Protecting people from illegal harms online

Volume 3:
How should services assess the risk of online harm?

Consultation

Published: 9 November 2023

Closing date for responses: 23 February 2024

Contents

Section

7. Introduction... 3

8. Governance and accountability 5

9. Service Risk Assessment Guidance 39

10. Record keeping and review 87

7. Introduction

7.1 In this volume, we explain our proposals about what governance services should put around managing risk, what services should do to assess the risk of illegal harm, and how they can meet their record keeping and reporting duties.

7.2 This volume is structured as follows:

- **Chapter 8, 'Governance and accountability'**: sets out our proposed recommendations regarding how services should approach governance and accountability.[1]
- **Chapter 9, 'Service Risk Assessment Guidance':** covers our proposed recommendations regarding our overall approach to Services' Risk Assessment Guidance and Risk Profiles; and
- **Chapter 10, 'Record Keeping and Review Guidance':** sets out and explains our proposed approach on how services should make and keep written records of their risk assessments and measures taken to comply with their safety duties, as well as regularly reviewing their compliance with all their duties.

7.3 We have produced two draft guidance documents to help in-scope services comply with their obligations in respect of risk assessment and record keeping and review under the Act. These are:

- **Annex 5, 'Service Risk Assessment Guidance'**: covers Ofcom's guidance for services when conducting their own risk assessment for illegal harms. All services need to ensure their assessments are "suitable and sufficient", and they must take appropriate steps to keep them up to date. Our guidance lays out the steps that services can take to ensure that they are meeting these requirements, including a proposed universal four-step process that has been informed by industry best practice in risk assessments. To further assist services with their risk assessment, we have produced a set of 'Risk Profiles' which provide a short, accessible summary of the factors we consider are associated with a heightened risk of illegal harms. These Risk Profiles draw on the analysis set out in our Register of Risks and serve to summarise the main findings of the Register. They are intended to provide a useful aid for services when they are undertaking their risk assessment.
- **Annex 6, 'Record Keeping and Review Guidance'**: provides greater clarity for services on how to make and keep written records of their risk assessments and measures taken to comply with their safety duties, as well as regularly reviewing their compliance with all their duties. This guidance applies across multiple duties under the Act.

7.4 The recommendations on governance and accountability which we discuss in this volume are captured in our Codes of Practice (Annexes 7 and 8). We discuss our approach to Codes of Practice in more detail in Chapter 11 of Volume 4. This includes, in particular, our proposals in relation to what sort of services measures should apply to. We explain there

[1] We note that our governance proposals are part of our Codes of Practice proposals. We present them in this volume, given they closely align to what services should do in relation to their risk assessments and record keeping and review. Our proposed measures for governance can be found in Annexes 7 and 8.

how we are defining concepts used in our governance and accountability recommendations, including our proposal to define ‘large’ services as those with more than 7 million monthly UK users, our proposal to define ‘multi-risk’ services as those that identify as medium or high risk for at least two different kinds of illegal harm in their latest illegal harms risk assessment, and how we are defining ‘general search service’, ‘vertical search service’ and whether or not a service manages its index itself or procures it from another search service provider.

7.5 We are consulting on our Codes and these draft guidance documents and invite feedback on our approach to developing them, as well as the drafts themselves. We have set out specific consultation questions in each chapter in this volume on issues where we would particularly welcome feedback and further supporting information to inform our final versions of the codes and these guidance documents. See Annexes 1-4, for more information about how to respond to our consultation.

7.6 Having reviewed responses to this consultation, we will then publish our final decisions in a Statement and our final versions of our codes and these guidance documents. Once we publish our Statement, services will have three months from the date of publication of the Services Risk Assessment Guidance to conduct their illegal content risk assessment and produce written records of these risk assessment and any measures taken to comply with a safety duty.

8. Governance and accountability

What is this chapter about?

Governance and accountability processes are key to a service's ability to properly identify and manage online safety risks.

This chapter sets out our proposed recommendations regarding how services should approach governance and accountability in relation to their illegal content duties under the Act. It covers measures related to governance arrangements; senior accountability and responsibility; internal assurance and compliance functions; and staff policies and practices.

For proportionality reasons, we propose that most measures only relate to large and/or multi-risk services.[2] However, we propose that the requirement for a senior accountable officer applies to all services (U2U and search).

What are we proposing?

We are making the following proposals for all services:

- **Name a person accountable to the most senior governance body for compliance with illegal content duties and reporting and complaints duties**.

We are making the following proposals for all multi risk services and all large services[3]:

- **Written statements of responsibilities for senior members of staff who make decisions related to the management of online safety risks**.
- **Track evidence of new kinds of illegal content on their services, and unusual increases in particular kinds of illegal content**, and report this evidence through the relevant governance channels. U2U services should also track and report equivalent changes in the use of the service for the commission or facilitation of **priority offences**.
- **A Code of Conduct that sets standards and expectations for employees around protecting users from risks of illegal harm.**
- **That staff involved in the design and operational management of the service are sufficiently trained in the service's approach to compliance**.

We are also making the following proposals for large services:

- **The most senior body in relation to the service should carry out and record an annual review of risk management activities in relation to online safety, and how developing governance risks are being monitored and managed**.

[2] For further detail, 5please see our Introduction to Volume 4, where we define 'large' and 'multi-risk' services.
[3] This is with the exception of large vertical search services. This is because we are not aware of evidence of such services showing illegal content and by their nature vertical search services are unlikely to have content that is as rapidly changing as U2U services and the search results are more under their control than for U2U. content. We also propose to exclude vertical search from the measure relating to reporting annually to the most senior governance body, for the same reasons.

- **Large multi-risk services should have an internal monitoring and assurance function[4] to provide independent assurance that measures taken to mitigate and manage the risk of harm to individuals identified in the risks assessment are effective on an on-going basis, reporting to an overall governance body or audit committee**.

Why are we proposing this?

Robust governance processes are an effective way of ensuring good risk management and we therefore expect that widespread adoption of such governance processes will make a material contribution to reducing online harm. Although there is the potential for significant costs in some areas, we consider that good governance is sufficiently important and beneficial to justify these costs. We also consider that the costs of deploying good governance to prevent risks from materialising will often be less significant than the costs services would incur remedying risks that have already materialised. Targeting several of these measures at only large and/or multi-risk services will help ensure we are not imposing undue costs on services that pose a low risk of online harm. Many of the services we are targeting will already have existing governance and accountability arrangements which can accommodate these recommendations.

We are not yet making any recommendations regarding external audit requirements, or regarding linking remuneration and bonuses to online safety outcomes due to limitations in currently available evidence that demonstrates the effectiveness and costs of these proposals.

The proposals for organisations that operate large services are designed to be consistent with the operation of a ‘three lines of defence’ governance model, and can easily be mapped to the first (management), second (risk management and compliance) and third line of defence (internal audit).

What input do we want from stakeholders?

- Do you agree with our proposals in relation to governance and accountability measures in the illegal content Codes of Practice? Please provide underlying arguments and evidence of efficacy or risks to support your view.
- Do you agree with the types of services that we propose the governance and accountability measures should apply to?
- Are you aware of any additional evidence of the efficacy, costs and risks associated with a potential future measure to requiring services to have measures to mitigate and manage illegal content risks audited by an independent third-party?
- Are you aware of any additional evidence of the efficacy, costs and risks associated with a potential future measure to tie remuneration for senior managers to positive online safety outcomes?

Introduction

8.1 As set out in our Register of Risks, there is a wide range of risks of relevant harms on regulated services. A service's risk assessment should identify those risks, but they cannot be managed without a proper internal process. As set out in Step 3 of the Risk Assessment Guidance there will also always be residual risks which require monitoring and management.[5]

[4] Where appropriate, this could be fulfilled by an existing internal audit function.

[5] See Annex 5 (Service Risk Assessment Guidance).

8.2 Effective governance and accountability structures provide the foundation for how a service identifies, manages, and reviews online safety risks to its users. Having well-functioning governance and organisational design processes requires organisations to embed principles of accountability, oversight, independence, transparency, and clarity of purpose into their operations.

8.3 Good organisational governance and accountability helps services to understand and anticipate risks and to communicate them internally. These processes also help to identify appropriate risk mitigations. This increases the likelihood of risks to users being prioritised appropriately and factored into strategic decision making. It also increases the likelihood that mitigations are implemented effectively.

8.4 Having these processes in place also makes it more likely that services would be prepared to deal with changes in the online landscape that may increase risks to users, including sudden spikes in illegal content and sensitive events, as well as monitoring and reviewing the effectiveness of measures designed to reduce risk. In this way, governance and organisational design should be seen as a fundamental part of ongoing risk management.

8.5 Where governance measures do not exist, or where they are inconsistently or ineffectively implemented, there is a greater risk that a service will fail to manage risks. This has been demonstrated in high-profile examples of corporate organisational failure due to weak governance processes in other sectors.

8.6 Since good governance and organisational design are critical in good risk management, it follows that weak governance measures will increase the likelihood that internal controls to mitigate risk will be implemented inconsistently or inappropriately. For this reason, ensuring that services put in place well-functioning governance systems is one of Ofcom’s four strategic priorities for the early years of the online safety regime.

8.7 Ofcom’s proposed recommendations for governance and organisational design measures would ensure that services have appropriate assurance, oversight, awareness-building and reporting processes in place to support the management of risks identified in illegal content risk assessments. The aim of these proposals is that services are better placed to keep users safe from illegal content risks through effective risk management.

8.8 We have considered whether our proposals in this area would have implications for freedom of expression or privacy, or broader equality impacts.[6] We do not think they would. This is because governance and accountability are wholly concerned with the organisation and internal structure and processes of regulated services as businesses. However, a well-managed business is, in general, more likely to comply with its obligations under privacy, data protection and equality laws. As such, our proposals may help to safeguard these.

Relevant provisions

8.9 U2U and Search services have safety duties about illegal content under sections 10 and 27, respectively. In particular, services have a duty to take or use proportionate measures relating to the design and operation of their services to effectively mitigate and manage the risks of harm to individuals, as identified in their most recent illegal content risk assessment.

8.10 Regulated services have a duty to use proportionate measures designed to prevent individuals from encountering priority illegal content by means of the service. U2U services

[6] See Annex 13, which outlines our view of the impacts of our proposals on protected characteristics.

also have a duty to use proportionate measures to mitigate and manage the risk of the service being used for the commission and facilitation of a priority offence, and to have systems and processes to minimise the length of time for which any priority illegal content is present.

8.11 The above duties apply across all areas of the services and include a duty to take measures, if appropriate, relating to regulatory compliance and risk management arrangements and staff policies and practices.[7]

8.12 Our proposed recommendations around governance and organisational design would ensure that services have adequate structures, governance and accountability processes in place to maintain effective mitigation and management of risks of harm, in accordance with the safety duties.

Harms that these measures seek to address

8.13 Governance and accountability underpin the way that a service manages risk and ensures that efforts to mitigate them are effective. We consider that these processes are essential components of a well-functioning system of organisational scrutiny, checks and balances, and transparency around risk management activities. Effective governance and accountability processes should be effective in tackling all priority illegal harms.

8.14 Good practice in governance processes is multi-faceted. This means that there is no single governance and accountability policy option that can comprehensively address harms and risks that stem from non-existent or inconsistent risk mitigation and management. As such, we consider that it would be necessary for some services to implement several measures to address the harms and risks which may overlap in some areas. We consider that this would be the best way to ensure that illegal content risks are mitigated and managed effectively.

8.15 We have structured our proposed recommendations into thematic areas, ranging from the overall governance structure to policies for individual staff members. In each of these thematic areas, we provide more detail on the specific risks the relevant measures seek to address:

a) Annual review of risk management activities;
b) Senior accountability and responsibility;
c) Internal assurance and compliance functions; and
d) Staff incentives, policies and processes.

8.16 Our first Codes are aimed at establishing robust governance and accountability processes and represent a basis on which to build. We anticipate making further updates to our Codes through a process of iteration as our evidence base evolves.

8.17 Our review of good practice standards and principles in risk management and corporate governance across a range of different industries demonstrates the importance of clear, consistent, and codified assurance processes, governance structures, reporting mechanisms and internal communications in ensuring good safety practices and positive outcomes for users and consumers. Our recommendations are informed by this evidence base.

[7] See paragraphs 5.5 b) and c) in Annex 12 (Legal Framework). These paragraphs set out further detail on services' duties with respect to illegal content.

Annual review of risk management activities

Harms that this measure seeks to address

8.18 Services have duties under the Act to carry out risk assessments, to record details about how risk assessments are carried out and regularly to review compliance with their safety duties and their duties in relation to complaints and reporting.[8] However, this does not, on its own, necessarily secure that issues are scrutinised at a senior level.

8.19 Users are more likely to be exposed to illegal content risks where there is insufficient oversight and scrutiny of risk management activities. One of the key remits of a governance body is to monitor the effectiveness of a company's risk and governance practices.[9] Evidence also suggests that the structure of a governance body itself influences organisational approaches to risk management.[10]

8.20 The description or name given to a governance body in a service will vary depending on its size, structure, operating model, or preference. For the purpose of these Codes we refer to it as the body responsible for overall governance and strategic direction of a service. Some services may have a fully independent board with non-executive representation and a separate sub-committee for risk and audit, whereas others may have a senior leadership team providing necessary challenge and oversight.

8.21 Where a governance body fails to fulfil its functions, there is a risk that risk management activities are not adequately challenged or scrutinised. This could result in inadequately identifying and managing online safety risks with a consequential increase in illegal content and harm. There are numerous examples from other industries of how failures in board governance led to reduced safety for users and consumers.[11]

[8] Sections 9, 23(2), 23(6), 26, 34(2) and 34(6) of the Act.

[9] OECD, 2015. G20/OECD Principles of Corporate Governance. Subsequent references are to this document throughout; Milliman, 2023. Report on principles-based best practices for online safety Governance and Risk Management. This report was commissioned by Ofcom. Subsequent references are to this document throughout.

[10] Akbar, S., Kharabsheh, B., Poletti Hughes, J. and Shah, SZA., 2017. Board Structure and Corporate Risk Taking in the UK Financial Sector, *International Review of Financial Analysis*, 50, pp. 101-110. [accessed 04 September 2023].

[11] The OECD concludes that analysis of past incidents [including major safety incidents in high hazard industries] reveals that inadequate leadership have been recurrent features, "including the monitoring of safety performance indicators at Board level". Source: OECD, 2012. Corporate Governance for Process Safety OECD Guidance for Senior Leaders in High Hazard Industries. [accessed 04 September 2023]; This includes lawsuits filed against Boeing following the crashes of two 737 MAX airplanes in 2018 and 2019, in which shareholders claimed that a failure of the board to account for safety risks contributed to fatality events: "safety was no longer a subject of Board discussion, and there was no mechanism within Boeing by which safety concerns… were elevated to the Board or to any Board committee". Source: Consolidated complaint regarding Boeing accessed via the Washington Post, 2021 [accessed 05 May 2023]; The Health and Safety Executive offers several case studies of negative safety consequences when board members do not lead effectively on health and safety management. Source: Health and Safety Executive, (HSE). Case studies: When leadership falls short [accessed 04 September 2023].

Options

8.22 One of the roles of a governance body is to ensure that risk management frameworks within organisations operate effectively. They are therefore intrinsic to the proper functioning of how an organisation manages and mitigates risks.

8.23 In this area, we considered just one option, a recommended measure that boards or overall governance bodies of services carry out and record an annual review of risk management activities in relation to online safety, and how developing risks are being monitored and managed.

8.24 A review may form part of existing governance processes for annual review of strategic risks. This should include a review of risk oversight policy and procedures as related to online safety, including risk assessment processes, mitigations, trends and (where applicable) lessons learned from past mistakes.

Effectiveness

8.25 Regular review of risk management and regulatory compliance by a governance body is required for appropriate oversight over internal controls. Evidence supporting this principle can be found in corporate governance good practice principles and codes.[12]

8.26 It will be important for governance bodies within services to have a full understanding of risks as identified in an illegal content risk assessment, measures that a service has put in place to mitigate and manage those risks, and how a service intends to deal with developing areas of risk. This requires that governance bodies are made aware of relevant information regarding risk management in a service (provided, for example, by internal assurance functions) and have appropriate reporting lines with senior management.

8.27 We found evidence that effective use of data and information to report on risks to boards is associated with good risk management.[13] In Ernst & Young's 2021 Global Board Risk Survey of 500 companies, over 70% of companies regarded as highly effective at risk management provided timely and insight-driven risk reporting to their board and leveraged data and technology to be more predictive in their risk reporting.[14] The communication of this

[12] Under the UK Corporate Code, companies with a premium listing on the London Stock Exchange are already required to follow principles related to board oversight. This includes Provision 29 which states that boards "should monitor the company's risk management and internal control systems and, at least annually, carry out a review of their effectiveness and report on that review in the annual report". Monitoring and review activities are intended to cover all material controls, including financial, operational and compliance controls. Source: Financial Reporting Council, 2018. The UK Corporate Governance Code, pp.10 The OECD's Principles of Corporate Governance similar suggests that a key function of the Board should be "reviewing and guiding corporate strategy, major plans of action [and] risk management policies and procedures". The Principles suggest that while committees or other sub-bodies may have specific responsibilities for different areas of risk, "the board should retain final responsibility for oversight of the company's risk management system and for ensuring the integrity of the reporting systems". Source: OECD, 2015.

[13] This study by the Financial Reporting Council with participants from over 40 listed companies concluded that it was important for Boards to have a whole view of risk (including "gross" or inherent risks) to engage in meaningful discussion. Several organisations specified that they reported on emerging risks as well as more conventional risk registers to improve Boards' oversight across risk areas. Source: FRC, 2011. Boards and Risk A summary of discussions with companies, investors, and advisers.[accessed 4 May 2023].

[14] By comparison, only 16% of companies with developing risk management approaches provided such information to the Board, with 5% using data and technology to be more predictive in their insights. Source: Isaac Sarpong, 2022. The Board Imperative: How can data and tech turn risk into confidence?, *Ernst & Young* [accessed 4 May 2023].

information to the board is important for assessments of the adequacy and effectiveness of internal controls, and whether any changes are required to improve risk management.[15]

8.28 Best practice guidance and codes for governance bodies and boards points to the importance of setting a regular schedule around the review of risk management activities. Often this is framed as part of an annual cycle, tying in with financial and company results reporting or public disclosure.[16] As an example, the UK Corporate Code requires a company's board to use an annual report to confirm it has completed an assessment of emerging and principal risks, what procedures are in place to identify emerging risks, and an explanation of how these are being managed or mitigated.[17]

8.29 Services which responded to our 2022 Illegal Harms Call for Evidence described existing structures by which risk management activities are subject to review by a governance body or equivalent. Google explained how risks related to content issues are reported by senior management to the Audit and Compliance Committee for Alphabet at least annually[18], which helps ensure Board-level accountability for user safety.

8.30 We also note that there are various factors which influence the efficacy of a governance body in providing appropriate oversight on risk management activities that go beyond review and reporting schedules. For example, several studies that we have reviewed reference the potential importance of elements such as independence[19] and diversity[20] of governance body members in ensuring that a Board or governance body discharges its duties effectively. We consider these factors as relevant to the extent that they enable a governance body to carry out reviews of risk management activities effectively.

Costs and risks

8.31 Most large services will be run by companies that already have an existing governance body or board which is ultimately responsible for oversight of risk management and compliance activities.[21] Such services will have additional on-going costs relating to preparing an annual paper for the relevant governance body which sets out the service's risk management activities for online safety specifically, which the governance body will then need to review and scrutinise. If this were done by the main board of a large company, then we estimate

[15] UK Government Finance Function (GFF), 2021. Good Practice Guide: Risk Reporting. [accessed 4 September 2023].

[16] Best practice for overall governance bodies is to maintain an annual cycle of planned activity, to ensure that there is time for full consideration of specific exposures. Source: Milliman, 2023.

[17] FRC, 2015. The UK Corporate Code also specifies that boards should carry out a review at least annually of the effectiveness of the company's risk management and internal control systems. Companies should also report on that review in the annual report.

[18] Google's response to 2022 Illegal Harms Call for Evidence.

[19] Guluma, T. F., 2021. The impact of corporate governance measures on firm performance: the influences of managerial overconfidence, Future Business Journal, 7 (50) [accessed 04 May 2023].

[20] Creary, S.J., McDonnell, M-H., Ghai, S., Scruggs, J., 2019. When and Why Diversity Improves Your Board's Performance, *Harvard Business Review, 27 March.* [accessed 04 May 2023].

[21] Of the services we are aware of that have a relevant user base of more than 7 million, which is how we propose to classify services as large, most are ultimately owned by listed companies, typically in the US. Listed companies in the US are required by their respective stock exchange to have an audit committee which is required to discuss "policies with respect to risk assessment and risk management". Source: New York Stock Exchange 2009. NYSE Audit Committee Responsibilities [accessed 18 September 2023].

that this would cost £16,000 to £36,000 per year.[22] In addition to this cost, some services may need to consider whether their governance body has the right expertise, in terms of risk management or technical expertise, to be able to understand what they are overseeing, and may need to change the composition of the body if necessary. If the governance body scrutinising the online safety risk management and compliance activities were instead a lower governance body or a specialist committee, the costs will tend to be much lower than the costs of the main board considering doing this, not least as such bodies are likely to have fewer members.[23]

8.32 Smaller companies would also be expected to have much lower costs as their boards tend to be smaller and salaries lower. Any large services will lower risks will also tend to have lower costs, as reporting the annual review of risk management activities related on online safety to the governance body will be simpler if the risks are low.

8.33 Services that do not currently have a governance body suitable for this responsibility would need to establish such a body to consider online safety risk management and compliance activities. Where suitable people are available within the organisation, then the costs may be similar to those above. However, organisations that do not have suitable people internally would need to identify external people who would sit on the governance board. In addition to the costs above, they would incur the costs of hiring those people, who may then have to spend time understanding the service.

8.34 Many smaller services will not have a formalised overall governance body, although they may have external advisors or funders who scrutinise risk management arrangements. The set-up and on-going costs of a governance body to scrutinise risk management plans is likely to be a higher share of revenue for a smaller service.

8.35 Moreover, the remit of a governance body also goes beyond overseeing risk management and mitigation. For small organisations and start-ups, the decision to establish a board or other governance mechanism may come at a stage when the organisations need greater expertise on how to achieve sustainable growth, business contacts and long-term strategic support.[24] It is therefore unlikely to be proportionate to require smaller services to establish a governance body solely for the purposes of compliance with illegal content duties and reporting and complaints duties.

8.36 The annual report and governance body's scrutiny may identify problems with the way online safety risk management and compliance is currently conducted. Changes may be

[22] This is derived from the assumptions set out in Annex 14 combined with the following specific assumptions. We assume it takes 10 to 20 days to prepare the paper for the board and that on average each director on the board spends 1 to 2 hours in total to read, consider and discuss the paper. We calculate the fraction this represented of a board director's time, assuming on average directors spend 250 hours a year on board related activities for each company they are a director for, based on PwC's 2022 Annual Corporate Directors Survey. For average total remuneration of board members, we assume $316,091 per year, based on the average for 2022 of S&P 500 independent board directors, from 2022 U.S. Spencer Stuart Board Index a report by Spencer Stuart (a leadership consultancy). Many of the largest services are owned by US companies. For the number of board members, we assume boards have on average 11 members, based on the average S&P 500 board size, from Diversity, Experience, and Effectiveness in Board Composition, Harvard Law School Forum on Corporate Governance, Merel Spierings, 14 June, 2022. [accessed 27 September 2023].

[23] Some specialist board committees (such as audit, compensation and governance) typically have 3 or 4 members, based on the National Association of Corporate Directors Public Company Governance Survey 2019–2020. The 2022 U.S. Spencer Stuart Board Index a report by Spencer Stuart (a leadership consultancy) found that 12% of S&P 500 companies have a risk standing committee. [accessed 27 September 2023].

[24] International Financial Corporate, 2019. SME Governance Guidebook. [accessed 17 May 2023].

needed as a result. The purpose of this measure is that this scrutiny leads to better online safety risk management, and improved online safety for users. This could be thought of as an indirect cost of this measure. We have not separately estimated any such indirect costs, as they would vary depending on the specific problems identified and how they were addressed.

8.37 Steps taken to improve online safety as a result of this measure may affect other measures that we propose recommending, such as those to do with content moderation. We assume that if any such costs were incurred it will only be because there would be benefits in terms of enhanced online safety which would make such changes proportionate. Any costs could also be regarded as relating to those other measures, rather than as a result of this governance process. This is the case for all the governance measures considered below.

Provisional conclusion

8.38 We propose to recommend that large services should ensure that boards or overall governance bodies carry out and record an annual review of risk management activities in relation to online safety, and how developing risks are being monitored and managed.

8.39 Consistent with the literature around best practice, we consider that this recommendation would result in better risk management, reducing the level of harm users are exposed to. We consider that this would deliver significant benefits in terms of end-to-end risk management of illegal harms for such services.

8.40 We propose to recommend this measure even for large services that have not identified a medium or high risk for any kind of illegal harm. This is because a failure in oversight of risk mitigation and management at large services, even those that currently do not identify any risks for users, would affect a greater number of users and could have a very significant adverse impact. Hence the importance and benefits of the measure would be bigger for larger services than for smaller services. For these purposes we propose that a large service is a service with more than 7 million monthly UK users, as discussed in Chapter 11.[25]

8.41 Moreover, we consider that the role played by governance bodies in ensuring that a service's approach to addressing illegal content risks is appropriate is likely to be more significant in large services with complex operations, because of the need for high-level oversight of coordination and consistency in risk management. Even if large services do not currently identify any medium or high risks, the likely complexity of their services and the possibility that some risks have not been examined properly means that it will be important for a senior governance body to review their approach to risk management.

8.42 We propose not to recommend this measure to large vertical search services just because they are large. By their nature vertical search services are unlikely to have content that is as rapidly changing as U2U services and search results are more under a service's control than for U2U content. We are also not aware of evidence of such services showing illegal content.[26] Any benefits of applying this measure would therefore be low for vertical search services.

