

Access to data for researchers: A state of play 2025

Centre for Democracy & Resilience

Introduction

The ability of researchers to access platform data is foundational to democratic accountability in the modern age. Access to data is indispensable for safeguarding democratic processes and maintaining situational awareness. While the work of independent researchers is necessary for detecting, identifying, and understanding the “ABCDEs” - actors, behaviours, content, degree, and effect - associated with malign content that threatens our social fabric, such research remains critically dependent on data availability. In 2024, the decommissioning of established monitoring tools like CrowdTangle created a significant data gap that many organisations still struggle to fill.

This paper is the second mapping of data accessibility for researchers; the first was published in December 2024.¹ The paper draws on the findings of a survey conducted by GLOBSEC in fall 2025 among 34 experts from academia, think tanks, NGOs, research centres and organisations across the EU and neighbouring countries. The survey mapped out what data access researchers actually have, understand the methods and tools they have and rely on to obtain data relevant for their work, identify the obstacles encountered when requesting data from social media platforms,

and assess what conditions and support are necessary for Article 40(4) of the Digital Services Act (DSA) to become an effective, equitable, and timely mechanism. Taken together, the responses reveal a growing gap between the DSA's stated objectives and researchers' experience, highlighting persistent barriers such as platform obstruction, procedural delays, and limitations in available tools that continue to undermine the regulation's ability to deliver meaningful access to data for public-interest research.

The primary purpose of this report is to provide EU and national policymakers, Digital Services Coordinators (DSCs) and other stakeholders with an updated picture of researchers' data-access realities ahead of the full implementation of Article 40(4). It offers a forward-looking assessment of researchers' expectations and the practical conditions required for the vetted-researcher mechanism to become an effective and inclusive tool for studying systemic risks.

¹ <https://www.globsec.org/what-we-do/publications/access-data-researchers-state-play>

Methodology

This report presents the results of an online survey conducted between September and October 2025 as part of GLOBSEC's ongoing work under the Central European Digital Media Observatory (CEDMO) and EDMO hubs. The survey was designed to capture researchers' experiences with accessing publicly available data under Article 40(12) of the DSA, as well as their expectations and preparedness for the forthcoming vetted researcher data-access mechanism established by the Delegated Act implementing Article 40(4).

The questionnaire focused on several core themes:

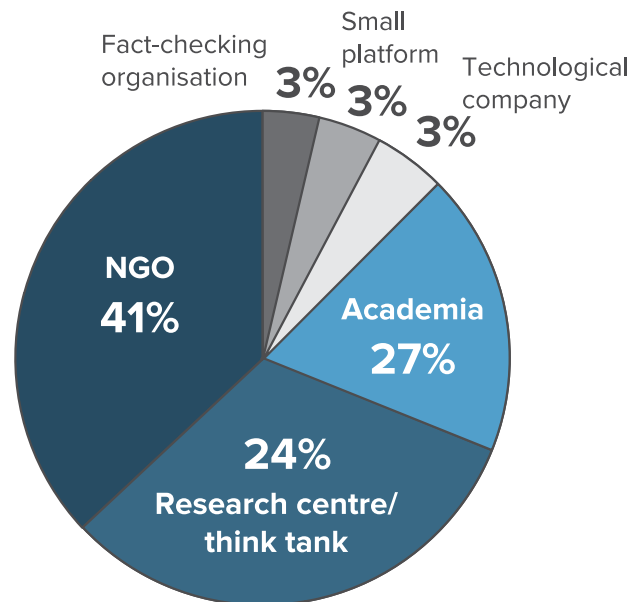
- the use of social media monitoring tools and scraping practices;
- experiences with applying for data from Very Large Online Platforms and Search Engines (VLOPs and VLOSEs);
- awareness and understanding of the vetted researcher regime;
- organisational readiness to apply for vetted status once the mechanism becomes operational;
- perceived feasibility of the Delegated Act's procedural timelines; and
- anticipated legal, technical and administrative barriers.

Profile of respondents and their research context

The expert survey gathered responses from 34 experts working across a range of institutions engaged in research on information operations, foreign information manipulation and interference (FIMI), platform governance, online harms and related digital-policy fields. Respondents represented a diverse cross-section of the research ecosystem, reflecting the multidisciplinary nature of work on systemic risks arising from online platforms.

