Summary of Reasons – Designated Internet Services Code

31 May 2023

eSafety decision

The eSafety Commissioner (**eSafety**) has decided not to register the *Designated Internet Services Online Safety Code (Class 1A and Class 1B Material)* (**DIS Code**). The DIS Code does not meet the statutory requirements set out in section 140 of the *Online Safety Act 2021* (Cth) (the **Act**) because it fails to provide appropriate community safeguards in relation to matters which are of substantial relevance to the community.

Accordingly, the eSafety Commissioner will proceed to prepare an industry standard to cover providers of Designated Internet Services (**DIS Providers**). In accordance with the requirements of section 148 of the Act, eSafety will publicly consult on a draft industry standard.

Background

The Act permits eSafety to register an industry code that has been developed and submitted by a body or association that represents a particular section of the online industry. To register an industry code, eSafety must be satisfied that it meets the requirements under section 140 of the Act, including that it provides appropriate community safeguards for any matters of substantial relevance to the community.

On 11 April 2022, eSafety gave a notice to the Australian Mobile Telecommunications Association, BSA | The Software Alliance, Communications Alliance, the Consumer Electronics Suppliers' Association, Digital Industry Group Inc and the Interactive Games & Entertainment Association (the **Applicants**) under section 141 of the Act requesting that they develop an industry code dealing with certain matters (the **Notice**).

On 18 November 2022, the Applicants submitted a draft of the DIS Code to eSafety pursuant to the Notice. In February 2023, eSafety gave a statement of preliminary views on that draft to the Applicants and invited the Applicants to submit a final version addressing feedback in eSafety's statement.

On 31 March 2023, the Applicants submitted the DIS Code to eSafety for registration, with a covering document entitled 'Request for Registration of Online Safety Codes' (the **Request**).

Scope of the DIS Code

The DIS Code applies to providers of Designated Internet Service (**DIS**), a very broad category of online services that are defined in the Act as services which allow end users to access material using an internet carriage service or which deliver material to persons having equipment appropriate for receiving that material, where the delivery is by means of an internet carriage service.[1]

DIS include most apps and websites, but exclude those provided by social media services, relevant electronic services (online services that enable users to communicate with other users) or other identified services.

Importantly, the DIS Code expressly applies to end-user managed hosting services (such as online file/photo storage, and other online media hosting services) as well as general purpose, enterprise, classified or high impact DIS.

The DIS Code contains measures proposed by the industry associations to address, minimise and prevent harms associated with access and exposure to the most harmful forms of online material. Material intended to be covered by the DIS Code includes:

- **class 1A material**, which is comprised of child sexual exploitation material, pro-terror material, and extreme crime and violence material, and
- **class 1B material**, which is comprised of crime and violence material and drug-related material,

in each case as described in Annexure A to the DIS Code Head Terms, which reflects the *Classification (Publications, Films and Computer Games) Act 1995* (Cth) (**Classification Act**) and related instruments.[2]

These types of material are subcategories of class 1 material under the Act which is material that has been or would be refused classification under the Classification Act. Serious harms are associated with these kinds of material whenever it is produced, distributed or consumed.

A future industry code or industry standard will be developed to address class 2 material under the Act, which includes material that has been or would be classified X 18+, R 18+, Category 1 Restricted or Category 2 Restricted under the Classification Act.

[1] Section 14 of the Act

[2] Importantly, the nature of the material, including its literary, artistic or educational merit, and whether it serves a medical, legal, social or scientific purpose, is relevant to the assessment of class 1B material – see section 11 of the Classification Act. Material only falls within class 1B if there is no justification for the material.

Structure of the DIS Code

The compliance measures proposed in the DIS Code apply on a tiered basis, based on each DIS provider's self-assessment of the risk profile of its service:

- **Tier 1**: a service with a higher risk to end-users that class 1A and 1B material will be accessed, distributed or stored on the service
- **Tier 2**: represents a moderate risk of this occurring
- **Tier 3**: represents the lowest risk of this occurring.

Certain categories of DIS Providers are not required to conduct a risk assessment because they are deemed to have a particular risk profile. They include:

- **High impact DIS**: deemed to be Tier 1. These are websites which have the purpose of providing, or allowing end-users to post material which is 'high impact' (i.e. content which is restricted under the Classification framework such as pornography or 'gore/shock' material).
- **Classified DIS**: deemed to be Tier 3. These are websites or apps providing:
    - general entertainment (such as film or computer game) that has been or would be classified R18+ or lower, or
    - news or educational content that would not be classified Category 1 or Category 2 Restricted under the Classification Guidelines for Publications.
- **General purpose DIS**: deemed to be Tier 3. These are websites or apps that primarily provide information for commerce, charitable, professional, health, scientific, academic research, government, public service, emergency, or counselling and support service purposes) as well as web browsers.

