WHITE PAPER

Automated Data Governance 101

An Introduction to Proactively Addressing Your Privacy, Security, and Compliance Needs — and Maintaining Customer Loyalty

2019

IMMUTA™

Contents

"Data privacy," "data security," "data protection" – whatever we call the way we control our data, it isn't working. Data is as vulnerable as ever. And this is true for both consumers hoping to keep their data safe, and for enterprises seeking to govern their corporate and customer data.

Why?

When it comes to the enterprise, data science and data governance programs are built on competing objectives, and enterprises have no choice but to try to satisfy both demands. Compliance and privacy programs, for example, require increasingly strict limitations on who can use what data and why, thanks to increasing regulations on data. On the other hand, data scientists need access to as much data as possible, as quickly as possible — otherwise their analysis could fall flat.

This is a problem that businesses simply don't know how to solve.

From a data science perspective, organizations have no choice but to become data driven. From finance to healthcare to retail and more, enterprises are investing in more data science, more machine learning, and more advanced analytics to strengthen their competitive edge. That's why an astounding 70 percent of Fortune 500 CEOs now claim to be running technology companies — so important is data to the future of their business.

At the same time, governments are placing stringent limits on how data is collected, put to use, and when it must be deleted to ensure their citizens are fully protected. This is true from California to India and the European Union (EU), as we'll discuss below. At the same time, consumers are demanding more guarantees that their data is being used responsibly. Enacting stricter, standardized controls on data across the enterprise is no longer optional.

Any way you look at it, we're at a crossroads: governing data and putting data to use are two dueling objectives, and businesses are stuck in the middle.

Can this problem be solved? In a word: yes. This white paper will show you how.

Our founding team spent over a decade working in the U.S. Intelligence Community, tackling some of the most complex and sensitive data problems in the world. Our team of experts in distributed systems, cryptography, enterprise analytics, and law are all committed to our mission of enabling the legal and ethical use of data. This paper outlines lessons we've learned about how data science and governance programs can, if implemented properly, reinforce each other's objectives, and in the process, strengthen the core goals of your business and maintain customer loyalty.

What's in a Name?

The term "data governance" is a loaded term — it can convey everything to everyone, and yet at the same time, nothing to no one. The way around this dilemma is to think about **data governance** in terms of the three problems it solves:

1. **Privacy:** Governance must ensure data is put to use in accordance with the rights and expectations of those who generated it.
2. **Security:** Safeguards the confidentiality, integrity and availability of that data and the systems in which it's used.
3. **Compliance:** Requires aligning how data is used and stored to legal and policy mandates.

Data governance is the process of addressing these three needs, and requires ongoing collaboration across different teams with different sets of expertise. True success, as we'll describe in this paper, requires nothing short of automation.

THE OLD WAY OF GOVERNING DATA:

A Passive System That Can't Scale

Passive processes hold your data and your business back.

What's a passive process? A process that's entirely reactive. In the world of data governance, passive processes involve waiting to evaluate requests for data until they're actually made. Traditional signs of passive processes include time-consuming meetings, long policy memos, custom permissions, policies that vary per database, or the creation of new copies of data to satisfy compliance or privacy concerns.

Time to Data

Here's one way to tell if the passive approach to data governance is plaguing your organization:

How long does it take between

1) when your organization collects data, and
2) when that data can be accessed and used?

☐ Days? ☐ Weeks? ☐ Months?

For most businesses, the answer is almost always "WAY TOO LONG." Every increment of time that passes between collection and use makes your organization's data less reflective of the present (and therefore less valuable). The more passive processes you use to govern data, the less useful your data is and will become.

NEW REGULATIONS:

The Straw That Broke the Camel's Back

A passive approach to data governance doesn't work well — it's reactive, time consuming, and dilutes the speed and value of enterprise data science initiatives. But for some businesses, being passive has worked "well enough." No need to spend time and resources fixing something that's not entirely broken, right? Wrong.

