



Cyber, Privacy + Data Security

NZILA 2023

Gerard Ward

Objectives

- It is crucial for organisations to plan for breaches.
- Emergent tools for cataloguing data assets are termed: *Privacy tools*.
- This session discusses:
 - Trends in Privacy Tools and impact on security.
 - Us of emergent Artificial Intelligence (AI) technologies.
 - Challenges of the NZ reporting regime viz-a-vie other countries.
 - Maturity of the Privacy Tools, and how they will shape reporting.
- My focus is the technical trends, Joseph will frame the legal consequences.
- *Context:* by 2026, fines for the mismanagement of subject rights are projected to be over USD 1 billion (Gartner 2022).

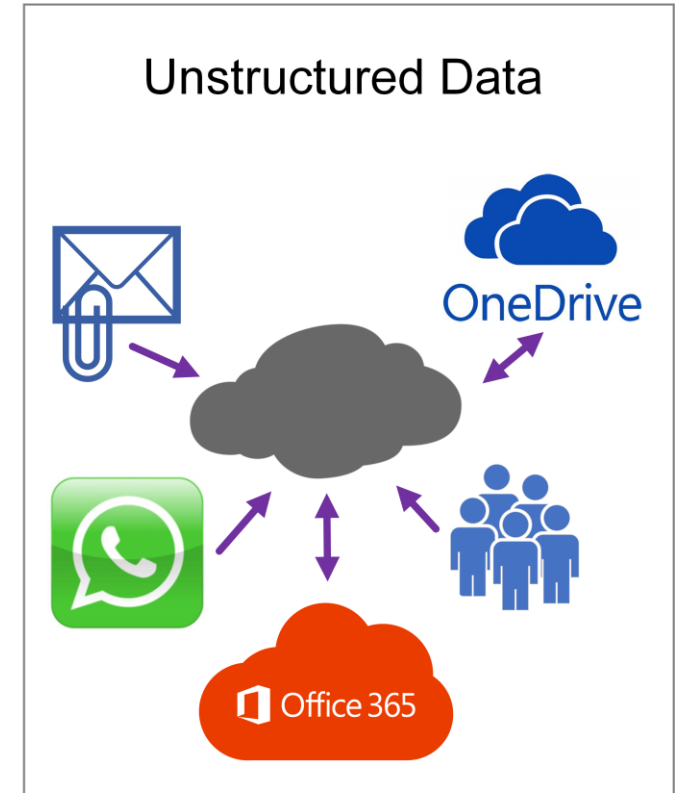
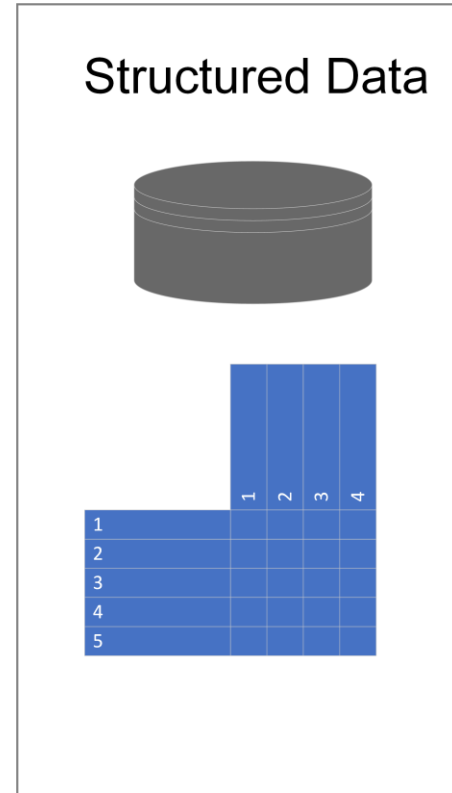
Context: *Cyber security stock-take exposes gaps**

- Published ***5 July 2023**
- The *Australian Prudential Regulation Authority* (APRA) assessed ~24% of financial institutions against its *CPS 234* standard.
 - Tests regulated entities cybersecurity defenses.
- The most common control gaps identified were:
 - **Incomplete identification and classification for critical and sensitive information assets;**
 - **Limited assessment of third-party information security capability;**
 - **Inadequate definition and execution of control testing programs;**
 - Incident response plans not regularly reviewed or tested;
 - Limited internal audit review of information security controls; and
 - **Inconsistent reporting of material incidents and control weaknesses in a timely manner.**

* <https://www.apra.gov.au/news-and-publications/cyber-security-stocktake-exposes-gaps>

The Unstructured Data Problem

- APRA emphasises compliance with the **identification and classification of critical and sensitive information asset**.
- This includes unstructured data.
- Unstructured data is often associated with productivity tools.



