Office of the Victorian Information Commissioner

Victorian Privacy Network

July 2026

A reminder – Today's session is being recorded.

Acknowledgment of Country

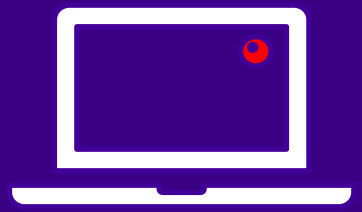
Sean Morrison
Victorian Information Commissioner

We acknowledge the Wurundjeri people of the Kulin Nation as the Traditional Owners of the land from which we are presenting today.

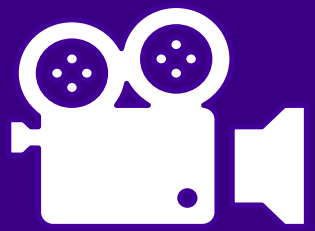
We pay our respects to their Elders, past and present, and Aboriginal Elders of other communities who may be with us today.

Housekeeping - What to be aware of

Cameras and mics have been muted for attendees**.**



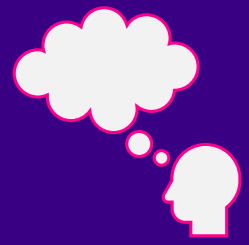
Today's session **is being partially recorded.**



A copy of the **recorded sections** will be made available on our Vimeo website.

Housekeeping – How to engage

If you want to ask a **question,** type your question into the **Teams Q&A channel**.



Please leave us some **feedback** at the end of the meeting!

Today's agenda

Agentic AI: identifying and managing privacy & security risks
Dr Jodie Siganto
Privacy 108

Children's Privacy by Design
Carly Kind - Privacy Commissioner
Office of the Australian Information Commissioner

Machinery of Government changes – Privacy and Information Security considerations
Emma Stephens
Office of the Victorian Information Commissioner

Agentic AI: identifying and managing privacy & security risks

Dr Jodie Siganto

Privacy 108

Agentic AI: identifying and managing privacy & security risks

July 2026

Dr Jodie Siganto

Privacy108
We Protect Privacy

SESSION OVERVIEW

Today's session

- What agentic AI is, and how it differs from other types of AI
- Privacy and security risks of AI generally – a quick refresher
- The specific risks agentic AI introduces for government service
- What the privacy officer can do
- Where this technology is heading, and what to stay alert to

What is 'agentic AI'?

A simple definition

Agentic AI refers to AI systems that can plan, decide and take multi-step action toward a goal, largely on their own – rather than simply answering a question or drafting text when asked.

Why this matters for privacy

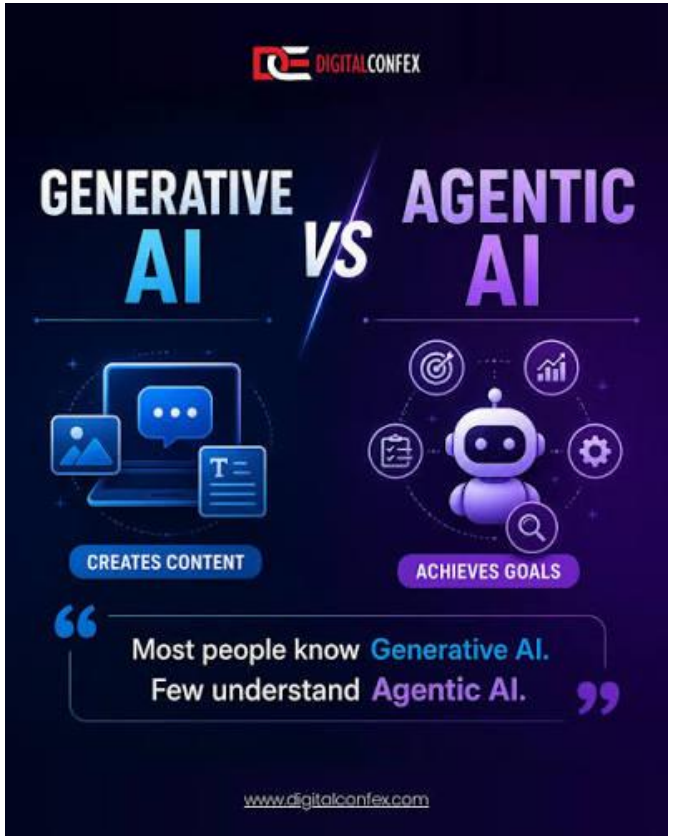
The more autonomy a system has, the less opportunity there is for a human to catch a problem before it happens.

