A Module Playbook for Risk Assessment and Audit

Chris Riley and Susan Ness

Executive Summary

Last year, the authors introduced the co-regulatory approach of modularity, which offers a positive path forward for international internet governance through the use of multistakeholder, multinational bodies to design and undertake common operational functions across legal jurisdictions, despite different legal systems and regulatory frameworks.[1]

The Module Playbook for Risk Assessment and Audit provides an example of how modules would work in practice, serving as a roadmap for common modules to function under different laws. It describes the building blocks and bodies needed for a co-regulatory system to conduct risk assessments and independent audits of the assessments. It is a companion to the Module Playbook for Platform-to-Researcher Data Access, which describes the building blocks for a module and details the steps and recommended bodies to create and operate a system for vetting researchers and their projects seeking access to platform data.[2]

The introduction of modularity is both timely and time sensitive, as many democracies are adopting risk assessments and audits as part of their regulatory toolkit. The European Commission currently is drafting delegated acts to implement the Digital Services Act (DSA), while Ofcom is poised to implement the UK's Online Safety Bill (OSB) upon enactment. Wisely, these frameworks delegate to regulators the task of implementation, providing an opportunity for well-designed multistakeholder modular solutions to be recognized.

[1] https://thehill.com/opinion/technology/3479764-a-safe-open-internet-with-transatlantic-rules-is-easier-than-it-sounds/; https://www.lawfareblog.com/modularity-international-internet-governance

[2] https://techpolicy.press/a-module-playbook-for-platform-to-researcher-data-access/

Many international entities and multistakeholders are hard at work developing risk assessments and audit standards that could become building blocks for a modular regime for online digital services. Additionally, the International Accounting Standards Board, the independent non-profit body that oversees international financial accounting rules, could serve as a structural model for a risk assessment and audit module.[3]

Regulators, platforms, prospective auditors, and the public would benefit greatly from shared expertise and elimination of cross-border duplication of standards and protocols for risk assessment and audits. Governments also would save financial and staffing resources by recognizing multistakeholder modules.

A proposed modular governance system would include:

1. Creation of a Risk Assessment Standards Board (RASB): a multinational, multistakeholder independent body with members drawn from democratic jurisdictions that are likely to recognize the modular system, and collectively possessing computer science, online harms, human rights, and legal, auditing and internet governance expertise. A secretariat would assist the board.

2. Drafting and adopting Risk Assessment Standards (RAS): The RASB would create with public input a risk assessment guidebook suitable for a broad range of digital platforms, and based on the scoping considerations within the DSA and risk profiles and guidelines for self-assessment in the OSB.

3. RASB-developed criteria for auditor selection (in addition to independence), as well as protocols and review processes for conducting audits: We currently do not propose that the RASB certify auditors. Instead, it is likely that a separate module or industry/public board would be formed to oversee a certification process and to discourage gaming the audit system. The RASB (or another entity) could develop a process to review a subset of audits for consistency and quality. Privacy rules might limit access to non-public audits; and

4. Review and update of Risk Assessment Standards: The RASB should adopt a process to regularly review and update the RAS to ensure it remains fit-for-service.

This structure is similar in many ways to the operations of national accreditation bodies, such as the ANSI National Accreditation Board, which takes on many such processes including for international standards from ISO.[4] The structure also aligns closely to the proposed DSA implementation in the European Union, incorporating many tasks that the European Commission was required to undertake. Modularity as a framework can reduce implementation costs incurred by individual governments by shifting the costs onto other bodies and structures.

[3] https://www.ifrs.org/groups/international-accounting-standards-board/
[4] https://anab.ansi.org/about-anab

It also reduces the likelihood that assessments and audits become ineffective “check box” exercises.

There is at this moment an opportunity – or perhaps an urgent need – for the broader community working on these issues to form a multistakeholder module and to develop Risk Assessment Standards to guide assessments and audits. Official recognition by one or more governments of the module’s contributions to their regulatory processes would add crucial validity to a co-regulatory modular regime. It would also encourage other jurisdictions to join on, aligning around common standards and protocols, much like countries have aligned around the IASB Standards for financial accounting. Governments would continue to control enforcement under the law, maintaining their regulatory sovereignty.

1. Context

The introduction of modularity is both timely and time sensitive, as more and more countries consider adopting risk assessments and audits as part of their regulatory toolkit. Notably, platform assessments of “illegal harms” and “children’s risk” are featured prominently on Ofcom’s recent roadmap for implementing the UK’s Online Safety Bill.[5] And the European Commission has opened a consultation for stakeholder input on risk assessments and audits under the Digital Services Act.[6] Regulators, platforms, prospective auditors, and the public would benefit greatly from sharing expertise and avoiding cross-border duplication in developing these strategic and substantive requirements. Now is the time – before a multiplicity of rules are set in concrete – to move forward with a multistakeholder, multinational modular approach to the design and implementation of standards and protocols for assessments and audits.

