



Redefining **Data Lifecycle Management** for the Modern Enterprise

Technology Considerations for
DLM, Retention, Minimization,
Deletion or Responsible Data
Disposition



An abstract background on the left side of the slide, featuring a light purple and pink color palette. It contains various geometric shapes, lines, and dots in shades of blue, green, and yellow, creating a complex, circuit-like or network-like pattern. The title text is overlaid on this background.

Redefining Data Lifecycle Management for the Modern Enterprise

Redefining Data Lifecycle Management for the Modern Enterprise

Data lifecycle management is deceptively complex.

Often reduced to finding and deleting stale or duplicate data, implementing data retention or lifecycle at scale requires a number of specializations to identify a broad sweep of shadow or ROT data; define and manage retention schedules; investigate violations, including the ability to copy and tombstone data; and responsibly and safely execute deletions.

This includes checks on legal hold or runtime impact, gathering approvals, audit capture, and automation.

Whether the driver is **data minimization** requirements in **regulations** like GDPR or CCPA, **attack surface reduction** in security, **records retention** in healthcare or financial services, **cost reduction**, or general **data hygiene in AI and BI** — data retention is becoming more critical in the enterprise.

Here are some technology considerations to do it effectively, responsibly, safely, and at scale across a modern enterprise

Step 1: Finding ROT, Target Record Types and DSR Deletion Data

In data retention, the acronym ROT is often used to define the target data of a deletion program. ROT stands for Redundant, Obsolete, and Trivial data, which can include a variety of low-value data that customers want to remove.

Some vendors limit this to:

- Duplicate files
- Stale files based on last access

However, in practice, organizations need to find a broader array of ROT data. Some considerations and examples end users should evaluate in a data retention vendor include:

Support for all types of Structured and Unstructured Data

Files are an important target for data retention programs — but they are not exhaustive. Many organizations care about unstructured data like emails or semi-structured data in data lakes, along with structured data such as databases and even redundant tables in Snowflake, Databricks, or Hive.

BigID is among a small number of companies that can support both structured and unstructured data.

Diverse Kinds of Redundant Data

Data that is an exact duplicate is common, but not the only kind. Most redundant data is not an exact copy — it's modestly modified. This is especially common with productivity files like spreadsheets and slide decks.

Even when considering only duplicate files, there are key differences among vendors. Do they rely on third-party platforms like AWS to label duplicate data (e.g., when a copy is generated inside S3), or can they determine duplication independently?

End users should evaluate:

- Can the vendor calculate whether a file, table, or other asset is a duplicate without depending on a specific data source?
- Can it detect derivative (i.e., similar but not identical) data?

BigID is perhaps the only vendor that can both determine if content is a duplicate using its own scanning intelligence and identify redundant — but slightly modified — content.

Obsolete, Expired and Stale Data

Usually when end users think of obsolete data they think of either stale data (has not been accessed in a period of time) or old and expired data (created and / or last modified beyond a certain mandatory holding period of time). However, in practice companies (and more specifically different teams) interested in retention will want more flexibility in deciding what is obsolete, expired or stale.

They want to base decisions on last access data; last edit date; period since creation; last activity or some combination. Increasingly they may also want to consider expiring consent to use as a consideration for actioning a purge.

BigID is the only data security and compliance solution with data retention that can support all of the above plus many other considerations either in isolation or in combination to trigger an action to delete or archive. Moreover BigID can do this by data or specific record type.

Identifying Trivial Data

There is no universal definition of trivial data. It's in the eye of the beholder. Therefore the only effective way for companies to define what's trivial to them is via effective data classification.

However, even here not all data classification technologies are created equal. Most only support classifying presence or absence of a specific data attribute. For finding trivial data, companies want to either look for specific types of docs or databases based business types. This requires classification that can be done accurately and consistently based either on a business taxonomy or by locating tables and files that resemble an example identified by the company.

BigID may be the only vendor that offers both types of classification to help companies more effectively identify trivial data.

