Compliance, Ethics & Sustainability

An international journal with a European focus

JAARGANG 25 - JULI 2025 3

DEN HOLLANDER

Uitgeverij Den Hollander BV
Postbus 325 7400 AH Deventer
tel.: 0570 - 751 225
e-mail: info@denhollander.info
www.uitgeverijdenhollander.nl

Abonnementsprijs
Zie:https://denhollander.info/

Nieuwe abonnementen
Abonnementen kunnen via de website worden afgesloten
Abonnementen kunnen worden gestart per 1 januari van een kalenderjaar. Valt de aanvraag van een abonnement niet samen met het begin van een kalenderjaar, dan wordt op aanvraag het tarief voor het lopende jaar naar rato berekend indien het abonnement ook in het jaar daaropvolgend wordt afgenomen.
Op al onze abonnementen zijn de Algemene voorwaarden van toepassiing

Beëindiging abonnement
Abonnementen kunnen alleen schriftelijk, tot 1 december van het lopende abonnementsjaar, worden opgezegd. Bij niet-tijdige opzegging wordt het abonnement automatisch voor een kalenderjaar verlengd.

Adreswijziging
Bij adreswijziging wordt u verzocht deze zo Spoedig mogelijk en bij voorkeur schriftelijk door te geven aan de uitgeverij onder vermelding van: adreswijziging Tijdschrift voor Compliance.

Citeerwijze
CE&S 2025, nr. 3

DEN HOLLANDER
UITGEVERIJ

COMPLIANCE, ETHICS & SUSTAINABILITY
Journal

JAARGANG 25 - JULI 2025 - NUMMER 3

Voorwoord

E. Nkune, P. Wijma MSc and prof. mr. dr. B. Snijder-Kuipers

Privacy, Data Protection and Cyber security will most likely be points, high on the agenda for most Risk Committee meetings. These are the themes for this, the third edition of the Compliance, Ethics and sustainability Journal for 2025.

In their article "*Apple Inc v Secretary of State for the Home Department*", **Julian Hayes**, **Megan Curzon** and **Jenna Gayle** examine the legal stand-off between the international tech giant Apple and the UK Home Secretary. Apple is challenging the Home Secretary's decision to impose a Technical Capability Notice, which is a requirement to maintain the capability to remove end-to-end encryption from its iCloud storage. The authors explore the resulting diplomatic furore and discuss the conflict between privacy and security. They examine the legislation, discuss the pros and cons of encryption and potential solutions to an age-old dilemma.

The increasing prevalence of cyber-attacks across the globe has led to a consequential increase in legislation to protect critical national infrastructure, including in new industry sectors not traditionally seen as critical. However, for many organisations the focus on reporting cyber security incidents, as opposed to the perhaps now established process of reporting personal data breaches, is something that may not be familiar to some compliance professionals. **Kelly Hagedorn**, **Alice Portnoy** and **Hanna Hewitt** explore this issue in their article, "*Navigating a New Era of Reporting Cyber Incidents In the UK and EU*", which seeks to map out the changing reporting landscape in both the UK and EU, providing an overview of what compliance professionals need to consider when updating their processes and procedures around cyber incident reporting.

Can the use of AI solve all of our compliance investigation problems? The answer to this question is examined by **Carys Whomsley** in her article, "*Public Data, Private Risks*". She argues that whilst their utility is there to be seen, the use of chatbots powered by Large Language Models comes with an inherent privacy risk. Whomsley also highlights some of the other issues that investigators will need to address if they want to harness the power of AI in open-source investigations, including hallucinations and copywrite issues. The article examines these issues in turn and suggests how these tools will need to evolve to be sufficiently robust to have a prime place in a compliance investigators armoury.

The co-existence of whistleblower protection and strict data privacy regulation often creates challenging situations for organisations and whistleblowers. *Blowing the whistle* can easily involve the dissemination of personal data, potentially constituting a data breach under GDPR. "*Privacy vs. Whistleblowing: Can Data Breaches Be Justified During Public Disclosure?*" This question will be answered in this article by **Anastasia Avramenko**. Balancing public interest, fundamental rights of individuals and the harm from data breaches, the author examines EU legislation and maps a suggested way forward through this complex landscape.

When discussing sensitive personal data, the compliance officer likely thinks first of the financial industry, where transactions reveal detailed profiles, especially when combined with publicly shared social media insights. However, one other category of personal data, literally close to our hearts and equally important, is the data about our health. The healthcare industry shares some data security challenges with other industries but also comes with its own particularities. The article "*Europe's Health Industry and Regulation, Anonymisation and Security*" by **Tanya Chib, Dr. Anna Hakkers & Renate van Kempen,** combines the insights of three experts on regulation, cybersecurity and anonymisation in this vital industry, and provides the reader with an overview of the relevant regulations and some of its practical challenges regarding data security.

A large number of books and articles appear on the topic of business ethics that address pressing issues in a practical way and make concrete recommendations for promoting the ethics and integrity of organisations and their employees. Not everyone knows where to find these publications or has time to read them. That's why **Edgar Karssing** discusses articles and books in this area and writes about them. In this issue Karssing published the second part out of two articles on power and ethics. In the two articles Karssing highlights several questions and perspectives related to this theme. In this second article the attention is focused on (i) the background of power by Robert Greene, (ii) the seven rules of power by Jeffrey Pfeffer and (iii) the ethics of political action.

We hope you find this issue insightful and engaging.

Edward Nkune, Pauline Wijma & Birgit Snijder-Kuipers

Apple Inc v Secretary of State for the Home Department: the latest iteration of the never-ending security versus privacy debate

Mr J. Hayes, M. Curzon and J. Gayle[1]

The precise subject matter of the legal stand-off between international tech giant Apple and the UK Home Secretary, currently before the Investigatory Powers Tribunal ("IPT") in London, is unknown. Apple is forbidden from disclosing it, the Home Secretary refuses to disclose it and, when hearing cases, the IPT itself must ensure information is not disclosed which could jeopardise national security, the prevention or detection of serious crime, the UK's economic well-being or the ongoing work of the intelligence services. From a media leak in February and a preliminary IPT ruling in April disclosing the barest of case details, we know only that Apple is challenging the Home Secretary's decision to impose a Technical Capability Notice ("TCN") requiring it to maintain the capability to remove end-to-end encryption ("E2EE") from its iCloud storage. It is Kafkaesque that something about which so little is known has caused Apple to pull its Advanced Data Protection encryption entirely from the UK, provoked a US/UK diplomatic spat,[2] and called into question the UK's European Convention on Human Rights ("ECHR") compliance. In fact, the dispute is the latest manifestation of a decades-long tussle over the merits of encryption, pitting against each other the competing priorities of security and privacy. That the dispute continues to generate such fierce argument is a sign not only of the intractability of the debate but also of the failure of the protagonists to level with the public about the trade-offs involved, over many years.

1. The origins of encryption & the dilemma it poses

In cryptography, encryption is the process of transforming comprehensible information (plaintext) into scrambled code (ciphertext) which only the intended recipient can decode. Its earliest known roots are in ancient Greek and Roman history but today it is a ubiquitous feature of our daily digital communications, from banking and e-commerce to video-conferencing and instant messaging.

Modern, computer-based encryption was developed in the 1960s, with banks some of the first to recognise its commercial application. From the outset, the US Government was anxious that encryption technology would fall into the hands of foreign adversaries, causing them to 'go dark', thereby impeding the interception work of the security services. In what were the first skirmishes of the so-called 'crypto-wars', the US and UK authorities battled with liberal-minded cryptographic researchers and proponents of wider access to encryption, attempting to curtail the level of commercial security it provided and the export of encryption technology.

However, growing mistrust of the state and security services, culminating with the 2013 revelations by former National Security Adviser, Edward Snowdon, of covert global telecommunications surveillance by the US and UK, alongside increasingly disruptive ransomware incidents, heightened the sense of digital insecurity and propelled the roll-out of E2EE across online platforms. Unlike standard encryption, where a service provider retains a decryption key which can unlock the cipher text sent between user devices over a messaging platform, where true E2EE is used, third parties are no longer able to unlock the cipher text in transit between user devices. The only way law enforcement agencies could access E2EE messages would be by somehow gaining access to the devices of the sender or recipient of the message.

The increasing deployment of E2EE and the obstacles this threw up to detecting unlawful online activity have aroused intense political hostility. At their most strident, UK Government Ministers vilified major US tech companies as enablers of child abuse.[3] In 2022, the UK Government launched a much-derided media campaign alleging that rolling out E2EE would be "*like turning the lights off on the ability to*

1. Julian Hayes, Megan Curzon and Jenna Gayle, BCL Solicitors, London. Julian is a Partner and a specialist in surveillance and data protection law. Megan is an associate in the business crime and regulatory team and Jenna is a legal assistant working across all of BCL's practice areas.
2. www.pymnts.com/cpi-posts/trump-criticizes-uk-demand-for-apple-user-data-calls-it-something-you-hear-about-with-china/
3. www.bbc.co.uk/news/technology-65686989

identify child sex abusers online".[4] Child safety campaigners have called for social media companies to explore alternative methods of maintaining online safety before resorting to E2EE,[5] whilst UK and EU law enforcement bodies called for 'security by design' when deploying E2EE so as to maintain the ability to both identify and report illegal activities – effectively a call for the creation of 'backdoors' into E2EE services.[6]

It is certainly true that criminals and other 'bad actors' have been quick to spot the clandestine opportunities presented by E2EE services, whether on mainstream platforms or those such as the EncroChat encrypted messaging network which were dedicated to organised criminal activity.[7] The French authorities' eventual infiltration of EncroChat led to 6,558 arrests and the seizure or freezing of over 900 million euros worldwide.[8]

Yet trying to roll-back the deployment of E2EE is, in reality, inimical to law enforcement interests and a fool's errand. The business models of many widely-used platforms, such as WhatsApp and iMessage, are built around privacy.[9] It is the platform providers' success in attracting subscribers and customers which has ensured deep pools of user data in which law enforcement and intelligence agencies have often been able to fish. Even if service providers cannot disclose what their users said in E2EE messages, they will often retain 'communications data' – the 'who', 'what', 'when', 'where' and 'how' of a message. This can include the names and contact details of the sender, their IP addresses and information about the device they used to communicate, all of which can contain vital clues for investigators. Far from 'going dark', some commentators argue that tech companies have, in fact, created the conditions for a golden age of state surveillance.[10]

Equally, cryptic allusions by politicians and state agencies to finding 'technical solutions' to the security versus privacy conundrum risk fuelling suspicion that they are really seeking backdoors into E2EE services. Yet, as privacy campaigners have repeatedly warned, it is not possible to create a backdoor available only to 'the good guys'. Weakening E2EE protection for law enforcement purposes undermines it for all purposes. With UK Government statistics suggesting half of businesses experienced some form of cyber security breach or attack in the last 12 months,[11] and the National Crime Agency reporting that cybercrime "*costs the UK billions of pounds, causes untold damage, and threatens national security*",[12] measures which undermine E2EE would only worsen the situation. Proposals to create backdoors into E2EE therefore overlook the risk that this would increase levels of malicious cyber activity and erode our trust in the security of digital products generally.

Further, while societies generally agree on the goal of protecting children and vulnerable people from predators, and in guarding against terrorist activity, in an increasingly polarised world the application of covert surveillance methods to achieve more politically contentious objectives risks provoking controversy. Present day examples include the use of surveillance to apprehend illegal immigrants[13] or intercepting the electronic communications of women suspected of seeking an abortion following the US Supreme Court's reversal of *Roe v Wade in 2022*[14].

Moreover, there is evidence that, to undermine freedom of speech and subvert domestic political opposition, less open and tolerant regimes justify demands to break E2EE by reference to well-intentioned surveillance measures taken in the name of security, particularly in traditionally liberal democratic states.[15]

2. Lawful exceptional access measures

To ensure continued access to online communications despite E2EE, governments have historically tried various solutions. In the 1990s, the US National Security Agency advocated for the creation of a backdoor known as the 'Clipper Chip' into voice and data messages on encrypted phones. The key to accessing this data would be held in escrow by a trusted third party who would release it to the authorities where lawful permission was given. To dispel the concerns of the privacy community, the public would be informed about the presence of the backdoor into their devices and commercial incentives would be offered to phone manufacturers to adopt it. However, the Clipper Chip attracted significant criticism, including from libertarians and Silicon Valley businessmen. It was attacked both for its technical vulnerability and its potential for abuse;

4. www.bbc.co.uk/news/59964656
5. www.iwf.org.uk/news-media/blogs/not-all-encryption-is-the-same-social-media-is-not-ready-for-end-to-end-encryption/
6. www.europol.europa.eu/media-press/newsroom/news/european-police-chiefs-call-for-industry-and-governments-to-take-action-against-end-to-end-encryption-roll-out
7. news.sky.com/story/encrochat-what-it-is-who-was-running-it-and-how-did-criminals-get-their-encrypted-phones-12019678
8. www.europol.europa.eu/media-press/newsroom/news/dismantling-encrypted-criminal-encrochat-communications-leads-to-over-6-500-arrests-and-close-to-eur-900-million-seized#.
9. www.whatsapp.com/privacy
10. slate.com/technology/2015/07/encryption-back-doors-arent-necessary-were-already-in-a-golden-age-of-surveillance.html
11. www.gov.uk/government/statistics/cyber-security-breaches-survey-2024/cyber-security-breaches-survey-2024#chapter-4-prevalence-and-impact-of-breaches-or-attacks
12. [11]
13. theintercept.com/2019/12/22/ice-social-media-surveillance/
14. therecord.media/anti-abortion-group-massachusetts-accused-intercepting-messages
15. www.medianama.com/2024/04/223-whatsapp-india-operations-end-to-end-encryption/#.

it failed to gain public support and was eventually made redundant by even stronger cryptographic technology.

Under the UK's Investigatory Powers Act 2016, the Secretary of State may authorise the intelligence services and certain law enforcement authorities to compromise E2EE by engaging in 'equipment interference', that is, by hacking. (Similar laws exist in various EU member states.[16]) In the UK, equipment interference may be authorised in the interests of national security or for the purpose of preventing / detecting serious crime. In 2023, approval was given for 3,101 targeted equipment interference warrants.[17]

Authorised equipment interference might be achieved straightforwardly, by downloading the content of a phone which is carelessly left unattended. However, it might also be achieved by exploiting software vulnerabilities in a device. The notorious Pegasus spyware, implicated in the targeting of multiple politicians, journalists and lawyers by its developers,[18] leveraged such vulnerabilities to provide covert access to E2EE messaging platforms such as WhatsApp and iMessage.[19] However, even where equipment interference is lawfully authorised and carefully superintended, it poses ethical dilemmas; failing to notify service providers of the software vulnerabilities they use leaves the software and its users open to exploitation by bad actors for more nefarious purposes.

By way of alternative, governments have sometimes granted themselves the power to require service providers to retain a key to encrypted messaging. The Home Secretary's TCN to Apple earlier this year is an example of this approach. Whilst controversial, this notice did not 'come out of the blue'. The power to issue a notice requiring the maintenance of the capability to remove electronic protection has existed since 2018, although the outcry which greeted the Apple TCN suggests few, if any, such notices have ever been issued to US service providers.[20]

Elsewhere, in 2018, the Australian Parliament passed the Telecommunications (Assistance and Access) Act enabling the authorities to compel providers to create backdoors into E2EE services for the investigation of offences carrying penalties of three or more years' imprisonment. In 2021, Apple itself floated but then postponed a plan to scan images in the iCloud and on users' devices for child sexual exploitation and abuse content ("CSEA") material and to report them to the US authorities. In 2022, the European Commission proposed an EU-wide Regulation on combatting child sexual abuse which mandated the detection, reporting and removal of CSEA material by hosting services and interpersonal communication service providers. The precise measures to be included in the Regulation are not yet agreed but they would almost inevitably involve the weakening of encryption.

More recently still, the UK's Online Safety Act, which received Royal Assent in 2023 making it an Act of Parliament, controversially gave the regulator, Ofcom the power to use 'accredited technology' to identify CSEA communicated privately.[21] At the time, the Government denied that the measure could force platform providers to compromise E2EE on their services, but it is widely believed that the provision was introduced to pave the way for law enforcement to roll-out automated on-device scanning, also known as 'client-side scanning' ("CSS"). CSS involves downloading software onto individual devices such as a smartphones, tablets and computers to conduct algorithmic scanning of text, images, videos and files for prohibited content before it is sent via the device. Where prohibited material is discovered, the CSS software may prevent it from being sent and may alert third parties such as the police.

Proponents of CSS hail it as a minimally intrusive technological solution which protects the public, avoids traditional security concerns about secret backdoors into encrypted communications via 'ghost protocols or 'escrow keys', leaves E2EE intact, and reconciles the competing demands of law enforcement and privacy campaigners. Opponents of CSS counter that it constitutes an insidious form of bulk surveillance, effectively placing bugs in everyone's pockets and altering the way we interact with our electronic devices and each other; that it is prone to false positives and manipulation by sophisticated criminals and hostile states; and that it is liable to 'scope creep' – once accepted in principle, the temptation to scan for other offences and socially objectionable behaviour will become irresistible.

3. Data protection, human rights and encryption

Apart from the ethical dilemma posed by E2EE, current data protection standards and human rights jurisprudence, applicable in both the UK and Europe, militate against its general weakening.

The GDPR and the similarly-worded UK GDPR UK require that those controlling or processing personal data (broadly, any information relating to an identifiable individual) implement appropriate technical measures to ensure a level of data security appropriate to the risk, including the encryption of personal data.[22] As the UK's data watchdog, the Information

16. www.europarl.europa.eu/RegData/etudes/STUD/2017/583137/IPOL_STU(2017
17. https://ipco-wpmedia-prod-s3.s3.eu-west-2.amazonaws.com/E03270100-HC_603-IPCO-Annual-Report-2023-Web_Accessible.pdf, p.89
18. www.bbc.co.uk/news/world-57891506
19. www.mcafee.com/learn/what-is-pegasus-spyware/#.
20. www.theguardian.com/us-news/2025/feb/26/tulsi-gabbard-uk-apple
21. Online Safety Act 2023, s.121
22. Article 32(1) of the GDPR and UK GDPR

Commissioner's Office ("ICO") has stated, "*[s]ystems that do not use E2EE can be abused, creating the risk for financial fraud, exposure to harmful content and other harms…[from] a data protection perspective, E2EE acts as a key enabler for compliance with the requirements of data protection law…positioning E2EE and online safety as being in inevitable opposition is a false dichotomy.*"[23] Bearing this unambiguous regulatory endorsement in mind, it is difficult to see how those handling personal data as controllers or processors could properly discharge their duty to process personal data securely without using E2EE wherever possible. As the ICO has trenchantly stated, "*[m]easures that would introduce widespread "backdoors" to encrypted channels or otherwise enable indiscriminate widespread access, would create systemic weaknesses unacceptably undermining security and privacy rights…*"[24]

Human rights, too, play a role; in *Podchasov v Russia in 2024*,[25] the European Court of Human Rights ("ECtHR") held that the right to respect for private and family life and correspondence under Article 8 of the Convention had been infringed by Russia's insistence that the provider of the Telegram messaging app disclosed technical information to facilitate the decryption of E2EE communications by the Federal Security Service ("FSB"). The weakening of E2EE for Mr Podchasov's communications would amount to its weakening for all Telegram users which would be disproportionate to the anti-terror aims said to be pursued by the FSB.[26]

Although Convention rights are generally only directly enforceable against public authorities rather than service providers, where a service provider is compelled by a public authority to act in a certain way, the interference with the individual's rights can be imputed to the public body which imposed the requirement on the service provider.[27] Given the ECtHR's decision in *Podchasov*, any state requirement to weaken E2EE in a way which simultaneously weakens E2EE for all users would arguably be disproportionate and unlawful.

4. Alternatives to compromising E2EE

Acknowledging the clear social utility of tackling online crime, particularly CSEA and terrorist-related activity, E2EE platform providers point to alternative, less intrusive electronic methods which they already deploy to identify suspicious activity, rather than undermining E2EE. For example, analysis of unusual patterns of online behaviour, such as communications between adults and children who have no other connection to each other, are 'red flags', prompting further enquiry and warnings to law enforcement agencies. One potential method for preventing harms is 'age-gating' services to restrict the availability of potentially harmful content to children and teens, and to keep bad actors from those at risk. Similarly, service providers such as Meta promote parental control over direct messaging to their children[28] and encourage the reporting of negative online behaviour by their users.[29]

Unfortunately, such alternative methods are not problem-free. In contrast to accessing and reviewing message content, which can provide conclusive proof of illegal activity online, metadata analysis is essentially a tool for gathering intelligence. It can only suggest potential illegal activity and, on its own, may even be insufficient for law enforcement to obtain a search warrant for further investigation. Likewise, 'age-gating' is liable to subversion since children and adults lie about their age. When it comes to reporting negative behaviours online, while this can be a useful warning tool, it places the onus on victims to identify abuse when they may be too embarrassed or afraid to do so, or they may not even realise what is happening.

More sophisticated technological solutions, for example 'homomorphic encryption', which could allow for limited analysis of still encrypted data and potentially assuage privacy concerns, remain at a developmental stage and are currently too slow to operate at scale.

