“...organizations often retain dark data for compliance purposes only. Storing and securing data typically incurs more expense (and sometimes greater risk) than value.”1

DATA ON THE MOVE
For decades, data management has been a dysfunctional interdepartmental challenge. Compounding yesterday’s data management problem is that today’s data is on the move. With the steady migration of existing and new workloads to SaaS and Cloud, the data center perimeter is fading fast.

The challenge in such a borderless environment is that as data fragments and the volume of machine and human generated data grow at exponential volumes, visibility and control shrinks, leaving many blind spots. Developers might be using unmasked data without thinking about the regulatory exposure. Analytics staff might catapult whole data sets up to the cloud seeking differentiating insight. Other groups might have Excel spreadsheets full of PII.

And while redundant copies of the same data provide for better perceived productivity for knowledge workers and developers, it also creates gaping holes in security and additional points for data exploitation. Policies might be in place but without technical oversight, violations can go unnoticed for quite a while. When visibility and oversight drops to zero, poor data management strategies become a potential threat, slowing digital initiatives, increasing security risk and exposing the organization to regulatory penalties.

As we evolve the way we do business, very frequently the mishaps wind up in the news. Organizations must decide what their tolerance for risk is, and manage the risk with better data management and governance. Regaining and keeping control seems complicated because data is generated in many discrete systems, but there is one IT service that touches all of them: backup and recovery. Leveraging backup for more than just an insurance policy is very appealing because it allows IT teams to reuse something they are already doing. Rubrik now makes this possible.

PRIVACY, COMPLIANCE, AND SECURITY
On the backdrop of the fundamental and chronic data management problem, regulatory bodies are increasingly moving to formalize the rights of the individual through legislation and enforcing those regulations with penalties. At the end of May 2018, the General Data Protection Regulation (GDPR) came into effect, and the California Consumer Privacy Act (CCPA) was signed into law one month later. Other US states and countries have their own versions pending.

GDPR and other regulations and standards like Health Insurance Portability and Accountability Act (HIPAA) and Payment Card Industry Data Security Standard (PCI-DSS) might regulate different types of data, but all of them stipulate that data under management should be stored securely. Failure to do so means potential exposure, and breaches with data loss must be disclosed. A breach disclosure could then result in a time consuming audit or legal investigation, damage to reputation and brand, and customer loss.

1 https://www.gartner.com/it-glossary/dark-data
$3.92 million
Average total cost of a data breach

$8.19 million
United States – the most expensive country

$6.45 million
Healthcare – the most expensive industry

25,575 records
Average size of a data breach

Source: Ponemon Institute 2019 Cost of a Data Breach

**IMPROVE RISK MODELING**

Blind spots multiply risk and leave organizations in the unsteady position of not being able to report on or factor in potential exposure. Inability to accurately model or project risk is an untenable position in 2019 and beyond. It is no longer enough to simply adhere to checkbox compliance regulations. Establishing the level of acceptable risk needs to be an informed decision, and is organization specific with an understanding of the business itself. This avoids rigid execution of risk management or a policy that overextends itself. The forward-looking company seeks a more integrated approach to risk than merely satisfying auditors and acting in a purely reactionary manner to the latest regulation du jour.

Understanding the business avoids a clumsy binary approach that has alarm bells going off for something that is not so immediately pressing. It should be possible to discern (and have a measured response) among situations where routine data management hygiene or remediation is needed, an incident that involves exposure but no loss, and a breach where data exfiltration is underway.

**QUANTIFY THE COST OF PRIVACY**

Safeguarding sensitive data and the privacy of customers has become a business critical requirement. Establishing the metrics to understand and quantify the cost and value of privacy is however quite challenging. What is clear is the cost of non-compliance is significantly higher than the cost of compliance. Many organizations adopt investigative, preventive and data governance technologies to protect their systems and data from ransomware attacks, and improve security and compliance. Technology has a cost, but it scales FTEs much better by automating the process. There are also intangibles such as reputation and trust. Fines are a clear risk from not having visibility, but what about the risk to innovation and transformation initiatives? Too many siloed tools and incorrectly utilized FTEs serve as a distraction from CIO, CEO, and Board goals that results from a less than ideal solution, especially when FTEs and budget are constrained. Together not everything can be quantified, but it helps model the risk and justify the expenditure.

**BLINDFOLDS ARE FOR PIÑATA PARTIES**

Expecting the end users, an administrator, or even business application owner to detect and alert when something is wrong is simply unrealistic. The application owner might think everything is done correctly when in fact their application is writing to a poorly secured NAS share. Another common source of surprise comes when conducting due diligence post merger or acquisition. Uncovering data exposure across many data silos without automation is not practical or even possible. Out of place data can be accidental or intentional. Either way, investigation and remediation is necessary. This is more than mere data discovery or classification, it’s more than compliance or just governance. This is risk identification and reduction into the dark areas of masses of data.

