iapp

Building the foundation: Records retention before AI

Tuesday, 26 May
11:00–12:00 PDT
14:00–15:00 EDT
20:00–21:00 CEST

filerskeepers

lawstronaut

Building the foundation: Records retention before AI

Speaker

Wanne Pemmelaar
CEO & Co-founder
filerskeepers | Lawstronaut

filerskeepers

lawstronaut

Wanne is a top-tier data and tech lawyer and 3x entrepreneur who loves building tech solutions to legal problems he has experienced firsthand. Wanne has over a decade of work experience as a legal practitioner. When it comes to developing products, he insists on stellar design and a commitment to user empathy. As a data and tech lawyer, Wanne's passion is assisting companies to navigate the legal landscape against the backdrop of exponential data and tech developments.

www.filerskeepers.co

Introduction to filerskeepers

Founded in 2018

337
jurisdictions online

358,402+
retention periods

4000+
multinational customers

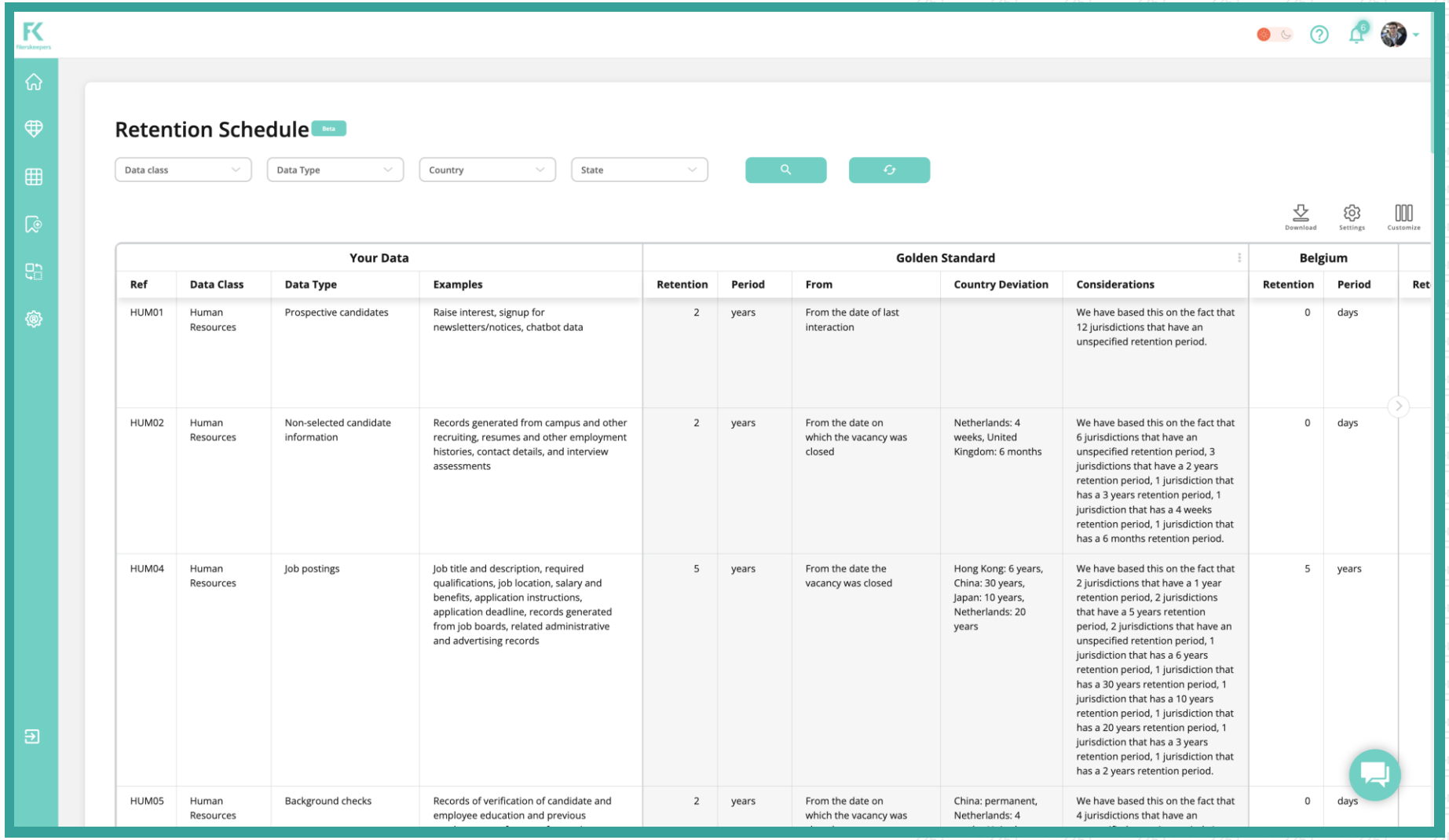
300+
law firm customers

80+
partners

95
team members

filerskeepers

www.filerskeepers.co



Your Data				Golden Standard					Belgium		
Ref	Data Class	Data Type	Examples	Retention	Period	From	Country Deviation	Considerations	Retention	Period	Ret
HUM01	Human Resources	Prospective candidates	Raise interest, signup for newsletters/notices, chatbot data	2	years	From the date of last interaction		We have based this on the fact that 12 jurisdictions that have an unspecified retention period.	0	days	
HUM02	Human Resources	Non-selected candidate information	Records generated from campus and other recruiting, resumes and other employment histories, contact details, and interview assessments	2	years	From the date on which the vacancy was closed	Netherlands: 4 weeks, United Kingdom: 6 months	We have based this on the fact that 6 jurisdictions that have an unspecified retention period, 3 jurisdictions that have a 2 years retention period, 1 jurisdiction that has a 3 years retention period, 1 jurisdiction that has a 4 weeks retention period, 1 jurisdiction that has a 6 months retention period.	0	days	
HUM04	Human Resources	Job postings	Job title and description, required qualifications, job location, salary and benefits, application instructions, application deadline, records generated from job boards, related administrative and advertising records	5	years	From the date the vacancy was closed	Hong Kong: 6 years, China: 30 years, Japan: 10 years, Netherlands: 20 years	We have based this on the fact that 2 jurisdictions that have a 1 year retention period, 2 jurisdictions that have a 5 years retention period, 2 jurisdictions that have an unspecified retention period, 1 jurisdiction that has a 6 years retention period, 1 jurisdiction that has a 30 years retention period, 1 jurisdiction that has a 10 years retention period, 1 jurisdiction that has a 20 years retention period, 1 jurisdiction that has a 3 years retention period, 1 jurisdiction that has a 2 years retention period.	5	years	
HUM05	Human Resources	Background checks	Records of verification of candidate and employee education and previous	2	years	From the date on which the vacancy was	China: permanent, Netherlands: 4	We have based this on the fact that 4 jurisdictions that have an	0	days	

filerskeepers

www.filerskeepers.co

Connect your AI to 40+ million laws and court cases!

Lawstronaut is the infrastructure layer that connects AI agents and software to millions of legal documents across jurisdictions — structured, updated, and ready to use.

40+ million legal records | **150+** jurisdictions | **Continuous** legal updates | **API and MCP** access

Get access

Why are we here?