Happily, multistakeholders already are developing risk assessment and audit systems and standards. For example, the Global Network Initiative, a multistakeholder organization, facilitates assessments of ICT companies in upholding human rights principles.[7] While the scope of digital platform governance is larger, GNI’s program offers valuable insights in both process and substance. Similarly, NIST’s 2023 AI risk management framework and the ISO 42001, a management systems standard for governance of AI systems, contribute to this work, since platforms increasingly use AI in their operations.[8] Also, audits exist in a broad range of industries, in particular financial and cybersecurity; their operating systems are instructive. We discuss these and other examples in greater detail in Section 4 on Building Blocks.

---

[5] https://www.ofcom.org.uk/__data/assets/pdf_file/0016/240442/online-safety-roadmap.pdf; https://bills.parliament.uk/bills/3137

[6] https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act-ensuring-safe-and-accountable-online-environment_en

[7] https://globalnetworkinitiative.org/

[8] https://www.nist.gov/itl/ai-risk-management-framework; https://aistandardshub.org/how-a-standard-in-development-iso-iec-42001-can-meet-collective-ai-governance-goals/

With 2024 elections scheduled in the world's largest democracies, platform efforts to mitigate the spread of misinformation and online harms will be under the microscope. A small window of opportunity remains to develop effective multilateral, multistakeholder-led risk assessment frameworks and processes to facilitate collaborative review and action for protecting election integrity. Aligning the core of such assessment systems across democracies through the framework of modularity would greatly facilitate the execution of this task without undermining sovereignty.

2. Problem statement

Effective risk assessment and mitigation are core to modern digital platform governance. While there are collections of best practices, there is no one right way to set a platform's content policy or to structure and resource its moderation systems that will always be effective, given the rapidly evolving digital landscape. Recommendation engines further complicate the challenge of regulating behavior to achieve desired outcomes in practice. So cogently assessing risk can be a spiraling scope exercise.

The governance paradigm – built on transparency, risk assessment, mitigation, and researcher access – is designed to be sustainable against that complex backdrop. However, that paradigm may prove ultimately ineffective if the assessment and audit processes devolve into "check box" exercises.

Ensuring the thoroughness of audits is difficult. Independence of the auditor is perhaps the most visible lever in promoting good outcomes. But unlike the case of traditional financial auditing systems, auditor competence cannot be readily assumed. There is no pre-existing playbook, formula, set of standards, protocols and certification, or accepted training guide on how to conduct a platform policy audit. The independence lever is necessary but not sufficient.

The first test of this risk management-driven paradigm will be conducted in Europe under the Digital Services Act, when the very large platforms and search engines so designated by the European Commission conduct initial risk assessments and mitigation, followed by the first independent audits of these efforts.[9] If the UK enacts the Online Safety Bill this fall, assessments under the OSB will occur shortly thereafter. Lessons from this first wave of test cases may be delayed if there are challenges to public transparency.[10] Meanwhile, as discussed in Section 4 on Building Blocks, scholars and NGOs are hard at work exploring the complexity of the underlying services and are crafting initial strategies and best practices for assessing risks.

---

9 https://digital-strategy.ec.europa.eu/en/news/digital-services-act-commission-designates-first-set-very-large-online-platforms-and-search-engines

10 https://dsa-observatory.eu/2023/01/30/counting-the-days-what-to-expect-from-risk-assessments-and-audits-under-the-dsa-and-when/

Luckily, signs of broad alignment may be emerging at the intersection of evolving law and understanding complex digital platforms. The DSA and OSB delegate most of the details of governance to subsequent rulemaking, including delegated acts and guidelines by the European Commission under the DSA, and further proceedings by Ofcom, the UK agency tasked with implementing the OSB. Such delegation is a feature of these laws, not a bug; it gives space to fold evolving research and public perspectives into governance, and – crucially – it creates the opportunity for modularity to be acknowledged as an effective co-regulatory system to address the complex tangle of risk assessment and mitigation.

3. Examples of implementing legislation

Both the DSA and the OSB have substantial guiding language related to platform assessment. The Digital Services Act includes a spectrum of directed obligations for platforms, based on the size of the online service provider. More open-ended assessment and mitigation responsibilities are required only for very large online platforms (VLOPs) and search engines (VLOSEs), which then must complete an independent audit of both the obligations and the self-assessment. See Exhibit I for details.

