APARAVI®

WHITE PAPER - JAN 2026

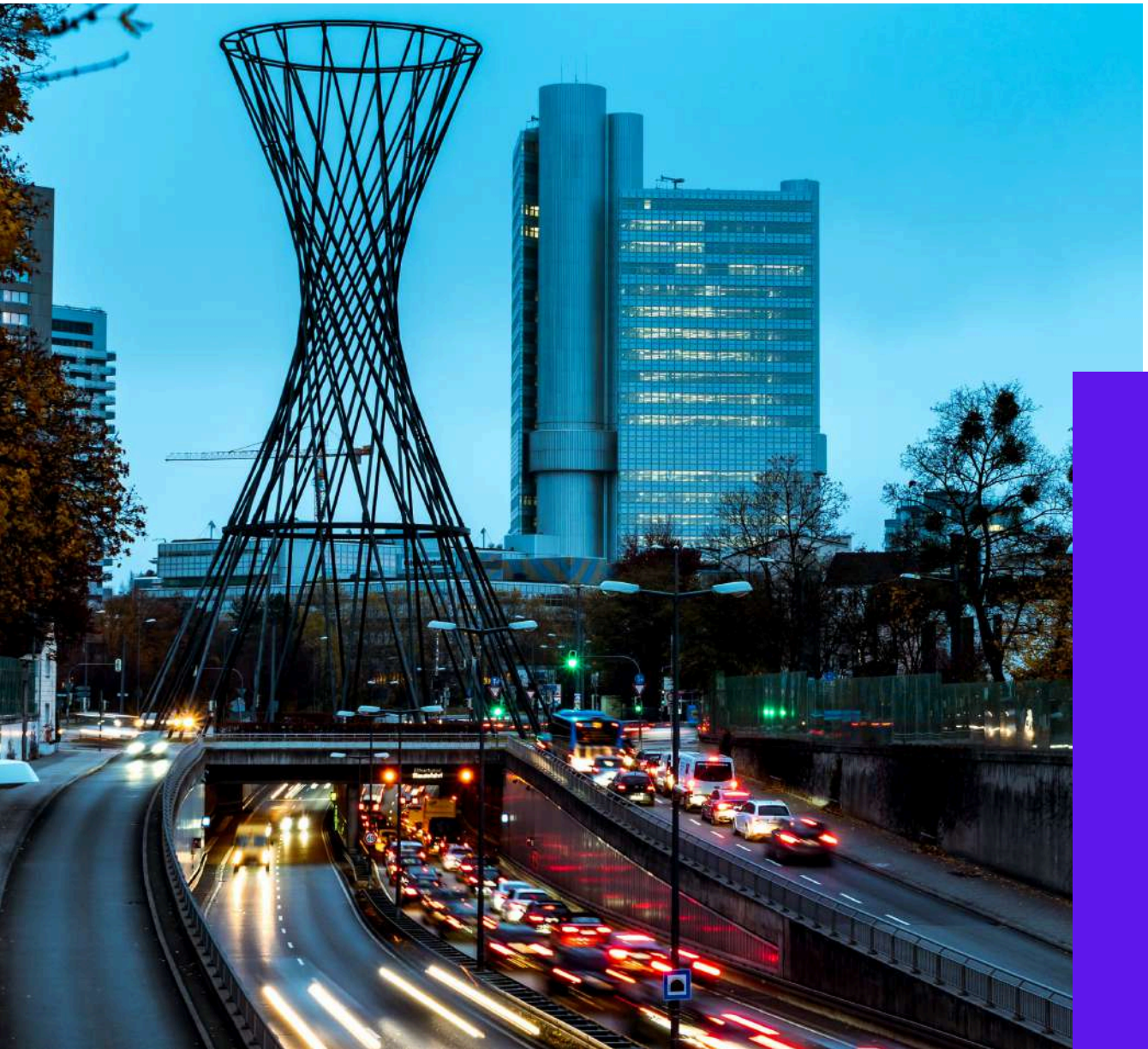
REINVENTING DATA PROTECTION FOR THE AI ERA

FROM GDPR COMPLIANCE TO PROACTIVE GOVERNANCE

A comprehensive guide to modern data governance
in the age of AI and unstructured data

Publication date: 01/01/2026

aparavi.com



EXECUTIVE SUMMARY

The Data Governance Imperative

In the current environment where AI dominates the digital, never-before-seen challenges are arising for handling the use of such huge data. Over 2,200 cases of enforcement involving fines of over €5.6 billion have been filed as of 2025, with DSAR cases rising a massive 246% since 2021. New regulatory Acts such as the European AI Act and Digital Omnibus are now imposing tough requirements on handling such data, especially that of the AI system. The existing regulatory mechanism, working efficiently with structured databases, is reaching its peak with the rise of unstructured data (email, docs, images), which currently comprises 80% of business data.

