

Best Practices for Intelligent Data Governance

A practical guide to establishing an Enterprise Data Governance program with automation and intelligence

About Informatica

Digital transformation changes expectations: better service, faster delivery, with less cost. Businesses must transform to stay relevant and data holds the answers.

As the world's leader in Enterprise Cloud Data Management, we're prepared to help you intelligently lead—in any sector, category or niche. Informatica provides you with the foresight to become more agile, realize new growth opportunities or create new inventions. With 100% focus on everything data, we offer the versatility needed to succeed.

We invite you to explore all that Informatica has to offer—and unleash the power of data to drive your next intelligent disruption.

Table of Contents

Introduction	4
The Rapidly Changing Role of Data Governance	4
What Is Intelligent Data Governance?	7
What Is the Same After the Revolution?	10
What Do We Need to Do Differently? How to Operationalize Intelligent	
Data Governance	13
What Are the Business Benefits of Intelligent Data Governance?	16
Conclusion	17

"Data in the 21st
Century is like Oil in
the 18th Century: an
immensely, untapped
valuable asset. Like
oil, for those who see
Data's fundamental
value and learn to
extract and use it there
will be huge rewards."

Joris Tonders, Yonogo

Introduction

If you're reading this, you're likely to have already recognized two key things:

- An understanding of data, business, and technical landscapes is critical to the success of your business.
- Building this understanding and keeping it up to date cannot reasonably come from manual activity alone.

You will have also discovered that connecting business and technical metadata across disciplines is paramount and that the true value of data governance lies in the actions you can take relative to the understanding you build. You will be looking to go beyond traditional methods of data governance and begin applying intelligence to the practice, simply in order to keep up with the scale of the data you are managing.

Your competitors are thinking the same thing.

Data governance has evolved significantly in the last few years. It's no longer enough to simply align on semantics or to support manual collaboration and encourage cultural change. Simply doing the bare minimum to comply with regulatory demands just won't cut it anymore. Enterprise-wide Intelligent Data Governance is required to keep you competitive, and this means going a whole lot further than creating a governance committee tasked with documenting definitions. This guide will help you to understand how we got to this point and plot the path that data governance has taken as it has changed from a localized compliance project to a must-do for every organization out there. It will lay out the challenges that face you and how automation and machine learning can help you to overcome them. And although this guide doesn't abandon the lessons of the past, it definitely looks to the future and spells out what organizations need to do differently in order to survive.

Let's take a closer look.

The Rapidly Evolving Role of Data Governance

What do we mean when we say that the role of data governance has begun to evolve rapidly? Well, in many ways, data governance has been evolving since its inception, but the demands we are now putting on the practice have forced it to transform at a much faster pace. The sheer volume and ubiquity of data—along with ever-increasing expectations around agility and time-to-market—mean that your organization will not be able to compete unless you have data governance capabilities in place.



Figure 1 - From Origin to Intelligence Revolution

Origin

Traditional approaches to data governance manifested as a series of controls and restrictions; measures put in place primarily to stop people who might mishandle data and create risk to their organizations as a result. Such efforts were often driven by regulations, focused on complying with specific rules to mitigate risk, rather than generating business value.

The problem with this approach was that it was disconnected from business outcomes, so it was only possible to measure the coverage of the regulation and how busy the data governance operations team was. The work done remained within its silo and did not reach the rest of the company—the day-to-day running of the business was barely affected.

Expansion

An increase in regulatory coverage coincided with a growing recognition of the value of trusted data, which expanded people's view of data governance and what it could be used to achieve. Regulations became broader and began to incorporate more than just technical views, with a new focus on intentions and processes along the way. Simultaneously, the world was waking up to the value that might be exposed if data was truly understood and became an asset to the companies that owned it

The amount of effort required to govern data could no longer be justified purely by a need for regulatory compliance. However, this backdrop provided the opportunity to drive forward the agenda of data value. 'We have to spend the money for compliance, let's make this about our digital transformation'. In order to get there, the practice had to be adapted:

- Enterprise Understanding-it needed to engage the many, not just the few
- Connected Views—it needed to go beyond data to contextualize and show value
- **Driving Outcomes**—it needed to be driven by more than a desire to 'do' governance

Value-Driven Governance, the world has changed



Figure 2 - Value-Driven Governance, the World Has Changed

"Today, more than 5 billion consumers interact with data every day-by 2025, that number will be 6 billion, or 75% of the world's population. In 2025, each connected person will have at least one data interaction every 18 seconds. Many of these interactions are because of the billions of IoT devices connected across the globe, which are expected to create over 90ZB of data in 2025."

