DATA LIFECYCLE MANAGEMENT
Identify and Prevent Data Sprawl

Summary
With the blur between on-premise and cloud networks, data finds its way into many unintended places. The sprawl of data across enterprise systems increases costs and puts information at risk of loss, or theft. An enterprise's reputation can be further tarnished based on noncompliance, loss of intellectual property, and/or public embarrassment. A good Data Lifecycle Management (DLM) strategy can help manage these risks to ensure data is protected and used in accordance with policy.

Proposal
A close partnership between Enterprise (Cloud) Architecture, Security, and Data Management teams should be formed to champion this strategy. The recommended steps involved are:

- Identify where the data is going, and what type of data is involved.
- Determine why it is happening, and how to avoid future occurrences.
- Collaborate to effectively define and implement a DLM strategy.

Identify
DLM tools can assist in identifying and classifying what kind of structured and unstructured data is throughout the company, where it resides, and provide insights for remediation. This would include software to assist in discovering data in both cloud and on-premise networks. Vendor recommendations are noted at the end of this document.

Why
"Data sprawl describes the staggering amount of data produced by enterprises worldwide every day; with new devices, including enterprise and mobile applications added to a network, it is estimated data sprawl to be 40% year over year, into the next decade."[1]

In many cases, data sprawl is a result of employees just wanting to get their jobs done. This can include copying data to different devices and locations as they follow the path of least resistance. The balance between access and security may be overly restrictive, so employees find a way around policy to "make things work." If access controls or processes take too long or are too cumbersome, determined engineers and users usually find their own way. Data can also be replicated automatically for backups or infrastructure scaling, then forgotten.

In other cases, legacy applications, databases, or storage systems aren't decommissioned and present a real risk to the enterprise. With pressure to innovate, and lack of time or resources to destroy or archive data, the circle of DLM (see image below) is left open, leaving data at risk.

DLM Strategy

Ideally, a closed loop process for the lifecycle can be defined, similar to the image below. By ensuring data is managed from beginning to end, the risks to the enterprise can be reduced or mitigated.

Circle of Data Lifecycle
Image Source: https://www.spirion.com/data-lifecycle-management/

If research shows a high likelihood and significant impact to the organization, an ROI analysis should be pursued on the recommended changes. This would include costs of resources, tools, and systems.

Subsets of DLM include Data Loss Prevention (DLP), and Data Retention Policies which can be addressed by partnering with the Information Security and Data Management teams.

A favorable way to start things off would be finding which systems and data are most vulnerable. Then harden, isolate, migrate, archive, or destroy the data. The other option is to do nothing if the data is deemed sufficiently secure, or if the business accepts the risks.

A good way to assist in managing data is to empower users through automation. Different approaches should be investigated so that data sharing adheres to policy, and is kept within predefined boundaries. By creating an automated, easy to use provisioning system with good DLM controls, users will be less likely to work around them.

The next step would be to review default policies in both on-premise and cloud infrastructure. This review should include traditional infrastructure such as application servers, databases, and enterprise storage systems. A cloud policy review for SaaS, PaaS, and IaaS solutions should be considered for systems such as Outlook, OneDrive, AWS, Azure, and other cloud database and storage platforms. It is also imperative to review policies for local workstations, and mobile devices.

Policies should be reviewed and updated regularly. Any changes to the DLM standard should focus on small, easy wins to gain momentum.

Automated replication of data should be documented, and have monitoring attached to it so problems are found and remediated quickly.

Legacy systems identified through automated discovery or the enterprise CMDB should be decommissioned when they are no longer used. The related data should follow the DLM standard.

Training can be provided to data owners to ensure they understand the DLM strategy, and their responsibilities from the beginning to end.

In all cases, regulatory and compliance standards should be adhered to and considered when designing a solution. Being a wise steward over the data will protect the enterprise and individuals.

Vendor Considerations

Evaluate tools/software/systems from vendors to assist in defining a good DLM strategy such as Alooma, Spirion, Komprise and Forcepoint. A software gap analysis will be created to find out where vendor's strengths and weaknesses are, and see how they align with the enterprise's mission and strategy. From there, a proof of concept should be setup with the vendor of choice to evaluate theoretical vs. practical application.

References:

1) https://www.komprise.com/glossary_terms/data-sprawl/