The current version of the Online Safety Bill sets out expectations for platform risk assessment with ample direction to Ofcom to develop guidance.[11] The OSB's assessment process follows the safety duties outlined in the law, as well as the codes of practice on safety that Ofcom will draft after public consultation. Risk assessment duties arise in the context of "risk profiles" under the law, with three named categories in the law: "user-to-user services", "search services", and those that could risk "harm to children"; see Exhibit I for details. The OSB does not require independent audits, but the agency will work with platforms to improve their assessments.

The DSA initially offers limited substantive or procedural guidance to assist the VLOPs and VLOSEs in fulfilling these additional requirements. It defers to the European Commission with support from the member state Digital Services Coordinators to offer guidance through a delegated act. As in the case of Ofcom providing guidance under the OSA, this arrangement creates an opportunity for a modular structure, enabling multistakeholders from multiple countries to help set best practices for risk assessments and for standards and protocols for required audits under the DSA.

4. Building blocks

Many existing collaborations can serve as building blocks for designing one or more modules to implement the required risk assessment and audit obligations. Most of the projects referenced in this section have some but not total overlap in scope with this playbook's focus on digital platform governance. But their existing processes and standards can serve as useful examples to draw upon. Some collaborations are sufficiently aligned (and sufficiently flexible) that their

[11] https://bills.parliament.uk/bills/3137

scope could be modified to enable them to function as the independent modular body that administers the risk assessment and audit functions outlined here.

Perhaps the most comprehensive research on the assessment and audit cycle is the recent Auditing Recommender Systems report from Stiftung Neue Verantwortung (SNV).[12] Recommender systems present complex risks and give rise to difficult questions to assess. The SNV report breaks down the assessment process, outlining the different elements of a digital platform that can contribute to risk, and the six distinct types of audits that could be used to assess risk: code audit, crowd-sourced audit, document audit, architecture audit, automated audit, and user survey.

At the organizational level, the Global Network Initiative (GNI), has undertaken similar work.[13] Founded in 2008, its mission is to help its member companies protect human rights and privacy. For many years, it has overseen self-regulatory assessments, independent audits, and mitigation efforts of its private sector members for their human rights impact. GNI organizes the periodic assessment and audit of its member companies in compliance with its Principles on Free Expression and Privacy. GNI's assessment toolkit describes this process in detail.[14]

While the scope of this playbook extends beyond the GNI focus on human rights and privacy, nevertheless, GNI's experience provides valuable insights on both the substance and the GNI operating structure in designing one or more modular mechanisms for assessment and audit of online services. With sufficient resources, GNI potentially could apply its same tools and methodology to oversee assessments and audits evaluating private sector compliance with other standards, not just GNI's own principles. However, this would be a significant expansion in scope if GNI were tasked with the scale of the DSA.

Brainbox has prepared an extensive analysis of DSA audit requirements for the Action Coalition on Meaningful Transparency (ACT).[15] While Europe is the focus of the analysis, the document references "related resources" including global perspectives on audits, citing the work of GNI, the United Nations, civil society organizations, and multistakeholder groups.

Looking across and building on many of these efforts, the World Economic Forum has assembled a Global Coalition for Digital Safety.[16] The WEF coalition recently released a report, Digital Safety Risk Assessment in Action: A Framework and Bank of Case Studies, which will provide substantial captured expertise useful for the development of assessments and audits.[17]

---

[12] https://www.stiftung-nv.de/sites/default/files/auditing.recommender.systems.pdf
[13] https://globalnetworkinitiative.org/
[14] https://globalnetworkinitiative.org/wp-content/uploads/2021/11/AT2021.pdf
[15] https://www.meaningfultransparency.tech/post/audit-frameworks-under-the-digital-services-act-an-act-briefing-note
[16] https://initiatives.weforum.org/global-coalition-for-digital-safety/home
[17] https://www.weforum.org/reports/digital-safety-risk-assessment-in-action-a-framework-and-bank-of-case-studies

The field of artificial intelligence (AI) provides many examples of relevant collaborations. As with platform governance, governance of AI systems is evolving towards a more process-based, flexible framework, recognizing that setting *ex ante* behavior guidance for such systems can be of limited utility. For example, the EU's forthcoming AI Act imposes on "high-risk" AI systems the obligation to undertake continuous risk identification, analysis, and risk management measures, rather than imposing specific behavioral obligations.[18] Risk management remains a central lever for intervention, as also evidenced by the U.S. National Institute of Standards and Technology (NIST) AI Risk Management Framework (AI RMF).[19] The work of ISO to develop a "management standard" for AI systems (see discussion regarding ISO/IEC 42001) similarly offers guidance to companies building and using AI.[20]

AI and platform governance share much in common – both in their underlying values and in practical implementation. For example, the AI RMF lists characteristics for trust that also apply to online content systems: "safe," "accountable and transparent," "explainable and interpretable," "privacy-enhanced," and "fair with harmful bias managed." All could as easily have come from a goal statement for platform services. Furthermore, modern digital platforms employ AI in their content moderation, both to reduce the cost of content moderation as well as to recommend content to users. Thus, AI risk management frameworks could directly contribute to risk assessment for these components.

