



Effective Strategies for Cleaning Up Legacy Data and Eliminating ROT Data

Contents

Introduction	3
Understanding the Problem	3
----- Definition of Legacy and ROT Data	3
----- Challenges Associated with Legacy and ROT Data	3
----- The Consequences of Not Addressing Legacy and ROT Data	5
Identifying and Categorizing Legacy and ROT Data	6
----- Strategies for Conducting an Initial Data Assessment	6
----- Techniques for Categorizing Data	7
Developing a Data Cleanup Strategy	8
----- Setting Clear Objectives and Goals	8
----- Building a Cross-Functional Team	9
----- Establishing Data Retention and Disposition Policies	9
----- Implementing Data Governance Frameworks	10
Executing the Data Cleanup	11
----- Tools and Technologies for Data Cleanup	11
----- Step-by-Step Process for Data Cleanup	12
----- Addressing Challenges and Obstacles During Execution	14
Ensuring Ongoing Data Management and ROT Data Prevention	15
----- Establishing Ongoing Data Management Practices	15
----- Training and Awareness Programs	16
Measuring Success and ROI	17
----- Defining Key Performance Indicators (KPIs) for Data Cleanup Success	18
----- Calculating ROI of Data Cleanup Initiatives	19
Conclusion	20
----- Recap of the Importance of Addressing Legacy and ROT Data	20
----- Encouragement to Act and Implement Data Cleanup Strategies	20
----- Final Thoughts on the Long-Term Benefits of Effective Data Management	21

INTRODUCTION

Effective Strategies for Cleaning Up Legacy Data and Eliminating ROT Data

In today's digital age, data is the lifeblood of every organization. However, the accumulation of legacy data and ROT (Redundant, Obsolete, and Trivial) data presents significant challenges that can impede operational efficiency, elevate security risks, and inflate storage costs. This whitepaper aims to provide a comprehensive guide on effectively cleaning up legacy data and eliminating ROT data, drawing on the extensive experience of working with various organizations across different industries.

Understanding the Problem

Definition of Legacy and ROT Data

Legacy data refers to outdated information that is still retained by an organization. This data is often stored in obsolete formats or systems and may no longer be in use. ROT encompasses data that is redundant, obsolete, or trivial and lacks any business value. Recognizing the types and characteristics of legacy data and ROT data is the first step toward effective management.

Challenges Associated with Legacy and ROT Data



Data Proliferation and Lack of Visibility

Organizations frequently struggle with data proliferation, where the volume of data grows exponentially without adequate visibility. This unchecked growth makes it difficult to manage and utilize data effectively. Over time, as data accumulates from various sources such as emails, documents and databases, it becomes increasingly challenging to keep track of what exists, where it resides, and its relevance. This lack of visibility can lead to significant inefficiencies and missed opportunities for data-driven decision-making.



Inefficient Data Management Practices

Many organizations lack standardized data management practices, leading to inconsistencies and inefficiencies. Without a structured approach, data cleanup efforts become sporadic and ineffective. Inconsistent data management practices can result in duplicated efforts, incomplete data, and a fragmented understanding of the organization's data landscape. The absence of a centralized data management policy often leads to each department handling data differently, further complicating cleanup efforts and increasing the risk of non-compliance with regulatory standards.



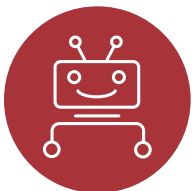
Resource Constraints

Limited resources, including time, budget, and personnel often hinder comprehensive data cleanup initiatives. Balancing these constraints while maintaining daily operations is a significant challenge. Many organizations find themselves stretched thin, with IT departments overburdened and lacking the bandwidth to tackle data cleanup projects comprehensively. Financial constraints can also limit the ability to invest in necessary tools and technologies, making it difficult to achieve the desired outcomes.



Fear of Losing Valuable Information

Organizations often fear that deleting or archiving data might result in the loss of valuable information. This fear leads to the retention of unnecessary data, exacerbating the problem of data overload. The hesitation to part with data, even when it no longer serves a purpose, stems from concerns about potential future needs, regulatory requirements, or historical value. This cautious approach can lead to a "save everything" mentality, where data accumulates indiscriminately, increasing storage costs and complicating data management.



Impact on AI Tools

The proliferation of legacy and ROT data can significantly undermine the effectiveness of advanced AI tools that integrate with an organization's data estate to help users find information and answer questions. When these AI tools ingest large amounts of outdated or irrelevant data, the quality of their assistance diminishes. Instead of providing accurate, relevant answers, they may surface redundant or outdated information, causing confusion and inefficiencies.