The key word is autonomy:

Generative AI generates content.

Perceive → Reason → Act → Learn/Feedback

An agentic AI system can perceive a situation, decide what to do, and then actually do it e.g. update a record, send a notice, escalate a case, with little or no human clicking 'go' each time.



The basic cycle: perceive, reason, act, learn

1

Perceive

Takes in information – a form, an email, data from a database, a sensor feed.

2

Reason

An AI model weighs that information against its goal, rules and policy constraints.

3

Act

Carries out the action via an API – a controlled digital 'doorway' between systems.

4

Learn

The outcome is fed back in, and may be stored in memory to inform future decisions.

Beyond a single agent: real deployments often use several specialised agents working together (an orchestration layer coordinates them) – a 'planner', a 'researcher', an 'executor' and a 'supervisor' agent that checks the others' work.
Keeping a human appropriately involved is so important it gets its own slide next.

Examples of agentic AI in government

1 Public enquiry handling

As part of the same process: an intake agent classifies an enquiry, a retrieval agent pulls relevant data, a human reviews sensitive cases, and a response agent replies or escalates.

2 Complaint intake and case creation

A complaint is triaged automatically, matched against current policy, and a case file is created and routed – with no manual data entry.

3 Multi-step eligibility processing

Each decision (e.g. 'is this document sufficient?') determines the next step in a benefits or eligibility workflow.

4 Back-office document processing

An agent coordinates several tools/APIs across systems to complete a task end-to-end.

Where agentic AI sits among other types of AI

Not everything labelled 'AI' works the same way – here's how the main types compare.

Type of AI	How it works	Typical government use	Autonomy
Traditional automation (scripts / RPA)	Follows fixed, pre-programmed rules; doesn't adapt.	Invoice routing, scheduled data transfers.	**None**
Predictive / machine learning	Learns patterns from past data to score or predict.	Fraud/risk scoring, demand forecasting.	**Low**
Generative AI (chatbots / co-pilots)	Drafts or summarises based on learned patterns.	Drafting correspondence, answering FAQs.	**Low–medium**
Agentic AI	Plans, chooses tools, takes action and adapts.	End-to-end complaint triage, multi-step case processing.	**Medium–high**

Source: adapted from Digital NSW's comparison of traditional automation, chat assistants and autonomous AI agents.

AI risk – the quick refresher

- Over-collection or use of personal information beyond what's needed.
- Bias and unfair outcomes, if training data reflects historical inequities.
- Hallucination – confident but false or fabricated output.
- Lack of explainability – difficulty telling a citizen why a decision was made.
- Data handled by third-party vendors or overseas infrastructure.
- Security exposure of data, models and integrations to breach.

All of this still applies to agentic AI – agentic AI adds a further, distinct layer of risk on top.

What's different about agentic AI risk (1 of 2)

Risk area	What it means in practice
Reduced real-time human check	Actions may occur before any person reviews them, narrowing the window to catch an error or privacy breach.
Cascading and 'conformity' errors	One agent's mistake can be copied or accepted by others without challenge, so an error can propagate across a case.
Memory and data retention	Agents may retain personal information across sessions – a new store of personal data needing its own retention rules.
Agent-to-agent data exchange	Data moving via APIs between agents as part of the same ‘system’ increases the points where information could be exposed or sent astray.

What's different about agentic AI risk (2 of 2)

Risk area	What it means in practice
Tool / API misuse	Broader access or permissions than needed could trigger unintended actions – e.g. deleting records, mass notifications.
Reduced traceability & contestability	Multiple automated steps can make it hard to reconstruct why a decision was made – undermining a citizen's right to review.
Blurred accountability	With autonomous, multi-agent action, it can be unclear who – which business area, which owner – is accountable for an outcome.
Vendor and supply-chain exposure	Models, hosting and tooling are often third-party – raising data sovereignty, subcontracting and records questions.

About the DTA's agentic AI guidance

Who is the DTA?

