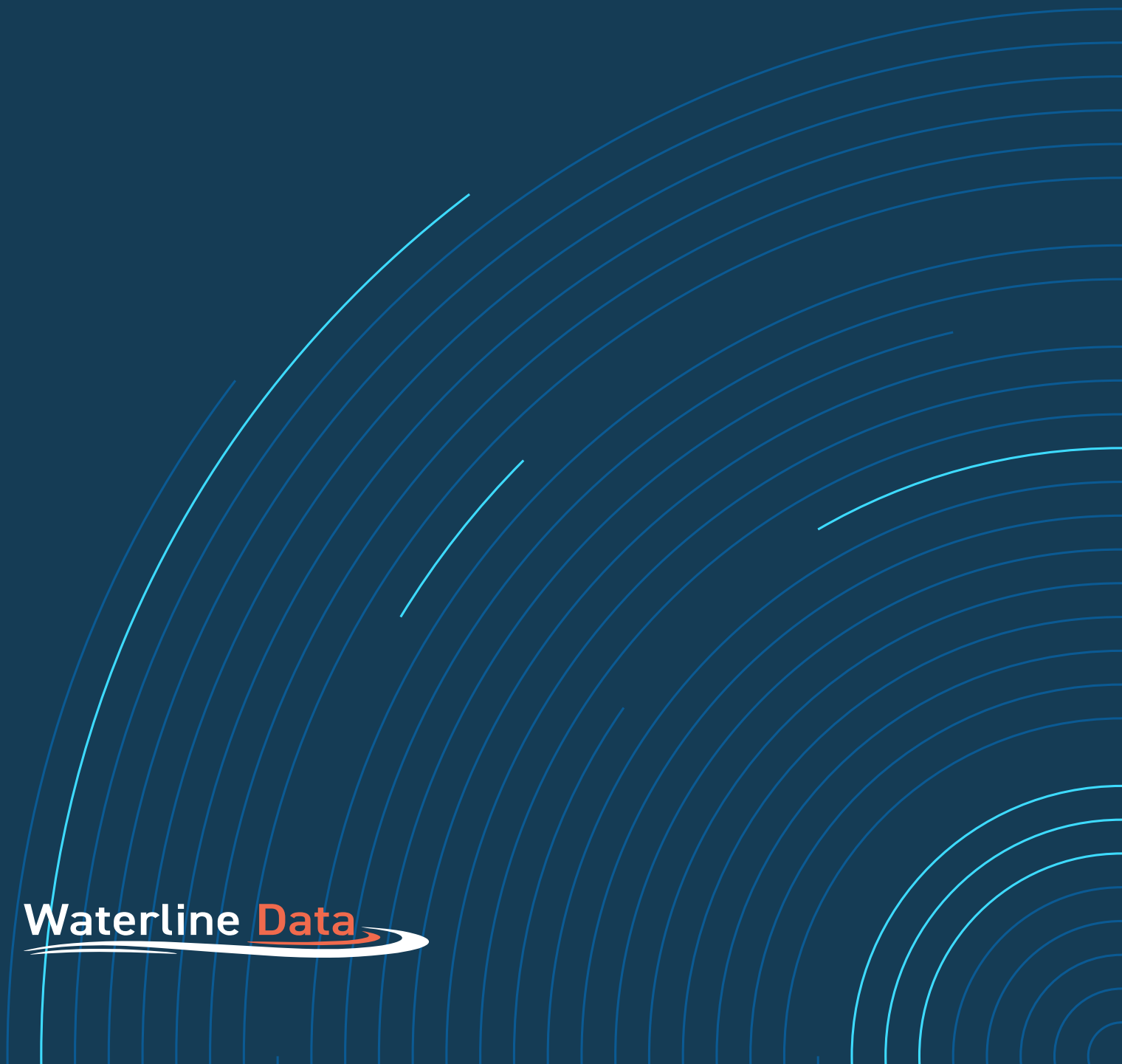
SOLUTION OVERVIEW

DATA CATALOGS FOR RISK AND COMPLIANCE



Waterline Data

Introduction

As governments increasingly recognize the importance of data and the potential for its misuse, the amount of compliance rules and regulations are continually increasing. The breadth and depth of compliance regulations is staggering. Examples of these regulations include, Basel I, Basel II, Basel III for financial institutions, GDPR (General Data Protection Regulation) often referred to as the "right to be forgotten" which is intended to give individuals control of their personal online information, HIPAA (Health Insurance Portability and Accountability Act) in the United States for protection of Personal Health Information (PHI), data sovereignty rules in Germany, China and other countries that forbid transfer of personal data out of the country, and yet even more data privacy regulations.

The fundamental reason these regulations exist is because at some point in time, the public trust was violated and new rules and regulations were created to keep everyone in line. And while most executives view data regulations as a nuisance, the risk of non-compliance can be hefty and must be avoided. GDPR is the latest regulation to hit the world and its fines can range from a minimum of €20million to 4% of worldwide revenues. None of this includes the negative cost of bad public relations when one of these rules is violated and makes it into public view.