Additionally, the presence of ROT data increases the risk of sensitive information being inadvertently shared with unauthorized users. Without proper data governance and cleanup, AI tools can potentially expose legacy data that should no longer be accessible or that falls under stricter access controls. This potential outcome not only compromises data security but also violates compliance regulations and internal policies.

The Consequences of Not Addressing Legacy and ROT Data



Increased Storage Costs

Unmanaged legacy data and ROT data consume valuable storage resources, leading to increased costs. As data accumulates, so do the expenses associated with storing and managing it. The costs extend beyond just physical storage; they include maintenance, associated hours spent by IT personnel, the cost of backups, and the energy consumption of data centers. Additionally, the more data an organization retains, the higher the potential costs for data migration during system upgrades or transitions to new platforms.



Compliance Issues

Compliance with data protection regulations is critical. Unmanaged legacy and ROT data can lead to non-compliance, resulting in legal penalties and damage to the organization's reputation. Regulations such as GDPR, CCPA, and HIPAA require organizations to manage and protect personal data responsibly. Failure to do so can result in hefty fines, legal actions, and a loss of trust among customers and stakeholders. Proper data management practices are essential to ensure compliance and avoid these risks.



Reduced Operational Efficiency

The presence of unnecessary data hampers operational efficiency. Employees spend more time searching for relevant information, and the overall decision-making process is slowed down due to data clutter. When data is disorganized and difficult to access, it impacts productivity and can lead to missed opportunities. Streamlined data management practices are essential for maintaining efficient workflows, enabling quick access to critical information, and supporting informed decision-making across the organization.

Higher Security Risks

Legacy and ROT data pose significant security risks. Outdated or redundant data may contain sensitive information, and the sheer volume of data makes it challenging to implement effective security measures. Legacy systems often lack the robust security features of modern technologies, making them attractive targets for cyberattacks. The presence of unnecessary data increases the attack surface, providing more opportunities for malicious actors to exploit vulnerabilities.

After a breach occurs, it's also essential for the organization to understand what data was exposed, with strict reporting deadlines for certain types of content. That means that every record will need to be reviewed to understand the scope of the event. If it seems expensive and burdensome to review legacy content and delete it during normal business operation, imagine having to do it in the context of a recent security incident with a 30-day deadline imposed by a regulator. A similar argument applies for reviewing content in response to a litigation or other regulatory investigation.

Identifying and Categorizing Legacy and ROT Data

Strategies for Conducting an Initial Data Assessment



Stakeholder Interviews and Surveys

Engaging with key stakeholders through interviews and surveys helps identify data management pain points and priorities. This process ensures that the data cleanup strategy aligns with business objectives. Stakeholders from different departments can provide valuable insights into how data is used, where pain points exist, and what the expectations are for data management. Understanding the diverse needs and perspectives of various teams helps create a more comprehensive and effective data cleanup plan.



Data Discovery Tools and Technologies

Leveraging data discovery tools and technologies provides insights into the organization's data landscape. These tools help identify and classify data, making it easier to manage. Data discovery tools use advanced algorithms to scan and analyze data across the organization, identifying patterns, relationships, and anomalies. By providing a clear picture of the data inventory, these tools facilitate informed decision-making and targeted cleanup efforts.



Establishing Criteria for ROT Data Identification

Defining clear criteria for identifying ROT data is essential. Criteria for ROT data identification can include factors such as data age (e.g., files not accessed in the last two years), usage frequency (e.g., documents never opened), and redundancy (e.g., duplicate files). Establishing these criteria helps prioritize cleanup efforts and ensures that valuable data is retained while unnecessary data is removed.

However, these criteria alone likely will not capture all the inputs needed for approving a deletion. For instance, just because a file is old and unused does not mean its retention period is over or that it is not relevant to the business. ROT criteria act as scalable filters for locating prime candidates for deletion. This is an important first step when facing large amounts of unstructured legacy content that are too vast to review manually from the start.

Techniques for Categorizing Data



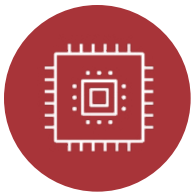
Metadata Analysis

Analyzing metadata is one approach to an initial data assessment. Metadata analysis involves examining the descriptive information associated with data files, such as timestamps, file types, and access history, as opposed to its content. By categorizing data based on metadata attributes, organizations can identify patterns, assess data relevance, and streamline data management practices. Due to the complexity of retention requirements, a single metadata field is hardly ever enough to make a final decision on a document. However, if multiple conditions are all met together (e.g., age, last accessed, last modified, file type, number of locations, size, system type, etc.), an organization can build confidence in its identification of ROT data.