Tier 1 services are proposed to have the most obligations under the code and, in particular, are the only service required to take proactive steps to detect known (pre-identified) child sexual abuse material.

Tier 3 DIS are not subject to any compliance measures under the code unless they implement a significant new feature that would take them to a higher risk category.

Two specific types of DIS are also not required to conduct a risk assessment because the Applicants consider services within each of these categories have similar risk profiles:

- **Enterprise DIS** (services provided within an organisation for its own internal use)
- **End-user managed hosting service** (online file/photo storage, and other online media hosting services).

Enterprise DIS and end-user managed hosting service are proposed to be subject to specific compliance measures in the DIS Code.

eSafety assessment of the DIS Code

Certain categories of DIS Providers can perform a critical role in reducing the accessibility and dissemination of class 1A and class 1B materials and eSafety agrees that a tiered approach to the obligations is appropriate.

The DIS Code proposed a range of minimum compliance measures that the Applicants submit provide appropriate community safeguards in relation to the matters identified in the Request.

eSafety agrees that the matters identified in the Applicants' Request (which are materially the same as those matters identified by eSafety in the Notice) are matters of substantial relevance to the community. However, eSafety considers that the DIS Code does **not** provide appropriate community safeguards in relation to the following matters:

1. Matter 1: Measures directed towards the objective of ensuring that industry participants have scalable and effective policies, procedures, systems and technologies in place to take reasonable and proactive steps to detect and prevent:
    a. access or exposure to
    b. distribution of, and
    c. online storage of,

    class 1A material.

2. Matter 2: Measures directed towards achieving the objective of ensuring that industry participants have scalable and effective policies, procedures, systems and technologies in place to take reasonable and proactive steps to prevent or limit:
    a. access or exposure to, and
    b. distribution of,

    class 1B material.

3. Matter 4: Measures directed towards achieving the objective of ensuring that industry participants have scalable and effective policies, procedures, systems and technologies in place to take reasonable and proactive steps to limit the hosting of class 1A material and class 1B material in Australia.

The DIS Code does not provide appropriate community safeguards in relation to Matter 1 because of the following:

1. the omission of a requirement on **end-user managed hosting services** to:
    a. deploy systems, processes and/or technologies to detect and remove known (pre-identified) child sexual abuse material and known (pre-identified) pro-terror material
    b. take action and invest in disruption and deterrence of class 1A material (including new/first generation child sexual abuse material)
2. there is no requirement for certain **end-user managed hosting services** (those which consider themselves to be not capable of reviewing and assessing materials on their services) to enforce their own policies or terms of use relating to class 1A and 1B material.

The proposed code also does not provide appropriate community safeguards in relation to Matter

4, due the concerns identified above, or in relation to Matter 2, because of the second reason above.

Omission of a requirement on end-user managed hosting services to detect and remove known child sexual abuse material and known pro-terror material

eSafety supports the commitments by Tier 1 DIS to proactively detect and remove known child sexual abuse material, and to take action to disrupt and deter known pro-terror material and first generation material. However, the DIS Code does not extend these obligations to end-user managed hosting services.

There is substantial evidence that online file/photo storage sites are used to facilitate dissemination of child sexual abuse material and pro-terror material.[3] Research shows that image hosting services and online storage sites have been the website types most frequently abused by offenders distributing child sexual abuse imagery.[4] While the services are described as hosting services, end-users can, and do, share log-in details, using the service to disseminate and distribute child sexual abuse material as well as other material.

eSafety's key concerns with the DIS Code include the absence of a requirement on end-user managed hosting services to deploy systems, processes and/or technology to proactively detect and remove known child sexual abuse material and known pro-terror material.

eSafety raised the absence of such a commitment with the Applicants during the code development process and suggested, among other things, that the DIS Code includes a minimum requirement for end-user managed hosting services to detect and remove known child sexual abuse material and pro-terror material.[5] In response, the Applicants submitted they are unable to agree with eSafety's suggested approach on these measures 'due to privacy concerns and key differences in service offerings'.[6]

eSafety recognises the importance of privacy and end-users' privacy expectations regarding their online file/photo storage. However, there are privacy preserving tools capable of detecting known child sexual abuse material and pro-terror material that are widely available and frequently used. These tools often rely on hash matching and do not review the actual content. There are also multiple options of systems, processes or technologies appropriate for less well-resourced businesses and for a range of service offerings.