In the field of financial accounting, the audit services industry has well-developed practices and procedures for evaluating company compliance with accounting rules. In most countries, financial accounting follows the IFRS Standards, set by the IFRS Foundation through a standards-setting board, the International Accounting Standards Board (IASB). In the United States, accounting practices instead follow the "GAAP" (Generally Accepted Accounting Principles), which are set by the Financial Accounting Standards Board (FASB), an independent non-profit organization.[21] These standards provide guidance for internal bookkeeping as well as the metrics by which auditors assess the company's books. These standards enable comparison of financial statements across businesses. In the US, FASB was granted legal recognition by the U.S. government through policy statements of the Securities and Exchange Commission, pursuant to language in the Sarbanes-Oxley Act of 2002, which explicitly gives the SEC authority to designate the standards of a non-governmental body as sufficient for securities law purposes.[22]

While financial accounting and content governance bear few similarities, financial accounting offers valuable insights in structuring a module for platform risk assessment and audits.

---

[18] https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0206 (Title III, Article 9)
[19] https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf
[20] https://aistandardshub.org/how-a-standard-in-development-iso-iec-42001-can-meet-collective-ai-governance-goals/
[21] https://www.fasb.org/facts
[22] https://www.sec.gov/rules/policy/33-8221.htm; https://www.govinfo.gov/content/pkg/PLAW-107publ204/pdf/PLAW-107publ204.pdf

Applicable concepts include the self-assessment and third-party audit systems, as well as the governance structure, which enables FASB to serve a critical governmental function, with oversight by the Securities and Exchange Commission.

In the field of digital privacy, an entire profession has been built to provide in-house product counsel assessment of privacy risks and internal work to mitigate harm. Privacy compliance audits have emerged as a tool to assist a company in gauging its compliance with relevant laws; however, such laws are generally more specific in their behavioral expectations than laws governing online speech and digital platform services. The burgeoning trust and safety professional field resembles the privacy industry in that it develops in-house and external expertise to evaluate private sector practices against a shared standard. Therefore, much can be learned about risk and mitigation from the work of organizations including the Trust & Safety Professional Association (which focuses on building a shared community of practice for trust and safety practitioners) and the Digital Trust & Safety Partnership (which focuses on company-level assessments of trust and safety practices).[23]

5. Objectives and operation of a modular governance system

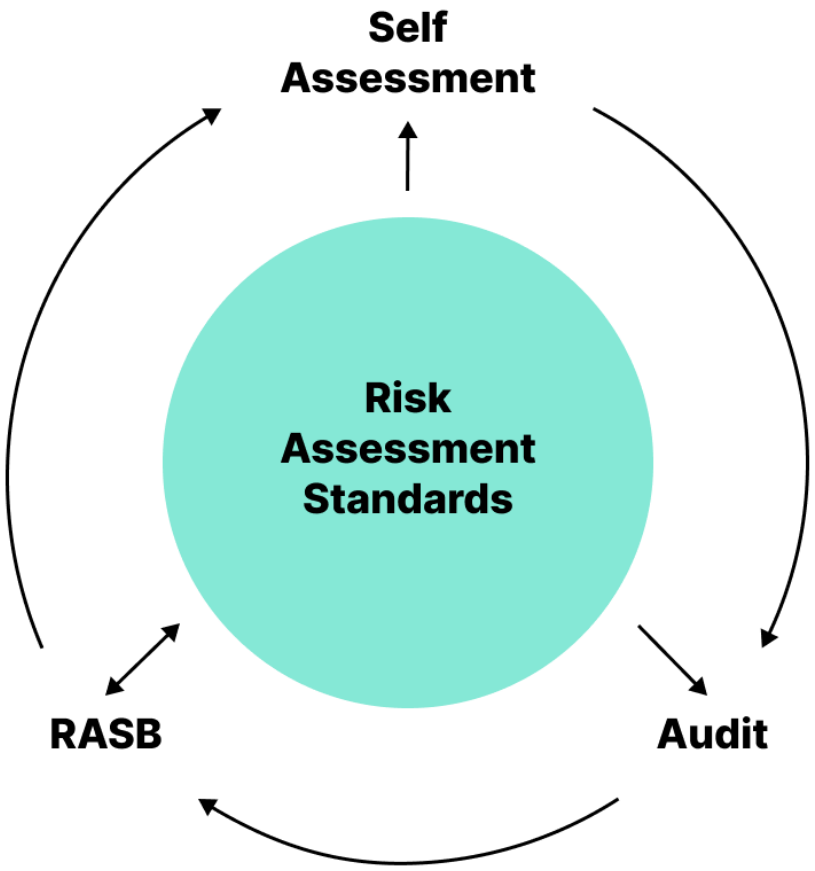
Building on these many starting points, a modular governance framework connects existing efforts to specific legislative and regulatory obligations across multiple jurisdictions. Core government enforcement functions operate alongside and in coordination with multistakeholder processes that implement the law. The result is a practical and multinational co-regulatory governance regime, with benefits for all stakeholders, as noted in prior modularity work.[24]