Lack of data insight makes it a liability. It costs money to store it, back it up, and in the event of a legal incident, the more data that must be trawled means significantly higher effort and associated costs.
COUNT ON COMPLIANCE (TO DISTRACT YOUR TEAM)
Governance, Risk, and Compliance follows the pattern of being able to demonstrate compliance to an auditor when needed or respond to a legal hold or legal discovery. Compliance to regulations does not necessarily equate to security. An auditor could claim an organization is compliant and a breach could happen the next day. The false security of compliance is well understood by InfoSec practitioners. While we should always strive for prevention, there is simply no such thing as 100% security. This has led to an evolution of a risk-based approach and a desire for more integrated risk visualization, modeling, and identification. One thing is certain, exposed data is more vulnerable.

As well-meaning as regulatory compliance is, the activity surrounding an audit can be time-consuming. For this reason, compliance must be dealt with while overcoming the constraints on visibility, organizational reluctance, and resources to conduct the activity. Insight automation and reduction of effort are key.

INTRODUCING POLARIS SONAR: AUTOMATED DATA GOVERNANCE
Rubrik Polaris Sonar is a SaaS application designed to discover, classify, and report sensitive data across the enterprise environment with zero additional infrastructure. Sonar simplifies the complexities of dealing with an ever-increasing number of regulatory requirements and gain insight into policy violations, track compliance progress, and help identify at-risk data. Natively integrated into the Rubrik platform, Sonar requires no agents, install, upgrades, or additional infrastructure, and can be centrally managed using a simple web interface. Sonar maximizes the value of data under management for real-time insights into security vulnerabilities and compliance violations.
DATA GOVERNANCE WITH POLARIS SONAR

The fundamental design of Rubrik Cloud Data Management (RCDM) allows for a natural extension of data management capabilities to data governance. As data is protected, file level metadata is gathered to enable search. Going a level deeper and gaining insight within the files themselves unlocks the data. Sonar reduces data exposure by applying machine learning to automate sensitive data classification, such as personally identifiable information (PII) and intellectual property, allowing enterprises to easily identify and lock down data exposure without impacting production. By leveraging data existing in RCDM, Sonar delivers unprecedented simplicity and rapid time-to-value.
EASE OF USE
To begin using Polaris Sonar, all a Rubrik customer has to do is turn it on and select one or all of the pre-defined templates such as US PII, HIPAA, PCI-DSS, and the locations that should be analyzed. That’s it. You just created your first policy. Polaris begins to scan the data in the background and flags matches to the defined policy. Since production workloads are already being protected, the data to be analyzed exists within the Rubrik CDM repository. Sonar sweeps through protected environments searching for defined patterns whether the data is on-premises, in a remote office, or in the cloud. This means no agents to install, no heavy scans on production workloads, and most importantly no heavy lifting for the IT staff.
CUSTOMIZABLE TEMPLATES

You know your data, and you know what should not be floating around. The pre-built templates cover some of the typical use cases, but modifying or tuning a policy is not rocket science. Either define a regex to match the patterns found in sensitive data, select a few words, or import an entire CSV dictionary. You are flexible. You can combine templates to create an overall policy that is just right for your organization.

![Create a custom analyzer](image1.png)

![Create a custom analyzer](image2.png)
RUN CONTINUOUSLY

After defining a policy the advantage of Polaris Sonar is that it scans continuously as new production data is created and protected with RCDM. Sonar allows for stepping back to each snap as well in the event that forensic examination might be required. This continuous background analysis improves predictability and accuracy of risk assessments and alerts to policy violations. Engaging with auditors proactively, demonstrating Sonar’s capabilities, and collaborating on an acceptable policy definition can be enough to satisfy many auditors.

In the event that a spot search must be performed, the same capabilities are available for a one-time search and discovery. This functionality can support GDPR data subject requests (GDPR Article 15), legal discovery, or other one time activities where it is required to find relevant data.
**EXPORT TO E-DISCOVERY**

Organizations have their own e-Discovery tools. Sonar is not designed to replace them. Our export to CSV function can save the results of a one-time discovery or background scan into a file that is the commonly supported format for e-Discovery tools. This alleviates the need to run an indexer or scanner on production workloads, and integrates with the requirements of legal discovery.

The combination of running continuously and export capabilities can provide the groups that need it with up to date reporting insight they need, and support discovery requests when the need arises.

To learn more about Polaris Sonar and how it can help your organization automate their data governance requirements, please visit our [website](#) or contact your local sales person.