- **Legaltech AI Struggles with Poor Data Quality** - Legal data is often: incomplete, outdated, hard to access and complex in language and context specific
- **Data is Key for AI Accuracy in Law High-quality**, comprehensive data is essential for training reliable AI models in legal applications.
- **Impact: Inaccuracies and Distrust in AI** - Poor data leads to AI errors, reducing trust and effectiveness in legal tools (see Moore v. City of Del City (2025)).
- **Lawstronaut Improves Data for AI agents -** The platform aggregates, organizes, and updates legal data to support better AI solutions.

lawstronaut

Not a crawler. Legal infrastructure.

Built by the makers of filerskeepers to turn fragmented legal sources into machine-readable legal intelligence.

Today and rapidly counting:

40M+

Legal records

legal documents and metadata at scale

150

jurisdictions

countries and states available today

Daily / weekly

updates

monitoring legal change as it happens

40

team members

legal, data and engineering expertise

The result: a legal data pipeline that discovers, parses, structures, versions and serves laws for software and AI.

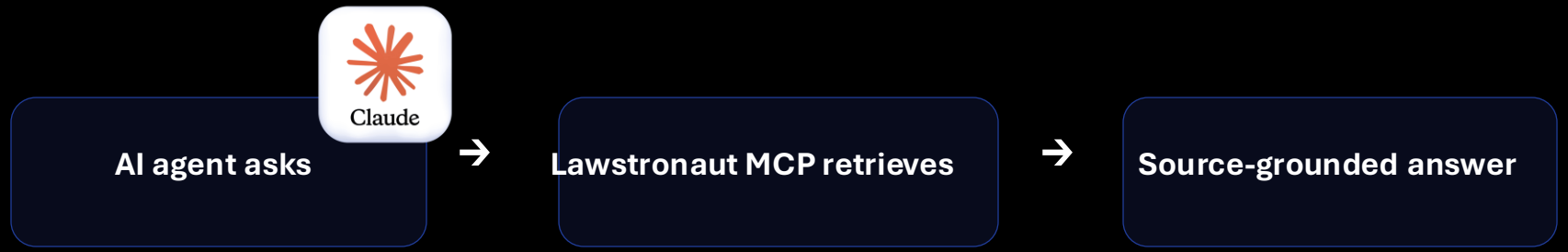
lawstronaut

lawstronaut.com

Plug legal infrastructure into any AI agent.

Lawstronaut MCP gives software a live, structured connection to the legal universe.

“What changed yesterday on AI, privacy or employment law across my key jurisdictions?”



Horizon scanning • cross-jurisdictional research • compliance monitoring • legaltech AI

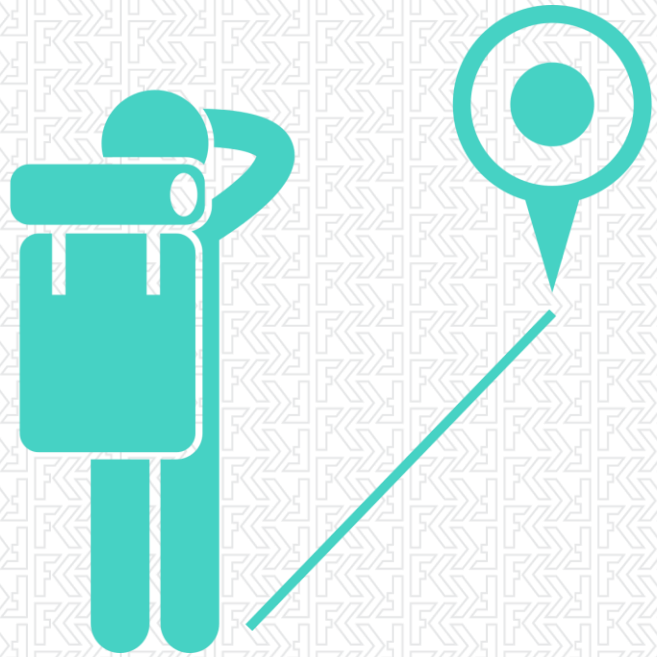
From “searching the law” to asking your systems what changed — and getting the underlying sources back.

lawstronaut

lawstronaut.com

DELETE

We've built an empire of keeping



filerskeepers

www.filerskeepers.co

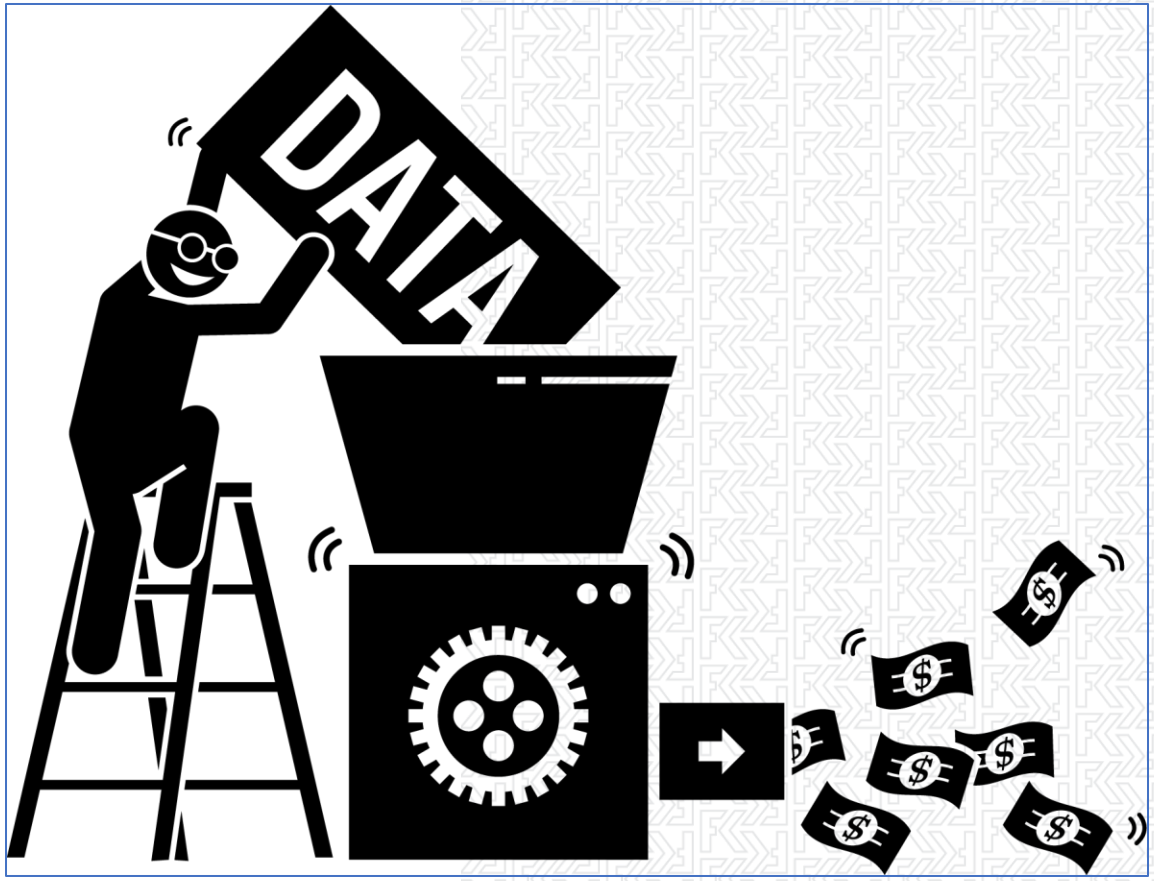
What makes deletion so hard?

