



Whitepaper: eModeration (user-generated content moderation services)

Brief

eModeration needed a whitepaper that explained to organisations operating Massively Multiplayer Online Games (MMOGs) some techniques that would help keep children safe when playing in these digital worlds.

The company wanted an informative and educational tone, and was keen to stress a sense of perspective and not appear scaremongering.

eModeration

Whitepaper: eModeration (user-generated content moderation services)

Brief

eModeration wanted a whitepaper to explain to companies that operate, or plan to operate, Massively Multiplayer Online Games (MMOGs) techniques that help keep children safe in these digital worlds.

The company wanted an informative and educational tone, and was keen to stress a sense of perspective and not appear scaremongering.

e moderation

Interaction and brand management

Contact information:

Tamara Littleton
CEO

T: 44 (0)20 7250 4736
F: 44 (0)20 7250 4737
M: 44 (0) 7776 138 642
E: tamara@emoderation.com

eModeration Limited
24 Greville Street
London EC1N 8SS
United Kingdom

Virtual World and MMOG Moderation:

Five techniques for creating safer environments for children

May 2008

Virtual World and MMOG Moderation: five techniques for creating safer environments for children

Executive Summary

Millions of children worldwide 'inhabit' virtual worlds and Massively Multiplayer Online Games (MMOGs) – online environments where they play games and interact with friends. Obviously, organisations that operate these worlds, together with parents, must ensure these environments offer children a positive experience, and devise and implement guidelines that help keep children as safe as possible.

It's very important however, to emphasise that millions of children enjoy these worlds every day without coming to any harm; and that the chances of being targeted by an adult with malicious intent are incredibly small. Allowing a child to explore a virtual world is the online equivalent of taking a child to playground: there is the possibility of danger, but it represents a very low risk that can be mitigated by simple safety rules and commonsense.

As in the real world, the fact that there is the possibility of danger in the virtual world means that every step should be taken to reduce risks – but, crucially, without unduly scaremongering both parents and children.

This whitepaper provides advice and tips on how organisations can work with experts in child safety, including children's charities and the Government, and use moderation services to create a safe and positive virtual world or MMOG.

eModeration recommends that when setting up a virtual world or MMOG, an organisation should:

1. Consult all available research when drafting parental guidelines
2. Utilise the expertise and experience of moderators
3. Use automated moderation filters in conjunction with humans
4. Make reporting inappropriate behaviour clear and simple
5. Get parents involved

eModeration is perfectly placed to provide comment on virtual worlds and MMOGs. The company provides user generated content moderation services for major brands and children's virtual worlds and games including Dizzywood (www.dizzywood.com).

Introduction

The popularity and growth of virtual worlds and MMOGs appears inexorable, especially those devised specifically with children in mind. Virtual Worlds Management, which tracks networking sites and virtual worlds, estimates that there are more than 100 youth-focused virtual worlds either live or in development, with 52 of them aimed at children under seven. Disney alone is understood to be developing up to 10 virtual worlds aimed at children.¹ Nickelodeon's online community, Nicktropolis, boasts more than 7 million users²; Mattel's Barbie Girls attracted 3 million users within 60 days of its launch, a figure that has risen to much closer to 10 million³; while 750,000 British children aged between 6 and 14 are estimated to inhabit Disney-owned Club Penguin⁴.

Such impressive statistics have not gone unnoticed by businesses. According to Virtual Worlds Management, in the first quarter of 2008, \$184 million was invested worldwide in virtual worlds, with more than one-third of this investment going to youth-oriented environments⁵.

As with any activity that involves children, safety is always paramount. Just as there are risks in the real world, there are also risks – albeit very remote ones – to children inhabiting online communities. But by implementing a number of moderation techniques and applying common sense, it is possible to mitigate these dangers as much as is possible. Children can play with confidence and enjoy all a virtual world or MMOG has to offer; while at the same time parents are afforded an acceptable level of reassurance that their child is safe.