Type of organisations

Respondents of the survey were researchers predominantly from the non-governmental and civil society sector, followed by academia and think-tank or research institutes. The distribution was as follows:



As in previous years, many organisations working at the frontline of monitoring malign content and information operations that filled in the survey are (smaller) civil society organisations whose access to platform data remains heavily constrained by limited financial, legal and technical capacities.

Geographic distribution

Respondents represented institutions based across the EU and beyond, with respondents coming from 21 countries overall. Central and Eastern Europe (CEE) was more represented, mirroring the geographic reach of the CEDMO and EDMO hubs and highlighting the significant concentration of FIMI-related research expertise in the region.

Key findings

- **The current data access landscape remains highly fragmented and unreliable.** Researchers still depend on scraping, manual monitoring and commercial tools, owing to inconsistent API access, restrictive clean room environments and prolonged waiting times. While some platforms, such as YouTube and LinkedIn, offer comparatively straightforward public data access routes, others provide limited, slow or non-functional tools that impede meaningful research.
- **Article 40(12) has not yet delivered predictable or efficient access to public data.** Researchers face prolonged delays, lack of responses from platforms, and opaque eligibility criteria. Applications are frequently rejected without substantiated reasoning.
- **The design of the vetted researcher status risks limiting its usefulness.** Overly long maximum timelines, administrative burdens, uncertainty about what constitutes a compliant secure environment, and vague definitions of eligible research organisations emerged as the principal reasons why researchers are reluctant to pursue vetted researcher status.
- **Researchers monitoring information space before elections, contributing to the European Commission's Rapid Response System and analysing FIMI operations consider the current timelines envisaged by the Delegated Act too slow to support public interest work.** Many emphasised that up to a nine-month decision period would render the data obsolete for monitoring FIMI, crises or ad hoc events that evolve rapidly.
- **There is strong demand for practical guidance, standardised templates and institutional support.** Researchers highlighted the need for legal assistance, GDPR guidance, secure storage models, training for research administrators, and clearer instructions from national DSCs and the European Commission.

Taken together, the findings paint a sobering picture: while the DSA holds genuine promise for transforming access to platform data, neither Article 40(12) nor the newly operative vetted researcher framework currently delivers reliable, timely, or fully inclusive access for public interest research. Determined enforcement by the European Commission and national DSCs, combined with clear guidance and meaningful platform compliance, will be essential to make Article 40(4) a genuinely workable instrument for independent scrutiny of systemic risks.

Legal framework: The DSA and its delegated act

Under the European Union's Digital Services Act (DSA), independent researchers are legally granted a right to access platform data to study systemic risks, with specific articles defining the distinct procedural conditions and eligibility requirements for obtaining either publicly available or sensitive non-public information.²

Under the Article 40(12), researchers, including those from civil society organisations, can gain access to data that is already publicly visible on a platform's interface. To facilitate this, Very Large Online Platforms (VLOPs) are legally required to provide technical tools, such as Application Programming Interfaces (APIs), that allow for the real-time analysis of this information. Furthermore, this article protects independent collection methods like scraping, ensuring that platforms cannot prohibit these techniques as long as the researcher is independent of commercial interests, discloses their funding, and demonstrates that the data is necessary and proportionate for studying systemic risks. Article 40(12) focuses on the immediate, tool-based analysis of public data.

On the other hand, the access to data under Article 40(4) is designed specifically for obtaining non-public data. To use this route, researchers must first be "vetted" by the Digital Services Coordinator (DSC). Once granted this status, the DSC, rather than the researcher directly submits, a "reasoned request" to the platform, compelling them to provide the specific data needed to identify and understand systemic risks within the EU. The Article 40(4) is thus a mediated process involving significant state vetting, intended to grant access to sensitive, non-public information specifically for deep-dive research into the internal mechanics of how platforms influence society.