5. Striking the right balance

The security and privacy tensions thrown up by communications technology provoke fierce debate, with hyperbolic warnings that vital intelligence sources will 'go dark' or that state surveillance risks our hard-fought liberties. As with most polarised disputes, the truth lies somewhere between the most extreme positions. Notwithstanding the anxieties of law enforcement agencies, advances in the means of human communication over the past 100 years have ensured ever richer sources of crime-fighting data potentially available for those with the desire and means to access it. Likewise, pessimistic warnings by campaign groups and activists of imminent state repression arguably underplay human characteristics of fair-mindedness and rebelliousness.

Any side of the argument which pushes its stance too hard must expect the other side to push back with at least as much vigour. Apple is pushing back in exactly this way by challenging the Secretary of State's recent TCN which it argues would undermine E2EE for everyone. The libertarian-minded US political establishment has lent its support to Apple's struggle, highlighting the importance of E2EE in the protec-

23. ico.org.uk/media2/about-the-ico/documents/4018823/ico-e2ee-paper-02112021.pdf
24. Ibid
25. ECtHR, *Podchasov v Russia,* Application no. 33696/19
26. *Podchasov v Russia*, para 79
27. ECtHR, *Ekimdzhiev v Bulgaria*, Application no. 70078/12
28. about.fb.com/news/2025/04/introducing-new-built-in-restrictions-instagram-teen-accounts-expanding-facebook-messenger/
29. about.fb.com/news/2024/01/teen-protections-age-appropriate-experiences-on-our-apps/

tion of all US citizens. A bi-partisan Congressional committee has even suggested the US Administration should threaten withdrawal from UK – US Data Access Agreement ("DAA"), a key means by which the UK obtains electronic data from US tech companies, to exert pressure on the UK Government to withdraw the TCN.[30]

It is a tribute to the work of the security services and law enforcement authorities that, for the most part, we go about our daily lives untroubled by those who would harm us or threaten our way of life. In an increasingly hostile world, this level of security is not easy to achieve; it requires constant vigilance and sources of intelligence, of which electronic communication is perhaps the most abundant and reliable. Refusal to acknowledge the need for such access, always within the boundaries of the law, is simply a denial of reality.

As the history of encryption suggests, each generation will weigh the risks of allowing state access to communications differently, sometimes favouring safety more and sometimes privacy. But reaching a stable accommodation between the two priorities requires trust between the protagonists underpinned by transparency and an informed public debate about the necessary trade-offs that are involved in maintaining our security. Pressing ahead with obscure and controversial measures, careless of international reaction, without widespread public support and in the teeth of data protection and human rights concerns risks undermining faith in the authorities we rely on to keep us safe and the digital means by which we now routinely communicate. The Home Secretary's decision to impose a TCN on Apple was undoubtedly well-intentioned. However, if the unintended consequences are greater mistrust of the state, a reduction in the willingness of tech companies to provide assistance and US withdrawal from the DAA, then the net effect will be a further diminution of the data available to those who would keep us safe. We will thus all have been mis-served by the TCN, with potentially significant implications for everyone's security.

30. judiciary.house.gov/committee-activity/hearings/foreign-influence-americans-data-through-cloud-act-0

Navigating a new era of reporting cyber incidents in the UK and EU

K. Hagedorn, A. Portnoy and H. Hewitt[1]

Cyber security continues to be an issue that gathers mainstream attention, and for good reason. Both the costs of, and length of time to recover from, a cyber incident are increasing. According to IBM's 2024 'Cost of a Data Breach' report, the average cost of a data breach in 2024 has risen to $4.88 million.[2] For some incidents (particularly those involving stolen or compromised credentials), the recovery period was recorded as being as long as 292 days.

The increasing prevalence of cyber-attacks, and the disruption they can cause, has led to governments globally introducing new legislation and/or supplementing existing legislation to protect the most critical infrastructure, whilst also encouraging information sharing to enhance overall awareness of cyber risks. In several jurisdictions, cyber security requirements mandated by law are now being imposed on new industry sectors not traditionally seen as critical, a reflection of the changing way the world operates.

For some compliance professionals, reporting cyber security incidents to authorities may not be a new concept. However, for many organisations the focus on reporting cyber security incidents, as opposed to the perhaps now established process of reporting personal data breaches, is something that may not be familiar.

This article seeks to map out the changing reporting landscape in both the UK and EU, providing an overview of what compliance professionals need to consider when updating their processes and procedures around cyber incident reporting.

1. Reporting personal data breaches

1.1. Reporting obligations under the EU GDPR and UK GDPR

Since it came into effect in 2018, the EU's General Data Protection Regulation (the "EU GDPR")[3] has imposed a range of obligations on organisations who are involved in processing personal data. Following the UK's departure from the EU in January 2020, the UK government transposed the EU GDPR into local law (the "UK GDPR").[4] As of the date of publication of this article, for the most part the UK GDPR mirrors the EU GDPR.

The EU GDPR and UK GDPR (collectively the "GDPR") apply to:

1. Organisations established in the EU or UK (respectively for each the EU GDPR and UK GDPR), that are processing personal data of any individual (referred to as a data subject).
2. Organisations *not* established in the EU or UK, that are processing personal data of data subjects and are either:
   a. offering goods or services to such data subjects in the EU or UK; or
   b. monitoring the behaviour of such data subjects so far as the behaviour being monitored takes place within the EU or UK.

For organisations to whom the GDPR applies, the GDPR introduced an obligation for organisations to report personal data breaches to regulators, individuals and/or data controllers (as appropriate) (Articles 33 and 34 GDPR). Personal data breaches are often, but not always, associated with a cyber security incident.

To whom an organisation should report a personal data breach under the GDPR will depend on the following factors, each of which are expanded on below:

1. whether a personal data breach has occurred;
2. whether the organisation acts as a data controller or data processor in relation to the personal data affected; and
3. whether the personal data breach is notifiable i.e., whether it meets the statutory thresholds.

(a) What is a personal data breach?

Article 4(12) GDPR defines a personal data breach as: *"a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, personal data transmitted, stored or*

1. Kelly Hagedorn (Partner), Alice Portnoy (Senior Associate) and Hanna Hewitt (Associate) are all attorneys in the Privacy, Cyber & Data Strategy Team of Alston & Bird.
2. Available here: https://www.ibm.com/reports/data-breach.
3. Available here: https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng.
4. Available here: https://www.legislation.gov.uk/eur/2016/679/contents.

otherwise processed". The GDPR does not distinguish between causes of a personal data breach i.e., if a personal data breach has occurred and it meets the thresholds for reporting set out under Articles 33 and 34 GDPR, a regulator and, where applicable, individuals should be notified. As such, notifications under the GDPR could be submitted for a broad range of issues including:

- an email attachment being sent to the incorrect recipient;
- a laptop being left in a public place;
- inadvertently making documents and/or folders accessible to individuals who should not have permission to see the relevant data;
- an unauthorised third-party gaining access to an organisation's systems (which may include deploying ransomware);
- deleting personal data that still needs to be retained meaning it is no longer available; and
- altering personal data meaning it is no longer accurate.

(b) What are the responsibilities of a data processor vs data controller?

By way of reminder, the obligations imposed on an organisation under the GDPR will depend on whether the organisation is considered a data controller or data processor. The GDPR defines a data controller and data processor as follows:
Controller *"means the natural or legal person, public authority, agency, or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data"* (Article 4(7) GDPR).
Processor *"means a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller"* (Article 4(8) GDPR).

A data processor is only required to notify a data controller of a personal data breach (i.e., a data processor is not responsible for notifying a personal data breach to a competent regulator and where required affected individuals). As such, the first step any data processor should take is to determine whether an incident has resulted in a personal data breach, as defined above. A data processor must subsequently determine whether the compromised personal data was processed on behalf of the data controller. If so, the data processor must notify the data controller of the personal data breach *"without undue delay"* (Article 33(2) GDPR). In addition, a data processor must *"assist the controller in ensuring compliance"* with its obligations under Articles 33 and 34 (amongst others) i.e., ensuring that a data controller has the information needed to notify regulators and as required, affected individuals, within the applicable deadlines (Article 28(3)(f) GDPR). Data processors should always review their contracts with their data controllers (a data processing agreement or otherwise) for provisions relating to personal data breach reporting. In particular, it is common to see a contractual timeframe stipulated for reporting an incident to a data controller, rather than the *"without undue delay"* timeframe specified by the GDPR.

A data controller is responsible for making notifications to a regulator and, where required, affected individuals as appropriate. This includes determining whether a personal data breach has occurred and is notifiable to the regulator and individuals.

(c) Is a personal data breach notifiable to a regulator and individuals?

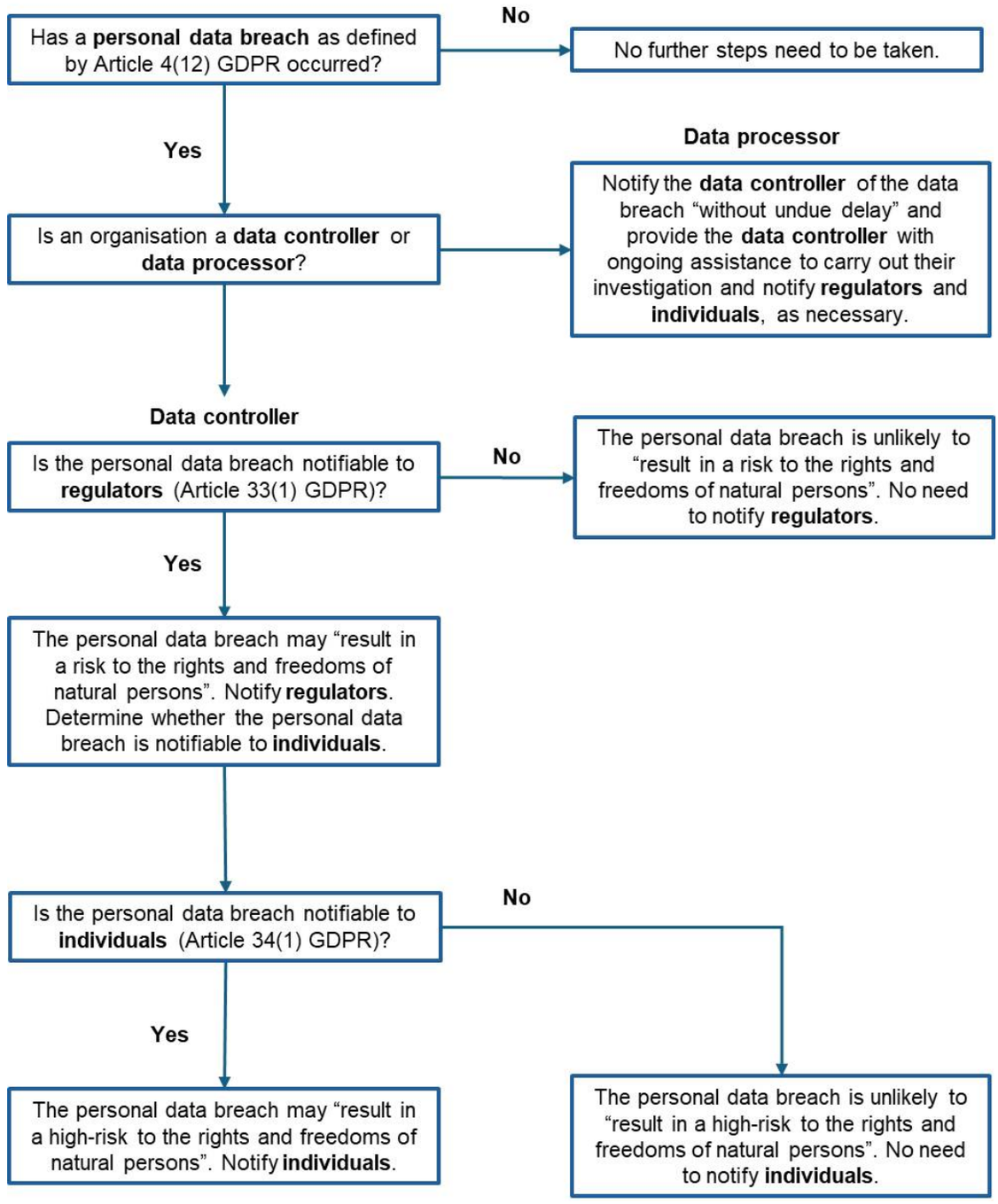
When determining to whom a personal data breach should be reported, organisations must assess whether the personal data breach poses a risk to the rights and freedoms of data subjects. Whilst the reporting threshold for notifications to both regulators and affected individuals uses the same metric of rights and freedoms of data subjects, the thresholds are slightly different, as set out in detail below.

A data controller must notify a regulator *"unless the personal data breach is unlikely to result in a risk to the rights and freedoms of natural persons"* (Article 33(1) GDPR) (emphasis added). As such, the default position is that a personal data breach should be reported to a regulator unless there is no risk to the individual. The GDPR states that a cyber security incident which compromises personal data and that could *"result in physical, material or non-material damage"* to individuals would constitute a personal data breach (Recital 85 GDPR). This includes for example, loss of control of personal data, limitation of an individual's rights in relation to their personal data, discrimination, identity theft of fraud, financial loss, unauthorised reversal of pseudonymisation, damage to reputation, loss of confidentiality or personal data protected by professional secrecy or any other economic or social disadvantage to the individual.

Individual notifications must be submitted if a personal data breach is likely to *"result in a high risk to the rights and freedoms of natural persons"* (Article 34(1) GDPR) (emphasis added). As such, the threshold for reporting a personal data breach to the regulator is lower than that of the threshold of reporting a personal data breach to an individual.

If the relevant thresholds as set out above are met, a data controller must notify a regulator within 72 hours of becoming aware of the personal data breach (Article 33(1) GDPR), and individuals *"without undue delay"* (Article 34(1) GDPR). In practice, individual notifications may take some time, particularly after a cyber incident where data may need to be reviewed to determine affected individuals.

High-level overview of reporting obligations under the GDPR



1.2. What is the position internationally?

When the GDPR became effective in 2018, these reporting obligations were novel. However, many jurisdictions internationally have followed the approach of the GDPR and now also require organisations to report personal data breaches to the regulator and/or individuals. As such, many organisations are now familiar with these reporting obligations and factor them into their incident response plans and internal escalation protocols, making sure that legal teams are involved in discussions early on to meet all reporting timelines.

Whilst the approach to reporting personal data breaches is similar, the deadlines for reporting to both regulators and individuals can vary significantly between jurisdictions. As such, organisations should always take care to ensure that they are familiar with reporting timelines in jurisdictions that they operate in.

2. Reporting cyber security incidents

2.1. Network and infomation Security directives

In 2018, the EU's Network and Information Security Directive ("NIS") became effective, applicable to organisations within specified sectors, to ensure a higher level of cyber security readiness in the most critical sectors. To supplement NIS, in 2023 the Network and Information Security 2 Directive ("NIS2") entered into force.[5] NIS2 replaces NIS and in doing so expands the scope of the original NIS Directive.

Both NIS and NIS2 are Directives, meaning that each EU Member State was required to transpose NIS and NIS2 into local law. EU Member States have already implemented NIS into local law, however the process of amending such legislation to cover NIS2 is ongoing (the process should have been completed in all EU Member States by October 2024; however, at the time of publication of this article, most Member States are still yet to implement NIS2). Given the jurisdictional variation caused by local implementation, organisations need to ensure that they review local NIS2 legislation in jurisdictions where they operate to ensure that any variation is captured, including regarding incident response planning. For the purpose of this article, we refer to reporting requirements set out under NIS2 itself, rather than requirements included by any EU Member State implementing legislation.

What is the position in the UK?

The UK implemented NIS through the Network and Information Systems Regulation 2018.[6] Following the UK's departure from the EU, it is no longer required to implement EU Directives into national law and therefore NIS2 is not applicable in the UK. However, in July 2024, the UK government announced that it would introduce the Cyber Security and Resilience Bill (the "Bill") to Parliament to address cyber security challenges faced by the UK.[7] This Bill – unpublished at the time of writing – is likely to align with the approach that the EU has taken with NIS2.

NIS2 requires 'covered entities' (see below) to report qualifying cyber incidents to the competent authority or the Computer Security Incident Response Team ("CSIRT") in the relevant jurisdiction. The CSIRT is a new concept under NIS2 (Article 10 NIS2) and they are responsible for governmental handling of incidents. EU Member States have the autonomy to establish the CSIRT as a new regulator, or as a branch within an existing regulator. As such, the regulator to which an organisation should report qualifying cyber incidents will depend on the applicability of implementing NIS2 legislation of an EU Member State to the organisation's activities.

Whether an organisation needs to report a cyber incident under NIS2 will depend on the following factors, each of which are expanded on below:

1. whether an organisation operates in a sector that falls within the scope of NIS2; and
2. whether a cyber incident is reportable under NIS2.

(a) Organisations within scope of NIS2

NIS2 seeks to increase the number of organisations that fall within scope of the legislation. Annex I to the legislation contains a list of high criticality sectors, and Annex II contains a list of other critical sectors. Some of these sectors were introduced under NIS, but others have been added under NIS2, as shown in the table below.

	NIS[8]	**NIS2**
Annex I (Highly critical sectors)	- Energy - Transport - Banking - Financial markets - Health - Drinking water supply and distribution - Digital infrastructure - Cloud computing	- ICT service management (business-to-business) - Public administration - Space

5. Available here: https://eur-lex.europa.eu/eli/dir/2022/2555/2022-12-27/eng.

Annex II (Other critical sectors)	- E-commerce - Search engine services	- Wastewater - Postal and courier services - Waste management - Manufacture, production and distribution of chemicals - Manufacture, production and distribution of food - Manufacturers of medical devices - Manufacturers of computer, electronic and optical products - Manufacturers of electrical equipment - Manufacturers of machinery and equipment - Manufacturers of motor vehicles, trailers and semi-trailers - Manufacturers of other transport equipment - Social networking services - Research

NIS2 categorises organisations that fall within these sectors as 'covered entities', further categorising these entities as 'essential' entities and 'important' entities (as defined below). For the purpose of reporting cyber incidents, the distinction is not material, however it is important to note that some obligations under the legislation do not apply uniformly to 'essential' and 'important' entities.

Essential entities are entities that fall within any of the highly critical sectors contained in Annex I of NIS and NIS2 (listed in the table above), and that exceed the 'medium-sized enterprises' ceilings under EU standards.

Essential entities will also be those entities that:

- provide public electronic communications networks or publicly available electronic communication services;
- provide trust services;
- provide top-level domain name registries and domain name system service providers;
- are a public administration entity of central government or at a regional level (as defined by EU Member States in accordance with national law);
- are the sole provider in an EU Member State of a service which is essential for the maintenance of critical societal or economic activities;
- provide a service, which if disrupted, could:
  - have a significant impact on public safety, security or public health;
  - induce a significant systemic risk, particularly in sectors where disruption could have a cross-border impact;
- are critical because of its importance at a national or regional level for the particular sector or type of service; or
- are entities that specific Member States identify as being essential when implementing NIS2 locally.

Important entities are entities that fall within any of the sectors listed in the table above, which do not qualify as essential entities.

(b) Cyber incident reporting under NIS2

NIS2 requires essential and important entities to report *"significant incidents"* (Article 23(1) NIS2). Under Article 23(3) NIS2, an incident will be considered "significant" if it has:

- caused/can cause severe operational disruption of the services (or financial loss for the entity concerned); or
- It has affected/can affect other natural or legal persons by causing considerable material or non-material damage.

To determine whether an incident will be *"significant"* an organisation should consider the importance of the affected network and information systems in the provision of the organisation's services, the severity and technical characterises of a cyber threat, underlying vulnerabilities that are being exploited and the organisation's experience with similar incidents (Recital 101 NIS2).

If an incident is *"significant"* an entity will need to report it to the competent authority or the CSIRT, as applicable.

8. These sectors are still in scope of NIS2.

If the relevant reporting threshold is met (as set out above), an organisation must adhere to the following notification timeline:

Notification type	**Notification timeline**	**Contents**
Early Warning (Article 23(4)(a) NIS2)	Within 24 hours	The Early Warning should indicate whether it is believed that the *"significant incident"*: – is caused by unlawful or malicious acts; or – could have a cross-border impact.
Incident Notification (Article 23(4)(b) NIS2)	Within 72 hours	The Incident Notification should: – update the information contained in the Early Warning notification; – provide an initial assessment of the *"significant incident"* including its severity and impact; and – where available, provide the indicators of compromise.
Final Report (Article 23(4)(d) NIS2)	Not later than 1 month after submission of the Incident Notification	The Final Report should include: – a detailed description of the incident, including severity and impact; – the type of threat or root cause; – any mitigation measures that have already been applied, as well as any that are ongoing; and – if applicable, the cross-border impact of the incident.
Progress Report (Article 23(4)(e) NIS2)	If an event is ongoing, no later than 1 month after submission of the Incident Notification	Following submission of the Progress Report, an organisation must then submit a Final Report within 1 month of finishing handling an incident.

A competent authority, or the CIRST (as applicable), can also request that an organisation provides an immediate report and/or status updates (Article 23(4)(c) NIS2). Given NIS2 is a new piece of legislation, it is not yet clear how the competent authority or the CIRST (as applicable) will seek to utilise this power.

2.2. Digital Operational and Resilience Act

In January 2025, the requirements of the EU's Digital Operational and Resilience Act ("DORA")[9] became binding. DORA is a Regulation and therefore is directly applicable in all EU Member States. The purpose of the legislation is to strengthen the resilience of financial entities (including their information and communication technology (ICT) supply chains) to ensure they can withstand, respond to and recover from IT disruptions including cyber incidents.

