

Feature

Alamy



DATA ANALYTICS AND DATA MANAGEMENT

## Data Spring Cleaning: How To Deal With ROT Data All Year Round

Clean, high-quality data is critical for business operations, but many companies have a data mess behind the scenes. Here's how to clean and prevent unwanted data.

Karen D. Schwartz | May 05, 2023



forgotten about given enough time. Data can become such a mess that there is even an industry acronym for it: ROT (Redundant, Obsolete, and Trivial). Data becomes redundant when it has duplicates stored across different systems or locations; obsolete when an organization neither needs nor is [required to retain the data](#) anymore; and trivial when the data is simply extraneous and without value.

## Why It's So Hard to Keep Data Orderly

Related: [10 Tips for Developing a Data Governance Strategy](#)

Clean, [high-quality data](#) serves as the bedrock for successful business operations and management. Obtaining high-quality data is a big undertaking and a critical first step – and only the first step. Too often, businesses fail to maintain good data practices. Poor data practices, such as failing to regularly organize and cull data, will result in a mess of ROT data over time.

It's gotten more complicated as businesses adopt more cloud-based resources, creating repositories of data in an exploding array of on-premises and cloud locations. "The ecosystem of where data is being created, how it's being stored, and how it's being managed has exploded by 1000%," said Juan Tello, U.S. chief data officer at Deloitte Consulting. "[Data] is now on-premises and in the cloud, with lots of fit-for-purpose solutions instead of a more monolithic approach."

Another reason for data disarray is the continued use of legacy systems, which often can't enforce modern data quality standards, Tello explained.

## The Costs of ROT Data

Having data, including ROT data, in so many different places not only results in chaos but can devour your budget.

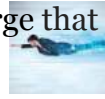
"If you're storing data on-premises and in the cloud, you're paying not only for storage in more than one place but making sure that data is backed up and replicated," said Rags Srinivasan, chief sustainability officer at Veritas Technologies.

According to a recent Virtana survey of IT leaders, 94% of IT leaders said [cloud storage costs](#) are rising. Fifty-four percent said storage spending is growing faster than overall cloud costs.

example, financial services companies are required to keep transactional history for only as long as legally required. The SEC and most FinTechs' SPOLs to purge that data after the Data Breach is up.



The Secret to Most FinTechs' SPOLs Migration  
MAY 11, 2023

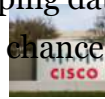


Want to Become a Data Hero? Here's How  
MAR 15, 2023

“They definitely don’t want an audit to come through and identify things they did badly in the past,” said Andy Pernsteiner, field CTO at Vast Data. “Keeping data longer than necessary can result in more liabilities, and companies don’t want to take a chance of being held liable for holding onto an asset beyond the legal requirements.”



Making the Case for Data Ingestion Pipelines  
MAY 11, 2023



NTT and Cisco Team Up to Bring Private 5G to Enterprise Players  
FEB 24, 2023

ROT data can also take a toll on the environment. As companies look to [become eco-friendly and sustainable](#), more want to reduce their carbon footprints. Data storage requires power and cooling, even if that data is stored in a cloud-owned data center. A 20TB hard drive, for example, uses about 14 watts of power, according to Srinivasan, and it will cost at least as much to cool the drive as it does to power it because of the heat it generates.

About

CCPA: Do not sell my personal info

Advertise

## How To Clean Up ROT Data

Privacy Policy

Contact Us

Terms of Service

Sitemap

Content Licensing/Reprints

Ad Choices

Cookie Policy

So how can companies ensure they manage data properly and retain only the data they need? These four tips will make a significant impact.

Follow us:

### #1: Identify your data locations

First, find out where the data is stored. That may be easier said than done since data can exist in many different locations, especially for companies that have existed for years. Structured data can reside in databases, data warehouses, data marts, and [data lakes](#), while unstructured data can reside in file systems and object stores. Then there are physical devices, like user laptops, smartphones, and USB drives.

### #2: Use a data catalog

Once you identify the data locations, move the data to a data catalog, which helps identify and eliminate duplicate and irrelevant data. Data catalogs can also fix data that has errors so that it can become more usable.

Comprehensive data catalogs should have automated data lineage creation, data profiles, and policies for granular control and [governance](#). The data profiling function examines, analyzes, and develops summaries of all data.

develops summaries of all data.

### **#3: Dive in and sort the data**

With a data catalog in place, decision-makers can gain visibility into which data is necessary and useful and which isn't. Organizations can also gain valuable insight into their incomplete data. For example, if you see that 10% of customer data lacks zip codes and another 5% lacks email addresses, you can then work to fill in those gaps.

In addition, data catalogs can show data attributes in new ways. For example, a data analyst could ask for all files or objects created by a specific department within a 2-year period, then filter that data to the files that have never been accessed. With these tools and processes, companies can more easily and safely move stale data offline or delete it altogether.

"It's not enough to know that you have 500 petabytes of data sitting on all your systems," Pernsteiner said. "You have to know how much of it is old, has GDPR-related information, or is data the organization doesn't care about anymore."

### **#4: Maintain processes and revisit policies**

Knowing what you have, organizing it, and implementing policies will help to reduce data bloat and ensure compliance. However, if you fail to maintain your data management practices, you can land you right back where you started (or at least close to it).

As such, revisit your data policies regularly and revise them as necessary. For example, if a company has a hard-and-fast policy that requires purging data that hasn't been accessed or viewed after six months, that policy in time may become unrealistic or extreme. The company may decide to use a more dynamic policy that is based on current business needs.

"If IT can prove that certain data has value to the business, there may be a value in keeping that data longer, and that might be a good reason to change the policies," Pernsteiner said.

### **About the author**



*Karen D. Schwartz is a technology and business writer with more than 20 years of experience. She has written on a broad range of technology topics for publications including CIO, InformationWeek, GCN, FCW, FedTech, BizTech, eWeek and Government Executive.*