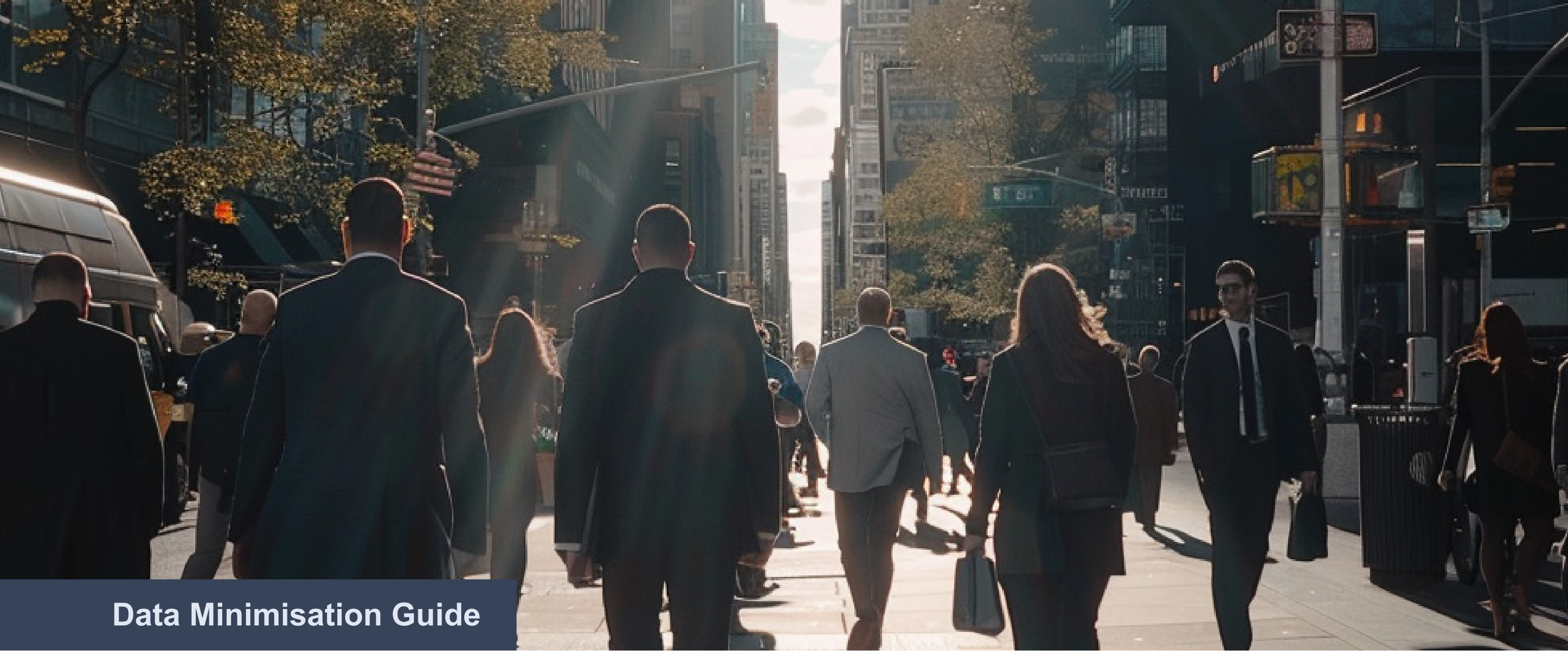


Data Minimisation Guide

The Essential Guide to Data Minimisation Best Practices for Information Professionals

Laws are changing around the world to require frequent disposal of high- risk information, to reduce the impact of (inevitable) breaches. Records and information professionals have become essential to cyber and privacy, because a key way to achieve data minimisation is with compliant disposition. Records lifecycle management has been a complex challenge, but now with Explainable AI, organisations can overcome the complexity of managing data at scale.