In this white paper, we discuss the need for an urgent data protection strategy that leans on automation, as well as AI-friendly governance. The business case with regard to ungoverned, "dark" unstructured data is described with regard to a shift from a reactive, human-run compliance process toward a proactive position, which is fitting for the era of AI.

80%
of enterprise data is unstructured

€5.6B
in GDPR fines levied to date

72%
surge in DSAR volumes since 2021

Key Insights and Recommendations

The AI-Era Data Challenge

Most organizations do not have the visibility to their unstructured data. 66% of the organization's employees confess that they do not know the value of the company's data, and 31% of them believe that the most important information about their company is stuck in a silo. This lack of transparency makes it difficult to comply with privacy laws, particularly when artificial intelligence is introduced as a method of consuming more data.

Regulatory Convergence

The enforcement activity on the GDPR is still strong (fines of €1.2 billion in 2024 alone) and is now aligning with new regulations. The EU AI Act (applying from 2025) obliges best data governance practices on high-risk AI, and there is also a Digital Omnibus package that updates the GDPR itself to align with AI and research, such as stating that the training of an AI model may be a "legitimate interest" under the GDPR.

Benefits of a Data Governance Framework

There are various benefits associated with a data governance framework. One of the benefits is that it ensures data quality. Data quality is necessary in businesses as it helps in decision-making. The data governance framework also ensures data security. We set forth an updated approach to unstructured data governance and DSAR management in four steps, transformed for the automation and AI era: (1) Automated discovery and intake, (2) Extended discovery & lineage tracking, (3) Layered transparency & oversight, and (4) Continuous remediation & policy enforcement.

THE CHALLENGE

The AI-Era Data Governance Challenge

Even now, seven years after the start of GDPR, it's still a challenge for many organizations to have visibility into and control over their data, especially unstructured data, which is so pervasive under GDPR as well. It's no longer just a database and CRM deal. Think about this: 80% of the data in the enterprise is unstructured (docs, emails, photos, logs, and so on), and it's growing at a rate of ~60% per year.

While performing GDPR compliance exercises, it has become clear that the lack of structured information inventory and management is a challenge in itself. According to one assessment, most companies confessed that they lack understanding of their unstructured data content and risk. Key questions such as "Whose personal data do we have, and where is it stored?" remain unanswered.

Key Insights and Recommendations

This level of opacity leaves critical compliance blind spots. GDPR requires (Art. 15) that any EU citizen request a complete copy of all their personal data that a company may hold, but if the personal data of a company is hidden away in PDFs, emails, or free-text files, can a company realistically comply with such a request? The truth is that ad-hoc processing has long been the norm. Quite often, companies have relied on one or two IT professionals to dig through various systems when a DSAR is received.

They could easily extract information from a customer database, but if the personal information is embedded in unstructured form, it often goes undetected. For instance, the HR department might download a candidate's CV to a local drive or an employee might save a customer's email or ID photo to their personal folder, it often slips their minds in the course of the usual data scan. This leads to an incomplete answer which may contravene the GDPR provision itself. Worse, if a person's personal information has not been extracted and erased when requested, the firm might unknowingly continue to store information they were supposed to delete (Art. 17 "right to erasure").

> Since unstructured data is difficult to search, processing GDPR queries becomes "a hard reality: manual, time-consuming, and error-prone." It may take weeks of work when a query is raised

66%
of employees don't understand data value

25%
of work week spent searching for information

$1,524
average cost per manual DSAR

With an average manual DSAR cost estimated at $1,500+ in administrative costs, it is clear that companies are ending up with tens or even hundreds of thousands of dollars simply to address these requests, and this is not sustainable as awareness on privacy continues to rise. Additionally, there is a skill gap, where most IT and compliance teams are not familiar with emerging best practices in handling unstructured data or AI data. In fact, only 1 in 5 privacy professionals is confident that their organization has its privacy compliance in hand.

REGULATORY LANDSCAPE

Regulatory Convergence: GDPR Meets the AI Act

The regulatory authorities in Europe have not been idle either since GDPR came into effect back in 2018. We are at a juncture today where we are seeing a convergence around data regulations. And these include changes within GDPR and similar acts such as AI, which are bound to shift the paradigm on compliance.