What is the position in the UK?
The UK has not implemented DORA, and as set out above, following the UK's departure from the EU, it is no longer required to implement EU Regulations into national law. However, financial services and critical third parties must adhere to other obligations as detailed below:

9. Available here: https://eur-lex.europa.eu/eli/reg/2022/2554/oj/eng.

1. ***Regulated financial services*** must notify the Financial Conduct Authority of any material cyber incidents. A material cyber incident is one that: (a) results in significant loss of data (or the availability or control of IT systems); (b) affects many customers; or (c) involves unauthorised access to IT systems (including the use of malicious software to infect IT systems).
2. ***Critical third parties*** must adhere to rules and guidance published by UK regulators of the financial services sector. In 2023, the UK government gave regulators of the financial services sector[10] powers to ensure that critical third parties manage potential risks to the stability of the UK financial sector, that could be caused by a failure in, or disruption to, the services provided by critical third parties. In November 2024, these regulators published a supervisory statement (SS6/24) ('Operational Resilience: critical third parties to the UK financial sector')[11], which sets out: (a) the regulators' expectations for how critical third parties should comply with duties and obligations under the Financial Services and Markets Act 2000 (as amended by the Financial Services and Markets Act 2023); and (b) the rules applicable critical third parties. The rules came into effect on 1 January 2025 and have been designed to closely align with other international standards including DORA.

Whether an organisation should report a cyber incident under DORA will depend on the following factors, each of which are expanded on below:

1. whether an organisation is a 'financial entity' or an 'ICT third-party service provider' (terms that are defined under EU law); and
2. whether a cyber incident is reportable under DORA.

(a) Organisations within scope of DORA

DORA applies to a group of entities collectively referred to as 'financial entities' (Article 2(2) DORA) and 'ICT third-party service providers'.

'Financial entities' include (Article 2(1) DORA):

- Credit institutions
- Payment institutions
- Account information service providers
- Electronic money institutions
- Investment firms
- Crypto-asset service providers and issuers of asset-referenced tokens
- Central securities depositories
- Central counterparties
- Trading venues
- Trade repositories
- Managers of alternative investment funds
- Management companies
- Data reporting service providers
- Insurance and reinsurance undertakings
- Insurance intermediaries, reinsurance intermediaries and ancillary insurance intermediaries
- Institutions for occupational retirement provision
- Credit rating agencies
- Administrators of critical benchmarks
- Crowdfunding service providers
- Securitisation repositories

'ICT third-party service providers' are any undertakings that provide ICT services to 'financial entities' (Article 3(19) DORA). ICT services are "*digital and data services provided through ICT systems to one or more internal or external users on an ongoing basis, including hardware as a service and hardware services, which includes the provision of technical support via software of firmware updates by the hardware provider, excluding traditional analogue telephone services*" (Article 3(21) DORA). The definition of ICT services is broad, and is intended to cover common ICT services e.g., product management services, software licensing, help desk support etc.

(b) Cyber incident reporting under DORA

Financial entities must classify ICT-related incidents and cyber threats using the criteria set out in Article 18(1) DORA. An ICT-related incident is a single unplanned event, or a series of unplanned linked events, that compromise the security of network and information systems, and which have an adverse impact on either: (a) the availability, authenticity, integrity or confidentiality of data; or (b) the services provided by the financial entity (Article 10(8) DORA).

If a financial entity classifies an ICT-related incident as a 'major ICT-related incident', it must report the incident to its relevant competent authority (Article 19(1) DORA). A 'major ICT-related incident' is an ICT-related incident that has a high adverse impact on network and information systems that support critical or important functions of the financial entity (Article 3(10) DORA). Each 'financial entity' will have a different competent authority to report to as set out under Article 46 DORA.

Financial entities may decide to outsource, in accordance with local law, their reporting obligations for major ICT-related incidents to specialised third-party service providers (Article 19(5) DORA). If financial entities choose to do this, they should be aware that they remain fully responsible for the fulfilment of incident reporting requirements. Note that ICT third-party service providers are also required, under DORA, to assist a financial entity following an ICT-related incident.[12]

12. For example, 'ICT third-party service providers' are required to assist a 'financial entity' at no additional cost, when an ICT incident that is related to the ICT service provided to the 'financial entity' occurs (Article 30(2)(f)

If an incident meets the major incident threshold (as set out above), a financial entity must adhere to the following notification timeline:

Notification type	**Notification timeline**	**Contents**
Initial Notification (Article 19(4)(a) DORA)	Within 4 hours of determining an incident is a major ICT-related incident, and in any event, within 24 hours of detecting the major ICT-related incident	EU DORA regulators are expected to develop common draft regulatory technical standards that will provide guidance on how to submit Initial Notifications.[13]
Immediate Report (Article 19(4)(b) DORA)	Within 72 hours of determining an incident is a major ICT-related incident	An Immediate Report should be submitted as soon as the status of the original incident has changed. Separate relevant status updates should be shared when new information is available.
Final Report (Article 19(4)(c) DORA	No later than one month from determining an incident is a major ICT-related incident	A Final Report should be shared when the root cause analysis has been completed, regardless of whether mitigation measures have been implemented.

Additionally, a financial entity must notify its clients of a major ICT-related incident if it has an impact on clients' financial interests. Such notification must be sent without undue delay and as soon as the financial entity becomes aware of the major ICT-related incident and must detail the measures taken by the financial entity to mitigate the adverse effects of the incident (Article 19(3) DORA).

Financial entities may also report 'significant cyber threats' on a voluntary basis and should also have communication plans in place to communicate major incidents to the public where appropriate.

2.3. EU Cyber Resilience Act

The EU's Cyber Resilience Act ("CRA")[14] came into force on 10 December 2024 and sets out cybersecurity requirements for organisations that design and sell 'connected devices' (see below) in the EU. Most provisions of the CRA will apply from December 2027.

Under the CRA 'connected devices' are products with digital elements ("PDEs") that are intended to have a direct or indirect logical or physical connection to a device or network (Article 2(1) CRA). The CRA classifies PDEs into different categories according to their level of cyber security risk:

PDE category under the CRA	**Examples of PDE**	**Level of Cybersecurity Risk**
Basic / non-critical	Smart TVs, home surveillance cameras, connected toys, video games	Low
Important	Biometric readers, smart home virtual assistants, personal wearable devices, firewalls, intrusion detection and prevention systems	High (the use of the PDE can impact individuals' health, security or safety)
Critical	Smart cards, smart meters	Critical (the use of the PDR can disrupt, control, or cause damage to other PDEs through direct manipulation)

DORA). 'ICT third-party service providers' are also required to fully cooperate with the competent authorities and the resolution authorities of the 'financial entity' (Article 30(2)(g) DORA).

13. Please see Annex I of the European Supervisory Authorities Final Report dated 17 July 2024, which addresses, amongst other issues, the draft regulatory technical standards on the standard forms, templates and procedures for financial entities to report a major incident and to notify a significant cyber threat.
14. Available here: https://eur-lex.europa.eu/eli/reg/2024/2847/oj/eng.

The obligations of the CRA are primarily aimed at PDE manufacturers, but the CRA also imposes obligations on EU-based importers that place PDEs on the EU market and on distributors that supply PDEs to users in the EU.

With regards to cyber incident reporting, beginning on 11 September 2026, PDE manufacturers are required to notify competent regulators of:

- actively exploited vulnerabilities affecting their PDEs; and
- severe incidents that have an impact on the security of their PDEs.

Where a manufacturer is required to report an incident, the following notifications must be submitted:

Notification type	**Notification timeline**	**Contents**
Early Warning Notification (Article 14(2)(a) CRA and Article 14(4)(a) CRA	Within 24 hours of becoming aware of the vulnerability or incident	The Early Warning Notification should indicate where in the EU the PDE was made available and, for incidents, whether the incident is suspected of being caused by unlawful or malicious acts
Incident / Vulnerability Notification (Article 14(2)(b) CRA and Article 14(4)(b) CRA)	Within 72 hours of becoming aware of the vulnerability or incident	For vulnerabilities: – General information on the PDE, the general nature of the exploitation and vulnerability, and any corrective or mitigating measures taken or that users can take to remedy the vulnerability. For incidents: – General information about the nature of the incident, and initial assessment of the incident, and any corrective or mitigating measures taken by the manufacturer or that users can take
Final Report (Article 14(2)(c) CRA and Article 14(4)(c) CRA)	For vulnerabilities: - No later than 14 days after a corrective or mitigating measure is available to remedy the vulnerability. For incidents: - Within one month after submission of the Incident Notification.	For vulnerabilities: – A detailed description of the vulnerability (including its severity and impact), information about the malicious actor that exploited or is exploiting the vulnerability, and details about the security update or other corrective measures made to remedy the vulnerability. For incidents: – A detailed description of the incident (including severity and impact), the type of threat or root cause that is likely to have triggered the incident, and applied / ongoing mitigation measures

A manufacturer will also be required to notify *"any severe incident"* to the CSIRT (Article 14(3) CRA). Under Article 14(5) CRA a *"severe incident"* is one that:

- negatively affects/is capable of negatively affecting the ability of a PDE to protect the availability, authenticity, integrity or confidentiality of sensitive or important data or functions; or
- has led/is capable of leading to the introduction or execution of malicious code in a PDE, or in the network and information systems of a user of the PDE.

3. What do organisations need to consider in light of the cyber security reporting requierements?

As the regulatory position in the UK and EU has demonstrated, governments are increasingly finding new ways to encourage organisations to strengthen their cyber resilience, especially in the financial services sector and other sectors deemed to be critical. Whilst the UK and EU are currently leading the way in this area, jurisdictions globally are mirroring this position.

The requirement to report personal data breaches (which are often, but not always, associated with a cyber security incident) is here to stay. What organisations now need to prepare for, is the likelihood that reporting of cyber security incidents will also become a standard requirement, both under the cyber security legislation cited above and other incoming legislation such as the EU AI Act.[15] As such, it is important for organisations operating in critical sectors to be aware of their obligations and accordingly update their processes and procedures around cyber incident reporting. This includes:

- reviewing incident response plans to ensure applicable legislation is referred to, particularly that reporting timelines and applicable regulators are correctly identified;
- familiarising those involved in incident response with additional reporting obligations and timelines, to ensure that all teams working on an incident are aware of the stricter reporting timelines imposed by new cyber security legislation (for example by way of table-top exercises with IT, Security, Legal and Executive teams); and
- as necessary, updating contracts with third-party suppliers and/or vendors to ensure that appropriate contractual provisions are put in place to strengthen cyber security posture.

If there is a possibility that new cyber security legislation applies to your organisation, it is always best practice to seek legal advice to determine what obligations your organisation is subject to (both with regards to incident reporting requirements and other cyber security obligations).

15. The EU AI Act entered into force in August 2024, but the requirement to report series incidents will become applicable from August 2026.

Public Data, Private Risks

How LLMs Might Reshape Compliance Investigations

C. Whomsley[1]

Since the explosion in use of generative AI tools in 2023, and more specifically, in the use of chatbots powered by Large Language Models ("LLMs"), corporate investigators and intelligence specialists have been testing their potential application in open-source investigations. This has given rise to questions surrounding the utility of these tools, particularly in retrieving public information which would otherwise be difficult to access, as well as questions surrounding the privacy implications of their use in the context of digital investigations.

Even the most popular chatbots, used by millions of people globally, have proven worryingly invasive in the information they provide on demand to their users. Despite efforts to minimise the accessibility of private information on their platforms and assurances of safeguards preventing their misuse to find out sensitive information, it doesn't take long to 'jailbreak' or corrupt the system. Even with fair use, through relatively straightforward prompts, these chatbots can divulge sensitive data such as personal addresses. These can be returned when found through URLs mined either as part of the chatbot's training data or identified through the chatbot's direct searches of the internet to provide responses to specific prompts (known as retrieval augmented generation or RAG).

Although this access to sensitive but public data could be considered invasive and problematic, AI chatbots could, as a result, revolutionise the way corporate investigations are carried out, particularly in the open-source intelligence ("OSINT") space.

1. The Rise of OSINT

In recent years, corporate investigations have become increasingly reliant on intelligence derived from public information found in open sources.

Thanks to the shift towards the digitisation of official records by governments, from corporate filings to legal judgments, as well as the expansion of media reporting online and the emergence of platforms providing access to print records, the in-person retrieval of information needed to carry out standard due diligence investigations is no longer required in many jurisdictions.

Beyond official records, the online world also provides more significant insight into individuals and businesses than has ever been possible from Human-Source Intelligence ("HUMINT"), largely through the world of Social Media Intelligence ("SOCMINT"). Prior to the advent of OSINT tools, traditional research on these platforms could already provide worrying and unprecedented insights into a person's lifestyle, background and source of wealth, for example, identifying their properties through their activity on exercise mapping app Strava, or their children's schools from their social media following lists.

Through the rise of contact detail scraping tools, relying on information unwittingly provided by individuals registering online accounts to data brokers, it is also now possible to identify links between accounts across not just social media, but with online services and apps such as Deliveroo, Vivino and Spotify, through the email addresses and phone numbers used to register them. This has meant that accounts believed by their operators to be unfindable can easily be identified and examined, even when created under an alias.

The proliferation of AI tools has led to the emergence of even more invasive research methods. It is now possible to run facial recognition searches on individuals, returning links to hidden social media profiles and appearances in photos on webpages in which they are not named.

This new access to information on people, their assets and their lives has meant that, for better or worse, the digital world provides the most expansive range of insights into individuals and businesses. OSINT investigations have, as a result, become the go-to approach to support most corporate investigations.

OSINT can, for example, support in the establishment and verification of a person's identity and any risks which could be associated with them, whether that may be from their source of wealth, key associations, background, or any reputational or legal issues they may have encountered. It can also support in the establishment of a company's ultimate benefi-

1. Carys Whomsley is a Managing Director and Head of Digital Risk at Digitalis, a London-based advisory firm who help clients protect and advance their online interests.

cial ownership, performance and reputation, as well as any non-disclosed issues lurking under the surface of an otherwise positive profile.

It therefore follows that compliance due diligence investigators now heavily rely on OSINT to ensure adherence to global Anti-Money Laundering ("AML") and Know Your Client ("KYC") regulations, particularly across the EU, UK and North America, where such a vast amount of data on individuals and businesses is readily available, and often free.

2. The Privacy Question

Enter Large Language Models. Following the release of OpenAI's ChatGPT, the use and performance of LLMs has skyrocketed. Initially drawing their knowledge from a static set of material constituting their training data, the most widely used models are now capable of 'reasoning' and accessing the internet directly to provide responses. The breadth of sources consulted in their training data, coupled with their new RAG capabilities, presents huge promise in the potential for LLMs to support in OSINT investigations.

In theory, the models could rely on their expansive training data as well as their instant access to up-to-date online information to provide actionable intelligence on the subjects of investigations quickly and at scale. Through their reasoning capabilities, they could also enable the identification of well-concealed red flags and establish associations which may have otherwise been difficult to find. They are also capable of analysing huge volumes of data at pace and could considerably speed up the process of carrying out standard red flag checks into individuals and companies, as well as support in carrying out more in-depth analysis into open-source material.

The unprecedented access to online information through these chatbots could lead to the creation of powerful tools to support compliance investigations. However, their use in these could give rise to significant concerns around privacy. OSINT research is often presented as the ethical alternative to HUMINT as it exclusively relies on publicly accessible information which, in theory, can be controlled by the subject examined. That said, a huge portion of any given individual's digital footprint has been unwittingly made available, and the extent of the detail available is rarely known to them.

In addition to the information unknowingly made public through, for example, social media posts mistakenly believed to be private, historic online accounts which have since been forgotten, and information accessible in official public records, sensitive information on individuals can also be accessed through paywalled, third-party data aggregating websites sourcing the information they provide to researchers from consented information.

Even in the age of the GDPR, without the right knowhow and tools, most people have little knowledge of the breadth of information about them accessible to investigators, and nor would they be able to find it for themselves. After the crackdown on personal data harvesting which followed the Cambridge Analytica scandal in 2018, could LLMs be in for a similar reckoning if the extent of their knowledge base was exposed?

Chatbots are also known to present significant privacy risks to their users, through their ability to collect and process user data. There have been, for the past few years, several warnings that any data provided to the key chatbots may enter their training dataset and be used to inform responses to queries by other users further down the line. Chatbots capture and preserve every prompt entered, as well as any information they can capture from a user's device and browser session. The release of DeepSeek, a Chinese model which overtook ChatGPT as the most downloaded free app at the beginning of this year, shone a light on the privacy implications of using chatbots to carry out research - with concerns that the chatbot's developers may be compelled to share any user data with the PRC's government upon request.

More worryingly still, the access to sensitive user data and integration of chatbots into critical systems makes them an attractive target for cybercriminals. A successful attack could leave sensitive data into the subjects of an investigation exposed, which could constitute a violation of privacy regulations.

3. Barriers to Reliability and Ethical Use

Aside from the privacy considerations, even with the unprecedented access LLMs can provide, it is so far unclear whether any tools which leverage them could prove more reliable than human investigators. At this stage in their development, there are seven key limitations restricting the usefulness of LLMs in carrying out compliance investigations.

3.1. Hallucinations and Accuracy

At the moment, the main issue reducing the reliability of LLMs in investigations is their propensity to "hallucinate", that is, to generate inaccurate or misleading answers to a user's prompt. Due to the authoritative tone with which chatbots present their responses, these can often go unnoticed.

When first recognised as a phenomenon a few years ago, such hallucinations were expected to disappear over time as chatbots improved in sophistication. However, AI leaderboard company Vectara and OpenAI have observed significantly higher hallucination rates in the releases of newer chatbots. GPT o4-

Mini, for example, was found to hallucinate 48 per cent[2] of the time when summarising publicly available facts about people.

If the hallucination problem does not get resolved, the application of LLMs to the world of investigations could be permanently derailed, as the need for constant human fact-checking would be more time intensive (and costly) than manual research and analysis.

3.2. Access Powers and Copyright

Despite the aggressive data scraping continually carried out to widen the LLMs' training data and improve both their knowledge and performance, they have been repeatedly held back by copyright disputes - from complaints over their presentation of content from paywalled sources[3] to criticisms over their creators' use of copyrighted lyrics[4] to train their models.

Most of the leading chatbot developers also use social media posts as part of their LLMs' training data, including text and images, considering these to be fair game as they are publicly available. This practice already attracted significant controversy following the announcement of Meta's change in its privacy policy in late 2024 to allow for social media post scraping, which could ultimately force the rollback of their access to these.

It is too early, at this stage, to understand what the impact of these early cases will be on the content returned in responses, and whether they will later face further restrictions in presenting content from these sources.

Despite their existing sweeping access to billions of pieces of content which could prove useful in carrying out investigations, their powers could therefore, paradoxically, retrieve fewer relevant results than could be found through traditional OSINT.

3.3. Bias

Chatbots are known to have fallen prey to the biases present[5] in their training data, which is largely based on content created by humans. These manifest in gender, racial, cultural and other biases, ingrained in the models despite their developers' best efforts to reduce their influence. Their optimisation for a positive user experience - to the point of sycophancy[6] - also makes them more likely to reinforce bias than to challenge it.

There is also a huge variation in the chatbots' capabilities across different languages and jurisdictions. While it's difficult to pinpoint an exact volume, training datasets for the most widely used LLMs in Europe are likely to have been based primarily on English-language content and sources, showing demonstrably lower capabilities[7] in other languages.

These limitations are likely to translate to reduced capabilities in accessing, interpreting and presenting data from non-English speaking sources in an objective manner, restricting the usefulness of these LLMs in global compliance investigations and with potentially devastating consequences for the individuals against whom they may have inherent biases.

3.4. Sourcing and 'Laziness'

Finally, the prioritisation of the sources consulted by LLMs to form their responses to specific prompts is also unclear, with the companies who released the chatbots reluctant to divulge any information on this.

At Digitalis, we have been carrying out research to seek to determine which types of sources[8] appear most frequently in chatbots responses since late 2023. Our research into a number of well-known/public individuals and companies has found an interesting trend. The leading chatbots appear to source much of the content of their responses from the first five URLs appearing in results for the corresponding search engines.

In terms of source type, those that appear the most frequently in chatbot responses are owned pages belonging to the subject of the research (for example, a person's LinkedIn profile or a company website), and Wikipedia - a source appearing in responses in almost every single instance in which a Wikipedia article exists for the subject of the chatbot query.

Although the reason for these preferences cannot be determined, it is possible that the reliance on material in a free-content online encyclopaedia and in owned assets could have been prioritised in response to the copyright and defamation controversies which have plagued the AI companies since the release of their chatbots.

Despite the significant breadth of sources which have been used in the training material underpin-

2. www.newscientist.com/article/2479545-ai-hallucinations-are-getting-worse-and-theyre-here-to-stay/
3. hls.harvard.edu/today/does-chatgpt-violate-new-york-times-copyrights/
4. thebarristergroup.co.uk/blog/ai-generated-content-and-copyright-evolving-legal-boundaries-in-english-law
5. www.winchester.ac.uk/News-and-Events/Press-Centre/Media-Articles/New-study-shows-AI-chatbots-reflect-human-biases-and-focus-on-threat-negativity-and-gossip.php
6. digitalis.com/news/fawning-chatbots-and-problematic-pattern-recognition/
7. www.nature.com/articles/d41586-024-02579-z
8. digitalis.com/news/how-chatbots-source-their-info-key-findings-from-our-study-revealed/

ning the LLMs, it therefore appears that chatbots are becoming 'lazy' and overly cautious in the insights they provide, hindering their usefulness in establishing red flags and key risks associated with the subjects of an investigation.