To address the criticality of information governance, and the growing challenges of managing data at scale, AI has entered the fray over the last five years, to automatically classify records against risk and retention rules with the aim of minimising personal and sensitive data holdings. But there are pitfalls to be aware of when considering AI for privacy and other data classification, not least of which relate to new legislation for Ethical AI. This guide provides a summary of the requirements for data lifecycle management, the technology approaches, and the risks. It also provides access to a Data Minimisation Best Practice Checklist for information governance professionals.

Data Minimisation – Why Less is More

Data minimisation is a security and privacy principle that requires organisations to limit the amount of sensitive information they hold, knowing that data in their systems could be breached or spilled at any time. Data minimisation is an approach to reduce the impact of data breaches and spills.

If, or more likely when, we experience a data spill, we want the spill to be as small as it can be. And we also want the size of the spill to be defensible – the data we were holding needs to be only what we needed, and nothing we couldn't justify holding onto.

To achieve data minimisation, we need to understand all our information holdings. We need to know what has risk, and what has value, and importantly what rules apply to that information (including retention rules, secrecy provisions, and other regulatory obligations).

Risk Management - Likelihood vs Impact

Risk is a combination of likelihood and impact. A risk may have a low *likelihood*, but a catastrophic impact if realised. And in that case, it's not a low risk. To date, most information security technology has focused on likelihood. However, we know we can't reduce likelihood to nil. There is always a zero-day exploit. Always a trusted insider, a potential misconfiguration, an Advanced Persistent Threat. There can never be zero likelihood of a breach or spill.

And there can be catastrophic impact, for national security, as well as for your stakeholders and clients.

Your Data Minimisation Obligations

The regulatory environment is rapidly changing with legislation around privacy and data protection evolving to not only protect consumers, but also strengthen national security in a worsening global threat environment. Organisations that do not take necessary steps to protect the data they hold may face heavy fines and legal action, as well as loss of stakeholder confidence and reputation.

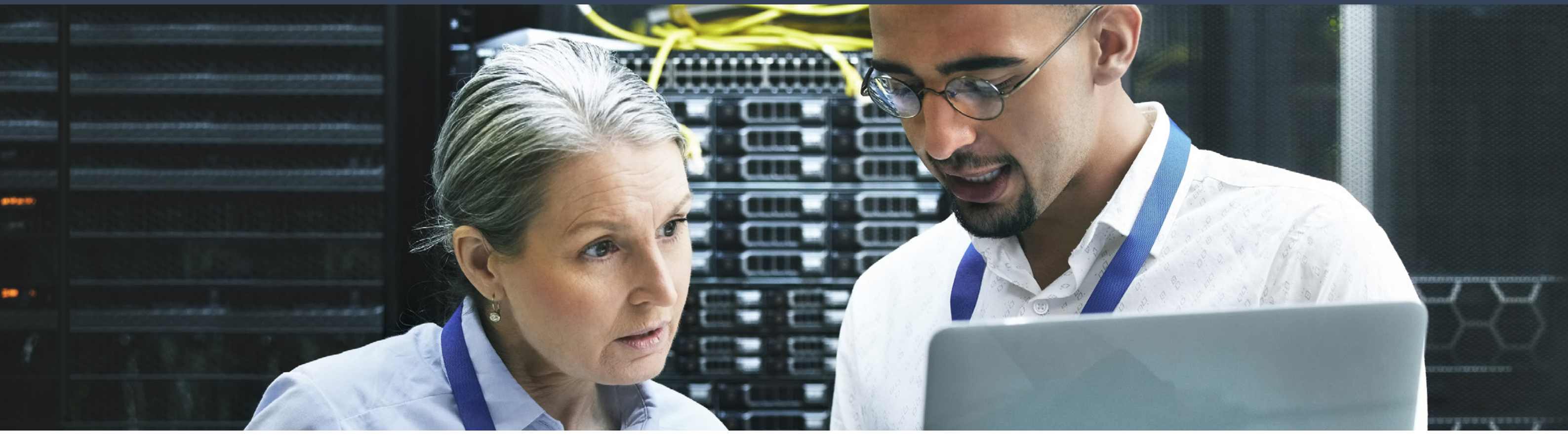
1. Privacy Obligations

Data Minimisation is a key GDPR privacy Principle, and is also a requirement under the national legislation like the UK *Data Protection Act* and NCSC 'bulk personal data' guidance, the *Australian Consumer Data Right*, and the US Consumer Data Privacy Laws such as the CPRA and VCDPA. The principle requires that you ensure that personal data you hold is limited to what is necessary – that is, that you do not hold more than you need for the stated purpose.

