Appropriate Filtering for Education settings

June 2016

Provider Checklist Reponses

Schools in England (and Wales) are required "*to ensure children are safe from terrorist and extremist material when accessing the internet in school, including by establishing appropriate levels of filtering*"[1]. Furthermore, the Department for Education published the revised statutory guidance 'Keeping Children Safe in Education'[2] in May 2016 (and active from 5th September 2016) for schools and colleges in England. Amongst the revisions, schools are obligated to "*ensure appropriate filters and appropriate monitoring systems are in place. Children should not be able to access harmful or inappropriate material from the school or colleges IT system*" however, schools will need to "*be careful that "over blocking" does not lead to unreasonable restrictions as to what children can be taught with regards to online teaching and safeguarding.*"

By completing all fields and returning to UK Safer Internet Centre (enquiries@saferinternet.org.uk), the aim of this document is to help filtering providers to illustrate to education settings (including early years, schools and FE) how their particular technology system(s) meets the national defined 'appropriate filtering standards. Fully completed forms will be hosted on the UK Safer Internet Centre website alongside the definitions

It is important to recognise that no filtering systems can be 100% effective and need to be supported with good teaching and learning practice and effective supervision.

Company / Organisation	Netsweeper Inc.
Address	57 London Road, High Wycombe, Buckinghamshire, HP11 1BS
Contact details	Ben.smith@netsweeper.com
Filtering System	Netsweeper
Date of assessment	15.06.2016

System Rating response

Where a supplier is able to confirm that their service fully meets the issue identified in a specific checklist the appropriate self-certification colour for that question is GREEN.	
Where a supplier is not able to confirm that their service fully meets the issue identified in a specific checklist question the appropriate self-certification colour for that question is AMBER.	

.

---

[1] Revised Prevent Duty Guidance: for England and Wales, 2015, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/445977/3799_Revised_Prevent_Duty_Guidance__England_Wales_V2-Interactive.pdf
[2] https://www.gov.uk/government/publications/keeping-children-safe-in-education--2

Illegal Online Content

Filtering providers should ensure that access to illegal content is blocked, specifically that the filtering providers:

Aspect	Rating	Explanation
● Are IWF members		Netsweeper is a long time member supporting the IWF, with council representation.
● and block access to illegal Child Abuse Images (by actively implementing the IWF CAIC list)		The Netsweeper product integrates with the IWF CAIC illegal content list. The IWF functionality is not exposed in the webadmin graphical user interface and cannot be disabled. Importantly, Netsweeper regularly submits URLs discovered by the global systems back to the IWF team, who review the candidate material to decide on inclusion in future updates of the IWF listings.
● Integrate the 'the police assessed list of unlawful terrorist content, produced on behalf of the Home Office'		With access to the CTIRU Netsweeper uses the UK Home office's terrorism blocklist to block terrorist content per Government guidelines. Netsweeper integrates the list into their worldwide 500 million user cloud delivery categorising new content and offering unmatched global protection against terrorist and extremist content.

Inappropriate Online Content

Recognising that no filter can guarantee to be 100% effective, providers should both confirm, and describe how, their system manages the following content

Content	Explanatory notes – Content that:	Rating	Explanation
Discrimination	Promotes the unjust or prejudicial treatment of people on the grounds of race, religion, age, or sex.		Netsweeper has a category called 'Hate Speech'. These sites portray views that are written, verbal, or illustrated, and are intentionally offensive to the general public. The intent of these sites is to degrade, intimidate, or incite violent or prejudicial actions against individuals based on race, ethnic affiliation, nationality, gender,

			sexual orientation, religion, disability, or profession. Advocacy or instructional sites that promote the harming of individuals or groups and encourage or promote peer abuse, videos of physical assaults, written harassment and threats are also included.
Drugs / Substance abuse	displays or promotes the illegal use of drugs or substances		Netsweeper has a specific category named 'substance abuse' which is blocked. These sites provide information about or promote the use of prohibited, illegal, controlled, or regulated substances for recreational rather than medicinal use. It can include sites that sell, encourage or advocate the use of any substance that produces hallucinations, as well as the cultivation, manufacture, and distribution of any intoxicant and related paraphernalia. Informational sites that are clearly intended to provide descriptions of drugs and substances, their negative effects, and addiction potential are not included.
Extremism	promotes terrorism and terrorist ideologies, violence or intolerance		Categories within Netsweeper that block this content include 'Extreme', 'Hate Speech', 'Criminal Skills' and 'Weapons'. Definitions can be found below: **Extreme** This includes sites that are considered far from normal and are categorized for their degree of intensity. The content features or promotes intentional, direct, and deliberate violence and destruction or the alteration of the human body and other living creatures. These sites may depict or promote torture, self-inflicted harm, mutilation, or other dangerous activities. Images and information that advocate and glorify eating disorders, suicide, death, gore, injuries or sites that