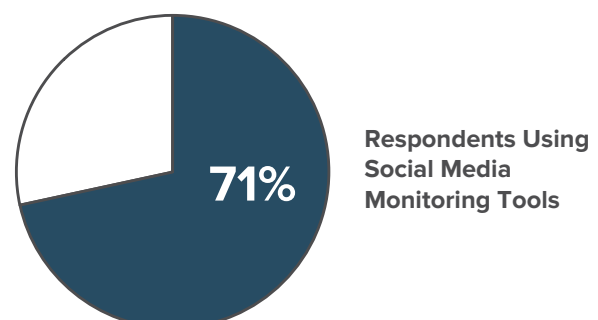
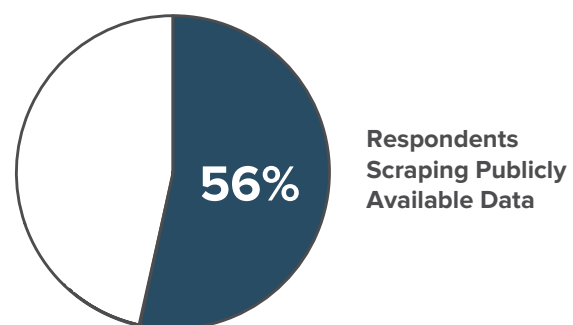
Furthermore, the delegated act on the access to data operationalising the Article 40(4) of the DSA entered into force on October 29, 2025, establishing

a structured EU mechanism for granting vetted researchers access to non-public platform data to study systemic risks. As the first applications under the vetted researcher framework were expected to be approved as of February 2026, this paper offers a state of play on how researchers currently obtain data from very large platforms, how Article 40(12) functions in practice, and how the research community perceives the workability of the emerging vetted researcher regime under 40(4).

How researchers get data

A majority of respondents (71%) reported using social media monitoring tools to support their research, reflecting a pattern consistent with the 2024 findings. These tools range from commercial social-listening platforms to custom-built scrapers and platform-provided interfaces and are employed for purposes such as tracking the spread of misleading content, monitoring harmful narratives during elections and crises, analysing audience engagement and virality, and conducting long-term or cross-platform trend analysis. 29% of respondents indicated that they do not use specialised tools, due to budget constraints or reliance on manual search and analysis.

Data access practices 2025



² <https://eur-lex.europa.eu/eli/reg/2022/2065/oj/eng>

Furthermore, scraping continues to be an important access channel for many organisations – 56% of respondents reported scraping publicly available data, motivated by the need for consistent, time-sensitive data flows, the high cost of commercial tools, the lack of accessible API-based alternatives, and the importance of cross-platform analysis during politically salient periods. Those who do not scrape cited legal uncertainty, insufficient technical capacity, and restrictive institutional policies as key barriers.

Deficit in data sufficiency

When asked about the sufficiency of available data, respondents rated their access at an average of 6 on a 10-point scale, where 1 indicated very poor data access and 10 indicated very good and fully sufficient access for effective research. While this marks a modest improvement compared with 2024, it continues to point to substantial and persistent gaps. Qualitative feedback shows that while researchers can obtain some relevant data, they **often lack the volume, granularity or timeliness required for effective analysis, particularly around elections, and that platforms' interfaces do not provide important metadata and engagement metrics. Organisations with fewer resources remain disproportionately affected. Overall, the 2025 results underline that Article 40(12) has so far delivered only limited improvements in real-time data availability, and that both scraping and commercial tools continue to serve as essential solutions.**

6/10 experience with data access in 2025

Rated by researchers on a scale of 1 (very poor) to 10 (fully sufficient) — up from 4.6/10 in 2024.

Access to VLOPs/VLOSEs under Article 40(12): existing experiences

Access to data using the Article 40(12) of the DSA **remains fragmented, administratively burdensome, and insufficient for real-time or cross-platform research needs.** 62% of respondents stated that they had not applied for direct access to social media researchers' APIs or content libraries, while 32% declared that they applied for APIs access and received the access.

The most common reasons for why people did not apply for access to APIs or other content libraries were that such access was not currently needed for their research, that they already relied on alternative tools or internal datasets, or that the data available through platforms would not be relevant to their ongoing work. In addition, **administrative complexity, unclear procedures, limited organisational capacity, along with low awareness of the possibility to apply or concerns about the expected quality and usefulness of platform-provided data** were reasons why researchers did not ask for access to APIs. A smaller number mentioned ethical or structural constraints that prevented the use of real platform data within their research projects.

Of the 32% of respondents who applied for API access and received it, 29% asked access to data from Facebook, 22% from Instagram, 14% from TikTok and YouTube, 3% from X and LinkedIn and 7% to other social media platforms.

