Summary of Reasons – Relevant Electronic Services Code

31 May 2023

eSafety decision

The eSafety Commissioner (**eSafety**) has decided not to register the *Relevant Electronic Services Online Safety Code (Class 1A and Class 1B Material)* (**RES Code**). The RES 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, eSafety will proceed to prepare an industry standard to cover providers of relevant electronic services (**RES 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, Digital Industry Group Inc, and the Interactive Games and 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 RES 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 RES Code to eSafety for registration, with a covering document entitled 'Request for Registration of Online Safety Codes' (the **Request**).

Scope of the RES Code

Service categories

Relevant electronic services (**RES**) is a broad category of online services that enable users to communicate with other users. The RES Code splits these services out into 10 subcategories with different requirements that aim to reflect the risk profile and capability of each subcategory.

1. **Closed communication RES**:

   These are services that enable a user to communicate with another user, but only if they already have that other user's contact details (e.g. phone number or email). This is a broad subcategory that includes email services, some online messaging services, some video conferencing services as well as carriage services (i.e. services offered by mobile phone operators that enable text messaging).

2. **Dating services**:

   These are services predominantly used for dating with a messaging function. This subcategory does not include escort or sex work services.

3. **Encrypted RES**:

   These are services that are either entirely or partially end-to-end encrypted. A partially end-to-end encrypted service may encrypt communications between users but not other parts of the service, such as profile photos and group names.

4. **Enterprise RES**:

   These are services being provided to an organisation to enable that organisation's users to communicate with each other.

5. **Gaming service with communications functionality**:

   These are services that enable users to play online games with each other and share material with each other (e.g. URLs, hyperlinks, images and/or videos)

6. **Gaming service with limited communications functionality**:

   These are services that enable users to play online games with each other but only allow limited sharing of material (e.g. in-game images, pre-selected messages)

7. **Open communication RES**:

   These are services that enable users to view, navigate or search for other users without already having their contact details. This subcategory mainly includes online messaging services and video conferencing services.

If a RES does not meet the criteria for any of the above subcategories, the service would need to undertake a risk assessment and would be classified as:

1. **Tier 1 RES**: highest risk
2. **Tier 2 RES**: medium risk, or
3. **Tier 3 RES**: lowest risk.

Covered material

The RES Code contains measures proposed by the Applicants 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 RES 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 RES Code Head Terms, which reflects the *Classification (Publications, Films and Computer Games) Act 1995* (Cth) (**Classification Act**) and related instruments.[1]

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.

eSafety assessment of the RES Code

There is a significant amount of evidence that RES are used to disseminate class 1A and 1B material, in particular child sexual exploitation material and pro-terror material.[2] RES Providers can play a critical role in reducing the likelihood that their services are used to distribute the most harmful online content.

[1] 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.

[2] See e.g. 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 2021, Live streaming of child sexual abuse: An analysis of offender chat logs; Australian Federal Police 2021, AFP warn about fast growing online child abuse trend; International Justice Mission 2020, Online Sexual Exploitation of Children in the Philippines: Analysis and Recommendations for Governments, Industry, and Civil Society; Ofcom 2022, The Buffalo Attack: Implications for Online Safety.

The RES Code sets out a range of minimum compliance measures for RES Providers that the Applicants submit provide appropriate community safeguards in relation to the matters identified in the Request.

eSafety agrees that the matters identified by the Applicants in the 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 RES 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.

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

1. there is no requirement on closed communication and encrypted RES Providers with capability to deploy systems, processes or technologies to detect and remove known (i.e. pre-identified) child sexual abuse material and known pro-terror material to take such steps
2. requirements on certain RES Providers to take action and invest in disruption and deterrence of child sexual abuse material and pro-terror material fail to address the omission identified above, due to enforceability concerns
3. there is no requirement on closed communication RES Providers (such as email providers) to have trust and safety personnel, and
4. there is no requirement on certain RES Providers (those which consider themselves to be not capable of reviewing and assessing materials on their services) to enforce their own policies relating to class 1A and 1B material.

The RES Code also does not provide appropriate community safeguards in relation to Matter 2 because of the third and fourth reasons above.

<u>Lack of requirement on certain closed communication and encrypted RES Providers to detect and remove known child sexual abuse material and known pro-terror material</u>

eSafety supports the requirements on some RES to proactively detect and remove known child

sexual abuse material (dating services, gaming services with communications functionality, open communication RES and Tier 1 RES) and known pro-terror material (open communication RES and Tier 1 RES). The RES Code provides that these requirements apply to the extent that a service is capable of reviewing, assessing and removing material.

