

Classification and Internet technical filtering

David Vaile, Executive Director

Cyberspace Law and Policy Centre
UNSW Law Faculty

March 2009

<http://cyberlawcentre.org/censorship/>

Outline

- Content categories
- Jurisdiction
- Technical censorship models
- Mandatory ACMA blacklist
- Opt-Out 'CleanFeed'
- Classification levels

Categories relevant to Content

- **Jurisdiction:** in or outside Australia?
- **Classified:** Actually, potentially, not
- **Classification level:** RC, X, R, MA, M, PG, G
- **Delivery medium:** mobile device, Internet, broadcast, DVD, book
- **Internet protocol:** HTTP/HTTPS, mail, encryption, P2P, Tor, other

Jurisdiction: the ancient challenge

- How to control global net from 1 country?
- Inside Australia: millions of items
 - Subject to direction, take-down
 - Prosecution (is ACMA list actually used?)
 - Appeal/review of sorts
 - Actual Classification, sometimes
- Outside Australia: billions of items
 - No power to direct
 - Can report crimes to Interpol, listkeepers etc.
 - No appeal or review? Deemed classification

Is it actually Classified/Prohibited?

- **Actual:** by Classification Board, known persons, expensive (\$500-700/page or site?), perceived expert/detached, potentially subject to review
- **Potential:** by ACMA, cheaper, opaque, somewhat less expert and detached, not prone to review
'Potential prohibited content' is content that has not been classified by the Classification Board, but if it were to be classified, there is a 'substantial likelihood' that it would be prohibited content.
- **Not Classified by human:** algorithmic or heuristic filters using probabilities and risk metrics: cheap, random, error prone
- **Classification too expensive for almost anything online;** ACMA deemed classification cheaper, but still too expensive for classification of the net.

Terminology

- **Conflation of terms for deprecated content:**
 - Legal/ 'Illegal'
 - 'Prohibited'
 - 'Unwanted', 'offensive'
- **Different policy issues, regimes, purposes**
- **'Clean Feed' cf. mandatory ACMA filter**

“two tiers - mandatory of illegal material and an option for families to get a clean feed service if they wish.’ Sen. Conroy 20.10.08
- **'Clean'/'filter': safe, or tampered/censored?**

2007 policy statement

'Labor's Plan for Cyber-safety'

Mandatory ISP Filtering

- A Rudd Labor Government will require ISPs to offer a 'clean feed' internet service to all homes, schools and public internet points accessible by children, such as public libraries.
- Labor's ISP policy will prevent Australian children from accessing any content that has been identified as prohibited by ACMA, including sites such as those containing child pornography and X-rated material.
- Labor will also ensure that the ACMA black list is more comprehensive.... by liaising with international agencies such as Interpol, Europol, the Federal Bureau of Investigation (FBI) and the Child Exploitation and Online Protection (CEOP) Centre and ISPs to ensure that adequate online protection is provided to Australian children and families.

Technical censorship models

1. 'Mandatory' ISP ACMA blacklist, based on Classification levels, legislation-driven
2. Opt-Out, pseudo-voluntary ISP 'CleanFeed': 'Dynamic'? Whose list? Whose criteria?
3. Opt-in, voluntary: PC filters, ISP option?
 - How does Mandatory compare with Opt-Out? Is it self-evidently of more concern?
 - Interested in: degree to which a model is or can be susceptible to 'rule of law', review, external monitoring, transparent governance, public interest policy analysis, precedent, reporting

Mandatory ISP ACMA blacklist

- Inclusion criteria: Classification levels
- Legislation-specified: 'Prohibited content'
- Listing: complaint-driven
+ notification by certain law enforcement sources
- 1,000 to 10,000 items
- Scope for false security? Material which would be Prohibited is not actively searched, most omitted
- Statutory secrecy, FOI exemption, inc. grounds, reasons, content, location, etc. as well as URL
- Potential for more scrutiny, transparency, checks?
Or intrinsically beyond open governance models?

Online Content Scheme

Schedule 5 and 7, *Broadcasting Services Act 1992 (Cth)*

- ‘... content accessed through the Internet, mobile phones and convergent devices, and delivered through emerging content services such as subscription-based Internet portals, chat rooms, live audio-visual streaming, and link services.’
- Complaint-based listing of ‘potentially prohibited’ material.
- No active search?
- Recent moves to have criminal material added by international agencies
- Controversy over listing of anti-abortion site by complainant intending to demonstrate the process.