- **Fear –** “What if we delete something we need?”
- **Accountability gap** – “No one gets rewarded for deletion, only punished if it goes wrong.”
- **Emotional attachment** – “my job depends on the data”
- **Complexity** – “We don’t know what the rules are and where all the copies are.”

filerskeepers

www.filerskeepers.co

The “pledge” of AI



filerskeepers

www.filerskeepers.co

Data itself
has little value

filerskeepers

www.filerskeepers.co

It is the use of data
that determines their value

filerskeepers

Not deleting is the real risk

- If deletion feels risky, aren't you doing something wrong?
  - Data overload → security risk, privacy exposure, discovery costs.
  - Inaction is no longer neutral — it's negligence.
  - Are you making most of your data?

filerskeepers

filerskeepers

www.filerskeepers.co

Real compliance!

filerskeepers

www.filerskeepers.co

RIM Becomes Mission-Critical in the Age of AI

- AI and ML depend on:
  - Clean
  - Classified
  - Structured
  - Governed
  - Trustworthy

…data.

filerskeepers

www.filerskeepers.co

filerskeepers

10

reasons your Copilot is failing.

RECORDS MANAGEMENT

THE PROBLEM

Your Copilot isn't broken. Your records are.

Microsoft 365 Copilot inherits whatever mess you've already built. Default sharing is set to **Everyone**. **ROT content** has no expiry date. **Retention schedules** don't exist or aren't enforced. **Sensitivity labels** are missing or inconsistent. **Duplicates** live across SharePoint, Teams, and OneDrive with no canonical source. **Drafts get cited as finals**. **Orphaned content** from leavers stays indexed. **Permission inheritance** is broken on sites no one has reviewed in years. **Content sprawls** in from local drives, email, and shadow tools with no provenance. And without consistent **information architecture**, Copilot ranks noise over authoritative sources.

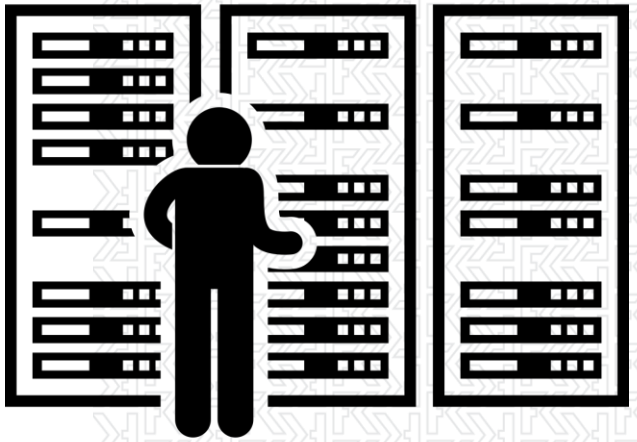
Ten problems. One root cause: **records management.**

Get the records right. The AI follows.

filerskeepers.co

AI = more data

- **AI generates data**
  - AI can create data
  - E.g. new images, data points
- **AI analyses more data sources**
  - AI is good at handling complex information
  - Wider variety of data sources (unstructured? No Problem)
- **AI demands more data for training**
  - Vast amounts of data necessary
  - The more sophisticated the model, the more data is needed



To keep or to delete?

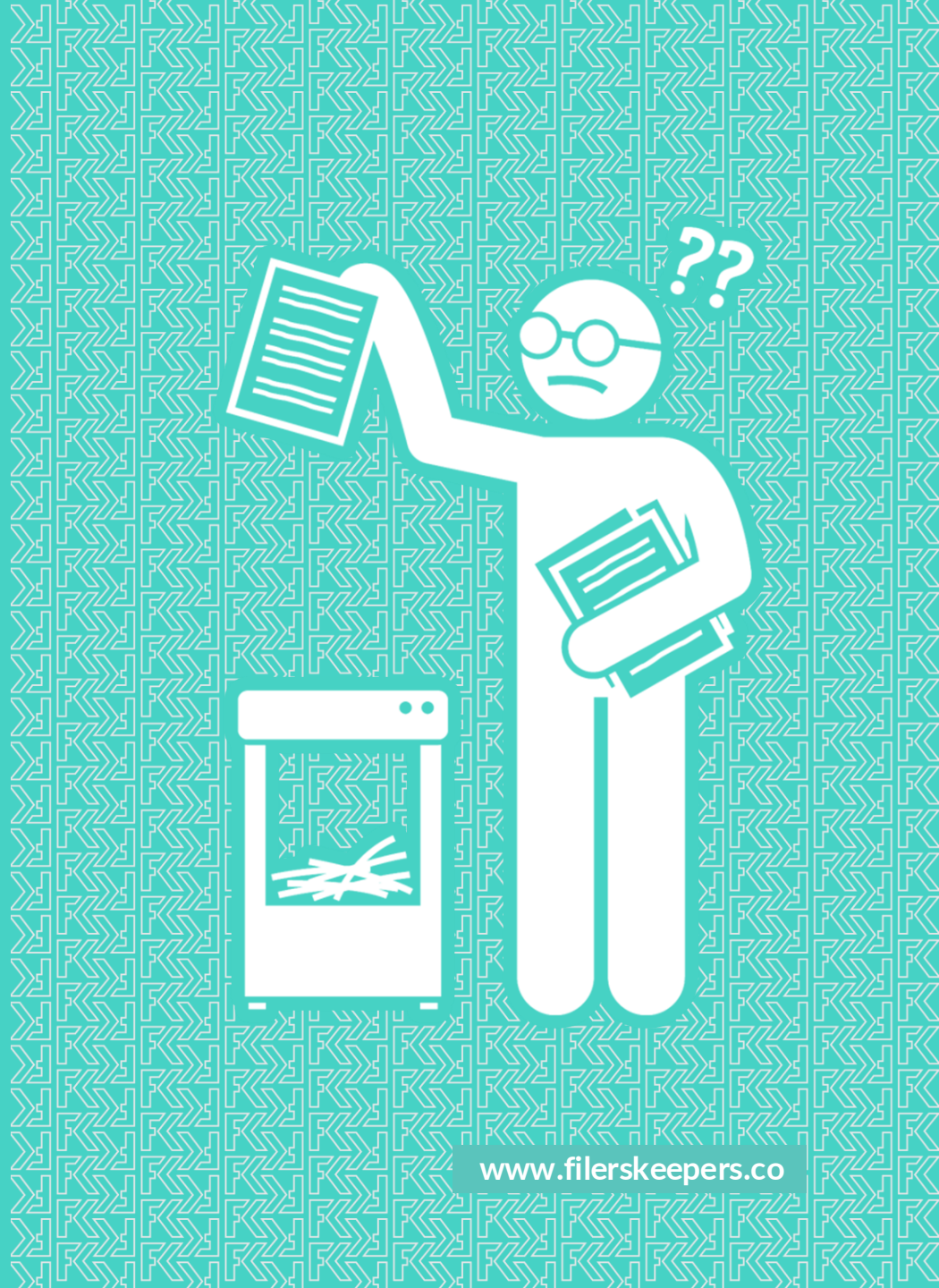
Sometimes companies want to forget:

- Laws (often data protection) tell them to destroy data
- The bigger the volume of data the bigger the breach
- Avoid holding on to evidence of breaches of contract or compliance violations
- Large amounts of data are perceived as unmanageable (and outdated data loses value)

Sometimes companies want to remember:

- Data represents tremendous value (the AI business case
- Laws tell them to retain data
- In litigations, having the right data available can mean the difference between winning or losing a case

But...
how long to store records?



filerskeepers

www.filerskeepers.co

The three retention issues of any organisation

- **Not knowing what the laws say**
- **Not knowing what to choose**
  - Laws are conflicting
  - There are so many
- **Not knowing where to start implementation**
  - Hardcopy vs digital
  - Global vs local
  - Cloud vs on premise
  - Structured vs unstructured
  - Current vs legacy

filerskeepers

Sounds easy?