The Digital Transformation Agency is an Australian Government agency and the whole-of-government coordinator for AI policy – it advises on, and publishes, the standards federal government agencies use to adopt AI responsibly.

Other guidance the DTA has issued

- Policy for the responsible use of AI
- AI technical standard (the parent standard)
- AI Impact Assessment Tool
- Standard for AI transparency statements

Where this guidance fits: published as an addendum to the Technical standard (June 2026) – it doesn't replace the parent standard, it extends it for agentic systems. All of the parent standard's statements still apply.

The 8 DTA statements at a glance (AGT.1–AGT.8)

AGT.1
Governance & safeguards

AGT.2
Memory management

AGT.3
Design for agentic workflow

AGT.4
Data & routing

AGT.5
Select agent, model & tech

AGT.6
Continuous evaluation

AGT.7
Tools & protocols

AGT.8
Ongoing monitoring

Governance & memory (AGT.1–AGT.2)

Statement AGT.1 – Governance & safeguards

MUST Assign human accountability for the decisions and outcomes agents produce (AGT.1.1).

MUST Identify effective human oversight and enable intervention when necessary (AGT.1.2).

MUST Ensure the system can explain its decisions without exposing personal data (AGT.1.3).

SHOULD Establish an orchestration layer to coordinate agents, tasks and tools (AGT.1.4).

Statement AGT.2 – Memory management

SHOULD Establish robust, auditable memory management: define what may be stored, for how long, where, and how it's purged (AGT.2.1).

SHOULD Prevent duplication, leakage or 'poisoning' of memory, and resolve conflicting or outdated information (AGT.2.1).

SHOULD Treat agent memory as a new personal-information store requiring its own retention schedule.

Design & data (AGT.3–AGT.4)

Statement AGT.3 – Design for agentic workflow

MUST Define business needs and goals, and narrow each agent's scope to reduce errors and hallucination (AGT.3.1).
MUST Define the logical sequence of tasks and agents, avoiding duplicated responsibility (AGT.3.2).
MUST Identify guardrails – a unique identity and least-privilege access for every agent (AGT.3.3).
MUST Embed self-correction, fail-safes and auditable kill-switch requirements (AGT.3.4).
SHOULD Select integration tools carefully and plan for scalability as complexity grows (AGT.3.5–3.7).

Statement AGT.4 – Data & routing

MUST Establish secure data management for agent-to-agent exchange – authentication, encryption, audit logs, and classification/sovereignty controls (AGT.4.1).
SHOULD Ensure messages and tasks are dynamically and consistently routed to the right, authorised agent (AGT.4.2).

Train, evaluate & integrate (AGT.5–AGT.7)

Statements AGT.5 & AGT.6 – Train & evaluate

MUST Ensure the agent, model and technology genuinely matches the use case's complexity and needs (AGT.5.1).

MUST Test agents for robustness against unexpected inputs and adversarial scenarios (AGT.6.1).

MUST Evaluate agent–tool, agent–environment and agent–memory interactions (AGT.6.2–6.4).

SHOULD Weigh trade-offs – cost, latency, data residency, vendor lock-in – when choosing models (AGT.5.2–5.3).

Statement AGT.7 – Integrate

SHOULD Select tools and communication protocols with defined scope, clear permissions and audit logging (AGT.7.1).

SHOULD Track emerging agent-to-agent communication standards for alignment and interoperability (AGT.7.2).

Monitor & decommission (AGT.8 + decommission)

Statement AGT.8 – Ongoing monitoring

MUST Continuously monitor agents and their environment – goal drift, permitted scope, unexpected behaviour, performance/cost alerts, tool misuse, memory issues, and changing authorisations (AGT.8.1).

SHOULD Establish a 'control tower' – centralised observability, dashboards and incident detection across all agents (AGT.8.2).

Decommission – existing Statement 41 / Criterion 147

MUST Securely decommission or repurpose computing resources dedicated to an agent – shut down its processes, remove agent-specific data and memory, and ensure tools, logs and resources are removed or repurposed.

Privacy officer role in managing agentic AI risk in practice

Get involved early

- Take part in use-case screening – test whether the proposed autonomy is genuinely needed.
- Require a privacy impact assessment before a pilot is approved – covering memory, agent-to-agent flows and vendor handling.