Other Kinds of Target Data for Retention and Deletion

ROT - even the more broadly defined kinds of ROT - is not the only kind of data organizations deal with when it comes to retention. Other examples include archive data, privacy DSR Deletion data, data that violates a legal mandate and lastly data that may violate a company retention requirement.

For archives, end users should ensure suppliers of data retention that can detect archives like BigID.

For DSR deletion requests, only one vendor - BigID - offers both an ability to locate data associated with a data subject, determine any legal hold, execute a deletion and then ensure data is not restored.

In fact BigID is unique among data retention vendors in providing a wide array of ways to target specific data types or record types (like CVs, health records etc).

Step 2: Validating Sufficient Scan Depth and Detail and Fidelity

Often overlooked in considering data retention or lifecycle tools is the ability of the platform to scan target data with sufficient depth, detail and fidelity to ensure data is not misclassified or passed over to make a decision around retention and deletion. This also includes ensuring the underlying platform does not skip data as part of its scanning process.

Firstly, any tool should have an ability to collect metadata around subject data like creation, access, activity or edit data plus the ability to scan content in full.

Secondly, the retention product should be able to ensure data is not skipped or passed over in the scan process. Some well known data access governance (DAG) products, for example, skip large files (which often contain zip archives) or older data (which represents the older data debt an organization may want to purge). Other tools - especially some of the younger DSPMs - cannot perform full scans. By and large, they only do assessment (or what some of them call Smart Sample) scans. These quick scans are designed to confirm presence of a type of data at folder or bucket level and by design skips to a next database, folder or bucket once a few examples of data are found.

In Data Retention use cases, however, end users usually want full scans to ensure important or sensitive data is not accidentally deleted. And while BigID can perform fast assessment scans, **BigID is also the only product in the market that can optionally perform a full scan without skipping large or older files.**

Thirdly, for those tools that also support labeling it is important to validate that they also don't overwrite last modified dates which could break certain kinds of retention policies. Some overwrite this data. BigID does not

Step 3: Building and Managing Retention Schedules and Policies

Decisions around data retention pivot around policies that define retention or deletion violations and rules. These policy definitions in the context of data retention are commonly referred to as retention schedules. Without an ability to define, import or manage retention schedules, data lifecycle management can't scale and lacks consistency and the explainability necessary to satisfy regulators or internal controls.

BigID is the only data security (or compliance or privacy) platform offering data lifecycle management with dedicated capabilities for importing and defining and managing retention schedules including support for a simple to author retention query builder.

Most companies of any size will already possess some retention schedules they previously defined or want to import from products like Iron Mountain or Filers Keepers. BigID supports a broad range of import (with edit) capabilities.

However, as the number of regulations and internal reasons for implementing retention or deletion grow, companies will also want to empower different teams in legal or compliance or security or risk to define and build their own retention schedule policies.

BigID is the first and only Data Security vendor to offer a full turn-key low code way for both technical and non-technical users to define retention schedules and retention based data / record queries. These retention schedules also support a diversity of metadata reference dates and logical conditions that can be used for activating a retention query or action. Examples of reference metadata supported by BigID retention queries and schedules include:

- Date of Creation: When the record was first made
- Date of Last Activity: The last time the record was changed or used
- End of Fiscal Year: The close of the relevant financial year
- Contract Expiration: When an agreement officially ends
- Employee Termination: The day an employee leaves the company
- Audit Completion: When an audit process is finalized
- Legal Resolution: When a legal case or dispute concludes
- Product Discontinuation: When a product or service is no longer offered
- Last Payment: The date of the final financial transaction
- Date of Birth: For records linked to a person's age
- Event Completion: When a specific event has finished

But while some organizations want the flexibility and customizability of building their own retention schedules, smaller companies often prefer recommendations and OOTB policies.

BigID also provides this native in the product, with policies and schedules continuously updated with changing regulatory requirements in the US and internationally.

Step 4: Investigation Workflow and Review Actions

Another often overlooked requirement for an effective data retention management product is the presence of workflow and actions to aid a user - usually a data owner - to investigate and validate that certain data has been flagged for erasure or archive or cold storage.