4. A Cautious Path Forward

In their current form, LLMs offer real potential for revolutionising compliance investigations, particularly in the OSINT sphere, but remain hindered by critical flaws. While their ability to access and process vast amounts of publicly available data at speed presents clear advantages, the risks associated with hallucinations, bias, opaque sourcing practices, and significant privacy concerns mean they cannot yet replace the discernment and accuracy of skilled human investigators.

As legal and ethical scrutiny around AI deepens, particularly with regard to personal data access and usage, LLMs will likely face increasing pressure to evolve in transparency, accountability, and performance. Until these models can consistently produce verifiable, unbiased, and ethically sound outputs, their role in compliance must be carefully introduced.

Privacy vs. Whistleblowing: Can Data Breaches Be Justified During Public Disclosure?

A.A. Avramenko[1]

Whistleblowing has recently been at the forefront of the public consciousness increasingly often. Wikileaks, Theranos, and Cambridge Analytics have become household names due to the efforts of internal whistleblowers. However, the history of whistleblowing stretches thousands of years into the past. Its first recorded instances occurred in Ancient Greece, where it was referred to as *parrhesia,* meaning "speak freely". Although it was not formally codified, *parrhesia* was extremely valued and considered a cornerstone of democratic society.[2] Today, the role of whistleblowing is not less significant. Within the European Union, there has been an increased focus on protecting those who *blow the whistle*, which resulted in the introduction of the Directive (EU) 2019/1937, commonly known as the Whistleblowing Directive[3]. While its main purpose is to "enhance the enforcement of Union law and policies in specific areas"[4], it also seeks to establish common standards for the protection of the whistleblowers by providing protection in cases of justified internal and external reporting as well as public disclosure. Simultaneously, the EU upholds one of the strictest data privacy regimes through the General Data Protection Regulation (GDPR)[5] by imposing strict obligations on the processing of personal data, with significant penalties for non-compliance. The co-existence of these two strict regimes introduces an inherent tension: *blowing the whistle* can easily involve the dissemination of personal data, potentially constituting a data breach under GDPR. This is especially relevant in the case of public disclosure[6], when the information reported becomes widely accessible. This presents not only a complex legal dilemma but an ethical discussion, creating a confrontation between two deeply valued principles: the pursuit of justice, accountability, and protection of the public interest through whistleblowing, and the safeguarding of individual privacy. This article seeks to examine the situations when a data breach, occurring as part of a public disclosure by a whistleblower, might be justified within the EU legal framework. It analyses the specific conditions for public disclosure, and examines whistleblower liability and its exemptions. The article also provides insights on navigating these complex legal and ethical dilemmas and analyses the extent to which organisations can hold whistleblowers accountable for such data breaches, considering the protections afforded by the Whistleblowing Directive and the case law of the European Court of Human Rights (ECtHR).

1. EU Legal Framework: Walking the Tightrope Between Disclosure and Data Privacy

The EU has established comprehensive legal frameworks to govern both the protection of whistleblowers and the safeguarding of personal data. These frameworks, the Whistleblowing Directive and the GDPR, create a complex interplay when a whistleblower's disclosure involves personal data. Understanding their core objectives, scopes, and points of intersection is crucial to analysing the justification of data breaches in public disclosures.

1.1. The Whistleblowing Directive

The Directive aims to enhance the enforcement of EU law by establishing common minimum standards for the protection of persons who report breaches of EU law. Its scope covers various sectors including public procurement, financial services, product safety, environmental protection, public health, consumer protection. The Directive also provides protection to whistleblowers who report breaches relating to the protection of personal data, which highlights the importance of data protection within the EU whistleblowing framework. Despite the limited material scope of the Directive, the EU Commission has been consistently encour-

1. Anastasia Avramenko is a specialist in Ethics & Compliance with a focus on whistleblowing and legal research.
2. Foxley, I. (2021). An ancient virtue. Index on Censorship, 50(2), 54.https://doi.org/10.1177/03064220211033791
3. Directive (EU) 2019/1937 of the European Parliament and of the Council of 23 October 2019 on the protection of persons who report breaches of Union law.
4. Article 1 of the Whistleblowing Directive
5. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)
6. WB directive 'public disclosure' or 'to publicly disclose' means the making of information on breaches available in the public domain.

aging Member states, when transposing the Directive, to consider extending its scope of application to other areas, and more generally to ensure a comprehensive and coherent framework at national level.[7] The majority of member states broaden the scope of breaches when reporting would qualify for the whistleblower protection; however, *ratione materiae*[8] is still very limited in most jurisdictions. The narrow scope of the Directive is caused by the fact that the Directive's fundamental purpose is not to protect individuals but to strengthen the enforcement of the EU laws by ensuring that breaches are identified and properly reported.

Organisations too are encouraged to go beyond the legislative requirements when establishing internal reporting channels and to broaden the *ratione materiae* when receiving wrongdoing concerns. According to the SUSA benchmark report[9], only 18% of the organisations have not expanded the scope of accepted reports. However, in case of any legal proceedings as a result of reports that do not follow under the Directive or national legislations, whistleblowers may not be granted the protection they expect when raising reports through internal channels.

The Directive also gives a prominent place to the GDPR and the protection of personal data. A significant portion of the Directive is dedicated to the right to the protection of personal data, the confidentiality requirement, and record-keeping.[10] In this context, it is important to note that data protection rules have to be respected not only by organisations, but also by persons acting as whistleblowers.

The Directive establishes a three-tiered system for reporting breaches: internal reporting, external reporting, and, finally, public disclosure as the last resort.
Public disclosure is the most exceptional route for whistleblowers. The protections afforded to whistleblowers who engage in public disclosure are not absolute, and the reporting must meet the strict criteria outlined in Article 15 and Recitals 79, 80 and 81 of the Directive.

1.2. The General Data Protection Regulation

The General Data Protection Regulation (GDPR) establishes a harmonized legal framework for data protection across the European Union. The Regulation applies to all processing of personal data, regardless of the context in which it occurs, including whistleblower disclosures.

At the core of the GDPR are certain foundational principles, outlined in Article 5, including lawfulness, fairness, transparency, purpose limitation, data minimization, integrity, and confidentiality. These principles bind public and private entities, referred to as "controllers", and extend to any natural or legal person who processes personal data. When whistleblowers disclose data through public or media reporting, the processing of personal data can fall outside the controller's formal oversight and may constitute unauthorized data processing.

According to Article 4(12) of the GDPR, a personal data breach is defined as "a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, personal data". A whistleblower's disclosure may fall within this definition if it involves exposing third-party personal data (e.g., employee records, client files, or correspondence).

When such a breach occurs, it activates an immediate set of legal obligations for the organisations involved, regardless of whether they instigated or condoned the whistleblowing. These organisations may face reputational and financial damage due to the whistleblower's actions. The question then becomes whether the organisations can hold the whistleblower accountable or whether their actions are protected by the Whistleblowing Directive.

2. Public Disclosure and Data Breaches: Assessing Justification

When a whistleblower resorts to public disclosure, especially if it involves the dissemination of personal data, the act triggers scrutiny under both the Whistleblowing Directive and the GDPR. The GDPR's foundational principles are frequently at odds with the nature of whistleblowing, which often requires exposing personal or sensitive information to demonstrate the wrongdoing. The justification for such a disclosure, and any consequent data breach, is pursuant to a careful assessment of specific conditions, primarily focusing on the concepts of necessity and proportionality.

2.1. Conditions for Public Disclosure under the Whistleblowing Directive

As previously outlined, Article 15 of the Whistleblowing Directive establishes a high bar for protected public disclosures. A person who makes a public disclosure qualifies for protection if:

– They first reported wrongdoing internally and/or externally, but no appropriate action was taken

7. Commission Communication 'Strengthening whistleblower protection at EU level' of 23.4.2018, COM (2018) 214
8. Legal term meaning "by reason of the matter"; Here refers to the material scope of the Directive - the types of breaches of EU law that whistleblowers can report and be protected for.
9. Vandekerckhove, W. (2024) The Lowlights Report – SUSA 2024 benchmark
10. See namely Recitals 14, 82, 83, 84 & 85 and Articles 16 & 17 of the Directive

in response to the report within the specified timeframes (The appropriateness of the follow-up should be assessed according to objective criteria); and
- They have reasonable grounds to believe that the breach of EU law may constitute an imminent or manifest danger to the public interest; or
- In the case of external reporting, they have reasonable grounds to believe that there is a risk of retaliation or there is a low prospect of the breach being effectively addressed due to the particular circumstances of the case (e.g., evidence may be concealed or destroyed, or an authority may be in collusion with the perpetrator).

Moreover, according to the Recital 91, "it should not be possible to rely on individuals' legal or contractual obligations, such as loyalty clauses in contracts or confidentiality or non-disclosure agreements, so as to preclude reporting, to deny protection or to penalise reporting persons for having reported information on breaches or made a public disclosure where providing the information falling within the scope of such clauses and agreements is necessary for revealing the breach. Where those conditions are met, reporting persons should not incur any kind of liability, be it civil, criminal, administrative or employment-related. Such protection should not extend to superfluous information that the person revealed without having such reasonable grounds".

If these conditions for public disclosure are not satisfied, any ensuing data breach would likely fall outside the protection guaranteed by the Directive, leaving the whistleblower potentially exposed to liability under GDPR and other laws.

Additionally, the centrepiece of whistleblower protection from liability for the disclosure itself is found in Article 21(2) of the Whistleblowing Directive. This provision states that reporting persons "shall not incur liability of any kind in respect of such a report or public disclosure provided that they had reasonable grounds to believe that the reporting or public disclosure of such information was necessary for revealing a breach pursuant to this Directive". Reinforcing this protection, Article 21(7) explicitly addresses legal proceedings that may be brought against whistleblowers, including actions for defamation, breach of copyright, breach of secrecy, breach of data protection rules, or unauthorized disclosure of trade secrets. In such cases, whistleblowers shall not incur liability and may seek dismissal of the case, on the condition that they had reasonable grounds to believe the disclosure was necessary.

The term "necessary" implies that the disclosure was essential to bring the specific breach of EU law to light. This necessity assessment extends to the content of the disclosure, including any personal data. If the wrongdoing could have been effectively revealed without disclosing certain personal data, or by using anonymized or aggregated data, then the disclosure of that specific personal data might not be deemed "necessary." Even though it is not explicitly mentioned in the Directive, this requirement of "necessity" introduces *a de facto* data minimisation consideration for the whistleblower. Data minimization is a core principle of the GDPR (Article 5(1)(c)) which requires that only personal data that are adequate, relevant, and limited to what is necessary to achieve a specific, clearly defined purpose can be collected and processed. While the GDPR's principle of data minimization is an obligation primarily related to data controllers, the "necessity" requirement in the Whistleblowing Directive in effect imposes a comparable reflective duty on the whistleblower making a public disclosure. They must assess whether the inclusion of specific personal details is essential to substantiate their claims before making them public, similar to the data controllers' responsibility of data minimisation.

2.2. Liability Exemption for Whistleblowers

The Whistleblower Directive, specifically Articles 15 and 21, provides immunity to whistleblowers from administrative and civil liability, including under the GDPR. Despite these strong protections, accountability is not entirely eliminated. Where a whistleblower discloses personal data without satisfying the necessity requirement under Article 21(2), or fails to meet the procedural thresholds of Article 15 concerning public disclosure, the act of disclosure itself may be considered an unprotected data breach. In such circumstances, the whistleblower may be directly liable under the GDPR, and the organisation may assert that the whistleblower's actions constitute an unlawful processing of personal data. If the data disclosed was not crucial to substantiate the reported wrongdoing, or if the whistleblower processed the data without a lawful basis attributable to a protected act, this may be viewed as GDPR violation independently of their whistleblowing status.

Moreover, Article 21(3) introduces an important limitation concerning how the information was acquired. According to this provision, whistleblowers do not incur liability for acquiring or accessing the information that is reported or publicly disclosed, as long as the acquisition or access itself did not constitute a "self-standing criminal offence." If the act of data acquisition or access qualifies as a distinct criminal offence, criminal liability is not precluded and remains subject to the applicable legal framework of the relevant Member State.

The term "self-standing criminal offence" refers to actions undertaken by a whistleblower to obtain information that are criminally prohibited, independent of any disclosure. Such acts may include hacking into computer systems, physically stealing documents or electronic devices, bribing individuals to access information, or unlawfully trespass-

ing to gather data. The determination of whether a particular act qualifies as a self-standing offence is grounded in national criminal law, and therefore its interpretation may vary significantly across Member states. Some jurisdictions may treat unauthorized access as an administrative violation, while others may criminalize such actions, potentially resulting in varying degrees of liability for whistleblowers. The Article 21(3) makes clear that the individual is not immune from criminal prosecution for that specific act. This remains true even if the disclosure meets all the criteria necessary for protection under Articles 15, 21(2) and 21(7).

Additionally, the European Commission's report on the transposition of the Whistleblowing Directive identified deficiencies in how member states have addressed liability exemptions. Most member states restricted the scope of this exemption by "excluding facilitators or third persons connected with the reporting person, certain types of legal proceedings or public disclosures."[11]

2.3. The Public Interest vs. Harm from Data Breaches

Even if a public disclosure meets the conditions of Article 15 and the information disclosed is deemed "necessary" under Article 21(2), the conflict with data protection rights necessitates a balancing act based on the proportionality principle - a general principle of EU law that requires any measure restricting rights to be appropriate and no more than necessary. In the case of whistleblowing, the potential harm caused by the data breach (e.g., violation of individuals' privacy rights, reputational damage to those implicated, potential for misuse of data) must be weighed against the public interest served by revealing the wrongdoing.

Whistleblowers making public disclosures usually do not meet the threshold of being lawful controllers or processors, therefore making their actions *prima facie* unlawful under the GDPR. However, the GDPR provides a degree of flexibility in Article 85, obligating Member states to reconcile data protection with freedom of expression and the public interest. The Whistleblowing Directive targets breaches of EU law that are harmful to the public interest. Article 2 and the Annex outlines the Directive's material scope, which includes various areas such as public procurement, financial services, environmental protection, public health, consumer protection, and the protection of privacy, personal data, and the security of network and information systems. This list provides rather clear insight into the concerns the EU legislature considers as public interest important enough to justify whistleblowing. Additionally, European Court of Human Rights jurisprudence further broadens this concept by recognizing the public interest in disclosures that start legitimate public debate or expose morally questionable conduct, even when the acts revealed are technically lawful.[12]

Alongside public interest, the severity of the data breach itself is key to the proportionality assessment. For example, the release of routine contact details poses a lower risk compared to the disclosure of sensitive data under Article 9 of the GDPR, such as information relating to health, racial or ethnic origin, or political beliefs. The latter would typically be viewed as a more serious privacy violation. Additionally, the number of individuals affected by the breach must be taken into account. A disclosure that impacts only a small group carries different implications than one that compromises the data of thousands or millions.

When this evaluation favours the whistleblower's actions, the public interest may outweigh the data breach consequences and disclosure can be justified under the principle of proportionality.

3. The European Court of Human Rights: Balancing Fundamental Rights

The European Court of Human Rights (ECtHR) plays a crucial role in shaping the understanding and protection of whistleblowers through its interpretation of the European Convention on Human Rights (ECHR)[13], particularly Article 10, which guarantees freedom of expression. This right is fundamental for whistleblowers, as their actions inherently involve imparting information, often to challenge powerful interests or expose misconduct. One of the most significant cases in relation to whistleblowing is Guja v. Moldova[14]. The landmark Grand Chamber judgment in this case established a set of six key criteria for assessing whether an interference with a whistleblower's Article 10 rights is justified and proportionate. These criteria, often referred to as the "Guja criteria,"[15] are:

- The channels used to make the disclosure: The Court generally indicates a preference for internal reporting channels to be exhausted first before resorting to external disclosure, particularly to the media, which is often seen as a last resort.
- The authenticity of the disclosed information: The whistleblower must have acted responsibly by carefully verifying, as far as circumstances permit, that the information is accurate and reliable.

11. Report From the Commission to The European Parliament and The Council on the implementation and application of Directive (EU) 2019/1937 of the European Parliament and of the Council of 23 October 2019 on the protection of persons who report breaches of Union law, Brussels, 3.7.2024 COM (2024) 269
12. Couderc and Hachette Filipacchi Associés v. France [GC], no. 40454/07, ECtHR 2015;
*Bucur and Toma v. Romania*no. 40238/02, ECHtR 2013;
Guja v. Moldova [GC], no. 14277/04, ECHtR 2008.
13. European Convention on Human Rights, Rome, November 4, 1950, Council of Europe Treaty Series No. 5.
14. Guja v. Moldova [GC], no. 14277/04, ECHtR 2008
15. Idem

- Good faith: The whistleblower's motives are important. The disclosure should be aimed at serving the public interest rather than being driven by personal grievance, animosity, or expectation of personal advantage.
- The public interest in the disclosed information: There must be a genuine public interest in the information revealed. The more significant the public interest, the stronger the justification for disclosure.
- The detriment caused to the employer or authority: The harm suffered by the employer (e.g., reputational damage, breach of confidentiality, financial loss) as a result of the disclosure must be weighed against the public interest in having the information revealed.
- The severity of sanctions imposed on the whistleblower by the employer, authority, or national court: The nature and severity of the penalty or measure taken against the whistleblower must be proportionate to the legitimate aim pursued by the interference.

These criteria are not applied as a rigid checklist but rather as factors in a balancing exercise. These criteria were further nuanced in the recent Halet v. Luxembourg[16] case. The Grand Chamber decision clarifies in detail its approach to the balancing exercise inherent in whistleblower protection, especially concerning the weight given to public interest and detriment.

The applicant in this case, Raphaël Halet, an employee of PricewaterhouseCoopers (PwC) in Luxembourg, disclosed sixteen confidential documents (Advance Tax Rulings and accompanying cover letters) to a journalist. These documents revealed highly advantageous tax agreements concluded between PwC, on behalf of multinational companies, and the Luxembourg tax authorities, forming part of the "LuxLeaks" scandal. For this disclosure, Halet was criminally convicted in Luxembourg for domestic theft, fraudulent access to a computer system, breach of professional secrecy, and laundering the proceeds of these offences, resulting in a fine of EUR 1,000. The applicant claimed that the conviction for this public disclosure violated his right to freedom of expression under Article 10 ECHR.

The Grand Chamber, overturning an earlier Chamber judgment that found no violation, meticulously applied, and refined the "Guja criteria":

- Public Interest: The Grand Chamber emphasized that the disclosed information on tax rulings was of considerable public interest, contributing to an ongoing debate across Europe on corporate tax avoidance, tax transparency, and social justice. Crucially, it rejected the notion that the information had to be "essential, new, and previously unknown" to be of public interest, noting that repeated disclosures can be necessary to prompt action.
- Detriment Caused: The Court re-evaluated the balance between the public interest served by the disclosure and the detriment caused to PwC. It found that the Luxembourg courts had attached decisive weight to the harm suffered by PwC without adequately considering the significant public interest at stake. The Grand Chamber concluded that the public interest in the information concerning the tax practices of multinational companies outweighed the detrimental effects of the disclosure on PwC's rights.
- Severity of Sanction: The criminal conviction and the EUR 1,000 fine, though seemingly modest, were deemed disproportionate given the importance of the public interest served by the disclosure and the chilling effect such sanctions could have on other potential whistleblowers.

As a result, the Grand Chamber found that there had been a violation of Article 10 ECHR. The judgment signals a robust protection for disclosures that serve a significant public debate, even when they cause harm to private commercial interests. The decision in Halet v. Luxembourg potentially strengthens the position of whistleblowers by giving greater weight to the public's right to know, especially on matters of significant societal concern like corporate tax practices, even when weighed against substantial private or commercial interests. The decision of the Grand Chamber shed more light on the steps that whistleblowers need to follow and the criteria that they need to meet in order to get protection. This could influence courts when interpreting similar concepts under EU law.[17]

When a whistleblower's public disclosure involves the personal data of third parties, Article 8 ECHR, which protects the right to respect for private and family life, home, and correspondence, is directly engaged. The ECtHR must then balance the whistleblower's Article 10 rights (and the public interest in the disclosure) against the Article 8 rights of the data subjects whose privacy has been infringed. The Court generally accords significant weight to the right to privacy. Key factors in such balancing include: the contribution of the disclosed information to a debate of general interest; how well-known the person concerned is and the subject of the report; their prior conduct; the content, form, and consequences of the publication; and the circumstances in which information was obtained.[18] These principles could be adapted to the whistleblowing context. The "necessity" of disclosing identifiable personal data – as opposed to anonymized, pseudonymized, or aggregated information – to achieve the public interest aim would be a paramount consideration in the proportionality assessment. The more sensitive and

16. Halet v. Luxembourg [GC], no. 21884/18, ECHtR 2023
17. Kafteranis, D. & Andreadakis, S. (2023) A New Perspective on The Protection of Whistleblowers Under ECHR: Halet v. Luxembourg, Studies in Logic, Grammar and Rhetoric 68, DOI:10.2478/slgr-2023-0035
18. Von Hannover v. Germany (No. 2), no. 40660/08 and 60641/08, ECtHR 2012

extensive the personal data disclosed, and the less direct its necessity for revealing the core wrongdoing, the more challenging it would be to justify the disclosure under Article 10 when weighed against the competing Article 8 rights of the individuals concerned.