Shape governance, transparency & design

- Confirm the accountable owner, oversight model and authority limits are proportionate – before go-live.
- Review and sign off citizen-facing transparency material and the review/appeal pathway.
- Set memory retention/purge requirements and confirm audit logs meet records obligations.

Stay engaged operationally

- Set a trigger for whenever scope, data access, autonomy or vendor changes – refresh the assessment.
- Contribute to incident response and kill-switch testing, including breach notification.
- Check the agent/asset register – named owners and up-to-date quotas.

Establishing human accountability – practical steps

Step	What the privacy officer should check
Name an accountable owner in writing	Record their name, role and business area in the PIA/project file – not just 'IT' or 'the vendor'.
Separate responsible from accountable	The technical/build team is responsible for day-to-day operation; **a specific human is accountable for the agent's decisions and outcomes.**
Confirm the owner has real authority	Check they are genuinely delegated to approve the decisions the agent takes on their behalf – not just named on paper.
Test accountability before go-live	Run a 'something went wrong' scenario and confirm the named owner can trace, explain and correct the decision – and is the contact for citizen escalations.

Data governance – practical steps

Step	What the privacy officer should check
Map every data flow before build	List every dataset the agent can read, write or generate, including any third-party tools and APIs it calls.
Apply data minimisation by design	Limit the agent's access to only the fields and records it needs for its task – not broad, standing access 'just in case'.
Set classification & sovereignty rules upfront	Confirm where data is hosted and processed, and apply the right classification and cross-border rules before any pilot.
Review vendor & subcontractor handling	Check contracts cover data use, retention, subcontracting and breach notification for any third-party model or tool.

Memory management – practical steps

Step	What the privacy officer should check
Treat memory as a new personal information store	Confirm what the agent remembers about a citizen or case, and bring it inside your records and retention schedule.
Define what may – and may not – be stored	Set explicit rules on which fields can enter short/long-term memory; exclude sensitive categories unless specifically justified.
Set retention periods and automatic purge triggers	Require a defined retention period and deletion mechanism, not indefinite storage 'in case it's useful'.
Cover memory in access, correction & deletion requests	Ensure a citizen's rights request also reaches what's stored in agent memory – not just the case file.

Establishing oversight and explainability – practical steps

Effective human oversight and the ability to explain a decision are two of the DTA's mandatory requirements – together they make an autonomous system contestable and trustworthy.

Step	What the privacy officer should check
Match the oversight model to the risk	Confirm whether each agent is HITL, HOTL or HOOTL and that the choice reflects the sensitivity and reversibility of its decisions – then document the rationale.
Test human intervention before go-live	Verify a human can genuinely pause, override or take over an in-flight action, and that escalation pathways are documented and accessible.
Require a plain-language reason for every action	Confirm the system can explain what it decided and why, in everyday language, for each individual action – not just in general terms.
Protect sensitive information within explanations	Where showing the full reasoning trail could expose personal or sensitive data, require an auditable decision record instead, and check metadata is captured for review.

Three models of human oversight

Choose the oversight model deliberately and proportionately to risk – and document it, rather than defaulting to whatever the vendor ships.

Human-in-the-loop (HITL)

A person actively reviews inputs and outputs and makes the final decision before an action happens.

Use for sensitive or irreversible actions – payments, entitlements, mass notifications.

Human-on-the-loop (HOTL)

The system runs autonomously while a person supervises in real time and can pause or override it.

Suits higher-volume tasks where full pre-approval isn't practical.

Human-out-of-the-loop (HOOTL)

The system decides fully independently; humans only conduct periodic audits or post-action sampling.

Appropriate only for lower-risk, well-tested processes.

If a citizen challenges an automated case decision, you need three things ready – a human they can appeal to, an audit trail of what happened, and a plain-language reason for the outcome.

Monitoring and decommissioning – practical steps

Step	What the privacy officer should check
Confirm real-time monitoring is active	Check dashboards and alerts cover goal drift, unexpected behaviour, performance and cost, tool misuse, and changes to authorisations – not just uptime.
Ask who is watching the watchers	Confirm a control tower or equivalent gives one place to see all agents' health, risk and compliance status, with a named person responsible for reviewing it.
Set a trigger for suspected drift or misuse	Require escalation and investigation the moment monitoring flags unexpected behaviour, conflicting agent actions or memory issues – not just at a scheduled review.
Require a documented decommissioning process	Confirm there's a plan to shut down an agent's processes, delete its agent-specific data and memory, and remove or repurpose its tools, logs and access when retired.