In most companies, before data (or record) can be removed or reduced permanently, the data owner needs to validate and approve the request. This requires specialized workflows for the data owner but also actions on the data like the ability to move the data into a segregated store or the ability to leave a stub or tombstone file when data is marked for deletion (with information on what happened to original data for any unsuspecting user).

But workflow for data retention is more than just a jira ticket.

It requires an ability to provide a locator to the data, it requires an ability to provide a preview of the data, it requires an ability to explain why the data was targeted and it requires a backend RBAC control to ensure the data owner and investigator never have access to data that they are not privileged to.

BigID is the only vendor in Data Security and Compliance with native end to end workflow built in with hooks for both Jira and ServiceNow for customers that want interoperability with other workflow tools they own, across their entire tech stack.

In addition, investigations will also require actions. This could include moving the data to a segregated sandbox, cold store or staging area. It can also include replacing the moved data with a stub or tombstone so others know it was removed. It can also archive the data as a zip or even restoring data after a legal hold or investigation. BigID is alone among Data Security and Compliance vendors with all these native capabilities.

Step 5. Moving, Tombstoning, and Restoring Data

Often overlooked in data retention or deletion programs is the need to sometimes move data from a source into a holding area.

Sometimes this needs to move data into a secondary story like a cold storage is for purposes of sequestering the data during a review or investigation. Sometimes it is driven by legal hold requirements. Sometimes it is needed as an alternative or precursor to deletion including archiving the data in a low cost cold storage.

Whatever the reason, a true end to end DLM or data retention offering should include a means to move data from a source to a destination. BigID is unique among data security and compliance tools to support this.

Concomitant with the need to move data however is the related need to leave behind a stub or tombstone explaining the disappearance of the target data (or record). The tombstone acts as a memorialization of the data and an explainer of why it was removed (retention, legal hold, archive are examples). BigID is unique among data security and compliance solutions that offers both data move and tombstoning.

Lastly, in order to support end-to-end data lifecycle management or retention, a technology will also need to reverse a move, ie a restore of the data or record to its original location (after a review or legal hold for instance). BigID again is unique in supporting this capability.

Step 6. Checks for Legal Hold and Referential Integrity Impact

Deleting data doesn't come without risk. One such risk is obviously deleting the wrong data. BigID addresses this with the industry's leading data identification / classification technology and the most advanced review and approval workflow in the category (plus the added ability to isolate data for a period before purging it).

However, there are other risks including accidental deletion of data already flagged for legal hold or equally bad the deletion of structured data needed to run an application. BigID also addresses these use cases.

Firstly BigID supports integration with an array of legal hold platforms to ensure data already flagged for a legal proceeding is not inadvertently deleted.

Secondly, BigID has developed proprietary and patented technology to detect potential rupture of referential integrity through the deletion of a database or table. The latter ensures structured data sources like a SQL database or table in Snowflake are not impacted by the deletion of dependent or connected data.

Step 7. Centralized and Decentralized Deletion or Archive

Executing a deletion on a data source is not hard. Doing it safely however, is.

We have already discussed the various safeguards a mature data retention product like BigID provide including:

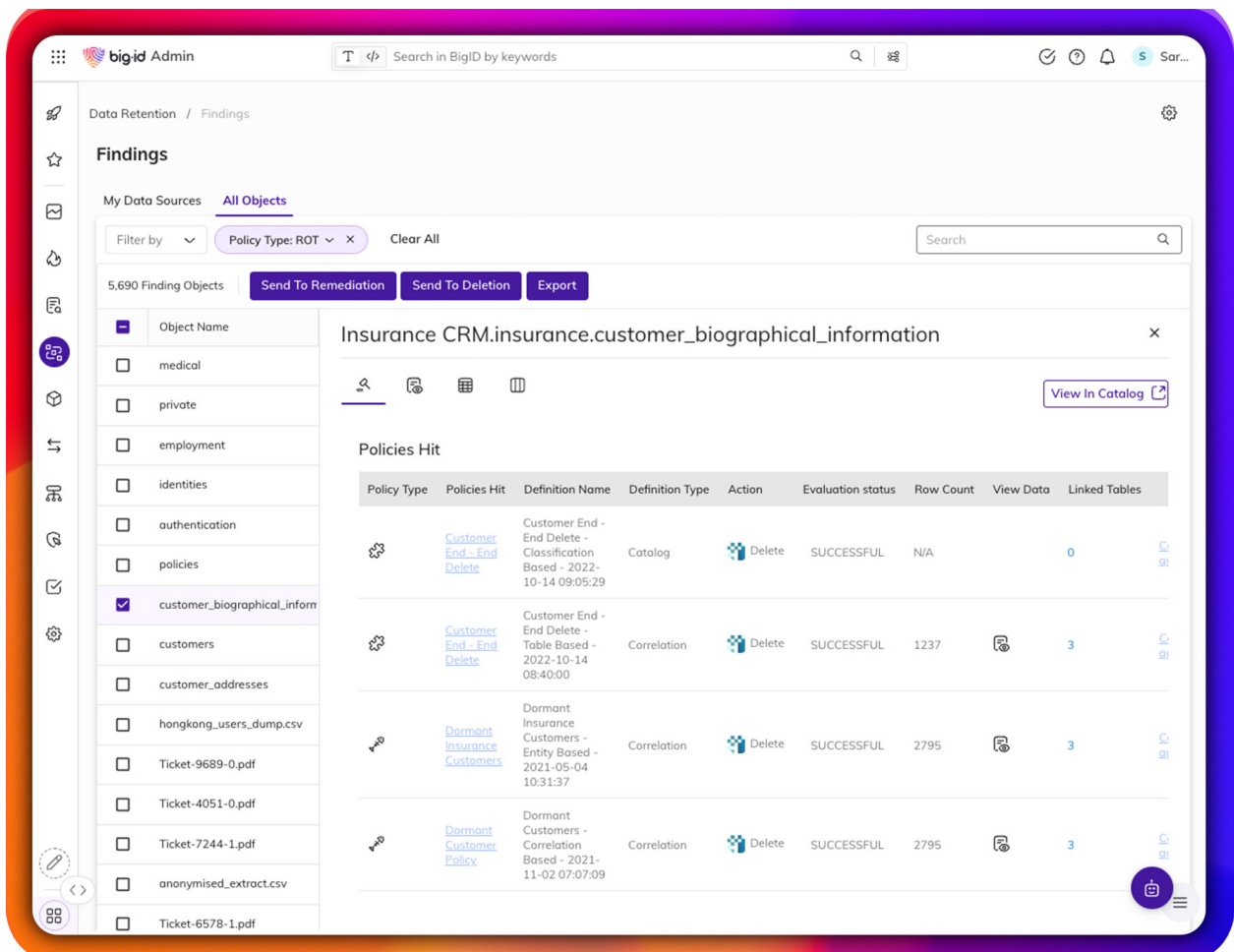
- Accurate classification and identification of target data
- Full content scanning without skipping content
- Flexible and business aware retention schedules and policies
- Proper review and approval workflows for data owners
- An ability to move, isolate, stage or cold store data before a final action
- Integrity checks for legal hold and referential integrity impact

Final consideration and check is the programmatic part of executing a deletion. No company wants a situation in which data is eliminated. Therefore, BigID has provided customers with three additional safeguards.

First, instead of outright deletion of data, BigID provides end users the option to put data into a cold storage archive as a first step.

Secondly, BigID provides an option to have a data owner provide a final review before executing a deletion.

Thirdly, for end users that want to centralize deletion (or archive) BigID also provides a dashboard for decision makers that presents all findings and impact analysis from a specific retention policy for final review and approval.