8.43 As we have set out, we consider the direct costs of the measure are likely to be low for large services which are already likely to have a suitable established governance body responsible

[25] Please see the commentary from paragraph 11.50 in Chapter 11 onwards.

[26] See [Register of risks, paragraphs 6.21(b)] for why we consider vertical search services to be low risk.

for oversight of risk management. They are also likely to be low for any large services that have not identified any risks.

8.44 Any indirect costs, as a result of actions arising from the governance body's annual review, are likely to be outweighed by the additional benefits for online safety of those actions. Given the importance of strong governance and the role it can play in reducing online harms, we therefore consider that, on balance, this measure is likely to be an effective and proportionate intervention for all large services (except vertical search services).

8.45 For services that are not large, including smaller services that identify some higher risks for users, we are not proposing to recommend this measure at this time. The benefits of imposing this on smaller services are likely to be lower because these services tend to be simpler and easier for management to ensure coordination and consistency in approach.

8.46 Moreover, it is likely, particularly for smaller services which find high risks to users, that an organisation is not mature enough to have a fully developed governance body. This is especially the case for micro and start-up businesses, or small-scale non-commercial services. This measure would imply significant staff and resource costs, and a change in the overall structure and dynamic of the service for these types of organisations. This could stifle innovation.

8.47 The evidence we have cited above relating to the efficacy of this proposed measure relates to priority offences including CSEA, terrorism and relevant non-priority offences. We therefore propose to include this measure in our Codes for U2U services and search services on terrorism, CSEA and other duties.

Senior accountability and responsibility

Harms that this measure seeks to address

8.48 Senior accountability for online safety is critical in building a culture that prioritises safety for users.

8.49 Users are more likely to be exposed to illegal content risks if there is a lack of accountability at senior management level for compliance with illegal content safety duties because it is indicative of an absence of senior oversight and responsibility for decisions which have a material impact on user safety. A lack of accountability would also raise the risk that risk management activities do not receive sufficient attention or oversight from an overall governance body or board.

8.50 Evidence from other sectors, including occupational health and safety, points to failures in leadership as a cause of high-profile instances of poor safety outcomes, where organisational leadership fails to adequately consider health and safety risks.[27] This is corroborated by findings from analysis of senior leadership failures in financial services in relation to the 2008 financial crisis, which demonstrate how a lack of senior accountability can result in reduced oversight and excessive risk-taking.[28]

[27] HSE, 2013. Leading health and safety at work. [accessed 4 September 2023].

[28] In the aftermath of the 2008 financial crisis, an inquiry into professional standards and culture of the banking sector by the Parliamentary Commission on Banking Standards concluded that many bankers had been allowed to operate with little accountability, and "claimed ignorance or hid behind collective decision-making".

8.51 Risks may also arise if services do not provide clarity on roles or responsibilities for managing illegal content risks. This lack of clarity could contribute to inconsistent application of risk management measures.

Options

8.52 Good practice guidance and evidence of governance failures in other sectors suggests that having senior management answerable for risk management decisions and responsible for risk management activities is important in ensuring that these decisions are properly considered and reported on.

8.53 Having senior members of staff accountable for illegal content risk management would be important in ensuring that users of online services are protected from harm. Accountability implies that services clearly set out their expectations of senior individuals with responsibilities and decision-making powers over how risks to users are managed and mitigated within a service.

8.54 In considering ways that this could be implemented in practice, we have assessed the following options:

a) Having a named senior individual accountable to the most senior governance forum for compliance with illegal content safety duties under the online safety regime.
b) Having written statements of responsibilities for senior members of staff who make decisions related to the management of online safety risks.

8.55 For the first option, the named individual would be accountable to the governance body which oversees risk management and compliance activities where one exists. Large services would need to have such a body under the previous measure. If such a body does not exist, for example in smaller services, the individual would report to the senior management team. A sole trader would not be required to report to anyone else but would be the accountable person for ensuring compliance with the illegal content safety duties.

Effectiveness

Option a) Having a person accountable to the most senior governance forum for illegal content safety duties and reporting and complaints duties

8.56 Senior accountability is a cornerstone of other regulatory regimes, including the Senior Managers & Certification Regime (SM&CR) jointly regulated by the Financial Conduct Authority (FCA) and the Prudential Regulation Authority (PRA).[29] Findings from a 2020 review by the PRA reported positive behavioural change and improvement in risk management practices among services which have implemented SM&CR.[30]

Source: FCA, 2013 The FCA's response to the Parliamentary Commission on Banking Standards [accessed 3 May 2023].

[29] The SM&CR is directly underpinned by legislation and serves different outcomes related to compliance with financial regulation, we consider the broad lessons and findings from the FCA's implementation of the regime as instructive for other areas of risk management and regulatory compliance. Source: FCA, 2023. Senior Managers and Certification Regime. [accessed 18 September 2023].

[30] Bank of England, Prudential Regulation Authority, December 2020. Evaluation of the Senior Managers and Certification Regime. [accessed 03 May 2023]. Subsequent references are to this document throughout.

8.57 Evidence of remediation efforts in some organisations which have experienced significant governance failings point to the importance of senior-level accountability.[31] For example, findings from inquiries into corporate governance failures at Equifax (a US based credit reporting agency) concluded that lack of management accountability was a contributing factor to the failure to protect consumer data and to exercise good risk management.[32] Correspondingly, remediation efforts by the company focused on senior management changes and strengthening of reporting and accountability frameworks to improve future safety and security of customer information.[33]

8.58 Responses to our 2022 Illegal Harms Call for Evidence demonstrated that several online services already have arrangements whereby they have a dedicated accountable staff member for regulatory compliance with online safety outcomes. This included Mojeek[34], Google[35], Trustpilot[36], X[37] and Glassdoor[38], which all described overall ownership for online safety compliance at a senior manager level. The Online Dating Association (ODA) also explained that in the case of small services, this role sometimes lies directly with organisations' CEOs or founders.[39]

8.59 Senior accountability in online safety is supported by evidence on best practice in governance and risk management, and existing corporate governance codes. Work commissioned by Ofcom from Milliman[40] includes Individual Accountability as the first principle of good governance, drawing on the Institute of Internal Auditors (IIA) Three Lines Model.[41] The UK Corporate Governance Code[42] and the 2018 Wates Principles,[43] developed to strengthen the corporate governance framework for the UK's largest companies, emphasise accountability of executive directors as a provision or principle to support effective decision-making.

8.60 We also consider it is important that accountability is owned at senior levels, and with those who have an overall decision-making remit within an organisation. Having direct reporting lines into an overall governance body, such as a board, Executive Committee or equivalent,

[31] Veetikazhi, R., & Krishnan, G. 2019. Wells Fargo: Fall from Great to Miserable: A Case Study on Corporate Governance Failures. *South Asian Journal of Business and Management* Cases, 8(1), pp 88–99. [accessed 6 September 2023].

[32] The report concludes that a lack of management accountability was a "significant factor" in the 2017 data breach of personal customer information at Equifax. Source: U.S. House of Representatives Committee on Oversight and Governance Reform, 2018. The Equifax Data Breach [accessed 03 May 2023].

[33] Equifax, 2018. Notice of 2018 Annual Meeting and Proxy Statement [accessed 03 May 2023].

[34] Mojeek's response to the 2022 Illegal Harms Call for Evidence.

[35] Google's response to 2022 Illegal Harms Call for Evidence.

[36] Trustpilot's response to 2022 Illegal Harms Call for Evidence.

[37] X's response to the 2022 Illegal Harms Call for Evidence.

[38] Glassdoor's response to 2022 Illegal Harms Call for Evidence.

[39] Online Dating Association's response to the 2022 Illegal Harms Call for Evidence.

[40] Milliman, 2023.

[41] IIA, 2020. The IIA's Three Lines Model [accessed 03 May 2023]. Subsequent references are to this document throughout.

[42] Under 'Division of Responsibilities', the Code states that the board and non-executive directors have a key role in holding "to account the performance of management and individual executive directors against agreed performance objectives" Source: FRC, 2018. The Waites Corporate Governance Principals For Large Private Companies [accessed 03 May 2023]. Subsequent references throughout

[43] Principle Three on Director Responsibilities states that "the board and individual directors should have a clear understanding of their accountability and responsibilities". It further states that the board has a role in providing clear lines of accountability and responsibility to support effective decision making. Source: FRC, 2018. [accessed 03 May 2023].

is likely to be crucial in ensuring that key decision-makers are properly held to account and that decisions are scrutinised at the highest level in an organisation. We assess that the benefits of having an accountable person would apply in the mitigation and management of all kinds of illegal harm.

8.61 For the avoidance of doubt, this measure would be separate from other aspects of the online safety regime that may make reference to individuals within organisations, such as in respect of senior managers' liability or information notices.

Option b) Having written statements of responsibilities for senior members of staff

8.62 Specifying responsibilities for senior decision-makers is an important feature of other regulatory regimes and has been found to be effective in improving outcomes. Statements of responsibilities for senior management is a concept drawn from the FCA's SM&CR, which requires organisations to specify areas of responsibility for senior members of staff in relation to certain controlled functions.

8.63 A statement of responsibilities is a document which clearly shows the responsibilities that the senior manager performs and how they fit in with the firm's overall governance and management arrangements.[44] From an online safety perspective, the purpose of statements of responsibilities would be to ensure that all key responsibilities for decision making in online safety risk management are assigned to senior management, and that there is clarity in how these responsibilities are owned within a service.

8.64 Those key responsibilities would include ownership of decision-making and business activities that are likely to have a material impact on user safety outcomes. Examples include senior-level responsibility for key decisions related to the management of risk on the front, middle and back ends of a service.[45] This would include decisions related to the design of the parts of a product that users interact with (including how user behaviour / behavioural biases have been taken into account), how data related to user safety is collected and processed, and how humans and machines implement trust and safety policies. Depending on a service's structure, key responsibilities in online safety may fall under content policy, content design and strategy, data science and analytics, engineering, legal, operations, law enforcement and compliance, product policy, product management or other functions.[46]

8.65 Findings from a 2020 review by the PRA of the FCA's SM&CR regime reported that many firms surveyed said the requirements of the regime had resulted in clearer articulation of authority and had improved focus on accountability and responsibility.[47] These findings mirrored a 2014 cost benefit analysis of the SM&CR, where large banks surveyed anticipated that statements of responsibility would impact behaviour around decision-making and risk.[48]

[44] Under financial regulation in the UK, there are specific considerations that regulated firms must take in drafting their statements of responsibilities, which would not apply to online safety regulation. These include specifying responsibilities in relation to controlled functions.

[45] Maxim, K., Parecki, J., & Cornett, C. 2022. How to Build a Trust and Safety Team In a Year: A Practical Guide From Lessons Learned (So Far) At Zoom. *Journal of Online Trust and Safety*, 1(4). [accessed 4 September 2023].

[46] Examples are listed by the Trust and Safety Professionals Association (TSPA). Source: Trust and Safety Professionals Association. Key functions and roles [accessed 4 September 2023].

[47] PRA, 2020.

[48] "Large banks and investment firms did consider it likely that the policies would result in behavioural changes as senior managers sought to ensure they would be protected in the event that misconduct or a regulatory breach was discovered, driven by the statement of responsibilities and the presumption of senior

8.66 International corporate governance principles support ensuring that senior decision-makers have clear responsibilities as part of good risk management. This includes the G20 / Organisation for Economic Co-operation and Development's (OECD) Principles of Corporate Governance, which suggests that the specification of accountabilities and responsibilities for managing risk is a "crucial guideline for management" within organisations.[49] We found corroboration of this principle among several good practice models for governance and risk management, including the Committee of Sponsoring Organisations of the Treadway Commission (COSO) Enterprise Risk Management (ERM) framework[50] and the IIA's Three Lines Model.[51]

8.67 Ofcom-commissioned work on best practice in risk management and governance for online safety by Milliman similarly highlights the importance of "having clearly defined roles and responsibilities for all senior managers" and individual accountability in forward-looking risk management systems. [52]

8.68 Several services suggested in their responses to our 2022 Illegal Harms Call for Evidence that they already specify responsibilities for senior members of staff in relation to online safety and risk management. In response to our 2022 Illegal Harms Call for Evidence, Google[53] and Trustpilot[54] highlighted the specific role their senior management plays regarding the direct oversight of risk, ensuring appropriate resourcing and cascading of responsibility in the management of risk, review of escalations and responsibility for reporting on risk and risk management activities to the Board.

Costs and risks

Option a) Having a person accountable to the most senior governance forum for illegal content safety duties and reporting and complaints duties

8.69 We anticipate that most services will choose to add accountability for compliance to the current portfolio of a senior manager or director who already oversees an online safety, compliance, or risk function. The costs of selecting and naming such an individual are likely to be negligible for such services. Any services that do not have a suitable individual would incur greater costs, as they would need to make changes to their internal structure for compliance. This is something they would need to do as a result of the Act coming into force.

responsibility. Such behaviour includes increased due diligence, monitoring and sign-off processes, as well as more formalised and considered decision-making. These actions are all likely to contribute to an increased likelihood that potential and actual regulatory breaches are identified and prevented." Source: Europe Economics, 2014. Cost Benefit Analysis of the New Regime for Individual Accountability and Renumeration. https://www.bankofengland.co.uk/-/media/boe/files/prudential-regulation/consultation-paper/2014/cp1414annex10.pdf?la=en&hash=C646BCBA58A94D0C6CEE7A0CC9C5995BA92A0435 [accessed 11 May 2023].

[49] OECD, 2015.

[50] This model focuses on how clarity of responsibilities among managers supports the proper functioning of internal controls. Source: COSO, November 2020. https://www.coso.org/_files/ugd/3059fc_5193266654244b96b9b2ed7b1270d1e2.pdf [accessed 18 September 2023]. Subsequent references are to this document throughout.

[51] The Three Lines model specifies that second line management roles require expertise, support, monitoring and challenge on risk-related matters including risk management. Source: IIA, July 2020.

[52] Milliman, 2023.

[53] Google's response to 2022 Illegal Harms Call for Evidence.

[54] Trustpilot's response to 2022 Illegal Harms Call for Evidence.

All services would anyway need to incur costs to familiarise themselves with their online safety duties in the Act, and DSIT's impact assessment for the Act already accounts for this.

8.70 However, we expect there to be some costs flowing from this individual being named. As a result of this measure, the named individual would have an increased focus on online safety risk management as a result of being personally accountable. We would therefore expect the named individual to spend more time considering the illegal content duties than they would otherwise.

8.71 The extent of this is likely to depend in part on how risky the service is. For a low risk service, it may have little impact on what the named individual does, and the costs may be negligible. For a higher risk service, having this accountability could have a more substantial impact on the person's role. This could result in some of that person's existing responsibilities needing to be backfilled by others. We could assume that for a larger, higher risk service the accountable person spends 10 days more a year considering online safety than they would otherwise, purely as a result of being named the accountable person rather than a change in role. This would result in an increase in costs of £8,000 a year for most higher risk services, assuming the individual is a senior leader.[55]

8.72 There could also be other costs flowing from the change in behaviour of the accountable person, as they focus more on the service's illegal content duties. These could be considered indirect costs of the measure, and could be significantly more than the costs discussed above. As discussed above, we think it likely that where such costs are incurred, they are likely to be proportionate as there are likely benefits from users being safer from illegal content. These costs could also be regarded as relating to some of the other measures we propose, such as content moderation. Such changes are the intention of this measure.

Option b) Having written statements of responsibilities for senior members of staff

8.73 Services would need to consider which senior managers are involved in making decisions that could have a significant impact on online safety risk management. One-off costs would include the development of a statement of responsibilities for each such individual and developing guidance relating to the handover of responsibilities. Ongoing costs would include maintaining a centralised reference of responsibilities and refreshing it (either as part of a general annual review or a revision each time an individual left or assumed a role) and developing and retaining policies for handover.

8.74 We anticipate that most services would be likely to have fewer than 10 such individuals, though the largest and most complex services could have more. We assume that on average it would take a few days to develop and agree each of these. This is because we assume this requirement can be partly related to existing objectives, job descriptions or performance management processes for senior managers. We assume that much of the time to develop the statement of responsibilities would need to be from senior leaders. If there were 10 senior individuals, we assume this would cost around £16,000 in the first year, with ongoing costs being lower.[56] We would expect costs to be lower for smaller or less risky services, as they would have fewer such individuals and their remuneration may be lower. However, for

[55] As we assume this person is a senior leader of a large service, we assume an annual salary of £150,000. We also assume an uplift on safety to reflect non-wage costs, as set out in Annex 14.

[56] This is based on assuming an average annual salary of £100,000. We assume that services already have senior staff undertaking relevant roles and so do not need to recruit people as a result of this measure. If this were not the case, costs would be much more substantial.

smaller services, the costs would likely represent a greater share of the services' revenue compared to larger services.

8.75 As discussed above, some services are already implementing this measure and hence would not need to incur these costs. They would only incur higher costs as a result of this measure if they wished to stop doing this in the future.

8.76 As with the other governance measures, there could also be indirect costs as a result of this measure, but we have not considered these here for the reasons discussed in paragraph 8.35.

Provisional conclusion

Option a) Having a person accountable to the most senior governance forum for illegal content safety duties and reporting and complaints duties

8.77 We propose that all services should have a person accountable to the most senior governance body for compliance with illegal content duties and reporting and complaints duties. Being accountable means being required to explain and justify actions or decisions regarding online safety risk management and mitigation and compliance with the relevant duties to the most senior governance body.

8.78 As set out above, the evidence we have considered suggests that clearly defining senior accountability materially improves risk management and associated safety outcomes. The benefits of this option are that all illegal risks on a platform are more effectively identified and managed and it would therefore ensure proportionate action is taken to address these risks, reducing harm to users.

8.79 We consider that this measure could also provide indirect benefits for some services. For example, by ensuring they have adequate risk management and governance frameworks in place from an early stage, which can evolve and expand as the business grows, smaller firms can address any online safety issues early and even save costs overall.

8.80 The direct costs of this proposed measure scale with the risk of a service, in the sense that the impact on the named person's time will be greater if a service has identified more or greater risks of illegal harm to users. It could be negligible for low risk services and be limited to a small number of thousands of pounds for riskier services. If there are indirect costs arising from any additional action taken as a result of the named person's involvement, these would only be incurred by services which find higher risks to users and would be expected to be outweighed by the benefits from reduced illegal harms resulting from that action.

8.81 Given that the costs of the measure are relatively low and the evidence referenced above suggests it would deliver significant benefits, we consider it proportionate to recommend all services to have a named senior position holder that is accountable for illegal content online safety duties and reporting and complaints duties.

8.82 As we expect this proposed measure to help with priority offences including CSEA, terrorism and relevant non-priority offences, we propose to include this measure in our Codes for U2U and search services on terrorism, CSEA and other duties.

Option b) Having written statements of responsibilities for senior members of staff

8.83 Additionally, we propose that large services and services that have assessed themselves as multi-risk should have written statements of responsibilities for senior members of staff who make decisions related to the management of online safety risks. A statement of responsibilities is a document which clearly shows the responsibilities that the senior manager performs in relation to online safety risk management and how they fit in with the service provider's overall governance and management arrangements in relation to the service.

8.84 For services that have identified significant risks in their illegal content risk assessment, our analysis suggests that there are considerable benefits from improved outcomes from having statements of responsibilities for members of staff who perform functions relevant to online safety risk management. For what it means to have significant risks, we propose meaning a service that identifies as multi-risk, for the reasons discussed in Chapter 11.[57]

8.85 For large services with significant risks of illegal content, there are likely to be particular benefits in having a more comprehensive oversight of risk management activities and greater clarity regarding individual responsibilities. The complexity of organisational structures within large organisations means that clarity of responsibilities will be important in ensuring that risk management activities are properly scrutinised by senior management.

8.86 The benefits of this proposed recommendation applying to large services with low risks of illegal harm would not be as great, as there would be less scope to reduce harms from illegal content. However, because large services have high reach and the potential to affect a lot of users, we consider that failures in oversight of risk management would have wider impacts on users. Considering the dynamic and rapidly shifting nature of illegal harm, we also consider that companies who are large and low risk should have measures in place – including regarding statements of responsibility – to ensure that they can manage new and escalating risks quickly and effectively.

8.87 As with measure [A1], we propose not to recommend this measure to large vertical search services just because they are large. By their nature vertical search services are unlikely to have content that is as rapidly changing as U2U services and search results are more under a service's control than for U2U content. We are also not aware of evidence of such services showing illegal content.[58] Any benefits of applying this measure would therefore be low for vertical search services, and as such would likely be disproportionate (except for vertical search services which are assessed as being multi-risk).

8.88 We assess that there are likely to be some ongoing costs associated with these measures for both large services and services which identify as multi-risk, but that these costs are likely to be low. For most services, we expect fewer than 10 members of staff would require a statement of responsibility as part of their role. We expect the number of staff affected, and hence the costs, to vary with the size and riskiness of a service. While the costs will tend to be higher for larger and riskier services, we would also expect the benefits to be higher, consistent with this measure being proportionate.

8.89 Given the benefits of ensuring senior level responsibility and oversight for online safety, and small costs associated with this measure, we consider it proportionate to provisionally

[57] See paragraphs 11.43-46 for further detail.

[58] See [Register of risks, paragraphs 6.21(b)] for why we consider vertical search services to be low risk.

recommend to large services (with the exception of large vertical search services) and services which identify as multi-risk (including vertical search services which are multi-risk). Although for small risky services the cost impact will tend to represent a higher share of total revenue, our view is that such a measure is proportionate given the evidence that clearly defined roles and responsibilities at a senior level helps improve overall risk management processes. We consider this an important aspect in ensuring the effective management and mitigation of all illegal harms.

8.90 We are not proposing to recommend this measure for smaller and low risk services, given that the additional benefits of specifying responsibilities for such services are likely to be small and having regard to the costs of establishing and maintaining a process for assigning statements of responsibilities.

8.91 The evidence we have cited above relating to the efficacy of this proposed measure relates to priority offences including CSEA, terrorism and relevant non-priority offences. We therefore propose to include this measure in our Codes for U2U and search services on terrorism, CSEA and other duties.

Internal assurance and compliance functions

Harms that this measure seeks to address

8.92 Users may be exposed to illegal content and services may be used for the commission or facilitation of offences in cases where services do not have a process for evaluating the effectiveness of measures to manage and mitigate risks of harm identified in their illegal content risk assessments. These risks may also arise where such processes are inconsistent, where chosen measures do not go far enough to address risks of harm identified, where measures are ineffective at addressing specific risks, or where measures become less effective in mitigating the risk of harm to users over time.[59]

8.93 Evidence from examples of high-profile organisational failures highlights the importance of effective internal controls in managing and mitigating a range of risks. Root cause analysis of major corporate scandals often point to weak or absent controls as a key contributing factor to organisational failure.[60] Weak controls are more likely to result in a failure to effectively mitigate risk, either because they are improperly implemented or not fit for purpose to address how risks may manifest.

[59] In the report, we stated that while many services have rigorous procedures to assess privacy implications of new products and features before launch and in use, this was less common with respect to safety risks. We concluded that it is important that platforms wishing to prevent the upload of terrorist content to put corresponding effort into making their services sufficiently robust against exploitation by these actors and embed user safety considerations into the product and engineering design processes. Ofcom, 2022. The Buffalo Attack: Implications for Online Safety. [accessed 4 September 2023].

[60] Di Miceli Da Silveira, A2011. Corporate Scandals of the Earlier 21st Century: Have We Learned the Lessons? [accessed 03 May 2023].

8.94 We found evidence supporting the hypothesis that poor internal controls played a role in high-profile instances of organisational failures related to fraud[61], data integrity[62] and product safety[63] across other sectors.

Options

8.95 For risk management activities to be effective, organisations need to establish policies and processes to confirm that internal controls are appropriate and effective to address identified risks of harm. This ensures that risks identified in risk assessments are properly managed and mitigated on an ongoing basis, and that compliance requirements are met.

8.96 We understand these activities to be part of assurance and compliance functions within organisations. 'Assurance' refers to the verification of risk mitigations and internal controls, including activities around effectively identifying, measuring, and managing risks.

8.97 In considering how services might assure the measures that they have in place, we assessed the following options:

a) Having an internal monitoring and assurance function to provide independent assurance that measures take to mitigate and manage the risks of harm identified in the risk assessment are effective on an ongoing basis, reporting to an overall governance body or audit committee.

b) Ensuring that services track evidence of new kinds of illegal content, and unusual increases in particular kinds of illegal content, including but not limited to evidence derived from reporting and complaints processes, content moderation processes, referrals from law enforcement and information from trusted flaggers and any other expert groups, and report these new kinds of illegal content or unusual increases in illegal content through relevant governance channels to the most senior governance body;

c) Requiring services to have measures to mitigate and manage illegal content risks audited by an independent third-party;

d) Requiring due diligence of third-party contractors or providers of services involved in the mitigation and management of illegal content risks to assure their approaches lead to good online safety outcomes; and

[61] Omoteso, K., Obalola, M., 2014. 'The Role of Auditing in the Management of Corporate Fraud' in Said, R., Crowther, D., Amran, A. (eds.) *Ethics, Governance and Corporate Crime: Challenges and Consequences*, Emerald Group Publishing Limited; Hamilton, S., Micklethwait, A., 2016. *Greed and corporate failure: The lessons from recent disasters*. Springer; http://web.nacva.com/JFIA/Issues/JFIA-2022-No1-6.pdf; This includes the case study of petrochemical operators Petrobras and PdVSA, where systematic violation of internal controls and the absence of controls in key areas led to a failure to prevent or mitigate fraudulent activity

[62] In 2014, the US Food and Drug Administration (FDA) sent a warning letter to Indian pharmaceutical company Wockhardt warning of repeated failures in oversight and controls that had contributed to the deletion of data related to failed tests. Source: FDA 2017. Warning letter to Morton Grove Pharmaceuticals, Inc. [accessed 4 September 2023].