4. Insights for Organisations and Potential Whistleblowers

Organisations play a crucial role in mitigating the risks associated with public disclosures and data breaches by whistleblowers. Data breaches caused by public disclosures present a huge risk for organisations and may cause serious reputational and financial damage. The most efficient way for organisations to manage public disclosures is to avoid them in the first place.

There is no need for whistleblower to go public if the issue can be resolved internally. Moreover, the whistleblower may not be granted protection when opting out for public disclosure if organisation has already properly addressed the concerns reported. Therefore, it is crucial to create Speak Up and Listen Up culture, as well as robust and efficient whistleblowing channels. The channels should be easy to access and should inform whistleblowers about the process, expected response times, confidentiality, protection, and available advice. Additionally, the feedback is crucial for maintaining trust. Beyond the EU requirements (acknowledgment within 7 days, handling information within 3 months), good practice includes immediate automated acknowledgment followed by personalized messages. Feedback at each stage should cover next steps, possible outcomes, timeframes, reasons for limited detail (if applicable), and information on support/protection. A simple check-in like *how are you now?* can be valuable. It is good practice to share awareness-raising materials, Speak Up Policy, and related procedures with the whistleblower. Employees should be clearly informed about the available reporting procedures, including possibilities of public disclosure, and their rights and obligations. It is much easier to implement a good Speak Up programme than manage consequences of public disclosure. If organisations implement effective, trustworthy internal channels that lead to action, it significantly reduces the likelihood of whistleblowers feeling compelled to resort to public disclosures, thereby minimizing the risk of associated data breaches. This proactive approach by organisations is a key practical measure to reconcile the tension.

If both an unprotected disclosure and a data breach occur, organisations must document the breach, the data involved, and the damage caused, along with notifying authorities when required by GDPR. This documentation is vital for any potential legal action against the whistleblower. Before pursuing legal action against a whistleblower, organisations should conduct a thorough legal assessment focusing on whether the whistleblower met all conditions for protection under the Whistleblowing Directive (Articles 6, 15, 21) and whether the information was lawfully acquired. There is a strong basis for organisations to take action if the acquisition was criminal (e.g., data theft, hacking). As going against whistleblowers may seriously harm the organisation's Speak Up program and lead to the loss of trust among employees, regardless of whether the public disclosure is protected, it is essential to assess the strength of evidence regarding the harm caused to the company by the disclosure and consider possible risks.

Individuals contemplating making a public disclosure, especially involving personal data, should be aware of legal thresholds and conditions for protection. If personal data must be included in a public disclosure, serious consideration should be given to redacting or minimizing that data to only what is absolutely and demonstrably necessary to reveal the breach. A whistleblower would need to demonstrate, if challenged, that they considered whether the wrongdoing could be exposed with less personal data, or through anonymisation or redaction. The more sensitive or extensive the personal data, the higher the bar for proving the necessity and proportionality of its disclosure. There appears to be a significant gap in explicit, detailed guidance from authorities specifically addressing how an individual whistleblower should navigate GDPR compliance (particularly regarding data minimization and establishing a lawful basis) when making a public disclosure that involves third-party personal data. While the Whistleblowing Directive places the onus on the whistleblower to have "reasonable grounds to believe" the disclosure was "necessary" (Article 21(2)), concrete guidance on how an individual can make this complex assessment regarding third-party data before a public disclosure is absent. Existing guidelines on whistleblowing primarily focus on how organisations should manage their reporting schemes, not on the individual reporter. The lack of explicit guidance for individual whistleblowers on GDPR compliance during public disclosures can be twofold. While it creates uncertainty, it also means organisations can argue that whistleblowers who recklessly disregard data protection principles in public disclosures cannot automatically claim their actions were necessary or proportionate. Therefore, given the legal complexities, seeking independent legal advice before making a public disclosure involving sensitive data is highly advisable.

5. Conclusion

The co-existence of whistleblower protection and strict data privacy regulation often creates challenging situations for organisations and whistleblowers. While the Whistleblowing Directive provides significant protections to individuals who expose

wrongdoing, organisations still retain the ability to hold whistleblowers accountable when their actions fall outside the protective power of the Directive, particularly when such actions lead to damaging data breaches. Even though that there are regulatory efforts to find a balance between promoting whistleblowing and safeguarding data privacy, striving for a harmonized approach requires more than just legislative texts. This can only be achieved if organisations support the development of "Speak Up and Listen Up" culture internally. Creating awareness around whistleblowing, removing negative stigma around the topic ("snitches get stitches") is the civic duty of all the organisations on order to promote whistleblowing as a positive phenomenon that should be at the core of all democratic societies. This includes fostering environments where internal reporting is genuinely encouraged, appropriately investigated, and leads to meaningful corrective action, therefore, removing need for risky public disclosures. Without this supportive infrastructure, the law alone may struggle to achieve the balance between protecting those who speak up, respecting public interest and safeguarding the fundamental right to privacy for all. The continued efforts of all the parties involved, and the commitment of organisations and individuals to ethical conduct is essential in navigating this complex challenge.

Europe's Health Data Shift: Regulation, Anonymisation, and Security

T. Chib, R. van Kempen and A.I. Hakkers[1]

The 2021 ransomware attack on Ireland's Health Service Executive[2], where attackers threatened to publish patient data, presaged a new era of healthcare vulnerability. As Europe implements ambitious data-sharing frameworks in 2025, this incident reminds us of the central challenge facing modern healthcare: how do we make data useful without making it dangerous?

To understand this fundamental tension, we examine it through three interconnected lenses that together form what we call the '*Privacy-Security-Utility Triangle*', or the '*Golden Triangle*':

The regulatory analysis reveals how Europe's new frameworks attempt to legislate the balance between usefulness and danger. In this article, we analyse where these regulations create genuine progress, such as patient control mechanisms and cross-border portability and where they multiply complexity without resolving core tensions. The convergence of four major frameworks over twenty-four months forces healthcare organisations to navigate conflicting requirements, and where making data useful for research may make it dangerous for privacy.

The anonymisation analysis examines why the regulatory promise of "anonymous" health data sharing fails against technical reality. We suggest how medical data's inherent uniqueness makes true anonymisation impossible while retaining utility. Every technique that preserves usefulness leaves fingerprints that enable re-identification, while every method that ensures anonymity destroys the very characteristics that make data valuable for research.

The cybersecurity analysis shows how every step toward better, faster care also creates new risks. As hospitals connect more systems and share more data, they become easier targets. Old machines, rushed innovation, and complex networks open the door to attacks that can shut down care or leak sensitive data. In healthcare's digital evolution, the question isn't whether we can make data useful without increasing risks. It's how dynamically we can balance creating systems resilient enough to bend without breaking, open enough to enable care without enabling attacks, and sophisticated enough to know when each matters most, and continue doing so as technology, threats, and care needs keep evolving.

These three perspectives converge to reveal that healthcare's digital transformation is not a technical challenge with a definitive solution, but rather a perpetual balancing act between competing imperatives that cannot be permanently resolved. The regulatory frameworks promise control through compliance, anonymisation promises safety through transformation, and cybersecurity promises protection through barriers, yet each solution creates new vulnerabilities even as it addresses existing ones.

1. Privacy in Healthcare: how regulations are reshaping rights and responsibilities

The European health data landscape is undergoing its most significant transformation since GDPR, driven by a deceptively simple question: how do we make data useful without making it dangerous? Four interconnected frameworks each attempt their own answer, yet their interaction reveals that every mechanism designed to unlock data's utility, such as standardised access, transparency, and interoperability, also simultaneously amplify risks.

1.1. European Health Data Space ("EHDS")

The European Health Data Space[3], which entered into force in January 2025 after three years of debate, represents Europe's most ambitious attempt to create a unified health data ecosystem. Its dual infrastructure: MyHealth@EU for patient data portability and HealthData@EU for research, promises to revolutionize both individual empowerment and collective medical advancement. Yet the decade-long implementation timeline stretching to 2035 itself acknowledges the enormity of harmonising 27 national health systems while introducing unprecedented patient controls.

1. Tanya Chib is a regulatory and privacy lawyer, Dr. Anna Hakkers is a cybersecurity expert specializing in Data Security and Renate van Kempen is a data anonymisation consultant focused on health data protection and re-identification risk assessment.
2. https://www2.hse.ie/services/cyber-attack/what-happened/
3. https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space-regulation-ehds_en

EHDS Implementation Timeline
2027: Member States designate Digital Health Authorities and Health Data Access Bodies.
2029: Basic health data (patient summaries, e-prescriptions) available cross-border via MyHealth@EU
2031: Expanded to include medical imaging, lab results, and discharge summaries
2033: European Commission evaluation and potential adjustments
2035: Full implementation including HealthData@EU infrastructure for secondary use

The EHDS does introduce genuinely transformative patient rights that, if properly implemented, could fundamentally alter the power dynamic in healthcare. Patients gain control through opt-out rights for secondary data use without needing to justify their decision. However, the exceptions for public interest research create ambiguity about when individual preferences can be overridden[4]. Nonetheless, EDHS enables granular access control, allowing patients to shield sensitive information about mental health or sexual health from certain providers while sharing other medical data. Real-time notifications promise to alert patients whenever their data is accessed, creating an audit trail of who sees what and why. Patients can demand immediate error corrections at no cost, addressing long-standing frustrations with inaccurate medical records that follow individuals across providers.

Yet these empowering provisions contain subtle contradictions that may undermine their effectiveness. The emergency override provision, while necessary for life-threatening situations, lacks clear boundaries on what constitutes sufficient emergency to bypass patient restrictions and remains undefined, potentially creating a loophole that normalises circumventing preferences. The technical complexity that makes data useful for granular control, also makes it dangerously inaccessible for vulnerable populations who need it most. Further, the framework's data sharing mechanisms could enable large technology companies to access aggregated health datasets for AI development, potentially prioritising commercial innovation over individual privacy despite regulatory safeguards.

The digital divide across Europe further threatens to transform EHDS from an equalising force into another source of healthcare inequality[5]. While Denmark's advanced digital infrastructure can readily support real-time notifications when patient data is accessed, healthcare facilities in less advanced countries may struggle. Vulnerable populations such as the elderly, disabled, and migrant communities will require extensive support to navigate granular controls, further adding costs that cash-strapped systems cannot afford. Without addressing these disparities, EHDS risks creating a two-tier system where digital literacy determines healthcare quality. The 2027 milestone will reveal whether Member States pursue genuine transformation or mere compliance. If stakeholders embrace both the letter and spirit of the regulation, Europe could pioneer a model balancing individual autonomy with collective benefit. If not, the EHDS risks becoming another hollow framework - well-intentioned but ultimately ineffective.

As exemplified above, the EDHS embodies the 'useful-dangerous' paradox at many levels. For example, while standardising access allows a Spanish tourist to get his prescriptions in Sweden, this does create new attack vectors. The framework's patient control features (opt-out rights, granular access, real-time notifications) make data useful for individual empowerment but dangerous through complexity, in cases of vulnerable populations where elderly patients might accidentally restrict critical information from emergency providers.

1.2. The EU-US Data Privacy Framework ("DPF")[6]

The implementation challenges highlighted above multiply when EHDS intersects with the EU-US Data Privacy Framework, established in July 2023 as the third attempt to enable transatlantic data flows. If EHDS shows how making data useful for patients makes it dangerous for privacy, the DPF reveals how making it useful for research makes it dangerous for sovereignty.

The DPF aims to establish "adequacy" for personal data transfers under GDPR, creating a mechanism for legal health data sharing between European and American institutions and organisations. The DPF's predecessor, the Safe Harbor and Privacy Shield, failed due to fundamental conflicts between EU privacy rights and US surveillance practices[7], so it remains to be seen if this third attempt can survive legal challenges.

Complexity arises because the DPF must coexist with multiple regulatory frameworks. Healthcare

4. Marelli L, Stevens M, Sharon T, Van Hoyweghen I, Boeckhout M, Colussi I, Degelsegger-Márquez A, El-Sayed S, Hoeyer K, van Kessel R, Zając DK, Matei M, Roda S, Prainsack B, Schlünder I, Shabani M, Southerington T. The European health data space: Too big to succeed? Health Policy. 2023 Sep;135:104861. doi: 10.1016/j.healthpol.2023.104861. Epub 2023 Jun 26. PMID: 37399677; PMCID: PMC10448378
5. https://www.eu-patient.eu/globalassets/ehds-analysis---final.pdf
6. https://www.dataprivacyframework.gov/Program-Overview
7. https://curia.europa.eu/jcms/upload/docs/application/pdf/2020-07/cp200091en.pdf

organisations transferring data between the EU and US must satisfy not only DPF principles but also Health Insurance Portability and Accountability Act ("HIPAA")[8] requirements, while navigating the patchwork of country-specific health data rules that EU Member States impose beyond GDPR[9]. Without clear guidance on how these frameworks interact, organisations are forced to layer Standard Contractual Clauses atop DPF certification, essentially creating their own reconciliation between competing regulatory demands. The absence of healthcare-specific provisions means medical data transfers operate in a regulatory grey zone where generic privacy principles inadequately address sector-specific realities.

The DPF represents progress but falls short of providing comprehensive solutions for healthcare data transfers. Its generic framework, combined with Member States' retained flexibility to impose additional health data restrictions, forces organisations into complex multi-layered compliance strategies. Healthcare institutions engaged in transatlantic collaboration must combine DPF certification with tailored contractual arrangements, conduct exhaustive jurisdictional analyses for each EU Member State involved, and essentially create their own sector-specific protections within the generic framework. This patchwork approach reflects a deeper failure: the absence of healthcare-specific provisions that acknowledge how medical data differs from commercial information in sensitivity, use patterns, and ethical considerations. Until regulators develop frameworks that genuinely address health data's unique nature, rather than treating it as merely another data category with higher risks, organisations must compensate through costly and inefficient workarounds that may still face legal challenges.

1.3. The European Data Act ("Data Act")

The Data Act[10], entering into force in January 2024 with general applicability from September 2025, introduces another layer of complexity that fundamentally challenges medical device innovation. While promising to create a thriving data sharing economy by granting users rights to access and share data from their pacemakers to fitness trackers, the regulation's "access by design" mandate demands architectural changes that many manufacturers may not have anticipated.

The Data Act's nuanced approach to access rights reveals tensions between promise and practice. Article 3.1 establishes that data must be "by default, easily, securely and directly accessible", yet manufacturers retain control over initial contract terms, creating immediate ambiguity about what constitutes genuine access. Users gain rights only to data generated by their use, not all data processed by devices, a distinction that becomes critical for AI-enabled medical devices that process far more information than they store.

Article 4's requirement for "continuous and real-time" data access poses particular challenges for implanted devices designed with closed architectures for safety reasons. Medical device manufacturers must somehow fit user-accessible data ports or wireless interfaces without compromising therapeutic functions, all while maintaining compliance with the EU Medical Device Regulation ("MDR") and In Vitro Diagnostic Regulation ("IVDR").

The framework's third-party sharing provisions under Article 5 create unexpected limitations. While patients can theoretically share their device data freely, the explicit exclusion of platforms designated as "gatekeepers" under the Digital Markets Act means patients cannot integrate pacemaker data with major tech companies' health platforms that they already might be using. This restriction, intended to prevent Big Tech dominance, may impair patients seeking consolidated health management tools.

More concerning, the Act's allowance for "reasonable compensation" for data access could transform the promise of free patient access into a cost barrier, particularly for complex medical devices requiring substantial infrastructure investments to enable sharing capabilities.

The Data Act's relationship with existing regulations adds layers to an already complex compliance puzzle[11]. The regulation explicitly preserves GDPR requirements, intellectual property rights, and Member States' prerogatives in public health and security, creating a multi-dimensional challenge where each regulatory framework pulls in different directions. A solution compliant with the Data Act might violate GDPR's data minimisation principles or compromise trade secrets protected under intellectual property law.

This ambitious scope reflects the EU's commitment to digital sovereignty but risks undermining the very innovation it seeks to democratise. Success requires more than technical compliance; it demands fundamental rethinking of product design, business models, and data governance strategies. Medical device companies must balance the Act's democratising vision with practical realities of device safety, inno-

8. https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html
9. Tschider, Charlotte and Corrales Compagnucci, Marcelo and Minssen, Timo, The New EU–US Data Protection Framework's Implications for Healthcare (September 27, 2024). Journal of Law and the Biosciences, volume 11, issue 2, 2024[10.1093/jlb/lsae022], Available at SSRN: https://ssrn.com/abstract=4983419
10. https://digital-strategy.ec.europa.eu/en/policies/data-act
11. Casolari, Federico & Buttaboni, Carlotta & Floridi, Luciano. (2023). The EU Data Act in Context: A Legal Assessment. SSRN Electronic Journal. 10.2139/ssrn.4584781.

vation cycles, and global competition. Those engaging early with the evolving regulatory framework and actively participating in guidance development may find opportunities within the constraints, and many may redirect innovation efforts to less regulated markets.

1.4. The EU AI Act[12]

While the Data Act struggles with making device data useful without compromising safety, the AI Act faces an even sharper dilemma. The EU AI Act, coming into force in August 2024 with full applicability by August 2027, adds a final layer of regulatory complexity as the world's first comprehensive legal framework for artificial intelligence systems. Its risk-based approach creates immediate categorisation challenges for medical AI, where the same technology might be prohibited in general use but permitted in healthcare contexts. Emotion recognition systems, normally banned, receive exemptions for medical purposes like psychological treatment, this flexibility that acknowledges healthcare's unique needs while creating interpretation challenges about what constitutes legitimate medical use.

The AI Act's risk categories translate into different compliance burdens. For example, high-risk AI systems, such as AI-assisted X-ray diagnosis, emergency triage systems, and medical training assessment tools face stringent requirements encompassing risk management protocols and human oversight. Low-risk systems like administrative AI for structured radiology reporting (unless integrated into medical devices) have minimal obligations. Approximately 75% of current AI medical devices will be classified as high-risk, requiring compliance with 16 distinct requirements within 36 months of the Act coming into force[13]. These requirements span continuous risk management, data governance protocols, extensive technical documentation, automatic event logging, fundamental rights assessments, and post-market surveillance systems, each adding layers of complexity and cost.

AI-enabled medical devices must satisfy both the AI Act's horizontal requirements and MDR/IVDR's vertical sector-specific regulations. Both frameworks demand risk management systems, technical documentation, and conformity assessments, but with different specifications, timelines, and interpretations[14]. Manufacturers of medical devices must make adjustments in a number of areas in order to comply with the requirements of the AI Act in addition to the requirements of the above-mentioned regulations.

The EU AI Act represents both a necessary step toward responsible AI governance and a potential impediment to medical innovation. While its comprehensive requirements may protect patients from harmful AI applications, the regulatory complexity, such as the overlap with existing frameworks, threatens to stifle the very innovation it seeks to regulate, or worse: create a compliance theatre with checkbox compliance.

These regulations are not merely coexisting but also compounding the useful-dangerous paradox. The EHDS makes data useful for cross-border care, DPF makes it useful for transatlantic research, Data Act makes it useful for device transparency, and AI Act makes it useful for algorithmic accountability. But each layer of usefulness multiplies dangers: more access points, more complexity, more attack surfaces, more ways for well-intentioned utility to become unintended vulnerability. This creates an urgent question: how can healthcare organisations enable all this data flow while maintaining privacy? The answer many are turning to is anonymisation, which appears to be deceptively simple until you examine what it actually takes to anonymise health data in a way that satisfies all regulatory frameworks simultaneously.

2. Anonymisation in healthcare datasets, an analysis to maintain utility

Anonymisation within the European healthcare sector has long been treated as a compliance afterthought, something that can be just "done" by removing direct identifiers only. The reality however is far more advanced. It is essential to distinguish between **pseudonymisation** and **anonymisation**, as the two are often confused or incorrectly used interchangeably. According to the European Data Protection Board, pseudonymised data remains personal data under the GDPR and does not meet the criteria for anonymisation[15]. The UK Information Commissioner's Office emphasises that anonymisation must render data incapable of identifying individuals, even when cross-referenced with other data[16].

In the European legal context, for data to be considered truly anonymised under GDPR, the risk of re-identification must be negligible when assessed against all means reasonably likely to be used[17]. This

12. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng
13. Emmanouil P Vardas, Maria Marketou, Panos E Vardas, Medicine, healthcare and the AI act: gaps, challenges and future implications, *European Heart Journal - Digital Health*, 2025;, ztaf041, https://doi.org/10.1093/ehjdh/ztaf04
14. Busch, F., Kather, J.N., Johner, C. *et al.* Navigating the European Union Artificial Intelligence Act for Healthcare. *npj Digit. Med.* 7, 210 (2024). https://doi.org/10.1038/s41746-024-01213-6
15. EDPB, guidelines 01/2025 on Pseudonymisation https://www.edpb.europa.eu/system/files/2025-01/edpb_guidelines_202501_pseudonymisation_en.pdf
16. ICO, *Anonymisation*. https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/data-sharing/anonymisation/
17. GDPR, recital 26: https://gdpr-info.eu/recitals/no-26/

bar is high, and many healthcare organisations are realising that anonymisation is no longer a fixed outcome, but a discipline requiring judgement, technical skills and constant adaptation.

This shift is driven by both regulatory expectations and operational needs. On the one hand, GDPR excludes anonymised data from its scope, creating strong incentives for sharing data in an anonymous format, particularly for secondary use, such as research, benchmarking or training (AI) algorithms. On the other hand, achieving this legal status is difficult in practice, especially when working with rich and complex health datasets. The technical challenges, institutional misunderstandings and a lack of consistent standards create real barriers to progress.