New data privacy regulations around the world are making the data governance ecosystem even more complicated. Almost every day, a new regulation is enacted or proposed that increases penalties associated with poor data governance. A few examples include:

- The EU's ***General Data Protection Regulation*** (GDPR), which came into force in May 2018 as the first and most stringent law in a new wave of global privacy regulations. With fines of up to four percent of global revenue, the GDPR has driven many global companies to rethink how they collect and use their data.
- The ***California Consumer Privacy Act*** (CCPA) was passed in 2018 and will go into effect January 2020. Frustrated with stalled federal efforts to create a national privacy standard in the U.S., state legislators in California implemented some of the strictest standards on consumer data in the nation, potentially affecting any business that collects the data of California residents.
- Not content to let Western economies lead the data protection charge, the Chinese government enacted the ***Cybersecurity Law*** in 2017. A host of regulatory developments related to the law have placed increased penalties on the misuse of data collected or stored within the world's second largest economy.

In the U.S., Congress has proposed dozens of bills in recent years to enforce new national standards for privacy, some of which could cost organizations upwards of $122 Billion USD per year.[1] Hundreds of similar proposals have been made and passed at a state level, from an industry-wide data protection act in Ohio, to restrictions on biometric data in Illinois, to limitations on retail data in New Jersey. Brazil, the eight largest economy in the world, in 2020 will begin enforcing its own version of the GDPR, the ***Lei Geral de Proteção de Dados*** (LGPD).

Throw a dart at a map and chances are you'll hit an area where new privacy regulations are making it harder for businesses to collect and use data.

How can data-driven businesses deal with so many different regulations on data? A passive approach to data governance will just break down under these circumstances. Manual, labor-intensive approaches to applying rules (a.k.a. policies) on data cannot keep up with a regulatory environment that's increasing in complexity and intensity.

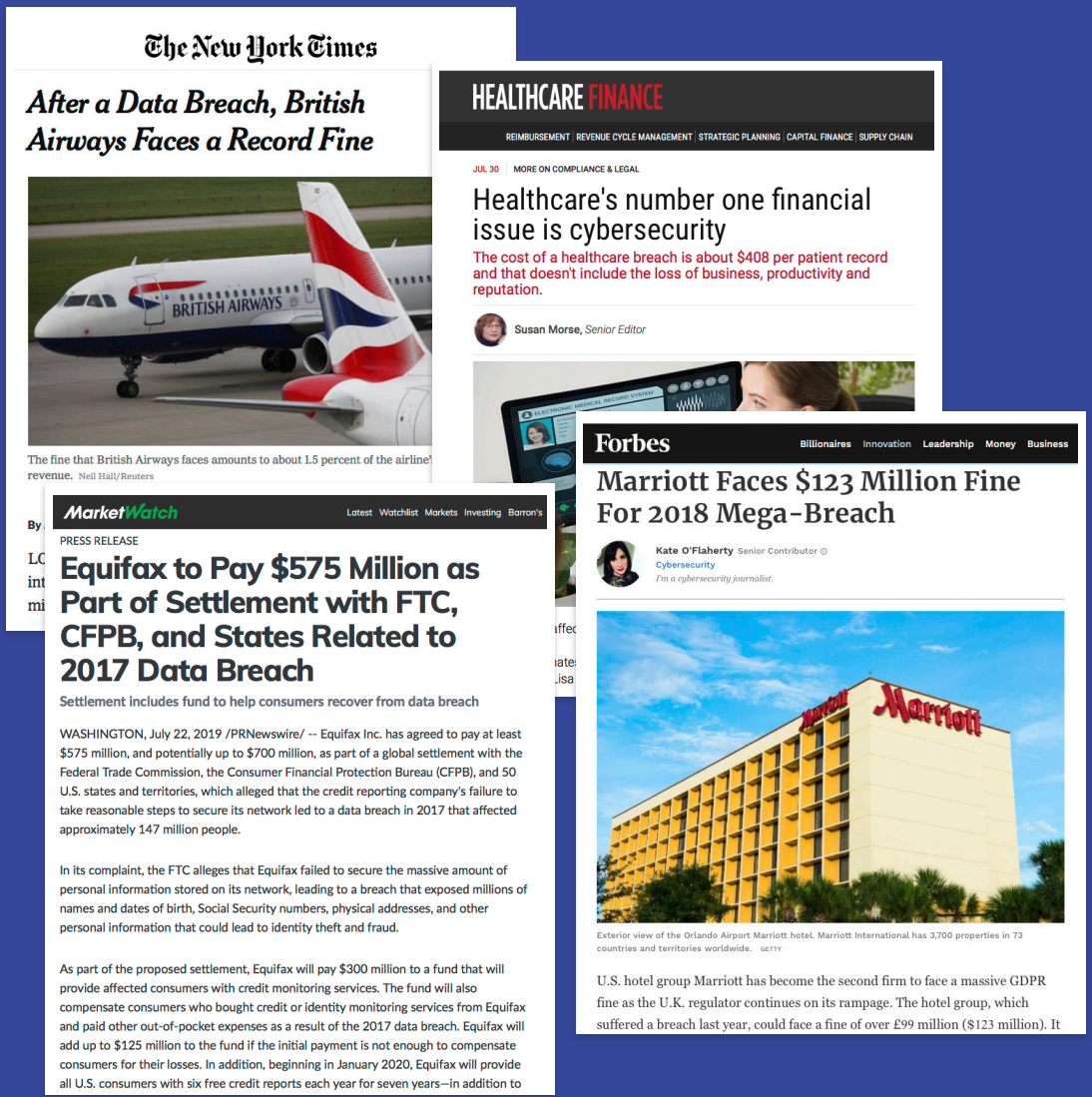
Businesses need a better way. And that's where automated data governance comes in.

