OECD*publishing*

NEW PERSPECTIVES ON MEASURING CYBERSECURITY

OECD DIGITAL ECONOMY PAPERS

June 2024 **No. 366**

OECD
BETTER POLICIES FOR BETTER LIVES

Foreword

Measuring the various aspects of cybersecurity across countries is challenging, in part because the actors in the cybersecurity ecosystem often do not have the incentives to share key data. At the same time, people, firms and governments need to feel secure to communicate online and use Internet-based services. This statistical report provides an overview of how cybersecurity is being measured across a variety of data sources and using different methodological approaches. Beginning with a checklist of measurement considerations, the report then discusses existing data from official and non-official sources, identifying when each data source is most useful. The report then provides two proofs of concepts for measuring uncertainty related to cyber risks, or "cybersecurity uncertainty". Measuring such uncertainty can complement existing statistics and help anticipate emerging cybersecurity trends, develop more targeted cybersecurity awareness programmes, and promote a more secure and resilient digital ecosystem.

This report was written by Simon Lange, Molly Lesher and Nicolas Benoit, with contributions from Mercedes Fogarassy, Anne Fornacciari and Pierre Montagnier. It benefitted from feedback from Audrey Plonk, Bénédicte Schmitt, Peter Stephens, Jeremy West and Bill Woodcock. This paper was approved and declassified by the OECD Digital Policy Committee on 4 December 2023 and prepared for publication by the OECD Secretariat.

Note to Delegations:

This document is also available on O.N.E Members & Partners under the reference code:

DSTI/CDEP/MADE(2023)14/FINAL.

This document, as well as any data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

Note by the Republic of Türkiye

The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Türkiye recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Türkiye shall preserve its position concerning the "Cyprus issue".

Note by all the European Union Member States of the OECD and the European Union

The Republic of Cyprus is recognised by all members of the United Nations with the exception of Türkiye. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

Table of contents

Tables

Figures

Boxes

1 Introduction

To fully benefit from the opportunities of the digital transformation, individuals, firms and governments need to trust that engaging in digital environments will bring more advantages than disadvantages. Managing cybersecurity risk is an important part of building trust among unknown actors in the anonymity of cyberspace. Cybersecurity incidents such as breaches of the availability, integrity or confidentiality of data, systems and networks, reduce trust, which can potentially slow down the spread of technologies that otherwise would have significant benefits for economic growth and social welfare.

As people increasingly use a wide variety of online services, firms become more reliant on digital technologies. At the same time, malicious actors are increasingly sophisticated, leading to a rapidly evolving threat landscape in recent years. Estimates of the financial damages caused by cybersecurity incidents vary from hundreds of billions to trillions of US dollars (OECD, 2022[1]), putting into sharp relief the differences and difficulties inherent in the measurement of cybersecurity. At the same time, there is a widely shared perception that cybersecurity incidents have been increasing in both frequency and severity. For instance, cyber incidents were ranked the single biggest global business risk in 2020 and 2022 in the Allianz Risk Barometer, with 44% of respondents citing them as one of their three primary concerns in 2022 (Allianz, 2022[2]).

Cybersecurity is often characterised as suffering from significant market failures (Moore, 2010[3]).[1] First, one firm's investment in cybersecurity will typically also benefit its clients and suppliers. If there is no way for these entities to share in the costs of increased cybersecurity, this kind of externality will result in an under provision of cybersecurity and a rationale for governments to take corrective measures. Second, asymmetric information about the cybersecurity of products and services can result in market failures, as is the case when consumers are unable to verify the degree to which a device or service is safe to use (OECD, 2021[4]). In contrast, critical infrastructure is in some cases directly provided by governments (i.e., there is no market process that could fail to begin with and thus a direct need for policy makers to factor cybersecurity into their decision-making processes).

Assessing a country's cybersecurity posture is an important step in developing a policy framework that fosters cybersecurity risk management practices that are well co-ordinated with and mutually reinforcing other policy domains that instil trust.[2] Policy makers need to know whether businesses, consumers and public services are exposed to an acceptable level of risk from cyber threats, and whether and how policies contribute to their safety. Ideally, they will want to trace the impact of policy measures from their design and implementation through to behavioural changes and impact.

Yet quantifying the threat landscape, intended and unintended cyber incidents, and the adequacy of cybersecurity policies and capabilities has so far been challenging. An early stocktaking exercise by the OECD finds that "the lack of empirical data on response-related activities [is] apparent" (OECD, 2012[5]). More recently, a report compiled by a working group in the context of the Paris Call for Trust and Security in Cyberspace (French Ministry for Europe and Foreign Affairs, 2018[6]) highlights the need for better monitoring of the evolution of cyberspace stability (GEODE, The Hague Centre for Strategic Studies and Cyberpeace Institute, 2021[7]). But it also acknowledges as a significant obstacle the availability and accessibility of relevant and independently identifiable data and a lack of standardisation.

This report identifies a checklist of considerations that should be considered when developing new cybersecurity indicators. It also assesses the availability and suitability of official data sources for cybersecurity measurement (e.g., sample surveys and administrative data) as well as non-official data sources (e.g., policy surveys and data from private sector firms). The report further outlines two innovative approaches to constructing indices of uncertainty in cyberspace using news reports and Google Trends data. The final section summarises the findings of the report and looks ahead to future work on measuring cybersecurity.

2 Checklist of cybersecurity measurement considerations

Measuring and monitoring levels of and trends in cybersecurity is a complex undertaking. While in principle many data sources are available, they speak to different aspects of cybersecurity and have different strengths and weaknesses. Data gathering in this area is also difficult given that most actors involved in the cybersecurity ecosystem do not wish to reveal challenges, the incidents they have experienced, or their overall level of cybersecurity preparedness to malicious entities. This section identifies a checklist of considerations that should be considered when developing new cybersecurity indicators.

What should be measured?

Defining the phenomena to be measured

Statistical definitions and standards are a key first step in any measurement exercise. "Cybersecurity" is a broad term that encompasses many different aspects of security in digital environments. From an international policy perspective, cybersecurity can be viewed as a broad and multifaceted challenge, aimed at supporting policies within the following four domains: economic and social, technical, law enforcement, and national and international security (including warfare and espionage). OECD work focuses on "digital security" policy issues, which refer to the economic and social aspects of cybersecurity, as opposed to purely technical aspects and those related to criminal law enforcement or national and international security (OECD, 2022[8]). From a measurement perspective, it is impossible to isolate the economic and social aspects of cybersecurity from the other domains with the data that is currently available. As a result, this report focuses on measuring cybersecurity in its broadest sense. Future work, including in the context of the review of the implementation of the OECD Digital Security Recommendations, may be able to isolate the digital security aspects of cybersecurity.

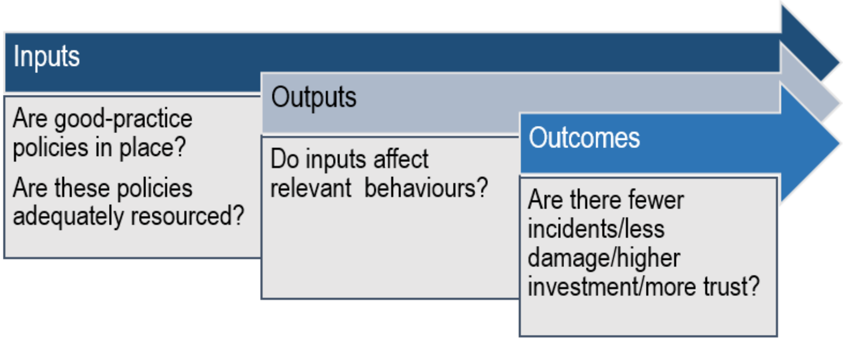
Identifying the actors: Government, firms and consumers

Achieving high levels of cybersecurity requires action on the part of a wide range of actors. Governments design and implement policies aimed at aligning other actors' incentives with an overall secure digital environment. They also need to ensure that their own data operations are secure. Firms need to take appropriate measures to protect themselves and their customers against security threats such as data breaches, and consumers need to exercise diligence online. Some actors within these larger groups have additional responsibilities. They include firms that provide digital services, publish software, or manufacture or maintain relevant hardware and critical infrastructure. They also include cybersecurity professionals and others whose work involves security-relevant tasks. Ideally, indicators of cybersecurity would relate to each of these actors (e.g., cybersecurity skills and the tools these actors have in place to identify and address vulnerabilities and incidents).

Policy impact: Inputs, outputs, and outcomes

The impact of policy measures to address cybersecurity threats runs from inputs to outputs to outcomes (Figure 2.1). For instance, new regulations that address cybersecurity risks are expected to change the behaviour of firms or individuals that fall under their purview, although the extent to which they do so depends on how they are designed and implemented. These behaviours, in turn, may or may not result in tangible outcomes – fewer incidents, less damage, reduced uncertainty, more trust, etc. Should indicators focus on inputs, outputs or outcomes? And where are the greatest knowledge gaps?

Figure 2.1. Assessing policy impact via inputs, outputs and outcomes


Inputs
Are good-practice policies in place?
Are these policies adequately resourced?
Outputs
Do inputs affect relevant behaviours?
Outcomes
Are there fewer incidents/less damage/higher investment/more trust?

While conceptually the distinction between inputs, outputs and outcomes is clear, in practice the line between them is not always obvious. For instance, from the perspective of the government, new legislation could be considered an output produced with inputs such as staff time and other resources and through activities such as drafting and consultations. However, the generic example above assumes that laws and regulations are an input into the production of actions taken by firms and individuals that might then lead to an outcome such as increased cybersecurity. For this report, the following distinctions are adopted:

- **Inputs** include all resources, activities and intermediate outputs (e.g., legislation) but exclude any behavioural changes on the part of relevant actors. They thus include first-line inputs such as funding and staff time, activities such as stakeholder consultations and awareness campaigns, and laws and regulations.
- **Outputs** include all behavioural changes that result from inputs as defined above. For instance, firms might adopt new measures to strengthen the digital defences and consumers might make different security-related choices online in response to an awareness campaign.
- **Outcomes** include realised security threats (e.g., the number of incidents and the harm that they do) and the level of trust in digital environments, among others. Outcomes are what ultimately matters.

Inputs – and to a lesser extent outputs – are often straight-forward to observe. Information on laws and regulations is typically publicly available, and household and enterprise surveys can provide information on which measures individuals and firms are taking. However, it is not always clear that a given set of inputs or outputs will lead to the desired outcome or that it will do so in the most efficient way. In extreme cases, some may even be counterproductive. Hence, their inclusion in a dashboard should be justified, ideally through evidence of their effectiveness.

Outcomes, on the other hand, are often very challenging to measure accurately for several reasons (see below). In addition, outcomes are determined not only by inputs and outputs associated with a specific intervention, but rather by third factors that are beyond the control of policy makers. For instance, an increase in the number of activities that come to rely on digital technologies is a key driver of the intensity of cybersecurity incidents. Hence, linking them causally to inputs and outputs will typically require advanced econometric methods.

Frequency, timeliness and comparability over space and time

Frequency and timeliness

Cyber threats tend to emerge rapidly and unexpectedly. This raises the stakes for compilers of statistical information – especially for data related to outcomes – to be available at high frequency and in a timely fashion. Indicators can differ substantially in terms of their frequency and timeliness. While some indicators, such as highly curated official statistics, might be available only annually or even less frequently, others can potentially be available at monthly, weekly or even daily intervals, though with less data cleaning. Similarly, some indicators become available only with a substantial time lag, while others might be available in a timelier fashion.

Comparability over space and time

Evidence on what works and what does not can be gleaned from cross-country comparisons of leading and lagging country performance. And because of the interconnectedness of digital economies, cybersecurity measurement has an important international dimension. Any cybersecurity indicator will thus be more valuable the more countries/jurisdictions it covers, but data availability and comparability differ from indicator to indicator. Hence, there is typically a trade-off between the number of countries/jurisdictions covered and the comparability of the data.

The salience of the trade-off between comprehensiveness and coverage also depends on the final use of the data. Missing data or differences in statistical classifications, for instance, are more easily accommodated within dashboards, while the construction of an aggregate index typically requires imputation (or else the entire series must be discarded). Imputation entails assumptions, and thus perhaps a less accurate depiction of the underlying phenomena that the index is trying to measure.

In addition, the nature of cyber threats – and with them the importance of different preventive measures – tends to evolve rapidly. This implies that an indicator that is highly relevant today, for instance because it provides information about the extent of cyber risk emanating from a prominent source, might not be as relevant in the future as technology and the strategies actors employ changes. As a result, a continuous cycle of assessing the appropriateness of cybersecurity indicators is important.

Interpretability, incentive-compatibility and transparency

Interpretability: Simultaneity and omitted variables

To ensure that a set of indicators is easily digestible for users, the indicators need to have a clear interpretation. Not all relevant indicators do. For instance, vulnerability databases are platforms that collect, store, and disseminate information about computer security vulnerabilities.[3] They play an important role in the management of cybersecurity risks and are a source of relevant data in this respect. But what does it mean for a vulnerability database to have many entries? It could mean that a lot of effort goes into finding and reporting on vulnerabilities. But it could also mean that much of the code that is published subsequently exhibits vulnerabilities (i.e., that it is written without the necessary level of diligence). The interpretation of an indicator like this would not be clear.

The issue here is one of *simultaneity*, a well-known phenomenon in empirical research in criminology, where the question frequently arises as to what extent policing reduces crime or more crime prompts a higher police presence. The likelihood is that they both cause each other. An extreme case is *reverse causality*. If policing has no effect on crime but more crime causes a higher police presence, an indicator such as the number of police officers per population would simply be a proxy for crime. While there are ways to deal with these issues in empirical research, an indicator that becomes part of an aggregate index or a dashboard should ideally be clearly interpretable. There is thus a case to be made to exclude indicators that are subject to simultaneity. In general, inputs (e.g., spending on cybersecurity) and outputs can be subject to simultaneity bias that may affect the interpretability of indicators.

A related concern that can affect interpretability is *omitted variables bias*. Some of the results in Section 3, for example, assume that contrary to what one might expect, a higher incidence of cyber incidents in firms across countries is associated with more preventive measures – that is, firms that appear to take more measures are also those that report more incidents. One explanation for such a finding is that there is a third variable that drives both actions to prevent incidents and observed incidents. An obvious candidate in this case is the degree of digitalisation. Firms that use digital technologies more intensively will also be at a greater risk to incur incidents. They will also be more likely to take preventive measures, investing more in cybersecurity risk management.

An index versus a dashboard of indicators

The choice of creating an index (a 'scalar') or a dashboard (a 'vector') also relates to transparency and interpretability. Dashboards provide more information, and they are valuable for those aiming to understand a complex phenomenon. They give users the option to weigh the importance of a set of indicators in line with their priorities and needs. Dashboards provide more transparency for users – each indicator is presented as is – and require fewer assumptions on the part of the producer, although judgments must still be made in terms of which indicators to include and how to present them. At the same time, dashboards require an active effort on the part of the user to process complex information, and less informed users may have difficulties in interpreting the importance of some of the indicators.

Table 2.1. Indices versus dashboards

	Index	Dashboard
Transparency	↓	↑
Ease of communication	↑	↓
Information content	↓	↑
Comparability across countries	↑	↓

Indices, on the other hand, reduce a complex phenomenon to a single number, which becomes easy to compare and communicate. Indices generate unambiguous rankings, which can create strong incentives to act.[4] However, doing so requires judgments on part of the producer that are typically not easy to justify, and the index "number" has no meaning in and of itself. Many indices adopt either the equal weighting of the different indicators or statistical methods (such as principal component analysis). Expert judgement is another technique sometimes employed by the producers of indices. Equal weights are often presented as a (false) agnostic choice, while statistical methods are presented as an effort to "let the data speak". However, the decision to adopt equal weights also constitutes a judgement by the producer and statistical methods can produce weights that experts disagree with.

Incentive-compatibility can result in under- and over-reporting

To the extent possible, any information on cybersecurity should be obtained and reported in a way that is incentive compatible. All actors can face incentives to underreport incidents. Firms, for instance, will be less willing to share information on incidents or challenges they experience if they anticipate that doing so would have adverse consequences for the firm or individuals associated with it. Similarly, as with other crimes, individuals might be reluctant to volunteer information about their status as victims of cybercrimes (e.g., out of shame), resulting in an underestimation of the true extent of incidents. Finally, governments that anticipate the use of information they provide in an index or dashboard and conclude that this use will affect their reputation or outcomes such as investment can also have an incentive to engage in window-dressing (Aragão and Linsi, 2020[9]; Kerner, Jerven and Beatty, 2017[10]).

While the above considerations lead to an expectation of significant underreporting, specific actors can also face incentives to overreport incidents. In particular, firms that offer cybersecurity services or products often have access to significant data that are informative about trends in cybersecurity. Yet unlike government sources, they also face incentives to make the challenge look larger than it is.

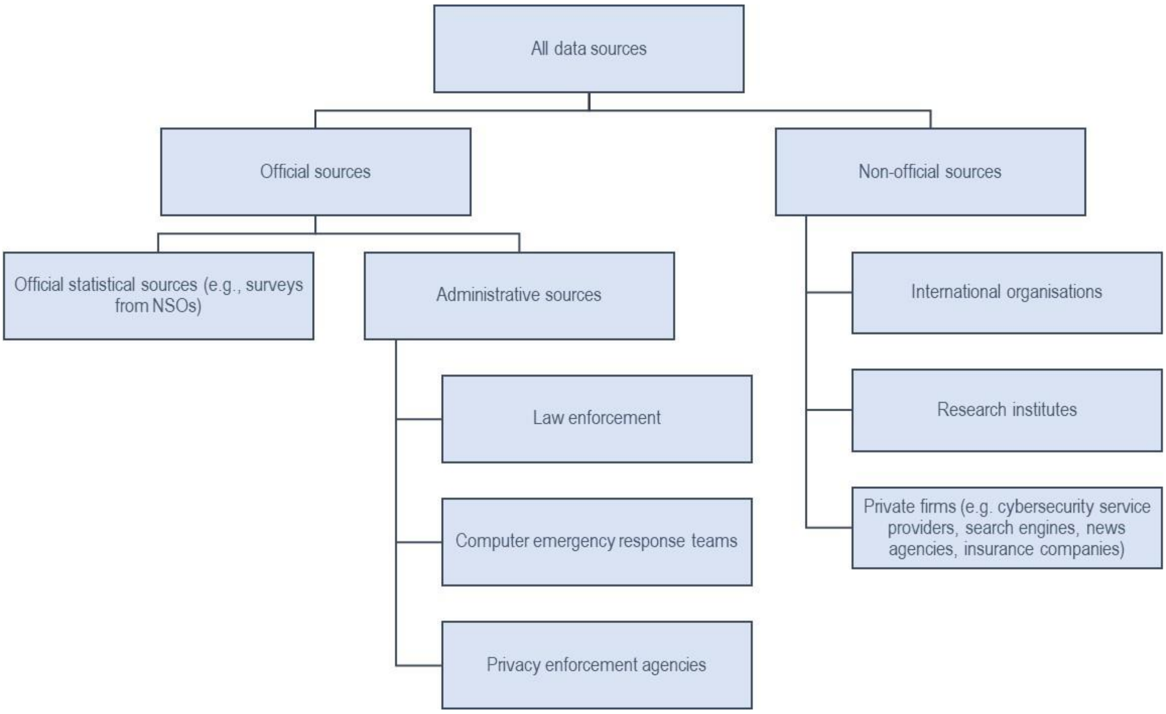
Transparency: Avoiding black boxes

To ensure they are not misused, any data series that are reported should have a clearly stated and retraceable methodology. If indicators are derived from a survey, the survey instrument should be made available as well as accompanying metadata such as sampling design, response rates, etc. Similarly, if an indicator is derived from qualitative sources, the inclusion and exclusion criteria should be clear. While this might seem obvious at first blush, there are several sources of data that will be discussed below that fail to meet these criteria.