			feature grotesque or frightening descriptions are also included. **Hate Speech** These sites portray views that are written, verbal, or illustrated, and are intentionally offensive to the general public. The intent of these sites is to degrade, intimidate, or incite violent or prejudicial actions against individuals based on race, ethnic affiliation, nationality, gender, sexual orientation, religion, disability, or profession. Advocacy or instructional sites that promote the harming of individuals or groups and encourage or promote peer abuse, videos of physical assaults, written harassment and threats are also included. **Weapons** This includes sites that provide information related to the promotion, support, sale, or discussion of weapons and any related device used in combat that can injure or kill, such as guns, knives, or swords. Information on how to build weapons or bombs will also be included in 'Criminal Skills'. **Criminal Skills** This includes sites with instructions or methods that promote, encourage, or provide skills considered to be illegal, criminal, violent or harmful to the general public, and are forbidden by law. This can include questionable material and sites that promote nonviolent, unethical, or dishonest behaviour such as academic cheating, or software hacking/key breaking. This does not necessarily reflect the laws of any particular region or country.
Malware / Hacking	promotes the compromising of systems including anonymous browsing and other filter bypass		Netsweeper has categories named 'Malware', 'infected hosts', 'phishing', 'viruses' and

|			
	tools as well as sites hosting malicious content		'adware'. These categories block websites sites that are associated with this. These are sites containing scripts, or code, that may be ran in a hostile or intrusive manner to a system.
Pornography	displays sexual acts or explicit images		Netsweeper has a 'pornography' category which contains URLs that reference, discuss, or display pornographic images, videos, or other sexually oriented material that is created for the purpose of arousing sexual interest. Soft and hard-core pornography, sadomasochism, bestiality, fetishes, erotic stories, adult magazines, sex toys, or any other sexual related products are included.
Piracy and copyright theft	includes illegal provision of copyrighted material		Two distinct Netsweeper categories satisfy this requirement; i.e., "Criminal Skills" and Peer2Peer. Criminal Skills includes sites with instructions or methods that promote, encourage, or provide skills considered to be illegal, criminal, violent or harmful to the general public, and are forbidden by law. This can include questionable material and sites that promote nonviolent, unethical, or dishonest behaviour such as academic cheating, copyright infringement or software hacking/key breaking. This category does not necessarily reflect the laws of any particular region or country. Peer2Peer (Torrents included): includes sites that distribute software and facilitate the direct exchange of files between users. Software that enables file searching, sharing and transferring across a network independent of a central server as well as web based sites of this nature are included.

Self-Harm	promotes or displays deliberate self-harm (including suicide and eating disorders)		The Netsweeper "extreme" category blocks sites categorised as self-harm sites, anorexia, bulimia and other content that prove harmful to children.
Violence	Displays or promotes the use of physical force intended to hurt or kill		Violence sites are included in the extreme category which includes sites that are considered far from normal and are categorized for their degree of intensity. The content features or promotes intentional, direct, and deliberate violence and destruction or the alteration of the human body and other living creatures. These sites may depict or promote torture, self-inflicted harm, mutilation, or other dangerous activities. Images and information that advocate and glorify eating disorders, suicide, death, gore, injuries or sites that feature grotesque or frightening descriptions are also included.

This list should not be considered an exhaustive list. Please outline how the system manages this content and many other aspects

Netsweeper has provided filtering solutions to the UK Education market for over 15 years and is trusted to protect the networks of over 30% of schools in the United Kingdom. Offering a global collective community experience, Netsweeper resides in over 63 countries, is localised in 30 plus languages, has categorised over 10 billion URLS and is used to filter over 500 million devices worldwide. The web as we know it is consistently changing and by navigating to http://www.netsweeper.com/live-stats/ one can see in real time the new content Netsweeper is categorising each and every day.

The value we bring to our customers is the Netsweeper collective platform where our customers experience the peer-to-peer benefits of our premium Cloud based categorization capabilities; through an intuitive easy to use interface enabling educational administrators to effectively deal with illicit web content on their networks.

