2021: The Year Dark Data Turns Light

BY KON LEONG ON DECEMBER 4, 2020 BIG DATA ANALYTICS, CCPA, GDPR 0

2021 will be the year companies begin to shed light on and harness the power of their dark, unstructured data. Notably, over 80% of companies' data is considered "dark," or unused and unanalyzed. The majority of which is unstructured, meaning it is information created by humans, for humans: emails, file shares, messages, etc. Despite the mass number of insights hidden in dark data, few organizations have had the capacity or knowledge to properly leverage this information for decision making. However, the efforts required to adopt solutions to govern unstructured data are well worth it, as the benefits of analyzing this information range from increasing efficiency to mitigating risks of insider threats.

It is not always that regulation drives transformation—sometimes it stalls it. However, in this case, regulatory developments over the past two decades have set the stage for companies to leverage unstructured data. With mandates set by laws like Sarbanes-Oxley, SEC 17a-4, and Federal Rules of Civil Procedure, which mandate oversight of electronic data, regulated industries have been forced to start learning how to navigate employee-created data. Now, new privacy regulations such as the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA) take it a step further by requiring companies to identify and remediate personal information.

Despite causing technological challenges, companies that have learned to adapt to these demanding regulations may now have a head start in controlling, analyzing, and deriving insights from their unstructured data. When properly leveraged, organizations use dark, unstructured data to answer questions that will better their efficiency and grow revenue. A notable difference, however, lies in the mode of management: While industry regulations and eDiscovery requirements often require data archiving—creating and securing a copy of each document—a more pragmatic approach for analytics may be "in-place" management, in which no copy is needed.

Early yet powerful use cases for leveraging dark data include:

- Who are the top performers (or as some call them, "STPs"—the same ten people)
- Who knows and has established relationships with a particular client?
- How is the workforce performing while remote?

Moreover, these same tactics can be applied for risk reduction. One of the reasons insider threats pose such a complex challenge is because of the ease of employee access to internal data and the multifaceted nature of human behavior that must be accounted for. Which is where analytics of unstructured data comes into play in uncovering—and ideally, preventing—insider threats, by answering questions such as:

- Has anyone accessed, moved, or deleted sensitive data in mass?

- Has there been any extreme negative sentiment towards the company expressed in communications?
- Do all sensitive data have proper access controls?

Instead of being entirely reliant on network security, organizations can preemptively track, analyze, and manage dark data at the content level to preventatively stop a potential insider threat before it is executed.

As Carl Jung said, "Knowing your own darkness is the best method for dealing with the darknesses of other people." Though the Swiss psychologist's words predated the age of information, they were prescient—now, by understanding dark data, companies can foresee acts of insider threat before they even happen.