Data sources

Different data sources also play a role in cybersecurity measurement considerations (Figure 2.2). At the highest level, there are two mutually exclusive types of cybersecurity data sources: official and non-official sources. Official sources refer to data collected by government entities for an official purpose and non-official data sources refer to all other data sources. Within official data sources, a useful distinction can be made between statistical sources and administrative sources. Statistical sources are all data that are collected for the explicit purpose of statistical reporting, typically through surveys and censuses. In contrast, administrative sources refer to data that are not primarily collected for statistical reporting but for other administrative purposes. In the case of cybersecurity, obvious sources of administrative data include law enforcement agencies, national computer emergency response teams, and regulators, among others. Non-official data can come from a wide range of different actors, including international organisations, academia, news agencies, private firms, etc.

Figure 2.2. A schema of cybersecurity data sources


All data sources
Official sources
Non-official sources
Official statistical sources (e.g., surveys from NSOs)
Administrative sources
Law enforcement
Computer emergency response teams
Privacy enforcement agencies
International organisations
Research institutes
Private firms (e.g. cybersecurity service providers, search engines, news agencies, insurance companies)

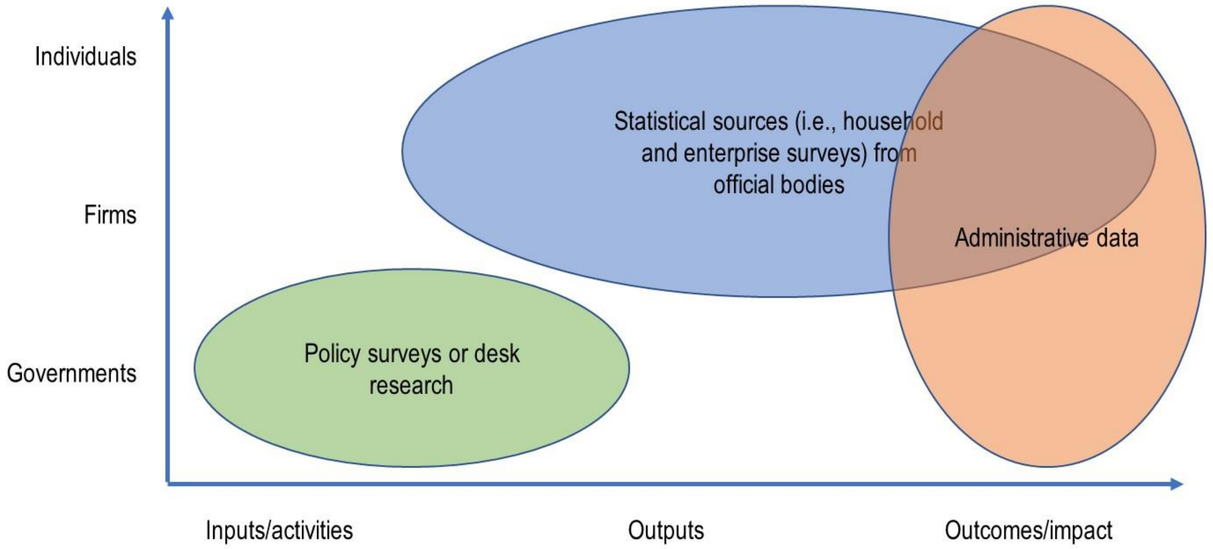
Data producers differ in terms of methods, the incentives they face and their focus

There are important differences across different data sources in terms of the methods of data collection. Official sources are typically based on either sample surveys administered to firms and individuals or administrative registers. Both are underpinned by laws and regulations, compelling individuals and firms to provide information and ensuring confidentiality. Non-government sources, on the other hand, lack recourse to legal instruments. They are more likely to use non-sample surveys (e.g., policy surveys administered to governments), desk research or expert assessments.

In addition, actors can differ in terms of the incentives they face when they publish data. Private firms that are primarily in the business of selling security solutions, for instance, might have an incentive to overstate the number of incidents in one way or another. Two reports commissioned by the United Kingdom illustrate this issue (Detica, 2011[11]; Anderson et al., 2013[12]). The first, which was commissioned by the Cabinet Office, put the total cost of cybercrime in the United Kingdom at GBP 27 billion, and the second, which was commissioned by the UK Ministry of Defence, put it at around USD 170 million. The former was authored by BAE Systems Digital Intelligence (at the time called Detica), a subsidiary of security company BAE Systems, while the latter was authored primarily by academics.

Different data sources also tend to cover different actors and steps in the chain from inputs to outputs to outcomes. For instance, official information and communication technology (ICT) surveys tend to be a major source of information on actions taken and outcomes at the level of individuals and firms, whereas information about policies are typically collected through policy surveys or desk research (Figure 2.3).

Figure 2.3. Stylised amenability of different data sources


Individuals
Firms
Governments
Statistical sources (i.e., household and enterprise surveys) from official bodies
Administrative data
Policy surveys or desk research
Inputs/activities
Outputs
Outcomes/impact

3 Cybersecurity data from official sources

At the highest level, this report distinguishes between two different sources of data: official statistics and non-official statistics. Official statistics are statistics collected, curated and disseminated by the national statistical system, which includes statistical organisations and units within a country that jointly collect, process, and disseminate official statistics on behalf of national governments (OECD, 2020[13]). While national statistical offices are typically at the heart of the national statistical system, any government entity that collects and disseminates data is also part of the national statistical system. Non-official sources include all other sources of data (see Section 4).

Official statistics differ from non-official statistics in three important ways. First, official statistics are usually collected within a legal framework. In the case of statistical sources such as sample surveys and censuses, for instance, governments might legally oblige respondents to participate to ensure a sufficiently high level of participation, although they typically tend to exercise restraint when it comes to enforcing compliance. Administrative data, or data collected primarily for non-statistical purposes but that can nevertheless be used for statistical production, are usually sourced from official administrative entities (e.g., regulatory bodies). Second, government statistical entities in OECD countries typically enjoy a high degree of independence. In particular, the production of official statistics is typically organised along the *Fundamental Principles of Official Statistics*, which underscore citizens' entitlement to public information, impartiality, and strict professionalism (ECOSOC, 2013[14]). Finally, national statistical systems have the explicit mandate to produce data relevant to citizens and governments. For these reasons, official statistics are a highly reliable and trusted source of relevant data on cybersecurity.

Official statistical sources

An overview of official statistics on cybersecurity

Sample surveys administered to firms and individuals are one of the main sources of information on cybersecurity. They include general-purpose surveys that capture aspects of cybersecurity, ICT access and use surveys, and surveys dedicated exclusively to cybersecurity. In the case of ICT access and use surveys, modules on cybersecurity are often included. For instance, Eurostat's 2022 model questionnaire for the *European Community Survey on ICT Usage and E-commerce in Enterprises*, which is used to streamline data collection across the European Union and close partner countries, includes a 27-question module on ICT security (Table 3.1) (Eurostat, 2022[15]). Similar modules were included in the past and it is envisioned that they will be included every other year going forward. Examples of surveys dedicated exclusively to gathering information about cyber incidents and cybersecurity include the *Canadian Survey of Cyber Security and Cybercrime* (CSCSC) (Statistics Canada, 2022[16])[5] and the United Kingdom's *Cyber Security Breaches Survey* (DSIT, 2023[17]), while Brazil has focused on measuring digital security risk management among firms (NIC.br, 2020[18]).

Sample surveys typically focus on outputs and outcomes. Eurostat's 2022 module on ICT security, for instance, enquires about cybersecurity frameworks and measures (D1-D4, D6 and D7) and incidents firms experienced (Table 3.1). The 2022 questionnaires of the CSCSC also enquires about whether other parties, including regulators, require the firm to implement certain cybersecurity measures and probes deeper into the reasons firms are taking specific decisions. And while measuring the costs associated with incidents and prevention is often seen as very challenging because of a lack of correspondence with accounting practices, it also includes items to this effect.

Table 3.1. Eurostat's 2022 module on ICT security

	D1. Does your enterprise apply any of the following ICT security measures in its ICT systems?
1	a) Authentication via strong password
2	b) Authentication via biometric methods used to access the enterprise's ICT system
3	c) Authentication based on a combination of at least two authentication mechanisms
4	d) Encryption of data, documents or e-mails
5	e) Data backup to a separate location
6	f) Network access control
7	g) Virtual Private Network
8	h) ICT monitoring system that allows to detect suspicious activity in the ICT systems and alerts the enterprise about it, other than standalone anti-virus software
9	i) Maintaining log files that enable analysis after ICT security incidents
10	j) ICT risk assessment, i.e., periodical assessment of probability and consequences of ICT security incidents
11	k) ICT security tests
	D2. Does your enterprise make persons employed aware of their obligations in ICT security-related issues in the following ways?
12	a) Voluntary training or internally available information
13	b) Compulsory training courses or viewing compulsory material
14	c) By contract (e.g., contract of employment)
15	D3. Does your enterprise have document(s) on measures, practices or procedures on ICT security?
	D4. When were your enterprise's document(s) on measures, practices or procedures on ICT security defined or most recently reviewed?
16	a) Within the last 12 months
17	b) More than 12 months and up to 24 months ago
18	c) More than 24 months ago
	D5. During 2021, did your enterprise experience any ICT related security incident leading to the following consequences?
19	a) Unavailability of ICT services due to hardware or software failures
20	b) Unavailability of ICT services due to attack from outside, e.g., ransomware attacks, Denial-of-Service attacks
21	c) Destruction or corruption of data due to hardware or software failures
22	d) Destruction or corruption of data due to infection of malicious software or unauthorised intrusion
23	e) Disclosure of confidential data due to intrusion, pharming, phishing attack, intentional actions by own employees
24	f) Disclosure of confidential data due to unintentional actions by own employees
	D6. Who carries out the ICT security-related activities in your enterprise?
25	a) Own employees
26	b) External suppliers
27	D7. Does your enterprise have insurance against ICT security incidents?

Note: Some questions include further explanations for respondents.
Source: Eurostat (2022[15]).

In general, the strengths and weaknesses of sample surveys are well-understood by statisticians. Strengths include the existence of well-developed statistical methodologies and a high-level of representativeness. Potential drawbacks also tend to be well understood and can be addressed to some extent. For example, to correct for a situation in which certain firms are more likely than others to respond to surveys and so a given sample would be biased, weighting observations based on information on the entire population (e.g., from administrative or census data) can be used. In this way, even surveys with a comparatively low response rate can still be used to obtain unbiased estimates of key statistics.

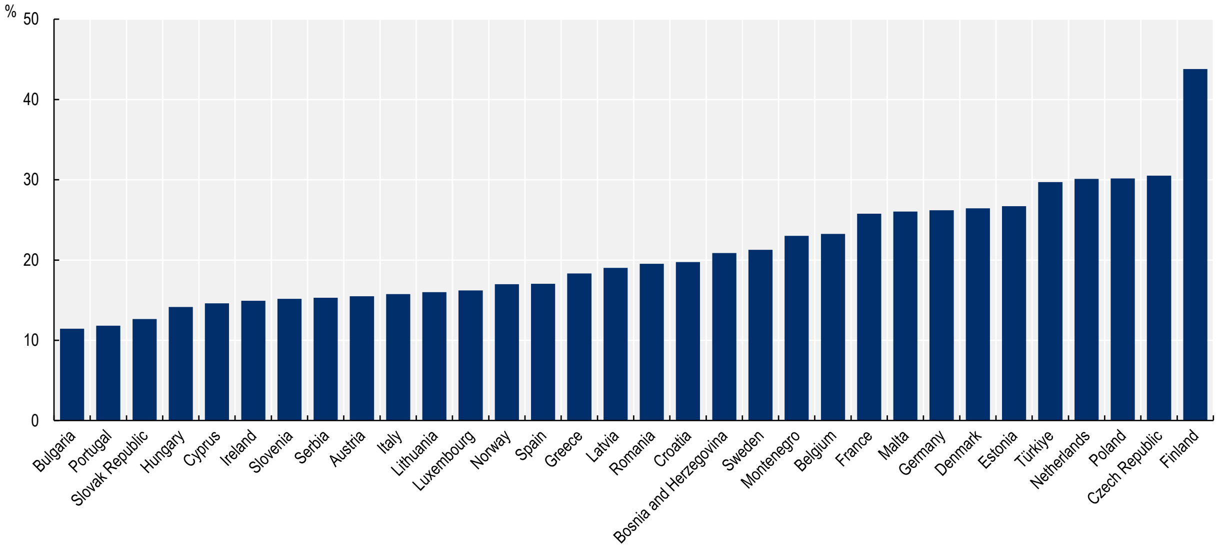
Low and falling response rates are a long-standing concern in nearly all official surveys. Even as early as 1999 and 2006, the *Journal of Official Statistics* and *Public Opinion Quarterly* devoted entire special issues to nonresponse. In the United States, response rates to most federal surveys have been declining for many years and the decline accelerated during the COVID-19 pandemic (U.S. Bureau of Labor Statistics, 2023[19]). ICT surveys are no exception, although the authors are not aware of any systematic assessment of the issue.

Limited coverage and concerns about respondents' ability to answer cybersecurity-related questions

Other concerns relate to the universe of entities covered in ICT access and use surveys. It is well-known that banks and insurance companies, hospitals, educational institutions and governments are often targets of cyberattacks. Yet there is some variation in the extent to which enterprise surveys cover industries and enterprises. Eurostat's *European Community Survey on ICT Usage and E-commerce in Enterprises* only covers business-sector firms, and it excludes finance and insurance (ISIC code "K") (see Figure 3.1).

Figure 3.1. Enterprises that experienced problems due to an ICT-related security incident at least once

Business-sector enterprises (excl. financial services) with ten employees or more in which employees have access to the Internet, 2021



Note: ICT-related security incidents include the unavailability of ICT services, the destruction or corruption of data, and the disclosure of confidential data.
Source: (Eurostat, 2022[20]).

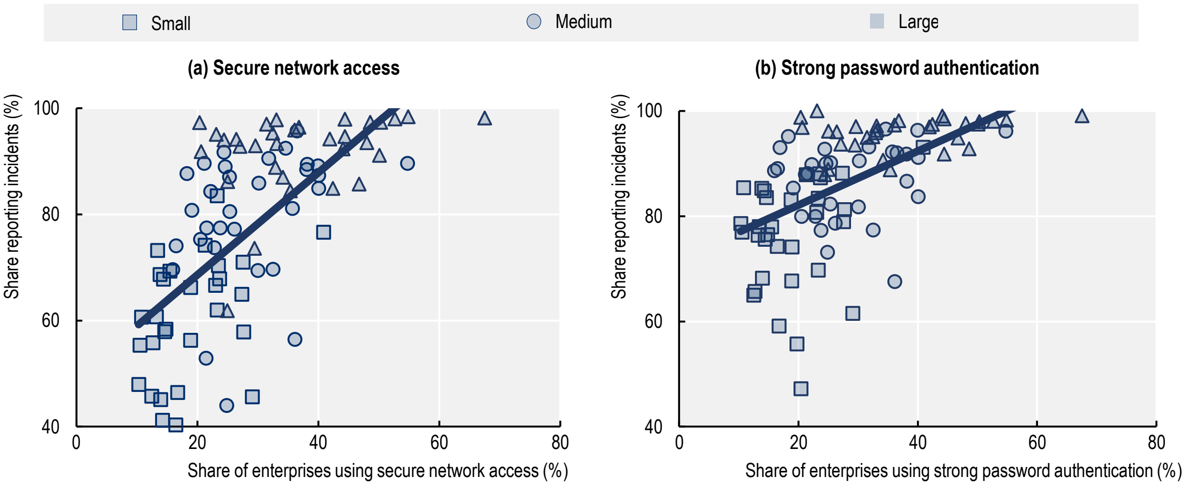
Further concerns about the drawbacks of surveys in the specific context of cybersecurity are discussed in OECD (2019[21]). They include, *inter alia*, a lack of (or systematic variation in) the technical knowledge of respondents. For example, do respondents understand the relevant technical terms? If not, are differences in technical knowledge correlated with other variables that affect whatever it is that is being measured?

The full extent of the problem is difficult to ascertain, but there are at least two pieces of tentative evidence. First, there is a significant positive relationship between firm size and the share of firms reporting incidents. While this could be driven by a higher proclivity to use digital technologies among large firms, it could also indicate that respondents for larger firms are in a better position to answer questions about incidents or that larger firms are better equipped to detect them.

There is often a positive correlation in survey data between indicators that capture the use of preventive measures (an output) and incidents (an outcome). Using data aggregated by country and firm size, Figure 3.2 illustrates this for two indicators – the share of firms that use network access control (i.e., the management of user rights in the firm's network) and the share of firms that require strong password authentication (i.e., minimal length of eight characters, mixed characters and periodic password changes). Both indicators can reasonably be expected to increase security and thus lower the share of firms that report incidents. Network access control prevents access to an organisation's network from endpoint devices that do not comply with corporate security policies, ensuring that viruses cannot enter the network from devices that do not comply with security standards. Strong passwords, on the other hand, prevent hackers from guessing or brute-forcing passwords.

Figure 3.2. There is often a positive correlation between the share of firms taking action and the propensity to report cybersecurity incidents

Share of enterprises using secure networks and strong password authentication (percent), 2021



Note: Squares, circles and triangles denote small, medium-sized and large enterprises, respectively, and the dark blue line is based on a least-squares linear fit. Network access control refers to the active management of access from devices and users to the enterprise's network. Strong password authentication refers to the use of passwords of minimal length of eight mixed characters that are changed periodically.
Source: Authors' calculations based on data from Eurostat (2022[20]).

What is going on? First, the positive correlation might be a case of reverse causality in that a higher incidence of cybersecurity incidents induces firms to take more measures going forward. If the experience of an incident leads firms to adopt more stringent measures, the result could be a positive correlation between incidents and measures taken.[6] However, the positive link seems too strong to be explained solely by simultaneity and it seems likely that other factors are at play. Second, omitted variables could also be behind the positive correlation in Figure 3.2. As already noted in Section 2, an omitted variable like digital intensity (i.e., the extent to which firms are already utilising digital technologies) could also play a role. If more firms are using digital technologies, more are likely to be at risk of ICT incidents and more are likely to take preventive measures.