Where this is heading

- Agentic AI is still an early, fast-moving field – standards, tools and vendor products are changing quickly.
- Expect more multi-agent, cross-agency ecosystems using emerging communication protocols – raising new cross-boundary data-flow questions.
- The human role is shifting from doing tasks to managing agents that do tasks – with workforce, training and accountability implications.
- Watch for scale creep – a well-governed pilot can quietly expand in scope, data access or autonomy without a fresh risk assessment.
- Keep a watching brief on regulatory developments, and be ready to revisit agency policy as guidance matures.

Key takeaways

- Agentic AI's defining feature is autonomy – it plans and acts, not just drafts – and that autonomy drives the extra risk.
- The privacy and security risks of ordinary AI still apply, plus new ones: **reduced human checkpoints, memory retention, cascading errors and blurred accountability.**
- Choose a human oversight model deliberately and proportionately, and make sure decisions can always be explained in plain language.
- The DTA's mandatory criteria span the whole lifecycle – privacy officers should expect to see evidence against each of them.
- Privacy officers have an active, ongoing role – not a one-off sign-off – from use-case screening through to decommissioning.
- This guidance will keep evolving – proceed cautiously, start small, and stay engaged as it develops.

Further reading

- Digital Transformation Agency (Cth), Agentic AI addendum to the AI technical standard for Australian Government.

- Digital NSW, AI agent usage and deployment guidance (October 2025).

Thank You

Any Questions?

CONTACT US

1300 41 20 50

hello@privacy108.com.au

www.privacy108.com.au

Privacy 108

Children's Privacy by Design

Carly Kind - Privacy Commissioner

Office of the Australian Information Commissioner

Australian Government
Office of the Australian Information Commissioner

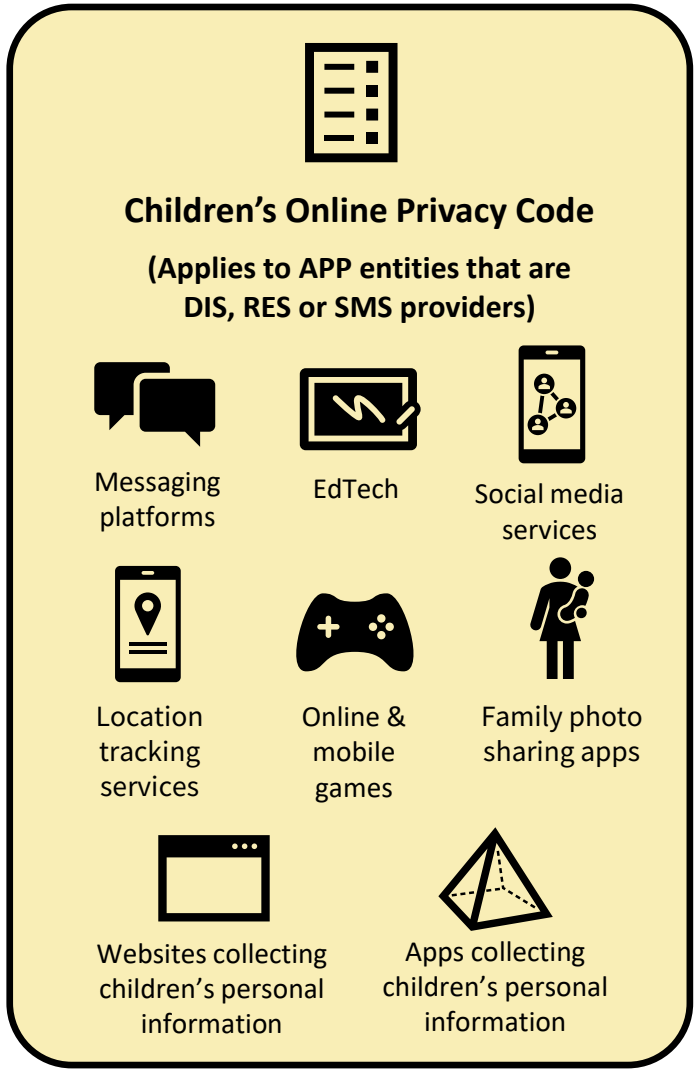
Children's Privacy by Design

Privacy Commissioner
Carly Kind

23 July 2026

OAIC

Children’s Online Privacy Code – Privacy by design



The Children’s Online Privacy Code (**the Code**) establishes special privacy rules which specify how the Australian Privacy Principles (APPs) apply when handling children's personal information online and imposes additional requirements to help protect children's privacy when using online services.