For risk assessment and independent audit review, a module could be established in the form of a multinational, multistakeholder independent board to develop: (1) effective, substantive guides for companies undertaking self-assessment; (2) standards and protocols for independent auditors reviewing risk assessments, built on the self-assessment guides; (3) capabilities to oversee the audit process to ensure minimum quality standards across different auditors and service categories; and (4) a process to update the standards and protocols based on experience.

This playbook proposes modeling a cross-border module for risk assessment and audit after the international accounting standards bodies that develop standards for internal corporate accounting and independent financial audit. While this is not the only possible approach, the duality of standards and a standards board offers great potential for clarity and sustainability. As detailed below, central to such a regime is the creation of a multinational, multistakeholder Risk Assessment Standards Board (RASB), to develop and maintain Risk Assessment Standards (RAS).

[23] https://www.tspa.org/; https://dtspartnership.org/
[24] https://www.lawfareblog.com/modularity-international-internet-governance

The RASB must design qualification standards for auditors and standards and protocols for conducting the audits. Both platform assessment and the independent audit would be performed consistent with the RAS. Finally, the RASB could be tasked with setting a process for review and update of the RAS, potentially by reviewing a sample of the audits and regularly updating the RAS based on that experience.



The Risk Assessment Standards are first created by the RASB; then they are used by platforms in self-assessment, and then again in the process of audits of those assessments. To restart the cycle, the RASB works to ensure audit quality and periodically updates the RAS as needed.

We now describe each function of the module in greater detail.

a. Create a Risk Assessment Standards Board

The first step in creating one or more risk assessment and audit modules is to assemble a Risk Assessment Standards Board. The Board should have a Chair and Vice-Chair responsible for convening the Board and ensuring compliance with its own rules and practices. Board members should reflect the multiple democratic jurisdictions that are likely to recognize the modular system. Board members should serve staggered terms, and should include individuals with the following subject matter expertise: computer science (including both fundamental knowledge and specific understanding of modern social media and digital platform and search operations); online harms (including the spread of mis- and disinformation, human rights, and threats to

democracy and democratic processes); and legal, auditing, and internet governance (including experience with inclusive stakeholder processes that bring together government actors with civil society and private sector stakeholders). Deep technical expertise is essential, given the wide range of complex, rapidly evolving platforms and risks. A secretariat or other permanent staff of the Board should include additional individuals with these skill sets.

Additionally, the Board should be structured with mechanisms and processes to incorporate input from a broad range of stakeholders. Explicit comment solicitation should be conducted at least annually for course correction and input for revisions (if any) to the adopted standards, best practices, and protocols. Additionally, the Board should endeavor to hold workshops and other in-person gatherings, open to a broader public, especially in jurisdictions whose laws are a focus for the module.

Note, this playbook does not propose a specific process for selecting the individuals in such a board. Rather, as suggested in the modularity framework overview, a group of experts working on implementation issues, which might include representatives of governments, would be assembled as a Risk Assessment Standards Board.[25] Several such groups might initially "compete" for recognition by civil society, platforms, and government actors in an informal process. At the end of the day, to achieve the goal of a common cross-border system, there should be only one recognized modular system approving the standards and protocols.

b. Establish initial Risk Assessment Standards

The RASB, after stakeholder consultation, should compile and adopt a risk assessment guidebook suitable for a broad range of digital platforms. It is a daunting task. Yet, the Digital Safety Risk Assessment in Action Report by the World Economic Forum's Global Coalition for Digital Safety, and the NIST AI Risk Management Framework and the ongoing ISO standards process, have undertaken similar ambitious efforts.[26] In addition, the SNV report offers substantial advice on how to begin drafting Risk Assessment Standards.[27]

While this playbook does not suggest what the standards should include, we offer a starting point: The questions to ask and methods to investigate are outlined in different and complementary ways in both the Digital Services Act and the Online Safety Bill.