Content Analysis

Content analysis involves examining the actual content of data to determine its relevance and value. This technique is the most resource intensive, but it is the most accurate in determining whether a file truly is ROT. Iteratively evaluating rules based on keywords or pattern matching can be accurate enough for certain types of content, while there is always a subset that may require manual review. Due to information volumes, it's never going to be practical to review the content of all files manually. If you do not have access to some sort of content analytic software, don't underestimate the power of statistical sampling. Taking a true random sample of a particular file population, enforced by confidence metrics, can often give workable intelligence on how to proceed with it.

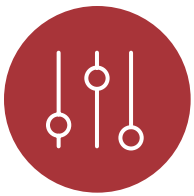


Automated Classification Tools

Using automated classification tools streamlines the process of categorizing data. These tools use algorithms including machine learning and artificial intelligence to classify data based on predefined criteria, enhancing efficiency and accuracy. These tools can significantly reduce the time and effort required for data classification, ensuring consistent and accurate results. Not everyone will be able to justify the purchase of these tools right at the outset. If that is the case, when using a manual approach or 'free' tools, make sure you track specific risks that are being uncovered and the amount of time it is taking you to get through the content. In other words, work as if you will have to do the entire classification process with the tools you have, but manage the project and track metrics as if you are building the business case for purchasing an automated classification tool.

Developing a Data Cleanup Strategy

Setting Clear Objectives and Goals



Aligning Data Cleanup Initiatives with Business Objectives

Data cleanup initiatives should align with the organization's broader business objectives. For example, if an organization's objective is to enhance customer satisfaction, data cleanup efforts can focus on improving the accuracy and accessibility of customer data. Aligning data cleanup initiatives with business objectives ensures that the cleanup efforts are meaningful and support the organization's overall mission.



Prioritizing Data Cleanup Efforts Based on Risk and Impact

Prioritizing data cleanup efforts based on the risk and impact of unmanaged data helps focus resources on the most critical areas. This approach maximizes the benefits of the cleanup initiative. High-risk data, such as sensitive customer information or data critical to business operations should be prioritized for cleanup to mitigate potential risks.

Building a Cross-Functional Team

Involving Key Stakeholders (IT, Legal, Compliance, Business Units)

A successful data cleanup strategy requires the involvement of key stakeholders from various departments. Engaging stakeholders from IT, legal, compliance, and business units helps ensure that diverse perspectives and expertise are considered. This collaborative approach facilitates a comprehensive understanding of data management challenges and enables the development of effective cleanup strategies. It also makes sure that everyone feels like they have a voice in the process. If you exclude a stakeholder for too long, not only could you be missing out on key insights from their perspective, but they may also end up being a blocker to the strategy you have developed, causing friction and wasted resources.

Assigning Roles and Responsibilities

Clearly defining roles and responsibilities within the data cleanup team ensures accountability and facilitates effective execution. Each team member should understand their specific tasks and contributions. Assigning roles such as data stewards, project managers, and IT specialists helps ensure that all aspects of the cleanup initiative are covered. Clear communication of roles and responsibilities fosters collaboration, accountability, and a shared sense of ownership. On smaller teams, assignments might be informal and agile, letting each of the roles grow over time and adapt to the obstacles and requirements you discover. However, on larger teams, you may want to use a more formal methodology such as a RACI matrix (Responsible, Accountable, Consulted, Informed).

Establishing Data Retention and Disposition Policies



Developing a Comprehensive Data Retention Schedule

A comprehensive data retention schedule outlines how long different types of data should be retained and should ensure compliance with legal, regulatory, and business requirements. For example, financial records may need to be retained for a certain number of years to comply with regulatory standards. A well-defined retention schedule helps manage data lifecycle and ensures that data is retained only as long as is necessary. A schedule's goal is to provide a one-stop shop of retention and disposition requirements that serve all organizational goals.

However, one of the biggest mistakes is to approach schedule creation purely as an exercise in documenting requirements. You also need to consider who will need to read the schedule, how often, how they can access it, how easily you can change it, how easy it is to understand, etc. Applying Design Thinking when creating the schedule is enormously important if you want it to be actually adopted and used.



Defining Criteria for Data Deletion and Archiving

Defining clear criteria for data deletion and archiving ensures that only non-essential data is targeted for removal, preserving valuable information. Criteria for data deletion and archiving should be based on factors such as data age, usage frequency, and business value. For example, files that have not been accessed in the last three years and do not have regulatory or historical value may be candidates for deletion.