- A company's systems are used and accessed from many countries
- The laws of all these countries will then apply
- Per country there can be hundreds record retention obligations
- These laws are often conflicting
- **Example: Payroll records**
  - Should be stored at least 50 years in Poland and Romania
  - Should be deleted after 7 years in the Netherlands (or is it 5)?

Companies are forced to make choices

Companies are waking up to the idea that:

- Keeping records forever is just not allowed in most countries.
- Compliance in one country can lead to non-compliance or loss of litigation position in another

While:

- A granular approach to records retention technically impossible
- Global systems often do not allow a per country/document approach

Result: companies will need to implement simple custom tailored golden standards to ensure compliance with most record retention requirements instead of all.

US states with a no longer than necessary requirement

- **Not just**
  - California (CPRA)
- **But also now**
  - Colorado
  - Minnesota
  - Virginia
- **And “No longer required to be retained”**
  - Arkansas
  - Oregon
  - Utah

filerskeepers

New FTC Coppa Rule Amendments

- **After the FTC orders re GoodRx, Drizly and CafePress etc. data retention keeps being an issue!**
- **FTC COPPA Rule Amendments - Data Retention & Policy Requirements**
  - Effective Date: June 21, 2025
  - Compliance Deadline: April 22, 2026
- **Key Data Retention Rules:**
  - Children's personal information cannot be retained indefinitely
  - Retention limited to time necessary for documented purposes
  - Data must be deleted after purpose is fulfilled
- **Data Retention Policy Requirements:**
  - Operators must establish and maintain a written data retention policy
  - Policy must specify:
    - Purposes for collecting children's personal information
    - Specific business need for retention
    - Timeline for data deletion
  - Published in the operator's privacy notice

Arkansas children and teens' online privacy protection act

Effective date: July 1st, 2026.

(iv) Only maintains the audio file long enough to *complete the stated purpose and improve or enhance the users' experience of the service and then deletes the audio file when it is no longer reasonably needed and does not make any other use of the audio file before deletion;*

(D) To retain the personal information of a child or teen for longer that is reasonably necessary to fulfill a transaction or provide a service requested by the child or teen except as required for the safety or integrity of the service or specifically authorized by law.

EDPB report on: AI Privacy Risks & Mitigations - Large Language Models (LLMs)

- **Retention is named a Key Risk! Unlawful Unlimited Storage of Personal Data**
  - Retaining personal data beyond necessary duration violates GDPR data minimization principles.
  - Increases risks of data breaches and unauthorized access.
- **Mitigation Strategies:**
  - Clear Retention Policies: Define specific time limits for data storage aligned with GDPR.
  - Automated Deletion: Implement mechanisms to automatically delete data once its purpose is fulfilled.
  - Anonymization/Pseudonymization: Minimize risks by reducing identifiable data in retained datasets.
  - Regular Audits: Conduct periodic reviews to ensure compliance with retention policies.
  - Continuous Monitoring: Adapt retention practices to evolving regulatory requirements.
- **Takeaway:** Robust retention policies, automated deletion, and ongoing audits are critical to mitigate privacy risks and ensure lawful data management in LLM systems.

International: Australia and China

50 retailers targeted in Fair Work record-keeping blitz

By Christopher Kelly | 17 April 2025

0 Comments

China Clarifies Cross-Border Data Transfer Rules: Key Takeaways from Official Q&A

April 18, 2025 Posted by China Briefing Written by Arendse Huld Reading Time: 9 minutes

China has released new clarifications on cross-border data transfer rules in an official Q&A, offering foreign businesses practical guidance on security assessments, personal information exports, important data identification, and easing compliance through certification and free trade zone (FTZ) policies.

www.filerskeepers.co

Data Retention

Is your privacy notice transparent enough?

The challenge true deletion

True compliant deletion is:

- Irreversible
- Auditable

True compliant anonymization is:

- Irreversible
- Auditable

Ever thought of synthetic data?

filerskeepers

Soft-deletion and pseudonymization?

Soft deletion: is a security mechanism

- Not deletion in the privacy compliance sense
- It is reversible

Pseudonymization:

- Also reversible
- Often easily identifiable when combined with other datasets

Ever thought of synthetic data?

Data deletion as the ultimate sign of trust

- Deletion is the ultimate sign of data protection and trust
- You are able to put the interest of the individual above yours
- This can increase trust of up to 50%!
- Deletion is very scary

filerskeepers

www.filerskeepers.co

filerskeepers

www.filerskeepers.co

Becoming Deletion-Ready

Deletion readiness

- **Clarity** – Know exactly what should go (know the rules and know the data)
- **Confidence** – Build legal and leadership trust in the process.
- **Capability** – Automate, verify, and audit.
- **Culture** – Normalize deletion as a positive act.

filerskeepers

www.filerskeepers.co

Why We Need Defensible Disposition
Turning “Whoopsie” Into a Controlled, Defensible Process

The Problem: How Destruction Looks to Courts & Regulators

- Unstructured, ad-hoc deletion looks accidental
- Gaps or missing data often appear as a “whoopsie”
- Regulators assume:
  - Negligence (“you didn’t retain properly”)
  - Improper motive (“you deleted to hide something”)
- In litigation, inconsistent deletion can be framed as:
  - Spoliation of evidence
  - Failure to comply with retention obligations
- Without a policy → every deletion is suspicious.

The Solution: Defensible Disposition

- It shows that destruction is:
  - ✓ Conscious (done on purpose, not by mistake)
  - ✓ Rule-based (linked to a retention schedule)
  - ✓ Validated (checked for legal holds & regulatory obligations)
  - ✓ Approved (legal, tax, compliance sign-off)
  - ✓ Documented (audit trail)
- With a policy → courts and regulators see discipline, not chaos.

What Makes It “Defensible”

- A defensible disposition program demonstrates that the organisation:
  - Follows routine, consistent processes
  - Performs independent checks before destruction
  - Keeps evidence of compliance (logs, approvals, certificates)
  - Ensures destruction is secure and unrecoverable
  - Applies the same rules to all formats (paper, systems, backups, local devices)
- This shows the company is not hiding, but managing responsibly.

The Defensible Disposition Process (High-Level)

1. Annual Review of Records & Data
    - Identify items that appear ready for disposal
    - Review storage, systems, and unstructured data sources
2. Retention Requirement Check
    - Confirm the item has met its required retention period
    - Validate using the official retention schedule
3. Legal / Tax / Regulatory Hold Review
    - Confirm no active legal holds, tax holds, or investigations
    - Obtain approvals from Legal, Tax, or Compliance

The Defensible Disposition Process (High-Level)

4. Prepare the Disposal List
    - Record codes, descriptions, date ranges, system locations
    - Circulate for sign-off
5. Secure Destruction
    - Destroy or delete so the data is not recoverable
    - Ensure copies, backups, and derivatives are included
6. Audit Trail Documentation
    - Capture approvals, methods, dates, certificates
    - Store as proof for regulators, audits, and litigation

filerskeepers

www.filerskeepers.co

Implementation of a retention schedule: basics

Starting with the exceptions: Legal Holds

- **Are you in control of your Legal Holds?**
  - Do you have a Legal Hold tool?
  - Do you have a full inventory of your Legal Holds
  - Do you know all custodians?
  - Have you lifted all legacy holds?