[63] "In the absence of any focus or controls on airplane safety, the Boeing Board pushed for achievement of production deadlines and competition with its chief rival, Airbus. In reviewing and approving the 737 MAX project, the Board never examined, considered, or questioned potential safety issues resulting from the re-design of the earlier generation 737 NG." https://blog.volkovlaw.com/2021/11/boeings-board-governance-failures-and-the-737-max-safety-scandal-part-iii-of-iv/ [accessed 4 September 2023].

e) Requiring that services use specific metrics to measure the effectiveness of measures to mitigate and manage illegal content risks.

8.98 We are proposing to recommend Options a) and b), and set out the discussion of the efficacy, costs and risks of both below. We follow with a brief discussion why we are not putting the other options forward as recommendations.

Effectiveness

8.99 Putting policies or processes in place to verify the effectiveness of controls enables measures to be scrutinised and evaluated for how effective they are at mitigating illegal content risks. As well as assuring the effectiveness of the measures taken, this may also enable services to identify and take action to address new forms of illegal content.

8.100 There are several ways that services could assure that the measures they have in place to address content risks are effective, including establishing internal functions to assess effectiveness or having third-party assurance of effectiveness. We will consider the merits of each of these in turn.

Option a) Having an internal monitoring and assurance function to provide independent assurance that measures taken to mitigate illegal harms are effective

8.101 Strengthening internal controls within an effective corporate governance framework is cited as an effective way to mitigate risk across several industries.[64] This is corroborated by best practice guidelines and controls on governance and internal assurance and audit. This includes references to monitoring and review of the effectiveness of risk controls in ISO 31000 on risk management.[65]

8.102 Ensuring that roles which provide objective assurance and advice on the adequacy and effectiveness of governance and risk management are independent is usually necessary to ensure objectivity, authority, and credibility. Independence in these functions can be established by having direct accountability between the function and the overall governing body, having unfettered access to people, resources, and data necessary to complete work, and having freedom from bias or interference in the delivery of findings on effectiveness of controls.[66] We do not envisage independence as requiring services to engage an independent third party (such as an external auditor) to confirm effectiveness of mitigations, although services may choose to do so.

[64] This consultation found overall support from respondents in favour of ensuring effective internal controls to improve the effectiveness and efficiency of corporate governance mechanisms. Notable responses to this consultation included comments from professional services organisations, which pointed to evidence that establishing and embedding a system for monitoring and reporting of internal controls improves the quality of financial reporting (PwC) and reduces the risk of corporate failure and fraud (Deloitte). We also found support for stronger internal control frameworks reflected in response to a 2022 BEIS consultation, which cites improved reporting and audit and better corporate governance as key outcomes. Sources: European Commission, 2022. Corporate reporting – improving its quality and enforcement [accessed 03 May 2023]; Department for Business, Energy, & Industrial Strategy, 2022. Restoring trust in audit and corporate governance. [accessed 18 September 2023].

[65] International Organization for Standardization (ISO),2020. ISO 31000 Risk Management. [accessed 11 September 2023].

[66] IIA, 2020.

8.103 The overall objective for independence of the monitoring and compliance function is to ensure that services find a way to achieve as much independent oversight and challenge as possible for each key task. For services where having dedicated members of staff in a monitoring and assurance function is not possible, there may be an option to structure the organisation to try to ensure that oversight of tasks within the monitoring and assurance function can be done by another individual in the firm who is not directly involved with that task.

8.104 In response to instances of serious corporate governance failures, many organisations have focused on ensuring the effectiveness of internal controls and oversight processes to improve outcomes. Remediation following corporate governance scandals has often focused on the strengthening of internal systems and processes – including assurance and compliance functions – to address areas where risk mitigation and management failed.

8.105 We found evidence that bolstering the independence of assurance functions has also been suggested as a way of ensuring that internal oversight on the effectiveness of controls is robust. This includes conclusions that failed internal assurance functions can be made more effective by ensuring that heads of function report directly to the overall governance body or 'supervisory board'.[67]

8.106 In response to our 2022 Illegal Harms Call for Evidence, some services referred to existing internal assurance or audit processes in place. This included Mindgeek[68] and Trustpilot[69] which referred to dedicated internal audit functions as part of their current Trust & Safety framework. Mindgeek[70] specified that internal audit included work related to process workflows, technical audit, and gap identification in compliance.

8.107 However, we also note evidence that questions the efficacy of internal assurance measures if risk management policies and processes are not properly implemented. The proper implementation of controls is a prerequisite for the effectiveness of risk management policies and processes, as demonstrated by case studies where scrutiny of internal controls failed despite the presence of internal assurance and audit functions.[71]

Option b) Tracking evidence of new and increasing illegal harm

8.108 Risks of illegal harm will change over time, highlighting the importance that services track evidence to identify any areas of emerging risk. User behaviour can also change over time meaning that users could become more exposed to the risk of encountering illegal content or experiencing harm. Examples may include users becoming desensitised or fatigued to warning messages.[72]

[67] Krahnen, P.K., Langenbucher, K., Leuz, C., Pelizzon, L. 2020. Wirecard Scandal: When All Lines of Defense Against Corporate Fraud Fail *Oxford Business Law Blog*, 23 November. [accessed 03 May 2023].

[68] Mindgeek's response to the 2022 Illegal Harms Call for Evidence.

[69] Trustpilot's response to the 2022 Illegal Harms Call for Evidence.

[70] Mindgeek's response to the 2022 Illegal Harms Call for Evidence.

[71] [accessed 03 May 2023]. This includes the case study of India's Yes Bank, which faced charges of money laundering. Despite referring to a clear risk management framework based on the IIA's Three Lines Model and having an internal audit department, ineffective implementation of policies in Yes Bank meant that serious financial risks were not managed or mitigated. Source: Teen, M.Y. (ed.), 2021. Yes Bank, No Governance. *Corporate Governance Case Studies* 10 accessed [03 May 2023].

[72] Burgess, M., 2018. The tyranny of GDPR popups and the websites failing to adapt, Wired, *29 August*. [accessed 6 September 2023]

8.109 Changes in the external environment may mean that kinds of illegal harm become more prominent on a service, even if a service has found low or negligible risks of the harm in their most recent risk assessment. As an example, there is evidence of foreign interference campaigns being organised around global events, including violent conflicts,[73] public health emergencies, and political processes.[74] The aim of these campaigns has been to use these events to focus influence and gain traction.

8.110 To ensure that risks are mitigated and managed consistently, services would ideally track and monitor for increases in volumes of illegal harm that Ofcom specifies in the Register of Risks, including kinds of illegal harm that have not been previously identified by a service, using relevant information. This exercise is often referred to as 'horizon scanning'[75], and in the context of online safety includes the analysis of trends in the internal and external environment in illegal harms and specific focus on sensitive events that may serve to increase illegal harms.

8.111 There will be differences in how this would apply to U2U and Search services. U2U services would need to track for new and increasing illegal harm, including evidence of their service being used to commit or facilitate an offence as well as illegal content. Search services would only be required to track for new and increasing kinds of illegal content.

8.112 We recognise that the way in which this works in practice is likely to be different for different services, depending in particular on how closely their complaints and content moderation processes use definitions which precisely track the UK's definition of illegal content and (for U2U services) the facilitation and commission of offences. A service may need to use proxies, sampling, market research among users and/or information from third parties where it does not have precise information.

8.113 Assessments of any identified trends or unusual increases of illegal content and illegal harm should be reported through governance channels. Reporting on new and escalating kinds of illegal harm is likely to be critical in achieving adequate governance oversight on risk mitigation and management. In its Good Practice Guide for Risk Reporting, the Government Finance Function emphasises that risk reporting best enhances decision-making when there are risk identification processes in place to capture new and emerging risks.[76]

8.114 To establish when they should report any identified trends or significant increases in illegal harm through governance channels, services would ideally establish a baseline understanding of how frequently illegal harm occurs on their service, based on their internal data and evidence. Where services do this, they should use this baseline to determine a threshold for unusual trends in evidence that may indicate that illegal harm is increasing. Where such increases are observed, they should be reported through governance channels.

[73]Nimmo, B., Torrey, M., 2022. Taking down coordinated inauthentic behavior from Russia and China , *Meta*. [accessed 6 September 2023]; UK Foreign Commonwealth & Development Office 2022. UK exposes sick Russian troll factory plaguing social media with Kremlin propaganda [accessed 6 September 2023]

[74] DiResta, R., Shaffer, K., Rippel, B., Sullivan, D., Matney, R., Fox, R.,; Albright, J.,; and Johnson, Ben 2019 The Tactics and Tropes of the Internet Research Agency, *New Knowledge;* Schliebs, M.,; Bailey, H., Bright, J., and Howard, P. J. 2021 GEC Special Report: The Kremlin's Chemical Weapons Disinformation Campaigns, U.S. Department of State, Global Engagement Center.

[75] Government Office for Science, 2017. The Futures Toolkit [accessed 6 September 2023]; Institute of Risk Management. An introduction to emerging risks and how to identify them [accessed 18 September 2023].

[76] Government Finance Function, 2021.

8.115 Internal controls to manage and mitigate risk need to be reviewed for effectiveness on an ongoing basis. This is to avoid the risk that measures become out of date as illegal harms change and evolve over time. For instance, illegal harm that becomes more prominent on a service due to external events that occur after a risk assessment may not be effectively mitigated and managed by measures implemented at the time of the original risk assessment. Although services will be required to keep their risk assessments up to date (such as by reviewing them on an annual basis), monitoring of real-time changes in illegal harms will be important for ensuring the ongoing effectiveness of mitigations.

8.116 To remain effective, internal controls to manage and mitigate risk must be informed by a monitoring, evaluation, and reporting process. This is achieved by establishing a process for the collection of up-to date- information about illegal harm on a service. The monitoring of trends in illegal harm needs to be followed up by analysis and interpretation of the trend data, which is then used to evaluate the effectiveness of the measures put in place to mitigate risks to users.

8.117 Several services highlighted in the 2022 Illegal Harms Call for Evidence how they track emerging issues to understand risks to users and relevant mitigations on an ongoing basis. Google specified that it has a designated intelligence team within its trust and safety function, which oversees processes designed to identify, escalate and where possible mitigate moderation issues related to user safety.[77] YouTube has a similar intelligence function which monitors for new and emerging user safety risks.[78] Across Google's organisation, there is a risk strategy programme which uses signals from news, social media, and other web sources to inform detection and mitigation of risks.[79] Meta pointed to a similar risk intelligence team which works on reviewing escalations across internal teams.[80] [**CONFIDENTIAL✂**.[81]]

8.118 We also received expressions of support for monitoring of trends in illegal content and other harms in response to our 2022 Illegal Harms Call for Evidence. Glitch supported the principle that services should be "constantly monitoring their platforms for new, emerging trends as well as those that remain pervasive and common".[82]

8.119 In some instances, it may be challenging for services to collect data that would allow them to make a judgement on how frequently illegal harm occurs on their service. This will likely vary depending on the kind of harm in question. Services should make a best judgement of how illegal harms may be increasing on their service based on relevant information and evidence, even if the evidence may be weaker for certain illegal harms.

8.120 We recognise that monitoring for indications of new and escalating kinds of illegal harm may not always lead to timely and effective mitigation and management of risk.[83] It is therefore

[77] Google's response to 2022 Illegal Harms Call for Evidence.
[78] Google's response to 2022 Illegal Harms Call for Evidence.
[79] Google's response to 2022 Illegal Harms Call for Evidence.
[80] Meta's response to 2022 Illegal Harms Call for Evidence.
[81] [**CONFIDENTIAL✂**]
[82] Glitch's response to the 2022 Illegal Harms Call for Evidence
https://www.ofcom.org.uk/__data/assets/pdf_file/0034/247759/Glitch.pdf
[83] The findings of the European Commission's 2008-2013 iNTeg-Risk programme on early recognition, monitoring and management of emerging risks found that "even when governance systems for emerging risks may seem well established upfront, it may still show to be inadequate and not adaptive enough to external context changes". Source: European Commission, iNTeg-Risk progammes [accessed 4 September 2023].

important that firms continue to assess and evaluate the effectiveness of the mitigations that they have put in place to ensure that they are adequate to address any emerging risks.

Costs and risks

Option a) Having an internal monitoring and assurance function to provide independent assurance that measures taken to mitigate illegal harms are effective

8.121 The costs of this measure would be considerable, with the main cost being the ongoing staff costs to run the monitoring and assurance function. There may also be additional costs associated with wider training and awareness raising of the remit of an internal assurance function among existing teams who would be expected to feed into the work of the function.

8.122 In considering how many staff services might typically need for their internal monitoring and assurance function, one reference point is the number of staff employed in internal audit functions. Internal audit functions evaluate an organisation's internal controls, especially their corporate governance and accounting processes. They also often involve considering an organisation's risk management processes and can involve looking at specific areas of a business, such as cybersecurity. Most organisations have fewer than five people working in their internal audit functions, though very large organisations can have hundreds of members of staff. Internal audit functions tend to be lower in non-profit organizations and privately held companies, and bigger in public sector, publicly traded, and financial services.[84] Smaller services are less likely to have an internal audit team, although they may have reporting and review function within operational teams.

8.123 The size of the internal monitoring and assurance function needed would vary by service. If a larger, riskier, more complex service needed to have ten additional people, then the costs might be £500,000 to £1,000,000 per annum. In contrast, for a smaller service that has a limited range of online safety risks and measures, it might be sufficient for it to have a single person in its monitoring and assurance function, and the annual costs might be £50,000 to £100,000.[85] Because the costs would tend to be higher for larger services that face more risks, we would expect the costs to scale with the potential benefits to some extent. We expect that the costs of this measure for smaller services would tend to represent a higher proportion of their annual revenue.

8.124 As demonstrated by responses to our 2022 Illegal Harms Call for Evidence, some large services already have internal assurance processes in place that deal with risks related to

[84] The 2022 Internal Audit: A Global View found that 51% of audit functions had 5 or fewer people. At the other extreme, 10% had 51 of more staff. This was based on 3,600 responses to the global survey. Another study, the 2019 Internal Audit Survey Insurance by PwC found that 48% of internal audit functions had 0-10 members, but this was based on responses by only 25 organisations. Source: PWC, 2019. Internal Audit Survey Insurance and Asset Management [accessed 18 September 2023]; Internal Audit Foundation, 2022. 2022 Premier Global Research. [accessed 18 September 2023].

[85] This is based on our assumptions set out in Annex 14. We use the salaries for the ‘professional occupations’ for the staff of the internal monitoring and assurance. We recognise that salaries will vary very considerably both between different organisations and also within the internal monitoring and assurance function at any organisation. This is the case for salaries within internal audit function, as shown by the benchmarking of different internal audit roles in [accessed 27 September 2023].

online safety.[86] Provided these processes are sufficient and services retain them for online safety purposes, they would not incur any additional costs from this measure.

Option b) Tracking evidence of new and increasing illegal harm

8.125 This measure would result in significant costs for services, including both one-off and on-going costs. The scale of these cannot be appropriately estimated, as it is likely to vary considerably from service to service depending on the kinds of illegal harm they face, how they are able to gather information and how much content they have.

8.126 All services will be required under the online safety regime to establish complaints processes. They will therefore have information from those processes, but they may choose to run them in a way that does not distinguish between illegal content and content that is violative of their terms of service. Services may need to rely on other sources of information, including but not limited to outcomes of content moderation processes, referrals from law enforcement or flags from expert groups to have a sufficient understanding of trends in illegal content risks on their service.

8.127 There will also be costs associated with monitoring, reporting and analysing the evidence of new and increasing kinds of illegal content, including both one off costs for establishing processes or automated collection systems, and ongoing costs related to staff to run these systems. We anticipate that services could collect and report this information on a regular schedule in line with other governance reporting mechanisms (for example, monthly updates to the online safety compliance function or equivalent body).

8.128 Although many larger services will also already have teams or individuals in place who are tasked with analysis of data for general online safety purposes, this may not extend to the analysis of trends in illegal harm specifically. We anticipate that most services will therefore face additional costs related to this measure, even if they currently monitor trends relevant to online safety and would have chosen to continue doing that in the future.

8.129 Smaller services are far less likely to have existing teams or systems in place for the ongoing analysis of information related to illegal harm. They may face challenges in accessing and analysing information in a systematic way if, for example, they outsource their content moderation operations to a third-party or they do not have appropriate data collection infrastructure or in-house expertise in data analysis.

8.130 As with the other governance measures, there could also be indirect costs as a result of this measure, but we have not considered these here for the reasons discussed at paragraph 8.35 above.

Other options considered

8.131 We also considered several other options regarding how services can assure their measures to mitigate and manage illegal content are effective. These could be alternatives to the options discussed above or could supplement them. However, we do not consider there is currently enough information on the effectiveness of other possible measures to be able to recommend them in Codes at this stage.

8.132 On external audits, we are aware that some services have undertaken some form of external audit for aspects of online safety. While we welcome this, we do not have clear evidence

[86] For example, Meta indicated that their audit arrangements already cover how the service can be used to facilitate harm or undermine public safety or the public interest. Source: Meta's response to 2022 Illegal Harms Call for Evidence

about the effectiveness of external audit for other services and how external audit might best be used alongside other governance measures. Further evidence about costs associated with this measure would also be needed, including costs related to the procurement of third-party audit services.

8.133 We also considered the option of requiring services to undertake due diligence on third-party providers of online safety services to assure that their approaches to mitigating and managing risks to users are effective. Further evidence is required to establish how due diligence would ensure that measures are sufficiently robust to protect users from illegal content, given how the landscape is currently in development.[87]

8.134 We do not currently have clear evidence to specify metrics that services should collect to measure the effectiveness of measures to mitigate and manage illegal content risks. This is partly because of the difficulties of specifying suitable metrics for the wide range of services in scope of the Codes, and lack of consensus on which metrics are the most accurate indicators of efficacy. We also note a risk of negative unintended consequences if metrics are not well specified.

8.135 As we do not intend to make recommendations in these areas, we have not made a full assessment the costs of these other options below. However, we note that costs related to these measures, including notably the option of requiring external audit, may be significant.

Provisional conclusion

Option a) Having an internal monitoring and assurance function to provide independent assurance that measures taken to mitigate illegal harms are effective

8.136 We propose to recommend that large multi-risk services should have an internal monitoring and assurance function to provide independent assurance that measures taken to mitigate illegal harms are effective on an on-going basis, reporting to an overall governance body or audit committee.

8.137 Our analysis has shown that there are benefits in ensuring that organisations have independent oversight over internal controls to ensure that governance and risk management are effective. This oversight helps organisations make objective, authoritative and credible judgments of the efficacy of their approach to risk management. As such, we consider this measure would deliver particular benefits in ensuring that risks of all illegal harm are mitigated and manage appropriately.

8.138 We have identified considerable ongoing costs associated with this measure. However, we consider that this measure is sufficiently fundamental to good risk management that it would be proportionate to recommend it for large services with multi-risk. We consider that this measure is likely to have greater benefit for services which identify greater risks of harm, given that robust assurance processes would increase oversight over risk management processes. We do not consider that the benefits would be as great for services which do not identify multiple risks of illegal harm, including large services.

[87] In discussion of the external audit market as related to algorithm auditing, the DRCF concludes that “where a market for algorithm auditing services exists, it is at an early stage of development. Efforts to proactively surface and identify new issues in algorithmic systems through auditing have been particularly slow to emerge.” Source: Digital Regulation Cooperation Forum, 2022. Auditing algorithms: the existing landscape, role of regulators and future outlook. [accessed 11 May 2023].

8.139 On balance, we therefore provisionally consider it proportionate to recommend establishing an internal monitoring and assurance function only to large services which are multi-risk. For the avoidance of doubt, this includes both large general search services which are multi-risk and large vertical search services which are multi-risk.

Option b) Tracking evidence of new and increasing illegal harm

8.140 We also propose to recommend that large services (with the exception of large vertical search services) and services that are multi-risk should track evidence of new kinds of illegal content on the service, and unusual increases in particular kinds of illegal content, or (for U2U services) equivalent changes in the use of a service for the commission or facilitation of priority offences. Relevant evidence may include, but is not limited to, that derived from: complaints processes; content moderation processes; referrals from law enforcement; and (for U2U services) information from trusted flaggers and any other expert group or body the service considers appropriate.

8.141 We propose to recommend that these services should regularly report any new kinds of illegal content or illegal content proxy, or (for U2U services) equivalent changes in the use of a service for the commission or facilitation of priority offences through relevant governance channels to the most senior governance body.

8.142 To understand this, the provider should establish a baseline understanding of how frequently particular kinds of illegal content, illegal content proxy, or the commission of facilitation of priority offences occurs on the service to the extent possible based on its internal data and evidence. The service provider should use this baseline to identify unusually high spikes in the relevant data.

8.143 Our analysis shows clear evidence that monitoring and assurance and tracking new kinds of illegal harm are both important components of good governance and risk management.

8.144 The benefits (in terms of reducing all kinds of illegal harm) are likely to be greatest for large services and services which identify considerable risks to users, given that these services are likely to require additional layers of monitoring and evaluation to ensure effective risk management. The evidence we've considered also suggests that services which don't put these steps in place will be less likely to monitor and respond to risks of all kinds of illegal harm effectively.

8.145 Additionally, given that illegal harm is highly likely to change over time, monitoring how harms are manifesting on services will be necessary to ensure that existing mitigations are effective and adequate to prevent all kinds of illegal harm.

8.146 As with other governance measures, we propose not to recommend this measure to large vertical search services just because they are large. By their nature vertical search services are unlikely to have content that is as rapidly changing as U2U services and search results are more under a service's control than for U2U content. We are also not aware of evidence of such services showing illegal content.[88] Any benefits of applying this measure would therefore be low for vertical search services, and as such would likely be disproportionate (except for vertical search services which are assessed as being multi-risk).

8.147 We have identified ongoing costs associated with these recommendations. We anticipate that these costs are likely to scale with service size, whereby larger services will likely face

[88] See [Register of risks, paragraphs 6.21(b)] for why we consider vertical search services to be low risk.

higher costs related to implementation. However, we recognise that these costs are likely to be a larger proportion of revenue for smaller services.

8.148 Given the important role these measures would play in helping services manage risk we consider the benefits that would be conferred by applying them to large services (with the exception of large vertical search services) and multi-risk services (including vertical search services which are multi-risk) justify the costs to which the measure would give rise. We therefore provisionally consider it proportionate to recommend these measures to these services. Given the significant benefits that increased assurance and oversight has on effective risk management and mitigation, we also provisionally consider this to be proportionate for small but risky services. The heightened risks to users will require these services to have processes to monitor and track on harms and effectiveness of measures to mitigate risks.

8.149 We are not proposing to recommend the measures listed above to smaller services who have identified low or no risks of illegal harm in their risk assessments. Many services in this position will be able to effectively manage and mitigate risks and to assure their measures are effective without incurring the costs of an assurance function that is independent from operational or business functions. Although all services are likely to experience changes in how their service is used over time, in our view the costs associated with establishing a system to track signals and to report them through governance channels are likely to be overly burdensome for smaller services that are low risk. We consider that they should be able to achieve good outcomes for users through the process of keeping their risk assessments up to date.

8.150 The evidence we have cited above relating to the efficacy of these proposed measures relates to priority offences including CSEA, terrorism and relevant non-priority offences. We therefore propose to include them in our Codes for U2U and search services on terrorism, CSEA and other duties.

Staff incentives, policies, and processes

Harms that this measure seeks to address

8.151 How staff are incentivised and trained in their roles can inform how they approach user safety considerations within a service. This includes setting policies and processes that inform how staff are rewarded and renumerated, how staff are informed of expectations of their organisation, and how staff are appropriately guided and instructed in achieving these expectations, including around compliance requirements.

8.152 Failing to effectively communicate or train staff on a service's approach to compliance with illegal content safety duties raises the likelihood that risk mitigation and management is not embedded in the day-to-day work of operating a service. Where staff do not understand or are not trained on a service's approach to addressing illegal content risks, risk mitigation measures could be applied poorly.

8.153 A failure to embed a culture of risk management across a service may result in the inconsistent application of measures designed to mitigate and manage risks of all kinds of illegal harm.

8.154 Alignment of objectives at all levels will be important to achieve good safety outcomes for users. This includes having a common understanding and expectation around risk management in relation to illegal content on a service. There is a possibility that without these efforts to align, staff in different areas of a service will not understand how a service is approaching regulatory compliance, or how it manages and mitigates risks of illegal content to users.

8.155 This is supported by evidence of how the absence of compliance training programmes has contributed to serious corporate scandals.[89]

Options

8.156 Good practice in risk management and governance suggests that communications and training are important tools for promoting a culture of risk awareness and compliance.[90] We consider that positive online safety outcomes and fulfilment of the safety duty requires regulated services to put in place processes for guiding and incentivising their staff to manage and mitigate identified illegal content risks appropriately.

8.157 In this area, we have considered the following options:

a) Having a Code of Conduct or principles for all staff that emphasise the importance of protecting users from illegal content risks;
b) Ensuring that staff involved in the design and operation of a service are trained in a service's approach to compliance with online safety duties sufficiently to give effect to them; and
c) Tying remuneration for senior managers to positive online safety outcomes.

Effectiveness

8.158 How a service guides, incentivises and rewards its staff is relevant to both risk management and regulatory compliance, given the influence these factors have on individual performance, decision-making and risk-taking behaviour.

8.159 Ensuring that there are principles for all staff that emphasise the importance of protecting users from illegal content risks makes it more likely that opportunities to mitigate those risks will be identified, considered, and adopted.

8.160 In contrast, where staff are not guided or incentivised to ensure that activities to manage and mitigate risks identified in a risk assessment are effective and appropriate, there is a possibility that risk to users will not be appropriately factored into everyday decision-making and operations either because of competing pressures and incentives on staff, or due to ignorance of compliance requirements. It follows that measures may be either inadequate to address risks to users, or improperly implemented. It also could be the case that any measures put in place are not monitored and evaluated for effectiveness over time.

Option a) Code of Conduct regarding protection of users from illegal harm

8.161 A Code of Conduct can be an effective way to communicate expected behaviours of all staff by an organisation. For the purposes of online safety, a Code of Conduct could include recognition of the potential risks of illegal harm to users of a service, a clear organisational

[89] In the case of Siemens, which in 2008 was subject to regulatory investigations for bribery, the failure to embed a programme of compliance and Code of conduct for staff has been cited as playing a "decisive role" in the scandal. Source: Primbs, M., and Wang, C., 2016. Notable Governance Failures: Enron, Siemens and Beyond *Comparative Corporate Governance and Financial Regulation.* [accessed 18 September 2023].