Regression analysis was used to further investigate this issue (Table 3.2). While regression results reported in columns (1) and (5) come from regressions that do not include any additional controls, columns (2) and (6) show that controlling for the share of firms that are highly digitally intensive reduces the coefficient estimates on preventive measures only moderately. Accounting for digital intensity only resolves a small part of the puzzle that is the positive correlation between preventive measures and incidents.

Table 3.2. Correlations between preventive measures and incidents are likely subject to omitted variable bias

Dependent variable: share of enterprises that experienced ICT incidents, 2021

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Network access control	0.43***	0.36***	0.10	0.25***	0.11					
	(0.05)	(0.10)	(0.09)	(0.08)	(0.10)					
Strong passwords						0.65***	0.40***	0.04	0.28**	0.04
						(0.09)	(0.14)	(0.14)	(0.13)	(0.15)
High digital intensity		0.11			-0.16		0.25**			-0.12
		(0.12)			(0.13)		(0.11)			(0.13)
Own employees for ICT security			0.34***		0.46***			0.40***		0.51***
			(0.08)		(0.13)			(0.07)		(0.13)
Size-class fixed effects				Yes.	Yes.				Yes.	Yes.
R-squared	0.42	0.42	0.52	0.47	0.53	0.35	0.38	0.51	0.44	0.53
Observations	92	92	92	92	92	92	92	92	92	92

Note: How to read the table: a coefficient estimate of 0.43 in column (1) indicates that a one-percentage point increase in the share of enterprises that implement network access control is associated with an increase in the share of enterprises reporting incidents by 0.43 percentage points. Standard errors are reported in parentheses. *, ** and *** denote statistical significance at the 10%-, 5%-, and 1%-level, respectively. Please note that there is no data on the share of enterprises.[7]

Source: Authors' calculations based on data from Eurostat (2022[20]).

If a firm has dedicated in house staff to manage its cyber risks, it might also be in a better position to correctly identify cybersecurity incidents when asked about it in a survey. Firms that have no dedicated in-house staff, on the other hand, might tend to underreport incidents simply because whoever responds to the survey may not be sure whether the firm has experienced one. Again, this would indicate that the indicator on incidents is underreported, at least for some firms. It is also possible that a low level of general cybersecurity expertise may be a reason for underreporting, reinforcing the need to assess cybersecurity competencies within firms to identify individuals who may need to upgrade their skills as well as areas that are not comprehensively covered.

From the firm perspective, several related indicators are included in the OECD ICT Access and Use Databases such as modifying security settings (I3D) and restricting location-sharing (I7D) (OECD, 2023[22]) (see Annex A). However, there are no cross-country comparable official statistics measuring specific cybersecurity technical skills required for cybersecurity professionals (however, there are some non-official statistics that have been collected and disseminated by private companies, see Section 4). Across 25 OECD countries, over 40% of firms reported conducting ICT security related activities in-house (OECD, 2019[23]). However, it unclear whether having dedicated staff is indicative of having more effective preventive measures in place, as these functions can also be sourced externally via on demand cybersecurity services.

This is tested with the regressions reported in Table 3.2 columns (3) and (7). An increase in the share of firms with in-house expertise to carry out tasks related to ICT security is associated with an increase in reported incidents by 0.34 and 0.40 percentage points, respectively. It seems unlikely that having in house cybersecurity expertise is in any way an inferior strategy. At the same time, the coefficient estimates on the variables capturing preventive measures become small and statistically insignificant, providing some support to the hypothesis of reporting bias in the outcome indicator.

Two more tests are conducted for both indicators. First, in Table 3.2 columns (4) and (8), size-class fixed effects are included to control for a broad range of other factors that vary with firm size. This has a negative effect on the coefficient estimates on the variables of interest although they remain positive, sizable and statistically significant. Finally, in Table 3.2 columns (5) and (10) all controls are included at once. Here again, the coefficient estimates on the variables of interest drop significantly and turn statistically insignificant. Only the coefficient estimate on the share of firms in house expertise to carry out ICT security-related tasks remains large and statistically significant and there is little gain in terms of the models' explanatory power *vis-à-vis* the models that only include the variable of interest and the share of firms with in-house cybersecurity capacity.

If firms have a strong incentive to conceal incidents as they are wary of potential reputational damages or liability issues, this suggests that indicators related to incidents are underreported, at least for some firms, because of a lack of trust in national statistical offices. However, this seems unlikely for most OECD Members for two reasons. First, national statistical offices enjoy a high degree of independence and are keenly aware of the need to be trusted by respondents to keep their data confidential. And while data leaks from national statistical offices are not unheard of, they are very rare. Second, if this were a major issue, one would expect that the overall share of firms reporting incidents in 2021 in Eurostat surveys would be lower. Instead, an average of nearly one in five small firms (10-49 employees) report at least one incident, rising to nearly two-in-five for large firms. The issue is not so much that few firms report incidents, but rather that the pattern of reporting and the relationships between incidents and other variables are difficult to disentangle.

Indicators from sample surveys on outcomes should be used only very carefully, while data on security measures are likely informative

What should analysts and policy makers conclude from this? As noted above, sample surveys remain one of the most valuable sources of information on the actions taken by firms and individuals and outcomes. But their weaknesses also mean that the resulting data can be difficult to interpret. It is not clear, for instance, whether the country comparison based on the Eurostat data shown in Figure 3.1 captures differences in the cyberthreat landscape or other factors related to the concerns discussed above.

At a minimum, data on incidents reported by firms in sample surveys should be interpreted very carefully. There are various reasons to believe that they are underreported, most likely because firms differ in their capacity to respond to questions about cybersecurity incidents. Data on outputs, however, are likely to be accurate and their use should be encouraged, not least to enable research that unpacks links between indicators of cybersecurity further. For instance, it seems worthwhile exploring the empirical relationships detailed above further in the underlying microdata, which would allow for controlling for firm-level heterogeneity in much more detail.

Administrative data

Administrative data are linked to laws and regulations

Contrary to sample surveys and censuses, administrative data are not primarily collected to inform official statistics, but rather to support administrative units. Nonetheless, there are many examples in which administrative data have proven to be a highly valuable source of information for official statistics. In the context of cybersecurity, sources of relevant administrative data may include law and privacy enforcement agencies as well as national government computer security incident/emergence response teams (CSIRTs/CERTs).

However, one drawback of administrative data is immediately clear: their production is closely linked to laws and regulations and there can be substantial variation across countries in terms of, say, firms' reporting requirements or classifications of cybercrimes. Except for countries in regional arrangements (e.g., the European Union), a country's domestic laws and regulations are unique, making the use of administrative challenging when comparing across countries. The usefulness of administrative data for the development of cybersecurity indicators hinges on the extent to which different national authorities use standardised classification systems (OECD, 2012[5]).

Data from law enforcement agencies typically suffer from substantial downward bias

Data collected for administrative purposes collected by law enforcement agencies can potentially speak to the incidence of cybercrimes. The US Federal Bureau of Investigation's Internet Crime Complaint Center (IC3), for instance, collects data of Internet crime from the public which are then published in an annual report (FBI, 2022[24]; FBI, 2021[25]). In Europe, the European Cybercrime Centre (EC3) was set up in 2013 with the objective to serve as the central hub for criminal information and intelligence, among others. The Internet Organised Crime Threat Assessment (IOCTA) is issued annually, and it is informed by operational information shared with Europol by European Union Member States and third parties, combined with expert insights and open-source intelligence (Europol, 2023[26]). However, the report itself does not include any data.

As noted in Section 2, data from law enforcement often tend to be subject to reporting biases, either because cybercrime goes undetected or because victims are reluctant to report crimes to authorities. As with other administrative data, it is also often unclear to what extent indicators are comparable across countries as different agencies might use different standards and definitions.

Data from privacy enforcement agencies

Privacy enforcement authorities often implement regulations that requires firms to report data breaches. A recent report based on an online survey of relevant privacy enforcement authorities analyses the availability and comparability of data originating from data breach notifications across 36 countries, 23 US States, and one US territory (Iwaya, Koksal-Oudot and Ronchi, 2021[27]).

The report finds that most countries and jurisdictions in the sample – even outside of the group of countries that have adopted the European Union's General Data Protection Regulation (GDPR) – have either introduced mandatory data breach notifications or were expecting to do so soon. Many governments therefore regularly collect data on data breaches and their characteristics, including causes and the type of data affected. Data on breaches were found to be publicly available in 80% of GDPR countries, more than 50% of non-GDPR countries, and more than 40% of US states and territories. However, the report also finds that the type of data made available varies. For instance, most privacy enforcement authorities reported that they used their own classifications, limiting the usefulness of their data for international benchmarking.

As highlighted by the 2020 Census of the General Policy Assembly, publicly available data may come either from mandatory or from voluntary data breach notification (GPA, 2020[28]). Out of 70 authorities responding, 43% reported having voluntary breach notification guidelines issued in their jurisdiction, whereas 79% have mandatory data breach notification. Mandatory data breach notification varies not only across country, but also across sector. For example, of the 55 authorities that participated in the 2020 GPA census and which have mandatory breach notification requirements, 62% say these requirements apply to the public sector, while 69% say the requirements apply to the telecommunications sector. Given the wide variety of data breach regulations, cross-country comparison of privacy enforcement agencies data is challenging.

Data from national security incident/emergency response teams (CSIRTS/CERTs)

For many private and public entities, their dedicated CSIRTs/CERTs are on the front lines of all digital security incidents, as well as behind the scenes trying to prevent and mitigate threats and risks. They are core to the overall strength of cybersecurity risk management which is why accurately measuring and quantifying CSIRT/CERT activities could be useful in assessing a country's overall cybersecurity posture. There is, however, a lack of standardisation across the global CSIRT/CERT landscape, from incident classification to benchmarking data and maturity measurement, and as a result, comparing capabilities amongst CSIRTs/CERTs is challenging.

Being able to compare data produced by and relevant to CSIRTs/CERTs could benefit policymaking. Elements such as the scope or mandate of a CSIRT/CERT, their constituency, and the types of incidents dealt with all vary and are understood differently, which make direct comparisons impracticable. In contrast, data such as the budget, skills, personnel, and types of co-operation can be easily compared (and in theory obtained), so collecting and beginning with those may be a good place to start building a database of CSIRT/CERT-relevant comparable indicators.

Another way to compare CSIRTs/CERTs is based on their maturity level, which is an idea gaining ground internationally and supported by global CSIRT/CERT networks such as FIRST and TF-CSIRT. However, this type of comparison must be taken as a correlation to a country's resilience rather than a causal factor as more analysis needs to be to link maturity with overall cybersecurity posture. Various frameworks have been developed for CSIRT/CERT maturity assessment, with the main tool to conduct such assessments being the Security Incident Management Maturity Model, also called SIM3 (ENISA, 2023[29]). CSIRT/CERT communities and networks around the world recommend being audited using SIM3 to assess how well a CSIRT/CERT governs, documents, executes and measures their work (Open CSIRT Foundation, 2022[30]). While the SIM3 tool is the widely known industry standard, the data produced from resulting audits as well as the groups which are audited are not made public.

4 Cybersecurity data from non-official sources

Non-official data sources comprise a wide range of actors: international organisations, civil society, private firms and even citizens. As a result, they come in many different forms, they cover various actors, and they can reflect inputs, outputs and outcomes. This section discusses the cybersecurity policy landscape and examples of how it is measured through a range of policy surveys and desk research before turning to indicators constructed from data sourced from private sector firms.

Policy surveys and desk research

As perceived threats have increased both in frequency and potency, the cybersecurity policy landscape has become more complex. Governments have responded by putting in place a wide range of policy instruments aimed at protecting organisations and citizens, ranging from high-level strategies to disclosure requirements, the adoption of processes and resources to deal with vulnerabilities, and education and training programmes. The OECD has been at the forefront of tracking policies in this area, marshalling good practice and providing guidance, with OECD countries adopting several recommendations relating to cybersecurity over the years.

Broadly speaking, the cybersecurity landscape often includes elements such as: national cybersecurity strategies, vulnerability management, incident reporting, policies that support the cybersecurity of critical activities, outreach campaigns and training programmes, and international co-operation arrangements. Understanding these main policy approaches serves as a basis for designing meaningful cybersecurity policy indicators.

The cybersecurity policy landscape

National cybersecurity strategies

National cybersecurity strategies are typically used to define the government’s priorities and objectives and the means of attaining them, and they situate digital security policy within the wider digital policy framework. Together with national broadband strategies, they are among the most well-established digital economy strategies and, by 2020, most OECD countries had one in place (OECD, 2020[31]). The main pillars of national cybersecurity strategies often involve: protecting critical infrastructure, capacity building, information sharing, and international co-operation (OECD, 2020[31]). A well-designed national cybersecurity strategy ensures consistency and co-ordination of digital security policies. The OECD Recommendation of the Council on National Digital Security Strategies provides high-level guidance on developing and implementing such strategies (OECD, 2022[32]).

Vulnerability management

Vulnerabilities[8] are technical weaknesses in products and information systems that can be exploited to damage economic and social activities. Identifying and remediating vulnerabilities are core cybersecurity activities, yet firms often face incentives not to reveal vulnerabilities (e.g., because their reputation or market valuation would be negatively affected) and consumers lack information about the quality of security products such as antivirus software (Moore, 2010[3]). Cybersecurity vulnerabilities are uniquely identified and logged in the Common Vulnerabilities and Exposures (CVE) database (MITRE, 2023[33]), which helps raise awareness of technical vulnerabilities and how to address them. The OECD Recommendation on the Treatment of Digital Security Vulnerabilities also provides guidance on vulnerability management (OECD, 2022[34]).

Incident reporting requirements

Some countries mandate the reporting of cybersecurity incidents. In the European Union, for instance, there are several different laws on cybersecurity incident reporting, including the EU Directive on Security of Network and Information Systems (the NIS and NIS2 Directives), which includes notification rules for cybersecurity incidents for operators of critical activities. Related to cybersecurity incident reporting are disclosure requirements for data breaches, and information disclosure laws and regulations often mandate disclosure of such incidents. Annual personal data breach reporting data are publicly available in more than 80% of countries covered by the GDPR, more than 50% of non-GDPR countries, and in more than 40% of US States, though the type of information available on data breaches varies (Iwaya, Koksal-Oudot and Ronchi, 2021[27]).

Policies that support the cybersecurity of critical activities

While many economic and social activities depend on digital technologies, some are critical as their interruption or disruption would have a serious impact on the well-being and prosperity of citizens. Some of these activities include financial services, health services, national defence, and electricity and water facilities, among others. National CSIRTs/CERTs and Security Operation Centres (SOC) play an important role in supporting the cybersecurity of critical activities.[9] When situated within government, they usually have both the expertise and a mandate to support government and industry in handling digital security incidents. In 2020, more than two thirds of the 194 countries surveyed by the International Telecommunications Union (ITU) had CSIRTs/CERTs, though their mandate and capabilities varied across different countries (ITU, 2021[35]). For instance, while some may only focus on technical aspects of monitoring, assessing, and defending systems, others might also engage in awareness raising, outreach, and research. The OECD Recommendation on Digital Security of Critical Activities sets out a range of policy recommendations to ensure that policies targeting operators of critical activities focus on what is essential for the economy and society without imposing unnecessary burdens on the rest (OECD, 2019[36]).

Outreach campaigns and training programmes

Outreach campaigns aimed at raising awareness of the importance of cybersecurity are commonly offered. The “Cyber Security Awareness Program” of the United States Cybersecurity and Infrastructure Security Agency (CISA), for instance, aims to increase the understanding of cybersecurity threats among the public (CISA, 2022[37]). In 2020, more than two-thirds of the 194 countries surveyed by the ITU ran awareness campaigns targeted at private sector or government actors and almost half had cybersecurity educational programmes (ITU, 2021[35]). Cybersecurity training programmes are also widely employed given that one breach impacts an entire supply chain, and so the lack of adequate digital security protocols in one vulnerable actor (e.g., a small and medium sized enterprise (SME)) can ripple across many firms. The German Federal Ministry for Economic Affairs and Energy’s initiative “IT Security in Industry” (*IT Sicherheit in der Wirtschaft*) aims to improve the digital security of SMEs (OECD, 2020[38]). Several OECD recommendations highlight the importance of awareness raising and training programmes, including the OECD Recommendation of the Council on National Digital Security Strategies (OECD, 2022[32]) and the OECD Recommendation on Electronic Authentication (OECD, 2007[39]).

International co-operation arrangements

Cyberspace transcends borders, rendering cybersecurity a global concern. International co-operation on everything from Internet governance to the interoperability of legal frameworks and threat monitoring and response are thus key for the effectiveness of cybersecurity. In 2020, about 60% of all countries participated in multilateral cybersecurity agreements, while 50% also participated in bilateral agreements (ITU, 2021[35]). The OECD Recommendation of the Council on Digital Security Risk Management (OECD, 2022[40]), the OECD Recommendation of the Council on National Digital Security Strategies (OECD, 2022[32]), and the OECD Recommendation on the Treatment of Digital Security Vulnerabilities (OECD, 2022[34]), among others, all underscore the importance of international co-operation on various facets of cybersecurity policy.

As noted in Section 2, monitoring, evaluating and benchmarking the impact of these policy instruments – linking their characteristics to indicators on outputs and outcomes – is a key step in ensuring that they achieve their objectives. And to do so analysts require qualitative data on the policy instruments that governments use. These data tend to come from policy surveys, which are often conducted by international organisations, as well as desk research, and which in turn form of the basis of quantitative measures.

Policy surveys and desk research provide information on the policy landscape

Policy surveys are surveys administered to relevant government authorities that aim to collect information about the institutional framework and potentially governments' actions. Policy surveys are often tailored to a specific normative policy framework.

In the space of cybersecurity, a prominent example of an index underpinned by a policy survey (and further complemented by desk research) is the ITU's *Global Cybersecurity Index* (CGI) (ITU, 2021[35]). The CGI, which was first launched in 2015, aims to "raise awareness of country-level commitments on cybersecurity, to identify strengths and areas of improvement, and share current cybersecurity practices" (ITU, 2021[41]). It is based on a framework that features five pillars – legal, technical, organisational, capacity development and co-operative measures. The data are collected through a questionnaire that is sent out to member countries and desk research is conducted to fill-in data gaps in case of non-response. The 2020 edition covers 194 countries.

Desk research can be either a substitute for a policy survey or a complement, for instance, in the case of non-response (as in the case of the CGI). The United Nations Institute for Disarmament Research's (UNIDIR) *Cyber Policy Portal*, is a repository of information on cyber policies of the 193 UN member states assembled through desk research (United Nations Institute for Disarmament Research, 2021[42]). Most of the data on the *Cyber Policy Portal* is compiled from official and publicly available sources. The Potomac Institute's Cyber Readiness Index 2.0 (PIPS, 2015[43]) is also based on desk research and expert assessments. It considers 70 indicators across seven elements (national strategy, incident response, e-crime and law enforcement; information sharing; investment in research and development; diplomacy and trade, and defence and crisis response), each of which obtains one of three scores (fully operational, partially operational or insufficient evidence).