Known child sexual abuse material and known pro-terror material is material that has been previously identified and verified as child sexual abuse material or pro-terror material. Such verification is typically carried out by well-recognised non-government organisations that are legally able to view 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.

eSafety's key concerns with the RES Code include the absence of a requirement on both certain closed communication services (e.g. email services) and some encrypted services to detect and remove known child sexual abuse material and known pro-terror material.

While eSafety recognises the importance of private communication and the expectations of end-users that their communication is private, there are privacy-preserving tools capable of detecting known child sexual abuse material and known pro-terror material that are widely available and also frequently used, including by many RES Providers. These tools often rely on hash matching and operate without reviewing the specific content of messages.

eSafety also recognises that, in other instances, there are technical and other barriers which prevent providers of encrypted RES and providers of certain closed communication RES (such as carriage service providers) from deploying particular tools to detect known child sexual abuse material and known pro-terror material.

While such providers would be unable to comply with a requirement to use specific tools, eSafety considers that these exceptions do not mean the RES Code cannot contain a general requirement of this kind on closed communication RES and encrypted RES. The RES Code already separately provides that, where a service is not capable of deploying technology or processes to detect and remove known child sexual abuse material and known pro-terror material, any requirement to deploy such technologies, systems or processes does not apply.

Importantly, there are some key closed communication RES (such as email) and also some services falling within encrypted RES which are only partially end-to-end encrypted, that would be capable of deploying privacy-preserving tools. eSafety considers that the absence of a requirement on RES Providers to deploy such tools significantly limits the safeguards the RES Code provides in relation to Matter 1.

Enforceability concerns with requirements 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 RES Providers to invest in the disruption or deterrence of child sexual abuse material and pro-terror material. This requirement covers

investment in technologies, systems or processes that identify the broader category of child sexual abuse material and pro-terror material, including new (i.e. first-generation) material.

eSafety considers that this requirement is important given the limited nature of the requirements on service providers (across multiple industry codes) to use technologies, systems or processes to proactively detect *known* class 1A material. Although this limitation is appropriate, given that tools to proactively identify new content have not yet been effectively tested and deployed at scale, the effectiveness and useability of tools capable of detecting first generation material are improving significantly. eSafety considers it important that RES Providers, like providers of social media services (**SMS**) and designated internet services (**DIS**), are required to invest in the development or deployment of such tools, which could have a significant impact if used at scale.

eSafety also recognises that, for some RES Providers including most encrypted RES, there are technical barriers in adopting technology to detect known child sexual abuse material and known pro-terror material, and that in some cases 'disrupting and deterring' such material may be a more effective requirement than a requirement to deploy such technology.

Nonetheless, the RES Code's requirement to 'take action' that aims to disrupt and deter child sexual abuse material and pro-terror material fails to provide appropriate community safeguards in relation to Matter 1 (when read in conjunction with the other relevant requirements). This is because this requirement applies equally to RES of differing capability and functionality. RES Providers with the capability to take more effective steps, including certain closed communication RES (such as email services) and some encrypted RES which are only partially end-to-end encrypted, are not required to take such steps under the RES Code. This omission results in a very low bar for compliance for many RES Providers.

Lack of requirement on closed communication RES to have trust and safety personnel

The Code requires nearly all categories of RES Providers to be resourced with adequate staff to oversee safety on the service and evaluate and adopt safety features and settings to minimise the risk of class 1A and class 1B material. However, this requirement does not apply to closed communication RES Providers.

eSafety is not concerned with the exclusion of carriage services from this requirement, but the exclusion of other closed communication RES including email services from this requirement is a significant limitation. The absence of a requirement to employ personnel to oversee safety and evaluate what safety tools a service could use is critical and the absence of such a requirement is expected to impact the ability of these providers to provide a safe service to end-users in Australia.[3]

[3] 'Australian end-user' is used throughout the industry codes but is defined in clause 2 of the Head Terms as an end-user in Australia to align with the language and scope of the Act. Both terms are used in this document.

Lack of requirement on certain RES Providers (those which consider themselves not capable of reviewing and assessing materials on their services) to take steps to apply their own policies relating to class 1 material

The RES Code requires most RES Providers to have systems and processes to deal with breaches of policies. However, it does not require those RES Providers that consider themselves not capable of reviewing and assessing material to take steps to apply these processes.

While eSafety recognises that the providers of such services may be unable to definitively ascertain whether class 1A and class 1B material is being stored, communicated or made available on their service, other positive steps foreshadowed in their policies could still be taken that can have a meaningful impact. There should be a requirement to take such steps and follow such policies.

Such steps could, depending on the RES Provider, include:

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

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

The concerns identified above do not represent all issues considered by eSafety in its assessment of the RES Code, however, they are the most critical concerns.

Next steps

eSafety will develop an industry standard applying to RES 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 RES Providers shortly. In accordance with the requirements of section 148 of the Act, eSafety will publicly consult on the development of the RES industry standard.