Establishing these criteria helps streamline data management practices and ensures that valuable data is retained. Although a retention schedule articulates the policy for retention, the deletion criteria act as a procedure, giving people (or automated rules) a workflow on how they can check any given file against the schedule and come up with a standardized, repeatable outcome on whether or not a file can be deleted. In other words, if a group of people all independently apply the criteria to the same file, they should all arrive at the same disposition decision.

Implementing Data Governance Frameworks

Creating Policies and Procedures for Ongoing Data Management

Establishing robust policies and procedures for ongoing data management is essential for maintaining data quality and preventing the recurrence of ROT data. These policies should outline best practices for data creation, storage, access, and disposal. Procedures should include guidelines for data classification, retention, and disposition. By creating comprehensive policies and procedures, organizations can ensure consistent and effective data management practices. The details of the procedures will depend heavily on the system where the data is created and stored. Leaning on IT or even business groups is essential to establishing procedures make sense for the types of systems with which people are interacting.



Ensuring Accountability and Compliance

Ensuring accountability and compliance with data governance frameworks requires regular audits and monitoring. This practice helps assess compliance with data management policies and identify areas for improvement. Monitoring tools can track data usage, access, and changes, ensuring that data management practices align with organizational standards and regulatory requirements. Accountability measures, such as assigning data stewards, help ensure that data management responsibilities are upheld.



Audits can often be seen as cumbersome and low value when it comes to data governance due to the sheer volume of data that is created. It can help to be tactical by aligning with stakeholders to embed auditing activities in other established business processes – e.g., partner with Compliance or InfoSec on information-based audits that are already taking place. It's also important to help the average user by making it easy to report non-compliance when they come across it.

For instance, if an employee is looking for a specific file on a share and they spot a problem, they are not going to report it if it requires completing a long form or finding out to whom they are expected to report the issue. However, if there is a dedicated slack channel, or even a reward system, it may be front of mind enough for them to stop the task they are doing and report the non-compliance. In this way, you can use normal data queries as mini audits for the systems used most.

Executing the Data Cleanup

Tools and Technologies for Data Cleanup

Data Discovery and Classification Tools

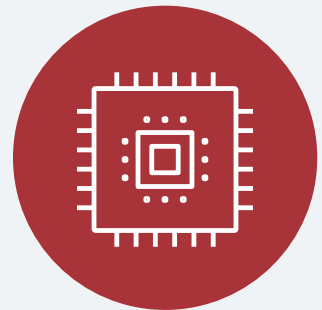
Data discovery and classification tools provide insights into the organization's data landscape. These tools help identify and classify data, making it easier to manage. By scanning and analyzing data across the organization, these tools provide a comprehensive inventory of data assets. They can identify redundant, obsolete, and trivial data, enabling targeted cleanup efforts. Advanced data discovery tools use machine learning and AI to classify data accurately and efficiently. These tools are rarely a silver bullet that organizes your entire data estate right from the start, but they often provide a powerful overview that can jumpstart your project.

Data Deduplication and Archiving Solutions

Data deduplication solutions help eliminate duplicate data, reducing storage requirements and improving data quality. These tools identify and remove redundant copies of data, consolidating storage and enhancing data accessibility. Archiving solutions provide secure and accessible storage for data that needs to be retained but is not actively used. Together, these solutions help streamline data management and optimize storage resources. However, the devil is in the details for these activities. It should be clear what constitutes a duplicate, whether it is a binary hash duplicate, or a near duplicate, or a versioning system that favors the most up-to-date version of a file and considers older versions to be duplicative. Making sure all stakeholders are clear on what is considered a duplicate will be important when evaluating tools that claim to deduplicate.

Automation and AI-Driven Tools

Leveraging automation and AI-driven tools enhances the efficiency and accuracy of data cleanup initiatives. Automation tools can help automate repetitive tasks such as data classification, deduplication, and archiving, reducing the manual effort required. AI-driven tools analyze data patterns and usage, providing actionable insights for data cleanup and management. One essential consideration is your organization's budget for bringing sophisticated pre-trained models to bear on its data estate, versus implementing a technology that requires additional training after installation. Although AI tools can make automated decisions, ensuring those decisions are accurate enough to perform well on your company's specific documents is crucial. Focus on the highest risk and highest value information types first and seek out vendors who specialize in that type of content for your specific industry.



Step-by-Step Process for Data Cleanup



Data Assessment and Categorization

The first phase of the data cleanup process is to prioritize the data you want to clean up. This step involves identifying which systems and data populations have the best chance of making progress towards the project goals. If the top priority is securing sensitive information, consider where it is most likely to be. If the goal is deleting a large quantity of data, look for locations with large video files where a few decisions could make a significant impact, for example.