GDPR Enforcement Trends

It took some time for GDPR enforcement efforts to get up to speed, but there's no sign of it slowing down, at least not with these big tech giants. By early 2025, more than 2,245 GDPR fines were handed out, amounting to approximately €5.65 billion. It's worth noting that 2023 marked a historic milestone with the first-ever billion euro fine imposed on Meta (Facebook) at €1.2 billion with regards to transferring personal data from the EU to U.S. servers.

The enforcement authorities began 2024 with hefty fines but with a cumulative value of approximately €1.2 billion, which marked a 33% drop compared to 2023. The big tech sector continues to be a focus, with, for instance, a €310 million fine imposed on LinkedIn and a €251 million fine on Meta. However, enforcement efforts have increasingly targeted other industries. A major Spanish bank paid a fine of €6.2 million, while an energy firm in Italy paid a €5 million fine for misuse of old customer data. These examples illustrate that no sector is immune: Finance, Utilities, Retail, and SMEs, all are liable under GDPR and are increasingly on the radar of regulators if they don't get their data act together. A significant observation here is that more than 80% of GDPR fines imposed in 2024 were linked to security lapses causing data leaks, and thus data security and GDPR are no longer two separate concepts.

2,245
total GDPR fines issued as of 2025

€1.2B
largest single GDPR fine to date

80%
of 2024 fines related to security breaches

Year	Total Fines	Number of Cases	Largest Single Fine	Key Trend
2023	€2.1B	~438	€1.2B (Meta)	Cross-border transfers
2024	€1.2B	~600+	€310M (LinkedIn)	Security lapses
2025 YTD	€5.6B+ (cumulative)	Incomplete data	€530M (TikTok)	AI data governance

The EU AI Act

At the same time, the European Union adopted the first global comprehensive AI regulation. The AI Act, adopted towards the end of 2023 and implemented from 2025 onwards, introduces a risk-management paradigm for AI systems. Although it is not a data-protection regulation, there is an overlap with data governance because it mandates handling the data that powers AI.

Digital Omnibus and GDPR Amendments

It's not alone, as it is making changes to AI rules; it has also set out a package of amendments commonly referred to as the Digital Omnibus with an intention to simplify and modernize existing digital laws, including GDPR. It's the first time since 2016 that changes are being suggested within GDPR itself. A proposed change tightens up the definition of "personal data" with a more nuanced focus on contexts. It would effectively mean that if it can't be reasonably associated with an identity by itself, then it doesn't qualify as personal data for that controller. It would be some recognition of pseudonymization, as it might exempt pseudonymous analytics data from GDPR if it couldn't truly identify people.

A further suggested amendment liberalizes processing grounds: it would specifically include "development and testing of AI models" as a legitimate reason for processing data, taking into account that AI innovation requires some leeway on GDPR grounds. Also, scientific research will get a lift on GDPR; it would be made clear within GDPR's Omnibus that additional processing for research or AI model-building will be considered consistent with an initial set of purposes if adequate measures are put into place.

The Digital Omnibus also addresses some challenges associated with GDPR compliance. For example, it aims at giving businesses a chance to charge a reasonable fee or even deny unfounded and excessive data access requests, particularly if they appear to be repeated. These are specific challenges that businesses have associated with DSARs. It would also attempt to simplify some transparency obligations for Article 13/14 notices, particularly where it is impractical or unnecessary to offer a comprehensive notice, for research in the public interest, or where a person probably already knows what they are asking.

Global Ripple Effects

It should be noted that Europe's actions have a global impact on privacy and AI regulation. The so-called 'Brussels Effect' ensures that several global businesses comply with standards set within Europe. Already, nations ranging from Brazil (LGPD) to California (CCPA/CPRA) have passed regulations modeled on GDPR. In AI, China has passed algorithm regulations, and regions such as Canada and Brazil are researching AI regulations. For global businesses, it often becomes simplest and most risk-averse to choose global best practice.

> Privacy and data governance considerations have to be baked into everything, including AI. It's no longer a checkbox or a project, but an operational capability.

AI PRIVACY RISKS

Hidden Risks in AI Workflows

When organizations start adopting AI, they hurry up and pour their data into ML and LLMs. Yet, they might overlook some previously unidentified threats with regards to data privacy, even if they have conventional data-protection projects. Leakage of PII via embeddings and ways to control cross-border data flow for AI services are two problems tightly intertwined.