[90] COSO, November 2020.

statement around protecting users from illegal harm, and expectations and guidelines for all staff in reporting instances of concern relating to illegal content on the service. Whatever the specific content, effective Codes of Conduct should be simple, concise, and readily understood by all employees, consistent with other policies and communications, and reviewed by multi-disciplinary teams.[91]

8.162 Responses to our 2022 Illegal Harms Call for Evidence from services highlighted documentation of values and behaviours expected of staff as part of a broader programme of good corporate governance regarding online safety. This included Google, which mentioned consistent principles, a Code of Conduct and Group guiding principles as part of governance and accountability.[92] Zoom[93] framed this in terms of its standard operating procedures (SOPs) which govern expectations for analysts dealing directly with content decisions. [**CONFIDENTIAL✂**][94] highlighted that it had clear policies and operational guidelines for how it governs its approach to user and platform safety and an expectation that responsibility for safe user experiences is shared across the organisation as a core value of the brand.

8.163 Ensuring that board members, senior management and staff understand and commit to organisational priorities is cited in good practice literature and guidance on corporate governance. As a general principle, this is underpinned by the idea that having a shared understanding of organisational values, reward mechanisms and expected behaviours is important in achieving commercial objectives. This extends to objectives regarding risk management, safety and running responsible operations.[95]

8.164 Codes of Conduct have been used to achieve compliance aims in other regulatory regimes. The FCA requires regulated firms to have Codes of Conduct for staff under its SM&CR scheme, which are in line with firms' duties to comply with financial regulations. Other jurisdictions highlight that Codes of Conduct are important as an expression of organisational efforts to link risk management and compliance in day-to-day operations.[96]

Option b) Compliance training for staff involved in the design and operation of a service

8.165 Training staff is an important way that a service can communicate compliance requirements and embed risk mitigation and risk management into company operation. Staff with roles and responsibilities for the design and operation of a service are likely to benefit most from training focused on compliance with illegal content safety duties, given the potential impact their work has on ensuring user safety.

[91] Deloitte. Suggested guidelines for writing a code of ethics/conduct. [accessed 4 September 2023].

[92] Google's response to the 2022 Illegal Harms Call For Evidence

[93] Zoom's response to the 2022 Illegal Harms Call for Evidence.

[94] [**CONFIDENTIAL✂**].

[95]FRC, 2014. Guidance on Risk Management, Internal Control and Related Financial and Business Reporting. [accessed 4 September 2023].

[96] In its evaluation of Corporate Compliance Programs, the US Department of Justice highlights that "any well-designed compliance program entails policies and procedures that give both content and effect to ethical norms and that address and aim to reduce risks identified by the company as part of its risk assessment process". It advises prosecutors that a Code of Conduct that is accessible and applicable to all staff is important in this regard, as an expression of an organisations efforts to link risk management and compliance to its day-to-day operations. Source: U.S Department of Justice, 2023. Evaluation of Corporate Compliance Programs [accessed 05 May 2023].

8.166 The outcome of an effective compliance training programme for online safety would be that staff have good understanding of both the general regulatory requirements and how the service is managing and mitigating risks to users. Staff should be trained sufficiently in both these areas and in illegal content safety duties to give effect to them in their roles.

8.167 Services referenced the general importance of staff training in the 2022 Illegal Harms Call for Evidence, and demonstrated their commitment to ensuring that their approach to online safety was understood by its employees. This included Google[97], which gives its employees specific training on "risk and compliance to raise awareness of requirements from new and emerging regulations which govern online content and behaviours" and Dropbox[98] which stated that engineers receive training to ensure they are aware of and accounting for safety concerns while software is being developed.

8.168 Evidence of remediation efforts in organisations which have experienced significant governance failings point to the importance of compliance training. A case study from Siemens relating to redress of governance failings focusses on strengthening compliance programmes through improved staff training.[99] This case study also provides supporting evidence to suggest that such changes lead to improved perceptions in how risks are managed.[100]

8.169 Regular risk culture training is also supported in good practice risk management guidance and industry frameworks, as a way of helping clarifying roles and responsibility and to ensure that management fully understand and appreciate the need to foster a healthy risk culture.[101]

8.170 Compliance training programmes should be supported by broader efforts on the part of a service to embed risk management awareness across the entire organisation. Services should frame staff training in this area as an important step in establishing a risk aware culture, and in supporting the effective management and mitigation of identified online safety risks.

Option c) Tying remuneration for senior managers to positive online safety outcomes.

8.171 We have also considered the option of recommending a measure that would require services to make remuneration for senior managers contingent on demonstrable efforts to mitigate and manage online safety risk, for example by requiring senior manager KPIs or objectives for online safety risk management to be tied to bonuses or other incentives. As part of its guidance on staff incentives, remuneration and performance management, the FCA highlights that how staff are rewarded and managed can have a major influence on behaviour which may translate to increased risks to customers.[102] The guidance stresses that

[97] Google's response to the 2022 Illegal Harms Call for Evidence.

[98] Dropbox's response to the 2022 Illegal Harms Call for Evidence.

[99] Following a 2008 bribery scandal, Siemens attempted to redress governance failings identified by strengthening its compliance programmes. This included ensuring that employees "in different levels have been provided with trainings specific to their roles and responsibilities". Source: OECD, 2010. Compliance Program@Siemens [accessed 18 September 2023].

[100] Siemens' response to the scandal has been "widely praised by many anti-corruption and ethics experts", demonstrating the value of compliance training as a core pillar of effective governance. Source: Dietz, G., Gillespie, N., 2012. Rebuilding trust: How Siemens atoned for its sins, *The Guardian,* 26 March [accessed 05 May 2023].

[101] Milliman, 2023; IIA, 2020.

[102] FCA, 2015. Remuneration. [accessed 4 September 2023].

firms need to carefully assess how they are incentivising staff to ensure that they take "reasonable care to organise and control their affairs responsibly and effectively".[103]

8.172 However, we do not currently have evidence to conclude that option is likely to be effective in achieving the safety duty. Although remuneration has been demonstrated to be an important factor in determining behaviour and better risk management in financial services, the evidence is less strong for ensuring good online safety outcomes. Such remuneration policies could be challenging to apply to services with different remuneration and bonus structures. Such policies would also need to be grounded in evidence regarding appropriate KPIs or metrics to determine how individual behaviour supports positive online safety outcomes. There could be unintended effects of these policies, including greater risk-aversion on the part of senior members of staff in areas not linked to online safety.

8.173 We are not proposing to include a measure regarding remuneration in this version of the Codes of Practice, but are seeking stakeholder views and feedback on the efficacy, costs or risks regarding this issue.

Costs and risks

Option a) Code of Conduct regarding protection of users from illegal harm

8.174 Services that do not have Codes of Conduct or principles for staff would need to develop them, and those that already have them would need to review and modify them if necessary. As one input to this, services could draw on their risk assessment, which services will need to do anyway under the OSA, and the areas of risk that this assessment identifies.

8.175 We anticipate this would include some time for senior members of staff to review and comment on the Code of Conduct. We envisage it being easier to develop for smaller services that tend to have less complex businesses. For most services, we envisage this cost being less than £10,000 initially.[104] There may also be some costs of reviewing and maintaining this over time, but we envisage this cost being much smaller.

8.176 The Code of Conduct or principles would then need to be sent to all staff who would be expected to read it. We assume that the document would be short and would not take a significant amount of time for staff to read and understand.

Option b) Compliance training for staff involved in the design and operation of a service

8.177 We assume that the total cost per person trained would be £2,000 to £4,000.[105] An important variable would be the numbers of staff needing to be trained. This is likely to vary

[103] FCA, 2018. FG18/2 Staff incentives, remuneration and performance management in consumer credit. [accessed 4 September 2023].

[104] This is using our cost assumptions set out in Annex 14 and assuming it takes less than 20 days to produce the Code of Practice.

[105] This is based on the assumptions in Annex 14 and assuming the training lasts a week. We also assume that the wage cost of the people being trained represents only half of the total costs of the training. Other costs included preparing the training materials, running the training and any related travel to the training. This is consistent with the Department for Education saying that the wage cost of staff being trained accounted for about half of all training expenditure in 2019, although this varies by size of the firm and sector. We assume this excludes the 22% uplift that we have elsewhere assumed for non-wage labour costs, so we have not also increased these wages by 22%. Source: Department for Education (DfE), 2020. *Employer Skills Survey 2019: Training and Workforce Development – research report*, pp. 38 and 40. [accessed 18 September 2023].

significantly between services. Larger services and those which more frequently change their design and operation would probably need to train many more people than smaller services.

8.178 If we assume that a service has 100 people involved in the design and operation of a service that it needs to train, then the cost of the initial training would be £200,000 to £400,000 and if a service only needs to train three people, then the cost of the initial training might be £6,000 to £12,000. We would generally expect the number of people that need to be trained to vary with the size of the service, meaning a smaller service would have lower costs. We also envisage that training would also be needed in subsequent years, including allowing for staff turnover, but that this would not be required on an annual basis. If we assume the training were done every two years, the average ongoing cost would be half the numbers above.

8.179 Consistent with responses to our 2022 Illegal Harms Call for Evidence, some companies are already training their staff on compliance with online safety responsibilities. Provided this training was adequate and these services wanted to continue with it in the future, this measure would not add additional costs for them compared to what they are currently incurring.

Provisional conclusion

8.180 We propose to recommend a measure in our Codes of Practice that large services (with the exception of large vertical search services) and services that are multi-risk (including vertical search services which are multi-risk) should have a Code of Conduct that sets standards and expectations for employees around protecting users from risks of illegal harm.

8.181 Additionally, we are proposing to recommend that the same set of services should ensure that staff involved in the design and operational management of the service are trained in a service's approach to compliance with the illegal content safety duty and the reporting and complaints duties, sufficiently to give effect to it.

8.182 Our analysis shows evidence that staff policies and processes, including having Codes of Conduct and targeted training programmes, are effective in ensuring that services communicate compliance requirements, and embed risk management and mitigation within organisational culture. This in turn is important for setting organisational direction regarding risk management, and supporting a culture of risk awareness among staff.

8.183 Large services with more complex operations and larger headcounts are likely to benefit from these measures, given that they provide a way to streamline expectations regarding risk management to all relevant staff.

8.184 As with some of the other governance measures, we propose not to recommend this measure to large vertical search services just because they are large. By their nature vertical search services are unlikely to have content that is as rapidly changing as U2U services and search results are more under a service's control than for U2U content. We are also not aware of evidence of such services showing illegal content.[106] Any benefits of applying this measure would therefore be low for vertical search services, and as such would likely be disproportionate (except for vertical search services which are assessed as being multi-risk).

[106] See [Register of risks, paragraphs 6.21(b)] for why we consider vertical search services to be low risk.

8.185 Services which identify considerable risk to users will require more checks and balances in place to ensure that they are effectively managing and mitigating identified risks. This includes aligning their staff policies with their approach to risk management, to ensure that employees across an organisation are aware of a service's duty to manage illegal content effectively on an on-going basis. These services will also likely benefit from communicating expectations around the importance of managing these risks to all staff. For the reasons set out in Chapter 11[107], we propose to apply these measures to multi-risk services.

8.186 We consider that costs related to the measure regarding Codes of Conduct are likely to be small for all services, and largely related to one-off set up costs. We consider that costs for the measure regarding staff compliance training are likely to vary significantly depending on the size of a service, and how many employees a service has involved in the design and operation of the service. As such, we expect these costs to scale with the size of the service.

8.187 On balance, having had regard to both the benefits and the costs they would result in, we consider it proportionate to recommend these measures for large services (with the exception of large vertical search services) and services that are multi-risk (including vertical search services which are multi-risk) , given the benefits that staff policies and processes which emphasize the importance of risk management and mitigation have on ensuring a consistent approach to organisational aims across a service. This is particularly true for large and complex organisations, where there is particular benefit in having clear expectations around risk management. Small services which find risks to users would also benefit from having a consistent and well-defined approach to risk management as expressed in a Code of Conduct or staff compliance training and are likely to face relatively lower costs of implementation.

8.188 By contrast, the need for low-risk services to have an aligned understanding of risk management and mitigation among all staff will be reduced, given that fewer of their activities will revolve around active risk management and mitigation. The benefits of implementing these measures are therefore unlikely to justify the costs of implementing them. For these reasons, we do not consider it proportionate to apply these measures to services which identify low risks to users.

8.189 The evidence we have cited above relating to the efficacy of this proposed measure relates to priority offences including CSEA, terrorism and relevant non-priority offences. We therefore propose to include this measure in our Codes for U2U and search services on terrorism, CSEA and other duties.

[107] See paragraphs 11.43-46 for further detail.

9. Service Risk Assessment Guidance

What is this chapter about?

This chapter covers our guidance about how services can fulfil their duties to assess risks (the 'Risk Assessment Guidance'), including our proposals for the process services should follow when doing their risk assessment and the types of evidence they should consider.

What are we proposing?

We are making the following proposals for all U2U and search services:

- **We will guide services to follow a four-step risk assessment process as the best way to ensure that their assessments are 'suitable and sufficient'.** These four steps are: (i) understand the harms that need to be assessed; (ii) assess risks by considering the likelihood and potential impact of harms occurring on their service; (iii) implement safety measures and record outcomes of the risk assessment; and (iv) report, review and update the risk assessment.
- **We will provide tables listing risk factors, which set out an explanation of what harms these risk factors are associated with and how these increase risks of harm.** We call these 'Risk Profiles'. Services should consult these tables when doing their risk assessment. The information in risk profiles is extracted from our assessment of the causes and impact of harms (see above).
- **We will guide all services to consider the following evidence when doing their risk assessment:** Risk Profiles (and relevant parts of Ofcom's Register of Risks), user reports, user complaints, user data including age (where relevant), retrospective analysis of incidents of harm and other relevant information that a service holds.
- **Where this evidence does not provide services with a sufficiently good understanding of their risk levels, Ofcom will recommend services look at some or all of the following pieces of additional evidence**: results of product testing, results of content moderation systems, consultation with internal experts on risks and safety measures, views of independent experts, internal and external commissioned research, outcomes of external audit or other risk assurance processes, consultation with users and user research, and engagement with relevant representative groups.
- **We will recommend services have a written policy in place to review their assessment at least every 12 months**, and to name a responsible person for overseeing this process (this links to the governance measures in Ofcom's Code of Practice).
- **We will recommend that services update their risk assessment whenever a 'significant change' to their service occurs and will provide general principles on how services should interpret what constitutes a significant change**. These principles will recognise the importance of the size of a service when considering if a proposed change may be 'significant'.

Why are we proposing this?

This approach reflects our understanding of best practice and current standards in risk management, and mirrors risk assessment processes that have been successfully implemented in other sectors. As explained above, we consider good risk assessment and management will make a material contribution to reducing online harm and that the costs of identifying and managing risks upfront will often be lower than the costs of remedying online harm after the fact. This approach is likely to be complementary to any risk management system that services already have in place, which will reduce the costs of our proposals and ensure they are proportionate.

What input do we want from stakeholders?

- Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views.

Specifically, we would also appreciate evidence from **regulated services** on the following:

- Do you think the four-step risk assessment process and the Risk Profiles are useful models to help services navigate and comply with their wider obligations under the Act?
- Are the Risk Profiles sufficiently clear and do you think the information provided on risk factors will help you understand the risks on your service?[108]

Introduction

9.1 The illegal content risk assessment duty is one of the broadest obligations in the Act. All user-to-user (U2U) and search services that fall within scope of Part 3 of the Act must complete an illegal content risk assessment.

9.2 Ofcom has a duty to produce guidance to assist services in complying with this duty. This chapter explains our proposed approach to the illegal content Risk Assessment Guidance, the key elements we are highlighting for consultation, and how the draft guidance will help services to protect users online. Our draft guidance is available to review at Annex 5.

9.3 The illegal content risk assessment duties include a range of different elements. U2U services must assess the risk of users encountering priority illegal content or other illegal content by means of the service, and the level of risk that the service may be used for the commission or facilitation of a priority offence. They must also assess the nature and severity of the harm which may be suffered as a result.

9.4 As part of the assessment, services must consider various characteristics of the service specified in the legislation – such as its user base, functionalities, business model, and systems and processes – and also take account of the relevant risk profile(s) produced by Ofcom.

9.5 Search services must complete a similar risk assessment, considering the risk of users encountering priority illegal content or other illegal content by means of the service. Likewise, they must assess the nature and severity of the harm which may be suffered. They must evaluate similar characteristics specified in the legislation (though not the user base) and also take account of the relevant risk profile(s) produced by Ofcom.

[108] If you have comments or input related the links between different kinds of illegal harm and risk factors, please refer to Volume 2: Chapter 5 Summary of the causes and impacts of online harm).

9.6 All services need to ensure their assessment is "suitable and sufficient", and they must take "appropriate steps" to keep it "up to date".

9.7 We consider risk assessments to be a critically important part of the online safety regime. The adoption of good practice in risk assessment is not only a legal obligation for services, but a key component of delivering the wider industry and culture change that will put safety at the heart of services' design and decision making. As the nature of harm and the way it takes place online will continually evolve, robust risk management processes are important to ensure services can respond more quickly and effectively, and to consider how changes to their service could impact the public.

9.8 Ofcom has made the adoption of good governance and risk management practices by online services a strategic aim in our approach to regulation. Our goal is that services prioritise assessing the risk of harm to users (especially children) and run their operations with user safety in mind. This means putting in place the insight, processes, governance and culture to put online safety at the heart of product and engineering decisions.

9.9 Overall, the purpose of the risk assessment is to ensure services have an adequate understanding of the risks that arise from their service, so that they can take suitable measures to manage and mitigate those risks (as required by the illegal content safety duty). The assessment is a crucial step in enabling services to identify potential harms that may arise on their service, to understand how its design and operation may give rise to particular risks, and to take meaningful action to protect the public.

9.10 The illegal content risk assessment duties also connect to a number of other duties under the Act and we set out where this is the case in the draft guidance.

9.11 The Risk Assessment Guidance does not represent a set of compulsory steps that services must take, but rather is intended to assist services in fulfilling their legal obligations. We consider that following our proposed guidance will put services in a stronger position to comply with their duties.

9.12 This chapter includes an explanation of:

a) our **approach to the guidance**, including our policy objectives and the research we have done;

b) the proposed **guidance summary**, including our guidance on carrying out a suitable and sufficient risk assessment (from 9.13);

c) the proposed **risk assessment methodology** relevant for all services (from 9.21);

d) the proposed **approach to Risk Profiles**, which are a key element of the risk assessment process and a key feature of the methodology (from 9.72);

e) the **evidence base** that different kinds of services will need to consider (from 9.93);

f) and our proposed guidance on how to keep a risk assessment **up to date** and identify any **significant change** requiring a new risk assessment (from 9.123).

Ofcom's approach to guidance

9.13 The draft guidance covers the illegal content risk assessment duties for U2U services, set out in section 9 of the Act, and for search services, as set out in section 26. A full description of these provisions can be found in Annex 12.

9.14 Ofcom is required to produce guidance for both U2U and search services. While there are different risk assessment duties for these two types of service, there are significant similarities between the requirements. The key difference is that some elements of the U2U risk assessment duty do not apply to search services. As such, our proposal is to produce a single set of Risk Assessment Guidance, emphasising where different elements apply and highlighting what different services need to do.

9.15 In addition to carrying out an illegal content risk assessment, some services will also have duties to carry out a children's risk assessment. We plan to consult on the duties relating to children separately in Spring 2024. These two sets of duties have substantive similarities, and we will ensure there is a coherent approach across the different risk assessments.

Policy objectives for the guidance

9.16 In preparing the draft Service Risk Assessment Guidance, we have focused on several key objectives:

a) Help services comply with their illegal content risk assessment duties, through clear, targeted recommended actions;

b) Ensure that services' risk assessments are effective in identifying and understanding risks, by drawing on best practice in risk management;

c) Prepare services to respond to those risks, which they need to do under the safety duties;

d) Ensure that the risk assessment duties can be implemented in a proportionate way and do not place an undue burden on services; and

e) Use the risk assessment process to create a clearer route to compliance across the regime, by integrating other resources produced by Ofcom into the guidance including the Register of Risks, Risk Profiles, Codes of Practice and record keeping guidance.

9.17 This chapter explains how we seek to achieve these objectives through the draft guidance.

Our approach to developing the guidance

9.18 To develop the guidance in a way that achieves these objectives, we have carried out:

a) Research into current and best practice, including:

i) Commissioning **research into best practice** risk assessment and risk management techniques in the context of online safety (a bespoke report by Milliman[109]), and reviewing existing literature on best practice;

ii) Gathering **evidence from industry**, including through Ofcom's online safety call for evidence, our experience of implementing the Video-Sharing Platform regulations, and engaging directly with services and industry bodies on risk assessments;

iii) Considering relevant research from Ofcom's **wider online safety research** programme, including industry studies and published transparency reports;

[109] Milliman, 2023. Report on principles-based best practices for online safety Governance and Risk Management. This report was commissioned by Ofcom. Subsequent references are to this document throughout.

b) **Impact assessments** to help us select a proposed approach that meets our objective in the least onerous way for services (summarised in this chapter);

c) Planning and coordination to integrate other requirements of the Act and Ofcom guidance into the risk assessment and risk management process (i.e., as part of the proposed methodology), providing a **clearer journey to compliance for services**.

9.19 Overall, we have rooted our proposed approach to the Service Risk Assessment Guidance in the evidence of best practice and current standards in risk management. We believe that this will help us fulfil our policy objectives and result in higher quality risk assessments. It is also an approach that has been implemented successfully in a range of sectors and which is likely to be complementary to any risk management systems that services already use. We also hope this will improve confidence among services undertaking a risk assessment for the first time.

Proposed guidance structure

9.20 We propose to structure the guidance in the form of a summary, followed by three sections of detailed guidance. The draft comprises:

a) A **guidance summary**, which sets out services' legal obligations; what a risk assessment must include; when and how it should be carried out; and what services need to do to ensure their risk assessment is **suitable and sufficient;**

b) Our proposed **methodology** for risk assessments, defining a **four-step process** which all services can follow to meet the duties and setting out how services should use Ofcom's **Risk Profiles,** which provide a short, accessible summary of the factors we consider are associated with a heightened risk of illegal harms;

c) Guidance on what **evidence** services should use to inform their risk assessment; and

d) Guidance on how to ensure that the risk assessment remains **up to date** and on the triggers to review or carry out a new risk assessment, including when services plan to make a **significant change.**

Our proposed approach to the guidance summary and suitable and sufficient risk assessments

9.21 The draft guidance summary includes:

a) An overview of the key illegal content risk assessment requirements, with emphasis on what the assessment must include (including all the elements set out in sub-sections 9(5) and 26(5));

b) A summary of when risk assessments need to be carried out (including reviews and new risk assessments); what risks services need to assess; how they should assess each risk; what methodology they should adopt (referring to the later detailed guidance); and what happens if a service fails to complete a suitable and sufficient risk assessment (referring to our enforcement guidance); and

c) Guidance on how services can ensure that their risk assessment is suitable and sufficient.

Suitable and sufficient illegal content risk assessments

9.22 Services' risk assessments must be "suitable and sufficient". The Act does not provide a specific definition for suitable and sufficient, so this is a matter for each service to consider in the context of its obligations under the regime as a whole. However, we consider this to be an important requirement which has two main components:

a) Services must ensure they complete all the relevant elements of a risk assessment specified in the Act; and

b) Services must carry out each of these individual elements to a standard that is suitable and sufficient for their service in the context of its obligations under the regime as a whole.

9.23 To address the first component, the draft guidance summary sets out the required elements of a risk assessment (largely captured under Sections 9(5) and 26(5)), and our proposed methodology includes practical steps that services can take to implement these requirements (see the next section on methodology).

9.24 To address the second component, we propose to take the following approach to the suitable and sufficient standard in our guidance:

a) Given the range of services within scope of the Act – with vastly different user base sizes, resources and risk levels – we emphasise that this is a **context specific requirement**. There is no one-size-fits-all approach; what may be suitable and sufficient for one service, may not be for another. This means that services will need to determine for themselves what approach they need to take. By the same token, Ofcom's approach to enforcing this standard must also be service-specific; our guidance cannot be exhaustive and we will need to assess whether services have met this requirements on a case-by-case basis. However, our objective is to help services meet this requirement and it is possible to provide guidance on key considerations.

b) Given the purpose of the risk assessment duty, we propose that a suitable and sufficient risk assessment should be **relevant to the specific characteristics of the service** in question and should **accurately reflect the risks**. It is important that the risk assessment provides services with an **adequate understanding of the risks** to implement appropriate measures in response.

c) We therefore propose that risk assessments should, as far as possible, be **based on relevant evidence** on the risk of harm on the service. In particular, services should consider evidence on the risk arising from the characteristics of the service specified in under Sections 9(5) and 26(5). The quality of the evidence and analysis underpinning the risk assessment is a key component of ensuring it is suitable and sufficient.

d) A key piece of relevant evidence is the Ofcom **Risk Profiles**, of which services have a duty to take account. Services should use the profile(s) to identify **relevant risk factors** (such as functionalities) and consider them as part of their risk assessment.

e) In addition to the Risk Profiles, our draft guidance sets out the types of evidence that services should consider (described in section 9.93 below). This approach is **designed to be scalable** for services of different types. The level of evidence and analysis required will depend on the nature and size of the service.

f) In addition to ensuring the quality of the risk assessment, there is also a link between the duty for services to carry out a suitable and sufficient risk assessment and the duty to take appropriate steps to keep the risk assessment **up to date.** If the risk assessment is not up to date, it may not be relevant, accurate and provide an adequate understanding of the risks, and therefore may not be suitable and sufficient.

Rationale and supporting evidence for our approach

9.25 Our proposed approach is supported by i) the best practice literature on risk management; and ii) the evidence we have gathered on existing risk assessment processes and other measures adopted in industry. Here we explain how these resources **support a flexible, scalable approach**, and our proposal that risk assessments be underpinned by an **appropriate evidence base** to support their relevance and accuracy. Later, we explain how these resources support our proposed methodology.