The DSA offers functional considerations for scoping risk assessment and thereby developing suitable Risk Assessment Standards. For example, the DSA distinguishes the design and functioning of a service, as well as use case for the service. (Article 34(1)). Similarly, the DSA calls out the policies and procedures of recommendation, moderation, terms of service, advertising, and data practices, among others, as worthy of attention. (Article 35(1)). The DSA

[25] https://www.lawfareblog.com/modularity-international-internet-governance

[26] https://www.weforum.org/reports/digital-safety-risk-assessment-in-action-a-framework-and-bank-of-case-studies; https://www.nist.gov/itl/ai-risk-management-framework; https://www.iso.org/standard/81230.html

[27] https://www.stiftung-nv.de/sites/default/files/auditing.recommender.systems.pdf

frames the scope of assessment in terms of “systemic risk,” and lists specific types of harm (i.e., to “civic discourse”) as a starting point for evaluating systemic risk.

Under the OSB, Ofcom will develop “risk profiles” reflecting a range of services that give rise to similar risks to be assessed. Ofcom has published a detailed four-step process to guide self-assessment, which could be integrated into the Guidelines that a Risk Assessment Board could develop:[28]

> ***Step one: Establish the context***
> *Establish the **risks of harm** that need to be assessed. Consult the **risk profiles** produced by Ofcom, which set out our assessment of key risk factors, and identify any gaps in your understanding and evidence.*
>
> ***Step two: Assess the risks***
> *Review **evidence** about your platform and your risks. Assess the **likelihood** of harmful content appearing and the **severity/impact** of harm. In addition, evaluate **existing mitigating measures**.*
>
> ***Step three: Decide measures and implement***
> *Decide how you will comply with the safety duties, including through Ofcom's **Codes of Practice**. Identify the measures you need to implement. Record the outcomes of the risk assessment. **Implement any new measures**.*
>
> ***Step four: Report, review and update***
> ***Report** via relevant governance structures. **Monitor** the effectiveness of your mitigation measures. Put in place **regular review periods** for your assessments, recognising any triggers to revisit assessments between these periods.*

There are no perfect solutions for assessing and mitigating the risks of digital platforms. But there are questions that can be asked to probe and gather the best understandings of risk and mitigation available at the time of assessment.

And, as previously noted, the work of compiling those questions and suggestions into a set of Risk Assessment Standards that are common across borders is best done by the RASB, an independent multistakeholder board supported by a secretariat function including computer scientists and social scientists who are up to date on relevant research, and can collaborate with the professional trust and safety industry.

Working through an established set of Risk Assessment Standards provides advantages above and beyond facilitating compliance with regulations. It allows digital platforms to conduct effective *ex ante* self-assessments as part of internal reviews in developing new features and functions. If the platform endeavors to identify risks before launch, it can mitigate harm before any audit or enforcement is even contemplated.

[28] https://www.ofcom.org.uk/news-centre/2023/how-we-are-approaching-online-safety-risk-assessments

c. Provide standards for auditors and review processes for audits for consistent quality

One of the most difficult questions facing the implementation of assessment frameworks is how to avoid audits becoming “check box” exercises. Current legislation does not discuss how to ensure audits and auditors avoid this outcome; at most, some guidance is offered for calibrating independence of the auditors, but not competence. As discussed below, the quality of the assessment cycle overall is well served if a module also specifies the qualifications of auditors and the standards and protocols to be applied. To the extent feasible (perhaps with spot checks, and/or where contextual elements indicate high need), the module could review a selection of audit outputs for quality.

In addition to ensuring that both assessments and audits are guided by the Risk Assessment Standards, the RASB should develop standards for auditor competence, focusing on the auditor’s personnel and processes. The RASB should also develop a process to evaluate an auditor against these qualifications. That process should include a review of the auditor’s training and internal systems as well as the qualifications of the audit team, looking for competency in technology design, online harms, and other subject matter as applicable. More than one specialized auditing team may be needed, given the wide range of skillsets necessary in complex risk assessment audits.

We do not currently propose that the RASB’s mandate include actual vetting and certification of auditors. However, a separate modular board, composed of industry and public members, could be established to operate a certification process. Having public members on the board will help to discourage anticompetitive audit industry practices. Prospective auditors might seek certification to enhance their candidacy or to dissuade platforms from selecting under-qualified, least-cost firms. Governments also might require auditor certification in specific circumstances.

The RASB should develop a process to review audit outputs for consistency and quality. At scale, this might exceed the capacity of any reasonably scaled institution. The result would be a slip in quality, undermining the entire purpose of the exercise. Rather than review every output, the RASB (or a separate modular body created for this purpose) could adopt a mechanism that tax agencies use for auditing returns: select a small portion at random to review, and factor into that calculation variables that may indicate higher need for review, such as a new auditor or new platform, or a particularly high-risk environment. Privacy rules may limit the board to reviewing only publicly-released audits.