The *National Cybersecurity Index* (NCSI) of the e-Governance Academy (eGA), a joint initiative of the Government of Estonia, the Open Society Institute and the United Nations Development Programme (eGA, 2022[44]), scores more than 150 countries, aiming to assess their cybersecurity capacity and thus to highlight issues that require government attention and capacity building. It is underpinned by a framework that comprises five steps: identification of national-level cyber threats, identification of cyber security measures and capacities, selection of important and measurable aspects, development of cyber security indicators, and grouping of the indicators. The focus of the NCSI is on measures implemented by central government, legislation in force, established bodies, co-operation formats and outcomes (here: policies, exercises, technologies, programmes, etc.). The NCSI comprises a total of 49 indicators organised into 12 capacities under 3 categories that are informed by a review of publicly available information.

Indicators based on policy surveys and desk research may fail to provide an accurate picture of a country's actual cybersecurity posture

Data collection through policy surveys and desk research faces notable challenges. First, they tend to focus on easy-to-verify, quantitative aspects of the cybersecurity policy landscape (e.g., whether specific legislation or agencies are in place). While important, such information does not capture effective implementation. Depending on analysts' needs, such data might also fall short of capturing information relevant to the quality of these aspects of the cybersecurity policy landscape. The effectiveness of these aspects will hinge on political will and the resources available to enforce them. In other words, data from policy surveys might tell you whether a person has all the right exercise equipment, but they do not necessarily tell you whether a person exercises regularly.

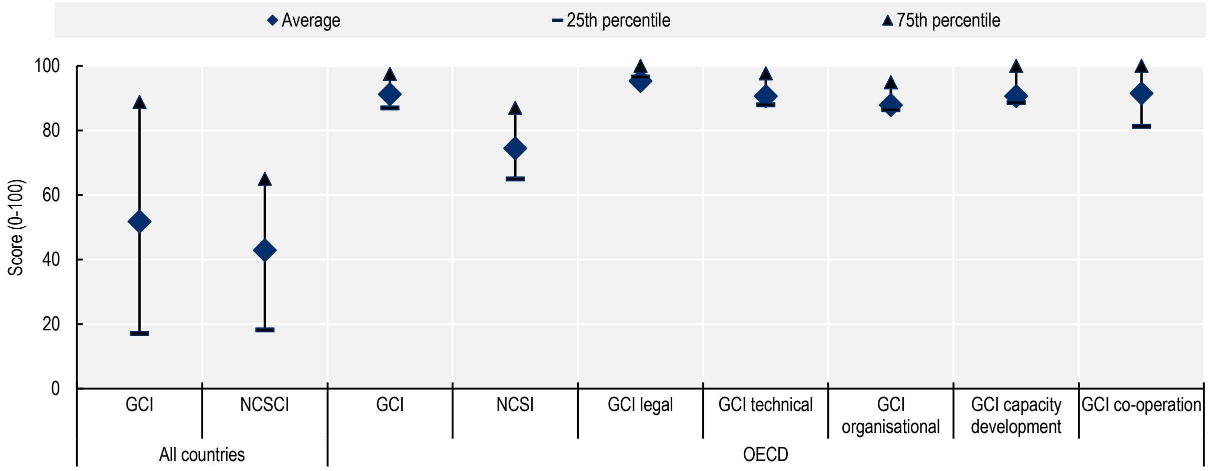
A specific concern in this regard is incentive-compatibility. Designing and implementing effective cybersecurity policies is difficult and requires that substantial resources be put in place. To the extent that an index of a government's preparedness to address cybersecurity threats is widely recognised and publicised, governments can have an incentive to ensure that they appear more ready to address cybersecurity threats while avoiding the actual costs of increased readiness. This is a specific version of the *Goodhart's Law*, which is often stated as "when a measure becomes a target, it ceases to be a good measure" (Goodhart, 1975[45]). In particular, incentives to engage in "isomorphic mimicry" may also come into play, ensuring the appearance of having in place the good-practice policies while avoiding the associated costs (Andrews, Pritchett and Woolcock, 2017[46]).

If governments need to participate actively in a policy survey, a second challenge emerges that could be called the "good-practice conundrum". In this situation, policy surveys require buy-in from governments, but governments can have different views on what is the right set of policies. Broad buy-in requires a consensus on what constitutes good practice – both in terms of policy principles and implementation. But to the extent that such a consensus exists and is supported by evidence, one would expect that all countries are also implementing good practice. As a result, one might find that modelling a policy survey based on the existing consensus of what constitutes good practice results in an index that exhibits limited variation across countries.

For example, in the case of the CGI there is very limited variation across OECD countries (Figure 4.1). The index, which in 2020 had a mean of 51.7 and a median of 50.5, exhibits substantial variation across all 194 countries, with an inter-quartile range from 17.1 to 88.8. However, OECD countries cluster heavily around a very high average score of 91.2. All but five of the 38 OECD countries are in the top quartile. While both the CGI and the NCSI are highly correlated (correlation coefficient of 0.85, p-value<0.0001), though the latter exhibits higher variation across OECD countries, which may be related to the fact that it does not depend on a consensus regarding the contents of a policy survey. A closer look at the sub-components of the CGI shows that the lack of variation is particularly pronounced in legal measures. However, the distributions of all five sub-indices have their mass concentrated close to the ideal score of 100. Just a handful of OECD countries score significantly lower.

Figure 4.1. There is limited variation in the ITU's *Global Cybersecurity Index* across OECD countries

Global Cybersecurity Index and *National Cybersecurity Index, 2020*



Note: Data for the NCSI was accessed on 3 June 2024.
Source: ITU (2021[35]) e-Governance Academy (2022[44]).

Notwithstanding the challenges described above, surveying the policy instruments that governments use and categorising and quantifying that information is a highly valuable activity. To establish an empirical link from policies to behaviours and then impact, one needs comparable data on what governments actually do. And policy surveys and desk research are the only ways in which that information can be marshalled. However, the discussion above should serve as a reminder that a survey instrument that yields relevant data – data that display sufficient variation to capture differences in approaches – needs to be carefully designed and implemented.

Data from private sources

Data related to cybersecurity is often collected by private companies, particularly those who are in the cybersecurity industry (e.g., cyber risk management companies, suppliers of security software, cyber insurance providers, etc.). Such data is undeniably valuable, but also difficult to obtain because private firms often do not have the right incentives to share such data. Moreover, data from private sector firms is understandably not collected with official statistics in mind, and as a result it may suffer from bias. As a result, it crucial to carefully assess the statistical representativeness of private data sources before constructing new cybersecurity indicators.

Private sector firms have a lot of data on cybersecurity, but the underlying data is often not public and aggregated data may be biased

There are a several for-profit organisations that provide intelligence and data on cybersecurity, including software, telecommunications, and cybersecurity companies. Examples, among others, include:

- The *Verizon Data Breach Investigations Report* (Verizon, 2023[47]) is published annually by Verizon, an American multinational telecommunications conglomerate, and highlights trends in data breaches by type, actor, region and industry. The report is based on data on breaches from different sources that are then harmonised. These data, which are not publicly available, come from various contributors (which also change over time) as well as Verizon itself and the authors acknowledge that they constitute a selected sample, stating that "some breaches, such as those publicly disclosed, are more likely to enter our corpus, while others, such as classified breaches, are less likely".

- The 2023-edition of *Cisco Security Outcomes Report* is based on data gathered from 4 700 security professionals across 26 countries as well as focus group interviews (Cisco, 2022[48]). The report states that the survey was conducted by a professional survey research firm in mid-2022 and relied on stratified random sampling. The appendix details the firm-size, industry and geographic distribution of the sample. However, there is no information about the sampling frame (i.e., the complete listing of all firms from which the sample was drawn), and the raw data are not publicly available.
- Gen Digital (formerly Symantec Corporation and Norton LifeLock), a multinational software company which also produces Norton AntiVirus and Avira Internet security products, publishes the *Norton Cyber Safety Insights Report*, which details trends in cybercrime, identify theft and other themes relevant to cybersecurity (Gen, 2023[49]). It is based on an online survey conducted by The Harris Poll, a survey company, with a sample size of about 1 000 adult respondents in eight countries, seven of which are OECD countries (Australia, France, Germany Japan, New Zealand, the United Kingdom and the United States). The survey data are re-weighted based on demographic and geographic characteristics to match the respective target population. However, response rates, which tend to be low in online surveys, are not reported and the raw data are not available publicly.

While such private data sources bring insights, they have several important limitations. First, while these reports tend to focus on highly relevant issues and some private actors have access to relevant data, there is typically a variable degree of transparency when it comes to the methods used (e.g., failure to report non-response rates). Second, the underlying raw data are typically not available for scrutiny. Third, such reports suffer from issues around incentive-compatibility. Many of the private firms that offer analytical reports are also vendors of services or software to guard against cyber incidents that stand to benefit from the appearance of an increase in cyber risk.

Overall, policy makers and analysts should distinguish carefully between data and intelligence provided by private providers and data and intelligence provided by government sources and non-profit organisations. While some of these private providers seemingly have access to large amounts of relevant data that are valuable to policy makers and analysts, in many cases, it is difficult to assess their quality.

Cybersecurity readiness depends on a skilled and well-staffed workforce

Measuring cybersecurity skills adds another layer to the overall measurement of a country's resilience and readiness to address cyber threats. As economic activities continue to digitalise, ensuring an appropriate cybersecurity risk management posture hinges on a skilled workforce. Analysing and tracking cybersecurity skills demand is useful for policy makers to anticipate and avoid skills shortages.

Despite the increasing and recognised importance of preparedness, resilience and risk management practices, firms around the world continue to face a lack of skilled cybersecurity professionals. The global cybersecurity workforce gap has been quantified by the ISC2 organisation, which estimates the difference between the number of cybersecurity professionals that firms require and the number of qualified people available for hire, to be nearly 4 million people across 20 countries[10], a 12.6% increase from 2022 (ISC2, 2023[50]).

Online job postings (OLJPs) are a useful complement to traditional labour market statistics to measure supply and demand of cybersecurity skills. Lightcast[11] collects information on OLJPs from over 65 000 sites worldwide to develop a comprehensive and real-time view of online labour market demand (Lightcast, 2024[51]). Research using Lightcast shows that despite the increasing importance of cybersecurity, firms continue to face a shortage of skilled cybersecurity professionals. Demand for cybersecurity professionals has steadily increased over the last decade (OECD, 2024[52]; OECD, 2023[53]; OECD, 2023[54]), with the volume of job postings in Australia, Canada, New Zealand, the United Kingdom and the United States observed at up to 16 times higher in 2022 than the number of openings recorded in 2012 (OECD, 2023[54]).

Lightcast research from 2022 shows the number of entry-level cybersecurity jobs requiring a bachelor's degree exceeded the number of bachelor's degrees awarded by cybersecurity programmes (OECD, 2023[54]). A similar trend is observed for cybersecurity certifications, where more job postings require certifications than there are certification-holders in the entire country. It should be noted, however, that Lightcast and other OJP data has limitations, as these sources may provide less comprehensive coverage of some sectors and occupations for which positions are not typically advertised or filled via online recruitment campaigns (OECD, 2021[55]).

Box 4.1. Quantifying the cybersecurity certifications gap in the United States

Cybersecurity certifications are numerous, and they each have their own target audience and training focus. The CyberSeek project has identified six prominent cybersecurity certifications completed by cybersecurity professionals and sought by employers in the United States: CompTIA Security+, Certified Information Systems Security Professional (CISSP), Global Information Assurance Certification (GIAC), Certified Information Systems Auditor (CISA), Certified Security Manager (CISM), and Certified Information Privacy Professional (CIPP).

Looking at the difference between certification holders and number of job postings requiring that certification can illustrate part of the gap in cybersecurity skills. With this information, policy makers can analyse the certification demand in the labour market and tailor the available training programmes within their country to match. However, obtaining data to measure this gap is challenging as the certifications are all administered by different companies who do not all consistently report comparable numbers of certifications completed, obtained, renewed or lapsed.

The CyberSeek project is a collaboration between Lightcast, NICE (an institutionalised programme working to advance cybersecurity education and workforce development led by the US National Institute of Standards and Technology (NIST)) and the Computing Technology Industry Association (CompTIA) to aggregate data about the most in-demand cybersecurity certifications in the United States (Table 4.1). In 2023, the CISSP was the most required certification in the US, with 85 566 job openings requesting it, followed by CompTIA Security+, and the CISA.

Table 4.1. Cybersecurity certifications gap

Data for the United States, December 2023

Certification	Certification holders	Demand for certification	Skills gap
CompTIA Security +	265 992	81 048	+184 944
CISSP	91 765	85 566	+6 199
GIAC	46 318	46 810	-492
CISA	35 812	61 020	-25 208
CISM	20 300	42 133	-21 833
CIPP	13 652	6 990	+6 662

Source: CyberSeek (2024[56]).

In addition to aggregating and quantifying the gap in certification holders compared to job openings, CyberSeek also maps job openings according to the NICE cybersecurity workforce framework, which provides a list of cybersecurity tasks and skills required in cybersecurity jobs. While the CyberSeek tool is only available in the United States, it provides an example of a measurement tool that could be replicated in other countries as a targeted effort to close the local cybersecurity skills gap.

Data from insurance companies holds promise, but publicly-available data lacks depth and comparability

Increasingly, private firms have the option to purchase insurance against damages incurred in the event of a cyber incident or attack. According to the Swiss Re Institute, the research arm of reinsurance company Swiss Re, the global cyber insurance market tripled in volume in the last five years, expanding to gross direct premiums of around USD 13 billion in 2022 (Swiss Re, 2023[57]). Premia are forecast to increase to

USD 22.5 billion by 2025. This implies that data collected by insurance companies are increasingly an interesting source of data on cybersecurity.

In fact, there are two different types of indicators that may be based upon data from insurance companies. First, data on premia are indicative of the perceptions of threats and precautions taken by firms in response to these threats. Second, data on claims and damages are more indicative of the actual cost of cyber incidents, although they would not include the costs of preventive measures.

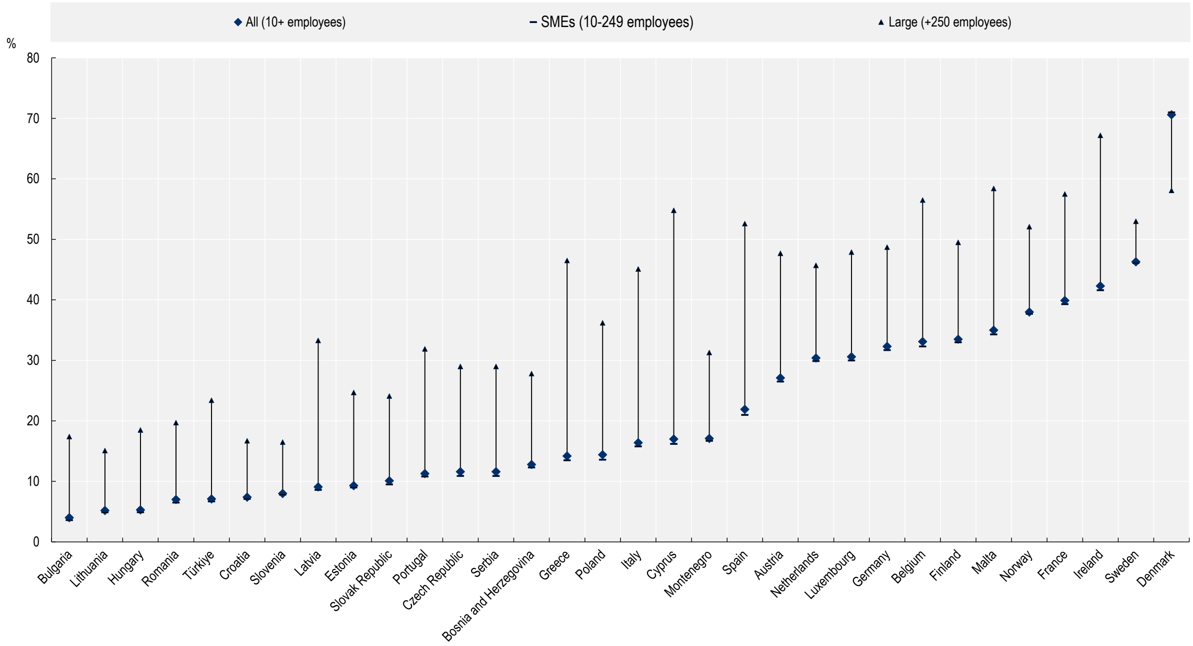
However, there are also several caveats. First, it is not clear to what extent insurance and reinsurance companies would be willing and are legally able to share their data. Sharing of anonymised and aggregated data may be viable, but there will certainly be limits and it is not clear what incentives insurance companies have to do so. Reluctance to share information between insurance companies and the public and private sector may stems from a fear that the anonymisation process might not be robust enough, and therefore data sharing comes with a risk that affected entities might be identified (OECD, 2017[58]).

Second, a better understanding of cyber insurance markets is required to be able to judge the extent to which any data would be provide a comprehensive picture. For instance, different insurance companies are likely to define incidents in different ways, creating challenges in combining data from different providers. Third, the fast growth in cyber insurance premia indicates that these markets have not reached maturity, suggesting that changes observed over time in indicators based on data from insurance companies reflect changes in the market rather than changes in the cyberthreat landscape.

Finally, what applies for changes over time within countries also extends to comparisons across countries, as the level of maturity of cyber insurance markets likely differs significantly across countries. According to data from Eurostat, the share of enterprises with ten employees or more that have had insurance against ICT incidents varies from 4% in Bulgaria to 71% in Denmark (Figure 4.2). While some of this variation might reflect differences in digital intensity of economies, a significant part of it will also be driven by differences in the maturity of insurance markets.

Figure 4.2. Uptake of cyber insurance varies significantly across countries

Share of enterprises that have had insurance against ICT incidents, 2022



Source: Authors' calculations based on data from Eurostat (2022[20]).

In a nutshell, data from insurance companies might in the future become increasingly valuable for statistical reporting. However, it will still take some time before insurance markets are sufficiently developed to allow for broad coverage of incidents and legal obstacles might have to be overcome before the data can be used to produce publicly available indicators.

5 Two innovative approaches to measuring uncertainty in cyberspace

Cybersecurity risk is defined as the "effect of uncertainty on or within information and technology", underscoring that uncertainty is a fundamental aspect of cybersecurity risk itself (NIST, 2020[59]). Uncertainty related to cyber risks, or “cybersecurity uncertainty”, is a multifaceted phenomenon that arises from the inherent unpredictability and ambiguity surrounding cyber threats and vulnerabilities. This uncertainty captures public perceptions of cyber risks, which is a proxy for but different from the actual cyber risk landscape itself, which is not directly measurable. By developing data-driven approaches to quantify uncertainty related to cyber risks, policy makers can complement cybersecurity indicators from official statistics and gain valuable insights into public sentiment, potentially anticipate emerging cybersecurity trends, and develop more targeted cybersecurity awareness programmes to safeguard public interests and promote a more secure and resilient digital ecosystem.