Those who had applied emphasised that **each project requires a separate application, creating a cycle of repeated paperwork, uncertainty and delays.** This mirrors challenges reported in 2024 and highlights why many researchers are closely watching the development of the vetted-researcher data access processes.

Examples of researchers' experience with obtaining access to APIs and content libraries:

"MCL and virtual enclave experience is terrible is extremely difficult to work with them. In a year we were not able to a single investigation with MCL and virtual enclave"

"Meta Content Library process was fast. We applied and were given access quite rapidly."

"With the platforms I have selected, the experience of applying for and getting to data was positive. YouTube is very straightforward, no issues. MCL takes about a month and a half but then works fine. TikTok took about two months and I've written an op-ed about the many issues with the data. Moreover, since this op-ed came out, our access has been revoked without any warning. When asking for clarification as to what happened, TikTok failed to clarify."

"We submitted custom data access requests for a random sample of content across 6 VLOPs. Because this was publicly-accessible data, we deemed that 40.12 was applicable (even if API/data access system currently do not have 'random' endpoints). LinkedIn: cooperated, sent us the data. Twitter: refused. FB & IG: never answered beyond 'we acknowledge receipt'. YT: never answered beyond 'we acknowledge receipt'. TT: after months, gave us access to an API that doesn't seem to offer what we need."

"My access is through the organization I belong to, and it is more or less unsatisfactory due to the content library's limitations."

Quality of data obtained

Among those organisations who had successfully obtained data access to platforms APIs or content libraries, experiences differed significantly across platforms. While some VLOPs provide relatively predictable public-data access routes, others rely on **restrictive environments, unclear eligibility rules or lengthy periods of non-communication.**

Meta

Access to Meta's ecosystem [Facebook, Instagram and the Meta Content Library (MCL)] was described as inconsistent and generally unsatisfactory. While a small number of respondents received access within days through existing cooperation frameworks, most emphasised the limitations of the MCL, including slow performance, very long query times, and the inability to download data outside Meta's virtual environment. Several remarked that MCL represents a substantial downgrade compared to the discontinued CrowdTangle, offering reduced granularity, poorer usability, and an interface designed primarily for commercial users rather than research. Some researchers were unable to conduct meaningful investigations even after months of access, and strict non-disclosure agreements were seen as additional barriers that complicated their work. The Meta Ads Library, also under Meta but administered separately, required additional verification steps but was still considered manageable and comparatively uncomplicated.

TikTok

Experiences with TikTok were the most complex and problematic. Respondents reported prolonged waiting times, stretching from several months to well over a year, multiple rejections based on eligibility criteria that did not match the platform's published rules, and contradictory communication from TikTok staff. Even after eventual approval, several were granted access only to the limited Virtual Compute Environment (VCE) rather than the Research API they had applied for, and provided credentials often did not work. **Data within the VCE was described as aggregated, narrow in scope and manually**

reviewed, making it unsuitable for many research purposes. In one case, TikTok revoked data access entirely without warning after the publication of critical analysis and subsequently failed to clarify the reason.

YouTube

By contrast, YouTube (Google) was consistently described as straightforward and user-friendly. Respondents who applied for YouTube access generally received it quickly and reported no major obstacles.

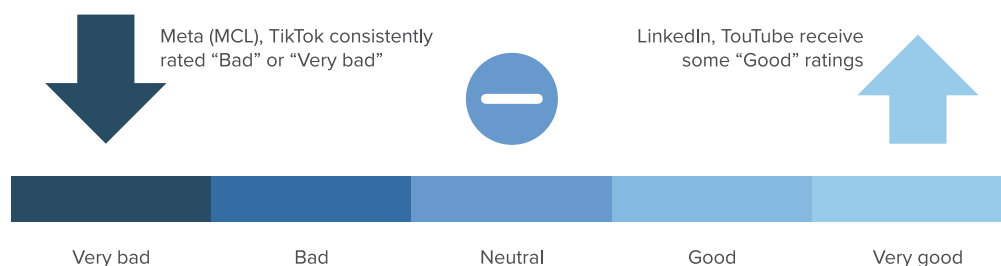
Other platforms

One respondent reported a positive experience with LinkedIn, noting timely cooperation and useful provision of requested data. Conversely, X (Twitter) was among the platforms that denied access, sometimes providing reasoning that respondents considered incorrect or not aligned with DSA criteria for systemic risk research. In several other cases, X simply acknowledged receipt but did not follow up.