It's well worth noting that when it comes to children's safety, the same rules apply in the online world as in the real world. There is a fine line between educating children and taking all reasonable steps to protect them, and scaremongering so that a child's otherwise positive experience of virtual worlds is diminished.

The moderation techniques recommended in this whitepaper have been designed, first and foremost, to help organisations that operate virtual worlds and MMOGs ensure the protection of children and minimise potential dangers. At the same time, they are also designed to help organisations create engaging, positive environments that encourage children to play and nurture their development.

¹ http://www.ft.com/cms/s/0/1fedc95e-1fb9-11dd-9216-000077b07658.html?nclick_check=1

² <http://www.nick.com/nicktropolis/game/index.jhtml>

³ <http://www.guardian.co.uk/technology/2008/apr/10/games.news>

⁴ http://technology.timesonline.co.uk/tol/news/tech_and_web/article3831738.ece

⁵ <http://www.redherring.com/Home/24182>

Technique one: consult all available research when drafting parental guidelines

As more and more children immerse themselves in online realities, children's charities and organisations and the Government have compiled advice and guidelines on how to keep children safe in virtual worlds and when interacting in MMOGs.

While the occurrence of a child being targeted by an adult with malicious intent is statistically low, the emotional impact on the child and family is unquantifiable. Therefore any organisation that plans to set up a virtual world or an MMOG must take every precaution necessary to reduce even the remotest risk, and must explain how they do this in their parental guidelines. Failure to do so may prompt the parent to instruct his/her child not to use the site and will not inspire confidence in the management of the site.

Each virtual world will vary in theme and content, but there are a number of 'golden rules' that children and parents should adhere to. The National Society for the Prevention of Cruelty to Children (NSPCC) lists on its website the following invaluable tips for children to stay safe while surfing the web:

- Never use a real name in chat rooms but pick a special online nickname
- Never tell anyone personal things about yourself or your family, such as addresses or telephone numbers or the school or any clubs you go to
- Never send photographs of yourself
- Remember that people online might not be who they say they are
- Never respond to nasty or rude messages
- Stop a conversation if you feel suspicious or uncomfortable about the way it is going, or if it is getting really personal
- Be careful with email attachments or links as they might contain nasty images. Do not open a link or attachment if you don't know who sent it.

While these tips refer to surfing the internet in general, they can also be applied to virtual worlds and can help form the basis of an organisation's 'guidelines for parents' page on its website.

In addition to these recommendations, eModeration advises organisations consider the following when drafting guidelines:

- Parents should tell their children to inform them of anything that makes them feel uncomfortable; and teach them to trust their instincts and report right away any behaviour that makes them feel nervous
- Parents should ensure their children know that they do not have to respond to any messages that make them uncomfortable
- Children should be made aware not to make plans to meet with someone they have spoken with online. Parents should explain to them that online friends may not be who they say they are
- Children should be made aware that the difference between right and wrong is the same on the internet as it is in real life
- Encourage parents to set clear guidelines about when and where their children may use the internet; and encourage children to share with their parents their internet experiences
- The email address used to activate a child's account should belong to a parent, not the child

Below are links to websites that contain information and research on children's online safety, and may prove useful when drafting parental guidelines:

The Home Office: Good practice guidance for the providers of social networking and other user interactive services 2008

<http://police.homeoffice.gov.uk/publications/operational-policing/social-networking-guidance>

The Home Office: Good Practice Guidance for the Moderation of Interactive Services for Children

<http://police.homeoffice.gov.uk/operational-policing/crime-disorder/child-protection-taskforce>

<http://police.homeoffice.gov.uk/publications/operational-policing/moderation-document-final.pdf>

NSPCC: Children and the Internet

http://www.nspcc.org.uk/whatwedo/mediacentre/mediabriefings/policy/children_and_the_internet_media_briefing_wda49338.html

The Children's Charity: Net Smart Rules

<http://www.nch.org.uk/information/index.php?i=135>

Child Exploitation and Online Protection Centre: Think You Know

<http://www.ceop.gov.uk/#section2>

Internet Watch Foundation: Protection Online

<http://www.iwf.org.uk/public/page.36.htm>

Technique two: use automated moderation filters

Ensuring children do not disclose personal details unwittingly, under duress or willingly, is absolutely crucial to prevent the 'jigsaw effect'. This is where an adult with malicious intent garners a child's personal details, often over a long period of time, in order to piece together as detailed a picture of that child as possible.