 "The Digitization of the World from Edge to Core" David Reinsel, John Gantz, John Rydning¹

Intelligence Revolution

During this expansion phase, it became clear that the manual, human-based data management techniques were no longer going to suffice. It was critical that data governance could scale through automation, accelerating results by maximizing human input. Just as machines began to transform manufacturing practices in the 18th and 19th centuries, artificial intelligence and machine learning have begun to transform the way we think about managing data in the 21st. This period can be seen as our intelligence revolution for data.

- When? NOW! Data has frequently been recognized as the 'new oil' since the explosion of the internet gave it center stage. The recognition of data as such an asset has played a huge part in driving forward the intelligence revolution for data. If you are just starting now, you are already behind the curve and the problem is only going to get worse until you get on top of it. If data that can be trusted is our best asset and that the global datasphere will grow to 175 zettabytes by 2025,¹ then you're going to need to take immediate action, just to keep up.
- Who? Everyone! It is universally accepted that companies need to be data-driven, but it is those that are ahead of the game who are already reaping the benefits. Seven of the top 10 most valuable companies in the world owe their success to data, with Apple as a prime example becoming the first ever \$2 trillion in August 2020.² No one can afford to ignore the call to become data-driven any longer.
- How? As quickly as you can! Catching up with the leading enterprises will take more than just human will and dedication—it will depend on pairing intention with artificial intelligence and machine learning. Increasing scale has limited the amount of impact that manual work can have; as data scales, you will need to employ automation techniques just to keep up. This is not to say that machines can do all the work—the skill lies in finding out just how much machines can complement the work your employees are doing today and continuing to push that boundary.

The Intelligence Revolution

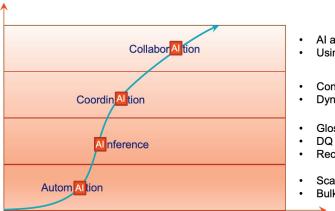


Figure 3 - The Intelligence Revolution

- Al as contributing actorUsing insight > take action
- Connecting People & Understanding
- Dynamic (just-in-time)
- Glossary and Context mapping
- DQ Rule Automation
- Recommendations
- Scanning, profiling, etc.
- Bulk loading, batch-processing

¹ IDC White Paper, sponsored by Seagate, Data Age 2025: The Digitization of the World From Edge to Core, November 2018

² https://www.nbcnews.com/business/business-news/apple-now-worth-2-trillion-making-it-most-valuable-company-n1237287

The intelligence revolution is accelerating through data management, and like everything in the digital world, it is happening much faster than ever before.

- Automation—Automation in today's data governance activity is essential. It takes the simple, manual actions of data governance and enables them to be completed by technology. This scales from low numbers to effectively unlimited volumes of scanning, profiling, indexing and more.
- Inference—Automation creates vast quantities of information in our data governance tools. If automation is helping with the scale of raw materials, inference helps us create meaning and value. It is critical to be able to link those materials together, and enable a coherent, consistent, holistic view.
- Coordination—As ML and Al capabilities increase, organizations are developing a better understanding of their workforce, both in terms of what they contribute and what they need in order to do their jobs. There is a strong move towards dynamically connecting communities and data/information to improve performance, efficiency, and creativity.
- Collaboration—As AI and ML continue to develop, we will grow the capabilities and trust in our software counterparts to the point where they are significant collaborators as well as providing scale and coordination.

What Is Intelligent Data Governance?

Intelligent data governance is the only viable solution to the revolution for data. It is a response to the ever-expanding volumes and types of data that organizations have to manage. The intelligence aspect of it is critical as these volumes and types cannot be effectively managed by manual data governance methods. This section of the white paper will consider three key challenges to any data governance program and how intelligent data governance can help you to overcome them.

Intelligent Data Governance

Data Governance Challenges

Manual Effort

Manual effort can be a killer of DG initiatives, where governance is seen as a burden rather than an enabling force.