9.26 Research on best practice in risk assessments from Milliman, commissioned by Ofcom, emphasises that both quantitative and qualitative analysis is an important part of the risk assessment process; expert judgement should be applied alongside analysis of past trends. Analysis of relevant information is a key component of the full risk management cycle, including identifying risks, classifying risks, assessing risks, managing risks through controls, and monitoring risk. Identifying, understanding and communicating the relevant quantitative and qualitative metrics is important for a range of risk-related activities.

9.27 The Government HM Orange Book states that risk analysis can be undertaken with varying degrees of detail and complexity, depending on the purpose of the analysis, the availability and reliability of evidence and the resources available.[110] The International Organization for Standardization (ISO) 31000 industry standard on Risk Management explains that "Risk assessment should be conducted systematically, iteratively, and collaboratively, drawing on the knowledge and views of stakeholders. It should use the best available information, supplemented by further enquiry as necessary". [111]

9.28 Comparable regulatory regimes in the UK (which rely on more mature risk management systems) also highlight the importance of setting different expectations for different kinds of in-scope firms. The Financial Conduct Authority (FCA) / Prudential Regulatory Authority's (PRA) Senior Managers and Certification Regime offers three categories for solo-regulated firms: Limited Scope, Core and Enhanced. Each category has specified guidance for firms which are suited to their capabilities and risk.[112] Similarly, the Health and Safety Executive provides different level of guidance for small, low risk businesses, and larger, higher-risk businesses and those in major hazard industries.[113]

9.29 International regulatory regimes for online safety also take a scalable or differentiated approach. The Digital Services Act sets a threshold for 'very large online platforms' and 'very large online search services' who face different requirements, including risk assessments. Platforms with 45 million or more users in the EU are considered to have a significant societal and economic impact, and pose particular risks in the dissemination of

[110] UK Government, 2023. The Orange Book, Management of Risk - Principles and Concepts. [accessed 20 September 2023]. Subsequent references are to this document throughout.

[111] International Organization for Standardization (ISO), 2020. ISO 31000 Risk Management. [accessed 11 September 2023]. Subsequent references are to this document throughout.

[112] Financial Conduct Authority (FCA), 2015. Senior Managers and Certification Regime [accessed 21 September 2023].

[113] HSE, Managing risks and risks at work. [accessed 20 September 2023]. Subsequent references are to this document throughout.

illegal content and societal harms. Such services are required to ensure that their assessments are "based on the best available information and scientific insights".[114]

9.30 The Australian eSafety Commissioner offers guidance which initially divides firms based on the number of employees they have into two categories – 'Start Up' or 'Enterprise' – aimed at early-stage companies with 49 or fewer employees, or mid-sized and enterprise companies with 50 or more employees respectively.[115]

9.31 Evidence from industry also supports that assessing relevant evidence is already part of some services' risk assessment practices, and that a flexible, scalable approach is most appropriate for the purposes of online safety.

9.32 The Digital Trust and Safety Partnership (DTSP) brings together policymakers, law enforcement bodies, relevant NGOs and industry experts to develop best practices for online services. Its Safe Framework adopts a "proportionate" tiered approach to its assessments, because "DTSP companies are varied in organizational size, scale, and resource capacity. Due to the diverse range of products and services they provide, they face a broad spectrum of content- and conduct-related risks, with varying levels of systematic impact on the digital ecosystem. Moreover, there are different degrees of maturity for Trust & Safety teams and practices across the membership's products and services." [116]

9.33 Its most recent report, based on an assessment of 10 companies, the DTSP uses a maturity scale to highlight 'core' content moderation practices which are in a 'mature' state across the industry, alongside other areas which were found to be 'less mature', emphasising the wide range of understanding and expertise in online safety practices.[117]

9.34 Based on our experience implementing the VSP regulations and on submissions to our 2022 Illegal Harms Call for Evidence, there is evidence of a range of different approaches to risk assessment (from mature, continuous risk management systems to no assessment process); to measuring and developing understanding online risk and harm; and to the metrics and sources of evidence that services use to do so. There are clear differences between large services which often provide detailed information about the metrics they gather to assess safety on their services, and smaller services, with fewer UK users, which have often never engaged in risk assessment nor considered why it could be important in their industry. For instance, among smaller services whose business models are likely to result in higher levels of risk, such as those hosting adult content, some state that they circumvent the need for a risk assessment by moderating every piece of content which appears on the platform.

9.35 Overall, the evidence we have reviewed highlights that analysing relevant information on risk and harm on a given service is a critical factor in assessing risk and implementing appropriate mitigations. Information from online services emphasises that there is significant variation in the information and resources available to different services.

9.36 In our draft guidance, we have therefore adopted a scalable approach which allows services to differentiate based on their size, nature and likely levels of risk. This will help

[114] Digital Service's Act, 2022. [accessed 20 September 2023].
[115] eSafety Commissioner. Assessment tools. [accessed 20 September 2023].
[116] DTSP, 2021. Safe Framework. [accessed 20 September 2023].
[117] DTSP, 2022. The Safe Assessments: An Inaugural Evaluation of Trust & Safety Best Practices. [accessed 20 September 2023].

services to carry out a suitable and sufficient risk assessment while minimising the risk of any undue burden on services. We have emphasised that risk assessments should be based on appropriate and relevant evidence. To this end, the subsequent sections focus on a universal, scalable methodology that we propose for risk assessment, and our draft guidance on what evidence services should collect to inform it.

Detailed guidance: proposed risk assessment methodology

9.37 This section focuses on our proposed detailed guidance on the methodology for carrying out a suitable and sufficient risk assessment. It defines a four-step process which all services can follow to meet the duties.

9.38 The objectives underpinning our approach were set out above at 9.16. To meet these objectives, we have considered how services can fulfil the specific requirements under sections 9 and 26 of the Act in the most effective, but least onerous way. As well as the detailed requirements of the Act, we have considered evidence of best practice in risk management and undertaken an impact assessment to help inform our proposed approach.

9.39 The methodology also integrates Ofcom's Risk Profiles fully into the risk assessment process. We set out our proposed approach to Risk Profiles in the next section.

Key elements of risk management based on best practice

9.40 In other sectors where risk assessments are already commonplace, they are often a component part of a **broader system of risk management and governance** within an organisation. A report by Milliman into good practice principles in risk management states that "An effective risk management system is based on defining the risk environment and the approach to managing risk, and implementing an iterative, ongoing learning process to manage risk."[118] According to ISO 31000, risk assessment is defined as the overall process of risk identification, risk analysis and risk evaluation.[119]

9.41 We consider effective risk management to be a critical factor for organisations to achieve good outcomes, both in terms of user safety and wider business objectives (commercial, reputational or related to sustainability and corporate responsibility). Where risk management systems are absent, inadequate, or inconsistently applied, there can be serious risks that an organisation will be unable to anticipate or respond to adverse events, or to protect users from harm.

9.42 Ofcom has reviewed a wide range of literature on best practice in risk assessments and risk management.[120] There is broad consensus around the key elements of risk management:

[118] Milliman, 2023.

[119] ISO, 2020.

[120] Alongside work by Milliman commissioned by Ofcom (2023) this includes, for example, relevant ISO standards, UK Government Orange Book, and guidance from the National Cyber Security Centre (NCSC).

Table 9.1: The key elements of best practice

Element	Activities	Illustrative outputs
Identifying risks	Exercises to identify risks that may affect an organisation, even if the risk is unlikely to occur or to materially impact business operations. This may involve interviews or surveys with relevant stakeholders, evidence-based methods such as literature reviews and analysis of historical data, scenario analysis and structured examination techniques such as Hazard & Operability Analysis[121] (HAZOP) or Structured What If Technique[122] (SWIFT).	A risk register which provides an exhaustive list of potential risks, classified by category or type.
Assessing risks	Conducting risk assessments and evaluating risks. This includes determining the likelihood and potential impact of events taking place that could affect or disrupt business operations. This typically feeds into an exercise to determine the severity or significance of events.	A risk assessment which scores, maps or evaluates risks according to pre-determined criteria; documentation which highlights priority risks an organisation faces based on the outcome of a risk assessment exercise (i.e. risks which have the most severe or significant consequences).
Managing risks	Putting in place risk mitigations, and internal and external controls that seek to reduce the likelihood of the events occurring, or to manage and mitigate their impact on business objectives.	Risk management plans detailing the controls in place to manage and mitigate risk. Plans should include consideration of any unintended consequences that controls may trigger.
Reporting risks	Ensuring that risk assessments and decisions on implementation of controls are recorded. Embedding risk management processes into governance structures by ensuring that risk management activities are regularly reported to risk governance bodies, and that there is effective oversight of risk across an organisation.	Records of risk assessments and measures taken to manage risks Policies and documentation laying out governance processes for risk management, including decision-making and oversight functions.

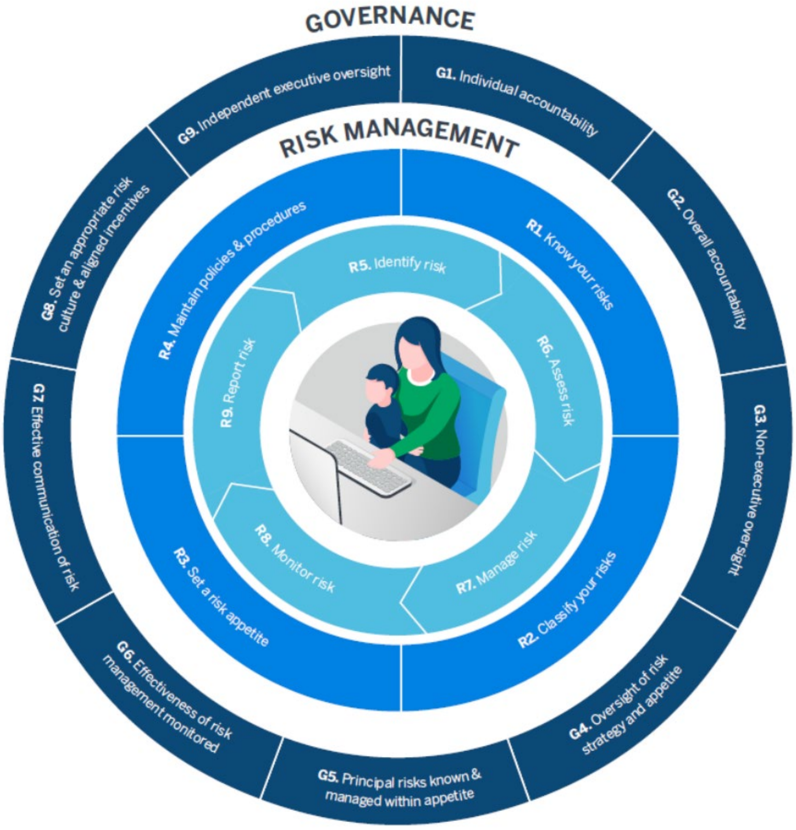
Source: Ofcom Analysis

9.43 Risk management is not a one-off, or a point-in-time exercise. Rather, it is a continuous and iterative process that involves feedback and communication at each stage. As such, risk management is often presented as a feedback loop, rather than a list of requirements and actions.

[121] International Electrotechnical Commission (IEC) and International Organisation 31010 on Risk Assessment Techniques, 2019. Subsequent references are to this document throughout.

[122] ISO/IEC 31010, 2019.

Figure 9.1: An organisation's risk management and governance system, the Milliman report


GOVERNANCE
G1. Individual accountability
G2. Overall accountability
G3. Non-executive oversight
G4. Oversight of risk strategy and appetite
G5. Principal risks known & managed within appetite
G6. Effectiveness of risk management monitored
G7 Effective communication of risk
G8. Set an appropriate risk culture & aligned incentives
G9. Independent executive oversight
RISK MANAGEMENT
R1. Know your risks
R2. Classify your risks
R3. Set a risk appetite
R4. Maintain policies & procedures
R5. Identify risk
R6. Assess risk
R7. Manage risk
R8. Monitor risk
R9. Report risk

Source: Milliman[123]

9.44 Risk assessment refers to the overall process of **identifying and assessing risks**, as outlined in Table 9.1 above. The outcomes of a risk assessment should be a comprehensive evaluation of the risks faced by an organisation. The findings inform how risks should be prioritised, managed, and reported within the organisation.

Proposals on applying risk management principles in the context of online safety

9.45 To achieve our policy objectives, we have incorporated many of these key principles and activities into the draft Service Risk Assessment Guidance.

9.46 Our key proposal is that services conduct their risk assessments as a **four-step process**. These four steps would be applicable to any type of risk assessment services carry out and would be designed for application to all in-scope services.

9.47 Within the four-step process, we have sought to embed all of the key legal requirements of the risk assessment duties. This is set out in Table 9.2 below. Throughout, we indicate how services should undertake each of their steps to help meet the requirement that their risk assessment is suitable and sufficient.

9.48 We consider that the risk assessment methodology should be implemented as an iterative and continuous process, supported by strong governance. This process should assist

[123] Milliman, 2023.

services to understand and manage risk on an ongoing basis, and remain compliant with the risk assessment duties.

9.49 In line with best practice, this continuous process also allows risk assessments to be integrated into other business activities, including product design and development, research and testing, and governance activities. This creates an opportunity to achieve our objective to, "use the risk assessment process to create a clearer route to compliance across the regime, by integrating other resources produced by Ofcom." We have indicated how the proposed risk assessment methodology can support compliance across the OS regime.

Rationale and supporting evidence for our approach

9.50 We found widespread support in best practice industry frameworks and standards for providing guidance on risk assessments as a staged and iterative process. This included:

a) The Milliman report highlights that a "structured approach to risk management is critical for successful implementation" and sets out an iterative five-stage process.

b) The ISO 31000 International Standard on Risk Management emphasises that risk management should be conducted systematically and iteratively.[124] The corresponding standard IEC 31010 on Risk Assessment Techniques provides a six-stage process.[125]

c) In health and safety, the UK Government's Health and Safety Executive describes five main steps to its recommended assessment method.

d) In cyber security, the National Institute of Standards and Technology's (NIST) Framework for Improving Critical Infrastructure Cybersecurity provides a framework "to help identify and prioritize actions for reducing cybersecurity risk" which is organised into five functions.[126]

e) In the field of human rights, human rights impact assessments (HRIA) analyse the effects that business activities have on right-holders and human rights principles. Guidance from the Danish Institute of Human Rights provides a five-phase process for conducting HRIAs, while BSR's Human Rights Assessment describes four stages to its assessment methodology. [127]

f) Ofcom has also already recommended that online services adopt a staged process for risk assessment and risk management under the Video-Sharing Platform regulations[128]. While this is not an obligation under the VSP regime, we have seen adoption of risk assessment methods by some notified platforms.

g) Finally, several respondents to our Call for Evidence also pointed to sequenced and iterative frameworks for risk assessment that they currently use to manage risk on

[124] ISO, 2020.

[125] ISO/IEC 31010, 2019.

[126] National Institute of Standards and Technology's (NIST), Framework for Improving Critical Infrastructure Cybersecurity. [accessed 21 September 2023].

[127] BSR, 2021. Human Rights Assessment: Identifying Risks, Informing Strategy. [accessed 21 September 2023]; Danish Institute of Human Rights, Human rights impact assessment guidance and toolbox [accessed 21 September 2023].

[128] Ofcom, 2022. Ofcom's first year of video-sharing platform regulation. [accessed 21 September 2023].

their services. Such frameworks were described by, for example, a search service, a dating app business, and a VSP.

9.51 While these individual frameworks differ based on the specific needs of each industry or organisation, there are clear common elements. All risk assessment models we have reviewed include stages which can be categorised as: i) planning and identifying risks; ii) collecting information, analysis and assessing risks; iii) putting in place mitigations, controls or protective measures; and iv) recording, reporting, monitoring risks and ongoing review.

The four-step risk assessment process

9.52 In our draft detailed guidance on methodology, we have proposed a process which reflects these four steps: i) understand the harms; ii) assess the risks; iii) decide measures, implement and record; and iv) report, review and update the risk assessment. We also include key common concepts from best practice which align to the risk assessment duties, such as:

a) Assessing risk through a matrix of likelihood and impact;

b) Assigning a risk level for each harm; and

c) Considering residual risk after mitigating measures have been applied.

9.53 Within each step, we have embedded specific activities to support services to meet their illegal content risk assessment duties. This includes taking account of Ofcom's Risk Profiles, which is a specific requirement of the Act.

9.54 We propose to include Risk Profiles at Step 1 because we expect the information provided in the Risk Profiles to give services a consistent starting point to understand risk on their services, based on Ofcom's sector-wide assessment on how harms manifest online (our Register of Risks). This will allow services to identify many of the key **risk factors** that could make harm more likely or impactful.

9.55 However, Risk Profiles are a starting point for services to think about the risks on their service. As explained above, risk is context specific and will vary based on the nature of the individual service. While Risk Profiles provide Ofcom's evidence-based guidance for services, the effect of an individual risk factor can vary. Different combinations of risk factors, user behaviour and existing mitigation measures will affect the level of risk that each service presents. As such, we propose to guide services to consider these risk factors alongside their own evidence and analysis at Step 2 to make an accurate assessment of their risk of each type of illegal harm.

9.56 In the table below, we summarise the four steps set out in the draft guidance, explaining each component and highlights the rationale in either the requirements of the Act or best practice in risk assessment. In the subsequent section, we explain our approach to Risk Profiles.

Table 9.2: proposed four-step risk assessment methodology

Step	Key activities included in the draft guidance	Explanatory notes	Rationale / Basis in the Act
1. Understand the harms	Identify the harms that need to be assessed	The guidance will explain the concept of illegal content and, for U2U services, introduce the requirement to assess the risk of the commission or facilitation of an offence (jointly referred to as 'illegal harm'). It will provide a breakdown of the different kinds of priority illegal content offences that need to be assessed in a risk assessment, based on Ofcom's Register of Risks.	To undertake a risk assessment, it will be necessary for services to understand the concept of illegal harm, and the kinds of priority illegal content and offences that must be assessed separately. It also reflects best practice on planning for the risk assessment and identifying and categorising risks. This step is relevant to components of the risk assessment duty for U2U and search services set out in the Act under section 9(5) and section 26(5) respectively.
	Consult Ofcom's Risk Profiles	The guidance explains that Ofcom's Risks Profiles help services identify an initial set of risk factors that apply to their service. These risk factors indicate which harms could be more likely on their service and should inform their assessment. We explain that services are required to take the Risk Profiles into account when conducting their assessment. In this section, we provide guidance on how services should use Ofcom's Risk Profiles. This includes direction for services to use a set of questions that we provide in Appendix A to identify which risk factors from the Risk Profiles apply to their service. We also direct services to record these risk factors.	Ofcom's Risk Profiles identify various risk factors which relate to particular harms. This will be a helpful first step for services in understanding which characteristics (including functionalities, user base and business models) may increase risks of illegal harm occurring. Services will be required to keep written records of their risk assessments, which should include the risk factors they identify. Services are required to take Risk Profiles in to account under section 9(5) and 26(5) of the Act. The risk factors which form the Risk Profiles are generally based on the "characteristics" of a service as defined at section 98(11). The **record-keeping duty** at section 23(2) and 34(2) of the Act requires services "to make and keep a written record, in an easily understandable form, of every risk assessment" under section 9 or 11, and section 26 or 28. Also see Annex 6 for Ofcom's guidance on Record keeping and Review.

2. Assess the risks of harm	Consider any additional characteristics that may increase or decrease risks of harm on your service	The risk factors included in the Risk Profiles do not include all of the characteristics that may give rise to risk. We introduce to services the need to consider their own evidence and other resources (including the Register of Risks where appropriate) to make a full and accurate evaluation of risk. This includes direction for services to consider if there are any other characteristics which apply to them, but which may not be present in Ofcom's Risk Profiles.	Ofcom's Risk Profiles are not intended to provide a bespoke analysis of risk as it exists on individual services. They are a starting point for services when considering which characteristics may increase risks to users. As such, services will need to refer to their own evidence when conducting their risk assessments. Different characteristics of a service may be risk factors for particular harms, so it is important that services consider these in their assessment of likelihood and impact. This will include assessing the impact of algorithms. Giving dedicated considerations to the service's specific characteristics is required under sections 9(5)(a), (b), (e), (f) and (h) of the Act for U2U and sections 26(5)(a), (b) and (d) of the Act for Search (which indicate the kinds of characteristics to be considered) and is an important part of completing a suitable and sufficient risk assessment.
	Assess the likelihood and impact of each kind of illegal harm	We guide services to assess the risk of each kind of priority harm by considering likelihood and impact in their assessment. To help then do this, we provide a list of relevant evidence inputs, and explain how services should make decisions about which kinds of evidence to use. This step is key in services understanding their specific characteristics, and any risk factors that could give rise to harm on a service. We also provide guiding questions for services on how they might assess and reach conclusions on the likelihood and impact of harm.	An assessment of risk of harm requires services to have a good understanding of the probability or likelihood of harm occurring, and the impact this harm has on individuals. As set out above, it will be important that services consult relevant evidence to make this evaluation, including the risk factors they identified through Ofcom's Risk Profiles in Step 1 and other evidence that is specific to their service. The full rationale behind our guidance on the evidence services should consider is set out below. This step is relevant to components of the risk assessment duty for U2U and Search services set out in the Act under section 9(5)(b), (c), (d), (e) and (g), and section 26(5)(a), (b) and (c). This includes the requirement for services to assess the risk of users "encountering" illegal content or (for U2U) the risk of the service being used to

			commit or facilitate a priority offence, and the "nature and severity" of harm.
	Assign a risk level to each kind of illegal harm, including by referring to our additional guidance on specific harms	The guidance will advise services to assign a risk level of low, medium or high to each kind of priority harm. We help services by providing an example of a risk matrix table, and by including a reference table listing the typical characteristics of each risk level. In this section, we also guide services to consider our additional guidance on assigning a level of risk to specific harms, where we have corresponding measures in our Codes of Practice. This will include CSAM and grooming. We guide services to make a full written record of their risk assessment, in line with Ofcom's guidance on record keeping and review.	Evaluating the risk of harm is a cornerstone of the risk assessment duty. This will also be important in directing services when they come to choosing and implementing safety measures to reduce risks to individuals from illegal harm. The proposed low, medium or high scale, based on likelihood and impact, is a common methodology found widely in best practice risk assessment literature. We have provided additional guidance on assessing risks of CSAM and grooming given the severity, sensitivity, and complexity of these harm areas. This is in line with our strategic approach to regulation, which focuses on addressing some of the most serious illegal harms to children.[129] The **record-keeping duty** at section 23(2) and 34(2) of the Act requires services "to make and keep a written record, in an easily understandable form, of every risk assessment" under section 9 or 11, and section 26 or 28. Also see Annex 6 for Ofcom's guidance on Record keeping and Review.
3. Decide measures, implement and record	Decide on measures to take to reduce the risk of harm	The guidance will state that services need to decide how to comply with the safety duty, whether by taking the measures recommended in Ofcom's Codes of Practice or otherwise. Services using Ofcom's Code of Practice will need to use the outcomes of Step 2 of the risk assessment (i.e. the level of risk	While we have included this step in the draft Risk Assessment Guidance for completeness, this is based on the safety duties at sections 10 and 27 of the Act, Ofcom's Codes of Practice, and the record keeping duties. This is included in the draft guidance because: i) Best practice in risk management includes assessing the impact of any mitigating measures and their effect of the level of risk and residual risk.

[129] See Ofcom's approach to implementing the Online Safety Act.

		assigned to each kind of priority harm) to inform the safety measures it implements. Services determining their own means of meeting the safety duty should record how their measures respond to the risks they have identified.	ii) Sections 9(5)(h) and 26(5)(d) of the risk assessment duties includes assessing “how the design and operation of the service (including the business model, governance, use of proactive technology, measures to promote users’ media literacy and safe use of the service, and other systems and processes) may reduce or increase the risks identified.” Therefore services’ safety measures are relevant to consider as part of their risk assessment. iii) Our objectives in developing the guidance include providing a clearer route to compliance across the regime, by integrating other resources produced by Ofcom into the guidance including the Register of Risks, Risk Profiles, Code of Practice and guidance on Record Keeping and Review.
	Consider any additional measures that may be appropriate	Services will be advised to consider any additional measures to respond to the risks they have identified. Ofcom’s Codes of Practice may not be comprehensive and services may be better placed to identify additional effective measures to prevent harm on their specific service.	As above. This is in line with best practice in risk management and mitigation, which emphasises assessing and managing residual risk, as well as inherent risk. This also reflects the dynamic and ongoing nature of risk management, of which risk assessment is a key part.
	Implement all measures to manage and mitigate risk	Implement the measures identified in the previous two steps	As above.
	Record the outcomes of the risk assessment and how the safety duties have been met	We will explain the record keeping duties which apply to services mitigation measures.	As well as keeping a record of their assessment of risks, services must keep records of their compliance with the safety duty as set out at sections 23(3)-(6) and 34(3)-(6) of the Act. We have included this in the draft Service Risk Assessment Guidance as it is relevant to services’ mitigation measures and because it offered services a clearer, coherent route to compliance across multiple duties (in line with our objectives). Also see Annex 6 for Ofcom’s guidance on Record keeping and Review.