- **The challenge of active Legal Holds:**
  - Do they expand across departments?
  - Do they apply internationally?

- **Side step: do you keep boxes with incomplete inventory b/c they may fall under a Legal Hold?**

To prioritize or not to prioritize?

- **Tackling only the top systems:**
  - Most companies know their biggest risks
  - Usually: HRIS/HRM, ERP, CRM and Office suite

- **Tackling the entire system landscape:**
  - Where to go you have fixed the material systems?
  - You need to score your IT landscape!

- **Alternatively, tackling by business function/department:**
  - Fix HR first, then Finance, then Sales etc

Whatever you do, don't promise your leadership you will solve all their problems at once.

filerskeepers

Prioritization of Systems

- Goal: making smart choices with regards implementing the schedule

- You cannot solve all issues at once!
  - **Assess Importance & Risk:** Systems scored on data volume, sensitivity, regulatory significance, retention risk, and business criticality.
  - **Criteria for Scoring:** Include data type (personal, financial, health), compliance requirements, and operational role.
  - **Scoring Mechanisms:** Consider global vs local, cloud vs on-premise, structured vs unstructured data, and current vs legacy systems.
  - **Weighted Scores:** Assign weightings based on criteria importance, calculate scores to reflect system alignment with retention needs.
  - **Rank & Plan:** Use scores to rank systems by priority. High-level planning includes project team size, stakeholder involvement, tooling options, project timelines, and success conditions.
  - **Validation:** optionally engage stakeholders to validate and adjust scoring and planning based on operational realities.

Objective: This structured approach ensures that high-priority systems with critical or sensitive records are addressed first, optimizing resource allocation and enhancing compliance.

filerskeepers

www.filerskeepers.co

Implementation of a retention schedule: structured systems

Implementation of retention schedule into structured systems

- Goal: executable policy for data purge in structured system
- Deliver executable policy for data purge in structured system with input to global schedule including:
  - Determining frequency to execute
  - Defining filters to apply to data (e.g. worker type, company, country, etc)
  - Providing legal retention obligations to be executed
  - Mapping the retention obligations to data objects to include in execution
- Process to check data purged
- Test and validate the structured system data purge policy using data retention obligations and mitigate residual risk/make adjustments
- Execute structured system data purge policy using data retention obligations
- Brief description on the structured system data purge policy, risks and agreed process (included in structured system data purge policy file)


workday®

The three issues when mapping to a structured system

- **Scope differences between schedule and system**
  - Example: the law makes a distinction between regular leave and sick leave.
  - Tools like Workday groups these two and calls it absence.
- **Data dependency**
  - We want to delete data A but we cannot delete data A without deleting A and B.
  - A and B have different retention periods
- **Limited triggers available within the system**
  - The laws says e.g. end of fiscal year
  - E.g. Workday only allows for termination date for terminees

Targeting structured data (step by step)

Goal: executable policy for data purge in structured system

Step 1: Understand how the structured system is currently used (outcome: high level understanding of the use of the system within Sony) [Discovery]

- Which modules/subproducts are used
- What are the important source systems (if any), what are the downstream systems
- Where is the system used
- By whom is the system used
- Discuss what the archiving and purging capabilities of the system are
- Gain insights in data volumes and what kind of deletion loads would impact the performance of structured system (if necessary involve structured system for this)

Step 2: Get insight into the the structured system data model and purgeable data model (outcome: detailed understanding of the system and the purposes for which data are used)

- Delivery of structured system data model and structured system purgeable data model in Excel, specified by what is actually in use by Sony
- Review by filerskeepers
- We would like one or more workshops with subject matter experts to:
  - get a narrative on the data models
  - Understand what if the data model and the purging data models are the same or different
  - Understand data retention triggers (the moment the retention period starts running)

Targeting structured data (step by step)

Step 3: Go over a Sony data retention schedule (outcome: understanding of the retention decisions made by Sony)

- Delivery of Sony data retention schedule (to create a uniform HR data retention standard within regardless of structured system)
- One or more workshops with subject matter experts to give a narrative on the Sony data retention schedule and receive feedback

Step 4: Draft executable the structured system purge policy [Mapping & Report]

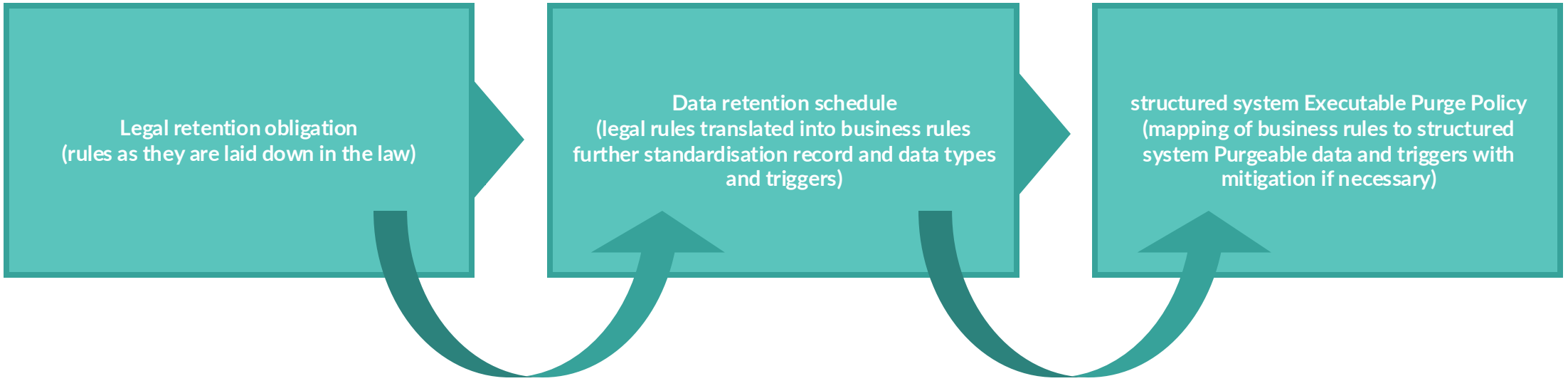
- Deliver executable policy for data purge in the structured system with input including:
  - Determining frequency to execute
  - Defining filters to apply to data (e.g. worker type, company, country, etc)
  - Providing legal retention obligations to be executed
  - Mapping the retention obligations to data objects to include in execution
- Process to check data purged
- Receive feedback from Sony subject matter experts and Legal on our draft report and purge plan.

Step 5: Test, validate and mitigate residual risk

- Test and validate the outcomes for executable the structured system purge policy
- Identify and mitigate residual practical risks and finalize document containing the end-to-end solution

Step 6: Update and finalize the structured system purge policy following input from Sony

Schematic end to end solution



This translation, mapping and standardisation is a legal exercise

This translation and mapping, adjusting of triggers is a legal and technical exercise

Building the end-to-end examples of translations of triggers and retention periods

Legal rules to business rules	Example - Law says: 10 years from the date the injured party became aware of the damages. Business rule becomes: 10 years from the date of termination of the employment agreement.
From Business Rule to structured system Purge Policy	Example – Business Rule says: 10 years from the date on which the fiscal year ends. structured system Purge Policy will say: 11 years from the date of creation

Purge plan

The Purge Plan are filters to be set within e.g. Workday to create reports of data which can be purged. They are grouped by record types in the schedule and by retention period. Deviations need a separate report within the filters.