The aim of the Code is not to prevent children from engaging online, but to ensure their personal information is protected within the digital ecosystem.

The Code will apply to online services where children experience the highest privacy risks, including most apps, games and websites that children and teenagers use daily.

The OAIC has developed the draft Code in consultation with children, young people, families, Australian Government agencies, industry, civil society and academics.

Rights-based approach

Children learn, play, socialise and connect with the world online - it's important that they can participate without fear or exploitation.

The OAIC continues to collaborate cohesively and consistently to address privacy in a manner that supports individuals' rights and builds children's knowledge so they can confidently engage with online services knowing their privacy will be protected.

Adopting a rights-based approach enables us to develop effective policy which improves children's interactions with the digital world.

Overview of Children's Online Privacy Code measures

Collection of children's personal information must be 'strictly necessary' by default	Collection, use and disclosure of children's personal information must be in the best interests of the child	Transparency measures (including requirements for age-appropriate privacy policies and notices)
Right to access, correct, and destroy personal information about children	New consent and assent requirements	Child-friendly inquiry and complaint handling processes
Notification requirements for mechanisms that monitor or control service use or monitor geolocation	Ascertaining the age of end-users of an entity's service	Governance and education

The best interests of the child

‘Best interests of the child’ refers to the fundamental principle in international law that all decisions made, and actions taken should be in children’s ‘best interests’ as per Article 3 of the United Nations Convention on the Rights of the Child.

The obligation requires APP entities to ensure that collection, use and disclosure of personal information is consistent with the best interests of the child.

The right to deletion

60% of children and young people said they want to be more in control of their personal information.

The objective of this obligation is to increase transparency and give children stronger control over their personal information.

Transparency obligations

Children, young people, parents and carers want online companies to be more up-front, direct and honest about how they handle personal information.

The obligations include ensuring privacy notices and policies are written in clear, accessible and age-appropriate language so children and young people can understand what is happening with their personal information and can easily seek help if required.

Assent

70% of children and young people said they do not like when information about them is shared without them knowing.

The obligation requires that, when a child under 15 enables an entity to collect, use or disclose certain personal information, the entity must seek the child's assent to that collection, use or disclosure, as well as a person with parental responsibility.

Joint regulatory oversight of SMMA

- The OAIC co-regulates the SMMA obligation with eSafety

- The OAIC monitors privacy acts and practices of entities subject to the SMMA obligation (i.e. age-restricted social media platforms and age assurance providers)

- The OAIC oversees the compliance and enforcement of the privacy provisions set out in Section 63F of Part 4A of the *Online Safety Act 2021*, which operate in tandem with the *Privacy Act 1988*

Children’s Online Privacy Code Consultation

Phase 1

Jan – Aug 2025

Initial discussions with children, parents and relevant organisations focused on children's welfare.

337 children, young people and parents/carers responded to our online worksheets in June 2025.

Phase 2

Apr – Aug 2025

Engagement with civil society, academia and industry stakeholders to gather insights and perspectives.

61 stakeholder responses received to our Issues Paper consultation.

Phase 3

Mar – Jun 2026

Engaged with 374 stakeholders across 14 roundtables with representation across academia, civil society, industry, government.

433 children, young people, parents and carers shared their perspectives on the draft Code.

Australian Government

Office of the Australian Information Commissioner

Presented by
Commissioner Carly Kind

www.oaic.gov.au

x.com/OAICgov

www.linkedin.com/company/office-of-the-australian-information-commissioner

OAIC

Machinery of Government changes – Privacy and Information Security considerations

Emma Stephens

Policy - OVIC

Office of the Victorian Information Commissioner

Machinery of Government changes – Privacy and Information Security considerations

Emma Stephens, A/Assistant Commissioner, Policy

OVIC

Official

What we'll cover

- Privacy considerations
- Information security considerations
- OVIC resources

Machinery of government changes

Examples:

- **Transfer** of functions and responsibilities between organisations
- Organisational **restructures**
- **Creation** of organisations
- **Cessation** of organisations

10 Information Privacy Principles

- Minimum standard for how Victorian public sector organisations should manage personal information.