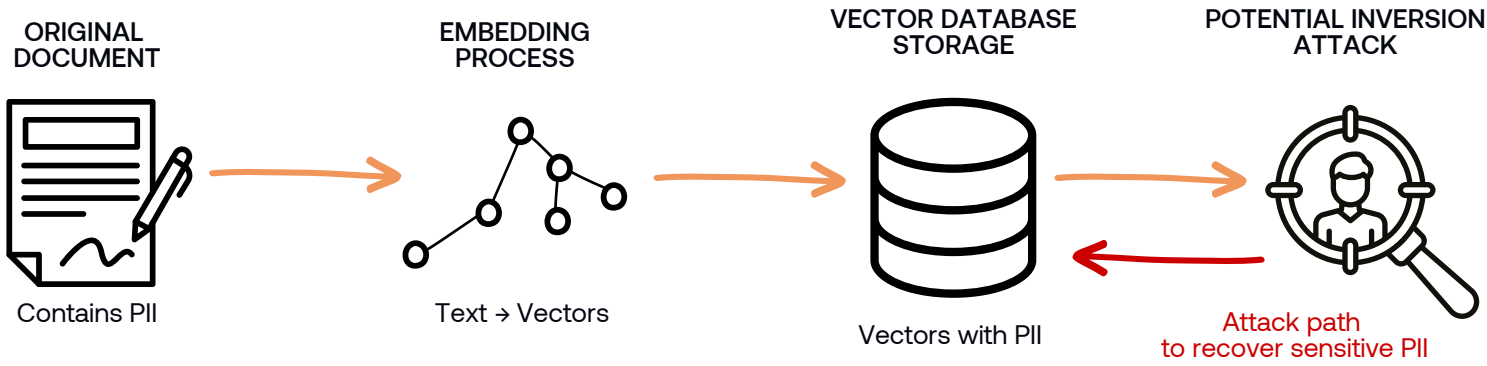
Embeddings and the Risk of Data Leakage

A number of modern AI models involving Retrieval-Augmented Generation, or semantic search, for instance, make use of embeddings. Embeddings are numerical representations of data like text or images within high-dimensional vectors. To illustrate, an LLM can search for answers within your organization's documents, all of which have been represented as vectors within a customer service chatbot. However, these vectors have an unwitting role as carriers of your personal data.

Although embeddings are designed with semantic meaning in mind, they end up encoding a considerable amount of information from the original text, including potentially private information. A vector can be looked at as a compressed picture of text. Should the text include someone's name or email address, for instance, there might be traces of it encoded within the vector. There have been inversion attacks done on embeddings obtained from an AI model, and it was shown that someone could backtrack and obtain the original information using approximate methods. It was shown that snippets from the original data, for instance addresses and phone numbers, were rebuilt from embeddings with great accuracy.

> This can be considered a new form of data leakage. Conventionally, if you have secured your databases and files, it is assumed that your data is secured. However, with these new developments, assuming your raw files are secured, because an engineer created an embedding index of your raw files and an attacker accesses it, your personal information can still be leaked.

Vector Embedding Privacy Risk Flow



Treating AI Embeddings as Regulated Personal Data

Furthermore, as embeddings are merely an array of numbers, they may not be considered personally identifiable information by conventional protection systems. But then again, there are tough GDPR questions: Do embeddings qualified as containing PII fall under GDPR guidelines? It probably does if it can be used as a tool for identifying a person. Should it be considered as an aspect of an individual's ‘personal data’ should they make an access request?

From an organizational perspective, it implies that AI data pipes should have privacy safeguards just as databases. When you store embeddings of documents that might hold PII data in vector databases such as Pinecone or Qdrant, and so on, it becomes imperative that your vector databases be treated as regulated data storage systems. Only authorized individuals or components should be allowed to interact with them. When an individual requests deletion under deletion rights, it becomes necessary that you remove not only documents but also companion vectors within your system.

Cross-Border Data Flows & AI

Another silent risk could arise from where the calculations are done. Many businesses rely on cloud computing-based AI or store data on third-party AI-based APIs. Where these systems lie outside either an EU jurisdiction for personal data within the EU or beyond any jurisdiction at all may put companies against rules on cross-border transfers. The highly publicized case involving a €1.2B fine against Meta in 2023 relates to cross-border transfers from the EU to U.S. servers without necessary safeguards, following Schrems II and invalidating Privacy Shields.