Drawing inspiration from the literature on measuring uncertainty (Baker, Bloom and Davis, 2016[60]) (Castelnuovo and Tran, 2017[61]), this section introduces two innovative approaches to measuring cybersecurity uncertainty. These approaches leverage data from non-official sources – news reports from leading newspapers and online search data – to provide insights into public perceptions of uncertainty related to cyber risks. This uncertainty might lead to adverse effects, such as a hesitancy in business investment and reduced adoption of online services due to concerns about security risks. However, these incidents can also act as catalysts for heightened awareness, driving innovation and stimulating investments in cybersecurity infrastructure and technologies to fortify digital defences.

News reports: towards a cybersecurity uncertainty index

News reports about major incidents reflect the state of cybersecurity and shape public opinion, which in turn might alter online behaviour. They are also a potential source of qualitative information that can be transformed into quantitative data about the state of cybersecurity: the frequency of news around cyber incidents and threats could be used as a proxy for cybersecurity uncertainty, which would be expected to have adverse effects on business investment and the uptake of online services.

Building an index of cybersecurity uncertainty based on news reports could in principle be modelled on the index of economic policy uncertainty developed by Baker, Bloom and Davis (2016[60]). To construct their index for the United States, the authors proceed in three steps. First, they obtain data on the relative frequency of newspaper reports that meet a pre-defined set of criteria from ten leading US newspapers. In their case, they focus on reports that contain the following search terms: “economic” or “economy”; “uncertain” or “uncertainty”; and one or more of the words “Congress”, “deficit”, “Federal Reserve”, “legislation”, “regulation” or “White House”. Second, they aggregate data across different newspapers and by month. Third, they normalise the series by dividing relative frequencies by the standard error of the series and then adjusting it so that it has mean 100.

The index is informative: the version for the United States spikes near tight presidential elections, the onset of military campaigns, major terrorist attacks, critical bank failures and major political battles over fiscal policy. It is also correlated with other measures of economic volatility and uncertainty and foreshadows declines in investment, output and employment. The index is also popular: it is carried by data providers such as Bloomberg and FRED, among others, and the method has been implemented by researchers at the global level in nearly 30 countries and within US states (Baker, Bloom and Davis, 2023[62]). Finally, Baker, Bloom and Davis also show that their index can be refined to capture specific categories of economic uncertainty such as monetary policy, trade, healthcare and national security. And there is little to suggest that their methodology could not be extended to measure policy uncertainty related to cybersecurity risks.

Measuring cybersecurity uncertainty: a proof of concept using news reports

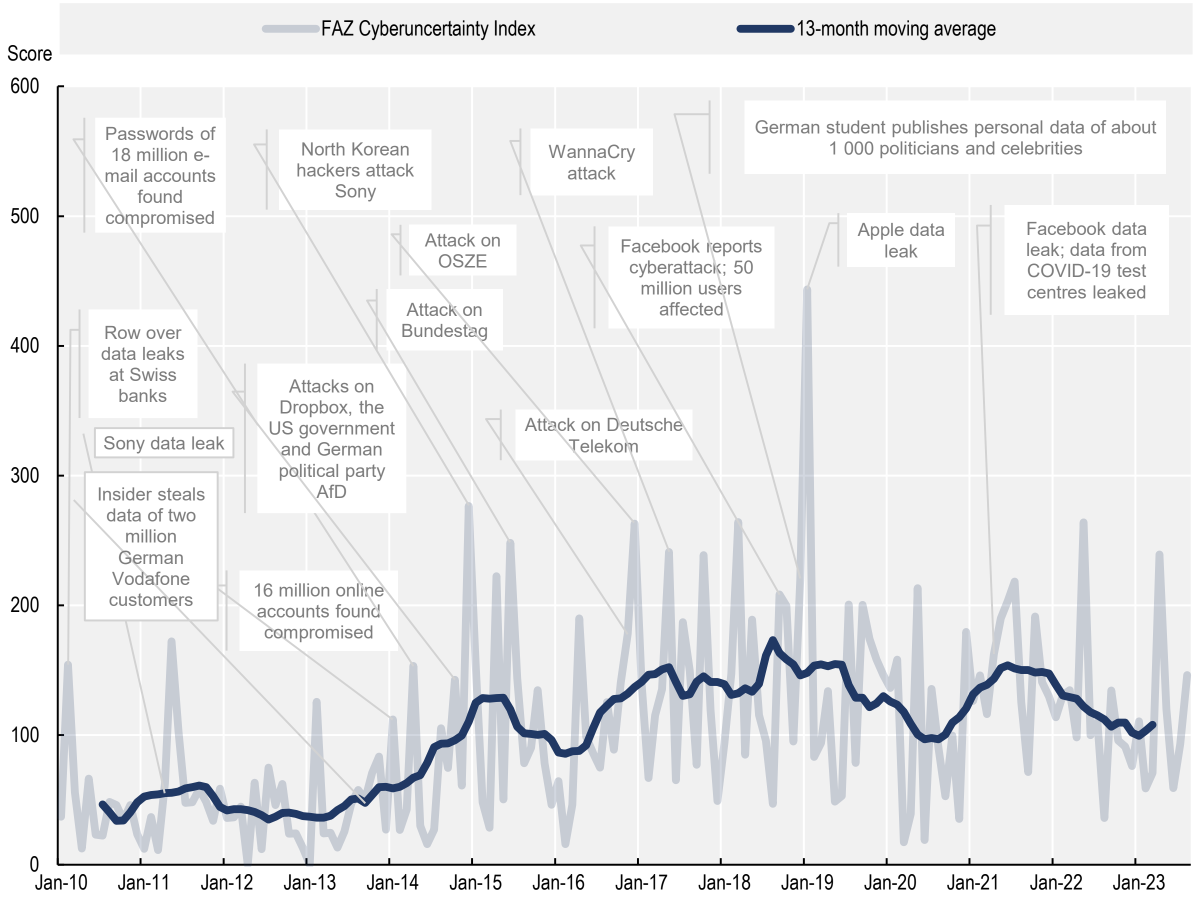
The data shown in Figure 5.1 follow the approach outlined by Baker, Bloom and Davis (2016[60]). It is based on data from the German newspaper the *Frankfurter Allgemeine Zeitung* (FAZ), one of Germany's major daily national newspapers.[12] Articles were selected based on the search criteria reported in row Frankfurter Allgemeine Zeitung Controlled Unclassified Information (CUI) III in Table 5.1 below. In other words, and for reasons explained below, the selection of articles was based on those articles that included one of the two terms "IT security" or "cybersecurity" (which are both used interchangeably in German) *and* at least one term indicating a threat, or any of several terms that by themselves already indicate a cyber incident event.

Table 5.1. Alternative search criteria for *Le Figaro*, the *FAZ* and *The New York Times*

	Search terms	English translation
Le Figaro	"cybersécurité"	"*cybersecurity*"
Frankfurter Allgemeine Zeitung I	"IT-Sicherheit" OR "Cybersicherheit"	"*IT security*" OR "*cybersecurity*"
Frankfurter Allgemeine Zeitung II	"IT-Sicherheit" OR "Cybersicherheit" OR "Datendiebstahl" OR "Datenleck" OR "Hackerangriff" OR "Datenpanne"	"IT security" OR "cybersecurity" OR "data theft" OR "data breach" OR "hacker attack" OR "data breach"
Frankfurter Allgemeine Zeitung III	(("IT-Sicherheit" OR "Cybersicherheit") AND ("Angriff" OR "Vorfall" OR "Panne")) OR ("Datendiebstahl" OR "Datenleck" OR "Hackerangriff" OR "Datenpanne")	(("IT security" OR "cybersecurity") AND ("attack" OR "incident" OR "breach")) OR ("data theft" OR "data breach" OR "hacker attack" OR "data breach")
The New York Times	"cybersecurity"	

Figure 5.1. The FAZ Cyber uncertainty Index

January 2010 to August 2023



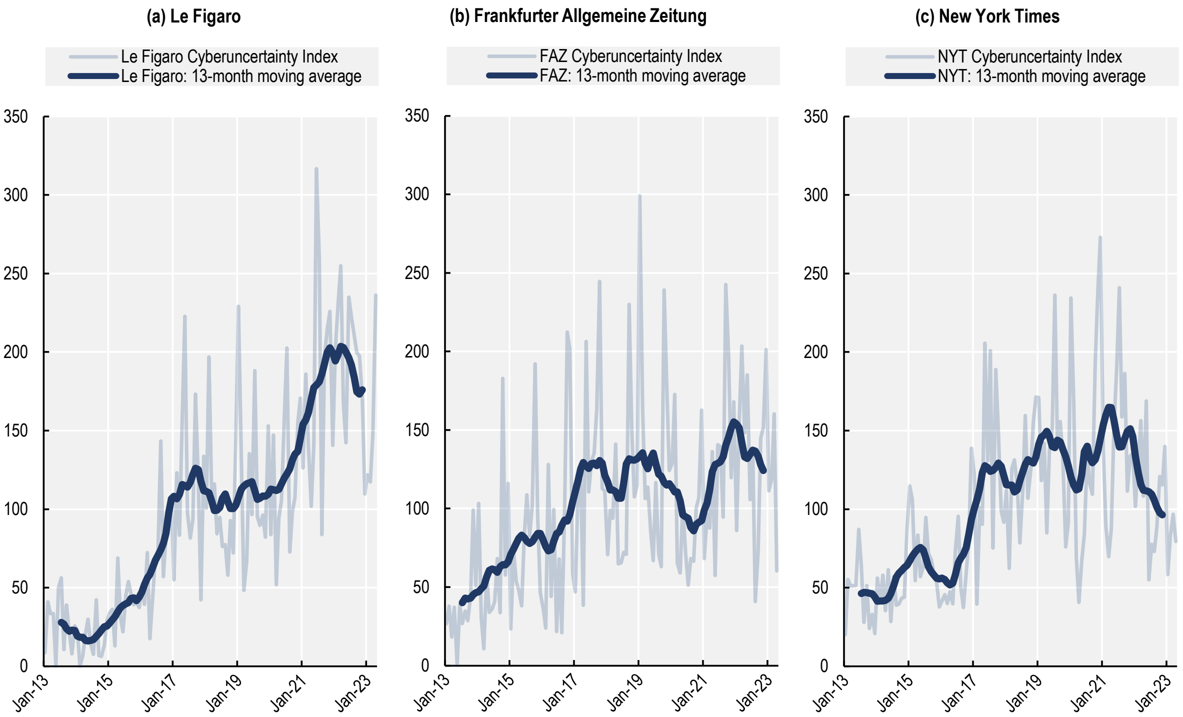
Note: The dark blue line indicates a 13-month moving average centred on the date for which it is reported.
Source: Authors' elaboration based on data from the *Frankfurter Allgemeine Zeitung*.

While the search strategy can be refined, the resulting series is informative in two ways. First, major incidents are reflected as spikes. This includes incidents that had repercussions at the global level, such as the WannaCry attacks in May 2017, but also incidents that primarily affected Germany, for instance the theft of the personal information of two million clients of Vodafone Germany in September 2013 and the cyberattack on the German parliament in June 2015.

Second, the series seems to exhibit a secular trend, estimated here as a 13-month moving average, but which is even easier to spot if the series is log-transformed. As seen in Figure 5.1, the FAZ cyber uncertainty index shows that the relative frequency increases somewhat from January 2010 to 2011 and then more significantly from early 2013 to 2015. It reaches a plateau in 2018, experiences some swings up and down, and then displays a decrease in cyber uncertainty following the Russian Federation's (hereafter "Russia") invasion of Ukraine in February 2022.[13] Similar secular trends are also evident in series constructed from two leading newspapers in France and the United States, *Le Figaro* and *The New York Times*, respectively – see panels (a) and (c) of Figure 5.2 which also reports a series for the FAZ for comparison. All three series, which are normalised to have a mean of 100 over the relevant period, are highly correlated, with correlation coefficients of 0.46 (Le Figaro-FAZ), 0.56 (Le Figaro-NYT) and 0.46 (FAZ-NYT).

Figure 5.2. Cybersecurity uncertainty indices for France, Germany and the United States

January 2013 to April 2023



Source: Authors' elaboration based on data from the *Frankfurter Allgemeine Zeitung*, *Le Figaro* and *The New York Times*.

However, for reasons explained below, the two series for *Le Figaro* and *The New York Times* are based on a simpler search strategy by which articles are selected based only on whether they contain the terms "cybersécurité" and "cybersecurity", respectively. And to support comparisons, the series for the FAZ is also based on the occurrence of one of two search terms, "IT-Sicherheit" OR "Cybersicherheit". Still, even these minimal examples tend to pick up on major cyber incidents. In the case of *The New York Times*, these include the SolarWinds-based breach of the US federal government in December 2020, the WannaCry ransomware attack in May 2017 (which was followed by the NotPetya attack that began to be felt from June 2017), and the aftermath of the attack on Sony in November 2014 by North Korean hackers. It is also interesting to observe that there is a secular trend in the series, with news reports referring to "cybersecurity" increasing in relative frequency from late-2016 and decreasing slightly from mid-2021.

Constructing a relevant indicator from newspaper archives requires subject knowledge and iterative refinement

Why are refinements of this kind required? The method described here to create an index of cyber uncertainty is grounded in the notion that certain news reports convey a sense that using digital technologies is unsafe, and that users are exposing themselves or others to harm. Articles that discuss specific incidents, especially those that affect individuals or organisations residing in the country of interest, should be included as true positives and it is also clear that many articles that do not discuss cybersecurity in any way or those that mention cybersecurity but in a different context should be excluded as true negatives (e.g., wedding announcement that indicate that one of the spouses works in the cybersecurity industry).

However, an initial scoping exercise reveals that there are borderline cases, especially those in which the threat itself is not the main subject of the article but serves as a motivation for it. Any search criterion will produce both false positives and false negatives (Table 5.2).

Table 5.2. Classifying articles based on cyber uncertainty sentiment

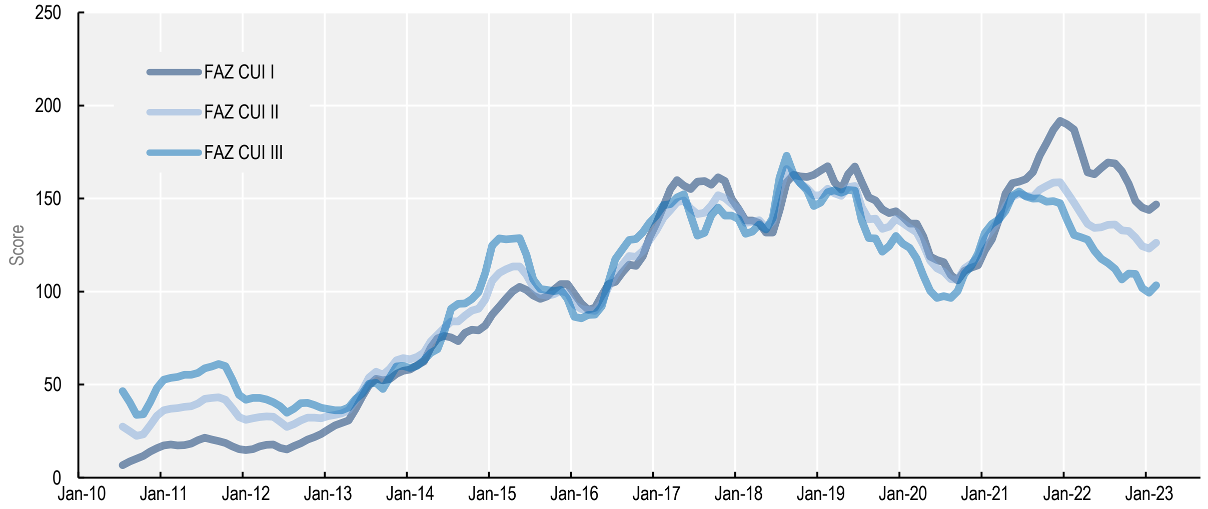
	Actual:	True	False
Predicted as:	True	**True positives** are articles that match the search criteria and convey to readers a sense of threat, risk or uncertainty related to cybersecurity. These include articles that discuss specific incidents but also articles that are motivated by reference to threats.	**False positives** are articles that match the search criteria but do not convey to readers a sense of threat, risk or uncertainty related to cybersecurity. Examples found in initial scoping exercises include business news that discuss the financial situation of cybersecurity firms and nuptial announcements in which one spouse is described as working in cybersecurity. Ideally, one would want to exclude these articles by refining the search criteria.
	False	**False negatives** are articles that do not match the search criteria but convey to readers a sense of threat, risk or uncertainty related to cybersecurity. Ideally, one would want to include these articles by refining the search criteria.	**True negatives** are articles that do not match the search criteria and do not convey to readers a sense of threat, risk or uncertainty related to cybersecurity. These articles were rightly excluded.

This implies that search criteria will have to be refined to minimise the share of false positives and false negatives. As it will be impractical to audit each article, this should be done based on the iterative auditing of a random sample of articles – that is, a random sample drawn from all articles that match the search criteria should be carefully read and classified as either being relevant to cyber uncertainty or not. This will allow the analyst to further refine the search criterion, a process that can then be iterated until the share of false positives and false negatives is deemed sufficiently low.

The scope for refining these series is illustrated in Figure 5.3, which reports 13-month moving averages from three different series constructed from the FAZ data (see Table 5.1). The three series are based on different combinations of search terms that reflect an increasing degree of sophistication.

Figure 5.3. Comparison of different cyber uncertainty indices based on data from the FAZ

13-month moving averages, January 2010 to August 2023



Notes: See Table 5.1 for the search terms used in the construction of these series.
Source: Authors' elaboration based on data from the *Frankfurter Allgemeine Zeitung*.

The resulting series are highly correlated – the correlation coefficient between CUI I and CUI II (CUI III) is 0.89 (0.59) – and visually appear to track each other closely over most of the period considered in this report. This suggests that simply considering the share of articles that mention the term "cybersecurity" in the respective language might already be moderately informative about trends in the public interest. There are, however, some notable deviations, with FAZ CUI III indicating higher uncertainty towards the beginning of the period and less uncertainty towards the end. In addition, FAZ CUI III seems to show a peak at the beginning of 2015, which is not as pronounced in the other two series.

Overall, the refinements implemented for the FAZ data seem to pay off. An audit of a random sample of 100 articles that were selected as relevant to cybersecurity threats suggests that an average of 65% (95%-confidence interval: 56%-74%). The same exercise for the (less sophisticated) index based on articles selected from *Le Figaro* resulted in a share of true positives of only around 42% (32%-52%).