The more comprehensive an understanding the adult has of the child's background and current circumstances, the easier it becomes for the adult to leverage the information and engineer a phone call, email/IM/text exchange or face-to face meeting.

Automated moderation filters can range from the rudimentary approach of text filtering and can, for example, be used to intercept the disclosure of a child's personal information – such as his/her real name, date of birth, mobile phone number etc. These filters can also be applied to pick up on the '@' symbol, or combinations of numbers, for example, preventing children from giving away their phone number, email or IM address. A text filter can also be set to recognise specific word patterns, and also URLs that might lead to their social network pages, which would otherwise lead a person with malicious intent to a wealth of useful information.

More recently, increasingly sophisticated filters have come into play including Anti-Grooming Engines. These filters are designed to detect very specific grooming tactics. They can flag to a moderator when a child is being persistently pursued for information, such as where they are from, what school they go to, and for his/her personal preferences, such as favourite football team and band or singer. A character in a virtual world or MMOG that relentlessly harasses another character for this kind of information, particularly if it is out of context with the game that is being played, could suggest untoward intentions and warrants further investigation

It is also possible to tackle overt bullying, abuse and harassment using filters. Moderation filters can be continually improved by adding information gathered by moderators to ensure that they adapt to changes in techniques used by instigators of this behaviour. This allows moderators to identify patterns of behaviour and then take appropriate action by blocking a user from participating, temporarily silencing them or in extreme cases banning them from the site.

It should be stressed that while automated tools are very effective, they are not perfect; they cannot adjust to changes in slang or cultural sensitivities, for example.

It's worth noting that children will frequently try to get around filters by creating their own language that they believe will be undetected. For instance, if a site doesn't allow users to refer to each other as boyfriend or girlfriend, they might refer to each other as 'trust friends'. It's also common for users that are aware of existing filters, to describe their locations. Salt Lake City, for example, would be described as the 'opposite of pepper, and area of water'. Another tactic used to evade filters relate to telephone numbers whereby the digits will be spelt out, so 012345 would be replaced by 'oh, fun, doo, free, door, hive'. Filters should be viewed as a first line of defence only and should not be seen as a quick and easy replacement for human moderators, who are vastly experienced and use a combination of training and instinct to detect any suspicious behaviour, which can sometimes be very subtle.

Technique three: utilise the expertise and experience of moderators

In Virtual Worlds and MMOGs there are two types of moderators. The first is the in-game moderator, who actively participates as a character or avatar on the site, helping and encouraging other characters to engage with the various activities within the game. This type of moderator may also act as an in-game host and are visible to all users.

The second type is the silent moderator, who stays in the background blocking offensive material from participants, warning users, defusing confrontation and reacting to abusive or illegal behaviour.

Moderators must set the boundaries for what is acceptable behaviour in a virtual world, and what type of information it is acceptable to divulge. eModeration has found that where clear guidelines have been set and the rules are consistently applied, children will very often not only adhere to the guidelines, but also remind other children of them and collectively help enforce the rules.

As in the real world, children can exhibit bullying behaviour towards other children when online. Intelligent automated filters will often detect such behaviour, but this doesn't remove the need for human moderators, who are trained to keep the peace and ensure a healthy playing environment. Without human moderators, children can find themselves in a 'Lord of the Flies' scenario.

eModeration has found that simply the visibility of a moderator can be enough to prevent instances of bullying or abusive behaviour. It is also important to remember that the role of the moderator is not to interfere in a heavy-handed manner. Their role is not dissimilar to that of a playground monitor – they are there to monitor behaviour and step in when it becomes inappropriate.

There are downsides to being visible within the game as it can detract from the game objectives and may encourage children to want to chat to the moderator instead. It is vitally important for moderators to keep a certain level of detachment from the children and not to be drawn into personal conversations. Obviously, it would be very unprofessional for a moderator to engage in any personal communications with a site user and this should be implicit in any internal guidelines.