Netsweeper offers category based alerting meaning you can customise and create alerts to be sent to safeguarding members of staff, head teachers, IT personnel. This can be triggered automatically, so for example you can create a 'Prevent' Report as illustrated below.
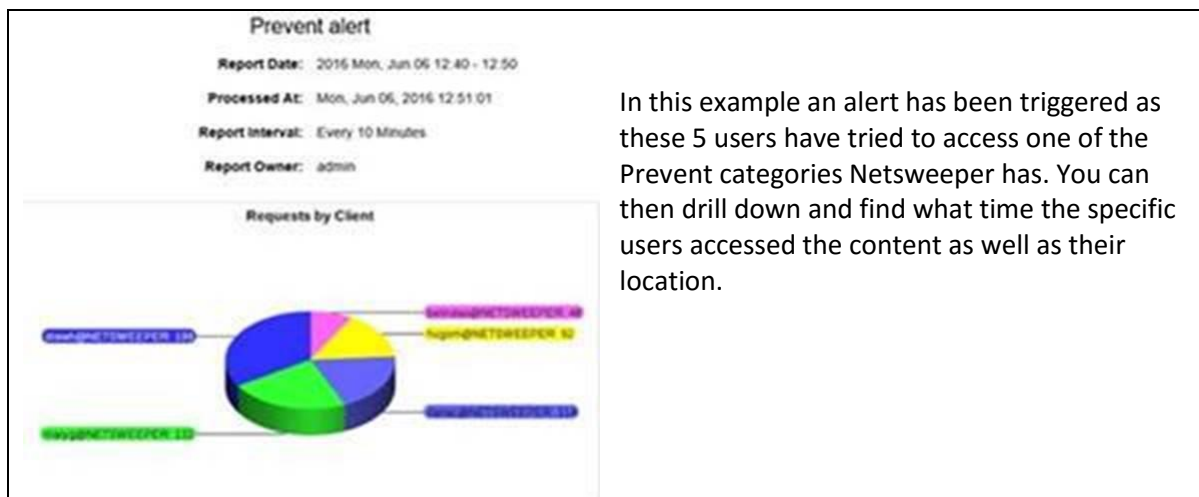
In this example an alert has been triggered as these 5 users have tried to access one of the Prevent categories Netsweeper has. You can then drill down and find what time the specific users accessed the content as well as their location.

Providers should be clear how their system does not over block access so it does not lead to unreasonable restrictions

> Netsweeper defines policies based on the categorisation of URLs. Policies will generally deny selected categories. Policies also have override lists, if a URL would be denied by a category, the lists can be used to amend that decision to be allowed. For example: If the policy denies the 'Social Networks' category, but the administrator wishes to allow Facebook, a simple entry in the local list to allow facebook.com is all that is needed.
> Users with administrative rights are able to modify their policy (categories and lists) according to their permissions.

Filtering System Features

How does the filtering system meet the following principles:

Principle	Rating	Explanation
● Age appropriate, differentiated filtering – includes the ability to vary filtering strength appropriate to age and role		Netsweeper is integrated with an existing directory system such as Microsoft AD, Novell LDAP, Apple LDAP, OpenLDAP or Radius Accounting to assign users based on their group or attribute to the correct filtering policy.
● Control - has the ability and ease of use that allows schools to control the filter themselves to permit or deny access to specific content		Netsweeper natively supports multiple tenancy and delegated administration with fine-grained permissions control. This control has two effects, it will simplify the web-admin graphical user interface removing elements of the interface that the user does not have permission for, and constraining the access to site data and policies. Nominated individuals will have delegated

		administration for clusters of sites, and/or individual sites. The ability to manage policies and lists, and report on the associated data will be provided.
● Filtering Policy – the filtering provider publishes a rationale that details their approach to filtering with classification and categorisation as well as over blocking		The Netsweeper collective community numbers over 500 million devices worldwide. This collective together with our technology and human oversight defines URL classification.

Netsweeper publishes classification of filtering and categorises on the Netsweeper website as well as a view in real time of new content categorised. This can be found at either http://www.netsweeper.com

or

http://www.netsweeper.com/live-stats
Netsweeper's core competency is using our patented techniques to categorise every URL that passes through our deployed systems. Netsweeper is both real-time and employ a hierarchy of data (URL-to-category), with our Category Naming Service (CNS) as a global-master database.

If any customer anywhere in the world accesses a URL, that URL is submitted to the local policy server, if that policy server cannot find a category match, it is automatically submitted to the CNS and looked up there. If the CNS already has the category mapping it is immediately returned to the local system and cached there for future use, a policy decision is then made by the policy server. |

		If neither the local system, nor the CNS has a category match, the URL is submitted to our "Artificial Intelligence" system that will interrogate the content at-and-around that URL, assess the content, detect if it references or contains malware, and assigns one or more categories to the URL into the CNS and then back to the local system, a policy decision is then made by the policy server.

The CNS allows us to adapt to trending URLs immediately due to its world-wide scope. If the local system hasn't seen a particular URL yet, then CNS probably has.

If the URL has been assigned one or more categories, local systems see immediate responses (sub-second).

If the URL is truly "new" then the AI will typically process the content within 20 seconds.