Back to Table of Contents

Purge Plan	Area / Purgeable Data Type	Description	Ref. No.	Retention Period	Geo scope
Purge Plan - HR01 - HR-Financial - GS	Compensation	Merit statements and merit, bonus, and stock notes.	HR01	15 years	All except Bulgaria, Chile and Vietnam
Purge Plan - HR01 - HR-Financial - GS	Compensation - Additional Compensation	Merit statements and merit, bonus, and stock notes.	HR01	15 years	All except Bulgaria, Chile and Vietnam
Purge Plan - HR01 - HR-Financial - GS	Compensation - Core		HR01	15 years	All except Bulgaria, Chile and Vietnam
Purge Plan - HR01 - HR-Financial - GS	Compensation - Merit Statements, and M	Merit statements and merit, bonus, and stock notes.	HR01	15 years	All except Bulgaria, Chile and Vietnam
Purge Plan - HR01 - HR-Financial - GS	Payroll	Worker's External Payslip Attachment and Comment	HR01	15 years	All except Bulgaria, Chile and Vietnam
Purge Plan - HR01 - HR-Financial - GS	Payroll	(DO NOT USE) Attachments for Worker's External Pa	HR01	15 years	All except Bulgaria, Chile and Vietnam
Purge Plan - HR01 - HR-Financial - GS	Payroll	Worker's External Tax Document Attachment and Com	HR01	15 years	All except Bulgaria, Chile and Vietnam
Purge Plan - HR01 - HR-Financial - GS	Payroll	Worker's External Payroll Document uploaded as "Oth	HR01	15 years	All except Bulgaria, Chile and Vietnam
Purge Plan - HR01 - HR-Financial - GS	Payroll	Worker's External Payroll Input Comment	HR01	15 years	All except Bulgaria, Chile and Vietnam
Purge Plan - HR01 - HR-Financial - GS	Payroll	Payment Elections for Worker	HR01	15 years	All except Bulgaria, Chile and Vietnam
Purge Plan - HR01 - HR-Financial - GS	Custom Object	Worker - End Date 30% Ruling	HR01	15 years	All except Bulgaria, Chile and Vietnam

Residual risks

Implementing in structured systems isn't all perfect, due to its functionality. Therefore, we list any residual risks we have come across.

Back to Table of Contents

Residual risks of non-compliance noted by filerskeepers	Impact	Mitigation
For terminated employees the retention trigger (start of the retention period) can only be from the termination date. Some laws can require [Company] to purge data well before the the termination date.	In some cases terminated employee records can be kept longer than necessary.	To be mitigated when Workday offers additional trigger moments for terminated employees in addition to termination date
Terminees: absence records include both sick leave data (HR-06) and regular leave data (HR-16) these are aften regulated by different laws and retention periods.	Philips could be non-compliant with either sick leave or regular leave retention periods if they deviate from one another. This will depend on the country specific regulations.	To be mitigated when Workday offers additional functionality.
A terminated employee in one country can get hired in another. WD sees it as an active employee while the law sees it as a terminee.	In some cases employee records can be kept longer than necessary. However, we can see why it would be nice for the business to have some more history on employees.	TBD
Documents are a separate purgeable data type, but these documents are not distinguished by the nature of the document. Workday lumps all different documents together and Philips will only be able to set one retention period.	This can be in breach of data minimization and shorter retention periods.	We will get rid of documents in accordance with the general personnel file retention periods as applicable. Philips can also consider document management solutions for Workday like
We noted that there are a few Covid19 records kept in Workday (purgeable data type Vaccinations), though the exact extent is not known to filerskeepers. As the pandemic is over now and countries have relaxed or abolished their Covid19 regulations, associated record requirements no longer apply. The only reason for keeping such records is to deal with aftermath (e.g. unemployment/benefits claims and insurance cases).	In some cases Covid19 records are no longer necessary and should be destroyed immediately.	Consider removing all Vaccination data.

Companies have two problems

- A legacy problem
  - ”We created a mess of our Sharepoints/Teams’ sites etc”
- A future-facing problem
  - ”How do we make sure we do not end up in the same mess again in a few years?”

Can eDiscovery and scanning software be used to find electronic records eligible for disposition?

YES, but…

Data discovery and scanning software main challenges

- Most vendors are focused on cybersecurity use cases:
  - Finding data elements of high risk
  - Aim: set confidentiality levels
  - One trick ponies
  - Low expertise on records management and compliance use cases
  - No way to map a schedule to a classification
- The right way:
  - Takes a data and records perspective
  - Searches for context and uses data lineage for this
  - Ingests records retention schedules and builds end-to-end data classifications which work
  - Knows how to remediate data

Records managers and IG Professionals

You are:

- Important to us
- Special to us
- Dear to us
- Inspiring to us

You are:

- The guardians of data minimization, retention, purpose limitation and security
- The essential tool in the toolkit of a Chief Privacy Officer, CISO, Head of Data

You know:

- Thousands of reasons to keep data
- When to delete
- Examples when talking to data protection authorities

filerskeepers

www.filerskeepers.co

Implementation of a retention schedule: unstructured systems

Provenance = The Story

- Archival / records-management concept, focused on origin, custody, context, and purpose:
  - Who created this information, under what authority?
  - Why was it created? For which business function or decision?
  - When and where was it created and used?
  - Who has held it? Has the chain of custody been secure?
  - What is its relationship to other records from the same creator?

It's like legal chain of custody + biography:
who sent it, why, what's inside, and proof it wasn't tampered with.

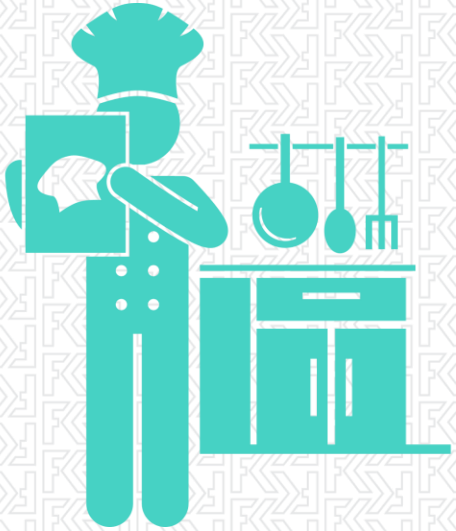
- Advantages
  - Wider scope
  - Richer Context
  - Deeper Trust (Authenticity and Reliability
  - Long-Term value

filerskeepers

www.filerskeepers.co

What is Ontology?