In practice, your organisation must only collect data sufficient to the purpose, and must periodically review and delete data it no longer needs for that purpose.

There are three key rules for minimisation:

1. **Don't over-retain**. Implement a timely and comprehensive retention and destruction function – 'destruction' can be permanent deletion or effective anonymisation.
2. **Apply records management**. Implement explainable and traceable methods for determining your retention and destruction policy, and a way to enforce right to be forgotten: finding every mention of a person, reviewing that content to see who or what else it contains, then, either with or without handing it over to them, destroying it irrecoverably (and being able to demonstrate that you have done so).
3. **Have a policy to follow**. Ensure the data retention policy is clearly documented, approved, and enforced. Information governance should be applied in the software development lifecycle, the product roadmap, the technology procurement planning and supply chain management, and the technical and architectural governance and oversight.



Records Management Obligations

Getting the balance right between retention and disposal can be very challenging, but it essential for all organisations. Legislation governing retention of data, such as the UK *Public Records* Act or Australian Archives Act and other statutory instruments are designed to protect the national interest, and the interests of your stakeholders. Keeping records for a legal minimum period helps make sure customers and staff can access justice if they need to take legal action in future. Retaining records also ensures your organisation can be audited and held accountable. But recordkeeping is not just for the benefit of the State and your stakeholders – it's for your organisation's benefit as well. Knowing how long data will be adding value to the business helps make sure you don't dispose of it precipitously.

You have an obligation to preserve records, as well as to destroy them compliantly. Getting the balance right can be very challenging, but it essential for all organisations.

National Security Obligations

As an information governance function, you also have a positive obligation to make life difficult for bad actors. Directives such as the UK *Networks and Information Systems Directive*, the US *National Security Memorandum on Critical Infrastructure Security and Resilience*, or the *Australian Security of Critical Infrastructure Act* apply to energy, transport, water, healthcare, and Relevant Digital Services Providers.

While these more formal Frameworks focus on critical and essential services, the core guidance is intended to apply broadly to all companies and organisations, and essentially states that you should decommission any information that is no longer used or that can't be linked to a business need.

Even if you are not technically in the scope of a national security Directive or supporting instrument, your Directors have defined responsibilities under laws such as the *Companies Act* and must act in the company's best interests to promote its success. This includes minimising detrimental impacts on your stakeholders, and protecting the company's reputation.

Data minimisation, even if not specifically mentioned, is key to both outcomes. The reason why is because data minimisation reduces bad actor success.

Other Benefits of Data Minimisation

Deterrence: Data minimisation obviously reduces the potential amount of harm when data is spilled. But, it also helps discourage further attempts on the data.

Ensuring sensitive data doesn't stay in the network any longer than it must by law means that any spills are smaller (and therefore less monetisable). When a hacker does not find commercial quantities of sensitive information, they will be less likely to attack again.

Response and Recovery: To do data minimisation, you must first be able to identify sensitive data at scale, and classify it. This means registering every data item, reading it, extracting its terms and topics, and classifying it against security, and privacy, and, vitally, records retention rules. You must know your own data at a high degree of fidelity to be able contact affected parties in a breach. Once you know your data, you can remonetise it and make more informed decisions.