4. Report, review and update risk assessment	Report on the risk assessment and measures via relevant governance channels	We will guide services in best practice arrangements for governance reporting.	Best practice literature commonly includes reporting and governance to help provide internal assurance and visibility of the risk assessment process. This step also aligns with Ofcom's Code of Practice measures on organisational structure and governance (applicable to large and high risk services).
	Monitor the effectiveness of your mitigation measures	We will explain best practice around ongoing risk management and mitigation, which would require services to ensure that their mitigations continue to be effective in addressing harm after implementation.	As above, best practice emphasises the need for risk assessment to be an ongoing, cyclical process. Monitoring the mitigation measures and their effectiveness also links to the requirement at sections 9(3) and 26(3) to take "appropriate steps" to keep a risk assessment "up to date", which may include monitoring mitigations' effect on risk.
	Recognise and act on triggers to review an existing risk assessment or complete a new risk assessment	We will explain the points where a service will be required to review risk assessments as laid out in the Act: • If Ofcom updates a risk profile • Before a significant change to the design or operation of the service	Sections 9(3) and 26(3) of the Act set out a legal duty for services to keep their risk assessment 'up to date' through 'appropriate steps' including when Ofcom makes a significant change to risk profiles. We have advised that services meet this duty by holding a written policy which details the appropriate steps taken. Further, we propose that this policy should at a minimum establish a minimum review period of 12 months and by appointing a responsible person ensure the assessment remains up to date, including when Ofcom changes risk profiles. The Act sets out a further legal duty requiring a service to carry out a further assessment relating to any proposed significant change. To help services meet this duty we have offered guidance on what may amount to a significant change and how they can identify whether a proposed change is significant or not.
	Put in place regular review periods for your assessments	We will recommend a minimum timeframe (outside of the triggers to review a risk assessment) that a service should undertake reviews of their assessments of 12 months.	As above, we propose that services should at a minimum establish a minimum review period of 12 months which aligns with international best practice (such as the Digital Services Act) and by appointing a responsible person ensure the assessment remains up to date, including when Ofcom changes risk profiles. Also see Annex 6 for Ofcom's guidance on Record keeping and Review.

Source: Ofcom analysis

Guidance on assessing likelihood and impact

9.57 As set out in earlier sections of this chapter, a key element of risk assessment best practice is making an accurate assessment of risk based on relevant information. Step 2 of our proposed methodology includes our draft guidance to services on how to assess risk by considering the likelihood and impact of illegal harm occurring by means of their service. Assessing risk in this way is both common practice and part of services' legal duties. We advise services to assign a low, medium or high level of risk and we provide specific guidance to support services in reach accurate conclusions in making this judgement. This includes:

a) General guidance, including:

i) Guiding questions on how to assess likelihood;

ii) Guiding questions on how to assess impact;

iii) A risk level table, illustrating the factors that could lead a service to conclude that they should assign a low, medium or high risk level; and

b) Guidance on risk levels for specific harms, namely CSAM and grooming.

9.58 Our general guidance aims to help services reach more accurate conclusion on the level of risk by explaining how to give appropriate consideration to the risk factors in Ofcom's Risk Profiles, which are in turn underpinned by the evidence in the Register of Risks; to the characteristics of their service, as specified in the Act; and to their own evidence (including core and enhanced inputs as relevant). These are not mandatory steps but aim to assist services in their analysis.

9.59 As part of the risk level table, we also provide draft guidance on the effect of a service's user numbers on its level of risk. In general, all else being equal, the more users a service has, the more users can be affected by illegal content and the greater the impact of any illegal content. We have therefore proposed that services which reach certain user numbers should consider the potential impact of harm to be medium or high.

9.60 For high impact, we propose a user number of more than 7 million monthly UK users. This aligns with how we propose to define a 'large' service, as discussed from Chapter 11, paragraph 11.51. It represents approximately 10% of the UK population, which is similar to the definition of very large service taken by the EU in the Digital Services Act.[130] It is also broadly similar to one of the factors feeding into the highest risk category in the Australian social media code.[131]

9.61 For medium impact, we propose a user number of between 700,000 and 7 million monthly UK users. This is approximately equivalent to having a user base of 1% to 10% of the UK population. In January 2023, 1,039 online brands were each visited by between 700,000 – 7

[130] The Digital Services Act classifies platforms or search engines as very large online platforms (VLOPs) or very large online search engines (VLOSEs) if they have more than 45 million users per month in the EU, a number equivalent to 10 % of the EU population.

[131] The Australian Social Media Services Online Safety Code relates to content that is similar to some of that included in our illegal codes. The measures that apply to social media services depend on which of three risk levels services fall into. There are a number of different factors that determine the risk level, which are given equal weight. One of those factors for the highest risk category is having over 3 million Australian users. This is roughly 11% of the Australian population, though the definition of user is narrower than we propose to use as it relates to Australian monthly active account holders.

million UK individuals aged 15+ on smartphones, tablets or computers[132], but only some fraction of these brands would have regulated services and be in-scope for the Act.

9.62 We are clear in the Service Risk Assessment Guidance that in some instances the number of users may be a weak indicator of risk level. They need to be considered alongside other risk factors. It is possible for a large service to be low risk, and for a small service to be high risk, depending on the specific circumstances of each service.

9.63 In the draft guidance, we also provide specific tables setting out how to assign risk levels for CSAM and grooming offences. We single out these harms specifically given the severity, sensitivity, and complexity of these harm areas; this is in line with our strategic approach to regulation, which focuses on addressing some of the most serious illegal harms to children.[133] In our proposed Codes of Practice, we have recommended specific measures aimed at addressing CSAM and grooming. We anticipate that additional guidance will assist services in making an accurate judgement of their level of risk for these harms and inform their approach to achieving their duties. We provide a full explanation of our rationale for these recommended measures in the relevant Codes chapters, with relevant supporting evidence set out in the draft Register of Risks.[134] These chapters explain why we attach weight to certain risk factors for these harms, and why other risk factors should be considered incrementally.

Impact assessment

9.64 The methodology proposed in the draft guidance introduces strong risk management practices that meet our policy objectives set out in section [9.16], and align with best practice in risk management. It has a clear structure and framework for all services to follow to enable them to meet all the elements of their risk assessment duties.[135] It will help them identify, understand, and respond to risks on their service.

9.65 We consider that following this methodology will deliver significant benefits. The evidence we have looked at shows that following best practice in risk assessment plays an important role in improving safety outcomes. This is because robust risk assessment makes services better able to identify risks, thus improving their ability to mitigate those risks.'

9.66 Implementing the proposed methodology will include costs (e.g. staff costs) and these costs will be higher where services do not have existing risk management processes in place.[136] However, the risk assessment duty is imposed by the Act and services will need to incur the costs of undertaking suitable and sufficient risk assessments to meet their legal obligations. In addition, our proposed methodology is intended to be flexible depending on service's risk levels, size and resources in order to minimise the cost burden. We intend

[132] Ipsos, 2023, Ipsos iris Online Audience Measurement Service, age: 15+, UK. We note Ipsos defines an online brand as consisting of its applications and websites.

[133] See Ofcom's approach to implementing the Online Safety Act.

[134] Chapter 14 on Automated Content Moderation, and Chapter 18 on Default settings and user support.

[135] This includes a means of addressing all of the elements of the risk assessment as set out in sections 9 and 26 of the Act, including how services should consult Ofcom's Profiles, how to take into account all of the relevant characteristics of their service and their user base, and how to make accurate judgments on the likelihood, nature and severity of illegal harms taking place by means of the service.

[136] We expect that the cost of setting up new risk management processes will be higher than updating existing ones.

that it could be integrated into existing risk management practices to improve the effectiveness of online safety risk assessments and minimise additional costs.

9.67 Overall, we provisionally consider that our proposed methodology is proportionate. The Service Risk Assessment Guidance does not represent a set of compulsory steps that services must take, but rather is intended to assist services in fulfilling their legal obligations, while giving them the flexibility to minimise costs. For services that have existing processes for assessing and managing risks (e.g. some large services), we do not expect our draft guidance to lead to significant additional costs.[137] While services with no existing processes in place (e.g. some small services) will face greater cost burden in undertaking risk assessments for the first time, we expect that they will need to incur the bulk of this cost in any case to comply with their risk assessment duties and our proposed guidance will support them in doing so by helping them to identify, assess, manage and record risks.

9.68 Lastly, the sections below explain how smaller services can benefit from the flexibility of the framework and how we have sought to minimise any unnecessary cost burden. To the extent that our proposed risk assessment imposes costs, we provisionally consider that these are justified by the significant benefits associated with high quality risk assessment processes.

9.69 Ofcom is also exploring new tools, techniques and services to further assist smaller services in complying with their risk assessment duties, which may reduce the overall cost impact.

Provisional conclusions

9.70 We believe that the proposed four-step risk assessment methodology is an appropriate response to Ofcom and regulated services' legal duties, will fulfil the policy objective we have set out, and represents a proportionate set of recommendations to services.

9.71 Having explained the rationale behind our proposals about the methodology services should follow when doing their risk assessment, we now move on to focus in detail on the following aspects of our risk assessment guidance: (i) our proposed approach to developing the Risk Profiles which services will need to consult during step 1 of the process; (ii) our proposals on the evidence services should consult when undertaking their risk assessment; and (iii) our proposals regarding when services should review and update their risk assessments.

[137] Costs may be incurred in aligning with our proposed four-step methodology, but as described above our draft guidance is a flexible framework to minimise this.

Our proposed approach to Risk Profiles

Proposed approach

9.72 The Act requires Ofcom to prepare and publish 'Risk Profiles' based on the findings in our Register of Risks (Register).[138] The Act gives Ofcom wide discretion about how to do this. In particular, we can group services in whichever way we consider appropriate.[139]

9.73 Services are required to take account of our Risk Profiles when they carry out their risk assessments and given this, we consider that their intended purpose is to help services conduct their risk assessment.

9.74 We are proposing an approach which uses the Risk Profiles to highlight:

a) What characteristics[140] of online services are likely to increase risk (we refer to these as **risk factors**), and

b) Indicate which kinds of illegal harms (see Appendix B of Annex 5) may be more likely to occur on their service as a result.

9.75 We are proposing to present a Risk Profile as a table with each row representing an individual risk factor (e.g. child users or livestreaming). For each risk factor, we provide a high-level description of how the risk typically arises, and the illegal harms that are most relevant to that risk factor**.** The table does not set out all the risk factors from the Register, instead it includes those which we have determined to be particularly important for services to consider.[141]

9.76 There is one table of risk factors for U2U services to consult ('U2U Risk Profile') and one for Search services to consult ('Search Risk Profile'). Services should consult the relevant table and decide which risk factors are relevant to them. The draft tables are available in Appendix A of Annex 5.

a) Some of the risk factors in the tables are things that only some services must take account of in their risk assessment, because they represent characteristics that only certain services will have (e.g. comments). We refer to these as **specific risk factors,** and services are expected to identify which specific risk factors apply to them. To help services do this accurately, we provide a list of Y / N questions, where each 'Y' answer corresponds to an additional risk factor in the table.

[138] Section 98(5) of the Act. The Register of Risks is Ofcom's own risk assessment of the impact of characteristics of services on the risks of harm to individuals from illegal content. For U2U services, this includes the risk of harm from the facilitation and commission of illegal harms, as well as users encountering illegal content. For Search, this includes only the risk of harm from users encountering illegal content. Details on our approach to the Register, as well as the full findings of our risk assessment are available in full in Volume 2.

[139] Taking into account the characteristics of services, the risk levels and other matters identified in Ofcom's risk assessment.

[140] Characteristics include a service's user base, business model, functionalities and any other matters we deem relevant to risk. Risk Profiles focus predominately on user base demographics, functionalities and business models. Step 2 of the risk assessment guidance provides information for services on user base size, governance, and systems and processes.

[141] For further details on how we determined this, see paragraphs 9.84 - 9.91 of this chapter.

b) Some of the risk factors in the tables are things that all services must take account of (e.g. user base demographics). We refer to these as **general risk factors.**

9.77 **After consulting the relevant table, services should have identified the list of risk factors (and associated illegal harms) that apply to them, which they must take account of in their risk assessment.** This list will always include all general risk factors for either Search or U2U, plus any specific risk factors indicated by their answers.

9.78 By taking account of our Risk Profiles in this way, services will have a good starting point for thinking about the level of risk their service may present for different kinds of illegal harms and which risk factors ordinarily contribute to that risk. As explained in Table 9.2, services should use this information to help them assess their risk level for each kind of priority illegal offence in Step 2 of their risk assessment.

Options considered and analysis

9.79 We considered several alternative options for our approach to Risk Profiles. We considered each option against two main objectives:

a) **Our approach should effectively present our evidence on what makes services risky.** All of the information in 'Risk Profiles' is based on our in-depth assessment of what increases the risk of illegal harm (Register of Risks). Our approach should allow us to accurately and robustly present the main findings from the Register of Risks.

b) **Our approach should be easy for all services to use.** All services are required to take account of 'Risk Profiles' as part of their risk assessment. This means our approach needs to be easy for all services to use, including small and micro businesses. It also needs to provide relevant information on risks to the wide range of services in scope.

9.80 We have considered whether it would be possible to produce separate **Risk Profiles for different 'types' of service**, rather than one for all U2U and another for all Search services. For example, a social media service would consult a 'Social Media Risk Profile' and online gaming services would consult an 'Online Gaming Risk Profile'. This would have allowed us to draw on terminology that is widely recognised in the sector and to group together services that appear to be similar. Given this, we would expect this option to have been relatively intuitive for services to use based on how they generally describe their service. **However, we rejected this option because it did not allow us to effectively present our evidence on risk.**

a) First, services that may fall into the same 'type' of service (and would therefore have the same Risk Profile) can have very different risks**.** For example, one social media service may allow direct messaging and child users, but another may allow neither. An online marketplace or online gaming service may allow both risk factors. We consider that services with the same risk factors - rather than those of the same service type - should consider similar evidence about risk to make an accurate assessment.

b) Second, while different sources of evidence may use the same terminology, the way they define a term can vary substantially. For example, a study may define a specific service as a social media service, and another may define it as a discussion forum. This makes it hard for us to generalise and categorise the evidence, and for services to select which 'Risk Profile(s)' applies to them.

c) In contrast, our proposed approach focuses on individual risk factors. This means services are expected to take account of risks based on their characteristics (e.g. direct messaging and child users), even if they are not ordinarily associated with their type of service. We expect this to give services a more accurate and robust understanding of the risks on their service.

9.81 We also considered producing a **'Risk Profile' for each kind of priority illegal harm**. Services would have 15 Risk Profiles to consult, each representing one of the illegal harms, for example an 'Intimate Image Abuse Risk Profile' or 'Fraud and Financial Services Risk Profile'. These would also match the 15 groupings of kinds of priority illegal harm used the Register of Risks and Risk Assessment Guidance.[142] This would have the advantage of following the same structure as the Register of Risks and we would have expected it to have effectively presented our evidence. It also would have allowed services to easily find the detailed evidence underpinning our conclusions. **However, we rejected this option because we considered it would not be easy for services to use.**

a) Services would need to identify which Risk Profile applied to them – to do this accurately, they would need some prior knowledge about the risks associated with each kind of illegal harm. This would be particularly hard for services with fewer resources or less internal expertise.

b) Under our proposed approach, services only need to know their functionalities or features to identify the relevant information about risks. We expect this to be easier for all services to use.

9.82 Overall, we consider that our proposed approach is better able to meet our policy objectives than the alternatives. Further details of the options we tested are included in Table 9.3.

9.83 There are two further points to bear in mind related to our approach:

a) First, **we expect to update our Risk Profiles to keep them up to date as our evidence base develops in the Register**. The risk factors for both U2U and Search will be updated as part of this process.

b) Second, **we include information about the links between risk factors and the 15 kinds of priority illegal harms at present.**[143] As our evidence base develops on other relevant offences, we will reassess this approach and adapt if appropriate.

How we determined the list of risk factors

9.84 Risk Profiles are a starting point for services to think about the risks on their service. As mentioned, the tables do not include all the risks identified in the Register, but only those

[142] The 15 kinds of illegal harm are: Terrorism offences; CSEA (grooming and CSAM) offences; encouraging or assisting suicide offence or serious self-harm; hate offences; harassment/stalking/threats/abuse offences; controlling or coercive behaviour offence; drugs and psychoactive substances offences; firearms and weapons offences; unlawful immigration/human trafficking offences; sexual exploitation of adults offences; extreme pornography offence; intimate image abuse offences; proceed of crime offences; fraud and financial services offences; foreign interference offence. For further information, see Volume 2, Chapter 5. The new self-harm offence is not yet in force and is not a priority offence, but as explained further in Volume 2, Chapter 5, we have included it in the same kind of illegal harm as encouraging or assisting suicide.

[143] The animal cruelty priority offence is not currently included in our Register, or as a consequence, in the Risk Profiles. As explained further in Volume 2, Chapter 5, we will consult in due course how we propose to include that offence.

we determined to be particularly important for services to consider based on our current evidence base.

9.85 A challenge we recognise is that Risk Profiles cannot fully capture the complexity and context of risk factors across all the harms considered. As explained in detail in the Introduction to the Register, risks can arise in combinations, and be different for specific kinds of illegal harm. Risk is also influenced by user motivations and other dynamics that may be unique to a harm or the nature of the service itself. For example, some risk factors linked to one illegal harm can provide protection from another harm.

9.86 Our methodology looks at all the risk factors in the Register chapters for the priority illegal harms and evaluated them in two stages.

9.87 First, we considered if the risk factor would apply to all services. We identified three risk factors that met these criteria (user base demographics, business model (revenue model and growth strategy) and commercial profile).[144] We refer to these as **general** risk factors, and they are risk factors for both U2U and Search services. We refer to all other risk factors in the Register as **specific** risk factors – these are mainly functionalities where some services may have them, and others may not.[145]

9.88 Second, we considered what risk factors to include in the tables, and what information about different kinds of illegal harms to highlight.

9.89 Given there were only three **general risk factors**, we include high level information about all three in both the U2U and Search tables. We also provide information about different kinds of illegal harms where possible.

9.90 We took different approaches to the **specific risk factors** for U2U and Search.

a) There are numerous U2U risk factors, and we therefore conducted a qualitative analysis to identify which risk factors were most strongly associated with the different kinds of illegal harms in our evidence base.[146] We only include information on these relationships in the U2U risk factor table.

[144] Commercial profile includes the capacity, pace of growth and maturity of a service.

[145] We include 'child users' as a specific risk factor, in addition to referencing age as part of the general risk factor of user demographics. This is because, unlike other demographic factors, there are services that allow child users, and services that do not. We wanted to ensure we were providing services with a clear expectation that if they allow child users, they need to take account of additional risks, in particular CSEA harms. In this version of Risk Profiles, we differentiate services based on if they *allow* child users. We recognise there are other considerations related to the presence of children on a service and will continue to monitor this approach to ensure alignment with our forthcoming work regarding age assurance, children's safety duties, children's access assessments and children's risk assessment.

[146] We determined that a qualitative methodology was better able to provide an accurate assessment of the evidence available given the complexity of the evidence and the lack of consistent or comparable numerical data across illegal harms. The methodology considered the strength of the evidence for different risk factors, common trends across illegal harms, and alignment with other aspects of our regulatory approach. For example, when considering "hyperlinks" as a risk factor, we considered how the evidence in the Register explained the relationship between hyperlinks and each kind of illegal harm individually, as well as considering the relationship between hyperlinks and illegal content more broadly. We also considered the relationship between hyperlinks and our wider regulatory approach, for example the Codes of Practice.

b) There are fewer Search risk factors.[147] We include all risk factors in the Search table and describe the general risk of harm, rather than linking a risk factor to individual illegal harms.

9.91 **While the list of risk factors we include in Risk Profiles reflect the evidence in the Register, it is high level and does not include all of the characteristics that may give rise to risk.** We include information for services on where they can find more extensive information on the risk factors and illegal harms within the Register. We are also clear in the draft Risk Assessment Guidance that services should see the Risk Profiles as a starting point for understanding their risks and that they should consider their risk factors alongside i) any other relevant characteristics of their service, and ii) relevant evidence about their specific service.

Table 9.3: Overview of approaches considered for Risk Profiles

Options	Relevant considerations
Option 1: **Risk Profiles by U2U and Search** ***(Proposed Option)***	This option, described in detail above, enables us to draw out the evidence robustly and accurately in the Register, and present it in a way that is easy for services to use. In terms of drawing out the evidence, this approach allows us to highlight similarities in risk across different services and illegal harms. For example, we can explain how direct messaging, whether between buyers and sellers within an online marketplace, as a feature of a game, or on a private messaging service, can increase the risk of illegal harms including child sexual abuse and exploitation (CSEA) and harassment. We expect this to help services interpret the evidence more broadly, and think more systematically about the types of illegal harms that may occur on their service. Additionally, this approach is easy for services to use, as they only need to know what characteristics their service has (e.g. livestreaming or hyperlinks) to determine what risk factors we expect them to assess for different illegal harms. We also expect this structure to be able to be updated with limited burden on services, as we can easily add or remove risk factors, or edit the descriptions for existing risk factors as our understanding of the harm changes. It may be a slightly more involved process to update the Risk Profiles if the priority offences change as we would need to consider the impact across all risk factors. This approach does require services to self-select relevant risk factors. To limit ambiguity or confusion, we have provided services with a draft glossary. would need to provide a linked glossary. While we consider that this option presents a robust summary of the evidence, this approach does not include information on all of the risk factors identified in the Register. It also provides a high-level summary of the risks. nor definitive. We have explained this in the tables themselves, and additionally have ensured that services are guided to consult Risk Profiles at an early stage of the Risk Assessment Guidance, emphasising that they serve as an *starting point* for services to think about what role different characteristics of their service may play in increasing or decreasing the risk of different illegal harms occurring.

[147] This is because the range of characteristics on Search services is narrower than on U2U, and there is less evidence available (including relatively limited information on the links between individual Search risk factors and specific kinds of illegal harms).

Options	Relevant considerations
Option 2: **Risk Profiles by service type**	An approach presenting different Risk Profiles based on U2U service type, for example a 'Social Media' risk profile and a separate 'Online Adult Services' Risk Profile would be intuitive to use because the concepts are widely recognised. This would have the advantage of allowing us to group together similar services (often with similar functionalities) at a high level to point out the kinds of illegal harms most typically associated with that service type. For example, we could highlight the risk for U2U online adult services and image-based sexual harms like extreme pornography, CSAM and intimate image abuse. However, we discarded this approach because of the following issues relating to evidence and ease of use. First, the evidence base by service type is often not robust or consistent enough to structure Risk Profiles around them alone. Some service types have considerable evidence, whilst others are limited, especially when assessing individual offences. Second, service definitions are not homogeneous or common – some are quite narrow, and several well recognised service types (e.g. social media services and online adult services) contain a complex and wide range of characteristics, each with different levels of risk. Thirdly, the evidence may define service types differently making it difficult to develop consistent definitions from the evidence base. In terms of ease of use, we found that this option did work well in cases where services have clear and specific purposes that map well onto a service type, as with the example above on online adult services. However, some services have what appears to be different U2U service types as part of one integrated user experience. Others may identify themselves as a different type than their users may or they may feel they cannot identify with any of the Risk Profiles presented – in either case, this could lead them to exempt themselves from considering relevant risks. We therefore considered that Risk Profiles structured by service type may lead to incorrect considerations of risk for both Ofcom and regulated services.
Option 3: **Risk Profiles by kind of illegal harms**[148]	The Act already requires services to consider the risk of all offences in scope. An approach providing different Risk Profiles based on the priority offences (for example, a Risk Profile for Intimate Image Abuse or CSEA) would align with the organisation of our evidence, and the kinds of illegal harms in the Act. It would also ensure services could find relevant information for each kind of illegal harm as there would be a Risk Profile for every kind of illegal harm they are expected to assess. However, we discarded this approach because of ease of use. This is because services would have to use the illegal harms as a starting point to think about risk, rather than first consider the role of their service characteristics. For example, a service would have to have prior knowledge of the risks associated with terrorism to determine if the 'Terrorism Risk Profile' was relevant to them. Services without accurate prior knowledge may make inaccurate judgements about what Risk Profiles apply to them – both under and overstating the relevance of certain offences. Given the range of resources and kinds of services in scope of the Act, we considered this would not be a workable approach. Finally, this option would make it difficult to keep up to date to reflect changes to priority offences, or to the underlying criminal laws, as they are based on our current understanding in relation to risks. We therefore considered that Risk Profiles structured by illegal harms type may lead to incorrect considerations of risk for regulated services.

[148] This option would use the same 15 grouping of kinds of illegal harms as is outlined in Appendix B of Annex 5. This corresponds to the chapters contained within the Register of Risks (Volume 2 Chapter 6).

Options	Relevant considerations
Option 4: **Risk Profiles by groupings of similar functionalities**	An approach providing different Risk Profiles based on functionality groups would be easy to use, as it would align with the way we have organised our evidence within the chapters and uses terms familiar to services. It would also be useful in that it would allow us to draw out links in the evidence across different types of illegal harms with regards to functionalities, e.g. the sharing of CSAM and extreme pornography. For example, we could provide services with a specific Risk Profile covering user networking functionalities (e.g. user search, user connections) and user navigation functionalities (e.g. content search, hyperlinks). However, we are not proposing this approach because it would not easily integrate the evidence on how individual functionalities within the same group may have different associations to either the same or different illegal harms. For example, livestreaming and commenting are both within the 'user communication' group and have distinct links to risk throughout the Register. Further, this approach would not present a robust view on the evidence related to risk factors associated with user base, business model or other characteristics. This is because it would all have to be structured through functionality groupings, and the evidence indicates that many of these risk factors interact with all functionality groups.[149] We therefore considered that Risk Profiles structured by illegal harms type may lead to incorrect considerations of risk for both Ofcom and regulated services.

Source: Ofcom analysis

Provisional conclusions

9.92 Services must take their risk profiles into account when carrying out a risk assessment. The proposed approach to Risk Profiles adapts to each service and comprises, for each of U2U and Search, a list of individual risk factors which are associated with an increased risk of harm. Having considered a number of options, we believe that this approach provides the most effective way for services to take account of matters which may affect risk and will help achieve our policy objectives. The approach enables us to highlight key relevant findings from our Register in a format that we expect to be easy for services to use as a starting point for conducting their own risk assessment.

[149] We retain aspects of this structure in Option 1 – when presenting functionality-based risk factors, we organise them based on the groupings described in this Option, which match those used in the Register. For more information on these functionality groups, see the draft Glossary.

Risk assessment evidence base – what different services should consider

9.93 As explained above, we propose that a key part of carrying out a "suitable and sufficient" risk assessment is ensuring it is based on relevant evidence. We believe that this would help ensure that the assessment accurately reflects risks as they exist on a service and would help provide services with an adequate understanding of the risks to implement appropriate measures in response. Our summary of best practice literature above emphasises that analysing relevant information on risk and harm on a given service is a critical factor in assessing risk accurately and implementing appropriate mitigations.

