



2025 WHITE PAPER

Best Practices to Find Stale, Duplicate, and ROT Data

A detailed overview of the challenges of finding and managing stale and ROT data. Learn best practices and how BigID can level up your data lifecycle management.



2025 White Paper

Best Practices to Find Stale, Duplicate, and ROT Data

2025 White Paper

Best Practices to Find Stale, Duplicate, and ROT Data

Today's organizations are grappling with the exponential growth of data across diverse environments. This surge has led to the proliferation of stale, redundant, obsolete, and trivial (ROT) data, which not only inflates storage costs but also poses significant risks to security, privacy, and regulatory compliance. Identifying, classifying, and effectively managing stale and ROT data is a critical challenge for businesses—one that can't be ignored.

BigID, the industry leading data security and privacy platform, provides a comprehensive solution for end-to-end data lifecycle management. By combining advanced AI and ML-based data discovery and classification with dedicated retention and minimization capabilities, BigID empowers organizations to efficiently identify, categorize, and remediate ROT data. This whitepaper explores best practices for finding and managing stale, duplicate, and ROT data, and how BigID can help organizations reduce costs, mitigate risks, and improve their overall data security posture.

The Challenge of Stale, Duplicate, and ROT Data

As data volumes grow, so does the complexity of managing it. Stale, duplicate, and ROT data can accumulate across on-premises, cloud, and hybrid environments, creating a range of challenges:

- **Increased Storage Costs:** Redundant and obsolete data consumes valuable storage resources, driving up infrastructure costs.
- **Security and Privacy Risks:** Stale data often contains sensitive information that, if left unmanaged, can become a target for breaches or misuse.
- **Regulatory Non-Compliance:** Retaining data beyond its required retention period can lead to violations of privacy regulations like GDPR, CCPA, and HIPAA, resulting in hefty fines and reputational damage.
- **Operational Inefficiency:** Manual efforts to identify and manage ROT data are time-consuming and error-prone, diverting resources from strategic initiatives.

To address these challenges, organizations need a proactive and automated approach to data lifecycle management.

Why Managing Stale and ROT Data Matters to Business Leaders

For senior executives, the stakes of managing stale and ROT data are clear:

Cost Optimization

Eliminating redundant and obsolete data reduces storage costs and optimizes resource allocation. By streamlining data management, organizations can allocate budgets more effectively to strategic initiatives.

Regulatory Compliance

Privacy regulations mandate that organizations retain data only for as long as necessary. Failure to comply can result in significant penalties. Automated data retention and deletion policies ensure adherence to legal requirements, minimizing the risk of non-compliance.

Enhanced Security Posture

Stale and ROT data often contain sensitive information that can be exploited in a breach. Proactively identifying and removing this data reduces the attack surface and strengthens overall security.

Improved Operational Efficiency

Automating data discovery, classification, and remediation reduces manual effort, enabling teams to focus on higher-value tasks. This leads to faster decision-making and improved productivity.

Trust and Reputation

Demonstrating a commitment to effective data management builds trust with

customers, partners, and stakeholders. It reinforces an organization's reputation as a responsible custodian of data.

According to the Veritas Global Databerg Report, up to 85% of an organization's data could be ROT.

Best Practices for Finding and Managing Stale, Duplicate, and ROT Data

1. Automated Data Discovery

The first step in managing stale and ROT data is gaining visibility into your data landscape. Organizations must comprehensively identify and catalog data across all environments, including on-premises, cloud, and hybrid systems.

BigID's AI-driven data discovery capabilities automatically scan and catalog data across diverse environments, ensuring no data is overlooked. Continuous monitoring ensures the data catalog remains accurate and up-to-date as data sources evolve.

2. Advanced Sensitivity Classification

Not all data is created equal. Classifying data based on its sensitivity, value, and regulatory requirements is essential for prioritizing retention and deletion efforts.

BigID leverages advanced AI and ML-based classification to accurately categorize data by content, context, and regulatory requirements. Tunable classification models adapt to evolving data patterns, ensuring ongoing accuracy and compliance.

3. Data Retention and Deletion Policies

Customizable and automated retention policies are critical for ensuring compliance with privacy regulations and minimizing storage costs.

BigID enables organizations to define granular retention and deletion policies aligned with legal and regulatory requirements. Automated workflows ensure data is securely deleted when no longer needed, while legal hold processes preserve data for investigative or compliance purposes.

4. Data Remediation

Once stale and ROT data is identified, organizations need a streamlined process for securely deleting it.

BigID's data remediation capabilities automate the deletion process, ensuring compliance with industry best practices and regulatory standards. Dynamic rulesets allow organizations to adapt retention policies based on data age, sensitivity, regulatory requirements, and other factors.

Take Control of Your Data Lifecycle with BigID

Organizations can no longer afford to overlook the risks and costs associated with stale, duplicate, and ROT data. Enterprises that adopt best practices and leverage BigID's industry leading platform for data privacy, security, compliance, and AI data management— can transform their data lifecycle management processes, ensuring security, compliance, and efficiency.

BigID is the first and only DSPM solution to deliver a fully integrated platform for data discovery, classification, and remediation. With intelligent detection, automated retention enforcement, and secure workflows, organizations can turn data minimization from a challenge into a competitive advantage.

BigID supports organizations at every step of the data lifecycle:

- **Discovery and Mapping:** Automatically locate and map sensitive, regulated, and high-value data across on-premise and cloud environments.
- **Automated Classification and Labeling:** Significantly reduce the time and resources spent on manual data management tasks by automatically identifying and classifying data accurately for better data-driven decision-making. Tag data by type, sensitivity, regulation, and purpose, enabling precise curation.
- **Enable Regulatory Compliance:** BigID's automated policies and workflows for data retention and deletion help you navigate the complex landscape of data privacy regulations such as GDPR, CCPA, and HIPAA. Ensuring that data is only retained for necessary periods and securely deleted when required, avoids costly penalties and reputational damage associated with non-compliance.
- **Reduce Risk:** Proactively identify and mitigate risks associated with stale and ROT data to strengthen your overall security posture. By minimizing the attack surface and protecting sensitive information with the proper security controls, mitigate the risk of data breaches, financial losses, and reputational harm.

About BigID

BigID enables security, compliance, privacy, & AI data management for all data, everywhere. BigID is enterprise-ready and built to scale: enabling a data-centric approach to comprehensive cloud data security & DSPM, accelerating compliance, automating privacy, and streamlining overnance. Customers deploy BigID to proactively discover, manage, protect, and get more value from their regulated, sensitive, and personal data across their data landscape.

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.

Find out more at <https://bigid.com>.

Know Your Data, Control Your Data.

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