who is actively interested in learning more about what's acceptable. Referring to guidelines as "Terms & Conditions" may scare off those who want to do the right thing.
- **Find the right location** – In the footer of every page of the site, Flickr has a link to its Community Guidelines. Not only does a link exist on each page, but the guidelines are listed as the first item in the Help area.
- **Copy write like you enjoy it** – Let's face it, even most lawyers don't like reading legal text. If not by design, then by implementation the very enjoyment of site participation is squeezed out. Nobody says guidelines can't be fun to read. Heck, make them so fun people want to share them. As an example, Flickr's guidelines suggest: "Don't be creepy. You know that guy. Don't be that guy".
- **Involve the community** – As your community grows, users will likely begin to feel a sense of ownership for the community itself. Start to involve them in the modification and extension of the guidelines. This has the dual effect of truly representing the community's boundaries, as well as providing buy-in from community members who would otherwise feel like they're being told what to do.

Remember, the more clearly defined and easily understandable a site's guidelines are, the more likely that users will abide by them.

Technique two: Build Automated Filters

No matter how clear a site's guidelines, there will most likely be the discussion that grows overly heated, the user who enjoys posting questionable content or the spammer trying to drive Google rank by posting links on the site.

The first line of defence against these issues is automating the moderation process through smart filters. Filters can *help* ensure that certain types of content never even appear to users of the site. However, filters are never perfect, and it is important to consider the following when thinking about developing a filtering system:

- **Develop multiple filtering types** – Typically filters are used only for profanity, simply declining submissions from user entries that contain words on a “banned word” list. While certainly important and helpful to moderation staff (who may never even see the rejected content), think about additional filters that might tell users immediately what's not acceptable. Additional filter types could block entries based on URLs, duplicate submissions, or even the prior history of the user.
- **Allow performance tuning** – No matter how smart filters are at launch, it is imperative to be able to update them quickly. Who knows when the client will find content that they didn't think to exclude originally, or when they might learn a few new inappropriate words?
- **Develop both accepted and non-accepted content methods** – One of the biggest complaints from site users about automated filters is that they all too often screen out content that is acceptable because some part of it contains something unacceptable. A profanity filter might flag and reject the word “assumption” because it contains profanity. Profanity filters should allow the site owner to make that distinction.
- **Choose a reaction** – Typically, content that doesn't pass a filter simply is blocked from being submitted. There are, however, other options to consider based on the core concept of the site. For instance, the filter can replace inappropriate content with place holders (“Hey ****, I don't like you”), or the content can be allowed to post to the site but moderators are notified more immediately.

To be completely clear – filters, no matter how smart, are a first line of defence only. They are far from perfect, and based on specific lists of content only. They can't adjust to changes in slang or trends or cultural sensitivities. Too many brands have declined human intervention only to discover large amounts of inappropriate content going live to their site. Humans are a crucial part of any intelligent moderation game plan.

Technique three: Embrace Your Technology

Filters and guidelines are only a starting point when creating a safe brand community site. In order to assure more complete brand protection, a human is going to have to verify most or even all of a site's content. Some sites operate on a pre-moderated method (nothing goes live before specifically being approved), while others operate on a post-moderation method (content goes live immediately, but is ultimately reviewed by a moderator who may accept, reject or edit the content according to client guidelines).

Whichever method of operation is chosen, at some point human moderators are going to have to review the content. Content will slip past the filters, users will ignore the guidelines, or online trolls will drop in just to cause mischief for the fun of it.

The moderation team is going to be busy. Even with a fully staffed moderation team, there's no reason to have them pulling out their hair trying to identify the biggest potential issues. Using some basic mathematics and logic, the moderators can be given tools to help prioritize what content to review first.

Building algorithms that help identify these potential threats is actually easier than it looks. The logic will have to be defined specific to each site, but here are a few examples of how we can put some basic calculations to work.