A small number of respondents accessed other platforms such as Telegram and Mastodon, where no major issues were reported, although these platforms do not fall under the same structured access frameworks.

Overall, the responses illustrate a fragmented landscape in which Meta and TikTok present the most significant challenges for researchers, both in terms of process and data quality, while YouTube and LinkedIn offer comparatively smoother and more predictable access. A visual breakdown of platform-specific satisfaction and waiting times is presented in the accompanying chart.

Satisfaction levels: mostly Poor to Neutral



Mostly poor to neutral overall

The waiting game: Days vs. Years



Extreme delays stall critical research

The application labyrinth: Delays, denials, and deterrents

The process for formally requesting data has proved to be a labyrinth of obstacles, falling into three recurring categories of failure - excessive delays, arbitrary denials, and deterrents that discourage applications altogether.

- 1. Excessive delays:** Already in 2024, the GLOBSEC report identified months-long waiting times to receive data from the VLOPs as a standard barrier. 2025 did not bring any significant improvement with waiting times lasting from few weeks to several months. The survey revealed one researcher going through a 21-months long process – applying in December 2023 and only receiving dysfunctional, limited access in September 2025 after navigating multiple flawed rejections and lengthy appeals with TikTok. Other researcher waited 10 months for any substantive response from Meta and YouTube. Such timelines render timely research, especially on fast-moving events like elections, impossible.
- 2. Arbitrary denials and gatekeeping:** 23% of respondents were denied access to data while platforms frequently reject applications with unsubstantiated or incorrect reasoning. For example, applications were denied on the ground that the research allegedly “did not contribute to DSA systemic risk assessments”. In another case, applicant’s institutional status as a non-profit academic institution, led to a rejection that was only overturned after a lengthy appeal process. In other cases, researchers received no substantive reply from the platforms, which acknowledged receipt but did not provide a decision. Together, these cases highlight significant shortcomings in platforms’ transparency, consistency and compliance with the spirit of Article 40(12), reinforcing the broader need for clearer reliable criteria for evaluating requests and enforceable standards governing data-access decisions.

- 3. Deterrents and burdensome requirements:** The application process itself is an obstacle course of deterrents. Researchers face disproportionate data-safety requirements for publicly accessible data, the inability to download data from platform “clean rooms” for independent analysis, the burdensome requirement to re-apply for access for each new project, and prohibitive legal barriers – including restrictive NDAs and liability clauses that universities and non-profits cannot sign.

Vetted researchers: Limited interest

59% of researchers are fully aware of the DSA’s vetted-researcher framework (Article 40(4)), while 26% have limited information and 15% of researchers have no information on the vetting process of researchers. With that, only 38% of inquired researchers were planning to acquire the status of the vetted researcher. Where differences emerge is in the intention to apply. NGOs appear more willing or prepared to apply than academic or research institutions. By contrast, academia and research centres/think tanks either responded “No” or expressed uncertainty, which may reflect institutional caution, longer internal procedures, or the need for complex approvals. Such lack or unwillingness to use this tool by the research community could be driven by the observed lack of information, but also by increasing attacks and exposure of researchers working on the topics related to threats in online information environment.

Scepticism and unmet expectations

While the DSA and the delegated act on data access were designed to establish conditions and frameworks, the survey data reveals that researchers regard these mechanisms – particularly the proposed timelines and vetting process – as **profoundly unworkable in their current form.**

The survey revealed deep scepticism, with many expecting the new framework merely to formalise, rather than resolve, existing barriers.

The problem with the timelines

The delegated act's proposed timeline – which allows a single data access request to take anywhere from 80 days to nine months – represents a critical design flaw. While the act does set procedural timelines at each stage, the overwhelming majority of researchers regard the overall timeframe as “not feasible at all” or “rather not feasible,” describing it as “unreasonably long” and incompatible with the pace of their work. Respondents engaged in election monitoring, crisis response, FIMI analysis, or security research were particularly emphatic: a process that could span up to nine months would render data “obsolete,” “irrelevant,” or “impossible to use” for research that must track rapidly evolving information environments.

Moreover, a nine-month waiting period is structurally incompatible with grant-funded projects whose total duration may be shorter than that, creating a “chicken-and-egg” problem: researchers cannot secure project funding without a guarantee of data access, yet cannot apply for data access without a funded project. This procedural impasse threatens to paralyse public interest research before it even begins.

