



## Dark Data: Goldmine or Minefield?

ON NOVEMBER 11, 2020

DIGITAL TRANSFORMATION

**As someone once said, confidence is the feeling you have before you understand the situation**

For many years, companies have found themselves in that blissful state with regards to their data. The fact that up to 80 percent of an organization's data is typically dark may then give you a sense of how far there is to go before companies can even begin to say they have a grasp on their data. However, for those truly steeped in the risks and untapped potential posed by dark data, it's safe to say they're well past confidence.

### **What makes data “dark”?**

Over the past several decades, the IT architecture has developed around data siloes: dozens and even hundreds of disparate repositories housing everything from structured data and log files to human-created data such as email and file shares. New siloes are created to manage data for functions such as compliance, records management, and eDiscovery. On the human side, there's been an exponential acceleration of data creation, leading [some to estimate](#) that 90% of the world's data has been created in the past two years alone. With that comes new data

siloes—for example the shift to remote work has led to massive adoption of collaboration platforms such as MS Teams, which at large companies are being used by thousands of employees. Yesterday's systems for managing company data are simply not built to scale to the billions of documents and communications that are now scattered across the enterprise—now on-premises and in the cloud. The result? Petabytes and petabytes of that is unmanaged, unharnessed, and altogether unknown.

## **New Rules**

The risks associated with dark data are numerous. Enter new business requirements and risks, such as privacy regulations that mandate an organization find and remediate any personal data when requested by an individual. Led by the General Data Protection Regulation (GDPR) and now the California Consumer Privacy Act (CCPA), these regulations simply do not respect siloes. Organizations are effectively required to search across the entire enterprise to find personal data relating to any John Doe, or risk millions in fines and class action lawsuits. Putting aside the complexities of searching across billions of documents in hundreds of repositories, the prosecution of these searches must also work hand-in-hand with governance policies around compliance, records management, eDiscovery, and more. More than ever, there is an essential need for a holistic governance layer across data to facilitate synergy across these functions.

## **Convergence**

Despite this cautionary tale, dark data has unlimited potential. Through the advancement of analytics, dark data can be repurposed into a powerful business asset. It can be used to drive new revenue sources, eliminate waste and reduce costs, and companies that are able to solve the governance equation take a giant leap forward towards harnessing data for analytics.

For instance, data created *by* humans *for* humans, such as electronic communications and documents, carries massive insight into a company's workforce and business. Yet because this data is unstructured and

ungoverned, in most organizations it lies around collecting dust (and risk). When attempts are made to harness it for analytics, results are less than optimal.

Imagine the entirety of a company's data as a beach and each grain of sand a document. The current paradigm for analytics is built around sampling this data to find a representative data set—you might say a sandbox—, cleaning and scrubbing it (this is where data scientists spend huge amounts of energy), and then exporting it to an analytics platform. This often takes many iterations and results in an incomplete data set, and as the final deliverable, an analysis that is neither in time nor dependable.

The issue here is that companies are still thinking in terms of sandboxes even when they are face-to-face with the entire beach. A system that considers analytics and governance flip sides of the same coin and incorporates them synergistically across all enterprise data is called for. Data that has been managed has the potential to capture the corpus of human knowledge within the organization, reflecting the human intent of a business. can offer substantial insight into employee work patterns, communication networks, subject matter expertise, and even organizational influencers and business processes. It also holds the potential for eliminating duplicative human effort, which can be an excellent tool to increase productivity and output. The results of this alone are a sure-fire way to boost productivity, spot common pain points that may not be effective to the workstream and can share insights to organizations where untapped potential may lay.

Companies that have successfully bridged information management with analytics are answering fundamental business questions that have massive impact on revenue: Who are the key employees (“TSTPs”—the same ten people that get everything done)?

- Who are the experts and decision makers in particular areas?
- Who has the strongest relationship with a particular customer?

- Sales Analytics—Communication patterns can uncover successful vs. unsuccessful Salesperson up to 9 months before the human eye, saving hundreds of millions in payroll
- Crisis Management—Instant replay during crisis, to make better and quicker decisions during key moments

## **Paradigm Shift**

With the increase in sophistication of analytics and its convergence with information governance, we will likely see a renaissance for this dark data that is presently largely a liability. Moving forward there should be a welcomed shift in the industry's perception on dark data, to make way for its unique business potential. Gone will be the days where organizations systematically ignore content that isn't proven to be a formal business record. Rather, the big data era tends towards managing everything – for better or worse.