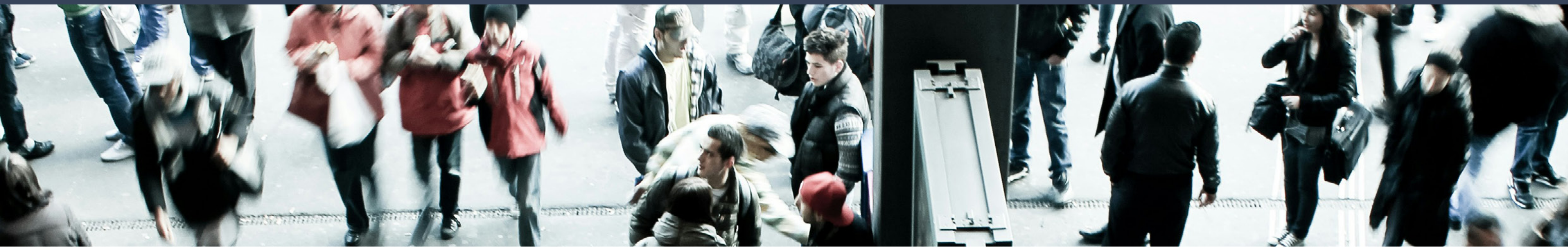
Transfer of Risk: Another tangential benefit of data minimisation is in your insurability, and transfer of risk. Cyber insurance providers tend to be opaque about how they assess cyber risk, and how they subsequently decide to cover you (and how much they charge). But we do know that if you hold a large amount of sensitive and high-value data, and especially if you don't have an effective retention policy, you may be considered 'high hazard' and pay higher premiums.

How do we put this knowledge into practice?

Having covered the importance of data minimisation and your obligations, let's talk about the lifecycle from data creation or capture through to eventual disposal, and everything in between.

Data Minimisation in the Information Lifecycle

- **Minimising collection:** minimise the amount of data you are collecting in the first place. This doesn't just mean not collecting extraneous personal details. It also means making sure you aren't keeping duplicates, caches, offline copies on devices, or excessive backups. It means not asking for the same data twice, as part of two separate functions or two different software applications.
- **Minimising access:** minimise the number of people with access, their privileges, and the duration of their access. Again, this usually comes down to governance and process, but information managers can also take steps to drive change. Seeing who is doing what to data, and being alerted to actions on sensitive and high-value data, helps to catch and kill malign or risky behaviour and privilege creep.
- **Minimising retention:** At the end of lifecycle management, there's more to disposal than just deleting things. There is a big process and governance role to play in managing records policies and retention rules. Applying your retention policies to all your data, in structured as well as unstructured systems, means you can meet your positive obligations to destroy data when reference ceases.



Technology Approaches to Data Minimisation

How can you implement these processes using modern technology?

Autoclassification: Autoclassification is the use of computers to determine what something is about, without needing to be told. Not just what it is, but what (and whom) it discusses, and what that means for organisational risk and reuse purposes.

Castlepoint pioneered 'Regulation as Code' for classifying data ethically, efficiently, and effectively, and without impacts on users or the governance team. Our Explainable AI determines how to classify items and records for risk, value, and regulatory obligations without needing a rules engine, supervised ML, or a file plan, because it matches the record contents directly to the applicable rules, frameworks, and policies.

> ***Castlepoint pioneered 'Regulation as Code' for classifying data ethically, efficiently, and effectively, and disrupted the first wave of automatic classification.***

Automated Decision Making (ADM): This is the process of speeding up outcomes by taking some of the human intervention out of the equation. ADM may or may not use AI.

With something as risky as data governance, it's very important not to take people completely out of the loop. Castlepoint uses ADM in a decision-support model, doing the work of collating and presenting, in a traceable way, the evidence that a human will need to make an informed decision. ADM has to be implemented carefully and compliantly.

AI and ADM Risks in Data Minimisation

These are very good outcomes possible when using AI and ADM, but it's important to consider and model what can go wrong with these technologies, as they are not always a panacea.