GDPR is the latest regulation to hit the world and its fines can range from a minimum of €20million to 4% of worldwide revenues

The Common Thread

While the details of these regulations differ greatly, there are several common and very fundamental themes that cut across these regulations. To one extent or another they all require that organizations answer some basic questions:

- Where the data protected by the regulation is located?
- Where did the protected data come from and where is it going? What is its lineage?
- Do you have control of who has access to the data, especially sensitive and private data?

These sound like pretty basic points. And most people who aren't used to dealing with data think to themselves, "But the data is in the computer, right? You must know where it is?" But as those with any reasonable amount of experience with data know, data is more like water. Data moves between and across systems and much like water, which can corrode your pipes, can have a nasty habit of leaking out into unknown locations.

Plus, with the introduction of "big data", the amount of data continues to double almost every two years. So keeping your data ready to be governed is much harder than it sounds. The result is that most organizations don't even have the baseline infrastructure to properly support data governance initiatives.

The result is that very often the solution to governing data is to lock it all down, put it into quarantine and limit access. This also has the unfortunate side effect of treating all data as if it is sensitive. People who have legitimate need to see some data, but not all data have to go through a centralized group to process their requests. Unfortunately, these centralized groups are often understaffed, so access to data is slow and frustrating and the concept of "Data Governance" now has a bad reputation. If this sounds like your organization, don't be alarmed. This is the norm in most companies.

Data Catalog Powering Agile Governance

To solve these problems, you should think about your data like any other physical asset. If you were managing data like raw materials in a factory making cars, the first thing you would do is take an inventory. Then you would track that inventory from raw materials (steel), to work in progress (car frame), to finished goods (the car). You would also put in control mechanisms so only the right people who were properly trained could touch sensitive (or hazardous) materials on the factory floor.

The same thing is true for data and this is where the Waterline Smart Data Catalog can help. The Waterline Smart Data Catalog automatically "fingerprints" data at scale by analyzing source data. If a business glossary exists, Waterline Smart Data Catalog organizes the data using machine learning to automatically tag and match data fingerprints to glossary terms. This is the process that builds an inventory of all your data and answers the first question:

"Where is the data protected by the regulation located?"

As part of this inventorying and cataloging process, Waterline identifies sensitive data that must be either masked, deleted or put under access control. It also captures data lineage, importing lineage from other systems and filling in the gaps automatically, answering the second question:

"Where did the protected data come from and where it is going to?"

Finally, because Waterline tags the data, sensitive or private data will have their tags passed on to access control tools like Apache Ranger or Cloudera Atlas (or others via a REST API) which can then be used to make sure only people who have access rights can see the sensitive data. This answers the final question:

"Do you have control of who has access to the data?"

The big advantage of this approach is that not only does it allow organizations to demonstrate they have an auditable data inventory and lineage, but because everything is documented in a searchable catalog, they can react quickly to new regulatory requests. Additionally, data governance is no longer a bottle neck in the organization, but becomes an agile group responsible for quickly getting the right data into the hands of the right users in a governed, compliant manner.

everything is documented in a searchable catalog

How Does It Work?

The Waterline Data Catalog is built on the premise that there are three parts to the data cataloging process. The first is to help data professionals discover, organize and curate the data; the second is to allow governance professionals leverage tagged data for compliance reporting and access control and the third is to expose the newly organized data for business professionals to use.

Discover: Automatically and incrementally "fingerprint" data and infer data lineage at scale by analyzing actual data values

Organize: Uses machine learning to automatically tag and match data fingerprints to glossary terms

Curate: Human reviewers accept or reject tags, machine learning fine tunes the tagging process and improves the matching algorithm

Compliance: Map your compliance policies to your data assets: acceptable use, legal holds, expiry.

Reporting: Simplify ongoing mandated reporting as the catalog uncovers "dark data" and catalogs it dynamically

Access control: Automate data access control via tag based security

Search: Search for data through the Waterline GUI or through integration via 3rd party applications

Rate & Collaborate: Users collaborate to create subjective crowdsourced ratings/reviews which combined with objective profiling metadata provides users with a view into data quality and usefulness.

Benefits for Risk Management and Compliance

Automating the end to end data cataloging process with Waterline Smart Data Catalog significantly reduces the cost of data governance while making it a process that can be used for competitive advantage.

- 90% reduction in people cost for tagging and inventorying data. Spend more time using data and less time searching for it.
- 80% time reduction to get data out of “quarantine” and into use
- Significant increase in data inventory and lineage accuracy reduces risk of financial penalty (e.g. 4% of worldwide revenue from GDPR fine exposure)

Conclusion

Data Governance doesn’t have to be the equivalent of a four-letter word that leaves business professionals muttering under their breath. Those organizations that can properly get data into production faster will reap the benefits that come from being first to market with new information, products and services. By automating the underlying process for inventorying, tagging and curating data and data lineage, new data can quickly be passed through quarantine and securely put into use while turning the governance process from a liability to an asset that delivers competitive advantage.

Waterline Data

waterlinedata.com

Sales

(800) 123-4567
sales@waterlinedata.com

Technical Support

Visit the Support Center
help@waterlinedata.com

Corporate Headquarters

201 San Antonio Circle Suite 260
Mountain View CA 94040
(650) 946-2104