You won't have all the answers at the outset, which is why an iterative approach is essential. Formulate a hypothesis based on your initial work about where you can make quick progress. However, don't allocate all your resources to it immediately. Take a deeper look to confirm your hypothesis through methods like statistical sampling or metadata analysis. If you gain confidence that the path will pay off, double down on it. If not, shift strategies and investigate the next highest priority area.

It's crucial to show a return on investment as soon as possible to secure continued funding and support for the long run. This iterative and agile approach helps ensure that you don't waste too much time on low-value areas, and you maximize your ability to quickly demonstrate value.



Identifying ROT Data Candidates for Deletion and Archiving

With proper preparation, this stage should be straightforward. The key is to establish clear, pre-approved criteria for data deletion before you begin. All stakeholders need to sign off on these criteria to ensure a streamlined process.

However, without proper preparation, this stage will stall very quickly. The biggest risk is that you rely heavily on manual review. Instead, try to move as close to criteria that everyone agrees on and that serves as a pre-approved classification/deletion trigger. Avoid queuing up data that "seems to be" redundant, obsolete, or trivial and then sending it to various stakeholders for manual review. Stakeholders are often busy with their primary responsibilities, making it difficult to prioritize disposition reviews.

When you encounter data that requires input from stakeholders, don't just send them a list of files. Instead, use these files as examples to further develop your pre-approved criteria. Discuss how stakeholders determined whether to keep or delete a file. If stakeholders want to keep a file that contradicts the criteria, be willing to adjust the criteria after careful consideration. The goal is to establish criteria with which everyone feels confident and comfortable. That is the only way they will be okay with pre-approved classifications and deletions. It's also important that the criteria are specific and objective; if three managers from three departments all apply the criteria to the same file, they should all arrive at the same ROT assessment.



Implementing Data Deletion or Archiving Actions

The final step is executing the data deletion or archiving actions. Documentation is critical at this stage for several reasons, including legal compliance. If you need to prove to a regulator or court that you destroyed certain data, you'll need thorough documentation.

Determine the appropriate level of documentation, whether it includes filenames and paths or just summary information like the number of files deleted and their system location. Always record the criteria or justification for deletion.

Tracking metrics & documenting deletions serve multiple purposes beyond compliance:



Project Management:

Monitor progress against goals, such as the number of candidates for deletion identified and how many have been deleted, allowing you to react early and pivot strategies if needed.



ROI Tracking and Communication:

Clearly communicate to stakeholders the business outcomes achieved through data cleanup.

Beyond tracking and documentation, consider additional steps:



Make Archiving Harder Than Deleting:

If users are prone to archiving rather than deleting, implement an approval process or require them to fill out a justification form. The extra work can be a deterrent to abusing an archive versus just deleting the content.



Quarantine / Recycling Bin Step:

Move identified ROT data to a quarantine location with a timer for deletion (e.g., delete after one year unless requested otherwise). This alleviates concerns about immediate deletion and builds confidence in the classification process. In practice, data rarely gets recalled from quarantine.

Finally, ensure your archive solution meets all your requirements.

Providers with industry experience can offer tailored processes and are familiar with relevant regulations, which is crucial for compliance and effectiveness.

Addressing Challenges and Obstacles During Execution



Handling Resistance to Change

Resistance to change is a common obstacle in data cleanup initiatives. Addressing this challenge requires effective communication, stakeholder engagement, and change management strategies. Consider offering rewards for active participants as much as outlining consequences for non-compliance. Communicating the benefits of data cleanup (i.e., what's in it for me?) and addressing personal concerns helps gain stakeholder buy-in. Involving stakeholders in the planning and execution phases fosters a sense of ownership and reduces resistance. Change management strategies, such as training and support, help smooth the transition and ensure successful implementation.



Misalignment with Current Company Objectives

One of the critical challenges in executing a data cleanup project is ensuring that it aligns with the company's current objectives. Misalignment can result in a lack of necessary resources and support, ultimately leading to incomplete or ineffective cleanup efforts. For a data cleanup project to be successful, it must be integrated with the organization's strategic goals and demonstrate how it contributes to overall success.

It's essential to clearly communicate how the data cleanup project supports the company's strategic objectives and to secure executive sponsorship. This alignment ensures that the project is seen as a priority and receives the backing it needs from all levels of the organization. Additionally, consider that there are multiple ways to approach the cleanup process. Evaluate various options to determine which has the best potential for aligning with the corporate strategy and achieving quick wins. By doing so, you can demonstrate value early on, build momentum, and maintain ongoing support for the project.