Known child sexual abuse material and pro-terror material is material that has been previously identified and verified as such material. Such verification is typically carried out by well-recognised non-government organisations, working on a global scale that are legally able to view

[3] OECD 2022, Transparency reporting on terrorist and violent extremist content online 2022; National Center for Missing and Exploited Children (NCMEC) 2023, 2022 CyberTipline reports by electronic service providers (ESP); WeProtect Global Alliance 2021, 2021 Global Threat Assessment; Australian Institute for Criminology
[4] See IWF Annual Report 2022 under trends and data analysis by site types. IWF refers to image hosting services as sites where users can upload images and make them available via a unique URL, and cyberlockers to include online file hosting services, cloud storage services or online file storage providers.
[5] See 'Invitation to respond and/or submit amended draft code – Designated Internet Services'
[6] Page 100 of the Request.

and verify the material. Material that has been identified and verified by such organisations is typically then 'hashed' (ascribed a unique digital fingerprint). Online services are then able to use hash matching tools to find and prevent the re-sharing of copies of the same image or video.

Many online services, including some key end-user managed hosting services already use such tools.

eSafety considers that the absence of a requirement on end-user managed hosting services to deploy such tools significantly limits the safeguards the DIS Code provides to the community in relation to Matters 1 and 4.

Omission of a requirement on end-user managed hosting services to take action and invest in disruption and deterrence of child sexual abuse material and pro-terror material

eSafety supports the inclusion of a requirement on Tier 1 DIS to *invest* in the disruption or deterrence of child sexual abuse material and pro-terror material. This commitment covers investment in systems, processes and/or technologies that identify material with the broader category of child sexual abuse material including, importantly, new/first generation material, that has not yet been verified as child sexual abuse material and hashed. However, the DIS Code does not extend these obligations to end-user managed hosting services.

eSafety considers a requirement to invest in technology, systems or processes to capture the broader category of child sexual abuse material and pro-terror material important. eSafety recognises that automated tools to detect new child sexual abuse material and pro-terror material created by offenders have not in many cases yet been effectively tested and deployed at scale. Commitments by DIS, as well as Social Media Services (**SMS**) and Relevant Electronic Services (**RES**), to invest in the development of such technology, systems or processes is important to increase the ability of service providers to pick up new child sexual abuse material as quickly as possible and to tackle emerging risks including deepfake content[7] and user-generated content. The effectiveness and useability of tools capable of detecting first generation material is improving significantly over time.

The omission of any requirement on end-user managed hosting service to invest in such technology, systems and processes is particularly problematic given the omission of the requirement (discussed above) to deploy steps to proactively detect child sexual abuse material and pro-terror material which has already been verified as such.

Concerns with the effectiveness of measures requiring end-user managed hosting services to enforce policies relating to class 1A and 1B material

The DIS Code requires end-user managed hosting services to have systems and processes to deal with breaches of policies. However, it does not require end-user managed hosting services to enforce their policies when they consider themselves not capable of reviewing and assessing class 1A and 1B material reported by end-users.

[7] See 'Deepfake trends and challenges - position statement'

eSafety recognises that technical capability is clearly relevant in determining which steps can be taken to implement these measures and that some providers may be unable to definitively ascertain whether class 1 material was being stored in their services.

However, eSafety is concerned that under the code, services will determine current capability based on their existing configuration and will not take reasonable steps to move towards compliance with design changes that are technically feasible and practical. Further, while there may be some technical limitations associated with particular features, other steps foreshadowed in a service provider's policy, could be taken and there should be a commitment to take such steps and follow policies.

Such steps could include:

- making appropriate enquiries into any expected breach of their policies,
- issuing warnings/notifications, or
- otherwise taking steps to deter end-users from storing and making available class 1 material (and in particular, known child sexual abuse material or pro-terror content).

Requirements to have policies or processes in place are not effective without a requirement to apply a policy or implement that process.

Next steps

eSafety will develop an industry standard covering DIS Providers that does provide appropriate community safeguards for end-users in Australia with respect to class 1A and class 1B materials.

eSafety will commence development of the industry standard for DIS Providers shortly. In accordance with the requirements of section 148 of the Act, eSafety will publicly consult on the development of the DIS industry standard.