2.1. Key challenges in anonymisation

While awareness of anonymisation has grown and regulatory expectations have become clearer, many healthcare organisations, including those in pharma, healthtech and clinical research, still face significant challenges in practice. These range from technical constraints to persistent misconceptions, and they continue to complicate the safe and effective reuse of health data. The main challenges observed across these sectors are:

1. Risk of singling out through rare combinations:
Health data often includes detailed combinations of attributes such as age, treatment date, diagnosis, and geographic location. These combinations, while not directly identifying, can be unique. For instance, a 42-year-old patient with a rare diagnosis admitted to a specific regional hospital in March may be the only person matching that profile. This kind of uniqueness can enable re-identification even in the absence of names or IDs[18].

2. Ambiguity about what constitutes "sufficient" anonymisation:
Although the GDPR sets a high bar, it does not define measurable thresholds. Phrases like "reasonably likely" and "negligible risk" leave room for interpretation. In practice, this legal ambiguity makes it difficult for organisations to assess whether a dataset is anonymous enough to no longer fall in scope of the GDPR. This is particularly problematic when data is reused or shared across borders where interpretations may differ.

3. Limitations of current tools:
While anonymisation tools are improving, most are optimised for structured data and still require expert configuration. Automated settings often fail to capture the entire complexity of clinical or longitudinal datasets. Also, there are no tools built yet that cover all anonymisation challenges and techniques, and many tools struggle with the fact that no dataset is the same. For example, applying a uniform generalisation rule on identifiers might protect identity but can destroy important useful information in treatment timing or outcomes. Hence, at the moment, most datasets require a tailored approach that tools alone cannot provide.

4. Misunderstanding anonymisation as a one-time action:
A common misconception is that once data is anonymized, it stays that way forever. However, as new external data becomes available, new techniques emerge that adversaries can use or new regulations are implemented, a dataset that was previously considered safe may no longer meet the anonymisation standards. It is important to bear in mind that anonymisation is not a one-and-done process. As highlighted by the Spanish Data Protection Agency in their publication on common misunderstandings[19], it requires periodic re-evaluation, especially if data is re-used or shared in a new context, out of scope of the original consent.

5. Inconsistent adoption of advanced techniques:
While some organisations have integrated advanced techniques like hashing, differential privacy and k-anonymity[20] to mitigate risk, implementation is often fragmented. One research team may apply a secure hashing method, such as SHA-512, while another uses a weaker approach, like simple randomisation to generate pseudonyms. The adoption of most advanced techniques is often hindered by a lack of technical expertise and proper tooling or other resources. This inconsistency undermines trust and increases the likelihood of either over-anonymisation and under-protection, so a higher risk of leaving room for reversing the techniques used and allowing the adversary to re-identify subjects.

2.2. Synthetic data as a solution to maintain utility?

Synthetic data is often presented as a privacy-preserving solution, but its limitations are frequently overlooked. Multiple studies show that disputes remain over whether synthetic datasets leak identifiable patterns or maintain analytic utility. A narrative review found concerns that synthetic data can "not replicate precisely the content and properties of the original dataset," leading to risk of data leakage[21]. Another recent study described a "data-sharing paradox" in healthcare: synthetic data is designed for sharing yet often remains overly re-

18. ICO, *Anonymisation*. ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/data-sharing/anonymisation/
19. AEPD, *10 Misunderstandings related to Anonymisation*. https://www.aepd.es/guides/10-anonymisation-misunderstandings.pdf
20. AEPD, *K-anonymity as a privacy measure https://www.aepd.es/guides/k-anonymity-as-a-privacy-measure.pdf*
21. Gonzales et al., *Synthetic data in health care: A narrative review*, PLOS Digital Health (2023) https://pmc.ncbi.nlm.nih.gov/articles/PMC9931305/

stricted due to ambiguous re-identification risk metrics[22].

Based on our experiences within many health-related organisations, synthetic data is currently seen as not yet mature enough for widespread healthcare use. For it to be safe and effective, it must be demonstrably free from identifiable patterns and retain sufficient clinical relevance, a delicate balance that has not been consistently achieved.

An example might be if, if synthetic patient data were generated for a rare neuromuscular disease cohort and researchers relied upon this data to develop diagnostic models. If the synthetic dataset closely mirrored a unique, real patient (e.g. combining rare genetic markers with clinical outcomes), there is a re-identification risk. Conversely, if too much distortion is applied, critical genotype–phenotype links may break, rendering the data scientifically useless. Without rigorous privacy and utility evaluations, the data may expose patients' identities or invalidate research outcomes.

A 2024 study in *Nature* confirms these risks, warning that synthetic data may allow membership inference or re-identification attacks, particularly when auxiliary datasets are available[23]. Until standard metrics and rigorous governance frameworks exist, synthetic data remains a useful tool, but not yet a plug-and-play alternative for anonymisation.

2.3. How to maintain as much utility as possible?

To make data useful without making it dangerous, healthcare organisations must begin treating anonymisation as a strategic capability, not just another technical step. Leading organisations are already moving in this direction by embedding anonymisation into their broader data governance programs. This includes incorporating risk assessments during project design, adopting context-based transformation strategies, and assigning clear responsibilities for reviewing anonymisation outcomes.

Crucially, anonymisation must be integrated early into the data life cycle, well before data is shared or published (anonymisation-by-design). This requires the involvement of qualified experts who understand both the technical risks and the regulatory landscape, as well as the clinical context of the data. Without this, organisations risk either over-transforming data and losing its utility value, or under-protecting it and exposing patients to privacy harms, therefore making it too dangerous.

Several sector-specific initiatives are helping to close this gap. MedTech Europe, for example, has developed a practical anonymisation framework to support its members in applying structured, risk-based approaches. While not yet formally published as standalone guidance, the framework was outlined in a recent MedTech Europe article and aims to offer clear starting points for organisations that need to operationalise anonymisation within compliant data sharing practices[24].

Ultimately, anonymisation done well enables the secondary use of health data, so maintaining as much utility as possible while safeguarding individual privacy. It allows organisations to meet GDPR expectations, build public trust, and support responsible innovation. But this balance cannot be achieved through a single action or via generic tools only. Proper anonymisation must be seen for what it really is: a discipline that requires ongoing investment, expert judgement and continuous attention. This ensures a maximum of utility without the data becoming dangerously harmful.

Even the most sophisticated anonymisation strategies require protection, especially when data is shared publicly or across borders. Protecting information in motion and at rest becomes just as critical as transforming it at the source. This is where cybersecurity steps in, not as a final barrier, but as a dynamic additional safeguard for possible and potential outside threats.

3. Cybersecurity in healthcare: balancing protection, privacy and progress

The Ireland Health Service Executive attack mentioned in the introduction exemplifies a pattern that has only intensified since 2021. When Germany's Düsseldorf University Hospital suffered a ransomware attack in 2020, the consequences went beyond data breaches when a patient died after emergency services had to be rerouted to a more distant hospital[25]. Such incidents reveal healthcare's fundamental challenge: every connection that enables better care also creates new vulnerabilities.

This question becomes more urgent as Europe implements ambitious data-sharing frameworks in 2025. Every technological advance that promises better care through connected systems simultaneously creates new vulnerabilities. The challenge isn't simply

22. Achterberg et al., *The Data Sharing Paradox of Synthetic Data in Healthcare* (2025) https://arxiv.org/html/2503.20847v1
23. Recent *Nature Scientific Reports* study on membership inference and re-identification risk in healthcare synthetic data https://www.nature.com/articles/s41598-024-72894-y
24. van Kempen, R. *Can Europe unlock the power of data while protecting privacy?*, MedTech Europe (2024). https://www.medtecheurope.org/medtech-views/policy-views/can-europe-unlock-the-power-of-data-while-protecting-privacy/
25. Ransomware's impact on patient care, including the first reported death: https://arstechnica.com/information-technology/2020/09/patient-dies-after-ransomware-attack-reroutes-her-to-remote-hospital/

technical. It reflects deeper tensions between competing imperatives that cannot be permanently resolved, only dynamically balanced.

3.1. The data paradox: connected care, multiplied risk

Healthcare data possesses unique characteristics that intensify such tensions. Clinical records combine personally identifiable information, protected health information, payment details, and increasingly, biometric markers within single datasets. For pharmaceutical and life sciences organisations, patient data coexists with research findings and manufacturing processes representing millions in intellectual property value[26].

The Northeast Radiology breach illustrates how data's inherent value makes it vulnerable. Unauthorized access persisted from April 2019 through January 2020, with discovery only occurring in March 2020. By then, 298,532 patient records had been exposed[27]. The breach's extended timeline reveals a crucial insight: the very characteristics that make health data valuable for longitudinal care make it attractive for theft. Rich, comprehensive records enable better treatment decisions but also command premium prices in criminal markets.
Recent statistics underscore this reality. Ransomware accounts for 54% of reported cyber incidents in EU healthcare, with 43% including confirmed data leaks according to ENISA findings[28, 29,]. These aren't merely IT disruptions. They represent fundamental breakdowns in healthcare's ability to simultaneously share and protect information.

The Medefer case in early 2025 demonstrates how modern interoperability solutions embody these contradictions[30]. An authentication flaw in APIs used by this NHS contractor exposed patient referral data for years. The system was functioning exactly as designed, efficiently sharing information between providers. Its vulnerability stemmed from utility. Making data flow seamlessly between organisations meant reducing friction, and reduced friction meant fewer barriers for both authorised and unauthorised access.

This paradox extends beyond individual breaches. Healthcare's push toward integrated care requires breaking down information silos that, however inefficient, provided natural segmentation against attacks. Modern initiatives like integrated data lakes and FHIR APIs promise transformative benefits: population health insights, precision medicine breakthroughs, rapid research capabilities. Yet aggregating scattered data into centralised, accessible platforms creates what security professionals recognise as high-value targets.

Consider how security context typically disappears as data moves through integration layers. Access controls, classifications, and audit trails that protect information in its original system often get stripped away when data flows through APIs into analytics platforms. Once decoupled from native protections, this information gets reused in dashboards, research datasets, and decision support tools with minimal oversight. The very processes that unlock data's potential simultaneously erode its protection[31].

3.2. The innovation trap: racing ahead of protection

Healthcare's digital transformation accelerated dramatically during the COVID-19 pandemic, with telehealth adoption jumping from peripheral service to core delivery mechanism virtually overnight. This transformation saved lives by maintaining care continuity during lockdowns. It also opened vast new attack surfaces faster than organisations could secure them[32].

The pandemic revealed how utility demands can overwhelm security capabilities. Healthcare organisations deployed collaboration platforms, remote access systems, and patient engagement tools at unprecedented speed. Each new capability addressed urgent care needs but also expanded the perimeter that security teams needed to defend. Traditional security models built on controlling defined network boundaries became obsolete when care delivery itself became boundaryless.

This acceleration continues with artificial intelligence adoption. Healthcare organisations rush to implement AI for diagnostics, treatment planning, and operational efficiency. Yet research indicates most lack AI-specific governance frameworks[33]. The

26. Pharmaceutical companies and life sciences organisations face unique challenges in protecting both patient data and intellectual property. See: https://www.techtarget.com/healthtechsecurity/news/366594393/Inadequate-Healthcare-Cybersecurity-Maturity-Jeopardizes-Patient-Privacy
27. Northeast Radiology settlement details: https://www.hhs.gov/press-room/hhs-ocr-hipaa-settlement-nerad.html
28. ENISA Health Threat Landscape Report, p. 3, p. 13: https://www.enisa.europa.eu/sites/default/files/publications/Health%20Threat%20Landscape.pdf
29. European Commission on healthcare cybersecurity: https://commission.europa.eu/news/bolstering-cybersecurity-healthcare-sector-2025-01-15_en
30. NHS API vulnerability investigation: https://www.computerweekly.com/news/366620174/NHS-investigating-how-API-flaw-exposed-patient-data
31. API security incidents in healthcare have risen sharply, with 79% of organisations experiencing incidents: https://www.hipaajournal.com/79-of-healthcare-organisations-experienced-an-api-security-incident-in-the-past-12-months/
32. COVID-19's impact on healthcare cybersecurity: https://pmc.ncbi.nlm.nih.gov/articles/PMC8059789/
33. HIMSS report on AI governance gaps: https://www.himss.org/news/report-health-system-cybersecurity-budgets-increasing-lack-ai-governance-threatens-security

pattern repeats: transformative utility drives adoption before protective measures mature. Each wave of innovation creates new imbalances that organisations struggle to address while already managing previous ones.

The innovation trap isn't simply about moving too fast. It reflects how healthcare's mission creates different risk calculations than other sectors. When a new technology might improve patient outcomes, the ethical imperative to adopt it can override security concerns. This calculation makes moral sense in individual cases but creates systematic vulnerabilities when repeated across thousands of decisions.

3.3. The architecture of vulnerability

Healthcare's technical landscape reveals how historical decisions about balancing access and protection accumulate into current vulnerabilities. Organisations operate environments mixing 20-year-old medical devices with cloud-native applications, multivendor systems with proprietary protocols, and safety-critical applications running on unsupported operating systems. Each element represents a past decision where immediate utility took precedence, creating what security teams now experience as a nearly unmanageable attack surface[34].

Medical devices are a great example of these accumulated trade-offs. A ventilator or MRI machine designed for a 20-year service life cannot be patched like consumer electronics. The device critical for care simultaneously endangers the network through unpatched vulnerabilities[35].
Legacy clinical systems present similar dilemmas. Picture Archiving and Communication System ("PACS") imaging archives, laboratory information systems, and specialised departmental solutions often run on outdated platforms because replacement would disrupt care delivery. These systems grew in isolation, use unique or outdated data formats and protocols that are not supported by modern security tools. A Digital Imaging and Communications in Medicine ("DICOM") imaging file that enables critical diagnostics may be invisible to security scanners designed for conventional documents.

The architectural complexity multiplies through third-party dependencies. Modern healthcare relies on intricate webs of vendors, from electronic health record ("EHR") providers to device manufacturers, imaging centres to laboratory networks. Each connection enables essential services but also creates potential entry points. Supply chain attacks exploit these relationships, compromising trusted partners to reach ultimate targets. The same interconnections that enable coordinated care, also enable coordinated attacks.

3.4. The maturity spectrum in the sector

Cybersecurity maturity in healthcare remains uneven across subsectors. Providers, particularly hospitals and clinics, tend to lag behind pharma and life sciences, where protecting intellectual property and complying with regulations have driven earlier investments in security. Medical device manufacturers are only recently shifting from a focus on safety to one that includes robust digital protection, responding to both rising connectivity and regulatory scrutiny.

The result is a patchwork with islands of resilience within a broader environment that is still catching up. But this unevenness also reflects deeper questions of strategic priorities. Organisations more advanced in protecting security tend to be those that have explicitly recognised the tension between utility and risk and invested accordingly.

By 2025, a significant shift is underway. Healthcare technologies are increasingly being built with embedded security from the start. Whether it's a medical device, mobile health application, or an EHR system, incorporating protective measures early in development is now the new norm. This evolution is being shaped by regulations such as the U.S. Food and Drug Administration's cybersecurity guidance and the EU's Cyber Resilience Act, as well as the cumulative impact of previous high-profile security lapses.

3.5. Toward adaptive resilience

Healthcare can accelerate progress by learning from sectors that faced similar challenges in securing complex and distributed systems, such as finance and telecommunications. Over the past decade, these sectors have built mature practices around API governance, zero-trust architecture, and real-time auditing across cloud environments. While healthcare's risks are unique in their clinical consequences, the underlying technical problems are often shared.

The key insight from mature sectors is that balance points shift constantly. What works during normal operations fails during crisis. What protects adequately today becomes vulnerable tomorrow. Organisations need capabilities to sense these shifts and adjust accordingly rather than seeking permanent solutions. Static security models that try to lock down systems inevitably fail because they conflict with care delivery needs.

This adaptive approach extends to governance mod-

34. Medical device security evolution: https://galendata.com/updating-cybersecurity-for-advanced-medical-devices-2024-insights-and-best-practices/

35. Government investigation into medical device vulnerabilities: https://www.reuters.com/article/technology/us-government-probes-medical-devices-for-possible-cyber-flaws-idUSKCN0IB0DQ/

els. Instead of centralised control that stifles innovation or distributed chaos that enables breaches, leading organisations create flexible frameworks. These establish baseline requirements while allowing justified variations. They emphasise outcomes over compliance, measuring whether data remains protected rather than whether specific controls exist.

Healthcare's digital future requires accepting that making data useful without making it dangerous isn't a problem to solve, but a tension to manage. Every advance in care delivery through connected systems creates new vulnerabilities. Every security measure that adds friction potentially delays treatment. Every confidentiality protection that segments data might prevent crucial insights.

4. Conclusion

The Privacy-Security-Utility Triangle we've explored reveals that Europe's health data transformation is not heading toward a destination but embarking on a continuous journey. Each vertex of this triangle, regulation, anonymisation, and security pulls in its own direction, creating tensions that cannot be resolved but must be dynamically balanced.

For healthcare leaders navigating this landscape in 2025 and beyond, the key insight is to stop seeking perfect solutions and start building adaptive capabilities. Success will belong to organisations that can sense when balance points shift, adjust protections without paralysing innovation, and maintain resilience while enabling transformation. In healthcare, both the music and the stakes keep changing. In this dance, standing still is not an option; only those who keep moving, learning, and adapting will thrive in Europe's bold new health data ecosystem.

Uit de boekenkast van de bedrijfsethiek (94)

prof. dr. E. Karssing[1]

In de bedrijfsethiek is een groot aantal boeken en artikelen verschenen waarin op praktische wijze prangende vraagstukken worden behandeld en concrete aanbevelingen worden gedaan voor het bevorderen van de ethiek en integriteit van organisaties en hun medewerkers. Niet iedereen weet deze publicaties te vinden of heeft tijd ze te lezen. Daarom kijkt Edgar Karssing geregeld voor het *Compliance, Ethics & Sustainability Journal* in de boekenkast van de bedrijfsethiek en bespreekt hij een artikel of boek. Deze bijdragen zijn geen recensies, maar een samenvatting van de belangrijkste conclusies en aanbevelingen van de auteur(s), die hij zal confronteren met zijn eigen observaties als onderzoeker, trainer en adviseur op het gebied van ethiek en integriteit.

Dit is deel 2 van een tweeluik over macht en ethiek.

1. Inleiding

Geregeld wijst iemand mij tijdens een college of workshop op het boek *De 48 wetten van de macht* van Robert Greene.[2] Volgens de achterflap 'een handboek om het wezen van de macht te doorgronden'. Het is geschreven als een zelfhulpboek met heel veel adviezen en praktische tips. Iedere wet wordt geïllustreerd aan de hand van talloze historische anekdotes en geeft aan wat je wel of juist niet moet doen om jouw macht te vergroten. Dit klinkt als een boek dat gelezen moet worden als je een artikel over macht in organisaties wilt schrijven. Voorbeelden van wetten zijn:

- Houd uw ware bedoelingen strikt geheim: zet iedereen voortdurend op het verkeerde been en onthul dus nooit de ware bedoeling van uw handelen.
- Laat anderen het werk doen en strijk zelf de eer op: gebruik de wijsheid, de kennis en de routine van anderen om uw eigen zaak te bevorderen.
- Ontwapen uw slachtoffer met selectieve eerlijkheid en edelmoedigheid: een enkel gebaar van eerlijkheid en oprechtheid kan tientallen slinkse manoeuvres verhullen.
- Zoek voor de ander de duimschroef die hem past: iedereen heeft een zwakke plek, een bres in de vestingmuur.
- Vermorzel uw vijand: alle grote leiders sinds Mozes beseften dat een gevreesde vijand compleet verpletterd diende te worden.

En dat gaat zo 500 bladzijden lang door... Waar blijft de ethiek? Greene schrijft hierover in zijn voorwoord:

> *'Macht is in essentie amoreel en één van de belangrijkste vaardigheden die u zich moet aanmeten is geen onderscheid te maken tussen goed en kwaad... Het ligt in de aard van de mens zijn daden te verhullen door zich op alle mogelijke manieren te rechtvaardigen, er steeds maar weer van uitgaande dat hij niets anders dan goeds in de zin heeft. U moet leren om steeds als u iets dergelijks hoort verkondigen inwendig te lachen en u er nooit toe laten verleiden iemands bedoeling en daden te waarderen aan de hand van een reeks morele oordelen die feitelijk alleen een excuus zijn voor de opeenhoping van macht.'*[3]

Is dit satire? Daartoe geeft Greene geen enkele aanwijzing, hij blijft het hele boek bloedserieus. Is Greene de nieuwe Machiavelli? Dat mocht hij willen. Machiavelli was, in de lezing zoals ik die in de vorige boekenkast meegaf, iemand die een oproep deed aan al degenen die het algemeen belang dienen.[4] Zijn oproep was om realistisch te zijn over hoe de wereld echt werkt, omdat je alleen op die manier het goede kunt realiseren. Je moet dus niet naïef zijn: macht en politiek handelen spelen een rol zodra je mensen bij elkaar zet. Mensen hebben verschillende belangen en dan kom je er niet altijd uit met alleen een goed gesprek op basis van redelijke argumenten. En ja, dan moet je soms streken uithalen, dan kunnen doelen sommige middelen heiligen. Maar geen zelfzuchtige doelen, altijd doelen die – direct of indirect – het algemeen belang dienen. Je kunt moreel tekortschieten door machtsmisbruik en politieke spelletjes; je kunt ook tekortschieten door het machtsspel, het politieke spel, uit de weg te gaan waardoor je niks voor elkaar krijgt. Machiavelli roept ons dus op tot het

1. Edgar Karssing is als hoogleraar Filosofie, Beroepsethiek en Integriteitsmanagement verbonden aan Nyenrode Business Universiteit. De auteur dankt Olga Crapels, Wim Lieve, Frank Segers en Raoul Wirtz voor hun commentaar op het concept van deze bijdrage. Voor reacties en suggesties: e.karssing@nyenrode.nl
2. R. Greene (1998/2024). *De 48 wetten van de macht*. Meulenhoff.
3. Greene (1998/2024), ibid.: 22.
4. E. Karssing (2025). 'Uit de boekenkast van de bedrijfsethiek 93'. *Compliance, Ethics & Sustainability Journal*, 25. april, 64-68.

versterken van politieke vaardigheden. Dit ga ik in deze boekenkast verder uitwerken. Hoe kun je jouw mogelijkheden vergroten om succesvol te zijn in het realiseren van jouw doelen, hoe kun je jouw politieke vaardigheden versterken, jouw politieke intelligentie vergroten, zodat je verschil kunt maken? En hoe kun je dat op een moreel verantwoorde manier doen? Hoe ziet de ethiek van machtsgebruik eruit?