AI workflows might include these transfers, for instance, personal data sent to an API for an LLM based in the US, or cloud AI platforms on which data might be processed by engineers based in India or China. These transfers will all require a legal ground (such as SCCs and transfer impact assessment under GDPR), requirements that might be missed by data science teams in the hurry to get analysis completed.

Current enforcement activity evidences that it is being carefully considered. A €290 million fine imposed on a ride-sharing service in August 2024, widely reported as Uber, relates specifically to cross-border transfers to third countries. Evidently, it would appear that it is not solely big tech businesses that risk penalty as a consequence of an inadequate legal structure if they are processing EU personal data across borders.

THE FRAMEWORK

Modern Data Governance Framework for RAG/LLM Systems

To properly safeguard data within this ever more intricate scenario, a holistic and updated data governance strategy should be adopted that encompasses RAGs and LLMs side by side with conventional IT. From our analysis, an optimal strategy would still comprise the traditional 'find, manage, fulfill, and delete' steps, as originally outlined within the 4-step DSAR-processing practice, but with these steps perfected by capabilities specific to AI. Below are the steps:

1 Unified Intake and Automated Discovery

As soon as a privacy request (such as a DSAR or deletion request) is received, it should launch automated discovery against all data silos. And this should include not only structured data sources (CRM, ERP, HR systems), but also unstructured sources (file shares, email archives, collaboration tools such as SharePoint and Microsoft Teams, cloud storage such as Google Drive, and so forth).

Contemporary data discovery solutions employ connectors and AI algorithms to search data assets for personally identifiable information. These solutions can analyze documents, emails, images (with OCR), a feat impractical and unnecessary without these tools. By applying these technologies, an organization can locate useful personally identifiable data within minutes instead of weeks. A specific tool, for instance, accesses on-prem and cloud data sources and analyzes millions of documents for personally identifiable information with minimal human intervention at speeds that can achieve 100,000 words per second.

Notably, the search should also extend to vector databases and AI model repositories, for instance, if you have a Qdrant index for document embeddings, your search should include finding out if there were any embeddings associated with the person. All findings will be centrally logged.

2 Contextual Identification and Lineage Tracking

After identifying data, the next process would be contextual identification, which aims at understanding what information relates to the requesting person as well as its location. In today's world, data belonging to a data subject might be contained in structured information, such as a customer database entry, unstructured documents, for instance, a PDF report that includes that same data subject, and even in AI-created responses.

Lineage recording becomes even more pressing in AI. We would like to record the "path" of personal data: for instance, Alice's email address was obtained from System A, then replicated into a CSV file, which was then used to train Model X, which then produces outputs that potentially include Alice's information. When we receive the delete request, we know based on lineage that we not only delete Alice from System A but also delete that CSV file and perhaps train Model X again.

3 Layered Transparency, Controls and AI Oversight

The third step relates to taking action on the data discovered, to supply the necessary information to the data subject and put proper controls in place if it is determined that the data will still be processed. According to GDPR articles 15-20, as a response to a DSAR notice, it is necessary to provide a copy of the personal data and additional information regarding, among other things, the purposes of processing and the source of the data.

As we enter the AI era, it would be necessary if an AI system were processing your data (for instance, making decisions on you), and it should be disclosed at least in summarized form. It would be a good practice to document algorithmic decision-making and be ready with these cited facts within responses or privacy statements.

From a governance perspective, role-based access control and need-to-know should be imposed on all personal data stores, and also on AI-related ones. It would be a good practice to assign permitted purposes for your data and make your systems review these before accessing.

A large piece of this puzzle is data quality and minimization, to make sure that data maintained and processed is as accurate as it can be and no more. A more recent trend is pre-embedding data sanitizing. When text data goes into an embed model, before it goes into the model, remove or mask PII. So instead, put stand-ins for names.

4 Automated Fulfillment and Deletion

The final stage would be to carry out the actions that are required. These would be delivering a full response to an individual for an access request and deletion of personal data. This would be greatly aided by automation. After identifying and validating the data, it can be compiled as a response report by an organization for a data subject.

As for deletion requests, it would be more useful to have automation. It would have been common knowledge within an organization that an organization might have been hesitant to proceed with completely deleting data because it might have broken an app, or an organization might have failed to locate all copies. The best practice here would be to have policy-based deletion.

A current issue might be deletion within AI models. You cannot hit an easy “erase” button on an AI model's memory. Actual methods include deleting the data from the training data set and retraining your model so it does not have that data. A third option would be to insert the data into a block list during testing. It would be imperative to record your actions.