Extent of the Problem

Around 80% to 90% of an organization's data is largely semi structured or unstructured and is never used (Gartner 2021).

Anywhere from 55% to 80% of the data a company stores is dark, meaning there is little to no visibility into it (Gartner 2021).

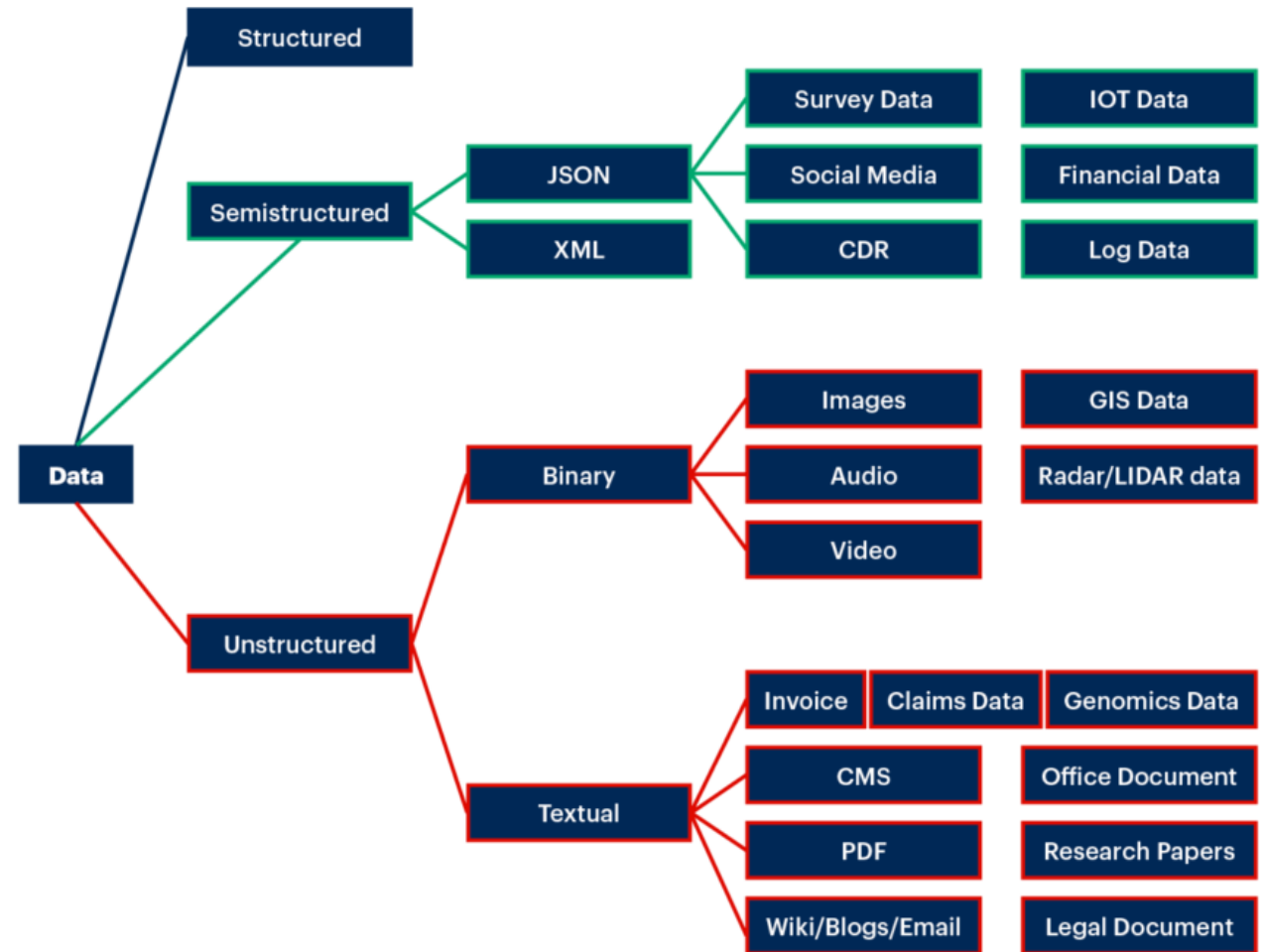


Diagram: Gartner 2021 - G00723116

Problem: Double Extortion

- Double extortion: Encrypts data plus exfiltrate trade secrets and Personally Identifiable Information (PII).
- Royal Mail UK 2023: £66m ransom demand, operational systems halted; HR including disciplinary records exfiltrated.
- Guardian Newspaper UK 2022: IT systems impacted; employee PII exfiltrated.
- Latitude Australia & NZ 2023: IT systems impacted; PII of 255,000 customers exfiltrated, losses totaling A\$76 million.
 - *“For a period of six weeks, new originations stopped, receivables declined, pricing actions were paused, and collections activities were significantly disrupted”* (Latitude’s CEO).
- The insurance problem includes the cost of reporting on unstructured data assets.
- Exfiltration is theft by stealth, so unstructured data is also complex for threat actors.

<https://www.itnews.com.au/news/latitude-financial-flags-76-million-in-cyber-incident-costs-599350>

Hype: Regex vs AI

- A regular expression (regex) matches a sequence of characters to a defined pattern.