Managing Resource Constraints

Managing resource constraints requires careful planning and prioritization. Allocating resources effectively and leveraging automation tools can help overcome these challenges. Prioritizing data cleanup efforts based on risk and impact ensures that resources are focused on the most critical areas. Leveraging automation tools reduces the manual effort required and enhances efficiency. Effective resource management ensures that data cleanup initiatives are executed successfully within the constraints of time, budget, and personnel.

Ensuring Ongoing Data Management and ROT Data Prevention

Establishing Ongoing Data Management Practices



Regular Data Audits and Assessments

Regular data audits and assessments are essential for maintaining data quality and preventing the recurrence of ROT data. These practices help identify and address data management issues proactively. Data audits involve systematically reviewing data assets to ensure compliance with policies and identify any anomalies. Regular assessments evaluate data usage, relevance, and quality, providing insights for continuous improvement. Implementing these practices ensures that data management remains effective and aligned with organizational goals.



Continuous Monitoring and Reporting

Continuous monitoring and reporting provide visibility into data management practices and help identify areas for improvement. These practices ensure that data remains accurate, secure, and compliant. Monitoring tools track data usage, access, and changes, providing real-time insights into data management activities. Regular reporting provides stakeholders with updates on data quality, compliance, and cleanup progress. Continuous monitoring and reporting help maintain accountability and support informed decision-making.



Implementing Data Lifecycle Management

Data lifecycle management involves managing data from creation to disposal. Implementing this approach ensures that data is managed effectively throughout its lifecycle. Data lifecycle management practices include data classification, retention, archiving, and disposal. These practices ensure that data is handled according to its value and relevance, reducing the risk of ROT data and enhancing data quality.

Implementing data lifecycle management helps maintain an organized and efficient data environment. The best part is, you will have already gained a lot of insights and intelligence on how data is used at your organization throughout the cleanup process, so you can apply those insights for establishing lifecycle management policies and procedures that take that into account.

Training and Awareness Programs



Educating Employees on Data Management Best Practices

Educating employees on data management best practices is crucial for maintaining data quality and preventing ROT. Training programs should cover topics such as data classification, retention, and security. Providing employees with the knowledge and skills to manage data effectively ensures that data management practices are consistently applied across the organization. Regular training sessions, workshops, and resources help reinforce best practices and keep employees informed of any changes or updates.

Promoting a Culture of Data Stewardship

Promoting a culture of data stewardship encourages employees to take ownership of data management practices. This approach fosters accountability and supports long-term data management goals. Encouraging employees to act as data stewards helps ensure that data is handled responsibly and consistently. Recognizing and rewarding good data management practices reinforces the importance of data stewardship and motivates employees to prioritize data quality and compliance.



Organizations must carefully weigh the decision between utilizing third-party training solutions and developing in-house programs for data management awareness. External vendors and software platforms typically offer polished, ready-to-deploy content and automated tracking features that can significantly reduce implementation time and administrative burden. These solutions often bring industry best practices and regularly updated content that reflects emerging trends in data management. However, they may lack organization-specific context and can be costly to maintain long-term.

In contrast, in-house training programs offer the advantage of being fully customizable to an organization's unique data landscape, internal policies, and corporate culture. They can incorporate relevant examples from the organization's actual data environment and be modified quickly as needs evolve. The main drawbacks of the in-house approach include the substantial initial time investment required for content development, the need for internal expertise, and the ongoing resource commitment to maintain and update materials. Many organizations find success with a hybrid approach, using third-party solutions as a foundation while supplementing with customized, organization-specific modules.

Measuring Success and ROI

Defining Key Performance Indicators (KPIs) for Data Cleanup Success



Storage Cost Reduction

Storage cost reduction is a critical KPI for data cleanup success. Measuring the reduction in storage costs helps demonstrate the financial benefits of data cleanup initiatives. Calculating storage cost savings involves comparing storage expenses before and after the cleanup. Reduced storage costs result from eliminating redundant and unnecessary data, optimizing storage resources, and improving data management practices. Remember to factor in not only the direct costs of the storage itself, but also the time and effort of IT personnel to manage and maintain that storage.



Compliance Metrics

Compliance metrics measure the effectiveness of data cleanup initiatives in achieving regulatory compliance. These metrics include the reduction in compliance risks and the number of compliance issues resolved. Tracking compliance metrics involves monitoring adherence to data protection regulations, identifying and addressing non-compliance issues, and assessing the impact of data cleanup on regulatory compliance. Effective data cleanup enhances compliance, reduces legal risks, and supports organizational integrity. Additionally, integrating data cleanup with cybersecurity efforts can further protect sensitive information by ensuring it is securely deleted and cannot be exfiltrated by attackers.