- **Using the timing** – Keep an eye out for a certain number of user submitted reports within a timeframe. For instance, the system might automatically move posts to the top of the moderation queue that has received more than 10 user reports within an hour timeframe.
- **Using the user history** – Certain users tend to be more helpful than others, and developing an algorithm that calculates the net worth of one user's report versus another user's report is straightforward. Look at how many times a user has submitted a report and had a moderator agree with the report. If there is a large similarity between reports by that user and moderation acceptance, over time that user's reports will jump to the top of the moderation queue.
- **Using the traffic** – Generally, heavily trafficked content is worth keeping an eye on, even if there hasn't been a specific user report. If the content is generating traffic for a positive reason there's reason to celebrate. If the traffic is generated due to less than desirable content, the moderators should check it out right away before it spreads any further.

These algorithms are typically easier to build after the community site has generated even a small amount of use. Consider what algorithms might work best before the site launch, but be open to modification once actual data has begun to be collected and show patterns.

The crucial issue to note is that finding patterns over time will only work if enough data is being collected in the first place. If it's not tracked, it can't be measured.

Technique four: Enlist Your Users

In social groups the majority of the group's members are interested only in a positive experience. Whether looking to a community forum for tech support or submitting a homemade video to a brand campaign, they are simply looking for a satisfying experience.

Given the opportunity, many community members are more than willing to lend a hand and help protect the safety and quality of a project. Enlisting users in the mission can not only help keep the weight off the moderator's shoulders, it can also engage users in the site itself.

- **Different tools for different users** – Any site attracts a variety of users, from the ultra-engaged to the passer-by. Some users may be willing to become volunteer moderators (if moderation tools can handle that); while others simply want to tell someone that they've found something objectionable by clicking a "Report this content" link. Ensure that the site, as well as the concept includes multiple tiers of participation in order to capture any level of interest.
- **Engage the most enthusiastic users directly** – Often a site will see a great deal of activity from the so-called "One Percenters" ⁽¹⁾, the small group of people within a social group that do a lion's share of the content creation. Seek out the One Percenters and contact them directly. Often a personal email from the site owner will further deepen their personal connection to the site. This in turn drives them to want to ensure that "their" site stays as positive and safe as possible. Potential volunteer moderators may be discovered through this process.
- **Volunteering doesn't equate to free labour** – It's too easy, and incredibly dangerous to think of volunteer users as nothing more than free labour. Make no mistake volunteers always desire compensation of some sort, and in a social setting this is rarely financial. Put significant effort into determining what type of compensation might best motivate volunteers. This could be an "ego score" showing the world who has done the most "work" on the site. Or this could be inside information that comes only from volunteering.

1) http://customerevangelists.typepad.com/blog/2006/07/understanding_t.html

Technique five: Make Moderation Actions Visible

Traditional thinking holds that great human moderation is best when hidden from view; a site's users are completely unaware of the magic that ensure a site stays upbeat. As in *The Wizard of Oz*, "Pay no attention to that man behind the curtain!"

For most UGC projects, this thinking isn't exactly correct. Moderator actions and responses can be invaluable both to the community and to the moderation process itself. When moderation controls are completely hidden, an implicit invitation is given to online trolls to try to abuse the system, which in turn creates extra work for the moderators. If the community knows that inappropriate content will be removed quickly because they've seen clear signs of that very thing happening, there will be less reason to test the boundaries. It's simply not worth the time. This not only improves the positive feel of the site, it decreases the workload on the moderation staff. Teddy Roosevelt once said: "Speak softly and carry a big stick; you will go far."

Not every UGC project is created equal, or has the same objectives. There are a few key issues to consider when determining how much visibility to give users about what the moderators are up to.

- **Moderators should be able to communicate** - A human moderator's task isn't simply about approving or declining content. Moderation should help train users on what is acceptable within the community. But like any training, everyone needs a little feedback. Moderators should be able to easily (and quickly) give reasons for whatever actions are taken. This can be simply preset multiple choice or free-form text response (which requires much more time from the moderators but may be appropriate for the project).
- **Level of notification** - Moderation messages should be easily accessible by the user so that they can understand what content has been accepted or declined and why. This notification can happen in a private area only the user can see, or on the publicly viewable site. Each has pros and cons, so let the project concept set the direction.
- **Editing versus deleting** - When content submissions are mostly acceptable but include minor problems (such as a few inappropriate words in an overall acceptable text-based submission), moderators have a choice whether to edit out the offending content, or to delete or reject the entry all together. Generally, deleting/rejecting content should be the default choice so that users don't feel that the moderator is speaking for them. Additionally, deleting/rejecting content will help to train users further that moderators aren't going to do the heavy lifting on their behalf.
- **Offer a chance to change** - Users make honest mistakes, so make sure to give them a way to make good on a moderator's actions. Content that has been deleted/rejected should offer users a chance to easily fix it and resubmit for consideration.