 - E.g., Drivers Licence → 8 varchar → 2 alpha by 6 numeric.
- eDiscovery relies on regex libraries.
- AI relies on Neural Networks (NN).
- A NN uses supervised training, or unsupervised training using data, or a combination of both.
- Neural Networks are a subset of Machine Learning (ML) that model and process complex data inputs.

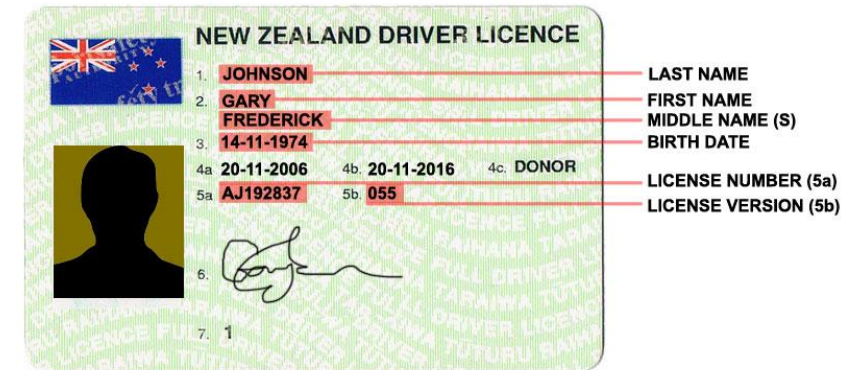


Diagram: no1canterburycollege.wordpress.com/2011/02/21/nz-driver%E2%80%99s-licence/

eDiscovery vs Privacy Tools

eDiscovery Tools

- Electronically stored information (ESI) should include including unstructured data: email, audio, video, IoT meta-data etc.
- Recent innovations: SaaS Cloud hosted, integration with other solutions, reduce manual process by utilising legal hold tracking solutions.
- Market is *Mature* (Gartner 2022).

Privacy Tools (AI-powered)

- Used for discovery, classification, Data-Subject-Request (DSR).
- Claim: *Privacy-by-Design*.
- Innovations: Often SaaS, promote proactive data management capabilities, claim Natural Language Processing (NLP) and Large Language.
- Market is *Embryonic*.
- Uptake < 1% of target audience (Gartner 2022).

1. *To protect against cyber threats, organisations should have robust cybersecurity measures in place.*

However, the sophistication of threat actors mean a prudent organisation must envisage what's next once they are breached.