Het boek van Greene zal hierbij verder geen enkele rol spelen. Mijn advies: bespaar jezelf de moeite, er zijn zoveel wel serieus te nemen boeken over macht en politiek handelen geschreven. Boeken die gebruik maken van wetenschappelijke inzichten in plaats van alleen maar anekdotes. Daaraan kun je beter tijd en geld besteden. Ik zal in deze boekenkast verschillende van dergelijke boeken gebruiken. In paragraaf 2 bespreek ik het boek *7 Rules of power* van Jeffrey Pfeffer.[5] Van 48 wetten naar 7 regels, is dat een verbetering? In ieder geval behapbaarder. Maar, belangrijker, Pfeffer is één van mijn helden in de 'evidence based management'-beweging, waarbij praktische adviezen altijd worden getoetst aan wetenschappelijke inzichten.[6] Pfeffer gebruikt net als Greene ook veel anekdotes, dat maakt zijn boek lekker leesbaar, maar altijd en alleen om wetenschappelijk onderbouwde adviezen te illustreren. In paragraaf 3 bespreek ik de ethiek van politiek handelen: hoe kun je op moreel verantwoorde wijze macht verwerven en inzetten? En in de slotparagraaf geef ik enkele suggesties hoe je jouw politieke vaardigheden kunt versterken door oefening, ervaring en intervisie. Tevens doe ik het voorstel om als compliance & ethics professional ethiek van macht in workshops met managers te agenderen.

En wat doen we met het boek van Greene? Je kunt het gebruiken als test bij mensen die het hebben gelezen, die jou erop wijzen. Wat vind jij van dit boek? Vind jij het een goed zelfhulpboek? Mensen die hun leven naar dit boek inrichten, kun je beter vermijden—zowel privé als zakelijk.[7] Zij gebruiken macht alleen maar ter meerdere eer en glorie van zichzelf, ook als dat ten koste van jou en anderen gaat.

2. De zeven regels van Pfeffer

In het eerste deel van dit tweeluik heb ik, met Mark Rutte als inspiratiebron, macht neergezet als de mogelijkheid om iets voor elkaar te krijgen. Politiek handelen betreft dan enerzijds het verwerven van macht en anderzijds het gebruik ervan. Beide aspecten verdienen aandacht. Hoe kun je machtsbronnen verwerven zodat je goed voorbereid de politieke arena betreedt? En vervolgens: hoe kun je het machtsspel beter spelen? Voor eerste antwoorden bespreek ik het boek *7 Rules of power* van Jeffrey Pfeffer. Hij geeft al decennia lang workshops over macht op gerenommeerde instituten als Stanford Graduate School of Business, Harvard Business School en London Business School. Ook heeft hij er meerdere boeken en artikelen over geschreven. *7 Rules of power* is zijn meest recente boek. Pfeffer wil juist nu – dat is: 2022 – een tegenwicht bieden tegen het idee dat we in tijden leven waarin alles verandert en dus ook oude opvattingen over macht niet meer relevant zijn. Volgens hem zijn de door hem gepresenteerde inzichten tijdloos, gelden ze niet alleen in de dominante witte-mannen-cultuur in de VS en zijn ze even relevant naar gender of kleur. Hij heeft even overwogen om het hele boek aan Trump op te hangen – 'Trump surely follows the seven rules of power I outline in this book'[8] – maar heeft dat uiteindelijk niet gedaan omdat Trump zulke heftige emoties oproept dat de lezer wellicht niet meer de regels op hun eigen merites kan beoordelen. Blijft staan dat de regels veel van Trumps succes duidelijk maken... Pfeffer geeft overigens aan dat macht op zichzelf meestal niet de kritieke succesfactor is voor het bereiken van doelen (je moet bijvoorbeeld ook over bepaalde kennis en vaardigheden beschikken om te presteren), maar dat macht dit als hefboom wel gemakkelijker maakt.

De zeven regels zijn:

1. Sta jezelf niet in de weg
2. Overtreed de regels
3. Kom krachtig over
4. Bouw een krachtige reputatie
5. Netwerk zonder ophouden
6. Gebruik je macht
7. Succes verontschuldigt bijna alles

Ik leg de zeven regels kort uit. Het navolgen van de regels biedt overigens geen garantie op succes, maar het vergroot wel de kans dat jij jouw doelen bereikt.

Sta jezelf niet in de weg

Vaak zijn wij zelf de grootste belemmering voor het verkrijgen van macht. Omdat we niet durven of willen. Omdat we bang zijn door de mand te vallen. Omdat we politieke spelletjes vies vinden: 'People who see power as evil or dirty may abjure power and be unwilling to "play the game"'.[9] Omdat we menen dat in een eerlijke wereld macht geen rol speelt. Om met het laatste te beginnen, Pfeffer is heel duidelijk: de wereld is niet eerlijk, mensen krijgen niet als vanzelf wat ze verdienen, daar moet je wat voor doen. Je moet voor jezelf opkomen. En dan niet bang zijn dat je door de mand valt, dat je tekortschiet. En als je het moeilijk vindt om positief over jezelf te denken, begin dan met positieve bijvoeglijk naamwoorden bij jezelf op te schrijven om zo jouw zelfbeeld krachti-

5. J. Pfeffer (2022). *7 Rules of power. Surprising – but true – advice on how to get things done and advance your career.* Swift.
6. Zie E. Karssing (2018). 'Uit de boekenkast van de bedrijfsethiek 70'. *Tijdschrift voor compliance.* 18. december, 398-405.
7. Vgl. M. Falconer (2022). *Book Review: The 48 Laws of Power by Robert Greene* mikefalconer.net/2022/03/19/book-review-the-48-laws-of-power-by-robert-greene/ en S. Joppich (2022). *Why Reading 'The 48 Laws of Power' Is a Huge Waste of Time* stephanjoppich.com/48-laws-of-power/
8. Pfeffer (2022), ibid.: xiv.
9. Pfeffer (2022), ibid.: 26.

ger te maken. 'Check with friends to see if your list is correct. Then ask yourself what descriptors you need to get rid of in order to project yourself in a more powerful way. Ask yourself what positive adjectives about yourself – language that gives credit to your accomplishments and credentials – you underutilize in your interactions with others'.[10] Bedenk: als jij al niet positief over jezelf denkt, dan zal een ander dat zeker niet doen. Je hoeft niet aan het politieke spel mee te doen, maar dat heeft een prijs: je zal minder succesvol zijn, je zal minder in staat zijn eigen doelen te bereiken. Want anderen spelen het spel wel. Pfeffer vindt daarom de discussie of politiek handelen in organisaties wel of niet wenselijk is niet relevant: het spel is er, je kunt wel of niet meedoen, en als je meedoet kun je maar beter jezelf goed voorbereiden.[11] Kortom, overwin je onzekerheden en durf kansen te grijpen. Moet je daarvoor je eigen persoonlijkheid geweld aandoen, je persoonlijkheid veranderen? Neen, de uitdaging is jouw vaardigheden versterken: 'Power skills and behaviors are just that – skills and behaviors that can be learned and practiced selectively as situations demand'.[12]

Overtreed de regels

Door je niet aan de regels te houden kun je anderen verrassen die dat wel doen. Je laat ook zien dat de regels voor jou niet gelden, dus (let op de 'dus') ben jij blijkbaar machtig. Met andere woorden, door regels te overtreden denken anderen dat jij wel machtig moet zijn. Pfeffer geeft aan dat dit waarschijnlijk één van de oorzaken is waarom Trump overal mee wegkomt: zijn gedrag is geen teken van boefheid maar van macht.[13] Ook geldt dat het gemakkelijker – en effectiever! – is om achteraf om vergeving te vragen dan vooraf om toestemming.[14] Er zijn natuurlijk grenzen aan het overtreden van regels. Zeker vanuit moreel oogpunt, maar ook vanuit effectiviteit: bepaalde regelovertredingen doen afbreuk aan jouw reputatie (zie hieronder ter illustratie een overzicht met politieke blunders).

Politieke blunders[15]
– Anderen in het openbaar vernederen
– Het schenden van de ethische code en gedragsnormen van de organisatie
– Ongebreidelde hebzucht, zelfs als grote sommen geld legaal worden verkregen
– Het versturen van negatieve berichten via zakelijke e-mails, websites en sociale media
– De baas omzeilen
– Vijandigheid tonen en wraak zoeken in een exit-interviewIndiscreet zijn in het privéleven
– Een ongepaste kantoorromance hebben

Pfeffer noemt een heel krachtig voorbeeld van een regel overtreden.[16] Hij geeft aan dat in onze Westerse cultuur zelfredzaamheid een belangrijke norm is en dat om hulp vragen een overtreding is van die norm. Veel mensen vermoeden ook dat hulpvragen worden geweigerd. In de praktijk pakt dit vaak heel anders uit en kom je met een hulpvraag dus een stuk verder. En wat is het ergste dat je kan overkomen? Dat je niet krijgt wat je vraagt … maar dat zou je zonder vragen ook niet krijgen.

Kom krachtig over

Wat je doet – in woord en gebaar – en hoe je er uitziet, spelen een grote rol in hoe anderen je zien en beoordelen. Je kunt jezelf dus beter wapenen voor de politieke arena door hier welbewust in te investeren. Zo stelde Henk van Luijk dertig jaar geleden als mijn eerste leidinggevende op Nyenrode dat je als bedrijfsethicus alleen met een Armani-pak en stropdas serieus wordt genomen. Bedenk dat mensen heel snel een eerste oordeel vormen en daarna vooral bezig zijn dit oordeel voor zichzelf te bevestigen (*confirmation bias*). Denk goed na over hoe jij je kleedt, welke woorden je gebruikt, welke gebaren. Pfeffer geeft het volgende rijtje van non-verbaal gedrag dat wijst op een hogere mate van macht, status en dominantie:

– Expressief door veel kleine gebaren
– Opener lichaamshouding
– Minder fysieke afstand (dichter bij anderen gaan staan)
– Meer gebruik van arm- en handbewegingen
– Luidere stem
– Meer succesvolle onderbrekingen van anderen
– Meer spreektijd
– Langer staren naar gesprekspartner
– Hogere visuele dominantieratio (kijken + praten > kijken + luisteren)
– Vaker ongeremd lachen[17]

Ook boosheid straalt macht uit. Ikzelf vind boosheid juist een teken van machteloosheid, van gebrek aan zelfbeheersing, maar dat geeft waarschijnlijk precies de kracht aan: ook boosheid is een overtreding van de norm (zie hierboven), dus een teken van onaangepastheid en dus een teken van macht… 'if the powerful are permitted to display anger more readily than the less powerful because – displays of anger fall outside customary norms for behavior, and only the more powerful are permitted to violate social expectations – then displays of anger can create perceptions of higher status'.[18]

Tot slot, leef je in, zorg dat je de doelgroep kent. Je kunt pas krachtig overkomen en indruk maken op een bepaald publiek als je weet wat het publiek verwacht, wat bij dat specifieke publiek krachtig en machtig overkomt.

Bouw een krachtige reputatie

Creëer een sterk persoonlijk merk en zorg ervoor dat mensen je associëren met succes en expertise.

10. Pfeffer (2022), ibid.: 22-23.
11. Pfeffer (2022), ibid.: 218, 221, 236-7.
12. Pfeffer (2022), ibid.: 36-37.
13. Pfeffer (2022), ibid.: 50.
14. Pfeffer (2022), ibid.: 53.
15. Dubrin, geciteerd in D. Buchanan en R. Badham (2020). *Power, politics, and organizational change.* 3rd ed. Sage: 270.
16. Pfeffer (2022), ibid.: 58-61.
17. Pfeffer (2022), ibid.: 81.
18. Pfeffer (2022), ibid.: 75.

Bijvoorbeeld door veel om te gaan met mensen die reeds een hoge status hebben: hierdoor straalt hun status op jou af. Zorg dat je een duidelijk verhaal hebt – in twee of drie zinnen – wie je bent, wat jouw expertise is en wat je wilt bereiken en vertel dit verhaal heel geregeld zodat mensen het onthouden. En het begint er natuurlijk allemaal mee dat mensen je überhaupt kennen. Maak jezelf dus kenbaar: via een eigen podcast, of schrijf een boek, schrijf geregeld een stukje op LinkedIn, spreek op conferenties of organiseer zelf bijeenkomsten. Wees niet bescheiden en laat anderen weten wat je hebt gedaan, wat je hebt bereikt. En ja, hier speelt een dilemma: iemand moet jouw verhaal vertellen anders zal het niet bekend zijn, maar als je dat zelf doet is het veel minder geloofwaardig: 'wij van WC-eend…'. Het beste is daarom als je anderen jouw verhaal kunt laten vertellen, zelfs als die anderen hierbij een belang hebben (omdat ze bijvoorbeeld jouw directe collega of medewerker zijn); uit onderzoek blijkt dat dit toch effectief is.[19]

Netwerk zonder ophouden

Volgens mij stond het ergens in mijn agenda als scholier, nu lang geleden: 'kennis is macht, kennissen machtiger'. Ik kan de bron dus niet meer achterhalen, maar ik zie op internet wel zinnen als: 'Bij een carrière komt het vaak niet aan op kennis, maar op kennissen' en 'Menigeen die het ver gebracht heeft, dankt dit meer aan zijn kennissen dan aan zijn kennis'.[20] De boodschap lijkt me helder: een goed netwerk is een belangrijke machtsbron. Maar heel veel mensen vinden netwerken niet leuk of zelfs immoreel (je denkt bij het opbouwen van een relatie immers alleen aan je eigen gewin). Het eerste kan natuurlijk waar zijn (afhankelijk van jouw eigen voorkeuren), maar het blijkt dat een goed netwerk heel belangrijk is, dus… als je het politieke spel wilt winnen dan moet jij je daar maar overheen zetten. En netwerken is helemaal niet per definitie immoreel: het doel dat je dient kan immers veel verder gaan dan eigen gewin en jezelf laten helpen is niet hetzelfde als uitbuiting. Je kunt elkaar immers helpen en daarmee wordt het een win-win-relatie. Pfeffer geeft ook aan dat jouw netwerkrelatie alleen maar sterker wordt als je juist ook voor die ander waarde creëert. Verder adviseert hij jouw sociaal kapitaal te versterken door als een makelaar mensen uit verschillende netwerken bij elkaar te brengen. En vooral te investeren in zwakke connecties ('weak ties'), mensen met wie je een hechte band hebt, bieden je vaak weinig nieuws. Je moet hierin wel efficiënt blijven, hoe belangrijk netwerken ook is, het blijft net werken, uiteindelijk moet het echte werk ook worden gedaan. Besef daarom dat je ook met kleine updates of het sturen van een interessant artikel de banden met 'weak ties' warm kunt houden.

Gebruik je macht

De regels van Pfeffer hebben vooral betrekking op het verwerven van macht; je moet deze macht ook gebruiken anders is het verspilling van alle tijd en energie die je hebt geïnvesteerd in jouw voorbereiding. Dat is echter niet wat Pfeffer bedoelt met deze regel. Hij stelt dat het gebruik van macht juist tot meer macht leidt: macht raakt niet op door gebruik; hoe meer je het gebruikt, hoe meer je ervan krijgt. Doordat je laat zien hoe machtig je al bent en dat je ook bereid bent gebruik te maken van die macht.[21] Doordat je jouw macht gebruikt om doelen te bereiken en daarmee jouw reputatie versterkt. En doordat je jouw macht kunt gebruiken om structurele machtsbronnen te versterken door bijvoorbeeld bepaalde regels of stemverhoudingen te veranderen.

Succes verontschuldigt bijna alles

Als je wint heb je vrienden/Rijen dik/Echte vrienden/Als je wint/Nooit meer eenzaam/Zolang je wint.

Dat zongen Herman Brood en Henny Vrienten in 1984 en ze hebben gelijk. Deze songtekst vat treffend samen hoe succes sociale acceptatie beïnvloedt. Mensen vergeten of vergeven vaak de acties die je hebt ondernomen om macht te verkrijgen, zolang je succesvol bent. En opnieuw is Trump anekdotisch bewijs voor deze stelling (naast al het echte wetenschappelijke bewijs). Hoe kan dat? 'People want to be close to money and power and are therefore willing either to forgive those who have them or avert their gaze from their possessors' misdeeds'.[22] Pfeffer benadrukt dat hij hiermee geen pleidooi houdt voor immoreel gedrag, maar 'if you are successful, rich, powerful, and have many powerful and rich friends – a great social network – then that success and those connections will likely protect you (notice I used the word *likely*, not *inevitably*) from falling from power and grace, almost *regardless* of what you do'.[23] In deel 1 van dit tweeluik heb ik de machtsparadox van Keltner besproken: wat nodig is om macht te verkrijgen – rekening houden met anderen – kun je zomaar kwijt raken op het moment dat je de macht hebt. We zien hier nog een verklaring voor deze paradox: als je eenmaal machtig bent zul je nauwelijks corrigerende negatieve gevolgen ondervinden van eventueel wangedrag.[24] Heel vervelend voor mensen die graag in een eerlijke wereld geloven en een belangrijk inzicht voor wie hecht aan ethiek. En een geruststelling voor de mensen die zich zorgen maken dat je op de weg naar macht vrienden kwijtraakt: dat valt reuze mee. Pfeffer noemt dit daarom de belangrijkste regel.

3. De ethiek van politiek handelen

Pfeffer benadrukt dat macht een neutraal instrument is dat zowel positief als negatief kan worden ingezet. Hij besteedt verder weinig aandacht aan ethische overwegingen. Aan het begin van zijn boek geeft hij aan dat hij zijn verhaal 'value-free' vertelt, net zoals iemand jou over de fysica van atoomener-

19. Buchanan en Badham (2020), ibid.: 160-161.
20. citaten.net/quotes/met-kennissen.html
21. Pfeffer (2022), ibid.: 137.
22. Pfeffer (2022), ibid.: 151.
23. Pfeffer (2022), ibid.: 156-157.
24. Pfeffer (2022), ibid.: 157.

gie kan vertellen: wat je met de kennis doet is helemaal aan jou.[25] Ik vind dat te gemakkelijk. Het is niet voor niks dat we tegenwoordig het woord 'oppenheimermoment' op betekenisvolle wijze gebruiken, als verwijzing naar het moment 'dat iemand zich realiseert dat hij een point of no return gepasseerd is in de ontwikkeling van een technologie die op termijn de mensheid kan vernietigen'.[26] Nu mag ik hopen dat gebruik van macht in organisaties, in tegenstelling tot een atoombom, niet meteen het voortbestaan van de mensheid op het spel zet. Maar gebruik van macht kan zeker ernstige schade toebrengen. We hebben gezien dat ethiek niet zonder macht kan, maar wat mij betreft macht ook niet zonder ethiek. Helaas geeft De Vaan in zijn boek *Machiavelli op de werkvloer* aan dat er geen 'ethisch handboek van het machtsspel' bestaat.[27] Hij stelt daarom voor om concrete praktijkvoorbeelden te bespreken: 'Door erover te blijven nadenken en er met anderen over te praten, scherp je je morele standaard telkens verder aan'.[28] Dat past goed bij mijn opvatting van beroepsethiek. Ethiek draait dan niet om morele verboden en geboden, om regels, maar is een zoektocht en leerproces waarbij beroepsbeoefenaren begrijpen dat ethiek en integriteit essentieel onderdeel zijn van hun vak en ze met elkaar de morele kaders binnen de context van hun werk onderzoeken. Dat geldt ook voor de ethiek van machtsgebruik, van politiek handelen: wat kan door de beugel en wat niet meer? Wat zijn legitieme zetten in het politieke spel, wanneer worden het rattenstreken of, in onvertaalbaar Engels: '*Do the ends justify the meanness?*'.[29] Dat zoeken en leren hoeft natuurlijk niet steeds helemaal opnieuw te beginnen, we kunnen gebruik maken van inzichten die eerdere leerprocessen hebben opgeleverd. In deel 1 van dit tweeluik gaf ik aan dat bijvoorbeeld Hetebrij in zijn boek *Macht en politiek in besluitvorming* verschillende criteria voor machtsgebruik heeft geformuleerd: doelmatigheid, rechtmatigheid, rechtvaardigheid en transparantie. In deze paragraaf laat ik zien hoe Buchanan en Badham in hun boek *Power, politics, and organizational change* de ethiek van politiek handelen oppakken.