An effective auditor-and-audit review process run by the RASB (or other modular body) should help to reduce government enforcement expense by encouraging easier-to-process audit outputs while improving the quality of the audits themselves.

d. Review and update of Risk Assessment Standards

Digital platforms change frequently, including their core technologies (such as recommendation systems), the external cultures and contexts in which they are used, and their content policies – largely in response to evolving culture and law. Also, best practices for substantive and procedural guidance in how to mitigate harm may evolve over time. As a result, a risk assessment framework must be regularly reviewed to see if it, too, should change.

We propose that the RASB, which develops and applies the RAS in review of audits, would be best positioned to regularly evaluate whether the RAS should be updated. The RASB should establish a process for regular review of the RAS as well as a process for ad-hoc evaluation where circumstances necessitate an off-cycle update.

6. Getting to yes

Official recognition by one or more governments of the contributions of a module to their regulatory processes will add crucial legitimacy to a co-regulatory modular regime. Unfortunately, platform guidelines for conducting self-assessments have not been finalized, nor have rules been adopted governing how independent auditors should evaluate these assessments. Yet there are small signs of progress in many jurisdictions beyond the EU and UK, such as in Australia, New Zealand, and the United States. This creates incentive and room for governments in multiple jurisdictions to align around common international standards and protocols, in the same way accounting practices in 167 jurisdictions align around the IASB Standards.[29]

In the European Union, current disparities in governance frameworks most likely will surface. Under the Digital Services Act, initial digital platform self-assessments will happen before platforms receive guidance from a not-yet-existing multistakeholder collaborative process. The self-assessment processes will vary widely, as will the outcomes. So too will the processes vary of the independent auditors trying to evaluate these assessments, alongside the other obligations under the Digital Services Act. Some form of mechanism, framework, or guidance would be helpful, as the European Commission works to identify the best examples of this experiment and helps ensure that the next iteration is more aligned.

This represents an opportunity for the broader community working on these issues to come together via a multistakeholder, multinational module process. Working with the first round of platform assessments, a module will have practical, real-world examples to learn from in developing a shared best practices framework – or an initial set of Risk Assessment Standards that can guide assessments and audits going forward.

But we need not wait for those initial self-assessments. Risk Assessment Standards can, and should, be drafted as soon as possible through the collective action of an experienced

[29] https://www.ifrs.org/use-around-the-world/use-of-ifrs-standards-by-jurisdiction/

multistakeholder community. A workshop of stakeholders could begin to assemble existing resources into a framework that provides the questions and metrics to enable companies to conduct their assessments and for auditors to evaluate those assessments. In fact, the World Economic Forum's Global Coalition for Digital Safety has begun to do this already.[30] Working with interested governments, multistakeholders should develop a plan to create a sustainable and well-resourced Risk Assessment Standards Board that could finalize such a framework and maintain it over time. Any institutional architecture that emerges from such a process must include the input of stakeholders from multiple sectors and countries, including government experts.

Initial and ongoing funding for such a body remains an open question. Its operations could be funded by many sources, including by a portion of the fees collected by governments to support enforcement of their law; by international organizations (reflecting the potential for facilitating international alignment); by voluntary contributions from platforms (justified through the potential for reducing their compliance costs), by philanthropy (given the potential for improved quality of accountability and trust globally); or, most likely, by a combination of the above.

Governmental action that recognizes and encourages the use of a module to improve assessment and audit is the most important factor in establishing legitimacy. For the European Union, recognition in a delegated act adopted by the European Commission would be most direct, but may not be necessary under the DSA. Similarly, in the United Kingdom, action by Ofcom to recognize the legitimacy of an established RASB – in the same way the Sarbanes-Oxley Act in the United States gives explicit recognition to the U.S. FASB – would go far to establish a rebuttable presumption of sufficiency of assessments, to the extent they align with the RAS.

International collaboration through modularity offers tremendous advantages to jurisdictions in strengthening the accountability goals of digital platform laws. The structures created can help prevent assessments and audits from turning into mindless checklists and compliance exercises. While the prospects of this worst-case scenario may seem less risky in larger jurisdictions with more resources to spend on oversight and enforcement, they remain a concern everywhere. And incurring large government enforcement costs is not efficient solution. Modularity proposes to mitigate this risk through the development and adoption of common cross-border assessment frameworks.

The nature of risk assessment inherently requires visibility into platform operations and extensive sharing of sensitive information with auditors; it is never an easy or inexpensive task for government enforcement. While the frameworks guiding the process and substance of these interactions can in theory be imposed from the top-down, a co-regulatory collaboration will provide the best solution for realizing regulator goals and values in practice through real-world implementations. The enforcement of laws, such as through administrative agencies in the United States or digital services coordinators in the European Union, allows ultimate authority to reside with governments, creating more room for collaboration within modules.