Moving forward on indices of cybersecurity uncertainty using news reports

As in the case of the index of economic policy uncertainty, any full-fledged version of a cyber uncertainty index should ideally be based on data from several newspapers and should carefully combine several search terms through logical connectives.[14] Constructing a full-fledged index in this way has several advantages. Most importantly, the methodologies are well established, and data would be available at a high frequency (i.e., monthly). As in the case of the index of economic policy uncertainty, the index could be constructed for different jurisdictions and even at the global level. It is also conceivable (although by no means guaranteed) that one could demonstrate that a well-calibrated cyber uncertainty index is statistically correlated with relevant macroeconomic variables such as investment in ICT equipment.

That said, there are several caveats and challenges. First, an index based on newspaper reports would only capture what newsrooms see as newsworthy – that is, only reported incidents and their aftermaths. Such an index could be interpreted as based on an expert assessment, where the experts are journalists and their editors. Second, while the series would lend themselves to the analysis of trends over time, they are less likely to be of much value in benchmarking exercises based on cross-country data. Data series for different countries would have to be based on different newspapers, which in turn have different reporting priorities and editorial styles. As in the case of the index of economic policy uncertainty, country-specific series would have to be constructed and reviewed by experts that are familiar with the relevant vocabulary in each country's language(s).

Finally, there are practical considerations. First, the data underlying the three country examples above were constructed by querying the online archives of the respective newspapers. But neither the online archive of *Le Figaro* nor *The New York Times* allow for searches that combine search terms through logical operators (e.g., 'AND' and 'OR') and search engines of other major newspapers do not allow "empty queries" (i.e., queries that would return the total number of all published articles in a given period). However, the authors are aware that certain service providers (e.g., Factiva) allow for more systematic searches that could be used to this end. Second, constructing cyber uncertainty indices for all OECD countries would require resources in terms of staff time (for instance, for auditing of articles) as well as knowledge of the language and culture of the member country in question.

Overall, the construction of indices of uncertainty based on reports in major newspapers seems to be a promising endeavour. First, it has great potential to generate relevant data at high frequencies for many countries. Second, it is straight-forward to check results carefully and ensure that the analytical steps taken can be transparently documented. Finally, by constructing indices from newspapers from different regions and carefully weighting the result, one could even imagine the construction of a global index of cybersecurity uncertainty. This may be relevant given the increasing scale and scope of cyberthreats.

Online searches: another approach to constructing an index of cybersecurity uncertainty

Online search data is another source of information to measure uncertainty. Research in economic psychology, for instance, suggests that people are more likely to search for information about the economy when they are uncertain about it (Dzielinski, 2012[63]). Moreover, the volume of Internet searches for uncertainty-related terms can potentially capture the level of perceived uncertainty in the economy. For instance, in the event of a significant cyberattack, one might expect an uptick in searches for terms like "cyberattack", "data breach", or "antivirus software". The signalling attributes inherent in online searches make them an especially valuable instrument for measuring uncertainty, given that they arise from the spontaneous behaviours of users (see Dzielinski (2012[63]) and references therein).

At the same time, increased uncertainty often foreshadows changes in other relevant economic variables, implying that timeliness and frequency are important considerations when developing indicators of uncertainty. And while other economic uncertainty indices often rely on data with lags, online search data can in theory be used to measure of the prevalence of search queries linked to uncertainty on a weekly or even daily basis (Dzielinski, 2012[63]). Google is the most widely used search engine in the countries analysed in this report and Google Trends data are widely available in OECD countries, making online searches using Google a good proxy for how concerned people are about specific topics (Castelnuovo and Tran, 2017[61]).

In recent years, research using Google Trends data to measure economic uncertainty has expanded. This research takes the form of Google Trends Uncertainty (GTU) indices, which rely on search volumes for keywords related to uncertainty. GTU indices are constructed through an aggregation process based on relative frequencies, where the importance of each keyword and its search frequency determine its contribution to the index. This approach results in the creation of GTU indices tailored to the specific context of the country or region of interest. For instance, such indices that measure economic policy or financial uncertainty have been developed for the United States and Australia (Castelnuovo and Tran, 2017[61]), The Republic of Türkiye (hereafter "Türkiye") (Bilgin, 2019[64]) and India (Pratap and Priyaranjan, 2023[65]).

Similarly, Dzielinski (2012[63]) focusses on measuring general economic uncertainty using Google Trends data to investigate its impact on the stock market. In contrast, Weinberg (2020[66]) constructs the GTU for the European Union using a weighted average approach.

Overall, the literature shows that GTU indices consistently exhibit a robust correlation with alternative measures of uncertainty specific to the countries under examination.[15] Notably, the Google Trends-based Economic Policy Uncertainty Index (Castelnuovo and Tran, 2017[61]) and the VIX index (Bilgin, 2019[64]) demonstrate a positive association with GTU indices, supporting the premise that they offer a reliable measure of uncertainty. Bilgin (2019[64]) finds that higher levels of uncertainty, as measured by GTU indices, tend to have a negative impact on economic growth and employment while Dzielinski (2012[63]) finds that stock market returns tend to fall during periods of high uncertainty.

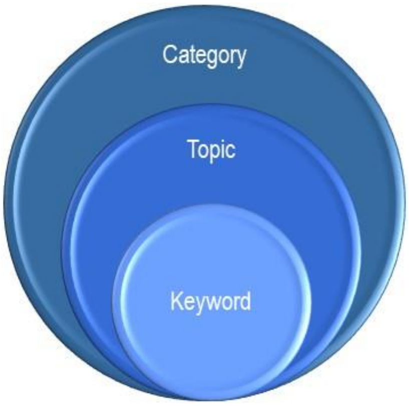
Unpacking Google Trends data

Google Trends provides a time-series index of the relative popularity of search terms, called the Search Volume Index (SVI). It is important to note that the SVI does not represent the actual number of times a search term is entered or the share of searches for a particular search term in total searches, but rather its popularity relative to the highest search volume observed in a given series and to other search terms in a particular region (Combes and Bortoli, 2016[67]). This implies that SVIs from different regions cannot be directly compared, as they are relative to the total number of searches in each region. Additionally, Google Trends data are sample based, so the results may vary slightly across otherwise identical queries. Finally, for privacy reasons, only queries with a significant volume are tracked, which can lead to breaks in series (Choi and Varian, 2012[68]).

Google Trends offers three ways to search and explore trends: keywords, topics, and categories. Keywords are the most specific, categories are the most general, and topics fall in between. Keywords provide a precise, language-specific way to investigate specific search queries. Google creates topics by grouping together related search queries based on their purpose and meaning, and by considering where users click. As a result, there is no fixed list of topics, and choosing topics requires exploring the Google Trends website (Woloszko, 2020[69]). Topics are more suited than keywords for comparing the relative popularity of a concept across countries. Unlike keywords, they are language-agnostic, allowing users to compare the same concept in each country.

While the algorithm behind the translation and expansion in search terms around the query is proprietary, Google provides the following example: if one searches for the word *London*, and then chooses the corresponding topic, the search includes results for topics like *Capital of the UK* and *Londres*, which is *London* in Spanish and French. In contrast, categories encompass topics and provide the most general perspective. The relationship between keywords, topics and categories is illustrated in Figure 5.4. Groups of keywords form topics, which in turn, combine into categories. This structure enables users to navigate from a granular search query to a broader, thematic query. However, Google Trends categories and topics lack transparency as the exact content is somewhat unclear (Woloszko, 2020[69]).

Figure 5.4. The relationship between Google Trends keywords, topics and categories



Using Google Trends data to construct a Google Trends Cyber Uncertainty (GTCU) Index

This report introduces an index measuring uncertainty related to cyber risks based on the methodology developed by Castelnuovo and Tran (2017[61]). The first step in developing the index is to identify possible search terms related to the topic of interest. In their paper, Castelnuovo and Tran identified 79 keywords for the United States and 78 keywords for Australia related to "uncertainty" by referencing official documents such as the Federal Reserve's *Beige Book*[16] for the United States and the Reserve Bank's *Monetary Policy Statement* for Australia. These keywords were subjectively selected based on their connection to uncertainty in the context of these documents. For example, if the *Beige Book* mentions "consumer uncertainty about financial markets", keywords like "bank deposit", "consumer confidence", "consumer uncertainty", and "financial markets" were chosen.

Nine topics related to cyber risks were used to build the GTCU index (Table 5.3). Topics are preferred over keywords to ensure the cross-country comparability of the fetched concepts. This was done by systematically checking that each keyword corresponds to an existing topic on the Google Trends Online Interface before fetching the data through the Google Trends Application Programming Interface using the topic IDs. Finally, the fetched data was further limited to search terms belonging to the “Computer Security” category to ensure that the data collected is relevant to the topic of interest.[17]

Table 5.3. Matching nine topic labels with corresponding IDs in Google Trends

Google Trends topic label	Google Trends topic ID
Cyberattack	/m/0p78w_d
Data breach	/m/03c18t5
Spyware	/m/075fh
Malware	/m/0582c
Ransomware	/m/0657nv
Computer virus	/m/01qgr
Phishing	/m/027b9k
Identity theft	/m/018npy
Cybercrime	/m/01y4q_

Note: Topic IDs can be directly identified through each topic URL in the Google Trends’ Online Interface: https://trends.google.com/trends.
Source: Author’s elaboration based on Google Trends.

The second step in constructing the GTCU index is to aggregate the Google Trends data. Given that Google only allows users to compare up to five topics at a time, to measure uncertainty for many topics it is necessary to combine data from multiple Google Trends searches. This can be done by selecting a benchmark term that is directly related to cyber incidents to capture the nuances of changing public perceptions in the dynamic cybersecurity landscape. This benchmark topic is then used to normalise the search volume of the other topics. This is done by randomly selecting four new topics from the pool and searching for them together with the benchmark term. This process is repeated until all the relevant topics have been searched.

The term "Cyberattack" was chosen as the benchmark term for the GTCU index. Following Castelnuovo and Tran (2017[61]), the data are normalised so that each country time-series has a mean of 100 and a standard deviation of 30. This process[18] makes it possible to compare the frequencies of different topics across periods and countries and to focus on the relative changes in their frequencies, which is a better indicator of the uncertainty related to cyber risks. Normalisation also de-links the frequency of each topic from the set of topics it was searched with to account for the fact that the maximum SVI value is set at 100 by Google (Castelnuovo and Tran, 2017[61]). More importantly, it mitigates the potential impact of country-specific search trends on the GTCU index. The final GTCU index is constructed by summing up the relative frequencies for a particular time period and country.

GTCU index series are then calculated monthly for France, Japan and the United States for the period January 2010 to October 2023. The GTCU index takes a value between 70 and 383 across all country time series. Within the minimum and maximum bounds, the GTCU index values reflect different degrees of search interest in or popularity of search topics related to cyberattacks, and as such corresponding levels of uncertainty associated with cyber incidents. The lowest GTCU index value indicates minimal to non-existent search interest in terms related to cyberattacks in a country, which suggests that there is not much uncertainty related to cyber risks. In contrast, the maximum value indicates the peak search interest within the dataset, which suggests that there is relatively high uncertainty associated with cyber incidents at this point in time.

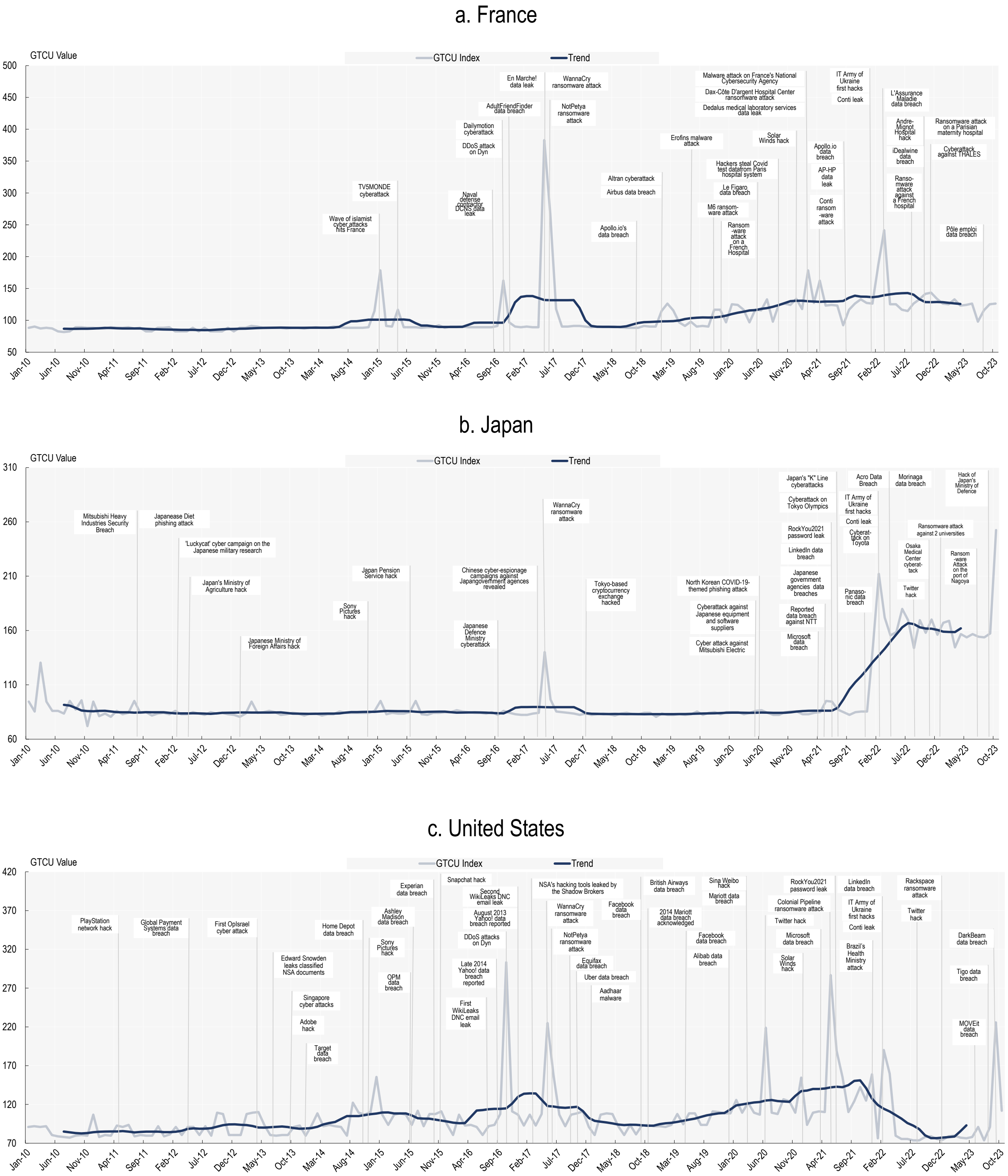
The GTCU index can be used to track changes in uncertainty over time, compare trends in uncertainty related to cyber risks across countries, and identify the factors contributing to perceptions of uncertainty related to cyber risks. For example, a sudden surge in the GTCU index for a country following a major cybersecurity incident could indicate a heightened level of uncertainty in response to the incident. Similarly, if the GTCU index for a country remains consistently high over an extended period, it may suggest the presence of persistent and deeply rooted cybersecurity threats within that country. Variations in the GTCU index between countries can be attributed to a range of factors, including differences in awareness about cyber risks, investment or exposure to cybercrime.

Three use cases of the GTCU Index: France, Japan and the United States

This section presents three illustrative examples of the GTCU index for France, Japan and the United States, spanning from January 2010 to October 2023. These examples provide a tangible demonstration of the application of the GTCU index and how they capture fluctuations in uncertainty related to cyber incidents over time.

Figure 5.5 displays the GTCU indices for the three countries, with each plot featuring two lines. The first line in each plot depicts the GTCU index itself, a proxy for the level of uncertainty related to cyber risks in each country. The second line represents the trend component derived from seasonal decomposition, effectively smoothing out short-term fluctuations and offering more insights into the underlying trends. While this section focuses on France, Japan and the United States as a proof of concept, it is important to note that the approach can be extended to the remaining 35 OECD countries. The comprehensive dataset and methodology allow for the calculation and visualisation of GTCU indices for each of these countries, offering a global perspective on perceptions of uncertainty in cyberspace.

Figure 5.5. GTCU indices for France, Japan and United States (2010-23)



Note: The dark blue line illustrates the trend component obtained through seasonal decomposition, capturing the underlying long-term behaviour or trend in the dataset.
Source: Authors' calculations using Google Trends data.

To assess the usefulness of the GTCU index as a measure of uncertainty related to cyber incidents, it is important to validate the results. To do so, it is informative to situate the GTCU index within the context of major cyberattacks. Figure 5.5 indicates that several GTCU peaks match major cybersecurity events that occurred since 2010, such as the WannaCry ransomware attack (all indices, May 2017), the Twitter hack (July 2020, US index), the French health insurance data leak (the “L'Assurance Maladie” hack) (French index, March 2022), and the cyberattack by on Toyota (Japanese index, March 2022). This suggests that the GTCU index is a valid proxy for uncertainty related to cyber risks.

The GTCU index results suggest that cyber incidents aimed at governments generate greater levels of uncertainty than those aimed at particular firms. For example, the “L'Assurance Maladie” hack in France in 2022 corresponds to a relatively high GTCU score in France (241), surpassing the impact of the ransomware attack against the French TV channel “M6” (116). The NotPetya ransomware attack, a seminal moment in state-sponsored cyber warfare and one of the most financially devastating cyberattacks to date, corresponds to a relatively high GTCU index (158) in the United States compared to the data breach at the firm Equifax (91). Similarly, the cyberattack on the Japanese Ministry of defence in August 2023 corresponds to a relatively high GTCU index for Japan (252) compared to the cyberattack on Toyota in March 2022.

In addition, given that cybersecurity incidents spill easily over borders, a cybersecurity incident in one country can relatively easily and quickly impact uncertainty about cyber risks in another country. For example, the “IT Army of Ukraine” first hack in February 2022 is reflected in a surge in the GTCU index in all three countries. This suggests that cybersecurity uncertainty is a global phenomenon, and that cybersecurity incidents in one country can have a ripple effect in other countries.

Figure 5.5 also reveals that public perceptions of uncertainty surrounding cyber incidents vary across countries. For instance, the GTCU index for France shows a gradual but consistent increase over the past five years, signifying that people in France are becoming increasingly concerned about cyber incidents. Conversely, Japan's GTCU index shows a more pronounced upward trajectory since 2021, with a noteworthy surge in the last two years. This surge underscores a heightened level of uncertainty related to cyber risks in the country. Meanwhile, the United States' GTCU index is characterised by more fluctuations, with intermittent peaks indicating varying degrees of concern about cyber risks.

Despite these differences, all three countries have experienced significant spikes in the GTCU index at different times, pointing to periods of heightened public uncertainty regarding cyber risks. This suggests that public perceptions of uncertainty about cyber incidents are increasing in all three countries, likely due to a combination of factors such as high-profile cyberattacks, increasing awareness of cybersecurity risks, and the growing complexity and sophistication of cyber threats.