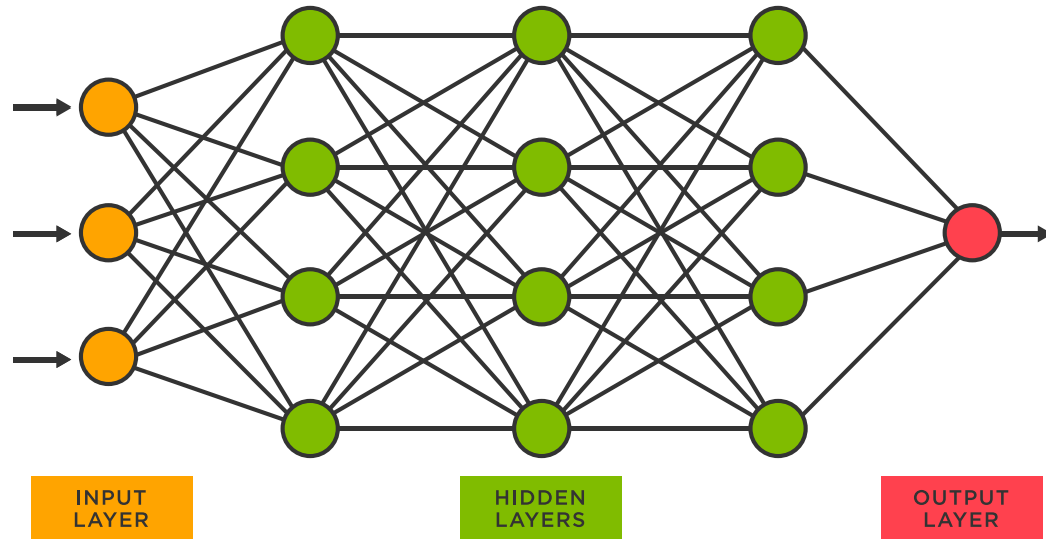
Technology vendors are now focused on providing tools that can assist firms to know their data assets in order to report to regulators and victims following an incident.

2. *To defend against cyber threats, organisations need strong cybersecurity defenses.*

But given the advanced skills of cyber attackers, it's crucial for organisations to plan for potential breaches.

Technology companies are now offering tools that help businesses identify their data assets, making it easier to inform affected parties and regulators following an incident.

Neural Networks: Black Boxes



- Blackboxes e.g., Large Language Model create hallucinations.
- Classifiers rely on the complex fine-tuning across multiple transformations.
- Small imperceptible perturbations can cause erroneous image predictions.
- Privacy in training data e.g., ChatGPT.
- *What training data and NN algorithms should be selected?*
- *The Insurance problem is Sprawl-on-Sprawl*
 - An increasing sprawl of unstructured data.
 - The layering of Internet-facing, complex technologies, to address the issues.


Diagram: <https://www.tibco.com/reference-center/what-is-a-neural-network>

New AI Business Processes

Neurotechnology [+ Add to myFT](#)

Brain implants give a voice to people who cannot speak

Scientists use electrodes and AI programs to turn thoughts into speech via a lifelike avatar



Ann, a participant in the study, uses a digital link wired to her cortex to interface with an avatar © Noah Berger

Clive Cookson 16 HOURS AGO 15

Two research teams in California have developed brain implants that they say are much more effective than previous devices at giving a voice to people who cannot speak.

- Results show a feasible path for restoring communication with people affected by paralysis.
- Participant's attempted speech was decoded at 62 words per minute, which approaches natural conversation at 160 words.
- These processes give rise to massive data volumes:
 - *Where is this data stored?*
 - *In what data formats?*
 - *What inference can be derived?*
- What happens if there is a regulatory requirement to identify or classify "*sensitive information assets*"?
- *APRA stock-take: Privacy-by-Design?*
 - A regulatory outlier e.g., Facebook, Amazon, and Twitter were evolved business models before GDPR in 2018.
 - Facebook had 2 billion users by the end of 2017 (Statista 2022).

doi.org/10.1038/s41586-023-06377-x

Image: www.ft.com/content/ac5da810-5079-4f10-861b-273acbf09bb3

Conclusion

- *APRA Stock-take*: regulation may outpace technology.
- Inconsistent reporting of APRA requirements....is it ushering in a brave new world.
- Privacy Tools are emergent.
- But Cloud and other parts of the tool-box are Internet facing.
- Internet facing creates expansive data networks, so new Cyber Security issues.
- *Insurance problem*: Sprawl-on-Sprawl, more complex volumes of unstructured data, reporting to regulators.