[30] https://initiatives.weforum.org/global-coalition-for-digital-safety/home

Susan Ness *is a distinguished fellow at the Annenberg Public Policy Center of the University of Pennsylvania (APPC) where she previously convened and co-chaired the Transatlantic Working Group on Content Moderation and Freedom of Expression.*[31] *She also is a nonresident senior fellow at the Atlantic Council. A former media and telecom regulator, she served on the Federal Communications Commission. She earned an M.B.A. from The Wharton School of the University of Pennsylvania and a J.D. from Boston College Law School*

Chris Riley *is a distinguished research fellow at the Annenberg Public Policy Center, and the Executive Director of the non-profit Data Transfer Initiative. Previously, he ran the global public policy team at Mozilla, working with his team in Brussels on the development of the DSA and the Code of Practice on Disinformation. He earned a Ph.D in computer science from Johns Hopkins University and a J.D. from Yale Law School.*

[31] http://www.annenbergpublicpolicycenter.org/twg

EXHIBIT I

EXAMPLES OF IMPLEMENTING LEGISLATION

European Union's Digital Services Act (citations from published final text of October 2022):[32]

The DSA sets its assessment mechanisms into motion with little initial substantive or procedural guidance. The DSA establishes the implementing regulator (the European Commission, in the case of the DSA) as a partner in providing guidance to platforms in their assessment, and reserves the right for future action to establish rules of the road for assessments and audits. This structure provides an opportunity for non-governmental stakeholders to contribute to modules and other processes that can develop from the outset standards and practices for assessment and audit, and to continue engaging on an ongoing basis. The DSA's relevant provisions are:

1. Very large online platforms (VLOPs) are obligated to conduct self assessments for a scope of "systemic risks" associated with their services; guidance on specific elements to examine are offered, including recommender systems, content moderation systems, and advertising. (Article 34) The platforms are directed to incorporate stakeholders, including civil society organizations, into these processes. (Recital 90)
2. VLOPs must undertake mitigation measures to address the risks identified in their assessment; the law offers a catalog of such potential changes touching on platform policies, processes, user interfaces, and technical systems. (Article 35(1))
3. The Commission reserves the right to provide guidelines related to mitigation practices, incorporating public consultations in its development. (Article 35(3))
4. VLOPs are required to conduct annual independent audits covering broad scope: all "obligations under Chapter III" (which encompasses articles 11-48). (Article 37) Notably, this includes auditing the self-assessments and the mitigation measures put in place to address them.
5. The Commission is directed to support voluntary standards developed to further compliance with the DSA, including auditing standards. (Article 44(1)(e))
6. The Commission has the authority to adopt delegated acts to establish specific rules for conducting audits, including methodologies and reporting formats. (Article 37(7)) These rules "shall take into account any voluntary auditing standards" of Article 44.

United Kingdom's Online Safety Bill, in progress (citations derived from January, 2023 text):[33]

Like the DSA, the UK's OSB directs the implementing regulator, Ofcom, to provide guidance for platform assessment. Relevant provisions in the OSB are:

1. Ofcom prepares "risk profiles" for applicable digital platform services - in this context, "user-to-user services" and "search services", and a third category for "harm to children"

[32] https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L:2022:277:FULL&from=EN
[33] https://bills.parliament.uk/publications/49376/documents/2822. If newer versions become available, they will be linked from the primary UK Parliament page on the bill: https://bills.parliament.uk/bills/3137

broadly construed. (Section 89(5) and 89(1)) These risk profiles are to be based on Ofcom's own assessments of potential harm deriving from illegal content. Ofcom's risk assessments must "identify characteristics … that are relevant to such risks of harm, and assess the impact of those kinds of characteristics on such risks." (Section 89(2))

2. Ofcom must then provide guidance to platforms on how to conduct risk assessments according to its risk profiles. (Section 90(1)-(3)) This guidance is expected to include both substantive elements (what risks to examine, based on Ofcom's assessments; and where to look for risk of harm) as well as procedural (how to conduct the assessment).
3. Ofcom is given the authority to revise the guidance over time. (Section 90(5))
4. The OSB establishes a core duty of platforms to conduct risk assessments following Ofcom's guidance. (Section 22)
5. A separate core duty of care under the law requires platforms to undertake proportionate measures to mitigate identified risk. (Section 23(2)-(4))
6. After a platform's self-assessment is completed,although Ofcom may conduct audits as one of its tools to enforce the law (Schedule 12), the bill doesn't specify a specific process for review, nor require independent audits of the assessments.