For several organisations, the lag would force pre-emptive applications based on speculative research questions, undermine their ability to plan grant proposals, or exclude them entirely from time-sensitive work such as FIMI monitoring and threat detection.

Anticipated barriers to application

Researchers anticipate new obstacles and administration when navigating the formal DSA

application process outlined by the delegated act, creating an uncertainty that may deter many from even applying. Key anticipated obstacles include:

- **Legal and administrative requirements:** Respondents noted that the need to prepare detailed project documentation, research protocols and secure data-environment specifications at early stage, before confirmation that the requested data exist or will be made available, was as a resource-intensive step. Taken together with necessary compliance, and the fact that this process would have to be repeated for every project creates a significant resource barrier for organisations.
- **Ambiguous eligibility criteria:** A lack of clarity on the definition of a “research organisation” under the Article 40(8) of the DSA. Many non-profit organisation, fact-checking organisations, and think tanks fear they may be considered ineligible despite to be granted status vetted researchers despite conducting critical research. There are also concerns about how platforms will interpret the “independence from commercial interests” clause, potentially using past fact-checking partnerships as a pretext for denial. Other concerns included lack of formal academic affiliation, not-for-profit legal status, or whether prior project partnerships with platforms could be used to question their independence.

The combination of pre-existing and anticipated barriers is **creating a systemic risk of inequity** in research capacity across the EU and beyond. In 2024 the mapping of access to data concluded that the reliance on expensive commercial monitoring tools - often the only alternative to direct platform access - already disadvantages smaller organisations and those in member states with lower purchasing power. The anticipated legal and administrative costs of the formal DSA process will only exacerbate this divide. The result will be a two-tiered system of research, where only the wealthiest and best-resourced institutions can afford to conduct the independent scrutiny that the DSA was meant to democratise. The clear identification of these deep-seated problems demands an equally clear set of concrete, actionable solutions for policymakers.

Recommendations

The survey data shows that researchers need help preparing submissions and require clearer information to navigate the process. The research community identified several areas where support from the EU or other bodies would be valuable:

- clear procedural guidance;
- annotated application templates;
- model documentation for GDPR and data-processing compliance;
- examples of successful applications, and step-by-step instructions;
- trainings on secure data environments, legal frameworks, and data storage and archiving requirements.

While researchers understand that the operationalisation and enforcement of the DSA including its delegated act on the access to data are a work of process, the current data available indicate a warning. To ensure objectives of the DSA Article 40, the following policy measures would be welcomed by the research community to translate the DSA into a functional, equitable, and effective system for independent oversight of digital platforms:

- **Timely access to data:** Shorter deadlines to approve or deny data access requests.
- **Provision of justification and appeal:** Under the delegated act, platforms must provide substantiated, evidence-based reasoning for any refusal of access and must offer a clear avenue for appeal. The European Commission and national DSCs should monitor and enforce compliance with these obligations.
- **Ensure usability:** Platforms should provide access to publicly available data via user-friendly interfaces with generous data quotas. “Clean room” requirements that prevent researchers from downloading and analysing publicly accessible data in their own secure environments are useless.

- **Clearer and broader eligibility for vetted researchers:** Vetting approval must be simplified and made more inclusive. Approval should be granted at an **organisational level**, not on a burdensome project-by-project basis and should remain valid for a minimum of two years. Eligibility criteria must be clarified to explicitly recognise the vital research contributions of non-partisan NGOs, think tanks, and fact-checking organisations on equal footing with academic institutions.
- **Data utility and functionality:** Access without utility is meaningless. It is imperative that all data provided under Article 40(12) are **downloadable in user-friendly, machine-readable formats (e.g., CSV)**. Datasets must include critical metrics, including detailed engagement data (reactions, shares, comments), the full content of comments, and granular advertising data, including exact monetary amounts spent and specific targeting criteria, rather than broad ranges.

The DSA’s core objective - enabling independent scrutiny of systemic online risks - is currently facing obstacles that are a combination of platform obstructionism and procedural friction. The current framework is not working for researchers and is perceived as slow and bureaucratic. Providing researchers with timely and functional access to platform data is fundamental to Europe’s democratic resilience and security.