> "Accelerating Privacy Regulation" Was Named the **#1 Risk Among Executives** Surveyed for Gartner's Q1 2019 Emerging Risk Monitor Report
>
> Gartner

1 https://itif.org/publications/2019/08/05/costs-unnecessarily-stringent-federal-data-privacy-law

The Consequences of Getting Data Governance Wrong

Governments are stepping in to ensure that businesses are governing their data correctly, and big tech companies like Facebook or Google aren't the only ones being fined. From large enterprises to small- and mid-sized businesses, companies of all stripes and across all verticals are being held accountable for poor data protection practices.



The New York Times

After a Data Breach, British Airways Faces a Record Fine

The fine that British Airways faces amounts to about 1.5 percent of the airline's revenue. Neil Hall/Reuters

HEALTHCARE FINANCE

REIMBURSEMENT | REVENUE CYCLE MANAGEMENT | STRATEGIC PLANNING | CAPITAL FINANCE | SUPPLY CHAIN

JUL 30 | MORE ON COMPLIANCE & LEGAL

Healthcare's number one financial issue is cybersecurity

The cost of a healthcare breach is about $408 per patient record and that doesn't include the loss of business, productivity and reputation.

Susan Morse, *Senior Editor*

MarketWatch

Latest Watchlist Markets Investing Barron's

PRESS RELEASE

Equifax to Pay $575 Million as Part of Settlement with FTC, CFPB, and States Related to 2017 Data Breach

Settlement includes fund to help consumers recover from data breach

WASHINGTON, July 22, 2019 /PRNewswire/ -- Equifax Inc. has agreed to pay at least $575 million, and potentially up to $700 million, as part of a global settlement with the Federal Trade Commission, the Consumer Financial Protection Bureau (CFPB), and 50 U.S. states and territories, which alleged that the credit reporting company's failure to take reasonable steps to secure its network led to a data breach in 2017 that affected approximately 147 million people.

In its complaint, the FTC alleges that Equifax failed to secure the massive amount of personal information stored on its network, leading to a breach that exposed millions of names and dates of birth, Social Security numbers, physical addresses, and other personal information that could lead to identity theft and fraud.

As part of the proposed settlement, Equifax will pay $300 million to a fund that will provide affected consumers with credit monitoring services. The fund will also compensate consumers who bought credit or identity monitoring services from Equifax and paid other out-of-pocket expenses as a result of the 2017 data breach. Equifax will add up to $125 million to the fund if the initial payment is not enough to compensate consumers for their losses. In addition, beginning in January 2020, Equifax will provide all U.S. consumers with six free credit reports each year for seven years—in addition to

Forbes

Billionaires Innovation Leadership Money Business

Marriott Faces $123 Million Fine For 2018 Mega-Breach

Kate O'Flaherty Senior Contributor

Cybersecurity

I'm a cybersecurity journalist.