- Some relevant IPPs:

    - IPP 1 – Collection

    - IPP 2 – Use and disclosure

    - IPP 5 - Openness

Information security considerations

- Information assets and systems should be adequately protected:
    - **Identify** information assets and systems that may be impacted by the MoG process
    - **Understand security value** of the information assets and systems
    - **Assess** the **information security risks**
    - Transfer information assets and systems in accordance with their security value and information security assessment

Notifying OVIC of significant change

- **A MoG may require you to notify OVIC of significant change** to your operating environment or security risks relevant to the organisation.
    - *Notification of significant change form* – available on OVIC website
    - Submit to OVIC Information Security Unit – security@ovic.vic.gov.au.

Example Scenario

- Organisation A's investigation functions will cease
- Other functions will transfer to Organisation C

- Organisation A conducts a brief stocktake of its information holdings, including information created or held by third parties who were supporting the investigation function. These existing assets **will not be transferred** to the newer, larger organisation – Organisation C.
- The information assets and systems relating to the investigative functions have had a recent **security value assessment** undertaken, resulting in an overall Business Impact Level (**BIL**) rating of **3**.
- Organisation A uses this BIL rating to inform a subsequent **information security risk assessment** helping determine the **most appropriate methods** for the **secure disposal** of the information and **secure decommissioning** of the system.
- Organisation A convenes its Information Security Lead (**ISL**) and its Records Manager (**RM**) to determine the next steps, including updates to any internal governance documentation.

Example Scenario

- Organisation A needs to transfer aged records to Organisation C

- Organisation A discovers a large amount of aged, yet actively relied upon, records in its custody.
- The last security value assessment of this information was conducted over ten years ago. Therefore, a refreshed assessment is undertaken to confirm the current security value. This results in an overall Business Impact Level (**BIL**) rating of 2.
- Using this BIL rating, Organisation A conducts an information security risk assessment to inform the most appropriate methods to securely transfer this information to Organisation C.
- Organisation A convenes its ISL and its RM, along with relevant representatives from Organisation C to determine the next steps, including updates to internal governance documentation. Organisation A arranges for the secure transfer of the information assets to Organisation C.

OVIC resources

Upcoming

- Factsheet for the Victorian Public Sector – Machinery of government changes: Privacy, information security and freedom of information considerations

Other resources

- Guidelines to the Information Privacy Principles
- Guide and form – Notification of Significant change
- Practitioner Guide: Assessing the security value of public sector information
- Practitioner Guide: Information Security Risk Management

Got any questions or feedback?

Contact OVIC's policy team - policyteam@ovic.vic.gov.au

OVIC

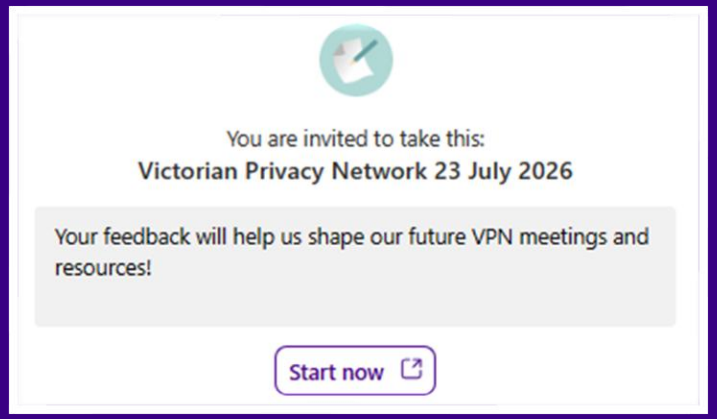
www.ovic.vic.gov.au

Closing remarks

Sean Morrison, Victorian Information Commissioner

Office of the Victorian Information Commissioner

For more information regarding this event or for any follow-up questions, please contact: **enquiries@ovic.vic.gov.au**



We'd love your feedback

Anonymous survey now available in chat

OVIC

www.ovic.vic.gov.au