Data Classification Can Make or Break Data Governance

By **Steve Zagoudis** on **November 16, 2022**

Read more about author Steve Zagoudis.

The cost of data loss is staggering – and avoidable. The news these days is full of massive fines levied against companies that failed to protect their customer data. An effective data classification approach is one of the best ways to ensure that companies can identify and protect their most valuable data.

Our experience continues to show companies often lack awareness of this critical data. Through all the processes, procedures, and technology, this data inventory is not sufficiently known. Bureaucratic processes focus on doing a lot of busy work, instead of solving the right problem. In this blog post, we will turn our attention to the most critical data assets in terms of data loss prevention – namely, confidential and personally identifiable information (PII).

Pause for a few seconds and ask yourself the following questions:

- What data is considered confidential or PII in your organization?
- Where are all the copies of confidential and PII data?
- Who owns them?
- Who uses them?
- Are they in sync?
- And are they adequately protected?

The ability to answer the above questions is a fundamental pillar of Data Governance. The answers lie in stewardship models, data lineage, data quality controls,

data loss protection, encryption, etc. In our quest for Lean Governance, it is paramount to solve the right problem. And do so in an efficient manner.

Data Governance needs to be built around classification. And governance or security implementations should be right-sized based on the class of data. Companies do not need to waste time and resources protecting data that is classified as public, or even internal. Loss or disclosure of these classes of data will do little harm. It is the confidential and PII data loss that leads to $100 million-plus fines.

Data classification is more of an art form than a science. Our methodology distinguishes between business attributes (business terms) and data objects (physical copies). The key to data classification is that you need to classify the business attributes, in accordance with your internal information security policies.

Let's use the Social Security Number (SSN), for example. Few business attributes are as vulnerable to data loss as SSNs. In most cases, this attribute is classified as PII. All data objects will inherit the PII rating. This means that all copies in transactional systems, data warehouses, data staging, and user-developed applications (spreadsheets) need to be protected in accordance with your data loss prevention approach.

But the work does not stop there. You need to extend this classification down to the unstructured data – the documents, reports, emails, etc. that are often not considered. The focus on structured and unstructured data is required to maintain awareness of the locations of SSNs across your company.

Metadata is the glue that binds this information together. The classification of the subject area and the business attribute are defined as metadata. The link between business attributes, structured data, and unstructured documents is also metadata. Likewise, the association to the governance stakeholders can be represented as metadata. This shows the emerging power of Governance Metadata Management. However, your choice of metadata tools is critical to your success. Focusing on confidential and PII data bring the scope within reason, and proactively eliminate potential data loss risk.