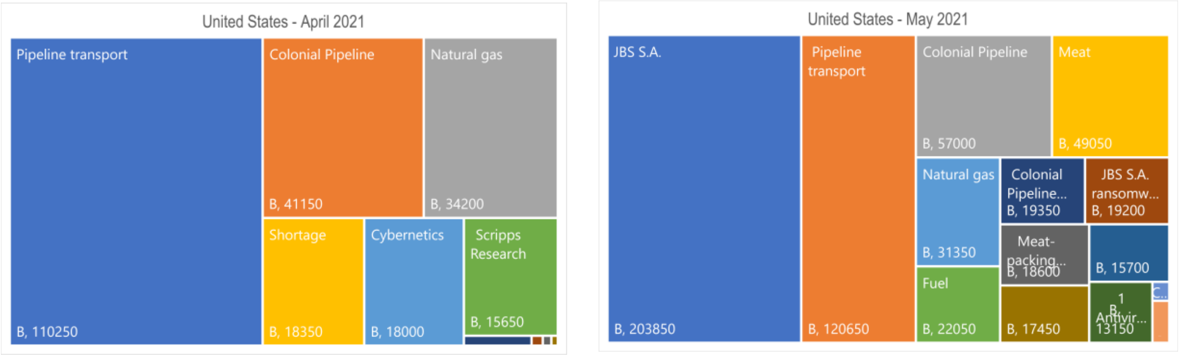
Top and rising related topics provide further insights on cybersecurity uncertainty

In addition to the GTCU’s ability to assess overall uncertainty related to cyber risks, top and rising related topics, which are frequently searched terms associated with a specific query, can shed additional light on cybersecurity uncertainty across countries. Analysing these topics could be valuable when observing persistently high levels of uncertainty in the GTCU index. For example, if a topic like "Phishing" remains among the top related topics for an extended period, it could signal heightened public concern about phishing specifically, prompting policy makers to take proactive measures. An illustrative example of this can be found in Annex B.

Rising related topics are search terms associated with the terms used to construct the index that experienced a significant increase in search volume, typically over the previous period. This feature detects unusual search activity associated with GTCU topics, providing real-time insight into emerging threats and growing concerns. Analysing rising related topics associated with the GTCU index can help predict or detect unusual search activity related to cyber incidents, as evidenced by a notable spike observed in the US index in May 2021 (see Figure 5.5, panel c).

During this period, the US index experienced a significant surge, reaching a value of 286, coinciding with the Colonial Pipeline ransomware attack that disrupted the fuel supply chain on the East Coast of the United States. Prior to the attack, several related topics saw substantial increases in search volume, such as “Pipeline Transport” and “Colonial Pipeline” in April 2021 (Figure 5.6). This suggests that there was highly unusual search activity related to the pipeline before the attack became public. This unusual activity involved searches for terms related to the pipeline that were associated with those included in the index (see Table 5.3).

Figure 5.6. Rising related topics to the GTCU index, 2021



Note: The values in each box indicate the search volume for each rising related topic. The letter “B” denotes “breakout”, defined as a search term that has witnessed a surge in search volume exceeding 5 000% compared to the previous month.
Source: Authors’ elaboration based on Google Trends data.

This surge in rising related topics also indicates a “breakout” (marked with a "B" in each of Figure 5.6’s boxes), signifying a substantial rise in interest as measured by Google Trends, with search volume increasing by thousands of percentage points compared to the previous month. For example, the rising related topic to the index, "Pipeline transport" broke out during both months, registering a value of 120 650 in May 2021. Given that this rising topic had no value in April 2021, the search value 110 250 indicates an increase in associated search volume of over 5 000%.[19]

Monitoring rising related topics within the GTCU index can complement early warning systems for cyber threats and vulnerabilities. By providing cybersecurity professionals with early signs of suspicious activity or heightened interest, the GTCU index can potentially speed up detection of cyber threats before they escalate into major attacks as a way to complement existing cybersecurity protocols.

Moving forward on indices of uncertainty in cyberspace using Google Trends

The GTCU index introduced in this report offers a valuable perspective on cybersecurity uncertainty. It provides a standardised assessment of global public sentiment regarding cyber risks and related topics, mitigating potential biases arising from country-specific variations. The index employs an aggregation procedure that minimises these effects, establishing a more objective and cross-country comparable measure of cybersecurity uncertainty.

Analysing index trends and exploring functionalities like rising related topics can help policy makers design targeted cybersecurity policies. For example, the GTCU index can inform the development of effective public awareness campaigns that align with genuine public perceptions of cyber risks and address potential emerging threats proactively.

However, it is essential to acknowledge the limitations of the GTCU index. Its reliance on Google Trends data can raise concerns due to the lack of transparency in the data construction methodology. This lack of clarity makes it difficult to understand how the methodology might have evolved over time, potentially introducing sampling noise. As suggested by previous research (Woloszko, 2020[69]), computing the mean GTCU index across different draws could help address this issue. The fact that the GTCU tends to map well with cybersecurity incidents also provides comfort in this data-driven approach.

Finally, validating the index by comparing it to other measures of uncertainty and investigating its predictive capabilities for relevant macroeconomic factors, such as cybersecurity-related investments or job postings, would be worthwhile. This would strengthen the GTCU's validity and provide valuable insights into its potential applications in a wider range of countries.

6 Conclusion

Cyber risks are shrouded in uncertainty, and indicators that measure the cybersecurity posture of countries and the effectiveness of the policies that they put in place to mitigate them are sparse despite the diverse sources of official and non-official data that exist about cybersecurity. While quantification efforts are almost always undertaken with caveats about their accuracy, cybersecurity data seem particularly prone to biases, and actors in the cybersecurity ecosystem do not often have an incentive to share accurate data as malicious actors might then exploit it.

Nonetheless, this report has provided some new perspectives on measuring cybersecurity. It outlines a checklist of measurement considerations inherent in cybersecurity measurement and reviews official and non-official data sources that are (or could be) used in creating cybersecurity indicators. It has also outlined two innovative approaches to constructing indices of uncertainty in cyberspace using reports from major newspapers and online search data from Google Trends. A summary of the report’s findings regarding cybersecurity data sources can be found in Table 6.1.

Table 6.1. Cybersecurity data sources: findings at-a-glance

	Official sources		**Non-official sources**		
	Sample surveys	Administrative data	Policy surveys and desk research	Newspapers	Online search data (Google Trends)
Inputs	Yes	Yes	Variable	No	No
Outputs	Yes	Yes	Yes	No	No
Outcomes	Yes	Yes	Yes	Yes	Yes
Country coverage	Medium	Low	High	High	High
Resource intensity	High	Medium	Medium/High	High	Medium/Low
Comparability	High/Medium	Medium/Low	High	No single-date cross-country comparisons, but comparisons of trends	High
Timeliness	Medium	Medium	Medium	High	High
Frequency	Low	High	Low	High	High
Representativeness	High	Variable	Variable	N/A	N/A
Transparency	High	High	Variable	High	Low
Interpretability	Variable	Medium	High	High	Medium

The report also highlights several key points, including:

- The need for standardisation in the various cybersecurity measurement frameworks, including in relation to surveys administered by national statistical offices, assessments of CSIRTs/CERTs, and for high-priority administrative data, to compare confidently across countries.
- Sample surveys are an important source of data on cybersecurity, but indicators derived from them on cyber incidents should be interpreted carefully and there are reasons to believe that they are underreported. In contrast, data on outputs in such surveys are likely to be accurate and their use

should be encouraged, not least to enable research that further unpacks the links between and among cybersecurity indicators.

- An index of uncertainty in cyberspace would capture perceptions of uncertainty related to cyber risks, which could then be used to predict relevant outcomes in the future (e.g., cybersecurity investments). Such a measure would be a timely and useful measure of uncertainty of cyber risks that would help policy makers better understand cyber risk dynamics and develop effective policies in response.

Measuring the various aspects of cybersecurity across countries is challenging, in part because the actors in the cybersecurity ecosystem often do not have the incentives to share key data. At the same time, people, firms and governments need to feel secure to communicate online and use Internet-based services. It is clear that official and non-official sources will need to be brought to bear to provide a comprehensive understanding of cybersecurity posture across countries. Areas such as CERT/CSIRT maturity, professional certifications, and specific firm-level cybersecurity competencies would benefit from common measurement approaches across a wide range of countries to assess country readiness in the face of cyber risks. Measuring cybersecurity uncertainty can complement existing statistics and help anticipate emerging cybersecurity trends, develop more targeted cybersecurity awareness programmes, and promote a more secure and resilient digital ecosystem.

References

Allianz (2022), *Allianz Risk Barometer 2022*, https://www.allianz.com/content/dam/onemarketing/azcom/Allianz_com/press/document/Allianz_Risk_Barometer_2022_FINAL.pdf (accessed on 10 May 2024). [2]

Anderson, R. et al. (2013), “Measuring the Cost of Cybercrime”, in *The Economics of Information Security and Privacy*, Springer Berlin Heidelberg, Berlin, Heidelberg, https://doi.org/10.1007/978-3-642-39498-0_12. [12]

Andrews, M., L. Pritchett and M. Woolcock (2017), “Looking Like a State”, in *Building State Capability*, Oxford University Press, Oxford, https://doi.org/10.1093/acprof:oso/9780198747482.003.0003. [46]

Aragão, R. and L. Linsi (2020), “Many shades of wrong: what governments do when they manipulate statistics”, *Review of International Political Economy*, Vol. 29/1, https://doi.org/10.1080/09692290.2020.1769704. [9]

Baker, S., N. Bloom and S. Davis (2023), *Economic Policy Uncertainty Index*, http://www.policyuncertainty.com/ (accessed on 10 May 2024). [62]

Baker, S., N. Bloom and S. Davis (2016), “Measuring Economic Policy Uncertainty”, *The Quarterly Journal of Economics*, Vol. 131/4, pp. 1593-1636, https://doi.org/10.1093/qje/qjw024. [60]

Bilgin, M. (2019), “A novel index of macroeconomic uncertainty for Turkey based on Google-Trends”, *Economics Letters*, Vol. 184, 108601, https://doi.org/10.1016/j.econlet.2019.108601. [64]

Castelnuovo, E. and T. Tran (2017), “Google It up! A Google Trends-based uncertainty index for the United States and Australia”, *Economics Letters*, Vol. 161, pp. 149-153, https://doi.org/10.1016/j.econlet.2017.09.032. [61]

Choi, H. and H. Varian (2012), “Predicting the present with Google Trends”, *Economic Record*, Vol. 88, pp. 2-9, https://doi.org/10.1111/j.1475-4932.2012.00809.x. [68]

CISA (2022), *CISA Cybersecurity Awareness Program*, Cybersecurity & Infrastructure Security Agency, https://www.cisa.gov/cisa-cybersecurity-awareness-program (accessed on 10 May 2024). [37]

Cisco (2022), *Security Outcomes Report, Volume 3*, https://www.cisco.com/c/en/us/products/security/cybersecurity-reports.html#~newest-reports (accessed on 10 May 2024). [48]

Combes, S. and C. Bortoli (2016), “Nowcasting with Google Trends, the more is not always the better”, *CARMA 2016: 1st International Conference on Advanced Research Methods in Analytics*, Vol. Editorial Universitat Politècnica de València, pp. 15-22, https://archive.carmaconf.org/carma2016/wp-content/uploads/pdfs/4226.pdf. [67]

CyberSeek (2024), *Cybersecurity supply/demand heat map*, https://www.cyberseek.org/ (accessed on 10 May 2024). [56]

Detica (2011), *The Cost of Cybercrime*, Detica, http://www.cabinetoffice.gov.uk/resource-library/cost-of-cyber-crime (accessed on 10 May 2024). [11]

Dourado, E. (2012), *Is There a Market Failure in Cybersecurity?*, Mercatus Center at George Mason University, https://www.mercatus.org/research/policy-briefs/there-market-failure-cybersecurity (accessed on 10 May 2024). [72]

DSIT (2023), *Cyber Security Breaches Survey 2023*, United Kingdom Department for Science, Innovation and Technology, https://www.gov.uk/government/statistics/cyber-security-breaches-survey-2023/cyber-security-breaches-survey-2023. [17]

Dzielinski, M. (2012), “Measuring economic uncertainty and its impact on the stock market”, *Finance Research Letters*, Vol. 9/3, pp. 167-175, https://doi.org/10.1016/j.frl.2011.10.003. [63]

ECOSOC (2013), *Fundamental Principles of Official Statistics*, United Nations Economic and Social Council, New York, https://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx (accessed on 3 June 2024). [14]

eGA (2022), *National Cyber Security Index*, e-Governance Academy, https://ncsi.ega.ee/ (accessed on 10 May 2024). [44]

ENISA (2023), *SIM3v1 self-assessment tool*, European Union Agency for Cybersecurity, https://www.enisa.europa.eu/topics/incident-response/csirt-capabilities/csirt-maturity/csirt-survey (accessed on 10 May 2024). [29]

Europol (2023), *Internet Organised Crime Threat Assessment 2023*, https://www.europol.europa.eu/publication-events/main-reports/internet-organised-crime-threat-assessment-iocta-2023 (accessed on 3 June 2024). [26]

Eurostat (2022), *Community Survey on ICT Usage and E-commerce in Enterprises 2022 (model survey)*, https://circabc.europa.eu/ui/group/89577311-0f9b-4fc0-b8c2-2aaa7d3ccb91/library/48d8c46d-3c8a-4ba8-9d14-405403ad1679/details (accessed on 19 May 2024). [15]

Eurostat (2022), *ICT Usage in Enterprises (database)*, https://ec.europa.eu/eurostat/web/digital-economy-and-society/database (accessed on 10 May 2024). [20]

FBI (2022), *Internet Crime Complaint Center (IC3)*, https://www.ic3.gov/ (accessed on 10 May 2024). [24]

FBI (2021), *Internet Crime Report 2021*, Federal Bureau of Investigation, https://www.ic3.gov/Media/PDF/AnnualReport/2021_IC3Report.pdf (accessed on 10 May 2024). [25]

French Ministry for Europe and Foreign Affairs (2018), *Paris Call for Trust and Security in Cyberspace*, https://pariscall.international/en/ (accessed on 10 May 2024). [6]

Gen (2023), *2023 Norton Cyber Safety Insights Report*, https://www.gendigital.com/media/qcymrc1i/2023-ncsir-us-global-report_final.pdf (accessed on 3 June 2024). [49]

GEODE, The Hague Centre for Strategic Studies and Cyberpeace Institute (2021), *Paris Call for Trust and Security in Cyberspace: Working Group 5 - Building a Cyberstability Index*, Paris Call, https://pariscall.international/assets/files/PWGR5-8-11-21.pdf (accessed on 10 May 2024). [7]

Goodhart, C. (1975), "Problems of monetary management: The U.K. experience", *Papers in Monetary Economics*, Vol. 1/1, pp. 1-20, https://doi.org/10.1007/978-1-349-17295-5_4. [45]

GPA (2020), *Navigating the global data privacy landscape: The 2020 GPA census*, Global Privacy Assembly, https://globalprivacyassembly.org/the-assembly-and-executive-committee/gpa-census (accessed on 10 May 2024). [28]

ISC2 (2023), *Cyber Workforce Study 2023*, https://www.isc2.org/Research (accessed on 10 May 2024). [50]

ISC2 (2022), *ISC Cybersecurity Workforce Study 2022*, https://www.isc2.org/Research (accessed on 10 May 2024). [73]

Ishak, N. (2022), *Is Russia Holding Back from Cyberwar?*, Vox, https://www.vox.com/2022/3/19/22986316/russia-ukraine-cyber-attacks-holding-back (accessed on 10 May 2024). [74]

ITU (2021), *Global Cybersecurity Index 2020*, International Telecommunications Union Publications, https://www.itu.int/epublications/publication/D-STR-GCI.01-2021-HTM-E (accessed on 10 May 2024). [35]

ITU (2021), *Global Cybersecurity Index 2020 – Frequently Asked Questions*, International Telecommunications Union Publications, https://www.itu.int/en/ITU-D/Cybersecurity/Documents/GCIv4/GCI-FAQ.pdf (accessed on 10 May 2024). [41]

Iwaya, S., E. Koksal-Oudot and E. Ronchi (2021), "Promoting comparability in personal data breach notification reporting"*, OECD Digital Economy Papers*, No. 322, OECD Publishing, Paris, https://doi.org/10.1787/20716826. [27]

Kerner, A., M. Jerven and A. Beatty (2017), "Does it pay to be poor? Testing for systematically underreported GNI estimates", *The Review of International Organizations*, Vol. 12/1, https://doi.org/10.1007/s11558-015-9239-3. [10]

Lightcast (2024), *Lightcast data (database)*, https://lightcast.io/about/data (accessed on 3 June 2024). [51]

MITRE (2023), *Common Vulnerabilities Exposure*, https://www.cve.org/ (accessed on 10 May 2024). [33]

Moore, T. (2010), "The economics of cybersecurity: principles and policy options", *International Journal of Critical Infrastructure Protection*, Vol. 3/3-4, pp. 103-117, https://doi.org/10.1016/j.ijcip.2010.10.002. [3]

NIC.br (2020), *Digital Security: An analysis of risk management in Brazilian enterprises*, Brazilian Network Information Center, https://cetic.br/media/docs/publicacoes/7/20221011154720/sectoral_studies-digital_security.pdf (accessed on 10 May 2024). [18]

NIST (2022), “National Vulnerability Database”, *Information Technology Laboratory (database)*, National Institute of Standards and Technology, https://nvd.nist.gov/ (accessed on 10 May 2024). [75]

NIST (2020), *Integrating Cybersecurity and Enterprise Risk Management (ERM)*, National Institute of Standards and Technology, https://doi.org/10.6028/NIST.IR.8286 (accessed on 10 May 2024). [59]

OECD (2024), *Building a Skilled Cyber Security Workforce in Europe: Insights from France, Germany and Poland*, OECD Skills Studies, OECD Publishing, Paris, https://doi.org/10.1787/23078731. [52]

OECD (2024), “ICT Access and Use by Businesses”, *OECD Telecommunications and Internet Statistics (database)*, https://doi.org/10.1787/9d2cb97b-en (accessed on 3 June 2024). [70]

OECD (2023), *Building a Skilled Cyber Security Workforce in Five Countries: Insights from Australia, Canada, New Zealand, United Kingdom, and United States*, OECD Skills Studies, OECD Publishing, Paris, https://doi.org/10.1787/23078731. [54]

OECD (2023), *Building a Skilled Cyber Security Workforce in Latin America: Insights from Chile, Colombia and Mexico*, OECD Skills Studies, OECD Publishing, Paris, https://doi.org/10.1787/23078731. [53]

OECD (2023), “ICT Access and Usage by Households and Individuals”, *OECD Telecommunications and Internet Statistics (database)*, https://doi.org/10.1787/b9823565-en (accessed on 10 May 2024). [22]