By automating intake all the way to deletion, it will be possible for enterprises to reduce the processing time and efforts required. As an example, a global organization that implemented an automated DSAR process noticed that it took less than a week to process requests compared to 4-6 weeks before. It also saved 50% of man-hours spent on searching for data.

ROI & VALUE

Automated Compliance: From Reactive to Proactive

Although deploying the above framework requires more than just preventing penalties, it completely changes an organization's approach to information and, as a result, brings tremendous benefits. To better grasp what implications deploying a proactive and automated compliance strategy might have on the three corners that form ROI on every business project, time, cost, and risk, it becomes necessary to discuss the issue of transition from a reactive position.

Time Savings and Productivity Gains

As discussed, manual data protection tasks consume enormous time. With DSAR volumes climbing and data volumes exploding, a manual approach simply doesn't scale. Automation and smarter data management can reclaim much of this time. For example, if employees currently spend 25% of their week searching for information and addressing data management issues, even automating half of that search effort gives back an eighth of the workforce's time.

One specific metric: automated data discovery can reduce search time by up to 98% for privacy inquiries. This means what took 10 hours of human hunting might be done in 12 minutes by the tool, with the human just reviewing results. Multiplied over dozens or hundreds of requests per year, the saved hours are huge.

98%
reduction in data search time

25%
of work week spent on data tasks

50%
reduction in staff hours for DSARs

Metric	Manual Process	Automated Process	Time/Cost Savings
DSAR Response Time	4-6 weeks	5-7 days	85% faster
Cost per DSAR	$1,524	$305	$1,219 saved
Data Discovery Time	10-15 hours	15-30 minutes	98% reduction
Staff Hours per Request	40 hours	8 hours	80% reduction
Annual Cost (100 DSARs)	$152,400	$30,500	$121,900 saved

Cost Reduction and Efficiency

Compliance functions as a cost center, but with intelligent automation, it becomes an equation with an opposite result. It costs approximately $1,500 per DSAR on average and as much as around $20,000 for more intricate matters. Processing 100 DSARs per year would generate at least $150,000 yearly. By automating responses and discovery, this expenditure would decrease by 70-80%, thus reducing costs by $100,000-$120,000.

IMPLEMENTATION

Best Practices for 2026

A vision for data governance in an AI age might be intimidating, but it can be broken down into steps. Below are some guidelines and tips on what needs to be done, gleaned from regulatory advisories and what works well in practice:

1. Know Your Data: Create a Unified Data Inventory

You can't protect what you don't know you have. First, start your efforts by expanding your data map to include every repository of data, not just your databases but cloud storage drives, collaboration sites, and storage solutions like data lakes and ML datasets. You can sub-categorize your data based on what it contains, customer info, employee info, and so on, as well as based on regulations, GDPR, CCPA, and so on. There are tools like data cataloging software that will help you classify your data with AI.

2. Embrace Automation and AI for Compliance Tasks

Use tech for your discovery, DSAR fulfillment, and monitoring. Stop counting on humans manually scanning documents, and use a privacy management software solution or eDiscovery tool capable of scanning multiple systems at once. Set up an automation for your privacy requests: if possible, link your request form with your back-end system so that it automatically sets off a scan and opens a case ticket. Use templates and an AI assistant for responding.

3. Integrate Privacy by Design into AI Development

Privacy needs to be a forward-thinking consideration, and not an afterthought, within any given AI project and data analysis endeavor. Run DPIAs for new AI systems that handle personal data. A DPA is mandated within GDPR on high-risk personal data processing, and it's most likely that most AIs will qualify. Within DPA, there should be an identifier on exactly how data feeds an AI, as well as its subsequent storage and handling of outputs. A remedy plan within DPA based on resultant risks.

4. Strengthen Access Controls and Accountability

Again, look at who can see your personal information. A principle of least privilege should be followed here. A role should receive no more information than it needs. This becomes even more vital for data scientists and consultants who have a habit of requiring access to as much data as possible. It should be ensured that if they are testing on production data, it gets masked or logged.

5. Data Minimization and Retention: Keep Data Lean

Battle the ‘data hoarding’ culture. Historically, data stores were retained for all eternity because they might be useful someday. But the more data you store, the larger your breach risk if things don’t go as planned. Adopt a ‘minimize and delete’ approach and get rid of data that isn’t necessary for sound business or legal reasons. Develop data retention schedules and schedule deletion actions wherever possible.