Buchanan en Badham hebben een echt lesboek geschreven: heel grondig en degelijk, leest niet lekker weg, maar biedt wel een uitgebreid overzicht van heel veel conceptuele, theoretische en empirische inzichten rondom macht en politiek handelen. Ze weigeren een eenduidige definitie van macht te geven, omdat macht daarvoor te veelzijdig is en we vanuit verschillende perspectieven verschillende aspecten van macht kunnen zien. Ze geven wel een mooie beschrijving van politiek handelen: 'power in action'.[30] En ze benadrukken het belang van reputatie: 'Reputation is defined not just in terms of the opinions of others, or the regard in which you are held. It is also defined by what you have done and achieved, on your actions and outcomes'.[31] Reputatie is een belangrijke machtsbron die we moeten koesteren; want net als bij vertrouwen, het komt te voet en gaat te paard. Met het onderstrepen van het belang van reputatie agenderen ze nadrukkelijk het spanningsveld tussen kortetermijnsucces en langetermijnimpact. Een kortetermijn succes ten koste van jouw reputatie kan je op de lange termijn in de weg zitten. Want reputatie is heel belangrijk. Als jij bekendstaat als eerlijk, transparant of principieel (ook al ben je dat niet), zullen jouw machtsdaden sneller als legitiem of moreel worden gezien. Omgekeerd: als jij een reputatie hebt als manipulatief of machtsbelust, zullen zelfs jouw goedbedoelde acties eerder wantrouwen opwekken.

Buchanan en Badham formuleren eerst een ethiektoets, een beslisboom, waarmee je door het beantwoorden van enkele eenvoudige vragen kunt vaststellen of een bepaald gebruik van macht wel of niet door de beugel kan. De vragen verwijzen naar klassieke ethische theorieën als het utilitarisme (gevolgenethiek) en de deontologie (beginselethiek). De drie vragen zijn:

- Profiteert de meerderheid van de belanghebbenden van de uitkomst?
- Doet het gebruik van macht recht aan de individuele rechten van belanghebbenden?
- Wordt het machtsspel op een eerlijke en rechtvaardige manier gespeeld?

In principe moet je drie keer 'ja' kunnen zeggen en dan is het gebruik van macht moreel gelegitimeerd; bij een 'nee' kun je deze legitimatie alsnog verdienen door bijvoorbeeld te verwijzen naar verzachtende omstandigheden. Vervolgens illustreren ze de ethiek-toets aan de hand van enkele voorbeelden. Ze zijn buitengewoon teleurgesteld over hun eigen toets: 'There is a danger that "ethical" in these examples is being used to describe behaviour that is careless, amateurish, naïve, and insensitive if not incompetent, while contextual awareness, prudence, astuteness, and professionalism are being labeled as "unethical".'[32] Omdat de toets teveel is gebaseerd op wensdenken: er is een soort moreel ideaal – geformuleerd voor een ideale wereld – waaraan je moet voldoen en anders schiet je tekort. En omdat met deze ethiektoets de brug niet wordt geslagen tussen de ethische theorie en de praktijk, waardoor de toets theoretisch wellicht verdedigbaar is maar voor de praktijk niet bruikbaar en dus niet relevant. Ook vinden ze, niet verrassend na hun nadruk op het belang van reputatie, dat met het beoordelen van een concrete handeling teveel naar die handeling als geheel op zichzelf staand wordt gekeken, terwijl juist ook de effecten op de reputatie zwaar zouden moeten wegen. Ze pleiten daarom voor een pragmatische aanpak waarin zorgvuldig en uitlegbare afwegingen worden gemaakt die recht doen aan concrete

25. Pfeffer (2022), ibid.: 4.
26. www.taalbank.nl/2023/08/07/oppenheimermoment/
27. M. de Vaan (2021). *Machiavelli op de werkvloer. Politieke vaardigheden voor managers.* Boom: 213.
28. De Vaan (2021), ibid.: 214-215.
29. Buchanan en Badham (2020), ibid.: 68.
30. Buchanan en Badham (2020), ibid.: 3.
31. Buchanan en Badham (2020), ibid.: 27.
32. Buchanan en Badham (2020), ibid.: 74.

situaties, aan de context waarin die situaties spelen en met oog voor reputatie-effecten op de lange termijn. Dit betekent dat een beslisboom helaas niet mogelijk is, dat is een te grove versimpeling om recht te doen aan de vele aspecten die relevant kunnen zijn; zeker ook omdat op voorhand niet altijd meteen duidelijk is welke aspecten relevant zijn. Een conclusie die ik deel en overigens niet alleen ten aanzien van morele vraagstukken rondom gebruik van macht; ik vermoed dat slechts weinig morele vraagstukken met een beslisboom zijn op te lossen. Of ze zijn zo eenvoudig dat een beslisboom eigenlijk niet nodig is. Kortom, morele vraagstukken vereisen nuance – een simpele beslisboom doet geen recht aan de complexiteit ervan.

We moeten dus in een zoektocht en leerproces tot een zorgvuldige en uitlegbare uitkomst komen. En dat begint met heel goed naar de concrete situatie te kijken binnen de context waarin het machtsspel wordt gespeeld. En dan speelt heel veel mee. We staan gelukkig niet met lege handen. Buchanan en Badham formuleren verschillende vragen, deels geïnformeerd door klassieke ethische theorieën.

- De intentie van degene die politiek handelt
  - Wat is het doel of de motivatie van degene die macht uitoefent? Wordt macht gebruikt om het algemeen belang te dienen of slechts het eigenbelang?
- De middelen die worden ingezet
  - Worden middelen open en eerlijk gebruikt of ligt de nadruk op manipulatie, misleiding en dwang? Is de inzet van macht proportioneel, dus gegeven het doel niet te veel en niet te weinig?
- De gevolgen van het handelen
  - Welke effecten heeft het machtsgebruik op anderen en de organisatie als geheel? Wordt schade toegebracht, of draagt het juist bij aan groei en rechtvaardigheid? Draagt het gedrag bij aan het behalen van organisatiedoelen zonder onnodige schade te veroorzaken?

Vervolgens introduceren Buchanan en Badham drie aandachtspunten die ervoor zorgen dat juist ook hun nadruk op langetermijn reputatie-effecten recht wordt gedaan: 'warrant', 'account' en reputatie. Daarbij wil ik overigens meteen benadrukken dat de eerste twee aandachtspunten wat mij betreft altijd belangrijk zijn: Een 'warrant' is een argument, een rechtvaardiging, voor het gebruik van macht. Het maakt politiek handelen uitlegbaar: het zijn verwijzingen naar goede redenen om de macht te gebruiken. Meestal verwijst een 'warrant' naar omstandigheden, intenties of hogere doelen. Dit kan bijvoorbeeld heel formeel, door te verwijzen naar iemands functieomschrijving ('ik ben hier de manager') maar ook moreel door te verwijzen naar kernwaarden, business principles of andere morele uitgangspunten. Een 'account' is de uitleg zelf, de verantwoording, waarbij iemand uitlegt of verdedigt waarom hij of zij een bepaalde handeling heeft verricht. Met een 'account' kun je achteraf begrip of acceptatie creëren voor gedrag dat als discutabel wordt ervaren. Een duidelijke 'warrant' vergemakkelijkt de verantwoording van het handelen en versterkt de legitimiteit ervan. Zowel 'warrant' als 'account' dragen bij aan iemands reputatie van geloofwaardigheid en betrouwbaarheid – is deze persoon te vertrouwen? Maar Buchanan en Badham benadrukken dat het effect van jouw handelen op je reputatie tevens een zelfstandige overweging is: je kunt niet altijd met een goed verhaal reputatieschade minimaliseren.

Helaas is er dus geen beslisboom, geen eenvoudige ethiektoets om politiek handelen te beoordelen. Ik moet dus De Vaan gelijk geven: belangrijk is om met elkaar concrete praktijkvoorbeelden te bespreken. Om je eigen morele standaard telkens aan te scherpen. Maar ook om met elkaar te leren wat belangrijke aandachtspunten zijn, welke vragen helpen om beter inzicht in de concrete situatie te krijgen en wat morele ondergrenzen zijn. Ik heb dergelijke oefeningen tijdens workshops gedaan. Het mooiste antwoord dat ik kreeg op de vraag welke criteria we kunnen gebruiken om politiek handelen te evalueren: 'mensen willen voor me werken'.

4. Tot slot

Heel lang was mijn belangstelling voor macht en machtsgebruik zeer beperkt. Natuurlijk, macht was geregeld een aanleiding voor ethische reflectie: als grondslag van verantwoordelijkheid, als oorzaak van immoreel gedrag, als onderwerp van aanklacht (machtsmisbruik). Maar macht als zelfstandig onderwerp van studie vond ik weinig interessant. In die zin heeft de literatuurstudie voor dit tweeluik mijn ogen geopend: macht is als fenomeen buitengewoon fascinerend, een heel rijk onderwerp dat vanuit heel veel verschillende invalshoeken kan worden bestudeerd. Ik ben ook zeker niet uitputtend geweest: bijvoorbeeld geen aandacht voor de machtstheorie van Foucault en geen aandacht voor institutioneel onrecht door structurele machtsongelijkheden (bijvoorbeeld naar gender, ras, of klasse).[33]

Er zijn heel veel verschillende definities van macht, een goed begin vind ik, a la Rutte, om macht heel neutraal te zien als de mogelijkheid om iets voor elkaar te krijgen. Mij staat daardoor veel helderder nut en noodzaak voor ogen van macht voor ethiek: zonder macht geen ethiek. En een diepgaander besef dat macht niet alleen nodig is, maar dat het machtsspel ook onvermijdelijk is. Dat je er wel voor kunt weglopen, maar dat je dan jezelf irrelevant maakt (en dus de doelen tekortdoet waarvoor je staat, waarvoor je gaat). Dan kun je maar beter leren hoe je goed voorbereid de arena inloopt en op vaardige wijze het machtsspel speelt. Een spel dat ook vanuit de ethiek kan worden bevraagd: de ethiek van macht. In de vo-

33. Buchanan en Badham (2020) besteden zeker wel aandacht aan Foucault, maar eerder als 'nabrander' dan als bouwsteen van hun eigen verhaal.

rige paragraaf heb ik aangegeven hoe we met elkaar in zoektocht en leerproces hieraan vorm en inhoud kunnen geven. Macht hoeft niet gelijkgesteld te worden aan eigenbelang ten koste van anderen.[34]

Pfeffer benadrukt dat alleen een literatuurstudie niet volstaat om je te bekwamen in het machtsspel. Daarvoor is oefening nodig, en ervaring in de praktijk. Door te experimenteren, met vallen en opstaan. En dan kan het helpen om lijstjes te maken met goede voornemens. Want sommige noodzakelijke stappen botsen met onze natuurlijk neiging om aardig te worden gevonden, onze twijfels ('Kan ik dit wel?') en onzekerheden ('Val ik door de mand?'). Ze botsen soms met de lessen uit onze opvoeding en scholing. Dan hebben we 'reminders and constant vigilance' nodig, om ons bij de les te houden.[35]
En bedenk, je hoeft het niet alleen te doen. En dan bedoel ik niet dat in het machtsspel bondgenoten jou kunnen helpen. Dat is zeker ook waar, maar je kunt politieke vaardigheden versterken door intervisie. Je kunt in kleine groepjes met elkaar reflecteren op situaties en ervaringen. Hoe kun je machtsbronnen verwerven, hoe kun je jouw machtsbronnen versterken? Wat kun je doen, welk handelingsrepertoire heb je tot jouw beschikking? Wat werkt en wat werkt niet? Is dit nog te verantwoorden of zak je door een morele ondergrens?

En ben je dan klaar om de politieke arena te betreden? De Vaan geeft in zijn boek *Machiavelli op de werkvloer* een laatste toets waarmee je eerst nog alles kunt nalopen voordat het spel begint.[36]

- 'Leidt mijn strategie tot succes?
  - Zal het spel verlopen zoals ik dat heb bedacht, en gaat het opleveren wat ik wil? Heb ik met alle denkbare aspecten rekening gehouden? Ken ik de belangen van anderen?
- Wat zijn de gevolgen voor mijn reputatie?
  - Wat gebeurt er met de beeldvorming? Zullen anderen me na mijn spel zien als iemand die "dingen voor elkaar krijgt" of als iemand die "lastige dingen omzeilt"?
- Past het spel bij mijn principes, mijn waarden en de dingen waarin ik geloof?
  - Waar trek je je grens, welke acties kunnen en welke niet? Hier past een waarschuwing: wees niet te streng en rigide in je begrenzing, enige rekkelijkheid is het politieke spel eigen. Bedenk ook wat je met je acties voor anderen of de organisatie kunt bereiken.
- Wat is mijn terugvaloptie?
  - Probeer alternatieve strategieën achter de hand te hebben voor als je spel geen of onvoldoende effect heeft. Overweeg net als bij het schaakspel verschillende opties bij mogelijke zetten van de ander.'

Tot slot. Ik heb aangegeven dat compliance & ethics professionals macht kunnen en zelfs moeten gebruiken, en daarvoor ook hun politieke vaardigheden kunnen en moeten vergroten. En dat ze kunnen reflecteren op de ethiek van macht en machtsgebruik. In tijden waarin sociale veiligheid steeds hoger op de agenda staat, wordt machtsmisbruik door managers steeds scherper veroordeeld. Dat is vanuit moreel oogpunt heel fijn, vanuit reputatieperspectief een risico. Compliance & ethics professionals kunnen een bijdrage leveren aan sociale veiligheid door de ethiek van macht in workshops met managers te agenderen. Dat is meteen een manier om eens op een andere wijze met elkaar te onderzoeken wat de kernwaarden van de organisatie kunnen betekenen voor gedrag op de werkvloer. En een workshop over macht is voor managers op zichzelf al interessant, want zo kunnen ze effectiever hun werk doen.[37] Dat klinkt als een win-win.

34. Vgl. Buchanan en Badham (2020), ibid.: 2, 7-13.
35. Pfeffer (2022), ibid.: 174.
36. De Vaan (2021), ibid.: 207; zie ook Buchanan en Badham (2020), ibid.: 257.
37. Het lesboek van Buchanan en Badham (2020) eindigt ieder hoofdstuk met oefeningen; hiermee heb je al heel veel mooie bouwstenen voor een workshop.

ca. 240 pagina's

€ 65.-*

In dit boek treft u een bundeling van publicaties en annotaties die betrekking hebben op de zorgplicht die op financiële ondernemingen rust. Deze zorgplicht bepaalt in toenemende mate het verdienmodel van financiële ondernemingen nu de zorgplicht steeds verdergaande eisen stelt aan de (kwaliteit) van de dienstverlening aan (zowel particuliere als zakelijke) klanten.

Deze zorgplicht komt in vier fasen naar voren. Iedere fase wordt in deze uitgave beschreven.

De eerste fase bestaat uit de productontwikkelingsfase. Dit is wellicht de meest cruciale fase aangezien in deze fase de productenkenmerken, wijze van distributie en de wijze waarop het product aan het publiek wordt gepresenteerd, wordt bepaald.

De tweede fase ziet op de distributie van het financieel product. De distributie dient aan te sluiten op de aard en complexiteit van het product en de doelgroep waarvoor het product is bestemd.

De derde fase ziet op de communicatie met de klant zowel in de precontractuele fase als gedurende de looptijd (de zogenaamde nazorg).

De vierde en laatste fase betreft de wijze waarop toezichthouders de naleving van de publiekrechtelijke gedragsregels (informeel) handhaven. Daarbij is van belang dat toezichthouders in de praktijk verlangen dat financiële ondernemingen bepaald gedrag tonen dat door hen als wenselijk wordt beschouwd maar niet per definitie wettelijk voorgeschreven is. Ook daarmee dienen financiële ondernemingen rekening te houden.

Dit boek beoogt niet een uitputtend maar wel een breed en opiniërend beeld te geven van de ontwikkeling van de zorgplicht en actuele thema's op dat gebied.

Redactie:
Dr. mr. F.M.A 't Hart

* De genoemde prijs is exclusief btw, exclusief verzendkosten en onder voorbehoud van wijzigingen.
ISBN: 978-90-77847-114
Bestellen via de website https://denhollander.info.

DEN HOLLANDER

Postbus 325 | 7400 AH Deventer | T 0570 - 75 12 25
info@denhollander.info | **https://denhollander.info**

Tijdschrift voor Levensmiddelenrecht

Het Tijdschrift voor Levensmiddelenrecht (TvL) is een onmisbaar tijdschrift voor professionals die op de hoogte willen blijven van belangrijke ontwikkelingen op dit rechtsgebied en zich willen bekwamen in hoe recht en beleid inwerken op de agri-food keten.

Het tijdschrift is een belangrijke bron van informatie voor degenen die zich bezighouden met regulatory affairs binnen de voedselindustrie, beleidsmakers en toezichthouders of consultants, advocaten, rechterlijkemacht en wetenschap.

NIEUW!

Een greep uit de onderwerpen die in TvL aan de orde komen:

- Marktordeningsrecht en mededingingsrecht gericht op de (beleids)praktijk,
- Voedselzekerheid en -veiligheid,
- Handel in levensmiddelen en veiligheidsstandaarden op nationaal en mondiaal niveau,
- Controle en handhaving door de overheid,
- Hoe bevoegdheden van de toezichthouders zich verhouden tot de verantwoordelijkheden van de levensmiddelenexploitant,
- Europese integratie en nationale regelgeving,
- De juridische positie van levensmiddelenproducenten en -exploitanten,
- Etikettering, claims, novel foods en duurzaamheid, toelating van pesticiden, diergeneesmiddelen en marktordening. Maar ook vraagstukken over begrenzing, zoals het onderscheid tussen levens- en geneesmiddelen en de praktische betekenis van het voorzorgsbeginsel,
- Duurzaamheid en de Green Deal,
- Levensmiddelenrecht in de praktijk en welke lessen die daaruit kunnen worden getrokken,
- Voedselveiligheidscrises en de (juridische) gevolgen daarvan.

Zie voor abonnement mogelijkheden: https://denhollander.info/

Het Tijdschrift inclusief archief is eveneens digitaal te raadplegen via content integrators Legal Intelligence en Rechtsorde BV.

Abonnement
U kunt zich abonneren via de website:
https://denhollander.info/

DEN HOLLANDER

Boomkampsweg 1 • 7245 WC Laren • info@denhollander.info • www.denhollander.info

Tijdschrift voor Kapitaalmarktenrecht

Binnenkort verschijnt het Tijdschrift voor Kapitaalmarktrecht (KMR) met artikelen over de ontwikkelingen op dat gebied doormiddel van voor de rechtspraktijk relevante artikelen, annotaties en korte opiniërende columns.

KMR richt zich als eerste Nederlandse tijdschrift uitsluitend op kapitaalmarktenrecht:
(i) publiekrechtelijke regulering van kapitaalmarkten (financieel toezichtrecht); en
(ii) civielrechtelijke onderwerpen met betrekking tot kapitaalmarkten

Voor wie

- Bedrijfsjuristen (met name bij beursgenoteerde ondernemingen)
- Wetgevingsjuristen
- Advocatuur
- Financiële instellingen waaronder:
  - *Banken*
  - *Beleggingsondernemingen*
  - *Handelsplatformen*
  - *Beleggingsanalisten*
  - *Credit Rating Agencies*
  - *ESG Rating Providers*
  - *Benchmark Administrators*
  - *Proxy Advisers*
  - *Clearinginstellingen & CCP's*
  - *Custodians*
- Toezichthouders
- (Register-) Accountants
- Universiteiten

Inhoud

Het focusgebied van KMR valt uiteen in de volgende hoofd- en subcategorieën:

- Algemeen
  - *Kapitaalmarktunie*
  - *Economische ratio & doelstellingen toezichtregels*
  - *Regelgeving- en toezichtstructuur*
  - *Transactietypen (zoals aandelen-emissies, obligatie-emissies, beursgang, securitisations)*
  - *Financiële instrumenten*
  - *Duurzaamheid (ESG)*
  - *Digitalisering*
- Openbaarmaking informatie
  - *Prospectus*
  - *Periodieke publicatieverplichtingen*
  - *Openbaarmaking voorwetenschap*
  - *Melding zeggenschap*
- Handel & afwikkeling
  - *Handelsplatformen*
  - *Handelsregels (inclusief transparantieregels en handelsverplichtingen)*
  - *Short Selling*
  - *Algoritmische handel & HFT*
  - *Clearing & Settlement*
- Markmisbruik
  - Handel met voorwetenschap
  - Marktmanipulatie
- Tussenpersonen & gatekeepers
  - *Beleggingsondernemingen*
  - *Beleggingsanalisten*
  - *Credit Rating Agencies*
  - *ESG Rating Providers*
  - *Benchmark Administrators*
  - *Proxy Advisers*
  - *Clearinginstellingen & CCP's*
  - *Custodians*
- Civielrechtelijke aspecten
  - *Contractuele en vennootschappelijke aspecten financiële instrumenten*
  - *Administratie & bewaring effecten*
  - *Giraal effectenverkeer*
  - *Aansprakelijkheidsvraagstukken*
  - *Civielrechtelijke gevolgen overtreding toezichtrecht*

Meer informatie: https://denhollander.info/kapitaalmarktenrecht

DEN HOLLANDER

info@denhollander.info • www.denhollander.info