The local policy servers can be configured with techniques to minimise the "new URL" wait. |
| ● Identification - the filtering system should have the ability to identify users | | Netsweeper sits in the core of the network and configured to proxy all traffic. Netsweeper is integrated with an existing directory system such as Microsoft AD, Novell LDAP, Apple LDAP, OpenLDAP or Radius Accounting to assign users based on their group or attribute to the correct filtering policy. For guest/wireless networks Netsweeper can utilise Radius Accounting packets which are generated by the Wireless Access Controller to identify when a user has authenticated. No additional software or agents are required. |

● Mobile and App content – isn't limited to filtering web traffic and includes the blocking of inappropriate content via mobile and app technologies		Netsweeper has a Packet Inspection System that is able to detect a variety of protocols using its unique outbound inspection system. Policies are defined based on 'category' which means that application protocols are 'categories' that can be turned off and on a per IP, per subnet, per group or global basis. Netsweeper can also be deployed as an SSL Decryption Proxy for applications that utilize HTTPS for its mobile applications further adding functionality to limit and control applications, such as limiting functionality on social network sites.
● Multiple language support – the ability for the system to manage relevant languages		Importantly, Netsweeper can categorise pages not only in English, but also in multiple languages. Netsweeper currently supports the following languages for URL classification: Arabic, English, French, German, Japanese, Persian, Polish, Russian, Simplified Chinese, Spanish, Turkish and Vietnamese. Future releases will include: Somali, Bangla, Croatian, Estonian, Swedish, Irish, Norwegian, Thai, Bulgarian and Traditional Chinese. Netsweeper's web site has a live stats page which gives an overview of some of the categories and languages. (www.netsweeper.com/live-stats/). As the popularity of a particular language increases, Netsweeper adds support for that language.
● Network level - filtering should be applied at 'network level' ie, not reliant on any software on user devices		Confirmed. Netsweeper can be configured and deployed as an explicit proxy, transparent proxy, out of band (port mirroring), inline or DNS URL Filtering system. Netsweeper Policy servers are deployed at the Internet gateways and communicate with a cloud

		based Category Name Service (CNS) in order to provide near real-time updates for URL classifications. Policy decisions are enforced on a per user, per group or global basis with time of day filtering and allowed or denied list overwriting the cloud based classifications.
● Reporting mechanism – the ability to report inappropriate content for access or blocking		Reports can trigger emails, thus a scheduled report can be considered an alert if the report contains data (if the report contains no data take no action).
● Reports – the system offers clear historical information on the websites visited by your users		Netsweeper provides a very flexible reporting tool. Out of the box a number of pre-defined quick reports are available, these can be adapted or removed as desired. Quick reports are typically graphical and provide visibility of for example "top 10 web sites" (there are many different quick reports) The reporting tool has the concept of demand reports and scheduled reports. Demand reports are typically one-off reports for a specific demand. For example "Can you tell me what web sites were accessed today between 10am and noon?" Scheduled reports run on a defined schedule. Scheduled reports are useful for informative info-graphics. For example, top 10 web sites visited this week, top 10 web sites denied this week. Reports can be graphical (pie-charts, bar-charts), or detailed (tabular text), or combined. Graphical reports can be multi-layered, allowing for interactive reports where further detail can be discovered by drilling down.

		There are various export options (image, PDF, CSV, etc…)

Filtering systems are only ever a tool in helping to safeguard children when online and schools have an obligation to *"consider how children may be taught about safeguarding, including online, through teaching and learning opportunities, as part of providing a broad and balanced curriculum".[3]*

Please note below opportunities to support schools (and other settings) in this regard

Netsweeper can also be used for tracking usage of subscription based learning tools. Using the Netsweeper reporting system schools can identify if they any students are using specific educational resources and the frequency.

Netsweeper can also provide custom messaging to students when they attempt to access specific content. For example: Netsweeper can provide an information page on hate speech, drug abuse, viruses, etc rather than just simply denying them. This allows schools to message specific topics and provide students with more information on how and where to obtain help.

[3] https://www.gov.uk/government/publications/keeping-children-safe-in-education--2

In order that schools can be confident regarding the accuracy of the self-certification statements, the supplier confirms:

- that their self-certification responses have been fully and accurately completed by a person or persons who are competent in the relevant fields
- that they will update their self-certification responses promptly when changes to the service or its terms and conditions would result in their existing compliance statement no longer being accurate or complete
- that they will provide any additional information or clarification sought as part of the self-certification process
- that if at any time, the UK Safer Internet Centre is of the view that any element or elements of a provider's self-certification responses require independent verification, they will agree to that independent verification, supply all necessary clarification requested, meet the associated verification costs, or withdraw their self-certification submission.

Name	Benjamin Smith
Position	Business Development Manager, Netsweeper (UK)
Date	15.06.2016
Signature	Benjamin Smith