The screenshot shows the BigID Admin interface. The top navigation bar includes the BigID logo, a search bar, and user information. The main content area is titled 'Findings' and shows a list of data sources. A filter is applied for 'Policy Type: ROT'. The table below shows the results of the search, with columns for Object Name, Policy Type, Policies Hit, Definition Name, Definition Type, Action, Evaluation status, Row Count, View Data, and Linked Tables. The table contains four rows of data, all with a 'Delete' action and 'SUCCESSFUL' evaluation status.

Object Name	Policy Type	Policies Hit	Definition Name	Definition Type	Action	Evaluation status	Row Count	View Data	Linked Tables
medical	Customer End - End Delete	Customer End - End Delete - Classification Based - 2022-10-14 09:05:29	Catalog	Delete	SUCCESSFUL	N/A	0		
private	Customer End - End Delete	Customer End - End Delete - Table Based - 2022-10-14 08:40:00	Correlation	Delete	SUCCESSFUL	1237	3		
employment	Dormant Insurance Customers	Dormant Insurance Customers - Entity Based - 2021-05-04 10:31:37	Correlation	Delete	SUCCESSFUL	2795	3		
identities	Dormant Customer Policy	Dormant Customers - Correlation Based - 2021-11-02 07:07:09	Correlation	Delete	SUCCESSFUL	2795	3		

In Conclusion

Modern enterprises face many more demands on how they manage and hold data:

- Data is now spread across structured and unstructured data sources in the data center and cloud.
- Organizations are facing unprecedented levels of data regulations like GDPR, CCPA, HIPAA, ITAR, that mandate what, when and how data is erased (or conversely preserved).
- Companies face more and more litigation requiring more complex legal hold requirements for data.
- They face exploding data volumes and commensurate cloud costs to store the data.
- And lastly, cyber attacks targeting sensitive data continue unabated leading companies to reimagine strategies for reducing their data attack surface by reducing shadow and data debt.

But no matter the reason, or the framing of the program whether data lifecycle management, minimization, deletion, records management or responsible disposition - finding a technology to safely and scalably effect the program is critical.

For this reason BigID is the first and only data security, privacy or compliance tool to deliver a mature and turnkey technology for DLM. While other tools promise to find and delete stale or duplicate data, BigID is the only technology that can:

- Find any manner of Redundant, Obsolete, Trivial, and regulated data or record
- Scan data in full without skipping files or rewriting important metadata
- Deliver a flexible and comprehensive retention schedule platform
- Offer RBAC protected review and disposition workflow
- Move, archive and tombstone data for investigation, isolation or cold storage
- Check for legal hold and referential integrity checks before erasure
- Add final guardrails and approvals ahead of executing a deletion

All this technology is available to customers in a range of deployment options and in full compatibility with BigID's other compliance, privacy and security modules including DSPM, DSR, CMP, Compliance and Controls, DAG, DAM, Labeling, Risk Assessments, Data Mapping and more.

To schedule a demo or free trial contact BigID at info@bigid.com.

About BigID

BigID helps organizations connect the dots in data & AI: for security, privacy, compliance, and AI data management. BigID enables customers to find, understand, manage, protect, and take action on high risk & high value data, wherever it lives.

Customers use BigID to reduce their AI & data risk, automate security and privacy controls, achieve compliance, and understand their data throughout their entire data landscape: from the cloud, on-prem, and everywhere in between.

BigID has been recognized for innovation as a World Economic Forum Technology Pioneer; named to the Forbes Cloud 100; the Inc 5000 for 4 consecutive years; the Deloitte 500 for 4 consecutive years; Market Leader in Data Security Posture Management (DSPM); Leader in Privacy Management in the Forrester Wave; and an RSA Innovation Sandbox winner.

Connect the Dots in Data & AI

Data Security • Compliance • Privacy • AI Data Management

Reduce risk, accelerate time to insight, and get data visibility and control across all your data - everywhere.

“Tools like BigID are the future.

Organizations should be leveraging these tools to remove the manual processes from data discovery, provide better visibility, and help with prioritization of controls.



Ryan O'Leary
Future of Trust: Battling Data Discovery Confusion