6. Upskill Your Team and Foster a Privacy Culture

It will take people and culture, not just tools. You should invest in training your employees on data protection and AI ethics. Educate your software developers and data scientists on basics about GDPR and Privacy by Design. You should offer workshops on new tools and technologies for compliance. You might want to set up ‘privacy champions’ within various departments. Those would be people who are expertise on queries regarding privacy within specific departments.

7. Leverage Privacy-Enhancing Technologies (PETs)

Looking ahead, you might want to integrate more advanced PETs into your data operations. Methods such as differential privacy (where statistical noise is added to data so that no specific information can be tracked) might enable knowledge extraction without direct access. Federated learning will enable AI modeling on data silos without consolidating your raw data. Your data can be protected end-to-end with processing, using either secure enclaves or homomorphic encryption.

ROADMAP

Implementation Roadmap

To achieve the above, we propose an implementation plan. It assists in breaking down the process into more manageable chunks of time:

FIRST 90 DAYS

Quick Wins & Assessment

- **Appoint Roles:** It has been seen that some jurisdictions might have given your organization an obligation to nominate a Data Protection Officer. Form a core project team.
- **Current State Assessment:** Perform a gap analysis on your current GDPR and data governance position. Take an inventory of your known systems and known gaps.
- **Apply Quick Wins:** Use known patches: for instance, encrypt unencrypted databases, turn on logging if it had been switched off, and update privacy policy statements if they appear antiquated. Awareness Training: Launch an organization-wide refresher course on data privacy awareness with an emphasis on data governance.

90–180 DAYS

Foundation Building

- **Data Inventory Expansion:** Execute data discovery sweeps across the organization. Document all systems and data repositories. Begin creating the universal data map, compiling data from multiple sources.
- **Tooling and Process Introduction:** Use a Privacy Management solution or integrate an existing Information Technology Service Management for processing requests. Set up the workflow process for DSAR requests, from receiving requests and getting approval.
- **Policy Updates:** Create or update data governance policies. Examples include data retention policy and AI usage policy. These should be endorsed and disseminated.
- **Start Data Minimization:** Start purging obvious sets of ROT data. Plan for their deletion.

180–365 DAYS

Integration & Refinement

- **Process Integration:** There should be an efficient functionality of your privacy request process. It should integrate with your information technologies.
- **Internal audit or tabletop exercise:** it can be conducted with a focus on being able to address a DSAR completely. The result will enable enhancing the data map and processes.
- **Advanced Controls:** Put into practice more advanced controls that have been determined as being necessary. Perhaps your discovery tool reveals a considerable amount of your sensitive information on SharePoint. You may want to incorporate DLP at this stage.
- **Training and Culture:** Create training sessions for data stewards on using the catalog, IT on why it is imperative not to bypass processes on new PETs or anonymization libraries.
- **Metrics & Reporting:** Develop KPIs for tracking the progress on: average time to meet a DSAR, number of records with data deleted per quarter, and reduction in the rate of growth of data storage. These should be communicated to the management.

RESOURCES

Resources & Next Steps

Embarking on this journey, it's useful to leverage external resources and tools:

Regulatory Guidance

It is essential that official guidelines, such as the European Data Protection Board's guidelines on DSARs, Breach Notification, and AI Processing, are readily available. The EU AI Act documents, as well as subsequent guidelines that are set to be released, are essential to ensure that your AI governance practices are met. The development of frameworks by organizations such as ISO and NIST, including ISO 27701 Privacy Management, NIST AI Risk Management, may serve as a complement to your AI program.

Technology Solutions

Analyze vendors that specifically provide solutions for handling DSAR, workflow, data discovery, and consent. The same applies to solutions that can assist with your inventory in terms of a data catalog and handling metadata. Cloud vendors also have inherent solutions such as a classifier, DLP, for your cloud platforms. Lastly, analyze vendors that assist with dataset anonymization/bias for your AI solutions.

Community and Knowledge Sharing

Analyze vendors that specifically provide solutions for handling DSAR, workflow, data discovery, and consent. The same applies to solutions that can assist with your inventory in terms of a data catalog and handling metadata. Cloud vendors also have inherent solutions such as a classifier, DLP, for your cloud platforms. Lastly, analyze vendors that assist with dataset anonymization/bias for your AI solutions.

Expert Consultation

In particular, when dealing with tricky domains such as machine unlearning or cross-border AI, it might be a good idea to reach out to lawyers and experts. There exist sandbox initiatives led by data protection authorities in various nations, where you can seek feedback on your novel solutions for compliance. Academia is also a helpful partner because researchers might have insights into PETs or AI audit solutions if you're willing to work together.