OECD (2022), *OECD Policy Framework on Digital Security: Cybersecurity for Prosperity*, OECD Publishing, Paris, https://doi.org/10.1787/a69df866-en. [8]

OECD (2022), *OECD Work on Digital Security*, https://www.oecd.org/digital/ieconomy/digital-security/oecd-work-on-digital-security-policy.pdf (accessed on 10 May 2024). [1]

OECD (2022), *Recommendation of the Council on Digital Security Risk Management for Economic and Social Prosperity*, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0415 (accessed on 10 May 2024). [40]

OECD (2022), *Recommendation of the Council on National Digital Security Strategies*, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0480 (accessed on 10 May 2024). [32]

OECD (2022), *Recommendation on the Treatment of Digital Security Vulnerabilities*, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0482pdf (accessed on 10 May 2024). [34]

OECD (2021), “Enhancing the digital security of products: A policy discussion”*, OECD Digital Economy Papers*, No. 306, OECD Publishing, Paris, https://doi.org/10.1787/cd9f9ebc-en. [4]

OECD (2021), *OECD Skills Outlook 2021: Learning for Life*, OECD Publishing, Paris, https://doi.org/10.1787/0ae365b4-en. [55]

OECD (2020), "Going Digital integrated policy framework", *OECD Digital Economy Papers*, No. 292, OECD Publishing, Paris, https://doi.org/10.1787/20716826. [38]

OECD (2020), *OECD Digital Economy Outlook 2020*, OECD Publishing, Paris, https://doi.org/10.1787/bb167041-en. [31]

OECD (2020), *OECD Glossary of Statistical Terms*, https://stats.oecd.org/glossary/ (accessed on 10 May 2024). [13]

OECD (2019), "Measuring digital security risk management practices in businesses", *OECD Digital Economy Papers*, No. 283, OECD Publishing, Paris, https://doi.org/10.1787/20716826. [21]

OECD (2019), *Recommendation of the Council on Digital Security of Critical Activities*, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0456 (accessed on 10 May 2024). [36]

OECD (2019), *Share of enterprises in which own employees carry out ICT security related activities*, https://goingdigital.oecd.org/indicator/60 (accessed on 10 May 2024). [23]

OECD (2017), *Enhancing the Role of Insurance in Cyber Risk Management*, OECD Publishing, Paris, https://doi.org/10.1787/9789264282148-en (accessed on 10 May 2024). [58]

OECD (2012), "Improving the evidence base for information security and privacy policies", *OECD Digital Economy Papers*, No. 214, OECD Publishing, Paris, https://doi.org/10.1787/5k4dq3rkb19n-en. [5]

OECD (2007), *Recommendation of the Council on Electronic Authentication*, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0353 (accessed on 10 May 2024). [39]

Open CSIRT Foundation (2022), *SIM3 Model & References*, https://opencsirt.org/csirt-maturity/sim3-and-references/ (accessed on 10 May 2024). [30]

PIPS (2015), *Cyber Readiness Index 2.0*, Potomac Institute for Policy Studies, Arlington, https://www.potomacinstitute.org/academic-centers/cyber-readiness-index (accessed on 10 May 2024). [43]

Pratap, B. and N. Priyaranjan (2023), "Macroeconomic effects of uncertainty: a Google trends-based analysis for India", *Empirical Economics*, Vol. 65, pp. 1599-1625, https://doi.org/10.1007/s00181-023-02392-z. [65]

Statistics Canada (2022), *Canadian Survey of Cyber Security and Cybercrime (CSCSC)*, http://www.statcan.gc.ca/en/survey/business/5244 (accessed on 10 May 2024). [16]

Swiss Re (2023), *What You Need to Know About the Cyber Insurance Market*, https://www.swissre.com/risk-knowledge/advancing-societal-benefits-digitalisation/about-cyber-insurance-market.html (accessed on 10 May 2024). [57]

The Economist (2014), *Ranking the Rankings*, The Economist Group Limited, https://www.economist.com/international/2014/11/08/ranking-the-rankings (accessed on 10 May 2024). [71]

U.S. Bureau of Labor Statistics (2023), *What Is BLS Doing to Maintain Data Quality as Response Rates Decline?*, http://www.bls.gov/blog/2023/what-is-bls-doing-to-maintain-data-quality-as-response-rates-decline.htm (accessed on 3 June 2023). [19]

United Nations Institute for Disarmament Research (2021), *Cyber Policy Portal*, https://cyberpolicyportal.org (accessed on 19 May 2024). [42]

Verizon (2023), *2023 Data Breach Investigations Report*, https://www.verizon.com/business/resources/Tb49/reports/2023-data-breach-investigations-report-dbir.pdf (accessed on 3 June 2024). [47]

Weinberg, L. (2020), "The Google Trends Uncertainty (GTU) index: a measure of economic policy uncertainty in the EU using Google trends", *Undergraduate Economic Review*, Vol. 17/1, https://digitalcommons.iwu.edu/uer/vol17/iss1/2 (accessed on 10 May 2024). [66]

Woloszko, N. (2020), "Tracking activity in real time with Google Trends", *OECD Economics Department Working Papers*, No. 1634, OECD Publishing, Paris, https://doi.org/10.1787/6b9c7518-en. [69]

Annex A. Cybersecurity indicators in the OECD ICT Access and Use Databases

The indicators related to digital security in the OECD ICT Access and Use Databases are include a fraction of the relevant cybersecurity indicators from Eurostat surveys. There are two firm-level indicators, the share of business experiencing ICT incidents (security breaches) (E3) and the share of enterprises with formal policy to manage ICT privacy risks (E7). While these indicators are available by default for enterprises with ten employees or more in the business economy (excluding financial services), they are also available by employment size class (enterprises with 10-49 employees, enterprises with 50-249 employees and enterprises with 250 employees and more) and by industry (manufacturing, construction, wholesale and retail trade and so on).

However, data availability is limited across the years and driven primarily by countries that co-ordinate data collection through Eurostat. Items on security incidents (e.g., E3) were included in 2021 and 2018. But questions about policies to manage ICT privacy risks have not been included in recent years.

Clearly, these indicators primarily relate to outcomes (share of incidents) and outputs (share of business with a policy to manage ICT privacy risks). However, it is important to note that both indicators capture only the extensive margin, not the intensive margin (the share of enterprises that experienced incidents rather than the severity of incidence experienced).

More indicators are available for individuals, including a battery of items related to security incidents (I3-I3D) which are available for most countries in 2015 and 2019. Note that, if available, the data only relate to individuals between the ages of 16 and 74 and can typically be broken down by age group, gender, education, labour market status and household income.

While these indicators measure the extensive margin of negative outcomes, other indicators provide a sense of the actions taken or not taken by individuals online that might be partly indicative of security concerns. They include the share of individuals that did not submit forms to public authorities out of concerns over data protection and security in the last 12 months (F4G), the share of individuals who have modified the security settings of their Internet browser in the last 12 months (H1J), the share of individuals who have provided personal information on the Internet in the last 12 months (I6) and the share of individuals who have managed access to their personal information on the Internet in the last three months. The latter question is broken down by means of managing access: restricting access to their geographical location (I7A), limiting access to their profile or content on social networking sites (I7B), refusing the use of personal information for advertising purposes (I7C), reading privacy policy statement before providing personal data (I7D).

In addition, there are two indicators that inquire about the changes to the settings of individuals' Internet browser to prevent or limit the number of cookies (I8) and the use of anti-tracking software (I9). However, the line blurs for the later badge of indicators between actions taken out of privacy concerns and actions taken out of security concerns.

Table A.1. Current availability of relevant indicators from the OECD's ICT Access and Use databases

Enterprises (10 employees and more, business sector excl. financial services) and individuals (aged 16-74), 2015-2021

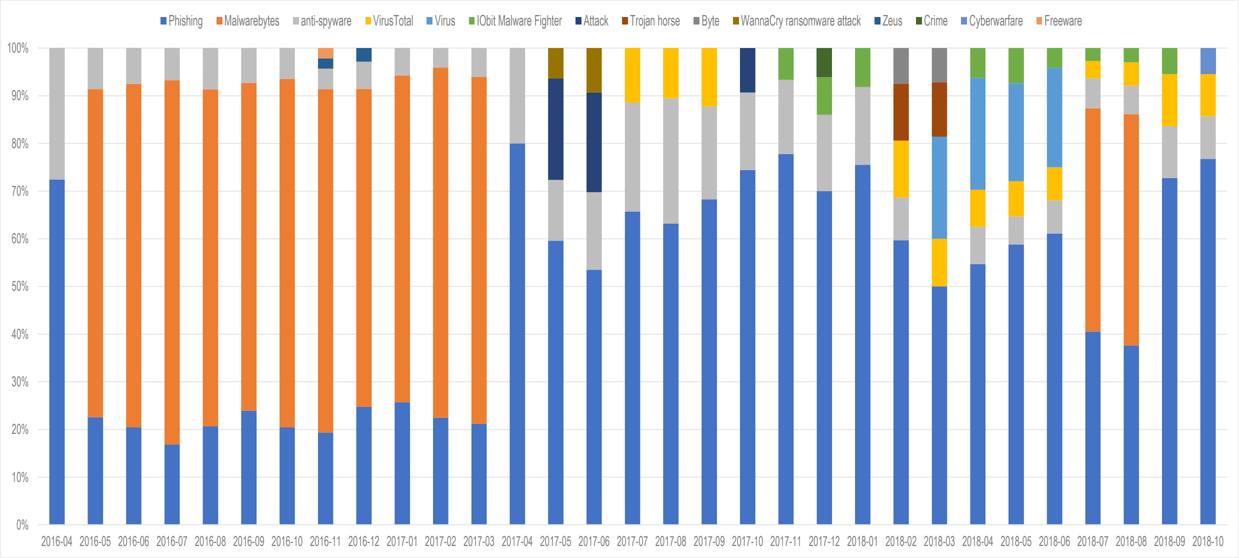
Indicator	Indicator caption	2015	2016	2017	2018	2019	2020	2021
Enterprises (40 countries, 38 OECD countries, Croatia, and Romania)								
E3	Businesses experiencing ICT Incidents (security breaches) (%)	5.0%	12.5%	7.5%	**80.0%**	10.0%	12.5%	**70.0%**
E7	Businesses with formal policy to manage ICT privacy risks (%)	5.0%	5.0%	7.5%	5.0%	10.0%	7.5%	2.5%
Individuals (42 countries, 38 OECD countries, Brazil, Bulgaria, Croatia, and Romania)								
F4G	Individuals did not submit forms to public authorities: personal data protection and security concerns - last 12 m (%)	64.3%	64.3%	**69.0%**	**66.7%**	**66.7%**	**66.7%**	**66.7%**
H1J	Individuals having caught a virus or other computer infection with impacts - last 3 m (%)	2.4%	4.8%	7.1%	2.4%	0.0%	0.0%	0.0%
I3	Individuals having experienced a financial loss from fraudulent payment - last 3 m (%)	**71.4%**	7.1%	9.5%	9.5%	**73.8%**	9.5%	4.8%
I3A	Individuals having experienced a financial loss from phishing/pharming - last 3 m (%)	**69.0%**	4.8%	7.1%	7.1%	**71.4%**	7.1%	4.8%
I3B	Individuals having experienced abuse of personal information/privacy violations - last 3 m (%)	**69.0%**	2.4%	7.1%	7.1%	2.4%	2.4%	2.4%
I3C	Individuals who have experienced security incidents - last 3 m (%)	**71.4%**	4.8%	9.5%	7.1%	**73.8%**	9.5%	4.8%
I3D	Individuals who have modified the security settings of Internet browsers - last 12 m (%)	**69.0%**	2.4%	4.8%	4.8%	**71.4%**	2.4%	2.4%
I6	Individuals who have provided personal information on the Internet - last 12 m (%)	0.0%	64.3%	2.4%	2.4%	0.0%	0.0%	2.4%
I7A	Individuals who managed access to personal data on the internet: read privacy policy statements before providing personal data	0.0%	64.3%	2.4%	2.4%	0.0%	**66.7%**	**69.0%**
I7B	Individuals who managed access to their personal information on the Internet: limit access to their profile or content on social networking sites - last 12 m (%)	0.0%	64.3%	2.4%	2.4%	0.0%	**66.7%**	**69.0%**
I7C	Individuals who managed access to their personal information on the Internet: refuse allowing the use of personal information for advertising purposes - last 12 m (%)	0.0%	64.3%	2.4%	2.4%	0.0%	**66.7%**	**69.0%**
I7D	Individuals who managed access to their personal information on the Internet: restrict access to their geographical location - last 12 m (%)	0.0%	0.0%	0.0%	0.0%	0.0%	64.3%	**69.0%**

Note: Years in which data on at least two thirds of the target population are available are marked in bold.
Source: Authors elaboration based on the OECD's ICT Access and Use databases (2024[70]; 2023[22])

Annex B. A case study of top related topics

There was a substantial and prolonged surge in France's cyber uncertainty index from September 2016 to February 2018 (Figure 5.5, panel a). This surge highlights a notable increase in uncertainty concerning cyber incidents within the country, aligning with a series of major cyber incidents like the DDoS attacks on Dyn in October 2016, the WannaCry ransomware attack in May 2017, and the 2017 Ukraine ransomware attacks (NotPetya). The most frequently searched topics alongside the combination of the nine cyber-incident topics forming the GTCU index can be found in Figure B.1. Most salient top related topics to France's GTCU index from April 2016 to October 2018. This figure reveals a marked shift in public concern from "Malwarebytes" to "phishing" following these attacks and persisting for several months.

Figure B.1. Most salient top related topics to France's GTCU index from April 2016 to October 2018



Note: Top topics data was cleaned to drop both irrelevant (i.e., those that would directly correspond to the nine topics used to construct the GTCU index) and repetitive topics.
Source: Authors' elaboration based on Google Trends data.

The sustained public focus on phishing attacks, evident in the persistent rise of the dark blue line in Figure B.1. Most salient top related topics to France's GTCU index from April 2016 to October 2018 highlights the need for tailored cybersecurity policies to effectively protect the most vulnerable populations – individuals and SMEs notably – that are at the frontline for these attacks. To address this growing concern, policy makers could develop comprehensive cybersecurity awareness campaigns that promote the recognition of phishing attempts, such as identifying common tactics, verifying sender information, and being aware of popular scams. Additionally, these campaigns could emphasise secure practices for information handling, including using strong passwords, enabling multi-factor authentication, and avoiding entering personal information on unverified websites. By focusing on these critical aspects, policy makers can enhance cybersecurity resilience and preparedness to counter evolving cyber threats.

Endnotes

[1] However, see Dourado (2012[72]) for a cautious note on market failures in the realm of cybersecurity.

[2] The Going Digital Integrated Policy Framework identifies five policy domains that are critical to ensuring trust in digital environments: digital security, privacy, consumer policy, digital risk management and small and medium-sized enterprises (OECD, 2020[38]).

[3] An example is the National Vulnerability Database (NVD) maintained by the National Institute of Standards and Technology of the US Department of Commerce (NIST, 2022[75]).

[4] Prominent examples of ranked indices whose publication has often triggered a policy response include the OECD's *Programme for International Student Assessment* (PISA), which rates pupils academic performance, and the now-discontinued *Ease-of-Doing-Business Index*, which was created by the World Bank and ranked countries according to how easy it is to set-up and run a business (The Economist, 2014[71]).

[5] From reference year 2022 onwards, the survey will be known as the Public Sector Survey on Cyber Security and Cybercrime (PSSCSC) with a focus on the Canadian public service and crown corporations.

[6] Note that in this case enterprises are asked about current practice but to report incidents in the previous year, which makes reverse causality more likely.

[7] Eurostat provides data on the share of enterprises that experienced ICT incidents in for 32 countries: Austria, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Cyprus, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Montenegro, the Netherlands, Norway, Poland, Portugal, Romania, Serbia, the Slovak Republic, Slovenia, Spain, Sweden and Türkiye (Eurostat, 2022[20]).

[8] Digital security vulnerabilities include code vulnerabilities, which are located in the code of products, and system vulnerabilities, which affect information systems and include primarily misconfigurations and the non-application of products' security updates, patches or other mitigations (OECD, 2022[34]).

[9] The European Union's NIS Directive mandates the establishment of CSIRTs.

[10] The 2023 ISC2 report provides a workforce gap estimate based on 20 countries: Australia, Brazil, Canada, France, Germany, India, Ireland, Japan, the Kingdom of Saudi Arabia, Mexico, the Netherlands,

Nigeria, the People's Republic of China, Singapore, South Africa, South Korea, Spain, the United Arab Emirates, the United Kingdom and the United States (ISC2, 2023[50]; ISC2, 2022[73]).

[11] Data are pulled from company career sites, job boards and job posting aggregators, and they are vetted to identify and avoid duplicate postings. Using natural-language processing technology and artificial intelligence tools, dozens of key elements from every job posting are extracted and standardised. Lightcast data provides insights on job titles, responsibilities, in-demand skills, experience and education requirements, and advertised salary.

[12] According to IFW, an organisation that tracks circulations of newspapers in Germany, the FAZ had a circulation of 254 000 copies in 2015, making it the second most read newspaper among Germany's high-profile national newspapers. In addition to the FAZ, other popular newspapers include *Süddeutsche Zeitung*, *Die Welt*, *Die Zeit*, *Frankfurter Rundschau* and *Die Tageszeitung*. Bild, a tabloid newspaper, had a circulation of more than 1.1 million copies in 2015.

[13] This seems to be consistent with initial fears of major cyberattacks in the run-up to Russia's full-scale invasion of Ukraine, but an absence of actual attacks in the wake of the invasion, which surprised many commentators. See, for instance, Ishak (2022[74]).

[14] In the case of anglophone publications the combination of terms “cybersecurity” AND (“breach” OR “threat” OR “incident” OR “risk” OR “attack” OR “breach”), for instance, should succeed in removing most marriage announcements.

[15] The literature emphasises that GTU indices can be sensitive to the choice of keywords and country-specific search trends, underscoring the need for careful interpretation when using keywords (Weinberg, 2020[66]).

[16] The Federal Reserve's *Beige Book* for the United States and the Reserve Bank's *Monetary Policy Statement for Australia* are official documents that gather information on current economic conditions based on interviews with key business contacts, economists and market experts (among other sources). Making them a good proxy of entrepreneurs' uncertainty as regards future business conditions according to the authors (Castelnuovo and Tran, 2017[61]).

[17] The full list of Google Trends categories can be found here: https://github.com/pat310/google-trends-api/wiki/Google-Trends-Categories.

[18] More information on the aggregation procedure can be found in the online appendix of Castelnuovo and Tran (2017[61]).

[19] Please see: https://newsinitiative.withgoogle.com/resources/trainings/fundamentals/google-trends-understanding-the-data/#:~:text=Rising%20data.&text=If%20you%20see%20%E2%80%9CBreakout%E2%80%9D%20instead,for%20the%20selected%20time%20frame.