- An ontology is not a metadata schema.
- It is a conceptual model of your organisation.
- It tells machines:
  - What things are
  - How they relate
  - What rules govern them
  - Why they exist
- File Plan = A static map
- Ontology = GPS + satellites + real-time traffic



filerskeepers

www.filerskeepers.co

The Knowledge Graph

- Ontology becomes powerful when activated as a knowledge graph.
- It lets the organisation:
  - Classify automatically
  - Apply retention automatically
  - Trigger legal holds automatically
  - Connect records, events, people, and processes
  - Surface insight across silos
- This is RIM on a whole new level.

filerskeepers

The challenge of email

- Email is a means of a communication not a record
- Technically one should do massive manual/automatic labelling or data discovery
- Most companies:
    - are not yet ready for data discovery on email
    - Implement blanket retention periods ranging from 6 months to 10 years on average
- Don’t forget commercial correspondence laws! In some countries you need to keep emails very long
- Commercial correspondence: two interpretations
    - Broad: anyone can create obligations for the organization
    - Narrow: only those who the power to represent the organization
- By role/function?

The challenges of unstructured systems

- Unstructured systems (e.g., emails, shared drives) lack organization, complicating retention schedule implementation.
- Key Technical Challenges:
  - Data Discovery: Scattered data is hard to locate and classify.
  - Inconsistent Metadata: Lacks uniform tags for retention rules.
  - Fragmented Storage: Data across multiple platforms hinders management.
  - Oversight Risk: Unstructured data may be missed, risking non-compliance.
  - Complex Deletion: Securely erasing data from diverse sources is difficult.
- Key Organizational Challenges:
  - Insufficient leadership buy-in
  - Records management and defensible disposal policies non-existent or not working
  - Legacy problem: Sharepoints and Team sites not properly designed

Data discovery and scanning software main challenges

- Most vendors are focused on cybersecurity use cases:
  - Finding data elements of high risk
  - Aim: set confidentiality levels
  - One trick ponies
  - Low expertise on records management and compliance use cases
  - No way to map a schedule to a classification

- The right way:
  - Takes a data and records perspective
  - Searches for context and uses data lineage for this
  - Ingests records retention schedules and builds end-to-end data classifications which work
  - Knows how to remediate data

So how long to keep prompts?

Why prompts & chats are personal data

Six paths to Article 4(1) — and one default position.

SELF-ID

User identifies themselves

Age, profession, location, life situation typed into a prompt is personal data of the user. (Art. 4(1))

ACCOUNT LINK

Linked via account

Email, payment, IP, fingerprint link every prompt to the account — even "capital of France" becomes personal data plus a behavioral trace.

THIRD PARTIES

Names dragged in

"Email my manager Sarah at Acme" processes Sarah's personal data without her consent or a lawful basis.

SPECIAL CATEGORIES

Article 9 in every health Q

Health, religion, sexuality, political views — explicit consent or another narrow basis required under Art. 9.

BREYER PRINCIPLE

Pseudonyms still count

Personal if the controller has means reasonably likely to identify. (Breyer C-582/14; SRB v EDPS 2025)

OUTPUTS TOO

Hallucinated bios count

Responses about the user — and about real third parties, accurate or fabricated — are personal data. (Garante 2024; noyb complaints)

Default position: *treat all prompts and chats as personal data — the exceptions are narrow and hard to verify.*

The new EU AI Act

The core of the Act

- Recital 66:

  "Requirements should apply to high-risk AI systems as regards risk management, the quality and relevance of data sets used, technical documentation and record-keeping, transparency and the provision of information to deployers, human oversight, and robustness, accuracy and cybersecurity. Those requirements are necessary to effectively mitigate the risks for health, safety and fundamental rights, and no other less trade restrictive measures are reasonably available, thus avoiding unjustified restrictions to trade."

Three forces pulling at chat history

Where each vendor lands on this trade-off shapes its retention policy.

TRAINING

Model quality

Longer retention means more data to improve safer, more capable models. The product gets better when chats stick around.

SAFETY

Abuse detection

Flagged content is kept longer than normal chats so classifiers can be evaluated, retrained, and audited.

PRIVACY

User control

GDPR right to erasure, data minimization, and basic trust push the other direction — keep less, delete sooner.

What this means for consumer chats

OpenAI

30d

After deletion. Indefinite before.

Anthropic

30d

Or 5 years if opted into training.

Google

18mo

Default. 72h minimum even when off.

The NYT case proved one thing: a privacy policy is only as strong as the next court order.

Enterprise tiers tell a different story

Customer data is treated as protected liability — DPA-governed, no training by default, admin-configurable.

OpenAI

Enterprise · Team · Edu · API

30d post-delete · admin-set

No training by default. ZDR available on the API. Carved out of the NYT preservation order from day one.

Anthropic

Team · Enterprise · API

30d post-delete · admin-set

No training by default. ZDR available on the API. The 5-year training opt-in is consumer-only.

Google

Gemini for Workspace

Inherits Workspace retention

Inside the customer's Workspace tenant under the Workspace DPA. No training, no human review.

Microsoft

M365 Copilot

Inherits Purview retention

Most mature governance. Stored in Exchange mailbox. Only platform with 'warm storage' for FINRA / SEC 17a-4.

Gotcha: *ZDR is the strictest — but removes the audit trail. Wrong choice for FINRA / MiFID II journaling.*

Orgs using AI (not high-risk)

How long should you keep your own chats and training docs? You're not in AI Act high-risk scope.

filerskeepers

www.filerskeepers.co

Treat AI chat logs as business communications

No AI Act mandate. The dominant pattern: apply your existing email/messaging retention.

LIGHT TOUCH **3 – 36 months**	General working chats — brainstorming, drafts, casual queries. No regulatory pull.
STANDARD **2 – 10 years**	Anything touching board mandated projects, contracts, financial decisions, HR matters, or regulatory exchanges.
SECTOR-MANDATED **5 – 10+ years**	Regulated communications in finance (MiFID II / SEC 17a-4), pharma (GxP), legal.

Three buckets, three retention logics

Even out of AI Act high-risk scope, Article 4 still demands you can prove AI literacy.

ART. 4 LITERACY

Training records

5 – 7 years

Who was trained, when, on what. Treat as mandatory-training records — same as anti-bribery or GDPR training.

GOVERNANCE

Policies & AUPs

In force + SoL

Internal AI policies, acceptable-use docs, governance frameworks. Keep while live plus the longest applicable limitation period.

CONTENT-BASED

Training corpora

Follow content

Prompt libraries, RAG sources, grounding documents. Retention follows the content's own classification — not the fact it's 'AI material.'

Separate AI retention schedule or not?

Some multinationals slot AI records into existing categories on the records retention schedule.

AI chat logs	→	**Communications retention**	*Email/chat schedule*
AI literacy training	→	**Compliance training**	*5–7 years*
AI policies & AUPs	→	**Governance documents**	*In force + SoL*
Vendor contracts (DPAs)	→	**Vendor contract retention**	*Term + 6–10 yrs*
AI incident logs	→	**IT/security incident retention**	*Per existing IR policy*