Operational Efficiency Improvements

Operational efficiency improvements measure the impact of data cleanup on business processes. These improvements include reduced time spent searching for information, enhanced decision-making, and streamlined workflows. Assessing operational efficiency involves tracking key metrics such as time savings, productivity gains, and process optimization. Data cleanup initiatives that improve data accessibility and quality contribute to enhanced operational efficiency and support business agility.

A subset in this area is the operational efficiency of non-standard business processes. For example, post-breach data review, eDiscovery review, regulatory requests, or data subject access requests are all more streamlined, less burdensome/costly, and less risky after cleaning up ROT data.



Enhanced Data Quality for AI Tools

One of the critical KPIs for data cleanup success is the improvement in data quality, particularly as it pertains to AI tools. By eliminating legacy and ROT data, organizations ensure that AI tools have access to accurate, up-to-date information, enhancing their ability to provide relevant and reliable assistance. This improved data quality directly translates into better decision-making and more effective use of AI capabilities, thereby increasing the overall value derived from these tools.

Calculating ROI of Data Cleanup Initiatives



Cost Savings from Reduced Storage and Maintenance

ROI comes from the cost savings achieved through reduced storage and maintenance expenses. These savings result from eliminating redundant data, optimizing storage resources, and reducing maintenance efforts. Remember that even data that wasn't deleted but is moved to a lower tiered storage (like archive) still has a net impact on storage costs. The ROI calculation involves comparing the costs of data cleanup initiatives with the resulting savings. Positive ROI demonstrates the financial benefits of data cleanup and supports continued investment in data management practices.

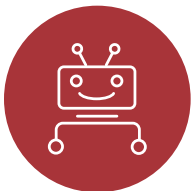
It may be worth trying to figure out a working per-TB cost ahead of time (i.e., what is your total cost of ownership for 1 TB of unstructured data storage), so you can calculate where your break-even point is (i.e., how many TBs do you need to delete to justify the cost of the initiative). Keep in mind the total cost of ownership will be different for different types of systems, such as 1 TB of on-premises file share storage vs. 1 TB of SharePoint 365 storage. If you know these differences in advance, you can use the information to prioritize a system where you will get more 'bang for your buck' in terms of deleting content.



Risk Mitigation Benefits

Risk mitigation benefits measure the reduction in security and compliance risks achieved through data cleanup. These benefits include reduced vulnerability to data breaches, enhanced data protection, and improved regulatory compliance. Assessing risk mitigation benefits involves evaluating the impact of data cleanup on security posture, compliance status, and overall risk exposure.

These benefits are notoriously hard to calculate. A regulatory audit that results in a monetary fine may never actually occur. If the fine were to occur, it's not always clear what the amount would be. For instance, should you assume the worst possible fine, or is that not credible in the eyes of executives? What seems to work best is to use the closest industry benchmarks that you can. For peers in your industry that were fined for similar offences, what was the cost and why? Is it a per-record violation fine, or was the fine based on some other metric? There are several entities that try to quantify the per GB cost of eDiscovery, as well as the per record cost of a security breach on an annual basis. Use these benchmarks to try to quantify the potential savings, but be transparent that these are risk avoidance ROIs and not hard savings.



Maximizing the Value of AI Investments

Cleaning up legacy data and eliminating ROT data not only leads to direct cost savings and risk reduction, but also maximizes the return on investment in AI tools. By ensuring that these tools operate with high-quality data inputs, organizations can fully leverage their potential to drive efficiency, improve decision-making, and provide valuable insights. This strategic advantage enhances the overall ROI of both the data cleanup initiatives and the AI investments, demonstrating the compounded benefits of maintaining a clean and well-governed data estate..



Productivity Gains

Productivity gains measure the improvements in employee productivity achieved through data cleanup. These gains result from reduced time spent searching for information, improved data accessibility, and streamlined workflows. Assessing productivity gains involves tracking metrics such as time savings, efficiency improvements, and employee satisfaction. Data cleanup initiatives that enhance productivity contribute to organizational success and support business objectives. In general, this must be measured in hours of work saved on a weekly (or some other) basis.

If you have access to general salary information, you can attempt to assign a dollar amount to such time saving. However, it is hard to be exact when it comes to the number of hours saved without conducting some kind of internal test. If you don't have the resources to do this (e.g., ask a group of users to find a set of records before and after the cleanup), you can also research benchmarking done for average number of hours spent by workers in your industry finding information, as many industry groups try to quantify this time on an annual basis as well.