AI and ADM are essential technologies for managing data risk and value at scale. But there are some key risks inherent:

- AI can be used maliciously, on purpose
- AI can accidentally apply unwanted bias
- AI and ADM can inadvertently create erroneous outputs
- AI and ADM systems can breach people's privacy and security
- And if there's an issue with an AI or ADM outcome, there may be no clear line of accountability, and no recourse for stakeholders

These standards were designed to protect the most vulnerable in our communities from bias, disadvantage, and harm caused by adverse outcomes. They essentially require any decisions arrived at using AI and ADM to be explainable, and transparent, so they can be challenged if they are unfair.

Ethical AI Obligations

Because of the likelihood and impacts of these types of harm, obligations are mounting for Ethical AI.

We have had best practices for AI for a long time. Some include:

- **Recommendation on the Ethics of Artificial Intelligence (UNESCO):** This states that AI systems should be auditable and traceable, and developed with Transparency and Explainability (T&E). AI systems must not displace ultimate human responsibility and accountability.
- **OECD AI Principles:** This states that AI Actors should commit to transparency and responsible disclosure regarding AI systems, to enable those affected by an AI system to understand the outcome, and, to enable those adversely affected by an AI system to challenge its outcome based on plain and easy-to-understand information on the factors, and the logic that served as the basis for the prediction, recommendation or decision.
- **G20 AI Principles:** The G20 AI principles draw from and agree with the OECD principles and recommendations.

The EU has led the way globally with its *Ethical AI Act*, which makes AI explainability mandatory. The *Act* regulates AI data quality, transparency, human oversight, and accountability, based on the risk classification of the AI system. *GDPR Article 25* also states that an algorithmic-based decision which produces legal effects or significantly affects the data subject may not be based solely on automated processing of data, and that Individuals have the right to contest it. In other jurisdictions around the world, various regulations and frameworks apply or are coming into force.

Visit Castlepoint's AI Explainability information hub for more information on explainable and ethical AI and download the checklist for assessing technologies for explainability.

To Wrap it Up

Understanding what data minimisation is, what your obligations are, and what policy, process, and technology approaches you can take to achieve compliance and success is an important first step.

As an information governance professional you play a key role in strengthening your organisation's cyber security. It is important for you to understand the threat environment, and the internal governance structures for minimisation policy and process.

You must know what rules apply to your information (including retention, secrecy provisions, and other regulatory obligations). While artificial intelligence helps you understand your data at scale, implementing autoclassification and automated decision making requires consideration of risks (supportability, sustainability, and ethics). In case of a data breach, ignorance is not a defence. Laws are changing around the world around privacy and data protection not only to protect people, but also in the interest of national security. Therefore, it is critical to implement data minimisation best practices early and ensure they are being maintained in line with evolving threats.

For further inquiries on how Castlepoint can help you, please contact us here.

Visit www.castlepoint.systems for more information.

Control your risk, command your data

Your data estate is full of unseen risk and unrealised value. Castlepoint's explainable AI informs you about every record in every system.

Discovery, Cyber Security, Privacy Management, Audit and Assurance, Records Management, Generative AI Governance.

Did You Know We Also Have an Actionable Data Minimisation Best Practices Checklist?

As information governance becomes a key part of cybersecurity, it is critical for you to be able to design, implement and maintain effective data minimisation best practices. We have developed a complete checklist of best practices you can leverage you ensure you are meeting your data obligations and minimising data compliantly. You can download the full checklist via this link or the QR code. You can also discover how other organisations are identifying and protecting their sensitive and high-risk data.

About Castlepoint

Castlepoint Systems' powerful AI-driven platform is the intersection between cybersecurity and data governance. It is the only solution using ethical AI to unlock information management, eDiscovery, governance, risk, and compliance for the fastest, easiest way to find, manage, protect, and de-risk all an organisation's information in a single interface. Founded in 2012 in Australia, the multi-award-winning solution allows governance and risk teams to have a complete view across the whole environment and optimise investment in security, eDiscovery, and compliance processes. With offices in Australia, Europe and the USA, Castlepoint's scalable and transparent AI enables organisations to safeguard communities and people from data mismanagement and risk.

Follow Us:

in castlepointsystems