Exterior view of the Orlando Airport Marriott hotel. Marriott International has 3,700 properties in 73 countries and territories worldwide. GETTY

U.S. hotel group Marriott has become the second firm to face a massive GDPR fine as the U.K. regulator continues on its rampage. The hotel group, which suffered a breach last year, could face a fine of over £99 million ($123 million). It

THE ONLY WAY TO GOVERN DATA:

Automated Data Governance

How can business replace the old passive approach to data governance? The answer is through what we call automated data governance, which introduces speed, agility and precision into the process of applying rules on data.

Automated data governance is built on five pillars:

1. Any tool

Data science must not be bullied into using specific tools for governance purposes – they're too independent and their technology landscape moves too quickly to build governance efforts around any single tool or application. Instead, data science should be able to use their tool of choice to access the data they need. As a result, ***automated data governance must be able to support any tool a data scientist uses*** – including tools that haven't even been developed yet.

2. Any data

Data comes in all shapes, sizes, and formats, and there's no way to know what type of data a project will need – or even where that data will sit. Some data might be in the cloud, some on premise, and sometimes projects will require a hybrid approach. ***Automated data governance must be able to support ALL data, regardless of where it's stored and regardless of the underlying storage technology.***

3. No copies

A passive approach frequently relies on creating new copies of data, usually with sensitive identifiers removed or obscured. New copies of data are then given to a group of data scientists, who then have access to that data for a specific period of time. Not only does this passive approach create inefficiencies — between determining what data to copy, how to copy it, and where to store the copy — but it creates new volumes of data that become harder to track over time. Yet again, a passive approach to data can't scale. Instead, ***automated data governance requires direct access to the same live data across the organization*** — data must never be copied for governance purposes.

4. Any level of expertise

Privacy and compliance personnel know the rules that must be applied to data, but they don't necessarily know the technology. (Chances are, they can't explain the difference between an Amazon S3 bucket and a Spark cluster, nor should they have to.) Instead, ***automated data governance requires that anyone, with any level of expertise, can understand what rules (e.g., privacy policies) are being applied to enterprise data.*** Data governance must empower both those with technical skill sets and those with privacy and compliance knowledge, so all teams can play a meaningful role controlling how data is used.

5. One policy, in one place

Policies can't live in different formats and in different places. It's all too common for different policies to be expressed in different ways, varying by the database and the underlying storage technology. This causes policy bloat and rules that no one organization, team, or employee will understand, let alone realistically manage. Instead, ***automated data governance requires that data privacy policies live in one central location***, so they can be easily tracked, monitored, and updated over time.

policy bloat

[pol–*uh*–see bloht]

The increasing complexity of policies governing data over time. Bloat causes policies to be harder to manage, and harder to understand, the longer they've been around.

> Without meeting these five pillars, organizations will never break free from the passive approach to data governance—their governance and data science programs will always be at odds, which will stunt their growth, increase risk, and dilute the power of their deepest sources of technical expertise.

While it sounds difficult to combine each requirement into a functional approach to data governance, it can be done. It's why we founded Immuta.

Our Approach to Automated Data Governance

Immuta was founded in 2015 by experts in distributed systems, cryptography, enterprise analytics, and law with decades of combined experience working in the U.S. Intelligence Community, which comprises some of the world's most heavily monitored, secured, and high-stakes data environments.

> When a mistake could compromise a project — or even a life — speed, efficiency, and compliance mean everything.

And yet our team encountered all of the mistakes described above. Passive approaches to data governance are a trap – they're intuitive and easy to fall into, even for the most data-driven organizations. We knew there was a better way, and that's why we built Immuta. We integrated each of the five pillars of automated data governance into our platform.

How Did We Do It?

We started with an approach to data based on a **single point of policy enforcement**, meaning that all policies applied to data exist in one place, in one easy to manage format that even the least technical governance personnel can understand. This approach allows any tool to access any data through Immuta, no matter where or how that data is stored. And it means no new copies of that data are ever required, because Immuta sits between the raw data and the users, ensuring that the right policies are applied to the query results in real time.

Second, we rely on dynamic policy enforcement, which means that policies are applied to data as it's accessed by individual users. As a result, Immuta users are freed from needing to think of entire databases as compliant – one of the central mistakes of a passive approach. Instead, Immuta ensures the right policies are applied to the right data for each query, enforcing the policy on the data in real time. A user acting in one project might have one view of a dataset – sensitive IDs might be completely removed, for example – but that view might change if they switch to another project, which requires different levels of access. This enables Immuta's highly powerful, granular policies to take advantage of context.