A very effective way of maintaining a healthy distance is for the moderator to use a 'host' or 'character' avatar that only speaks in character to discourage attempts to personalise the moderators. In this way, moderators can blend in to the game itself, letting children know they are there without becoming the focus of the user's experience.

Becoming a character or a host also helps enhance the playing experience for children. Moderators can help them overcome in-game challenges and obstacles, and hint at new things within the game for them to experience. For example, on Dizzywood.com, eModeration's moderators appeared on the site as a leprechaun for St Patrick's Day. The leprechaun would interact with the children and reveal hints in rhyme that would help them win rewards, such as gold coins. Here, the moderator is not only monitoring behaviour, but also adding an extra fun and interactive element to the game. Different moderators would take turns to control the leprechaun, but each adhered to set guidelines to ensure the character was consistent and equal treatment was applied to each child.

Essentially, the role of the moderator is not to tell players off publicly, but to defuse situations before they become problematic; children need to feel that moderators are there to keep them safe and not to control them. For example, if a moderator witnesses a player who is repeatedly swearing, the moderator can 'mute' that player before explaining to him or her why that action has been taken. Moderation messages should be delivered in a clear, accessible way so that the child can understand why what s/he has said or done is unacceptable. It is always worth remembering that children do make genuine mistakes, so it is very important to give them the chance to make good on a moderator's actions.

Technique four: make reporting inappropriate behaviour clear and simple

Children invariably enjoy their time in virtual worlds, but if for any reason they feel uncomfortable the site must provide a very easy way for children and parents to report such instances. For example, if a child or parent encounters bullying or abusive behaviour, they should be encouraged to alert a moderator via an easily-accessible report button, flag or email address.

The site should also have a very clear policy on what the community rules are so that children understand what is, and what isn't, acceptable behaviour before they play. When drafting the policy, it is important to ensure that the language is very easy to understand and the tone is not too intimidating for children. Remember, if they understand the rules and why they are there, they are far more likely to adhere to them and encourage other players to do the same. Also, encouraging parents to discuss the policy with their children is likely to give children a greater understanding of the rules and why they must be followed.

Parents should also emphasise to their children the need for them to trust their instinct; if they have any worries at all, they should err on the side of caution and report any uncomfortable situations to a moderator.

Technique five: get parents involved

As discussed throughout this whitepaper, parents have a very important role to play in ensuring virtual worlds and MMOGs are a safe and positive place for their children to play.

Make sure there is a 'guidelines for parents' page clearly visible on the site and, using the tips from this whitepaper and from the organisations listed earlier, draft guidelines appropriate for the site. Again, these organisations and moderation companies will be very willing to assist in this area.

It's a worthwhile exercise for parents to sit with their children and play in the virtual world together, at least to begin with. In doing so, it helps children learn how to play safely, while parents can familiarise themselves with the site and are afforded a greater understanding of how their children spend their time online.

It is very important that parents are encouraged to adopt a balanced approach when it comes to educating their children on the dangers of virtual worlds, and do not unwittingly frighten them before they've even played the game. A good analogy is making them aware of the 'stranger at the school gate' – children should be aware of the danger of talking to strangers, but not be scared each time they finish school.

The golden rule for children to remember is that when online, never share personal information; this way you are minimising the risk of a child being contacted outside the game or offline. From information provided by the site, parents should make clear to children what is, and what is not, information which can be shared.

About eModeration

Founded in 2002, eModeration Limited is an international, specialist user-generated content moderation company. It provides 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 Internet Taskforce for Child Protection on the Internet which brings together government, law enforcement, children's agencies and the internet industry, who are all working to ensure that children can use the internet in safety. She was also the Chair of e-mint, the online community for community professionals from 2006-2007.

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 mainly in the US and Europe with specialist editorial and community moderation skills, which are matched uniquely to the client. The company can moderate 24/7/365 in more than 18 languages. All its 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