9.94 In the sections above, we also set out our proposal to adopt a scalable approach which allows services to differentiate based on their size, nature and likely levels of risk. Step 2 of the proposed methodology, "Assess the risks", is where we specify that services should focus on assessing evidence relevant to their service.

9.95 In addition, we also highlighted that Ofcom's draft Risk Profiles will provide services with an important starting point when considering risk, but do not provide information on all of the characteristics of services that can give rise to risk. It is important that each service undertakes its own analysis of its specific risks, using appropriate evidence.

9.96 Ofcom has considered several options on how the draft guidance on further evidence can achieve these objectives. Our proposal is a scalable approach comprising **minimum standards of evidence** that all services must meet when doing their risk assessment and clear guidance on where services need to take a **more comprehensive approach**. We refer to this as the "core and enhanced" approach.

9.97 The rationale and evidence underpinning this proposal is set out below. We first outline why we have proposed this approach, the concepts of core and enhanced inputs, how services should decide what evidence they need, and then explain how we have identified the types of evidence in each list.

Options considered and analysis

9.98 We considered a number of options for implementation of a scalable, evidence-based approach:

a) **Non-prescriptive approach:** Ofcom takes a non-prescriptive approach which emphasises that accountability rests with services to ensure their risk assessment is suitable and sufficient, and they should define their own method to evaluate risks proportionately based on their size, capacity, nature and Risk Profile analysis.

b) **Core and enhanced approach:** All services are advised by Ofcom to consider a "core" set of inputs in their risk assessment. In many cases, consulting the "core" inputs to assess the level of risk of each illegal harm should be enough for services to conduct a "suitable and sufficient" risk assessment. However, to ensure that their risk assessment meet this standard, some services will need to consider additional "enhanced" inputs to help them assess the risk, likelihood and impact of a certain harm appearing on their service. In other words, services which operate in a more complex risk environment would be advised to consider a wider range of sources when doing their risk assessment than services which operate in a simpler risk environment. Under this approach we would not tell services upfront whether they should follow the core or enhanced approach. Rather services would be encouraged to take an iterative approach: once they have reviewed the core inputs and any existing evidence they hold, they should consider whether they have an adequate understanding of the risks. If they do not and further information is required, then they should consult the list of enhanced inputs and identify which types of information from that list could help them improve the risk assessment.

c) **Risk triage and tiering system**: Drawing on the model developed by the Australian eSafety Commissioner and the Digital Trust and Safety Partnership (DTSP), this approach would advise services to undertake a preliminary assessment of their characteristics to identify any key risk indicators (e.g. user base, functionalities) and size (e.g. user numbers, revenue, staff numbers). Services would then be allocated into one of a number of tiers, which indicate the evidence inputs recommended for their risk assessment. The risk triage option could also include UK user number thresholds, as applied under the Digital Services Act in the European Union.

9.99 We assessed the benefits, costs, and other impacts of each approach, including whether it would be likely to meet the policy objectives for the Service Risk Assessment Guidance. For our draft guidance we have assessed that option b, the core and enhanced approach, is most appropriate. Our key considerations are set out below.

9.100 We do not believe option a would be suitable because giving too little guidance in this area risks failing to **raise the standards of risk management** in online safety. Many services need to engage in this activity for the first time. It is therefore important that the chosen approach sets down both minimum standards and clear indications of where a more robust approach is required by services to achieve a “suitable and sufficient” risk assessment.

9.101 Equally, all of our evidence emphasises the **dynamic and complex nature of risk** in the online environment. Risk varies based on a complex interaction of service and user characteristics and is a continually changing landscape. Any system which is particularly rigid or based on fixed thresholds could mean that some higher risk services are steered towards an insufficient approach, whereas some lower risk firms are steered towards an excessive or burdensome approach, based on inflexible, industry-wide boundaries. **Flexibility, scalability and proportionality** are important qualities of option b above option c.

9.102 In our view, option b is the option which best reflects the evidence we have reviewed and will most clearly **help services to fulfil their legal duties**. We expect that the core and enhanced approach will help services to assess the matters described under sections 9(5) and 26(5) of the Act in the way that is most meaningful to their service.

9.103 To achieve the benefits of option b, we also concluded that it was important to emphasise the iterative, step-by-step nature of the approach. This is especially important to ensure that our proposals represent a proportionate approach for smaller services.

Core and enhanced evidence inputs

9.104 Evidence inputs refer to the kinds of information we expect services to consider as they conduct their risk assessment. They include information like Ofcom’s Risk Profiles, user reports, data gathered from content moderation systems, or views from relevant experts.

9.105 Under step 2 of the proposed risk assessment methodology, services will assess this evidence and may develop and iterate their approach to consider a wider range of evidence to build a sufficient understanding of risk.

9.106 We propose that core inputs include relevant information specified by the Act – namely Risk Profiles, user complaints, user data (including on age) – and other types of information that all services in scope can easily access and assess, such as retrospective analysis of incidents of harm. In addition, services following the core approach should review any other evidence they hold which is relevant to their risk assessments. Omitting this information may mean that the risk assessment fails to be suitable and sufficient.

9.107 It is rational and reasonable to ask all services to consider these types of information. Services should hold this evidence or be able to access it easily, and no incremental costs are likely to be incurred as a result of Ofcom’s approach. Failing to account for this information could risk that their risk assessment would fall short of the suitable and sufficient standard.

9.108 The enhanced inputs represent a more robust list of recommended evidence types which have been drawn from evidence of industry practice. These inputs are most relevant for services whose characteristics may result in a more complex or significant manifestation of risk. Services are likely to have access to the core inputs already but may need to do further research or gather additional evidence for the enhanced inputs. Examples include reviewing external expert studies, user behaviour research, or engaging relevant representative groups. One scenario in which the enhanced inputs would be relevant is where analysis of the core inputs resulted in ambiguous results about a service's risk levels. For example, where a service consulted Ofcom's Risk Profiles and found it had multiple risk factors for a particular harm but where user reporting data provided no or limited indications of that harm existing on the service.

9.109 As noted above, in many cases we expect that some services will already hold evidence inputs which feature on the enhanced list. If they do, they should include them in their risk assessment.

9.110 All inputs should help services to assess the matters described under sections 9(5) and 26(5) of the Act to a suitable and sufficient standard.

Draft guidance on how to use the core and enhanced inputs

9.111 The draft guidance includes a section on how services should decide what evidence to collect. This is intended to support the scalable approach and to assist services when they consider which enhanced inputs they should consider.

9.112 We state that the guiding principle when deciding what evidence to collect should be whether it will improve the accuracy of the service’s risk assessment and its understanding of harm. Every service should consider the core inputs when they are assessing the likelihood and impact of each illegal harm on their service. Thereafter, services must decide on whether they have sufficient information to reach accurate conclusions on the level of risk for each kind of illegal harm. This can be an iterative process.

9.113 In the draft guidance, we set out a sequence of guiding points, designed to help services decide if gathering further evidence will help them to carry out a more accurate risk assessment. These points are based on the following factors:

a) **The service's risk factors**: All services must consult the Ofcom Risk Profiles and this will enable them to identify some of the risk factors on their service and how they may affect risk. The number of "additional risk factors" provides an initial indication of risk (with more risk factors in general suggesting a potential higher level of risk). A service should assess this indication against the core evidence and, if it is not corroborated, consider gathering additional enhanced evidence inputs.

b) **Ability to answer prompt questions regarding likelihood and impact in the draft Risk Assessment Guidance**: The risk assessment guidance contains prompt questions to help services make decisions on the likelihood and impact on illegal harms. If they are unable to answer these questions with confidence based on the core evidence alone, they should consider gathering additional enhanced evidence inputs.

c) **Confidence in the accuracy of their conclusions**: Based on the information from the Risk Profiles and their consideration of the core evidence, the service should establish how confident they are in the accuracy of their conclusion on the level of risk. The guidance provides a Risk Level Table to help services assign themselves a risk level for each kind of illegal harm. If the core evidence inputs does not provide enough information to use the table and assign themselves a risk level confidently, they should consider gathering additional enhanced evidence inputs to help them reach this conclusion.

d) **Consideration of if/how the enhanced inputs could improve the assessment**: Different sources of information will be more relevant or valuable to different kinds of services or risks (e.g. product testing may provide useful insight on the effect of a recommender system, while independent research may offer more insight on how children are affected by grooming). Services should consider which enhanced inputs could improve the accuracy of their assessment to both decide whether to collect additional evidence, and to decide which inputs to focus on.

   i) The Register of Risks is a good starting point for services because it provides greater context on how harm can manifest on a service and how we identified the risk factors highlighted in the risk profiles. It may help them reach a more accurate conclusion on risk or help highlight where further evidence is needed.

e) **Expectations for larger services**: All else being equal, we will generally expect services with larger user numbers to be more likely to consult the enhanced inputs (unless they have very few risk factors and the core evidence does not suggest medium or high levels of risk). This is because the potential negative impact of an unidentified (or inaccurately assessed) risk will generally be more significant, so a more comprehensive risk assessment is important. In addition, larger services are more likely to have the staff, resources, or specialist knowledge and skills to provide the information, and are more likely to be the subject of third-party research.[150]

9.114 Services should focus on reaching objective, well-evidenced conclusions. In circumstances where Ofcom review risk assessments as part of our implementation of the regulations, we will be looking at whether services have made sound, justified decisions when assembling their evidence base.

[150] We note that some services, with large reach, do not necessarily have many risk factors or have more staff or resources (e.g. not for profit services).

Core inputs

9.115 Our approach to defining the list of inputs included:

a) Identifying a list of potential inputs, based on information specified in the Act, our review of best practice literature, and evidence from industry.

b) Assessing the incremental benefits and costs of the inputs. Owing to the limitations on available data on financial costs, we assigned benefit and cost categories (low/medium/high) based on a qualitative assessment of the nature and the expected order of magnitude of the benefits and costs.

c) Assigned relevant inputs to the core or enhanced list. Core inputs were those inputs with low cost and medium/high expected benefits. Other inputs were assigned as enhanced, noting the iterative nature of the risk assessment process, with services guided to consider what additional inputs they need to consider to carry out a suitable and sufficient risk assessment.

9.116 We consider that it is proportionate for all services to use the proposed core inputs as a minimum standard, based on which they can decide whether they need to do more. They will deliver medium or high benefits in terms of helping the service to identify and understand the risks on their service. On the other hand, we expect the incremental cost of our proposal to be low because it either require services to undertake activities that are necessitated under the Act or requires them to use/process any information or data that they might already have.

9.117 The core inputs are set out in **Table 9.4** below. Descriptions of each are included in the draft Risk Assessment Guidance.

Table 9.4: Considerations for core inputs

Input description	Benefits	Summary evidence & conclusion
Risk Profiles		
Services are required by the Act to consult Ofcom's Risk Profiles. We advise that services consult the relevant parts of Ofcom's Register of Risks (as necessary) to help them assess the kinds of harms highlighted by the relevant Profile.	The Act requires services to take account of the relevant Risk Profiles, which have been informed by the Register of Risks. The Register has been extensively researched and developed specifically to inform Ofcom's and services' understanding of the risks of illegal content. We advise services to consult the Register of Risks where they want to deepen their understanding of specific harms and the more complex interaction of risk factors synthesised in Risk Profiles.	This input is rooted in the requirements of the Act and there is a logical connection to services' risk assessment. The Register of Risks is itself a summary of a range of relevant evidence for online services
User complaints, including user reports		
The Act stipulates that services must "operate a complaints procedure in relation to a service that—(a) allows for relevant kinds of complaint to be made […], (b) provides for appropriate action to be taken by the provider of the service in response to complaints of a relevant kind, and (c) is easy to access, easy to use (including by children) and transparent." In addition, the Act sets out a duty for all services to operate "systems and processes that allow users and affected persons to easily report content which they consider to be content of a kind specified below [illegal content]."	Incorporating user complaints about illegal content (or other complaints received under the online safety complaints duties) as an input in a risk assessment will enable services to consider where users are dissatisfied with the service and identify possible areas where existing mitigations to combat illegal harms are proving ineffective. Complaints and reports provide a clear and direct channel through which users can communicate their experience of illegal harms, and therefore assist the service in assessing likelihood and impact.	This input is rooted in the requirements of the Act and there is a logical connection to services' risk assessment.

Input description	Benefits	Summary evidence & conclusion
Relevant user data including age		
The Act requires services to consider their user base as part of the risk assessment process. Services should therefore assess its risks based on relevant user data. According to the Act, user data includes "(a) data provided by users, including personal data (for example, data provided when a user sets up an account), and (b) data created, compiled or obtained by providers of regulated services and relating to users (for example, data relating to when or where users access a service or how they use it)." We consider that user data would include any data held as a result of age assurance and age verification processes.	Assessing the user base is a specific requirement of the risk assessment duties and user data, in combination with other inputs into the risk assessment, will help services understand if any particular groups are at risk of certain kinds of illegal content on their service. For example, Ofcom's Register of Risks and Risk Profiles indicate that children, women and users from historically marginalised groups are likely to be at greater risk of certain illegal harms.	This input is rooted in the requirements of the Act and there is a logical connection to services' risk assessment. Any assessment of users' personal data (including any data that is not anonymised), will require services to comply with their obligations under UK data protection law. Services can use aggregated, anonymised data to assess risks for particular groups of users, but this will likely be based on data collected/inferred about individual users. Services will need to make judgments on the data they hold to ensure it is processed lawfully, including providing appropriate transparency to users when the data is collected or further processed.
Retrospective analysis of incidents of harm		
Retrospective analysis or 'lessons learned' following incidents of harm can be directly relevant considerations for a risk assessment. Services should have some kind of process in place to diagnose where and how things went wrong following any significant instances of harm. A significant incident could include, for example, a major incident that causes serious harm, a prominent trend in illegal content, or	Retrospective analyses will help services assess the impact of different kind of illegal harm, particularly those harms which are less common but high impact. Such case studies may allow services to examine how particular aspects of the service's design (such as user characteristics, functionalities, recommender systems) may have played a role and where mitigating measures (such as content moderation,	This input is a rational step in the risk assessment process and does not require significant investment beyond analysis as part of the assessment. It is in-keeping with the requirements that risk assessments be "suitable and sufficient" and that services take appropriate steps to keep them "up to date".

Input description	Benefits	Summary evidence & conclusion
an individual piece of content which becomes widely disseminated.	terms of service, user reporting) and associated processes could have been more effective.	
Any other relevant information held by the service		
Services should also assess any other evidence they hold that is relevant to harms on their service. This could include any existing harms reporting, research held by the service, referrals to law enforcement, data on or analysis of user behaviour relating to harms or product testing. Any types of evidence listed under Ofcom's enhanced inputs (e.g. the results of content moderation, product testing, commissioned research) that the business already collects and which are relevant to the risk assessment, should inform the assessment. In effect, if the service already holds these inputs, they should be considered as core inputs.	Assessing any evidence already held by the service that is relevant to online harm is likely to improve the risk assessment without necessarily incurring significant additional cost. We explain the benefits of the enhanced inputs in the next section. Failing to consider relevant evidence that the service already holds could result in an unsuitable or insufficient risk assessment. For example, the risk assessment may be less accurate as a result of missing important information, a risk may be undiagnosed, or a service's measures to control a risk may be less effective than expected.	This input is a rational step in the risk assessment process and does not require any additional investment beyond analysis as part of the assessment. The benefits and evidence underpinning a range of types of additional evidence are set out under the enhanced inputs. As above, any use of users' personal data (including any data that is not anonymised), will require services to comply with their obligations under UK data protection law. Services will need to make judgments on the data they hold to ensure it is processed lawfully, including providing appropriate transparency to users when the data is collected or further processed.

Source: Ofcom analysis

Enhanced inputs

9.118 Unlike the core inputs, which represent a minimum standard of evidence based on materials the Act identifies and information which it is reasonable to expect the service to hold, the enhanced inputs have been drawn from industry practice bolstered with our own research and that of expert third parties. A summary of evidence is included in the table below.

9.119 These measures are not explicitly required by the Act but will be important for some services to assess the likelihood and impact of each illegal harm, as part of a suitable and sufficient risk assessment. Some services will need to gather additional evidence as part of this process, by putting in place new systems or by organising specific kinds of data from across their business.

9.120 Including the enhanced inputs in the draft guidance helps meet our policy objectives by steering services to improve their understanding of risk on their service and in turn improve their mitigations for these risks and have better safety outcomes. We recognise that services may incur some costs; however, this will be proportionate if they are needed to improve the quality of the risk assessment. The service has the discretion to choose the lowest cost and most beneficial enhanced input(s) to fill any evidence gaps they have to produce a suitable and sufficient risk assessment.

9.121 The enhanced inputs are set out in **Table 9.5** below. More detailed descriptions of each are included in the draft Risk Assessment Guidance.

Table 9.5: Considerations for enhanced inputs

Input description	Benefits	Summary evidence & conclusion
Results of product testing		
We use 'product' as an all-encompassing term that includes any functionality, feature, tool, or policy that you provide to users for them to interact with through your service. This includes but is not limited to whole services, individual features, terms and conditions (Ts&Cs), content feeds, react buttons or privacy settings. By 'testing' we mean services should be considering any potential risks of technical and design choices, and testing the components used as part of their products, before the final product is developed. We recognise that services, depending on their size, could have different employees responsible for different products and that these products are designed separately from one another. Services may consider running tests on individual products ahead of launching them on their wider service. This could include analysis of user behaviour and taking into account the potential impact of behavioural biases. We recommend	Evaluating data and insights gathered from these tests will improve their risk assessment because testing may indicate the effect of any product changes and whether they may increase or decrease the likelihood of illegal content appearing on or being disseminated by the service, and its impact.	Many large services already publish this kind of evidence. For example, Meta publishes information on how it approaches product testing and Snapchat produces broken down privacy assessments of each product on the service, this is publicly available as part of its transparency information. Call for Evidence responses from Google, Yoti and Trustpilot all highlighted the importance of product testing and signalled that this is common practice across the industry in services of a certain size.[151]

[151] Google response to 2022 Ofcom Call for Evidence: First phase of online safety regulation, pages 16, 54, 61; Yoti response to 2022 Ofcom Call for Evidence: First phase of online safety regulation, page 2.

Input description	Benefits	Summary evidence & conclusion
that services which think that they need additional evidence to understand risk on their service to consider adopting and recording findings of product testing with a view to integrating the data into their risk assessment.		
Data gathered through content moderation systems relating to risk levels		
Most services are likely to have a content moderation system in place to detect and potentially remove content that violates their policies. The nature, scope and maturity of these systems varies significantly between services. Services seeking to improve their understanding of and response to illegal content may choose to use a more complex content moderation system which measures more complex kinds of exposure. For example, how long a piece of illegal content is on the service, or the virality of pieces of content, rather than only the number of user reports and steps taken in response.	Measuring and recording this data to feed into a risk assessment can help services to more accurately understand a range of risk characteristics, such as the likelihood of illegal content appearing on the service, its dissemination, and the impact on users (e.g., by accounting for time taken to be removed or the dissemination of the content). Assessing the effectiveness of content moderation decisions and the systems themselves also helps services to understand the level of mitigation provided by this measure in their risk assessment.	Content moderation against community standards and policies is already widespread within industry and forms a key element of online safety practice and trust and safety teams' practice. In response to Ofcom's Call for Evidence, OnlyFans, Google and Wikimedia, for example, outlined three different and iterative approaches to content moderation – human and automated – that have evolved as they have developed their services and informed their view of risk and harm.
Consultation with internal experts on risks and technical mitigation measures		
The guidance will propose that some services should consult with any internal experts on risks and technical mitigations. We know that many large services monitor the efficacy of interventions to reduce harm and address community guideline breaches on their service. A	Consulting experts in risk and harm will help services to better understand the likelihood, nature and severity of potential harm, as well as the factors which drive it. Consulting experts on how technical measures, systems and processes may help address risk can improve understanding	Larger services under the VSP regime and which responded to the online safety call for evidence described current practices which already align with this input. One VSP routinely gathers information from product, legal and trust and safety teams to conduct an in-depth review of

Input description	Benefits	Summary evidence & conclusion
thorough technical examination process for a mitigation should consist of regular thematic technical expert meetings supported with focused follow up work.	of the risk environment and also ensure that the experience of any previous interventions, and their efficacy, is factored into risk assessments and mitigation measures.	terms of service docs once a year. Another described the work of “a dedicated team of hundreds of trained staff, who assist in access control inputs, continually reviews its access control inputs internally, and has engaged a third party to independently review the same.” However, this is an enhanced input because services have vastly different staff resources dedicated to online safety. A number of smaller VSP we regulate have set out the limitations of their resources and capabilities. For some SMEs, we anticipate that the costs of employing dedicated experts for these activities are unlikely to outweigh the benefit (unless evidence suggests that they may be medium or high risk but need to do further work to fill the evidence gaps, and this is the most cost-effective way to achieve this). However, the balance may shift as the business grows.
Consultation with users and user research		
Consultation with users and user research can take many forms, which can be tailored to the service in question. Quantitative surveys, qualitative research, engagement with user groups, behavioural analysis, or dedicated work by user research professionals can allow services	User research is common among product design teams working in the digital sector and serves to enable services to better meet the needs of their users. Applying user research and consultation in the online safety context can improve services' understanding of their users' experience of harm	The user research discipline is used widely in the digital and technology sector and forms a core part of agile, user-centric product design and development. For example, one VSP reported that it has invested in online spaces to gather users' views in a way that influences product

Input description	Benefits	Summary evidence & conclusion
to improve their understanding of their customers' behaviour and the likely impact of any interventions. Engagement could be general or designed to target specific users, such as those with vulnerabilities or in certain age groups.	and how to implement mitigating measures in an effective and user-centric way.	development. This includes conducting surveys with users, alongside interviews and group sessions and social media 'listening' to better understand the responses to changes to policies and features.
Views of independent experts		
Some services should consult external experts about the risk of harm from illegal content on their services, or specialists in techniques to address and mitigate harm. Services should take steps to ensure the quality and accuracy of any third-party advice. Other kinds of expert consultation may also be relevant for services to consider. This could include views of experts on industry trends, regulatory standards and the views of certain trade bodies or technical experts in relevant fields.	Expert consultation will help a service consider how a particular harm manifests online in general and/or on their service specifically, which would in turn help them develop mitigation and management techniques which are targeted and effective. Consultation can also help a service maintain an up-to-date knowledge and understanding of particular risks, harms and mitigations. Consulting external experts may help bolster an otherwise limited evidence base relating to a specific harm or kind of illegal content and their related risk factors. In addition, this kind of input would help services consider targeted issues such as vulnerable groups who may be more likely or more severely impacted by certain kinds of illegal content.	A range of services have highlighted the value of consulting external experts. Ofcom's VSP report cites Bitchute's relationship with TechAgainstTerrorism, which offers expertise and support to reduce terrorist content on the service.[152] One large VSP commissions experts to evaluate its internal risk assessment, while other large platforms convene groups of external experts to provide advice on challenging policy and rights issues. Responding to Ofcom's call for evidence, the Business School for Social Responsibility highlighted the value of consulting experts in a risk assessment process, recommending that services undertake gap analysis between their reporting and complaints mechanisms, and the effectiveness criteria for non-judicial grievance mechanisms contained in the UNGP.

[152] Ofcom, 2022, *Ofcom's first year of video-sharing platform regulation*, Section 10.

Input description	Benefits	Summary evidence & conclusion
Internal and external commissioned research		
Large services often commission research into specific trends or harms which informs their approach to safety and moderation. This can allow services to draw on wider evidence beyond that which is gathered through the operation of their service, accessing additional expert resource can improve the accuracy of the risk assessment. Relevant external research may also include published reports from expert bodies. Examples include research from Ofcom, other regulators, government, academics, policy organisations, and charities or representative groups.	Expert research will allow services to improve their understanding of the factors which may drive the likelihood of illegal content appearing on the service, the impact of that harmful content, and how it may be mitigated effectively. This can bring a range of benefits, such as identifying new and emerging trends, understanding how harms manifests online in ways that may not be visible or apparent to the service, understanding new opportunities to mitigate or manage harm, or learning from the experience of other sectors, services and peers.	A range of services of different sizes have undertaken or commissioned and then published research. Meta regularly publishes such research, for example, reports into how Covid-19 impacted the use of Facebook. Google responded to Ofcom's Illegal Harms Call for Evidence to describe a dedicated Google Safety Engineering Centre in Dublin which worked with experts to tackle the spread of illegal and harmful content. OnlyFans gave information to the VSP regime on how it engages third party independent experts to monitor, assess and validate the design of its safety compliance programme. Grindr has worked with third parties to test machine learning models detecting illegal activity on the service, and published its own reports on gender-inclusive content moderation.
Engaging with relevant representative groups		
Services can engage external organisations representing specific groups to better understand the perspectives of specific users, demographic groups or communities. This may be especially relevant if the service has evidence that certain vulnerable groups will be particularly impacted by illegal content or any aspect of the service's	This input can help services to critically assess the risks to specific kind of users or groups by drawing on the advice and experience of representative groups to bolster their evidence base. This is particularly beneficial for certain groups, such as older people, who may be less	Engaging with representative groups is a cornerstone of policy-making both within and outside the digital and technology sector. One dating service highlighted its advisory council, which includes advocates and expert groups involved in the study and prevention of sexual assault, harassment, sex trafficking and other

Input description	Benefits	Summary evidence & conclusion
design, including planned design changes which require a risk assessment**.**	familiar or able to use functions on the platform such as reporting or complaints meaning that their experience of harm is underrepresented in the core inputs.	issues which are particularly relevant to platforms related to online dating. In the US, the service has partnered with an anti-sexual violence organisation to inform its thinking around reporting, moderation and response policies and procedures.
Outcomes of external audit or other risk assurance processes		
Services seeking to improve their confidence that their trust and safety processes or wider risk management systems are robust may commission a third party to audit aspects of their service or undergo another form of risk assurance process.	Independent audits can provide insights and analysis which services are unable to produce or assure themselves. They offer services the opportunity to be robustly assessed and to identify new ways of improving their trust and safety processes. Services and any third-party suppliers should take steps to ensure that any methodology applied is robust and that the assurance process provides an independent and objective assessment of performance and recommendations for improvement.	This practice has already been adopted among several larger services and provides additional, objective assessment of online safety measures. For example, the Internet Commission produces annual Accountability Reports which summarise its work to auditing online services and their organisational structures, systems and processes. The process is designed to help services understand where they can improve practices using an evaluation framework for digital responsibility. In recent years, Meta has committed to publishing an independent, third-party assessment of the metrics and reporting methods in its Community Standards Enforcement Report. An independent auditor was appointed and a third-party framework applied for the assessment.