PURPOSE CONTROLS:

No Automation Without Context

There's no factor more important in automation than context: in order to apply the right policy without slowing down for manual intervention, rules must take into account who's using what data, for what purpose, and when. That's why purpose restrictions are a central element to Immuta's approach to automated data governance.

With purpose restrictions, governance personnel can create and manage rules based on how and why data is being used, while users must declare the project under which they're using any given data source. Some purposes might require enhanced access and increased monitoring, for example, and some purposes might require diminished levels. Tying rules to context enables more powerful policies, which leads to more powerful automation.

Third, Immuta policies are built through a ***natural language policy builder***, which means that anyone can understand and create policies on their organization's data, all in one place. Our team of legal and technical experts created a policy engine that can accommodate the strictest requirements of nearly any data regulation—without requiring any technical coding skills. If you can write a memo, you can write a rule in Immuta. And that allows for real, lasting data governance efforts. Expertise between privacy, security and compliance science teams is no longer siloed—everyone can use Immuta to understand how data is being accessed and controlled.

Last is the need to prove compliance through ***unified audit logs*** and ***governance reports***. Because Immuta acts as a single access and control layer to all underlying data, audit logs are standardized across storage technologies and projects, making it easy to build reports and track data activities.

A few examples of how organizations are succeeding with automated data governance

Global Financial Institution

A global bank had multiple employees who each require different levels of data access and control, contributing to inefficient and non-compliant data sharing practices.

With the Immuta Automated Data Governance platform, the bank now:

- has fine-grained access control to the data it needs;
- can set project-, purpose-, and role-based restrictions that ensure users can only see the data they are entitled to see;
- access data in hours rather than months;
- is confident that sensitive data won't be accessed by the wrong users or under the wrong circumstances.

---

cognoa

Cognoa is a digital behavioral health company, so data privacy and security concerns are paramount. It needed a way to enforce data access roles, permissions, and policies beyond the standard resource or table based control levels. To meet data demands and compliance requirements, Cognoa enlisted the support of Immuta. The Immuta Automated Data Governance Platform natively applies purpose-based restrictions to Cognoa's data, dynamically enforces in real-time data access and policy restrictions based on the customer's data scientist's needs. Immuta also applies masking techniques to create a view of Cognoa's data wherever sensitive information is included, and automatically anonymizes identifiers presenting compliance issues such as names and birthdates.

Prior to the relationship, Cognoa's data scientists were constantly looking at historical snapshots of data of which the cleansed script could be up to a month old. With Immuta, Cognoa can now dynamically adapt in real time—and when data scientists make a query, it hits the system live. The company is also able to define and enforce detailed data access policies that guaranteed the security and anonymity of sensitive data as required by industry regulations.

> "We have a group of very talented data scientists who build our run-time engine for diagnostics software. Our legacy practice of providing them with all of the data they required to build models, while removing the ePHI and HIPAA sensitive information, was extremely time and labor intensive. It was essential to expedite this process, and also continue to anonymize sensitive information for reporting."
>
> — **Halim Abbas,** Chief AI Officer, Cognoa

LMI

The Immuta Automated Data Governance Platform is an essential component of an integrated data analytics platform designed and managed by LMI for the Office of the Secretary of Defense. Its Maintenance and Availability Data Warehouse (MADW) contains availability, cost, inventory, and transactional data on nearly every Department of Defense (DoD) weapons system and readiness reportable piece of equipment—more than one billion maintenance records from 46 authoritative data systems. The integration of availability, cost, inventory, maintenance, and supply data makes numerous analyses available to leaders across the DoD enterprise.

> “Partnering with Immuta has enabled data scientists to leverage government data in compliance with appropriate governance regulations and privacy safeguards more quickly. Given the size of these data sets, Immuta has been a force multiplier in accelerating the delivery of insight to LMI’s clients.”
>
> — **Joseph Norton, Ph.D.,** Director of Data Visualization and Product Development, LMI

To learn more about how automated data governance can help your organization better embrace data, stay ahead of the rapidly changing regulatory landscape, and increase customer loyalty, email Immuta governance@immuta.com for more detail.

IMMUTA™

7878 Diamondback Dr, Suite C, College Park, MD 20740 | immuta.com | (800) 655-0982