'Prohibited' / 'Potentially Prohibited'

(inclusion criteria for ACMA mandatory blacklist)

- Includes content *not legal* to access or possess: some that is or would be classified RC, esp. child porn.
- Main affect is offshore sites ('potential'); local sites shut down?
- Short lifespan of offshore illegal RC sites?
- Includes content *legal* to access or possess: is or would be classified MA 15+, R18+, X18+ and some RC.
- Blocks adults' access to 'legal' content considered unsuitable for children (is or would be classified MA15+, R18+, or X18+ in Australia)

'Prohibited': RC, X 18+

RC *Refused Classification*

- Criminal to possess/ access some ('illegal'), not others
- Child porn
- Fetishes, extreme violence/abuse, ...
- Not on TV, cinema, DVD etc.
- Major part of actual list.

X 18+

- Not criminal?
- Non-violent, sexually explicit
- For physical sale in ACT, NT, mail order (DVD etc.)
- Not on TV?

Some policy debate implies RC is what the filter is all about ...

'Prohibited': R 18+, MA 15+

R 18+

- If not behind RAS age verification (strong)
- Not criminal
- Rarely on TV.
In cinema, on DVD etc.
- Offshore sites became eligible to be 'Potentially Prohibited' Jan 2008
Communications Legislation Amendment (Content Services) Act 2007 (Cth)?

MA 15+

- If provided by commercial service (other than a news or current affairs service) & not consist of text and/or still visual images, provided by means of a mobile premium service, or is behind age verification (weak, self assertion)
- Not criminal
- Late TV? Film, DVD etc.
- 'Prohibited' Jan 2008

Prohibition creep?

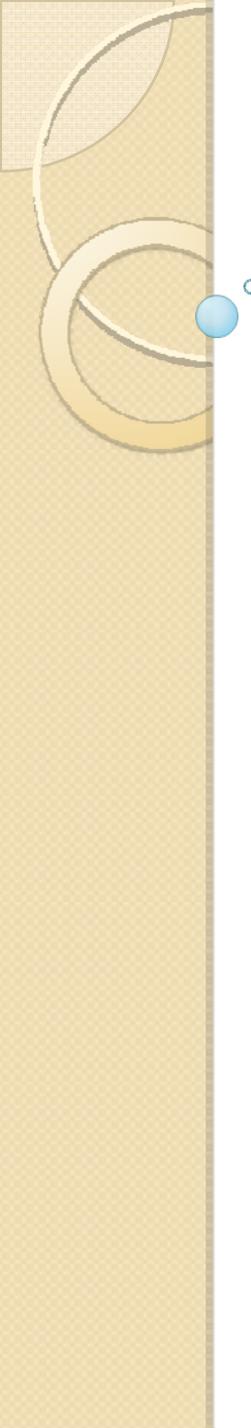
- R 18+ and some MA 15+ added to 'potentially prohibited' category in 2008 - Prohibition creep?
- Make net 'Safe for children (according to us)'
US history of campaign in this direction?
- Ruled unconstitutional in US, but no local 1st Amndt
- Public scrutiny of changes? Fed/state co-op process?
- *Classification (Publications, Films and Computer Games) Amendment (Terrorist Material) Act 2007*: Cth unilaterally requires RC for material that directly or *indirectly* 'advocates' a 'terrorist act', contra states.
- Is an ultimate goal known? Or is it a moveable feast?
- What issues arise if it is a moveable feast?

Opt-Out 'Clean Feed'

- 'Dynamic'? What rules for matching?
- Other static blacklists? Source?
- Whose criteria/block categories? Any relationship to Classification scheme?
- Evidence base: do criteria serve a known purpose?
- Is some content mandatory? Or private discretion?
- Scope for abuse by insertion of sites? Checks?
- Opt-Out, default In: Self-nominate as seeking blocked material. Convenience? Pseudo-voluntary?
- Governance model? Parliamentary review?
- Tens of millions: Does sheer size preclude review?

Generic issues for both systems

- Scope creep?
- Review of actual impact?
- Who sets criteria?
- Who does classification?
- Evidence of justification or effectiveness?
- Does process become opaque if implemented technically?
- Over-blocking: loss of access to proper material, temptation to break security
- Under-blocking: not give expected 'protection'?
- How would these be picked up?
- How would these be remedied?



David Vaile, Executive Director

**Cyberspace Law and Policy Centre
UNSW Law Faculty**

<http://cyberlawcentre.org/censorship/>