



CONFIRM. NYC

The Conference on Information
and Records Management

The New Information Ecosystem: The Intersection of AI, Privacy, and Records Management

Constantine Karbaliotis

Counsel, nNovation LLP



March 4, 2025



Today's Journey

- ✓ The evolving information landscape
- ✓ Data provenance and lineage in an AI world
- ✓ Records retention challenges in AI systems
- ✓ The records manager's critical role
- ✓ Call to action

The Evolving Information Landscape

From:

- Static documents had a clear lifecycle, ownership and predictable formats

To:

- Dynamic, interconnected data, which is in diverse and evolving formats in open ecosystems

The convergence is happening now:

- ⦿ AI systems requiring training data at an unprecedented scale
- ⦿ Privacy regulations demanding greater accountability
- ⦿ Copyright and other intellectual property concerns affect both the data inputs and their outputs
- ⦿ Technological innovation is outpacing organizational readiness

The disciplines you've developed for decades are now mission-critical in ways organizations don't yet fully appreciate.

Data Provenance and Lineage

Tracking Data Flow in AI Systems



Origin of Data

Tracking the source of each data element used in AI training.



Transformation History

Recording modifications, movements, and processing stages.



Chain of Custody

Ensuring accountability through permissions and consent tracking.

What it means in an AI context:

- 📍 Origin of each data element used to train or feed AI
- 📍 Complete history of transformations and movements
- 📍 Chain of custody and consent throughout data lifecycle
- 📍 Documentation of permissions, restrictions, and conditions
- 📍 Garbage in ... garbage out

Why Provenance Matters More Than Ever

AI amplifies existing issues:

- ⦿ Errors propagate and scale
- ⦿ Ethical concerns multiply with aged data that may not be relevant or contains historical biases that become encoded and hidden
- ⦿ Unauthorized data use has broader implications
- ⦿ Legal and compliance risks multiply:
 - Was there consent – did the original legal basis for processing cover what we are now proposing to do?
 - Are there copyright or confidentiality obligations attached to the training data?
 - How do we honour our retention schedule when data is used to train models?

The Provenance Problem: A Scenario

A financial services company implements a customer service AI chatbot using:

- Historical customer interactions
- Public financial advice forums
- Internal knowledge base documents

What can go wrong if the company...

- Cannot verify if training data included opt-out customers
- Discovers some forum data violated terms of service
- Is unable to track which responses came from which sources
- And gets regulators' demands to provide comprehensive data lineage

Records Retention in an AI World

New tensions to manage:

- 🔴 AI systems improve with more data vs. privacy principles of data minimization
- 🔴 Models may retain "knowledge" of deleted records
- 🔴 Versioning of AI models creates retention complexity
- 🔴 Training vs. inference data have different retention needs

When data is used to train an AI system:

- 🔴 Information becomes encoded in model parameters
- 🔴 Original records may be discarded after training
- 🔴 Model may "remember" information that should be forgotten
- 🔴 Cannot simply "delete" specific information from models

How do we apply retention policies to knowledge embedded in AI systems?

The Retention Dilemma: A Scenario

A healthcare technology company builds an AI diagnostic tool:

- ⦿ **Trained on 10+ years of patient records**
- ⦿ **Original data sources included images, notes, and lab results**
- ⦿ **Multiple models developed over several years**

What can go wrong when the company...

- ⦿ **Receives patient deletion requests**
- ⦿ **Can't determine which models contained which patient's data**
- ⦿ **Lacks processes to remove specific data influences from models**
- ⦿ **Lacks documentation of training data sources for older models**

Records Management's Critical Role

The sessions you will participate in today will help you to rise to the challenges of an AI-driven world:

- Help bridge technical and compliance teams
- Apply established information governance principles to new domains
- Develop frameworks for AI data documentation
- Meet the ethical and legal challenges AI presents in the retention context
- Implement practical data lineage processes
- Define retention policies for AI systems
- Leverage AI tools for discovery and retention compliance!

Records Management's Critical Role (2)

You can bring the right tools and frameworks to bear that align well with AI governance needs: ARMA International's Information Governance Reference Model (IGRM) domains link clearly to AI:

- **Accountability:** Clear ownership of AI systems and their data
- **Transparency:** Documenting AI decision-making processes
- **Integrity:** Ensuring AI training data is accurate and complete
- **Protection:** Safeguarding sensitive data used in AI systems
- **Compliance:** Meeting regulatory requirements for AI
- **Availability:** Ensuring AI systems and their records are accessible
- **Retention:** Determining how long to keep AI models and training data
- **Disposition:** Proper disposal or archiving of obsolete AI models

Records Management's Critical Role (3)

Be aware of and leverage the NIST AI Risk Management Framework:

- Map: Inventory data and systems (What records managers already do!)
- Measure: Assess risks (Including information risks you already manage!)
- Manage: Implement controls (Using governance you already understand!)
- Govern: Provide oversight (Through processes you already facilitate!)

Call to Action

Provide leadership:

- Join AI governance committees
- Collaborate with data science teams
- Advise on data lineage methodologies
- Develop AI-specific retention schedules
- Create provenance documentation requirements
- Ensure compliance with emerging AI regulations

And support the activities that lead to good AI governance:

- Inventory your organization's AI initiatives
- Identify where records management requirements are not being applied
- Articulate the specific risks this creates
- Propose practical governance solutions
- Collaborate across disciplines

AI without proper records discipline is a house built on sand.

About

- [Constantine Karbaliotis](#), Counsel at nNovation LLP, is a leading privacy and data protection expert with over two decades of international experience. As a former privacy officer for multinational organizations, he brings unique insights to his specialized practice in privacy law, cybersecurity, artificial intelligence regulation, and information management.



- [nNovation LLP](#) is a premium boutique law firm serving clients across Canada through its Ottawa and Toronto offices. We provide sophisticated legal counsel to businesses, public sector organizations, and international law firms seeking Canadian privacy expertise. Our virtual model enables us to deliver exceptional service while maintaining flexibility and efficiency for our clients.

ckarbaliotis@nnovation.com