Technique six: Moderation tools need love too

Organisations are spending tens or hundreds of thousands of pounds developing fantastic user generated content sites that deliver incredible experiences to users and excite them into submitting content to the site. They push the limits of the technology, they test the usability with the target audience, then measure success of the site usage over time.

Then there's the development of the moderation tools.

All too often the moderation tools are poorly designed afterthoughts to the project development. The first time anyone "tests" them is the first day the moderators actually start using the tool. Who wants to spend money developing a tool that only a handful of people can use, when that money could be "better" spent developing extra features for public users?

The reality is that bad user interface design on the back end equates directly to additional costs. Usability testing and smart interface design can improve and speed moderator workflows, which directly influences moderation costs.

- **Develop usability test plans** – Moderation tools are web development too. They need the same design and development approach as any other web development, for the same reasons: well-designed tools make for happy users. When users (i.e. your moderators) are going to be spending hours each day using these tools, cutting costs isn't nearly as important as helping them keep their sanity!
- **Allow multiple workflows** – When the moderation staff is more than one person, it's easy to step on each other's toes. Moderator actions shouldn't conflict with each other or cause confusion about who's working on what and when. Test the system with multiple users hitting the tools within seconds of each other to ensure stability.
- **Reporting tools** – Don't miss an opportunity to build-in reporting tools to help measure the success of the project from a moderation point of view. Simple tools that allow the moderators to report back on rejection rates and overall number of submissions give valuable insight and help manage volume.
- **Allow for change** – Even though content can be approved or rejected according to guidelines always build in the ability to change it after the fact. Users may request that content be removed if submitted in error, or guidelines may change.
- **Design the back-end alongside front-end** – Moderation is a fundamental part of a UGC site, so design the moderation tools as a fundamental part of the process. Consider the implications of design decisions on the moderation process from the kick-off to the launch.

Background

About eModeration

Founded in 2002, eModeration Limited is an international, specialist user generated content moderation company. They provide 24 hour community and content moderation to clients in the entertainment and digital publishing industry and major corporate clients hosting online communities and consumer driven projects.

eModeration's CEO and founder, Tamara Littleton, has an established background in editorial quality control, fault escalation and process management gained from previous work as the Product Delivery Director for Chello Broadband and Online Operations Manager for BBC Online where she managed the world's first ISO 9000 accredited team for digital publishing management and monitored over 400 BBC websites. Tamara Littleton is a member of the Home Office Committee advising the British Government on moderation of communities to help safeguard children and she is also the Chair of e-mint, the online community for community professionals.

eModeration's team of moderators and staff are the key to eModeration's success and excellent client list. eModeration draws on the expertise of carefully recruited and trained moderators located in the US and UK with specialist editorial and community moderation skills, which are matched uniquely to the client. All moderators are managed online from eModeration's headquarters in London, United Kingdom.

Tamara Littleton
 CEO, eModeration Ltd
 London, UK
tamara@emoderation.com
www.emoderation.com
 +44 (0)20 7250 4736

About Big in Japan

Big in Japan helps organisations understand, create, and nurture social connection with their consumers, both online and offline. Nothing pleases us more than being able to help organisations start conversations with their audience by building tools and crafting strategies that helps break down the wall between corporation and consumer.

The team of strategists and technologists at Big in Japan design social solutions to business challenges; solutions that aren't limited by technologies like blogs, RSS, and podcasts - rather they're supercharged by them.

If you're interested in training your team on social media concepts or are interested in building both technical and non-technical social strategies, we encourage you to contact us directly or visit our website for more information.

Jake McKee
 Lead Samurai, Big in Japan
 Texas, US
jake@biggu.com
www.biggu.com
 +1 214 550 3603