Source: Ofcom analysis

Provisional conclusions

9.122 We believe that the proposed scalable approach, comprising core and enhanced evidential inputs into the risk assessment, is relevant to the legal duties and will help achieve our policy objectives. The flexible and iterative nature of the recommended approach will ensure that the "suitable and sufficient" requirement can be met in a proportionate way by regulated services, while allowing governance bodies and Ofcom to effectively scrutinise the quality of risk assessments.

Up-to-date risk assessments and significant changes to services

9.123 This section focuses on how we are guiding services to meet their duties to keep their risk assessment up to date. The Act outlines the following duties regarding when services should review or carry out a new risk assessment:

a) a duty to take appropriate steps to keep an illegal content risk assessment up to date.

b) a duty to update their risk assessment if Ofcom makes any significant change to a Risk Profile that relates to the service of the kind in question.

c) a duty to carry out a further suitable and sufficient illegal content risk assessment relating to the impacts of that proposed change before making any significant change to any aspect of a service's design or operation.

Proposed approach in guidance: appropriate steps

9.124 Services should decide their own policy for reviewing the risk assessment and recording it. However, we expect services to be able to explain their approach and the steps they are taking to meet this duty. At a minimum, to meet the duty to take appropriate steps to keep an assessment up to date, we propose that:

a) Services have a written policy in place to review their assessment at least every 12 months; and

b) Services name a responsible person for overseeing this process.

9.125 The benefit of reviewing the assessment is to ensure that the analysis remains accurate and up to date considering incremental changes made to the service by the provider, or other changes such as in user behaviour, the emergence of unexpected risks or technological changes. While a service may not have made a change big enough to amount to a significant change that requires a new risk assessment,[153] several smaller changes to the service design or operation or environment mean that the assessment could be

[153] See our proposal on significant change below.

inaccurate and needs review so that it remains suitable and sufficient. Therefore, updating the assessment through regular reviews will improve a service's ability to identify and mitigate new risks and lead to better safety outcomes.

9.126 While reviewing a risk assessment will include costs such as staff costs, it should not be as burdensome as carrying out a new risk assessment. The review should use the existing risk assessment and take account of any new inputs or evidence to update it. The draft guidance relating to how to complete a review makes it clear that the steps we expect services to take are not as burdensome as carrying out a new risk assessment and is not very directive to give services the discretion to meet their legal duties in the most appropriate and proportionate way.

9.127 We provisionally conclude that our proposal for services to review the assessment *at least every 12 months* is the appropriate period to meet their legal obligations. It will help ensure that services can stay up to speed with the rapid and dynamic changes in the risk environment, thus making them better placed to address emerging harms. It is also in line with services' duty under the Act to carry out a Child Access Assessment "not more than one year apart" and is aligned with comparable regimes internationally, such as the Digital Services Act. In addition, it aligns with common governance cycles and accounting periods.

9.128 While services will need to incur costs anyway to meet their legal obligations to keep their assessment up to date, we recognise that the costs of our proposal are greater than would be the case if we allowed for less frequent review. However, we consider that if services reviewed less frequently than every 12 months, they could not be confident of maintaining a suitable and sufficient risk assessment in a changing risk environment, particularly given the fast-moving pace of online technology. This could pose a material risk to users and services could not be confident that they are meeting their legal duties.

9.129 We also recognise that recommending a maximum period between reviews of less than 12 months may improve the ability of services to identify new risks. However, we do not consider it proportionate to propose a shorter maximum period between reviews at this stage. This would increase costs and it is not clear whether its benefits would outweigh the additional cost burden. Therefore, it is more appropriate to give services the discretion to consider when they need to update their risk assessment within the 12 months period to comply with their legal duties. This point is strengthened because there are a number of other safeguards built into our policy on risk management which would allow for detection of risks that emerged in year: i) the requirement to review the risk assessment when risk profile changes; ii) requirement to carry out a new risk assessment before the service makes a significant change, and iii) for larger services, governance measure on monitoring for new illegal harms.

9.130 Based on the above, we provisionally conclude that our proposal is a proportionate way for services to meet their legal duties to take steps to keep their assessment up to date.

Proposed approach in guidance: a change to Risk Profiles

9.131 Services must also review their risk assessment if Ofcom makes a significant change to a relevant Risk Profile. Services only need to do this if the change Ofcom makes to the Risk Profile(s) is relevant to their service and impacts their assessment of risk.

9.132 To do this, services should use their most recent recorded risk assessment and review their analysis of their risk factors of each relevant harm and consider if any part of the assessment needs to be updated.

9.133 Reviewing the risk assessment following changes to risk profiles will involve costs; however, this is a direct requirement by the ACT. In any case - like the annual review - this process should be less burdensome than carrying out a new risk assessment as it should focus on taking account of the specific changes made by Ofcom to the Risk Profile.

Proposed approach in guidance: a significant change

9.134 To assist services in complying with this duty, we have provided draft guidance to help services decide if their proposed change is significant or not, and whether it therefore triggers the specific legal requirement to carry out a new risk assessment relating to the change.

9.135 The draft guidance offers broad principles supported by specific examples across a range of factors which may impact a services' design or operation in a way which can materially impact risk on the service. We opted for using a principle-led approach to give services flexibility as what amounts to a significant change can vary across the wide range of services in scope. We consulted with experts internally and externally to help understand the circumstances in which a change to a service may be significant enough to cause the risk assessment to become out of date and no longer provide a suitable and sufficient assessment of risk on the service.[154]

9.136 In particular, we propose that the kinds of changes we expect to amount to a significant change and trigger this duty include but are not limited to:

a) A proposed change which alters the risk factors which a service identified in its last risk assessment;

b) A proposed change which impacts a substantial proportion of a service's user base or changes the kind of users it expects to use the service;

c) A proposed change which impacts a vulnerable user group, such as children;

d) A proposed change which impacts the efficacy of the safety measures it has put in place following its last assessment to reduce the risk of illegal content appearing on its service; and

[154] This approach aligns with the methodology set out in the guidance for how to do a suitable and sufficient risk assessment, particularly regarding using risk factors and evidence inputs to assess risk.

e) A proposed change which impacts the service’s revenue model, its growth strategy and/or its ownership in a way that affects its service design.

9.137 When considering any of the principles above, we have directed services to consider the size of the service’s user base. For instance, as explained above in relation to assessing impact (in step 2 of the methodology), a relatively minor change on a large service is more likely to have a significant impact, while it could take a much larger change on a smaller platform to trigger the need to review their risk assessment.

9.138 The draft guidance is intended to help services to focus on the impact of a proposed change, to help them understand if that change is significant. It gives significant weight to the size of the service as it recognises that a small change on a large service could result in similar or greater impact than a large change on a small service. A real world example of the kinds of change we hope to capture in a large service could be, in 2018-19 Facebook made a significant change to its recommender system operations by introducing a new “meaningful social interaction” success metric, which would have resulted in significant changes to what users see. Further, based on the responses to our Illegal Harms Call for Evidence and wider engagement with industry, we understand that the larger and more complex a service may be, the more likely it is to have routine updates or system changes which we did not feel it was proportionate to capture under this duty.

9.139 We consider that our proposal will bring significant benefits as it will help services to meet their legal duties and improve safety outcomes for users. The proposed significant changes could potentially result in a material change to the risk environment and if a service did not consider them by carrying out a new risk assessment, there is potential that they would fail to identify and assess important new risks, and this could result in a risk of material harm to users.

9.140 The duty to carry out a new risk assessment before implementing a significant change to any aspect of its design or operation has the potential to be more burdensome and incur greater costs compared to the above duties to update the risk assessment. However, this is a direct requirement of the Act.

9.141 We provisionally conclude that our proposed “significant changes” to consider are necessary in order for services to be confident that they are complying with their legal duties, hence any associated costs are proportionate and are primarily based on the requirements of the Act, rather than on regulatory choices made by Ofcom. This is particularly given we have adopted a principle led approach (rather than directive) which affords flexibility to services to help them meet this duty as appropriate relative to its size, capability and specific circumstances that may affect risk. Overall, we think this approach is proportionate for services to help them meet a specific duty set out in the Bill.

Provisional conclusions

9.142 We believe that the proposed approach to reviewing, updating and carrying out new risk assessments is relevant to services’ legal duties and will help achieve our policy objectives. Our proposed approach is flexible and intended to be proportionate in terms of the size of the service and the risks they may create. We do not consider that our proposed approach creates any significant costs beyond what is necessary to comply with the requirements of the Act, and that there are clear benefits of our approach in assisting firms to achieve compliance, appropriately assess risks, and take steps to improve safety online.

10. Record keeping and review

What is this chapter about?

Providers of regulated U2U and search services have duties to make and keep written records of their risk assessments and the measures they take to comply with several duties set out in the Act, as well as regularly reviewing their compliance with relevant duties specified in the Act. This chapter introduces our proposed guidance about how services can fulfil these duties.

What are we proposing?

We are making the following proposals for all U2U and search services:

- **Written records can be made and kept in a durable medium of the service's choice.**
- **Where reasonably practicable, written records should be kept in English (or for services based in Wales, in English or Welsh**.
- **Written records are written in as simple and clear language as possible**.
- **A written record must be kept of current risk assessments and compliance measures and must be updated whenever a significant change is made**.
- **There are additional record-keeping requirements if the service takes alternative measures** to those set out in Ofcom's Code of Practice.
- **Written records should be retained in accordance with the service's record retention policies, or a minimum of five years, whichever is the longer**.
- **Reviews should be scheduled by services and occur with a frequency that allows for a continuous cycle of implementation, monitoring and review**.
- **Our expectation is that services should undertake a compliance review at least once a year,** but more frequent reviews may be appropriate if the regulated service becomes aware of compliance concerns, or implements new measures. Services should also carry out a compliance review if there is a significant change to any aspect of the design or operation of the service.

We are not proposing to exercise our power to exempt specified descriptions of services from the record keeping and review duty.

Why are we proposing this?

Our proposed guidance seeks to strike a proportionate balance between: accommodating the wide variety of services captured by the guidance; the need for Ofcom to have easy-to-understand, clear and sufficiently detailed written records; and minimising any unnecessary cost or burden on services.

What input do we want from stakeholders?

- Do you have any comments on our draft record keeping and review guidance?
- Do you agree with our proposal not to exercise our power to exempt specified descriptions of services from the record keeping and review duty for the moment?

Introduction

10.1 Providers of regulated user to user services and regulated search services are required to keep records of the measures they take to comply with some of their new duties and also to review them regularly. Ofcom is required to produce guidance to assist them to do so.

10.2 This chapter summarises the key aspects of the guidance at Annex 6 that we are proposing to produce. It also explains which record-keeping duties the proposed guidance covers and those that it does not. Record-keeping duties which are not covered by the proposed guidance will be covered by other guidance to be produced by Ofcom, as we explain in this chapter.

Scope of our proposed guidance

10.3 The draft guidance covers the record-keeping and review duties that apply to service providers under sections 23 and 34 of the Act. These are the duties to:

a) **keep written records** of their risk assessments and the measures they have taken to comply with certain duties specified in sections 23 and 34 of the Act, including the illegal content safety duties in sections 10 and 27 of the Act and the children's online safety duties in sections 12 and 29 of the Act. The specified duties are referred to as 'relevant duties';[155] and

b) **review** regularly and as soon as reasonably practicable, after making a significant change to their service, their compliance with the duties specified in sections 23 and 34 of the Act. We refer to these as the 'relevant review duties' and they include the duties in section 18 in respect of news publisher content and the duties in sections 71 and 72 in

[155] A 'relevant duty' for regulated user-to-user services means the duties set out in: section 10 (Illegal content); section 12 (children's online safety); section 15 (user empowerment); section 17 (content of democratic importance); section 19 (journalistic content); section 20 (content reporting); and section 21 (complaints procedures). A 'Relevant duty' for regulated search services means the duties set out in: section 27 (illegal content); section 29 (children's online safety), section 31 (content reporting) and section 32 (complaints procedures).

relation to terms of service, in addition to the duties to which the record-keeping duties apply.[156]

10.4 In our draft guidance, we provide guidance on the form that records should take, the matters that they should cover and when they should be made. We also provide guidance on the frequency with which providers should review their compliance with the relevant review duties and factors they should consider when deciding whether to conduct a review.

10.5 The draft guidance does not cover the record-keeping duties that apply to providers which provide an online service on which pornographic content is published or displayed by or on behalf of that provider ('Part 5 providers').[157] Guidance on these duties will be included in draft guidance for Part 5 providers that we expect to issue in December 2023.

10.6 The draft guidance at Annex 6 also does not provide specific guidance on the duties in section 36(7) and sections 23(2) and 34(2) of the Act to keep a written record of, respectively, of every children's access assessment conducted under section 36 and every children's risk assessment carried out under section 11 or section 28 of the Act. We expect to issue guidance for consultation on the duties under sections 36, 11 and 28 in March 2024 and will update our draft record-keeping guidance as appropriate at the same time.

10.7 We also expect to issue in March 2024 draft guidance for consultation under section 53(1) of the Act on content which we consider to be harmful to children, which will be relevant to providers' duties under sections 12 and 29 in relation to children's online safety. If we consider appropriate, we may update the draft guidance on the review duties as it applies to duties under sections 12 and 29 of the Act.

10.8 Finally, Ofcom is required to issue guidance in respect of other relevant review duties, including the user empowerment duty in section 15, the news publisher content duty in section 18, the terms of service duties in sections 71 and 72 and the duty in section 75 requiring the disclosure of information about the use of service by a deceased child. If we consider appropriate, we will revise the draft guidance to reflect the guidance we issue in respect of these relevant review duties.

[156] 'Relevant review duties' for regulated user-to-user services means the duties set out in: section 10 (illegal content); section 12 (children's online safety); section 15 (user empowerment); section 17 (content of democratic importance); section 18 (news publisher content); section 19 (journalistic content); section 20 (content reporting); section 21 (complaints procedures); section 71 and section 72 (terms of service) and section 75 (disclosure of information about use of service by deceased child users). 'Relevant review duties' for regulated search services means the duties set out in: section 27 (illegal content), section 29 (children's online safety), section 31 (content reporting), section 32 (complaints procedures) and section 75 (disclosure of information about use of service by deceased child users).

[157] Part 5 providers are subject to the duties in Part 5 of the Act. These include duties to ensure that children are not normally able to encounter pornographic content on their online service and to keep a record of the kinds of age verification or age estimation used and how they have been deployed.

Proposed guidance on the record-keeping duties

10.9 Good, well maintained and clear records and regular, timely reviews will assist service providers to keep track of how they are complying with their relevant duties and ensure that the measures that they have taken are fit for purpose. The records will also provide a useful resource for Ofcom in monitoring how the relevant duties are being fulfilled.

10.10 For the most part, we consider that our proposals in relation to appropriate record keeping are self-explanatory and we expect most providers to do this as a matter of course in operating their businesses. Accordingly, we do not set out different options in detail below, but simply state, in line with the requirements of the Act, what amounts to appropriate record keeping. We have specified certain retention periods, which we have proposed by reference to what we consider to be reasonable, taking account of the potential for such records to be required for monitoring or enforcement purposes.

10.11 Accordingly in our draft guidance, we are proposing that:

a) the records that providers must make and keep should be durable, easy to understand and up-to-date;
b) where reasonably practicable, records should be kept in English (or for providers based in Wales, in English or in Welsh). If this is not reasonably practicable, the records must be capable of being translated into English;
c) records should be updated to capture changes to a risk assessment or Code measure, but earlier versions should be retained so the provider is able to provide both current and historic records of how it has complied with the relevant duties;
d) records which are no longer current should be retained for a minimum of five years, unless the specific record has been provided to Ofcom. We consider this retention period is reasonable, given the potential for issues to arise after the event, and which may lead Ofcom to require the relevant record of the risk assessment or compliance measure taken at the time.

Risk assessments

10.12 Our draft guidance includes specific guidance about the records that service providers must make and keep about their risk assessments. This should be read in conjunction with the draft Services Risk Assessment Guidance at Annex 5.[158]

10.13 Providers are required to make and keep a written record of all aspects of every risk assessment they carry out, including details about how the assessment was carried out and its findings. We propose that the record should include how the provider has considered the required elements in section 9, section 11, section 26 or section 28 (as applicable) of the Act and the evidence the provider has relied on to assess the risks posed by the

[158] The Services Risk Assessment Guidance applies to the duties to carry out risk assessments in sections 9 and 26 of the Act. As set out in paragraph 10.6, we expect to issue for consultation draft guidance on the children's risk assessment duties in sections 11 and 28 of the Act in March 2024 and propose to issue further draft guidance for consultation on the record-keeping duty as it applies to children's risk assessments at that time.

provider's service. This should help to demonstrate how a provider's risk assessment is suitable and sufficient.

10.14 In relation to a record of a risk assessment made under section 9 or section 26, we propose that it should cover the matters set out at paragraph A5.201 of the proposed guidance, including:

a) confirmation that the service has consulted Ofcom's Risk Profiles, for example by recording the outcomes of the Risk Profiles questionnaire at Appendix A of the Services Risk Assessment Guidance;
b) a record of any risk factors from Ofcom's Risk Profiles which are relevant to the regulated provider's service;
c) a list of the evidence that has informed the assessment of likelihood and impact of each kind of priority illegal harm;
d) confirmation that the findings of the risk assessment have been reported through appropriate governance channels; and
e) information regarding how a service takes appropriate steps to keep the risk assessment up to date (for example, a written policy).[159]

10.15 We set out our rationale for a record of a risk assessment to include these areas in Table 9.2, Section 9 (Services Risk Assessment Guidance).

10.16 Our expectation is that the written record of a risk assessment or any revision to it will be made contemporaneously to ensure it is accurate and up-to-date.

10.17 Providers of Category 1 user-to-user services and Category 2A search services are required to provide in full to Ofcom their written records of risk assessments as soon as reasonably practicable.[160] We are proposing that these should be sent to Ofcom's dedicated email address (as published on Ofcom's website at the time of submission) as soon as the risk assessment, or any revision to it, is concluded.

Code measures

10.18 Providers must keep a written record of any measures taken to comply with a relevant duty[161] which are described in a code of practice issued by Ofcom under section 41 of the Act. In addition to describing each code measure that has been taken or is in use, we

[159] See Annex 6 (Record Keeping and Review guidance), paragraph A6.13 and Annex 5 (Services Risk Assessment Guidance) for more details.

[160] Section 23(9) and section 34(9) respectively. The Secretary of State is to make regulations under Schedule 11 of the Act, specifying the thresholds for user-to-user and search services to qualify as Category 1 and Category 2A providers respectively. Where Ofcom considers that a regulated provider meets these conditions, it must enter them in a register in accordance with section 96 of the Act.

[161] A 'Relevant duty' for regulated user-to-user services means the duties set out in: section 10 (Illegal content); section 12 (children's online safety); section 15 (user empowerment); section 17 (content of democratic importance); section 23 (journalistic content); section 20 (content reporting); and section 21 (complaints procedures). A 'Relevant duty' for regulated search services means the duties set out in: section 27 (illegal content); section 34 (children's online safety), section 31 (content reporting) and section 32 (complaints procedures).

propose that the record should include the relevant code of practice and the date on which the measure takes effect.

10.19 Where a measure described in a code of practice provides for a document to be made or information to be recorded (such as a policy document or statistical records), we propose that the document or information should be kept and maintained in accordance with our draft guidance as part of the record for the purposes of the record-keeping duty under section 23(3) or section 34(3) of the Act.

10.20 We propose that the written record of a Code measure should be made promptly after the measure has been taken or, where the measure is already in effect prior to the relevant duty coming into force, promptly after this date.

Alternative measures

10.21 It is open to a service provider to take alternative measures to those described in a Code of Practice, as appropriate to comply with a relevant duty. Where a provider opts to take an alternative measure, its record of the measure in question must cover the matters specified in section 23(4) and (5) and section 34(5) and (5) of the Act (as applicable).

10.22 We propose in the draft guidance that the record should be made promptly after the alternative measure has been taken or, where the measure is already in effect prior to the relevant duty coming into force, promptly after this date.

Proposed guidance on the review duties

10.23 Service providers are required to review compliance with the relevant review duties regularly and as soon as reasonably practicable after making any significant change to any aspect of the design or operation of the service.

10.24 Our draft guidance provides general guidance on matters that the service provider should consider when conducting a review and the frequency of such reviews. We propose that as a minimum, service providers should conduct a compliance review at least once a year, in line with the financial and annual reporting cycle for UK companies, as well as being in line with our guidance on service providers reviewing their risk assessments.[162] Where a service provider becomes aware of compliance concerns or implements new measures to comply with a relevant review duty, it may be appropriate to conduct earlier or more frequent reviews.

10.25 We also propose guidance about when there is likely to be a significant change to a regulated service, triggering a review under section 23(6)(b) or section 34(6)(b) of the Act, by reference to our draft Services Risk Assessment Guidance on when there is such a change for the purposes of the risk assessment duties. We consider that this latter guidance is likely to be of assistance to service providers when considering whether there is a significant change which may affect a provider’s compliance with other relevant review

[162] See Annex 5 (Services Risk Assessment Guidance), paragraphs A5.121-A5.123.

duties. As noted, we will be producing guidance on other relevant review duties, such as the user empowerment duty and the new publisher content duty, and therefore may issue further guidance on the question of when there is a significant change for the purposes of the review duties in due course.

Exemption from record keeping and review duties

10.26 Ofcom may (under sections 23(7) and (8) and sections 34(7) and (8)) exempt certain types of service providers from any or all of the record keeping and review duties which are the subject of this guidance. Ofcom can also revoke such an exemption. Details of any exemption (or revocation) must be published on Ofcom's website.

10.27 At this stage, we are not proposing to exempt any types of service from the record keeping and review duties, as we do not currently consider that there is a sufficiently strong evidence base to justify any such exemptions.[163] Additionally, our analysis of risk across the sector indicates a wide variation in levels of risk across services independent of type, which we believe would make an exemption by type of service difficult to implement in a reasonable and proportionate way.

10.28 We have considered whether there may be an undue burden arising from the record-keeping and review duties, in particular for smaller or lower risk service providers. However, we do not consider that the record keeping and review duties are onerous, in light of the proportionate approach we have set out in our draft guidance. We are also mindful of the importance of the risk assessment duties to the regulatory regime and hence the importance of having a record to demonstrate that a service provider's risk assessment is suitable and sufficient.

10.29 Finally, we note that the underlying duties to conduct risk assessments and take measures to comply with the relevant duties would not be removed by any exemption. Accordingly, we think it would be good practice for all service providers to keep written records and regularly review their compliance with their safety duties, particularly in the early days of the new regime when providers' understanding of their obligations is likely to be evolving.

10.30 Ofcom will keep its position on exemptions under review, and we welcome stakeholders' comments as part of this consultation.

Failure to keep records or review compliance

10.31 The requirement on service providers to comply with record keeping and review duties is an enforceable requirement under the Act. If a regulated provider fails to comply with its record keeping and review duties this may lead to enforcement action being taken by

[163] As part of this, we considered exempting small and/or micro businesses but did not have sufficiently strong evidence to understand the potential impacts if such businesses provide higher risk services.

Ofcom.[164] Any enforcement action would be taken in line with our Online Safety Enforcement Guidelines, which we are also consulting on as part of this Consultation.[165]

10.32 If Ofcom finds a service provider to be in breach of these duties, then we have the power to fine service providers up to £18 million or 10% of qualifying worldwide revenue, whichever is the greater.

Implementation

10.33 As part of our decision and statement on the matters which are the subject of this consultation, we expect to publish our guidance on the record-keeping and review duties, as well as other regulatory documents such as the Services Risk Assessment Guidance and our code of practice on the illegal harms safety duties. We must publish our first guidance on the record-keeping and review duties within 18 months of Royal Assent.[166]

10.34 The record keeping and review duties will take effect simultaneously with the service provider's obligations to conduct risk assessments and its obligations to comply with the relevant duties (for the purpose of the record-keeping duties) and the relevant review duties (for the purpose of the review duties).[167]

10.35 In regard to Category 1 user-to-user services and/or Category 2A search services, as soon as reasonably practicable after making a written record of their first risk assessment,[168] such regulated providers are required to provide this written record (in full), and in an electronic format, to Ofcom. The Ofcom email address for providing these written records will be determined by Ofcom at a later date.

Impact assessment

10.36 Ofcom is required by legislation to provide guidance on the record-keeping and review duties. As we have discretion over the nature of this guidance, we have carried out an impact assessment, as defined in Section 7 of the Communications Act (2003).

10.37 To the extent that the proposed guidance results in additional costs to those necessarily incurred by service providers in fulfilling their statutory duties and ensuring ongoing compliance, we consider that such costs are minimal and outweighed by the regulatory benefits of ensuring the availability of clear, well-maintained records and timely reviews of compliance. For the same reason, we consider that to the extent our decision not to grant any exemptions from the record-keeping and review duties creates burdens, these are

[164] See Part 7, Chapter 6 of the Act (Enforcement powers).

[165] See Annex 11 (Enforcement guidance).

[166] Unless we extend the period in accordance with section 194(3) – (6) of the Act.

[167] Note that these duties are subject to different contingencies and so are likely to take effect on different dates.

[168] This also includes any subsequent changes to the service provider's written records.

outweighed by the benefits of maintaining the application of the duties. In any event, as set out in paragraph 10.27, at this stage of the new regime, we do not have a sufficient evidential basis for exercising our power to grant an exemption.