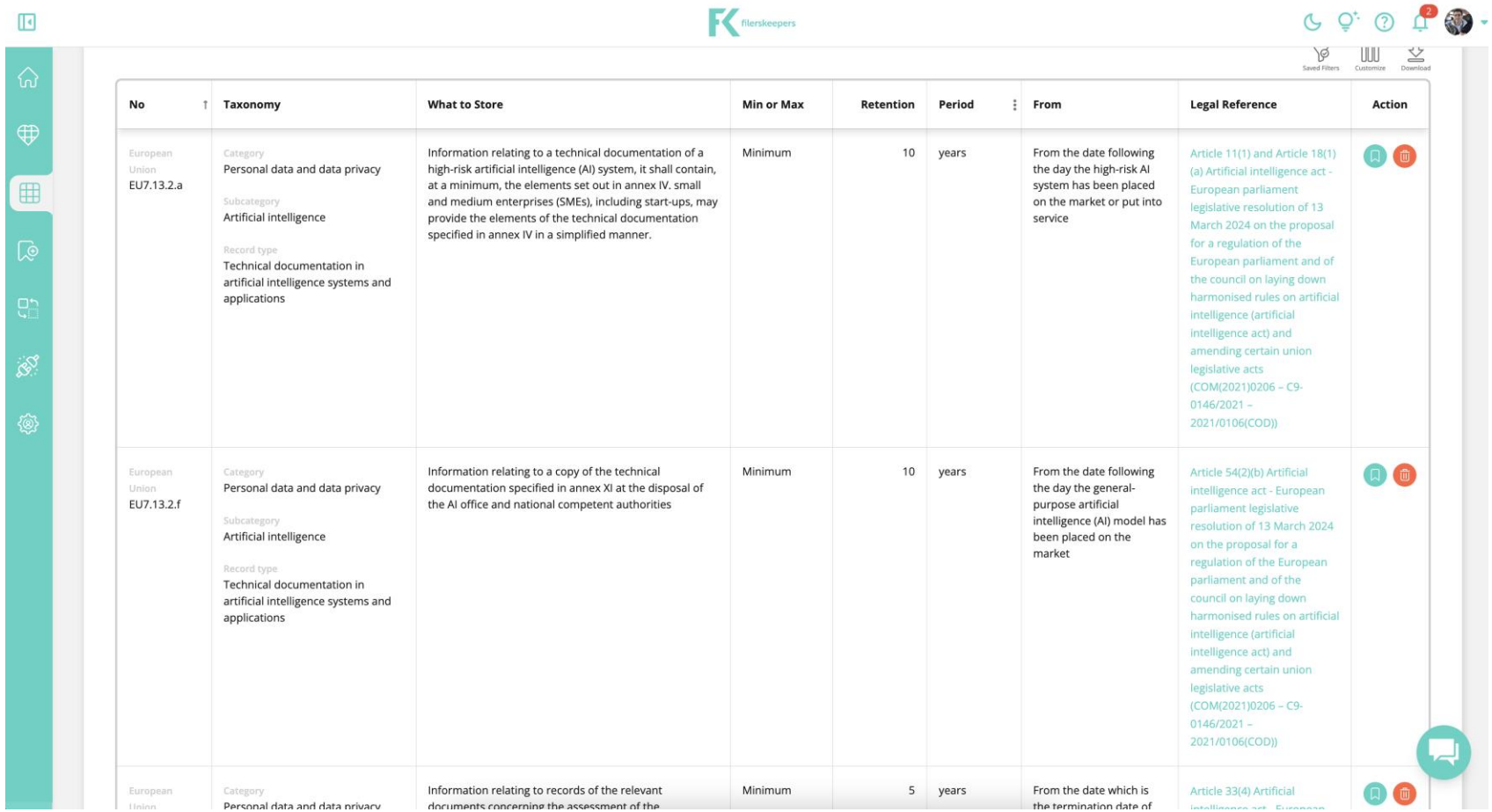
Why it works: the AI Act adds evidentiary needs without always providing retention numbers — GDPR pulls the other way.

What if...
the EU AI Act
Applies in full force?

filerskeepers

www.filerskeepers.co

Retention obligations under the EU AI Act



No	Taxonomy	What to Store	Min or Max	Retention	Period	From	Legal Reference	Action
European Union EU7.13.2.a	Category: Personal data and data privacy Subcategory: Artificial intelligence Record type: Technical documentation in artificial intelligence systems and applications	Information relating to a technical documentation of a high-risk artificial intelligence (AI) system, it shall contain, at a minimum, the elements set out in annex IV. small and medium enterprises (SMEs), including start-ups, may provide the elements of the technical documentation specified in annex IV in a simplified manner.	Minimum	10	years	From the date following the day the high-risk AI system has been placed on the market or put into service	Article 11(1) and Article 18(1) (a) Artificial intelligence act - European parliament legislative resolution of 13 March 2024 on the proposal for a regulation of the European parliament and of the council on laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD))	
European Union EU7.13.2.f	Category: Personal data and data privacy Subcategory: Artificial intelligence Record type: Technical documentation in artificial intelligence systems and applications	Information relating to a copy of the technical documentation specified in annex XI at the disposal of the AI office and national competent authorities	Minimum	10	years	From the date following the day the general-purpose artificial intelligence (AI) model has been placed on the market	Article 54(2)(b) Artificial intelligence act - European parliament legislative resolution of 13 March 2024 on the proposal for a regulation of the European parliament and of the council on laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD))	
European Union	Category: Personal data and data privacy	Information relating to records of the relevant documents concerning the assessment of the	Minimum	5	years	From the date which is the termination date of	Article 33(4) Artificial	

filerskeepers

www.filerskeepers.co

Three tiers of obligation

The 28 AI Act rules sort cleanly into three patterns — plus a residual bucket.

LONG ARCHIVE

10 yrs

8 rules

Documentation wall

Technical docs, QMS, EU DoC, notified body decisions. Clock starts from market placement.

HOT FLOOR

6 mo

2 rules

Operational logs

Automatic logs for Providers (Art. 19) and Deployers (Art. 26(6)). A floor — GDPR usually extends it.

MUST DELETE

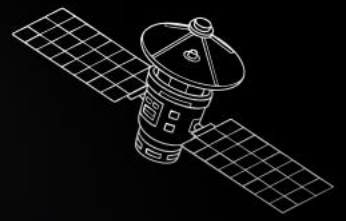
0 days

6 rules

Deletion duties

Special categories of personal data, sandbox data, rejected biometric ID authorizations.

Plus a fourth bucket: *~11 'unspecified' rules — open-ended in the Act, bounded by Art. 18 or GDPR.*



Train your AI on all the laws in the world.

Imagine a world where your AI knows every law, everywhere
— always up to date.

lawstronaut

lawstronaut.com

Let’s hear from you

Please drop your questions in the Q/A section.

Thank You!

Contact **filerskeepers**

Email: anything@filerskeepers.co

Contact **Lawstronaut**

Email: **anything@lawstronaut.com**

Scan here!

filerskeepers.co

lawstronaut.com

Web Conference
Participant Feedback Survey

Please take this quick (2 minute) survey to let us know how satisfied you were with this program and to provide us with suggestions for future improvement.

Click here: https://iappwf.questionpro.com/t/AbBPvZ8sik

Thank you in advance!

For more information: www.iapp.org

Attention IAPP Certified Privacy Professionals:

This IAPP web conference may be applied toward the continuing privacy education (CPE) requirements of your AIGP, CIPP/US, CIPP/E, CIPP/A, CIPP/C, CIPT or CIPM credential worth 1.0 credit hour. IAPP-certified professionals who are the named participant of the registration prior to the live webinar will automatically receive credit. After the broadcast date, individuals may submit for credit by completing the continuing education application form here: submit for CPE credits.

Continuing Legal Education Credits:

The IAPP provides certificates of attendance to web conference attendees. Certificates must be self-submitted to the appropriate jurisdiction for continuing education credits. Please consult your specific governing body's rules and regulations to confirm if a web conference is an eligible format for attaining credits. Each IAPP web conference offers either 60 or 90 minutes of programming.

For questions on this or other IAPP Web Conferences or recordings please contact:

livewebconteam@iapp.org