In taking the steps described in this whitepaper, organizations can mitigate GDPR risks in a deliberate way, not checkbox ticking, but with real improvement in how that organization manages and protects that information. It is a necessary change brought on by the intersection of the regulations around privacy law and the harnessing of artificial intelligence. It is a challenge that, when met, helps avoid the dangers of non-compliance but, more importantly, helps organizations position themselves to effectively use data as an asset.

CITATIONS

References

GDPR Enforcement Statistics:

- CMS Law GDPR Enforcement Tracker Report 2025: https://cms.law/en/int/publication/gdpr-enforcement-tracker-report
- GDPR Enforcement Tracker (live database): https://enforcementtracker.com
- Legal.io GDPR fine analysis: https://legal.io/articles/5568592/Total-GDPR-fines-in-2024-reached-%E2%82%AC1-2-billion-in-2024
- JumpCloud GDPR compliance costs: https://jumpcloud.com/blog/gdpr-ccpa-compliance-violations

Data Privacy Request Trends:

- DataGrail Privacy Trends 2024 Report: https://www.datagrail.io/press/datagrail-reports-worldwide-surge-in-data-privacy-requests/ (Source for 246% cumulative increase in DSRs from 2021-2023)
- DataGrail Privacy Trends 2023 Report: https://www.businesswire.com/news/home/20230330005251/en/ (Source for initial 72% year-over-year surge)
- Termly DSAR Statistics 2025: https://termly.io/resources/articles/dsar-statistics/ (Additional context on GDPR/CCPA request volumes)

EU AI Act Requirements:

- Alation EU AI Act data governance: https://alation.com/blog/eu-ai-act-2025-data-strategy/
- European Commission AI Act official documentation: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai
- EU AI Act Article 10 (Data and Data Governance): https://artificialintelligenceact.eu/article/10/

Digital Omnibus GDPR Proposals:

- IAPP analysis of Digital Omnibus key changes: https://iapp.org/news/a/eu-digital-omnibus-analysis-of-key-changes
- IAPP European Commission proposed reforms: https://iapp.org/news/a/european-commission-proposes-significant-reforms-to-gdpr-ai-act

DSAR Volume and Cost Data:

- Termly DSAR Statistics 2025: https://termly.io/resources/articles/dsar-statistics/
- JumpCloud GDPR/CCPA compliance violations (includes $1,524 DSAR cost): https://jumpcloud.com/blog/gdpr-ccpa-compliance-violations

Cross-Border Transfer Fines:

- Irish Data Protection Commission: https://dataprotection.ie
- Dutch Data Protection Authority (Autoriteit Persoonsgegevens): https://autoriteitpersoonsgegevens.nl

AI Act Fines and Obligations:

- European Commission AI Act official page: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

European Data Protection Authorities:

- European Data Protection Supervisor (EDPS): https://edps.europa.eu
- European Data Protection Board (EDPB) Guidelines: https://edpb.europa.eu/our-work-tools/general-guidance/guidelines-recommendations-best-practices_en

Note: All statistics and regulatory information current as of publication date (2025). For the most up-to-date enforcement data and regulatory guidance, please consult official sources such as the European Data Protection Board (EDPB), national Data Protection Authorities, and the European Commission.

ABOUT APARAVI

The Experts of Unstructured Data

Born in Switzerland. Operating Worldwide.

Aparavi gives businesses control over their unstructured data. Our mission is to empower businesses to discover, classify, and unlock the hidden value of their data, reduce risks, and minimize costs.

Aparavi has over 100 employees from 15 different countries, developing, operating, and licensing the data intelligence and automation solution in Europe and the USA. Our headquarters remain in Zug, a city in Switzerland, thanks to our founder, Adrian Knapp, who prides our business on the precision that exists in Switzerland.

We assist businesses in unlocking the potential of their unstructured data, irrespective of its location. Aparavi helps you identify, interpret, and work with your unstructured data. It also helps you optimize the management of your unstructured data when processed in automated data life cycles.

Contact Information

Aparavi Corporation

1351 3rd Street Promenade
Santa Monica, CA 90401 United States

Zug, Switzerland
Hammergut 6, CH-6330
Cham/Switzerland
Website: aparavi.com

Author

Roan Guilherme Weigert Salgueiro

Developer Relations & AI Engineer
Aparavi Corporation

Roan specializes in AI-powered data governance solutions and developer advocacy, helping organizations navigate the intersection of artificial intelligence, data protection, and regulatory compliance.