Presenting Success Stories and Metrics to Stakeholders

Presenting success stories and metrics to stakeholders demonstrates the value of data cleanup initiatives. These presentations should highlight key achievements, such as cost savings, compliance improvements, and productivity gains. Success stories provide concrete examples of the benefits achieved and help build support for ongoing data management efforts. Sharing metrics and success stories with stakeholders fosters transparency, accountability, and continued investment in data cleanup initiatives.

The key is to stay focused on the objectives that were originally communicated and how well you reached them. Visualizations help with communicating data quickly, but even more important is a narrative, or story. If you can, try to hit these broad stages:

1. Set the Stage

The stakeholder may have forgotten about the importance of the project in the first place. Provide a quick reminder of the objective of the project and how it relates to the company strategy.

2. Rising Action

Give details on some of the compliance issues you uncovered, risky files, or more concrete information on the extent of the problem. Don't go into exhaustive detail, just something that will help raise the stakes.

3. Climax

Jump into the ROI metrics that act as the key results for your objective. In other words, what was actually accomplished, and how much good it does for the overarching company goals (or the goals/strategy/pain points of the specific stakeholder to whom you are presenting).

4. Denouement

Talk briefly about the 'new normal' by outlining some next steps you have in mind that will make the changes and benefits last.

CONCLUSION

Recap of the Importance of Addressing Legacy and ROT Data

Addressing legacy and ROT data is critical for enhancing operational efficiency, reducing costs, and mitigating risks. Effective data management practices ensure that valuable information is retained, and unnecessary data is eliminated. The consequences of unmanaged legacy and ROT data, including increased storage costs, security risks, and compliance issues, underscore the importance of proactive data management. By implementing effective data cleanup strategies, organizations can achieve significant benefits and support their long-term success.

Implement Data Cleanup Strategies Now

Developing a comprehensive data cleanup strategy, leveraging advanced tools and technologies, and fostering a culture of data stewardship are essential steps in achieving effective data management. But even starting small, with a single business group or system is better than prolonged planning.

Proactive data cleanup efforts enhance data quality, support compliance, and improve operational efficiency. Acting today ensures that organizations are well-positioned to navigate the evolving data landscape and achieve their business objectives.



The Long-Term Benefits of Effective Data Management

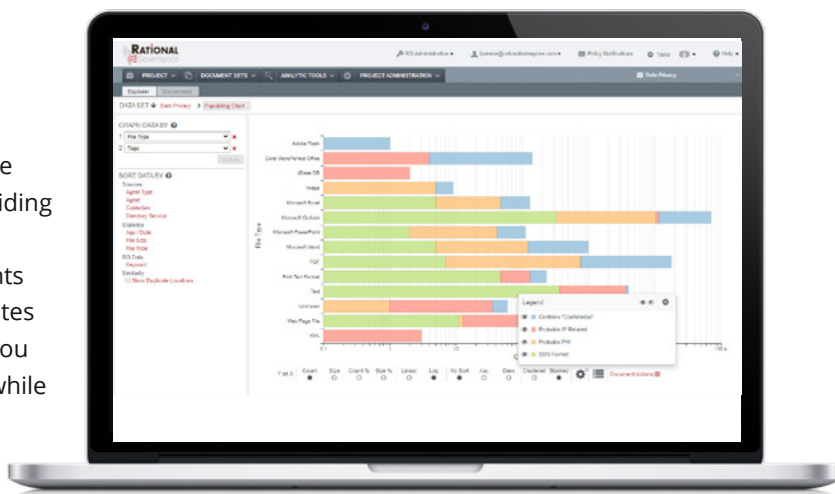


Effective data management provides long-term benefits, including improved decision-making, enhanced compliance, and optimized operational efficiency. By maintaining data quality and preventing the recurrence of ROT data, organizations can turn their data into a competitive advantage.

The ongoing commitment to data management practices ensures that data remains a valuable asset, supporting strategic goals and driving business growth. Embracing a proactive approach to data management enables organizations to harness the full potential of their data and thrive in the digital age.



Rational Governance empowers organizations to tackle their legacy and ROT data challenges head-on by providing comprehensive visibility and control over enterprise data from a central location. Through lightweight agents connected to your unstructured data sources, RG creates a complete index of content and metadata, enabling you to efficiently identify, analyze, and manage ROT data while implementing sustainable information governance policies that prevent future data accumulation.



Contact Us Today

Ready to transform your data landscape and unlock value while reducing costs and risk? Connect with our information governance experts today to take your legacy data cleanup efforts to the next level.

 [EMAIL US](#)

 [BOOK A DEMO](#)