Complexity

The complexity of the task at hand can paralyze initiatives or lead to unmanageable levels of customization.

Isolated Efforts

DG is often disconnected from the rest of the company with progress remaining isolated within the governance silo.

Figure 4 - Intelligence Data Governance

Intelligent Solutions

Scalable Automation

Automating from the beginning allows stakeholders to curate and tweak views rather than asking them to do all the heavy lifting.

Consistency

Starting with a robust but agile Data Governance framework allows initiatives to start quickly and adapt easily as they grow.

Extensible Model

Connecting DG to data programs across the organization (e.g. cataloging, quality, privacy etc.) allows for joined up progress to be made. "When people are excited and committed to the vision of data enablement, they're more likely to help ensure that data is high quality and safe."

- McKinsey³

Challenge 1: Manual Effort

The first and most urgent challenge you will come up against when trying to govern the data of a modern enterprise is the simple scale at which you must operate and the resistance of many of your workers to do so. Traditionally, data governance has been seen as a very manual undertaking, whereby individuals from across the company have to come together to pool their collective knowledge and ideas, often getting little in return.

Some companies have more than 50 million records. It would take YEARS to even count to 50 million, let alone document the records manually. Automation is the only way to handle such large volumes of data.

Solution: Scalable Automation

The manual effort challenge is most often overcome when individuals can see what they are gaining from becoming part of the data governance effort. If some of the work can be automated so that they are getting out a lot more than they put in, people will be willing to contribute. Therefore, the key to overcoming this challenge is not to ask too much of your employees' time to do the most menial of work, but only the work that really and truly cannot be automated.

For instance, if a company has an employee who is a subject matter expert on a particular system, the traditional approach to data governance would involve them documenting all they knew about every field within that system to try to create a holistic data dictionary. The intelligent way on the other hand would have the machine do the heavy lifting first. There is a lot that is common to more than just your organization, e.g., almost all companies hold some kind of personal datanames, addresses, telephones numbers, credit card details, why not begin by having the machine identify these for you? Common patterns and repeated data types along with name matching can account for a lot of information before the SME even has to be involved—and when they are, they will be curating and fine tuning rather than starting from scratch, a much more appealing prospect. This approach saves hundreds of man hours and makes that SME much more likely to remain engaged and supportive of the data governance initiative.

Challenge 2: Complexity

The second major challenge to data governance is the sheer complexity of the undertaking as we seek to govern more data types than ever before and use that knowledge across multiple disciplines for a variety of business outcomes. Some traditional approaches to data governance involve having a totally open model for documenting asset types and their relationships to one another. This can seem appealing in theory as the customizations applied will allow you to document any new data types you need to consider. However, practical implementations have shown us that this approach creates far more problems than it solves and builds a level of complexity into data governance that makes it unmanageable.

³ https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/designing-data-governance-that-delivers-value

Solution: Consistency

Rather than endlessly adding new categories to an ever-expanding framework, try to operate within a more robust overview, using types and relationships to get the specificity you need without reinventing the wheel constantly. The building blocks of any business are relatively consistent regardless of industry or use case—glossary terms, systems, policies, processes, etc. It is how you break these down and connect them up that will give you the precise overview of how your organization works.

If Automation and Intelligence are important to you-consistency has to be a priority.

Some data governance frameworks have unlimited customization capabilities, encouraging each customer to build what is 'right' for them – but how can they innovate and automate for their customers, when the foundation is different in every instance?

The more consistency you can keep, the more you will be able to leverage intelligence, automation and machine learning. The complexity of data management can easily become overwhelming if we allow it to, so rather than creating a new asset type for every new idea, look for commonalities over differences

- Is a report totally different in fundamentals to another collection of columns?
- Is a Personal Data Category really more than a glossary domain grouping together terms?
- Is a Data Controller/Processor really more than a legal entity with a particular relationship to the personal data within a process?
- Could what some describe as a business outcome dependent on data assets, similarly be seen as a capability enabled by those assets?
- Are APIs so different to interfaces between systems that they could not be considered a type of the same?

Challenge 3: Isolated Efforts

The third major challenge facing data governance initiatives today is their isolation from the broader success of the organization. As discussed in 'the beginning', data governance was far more about creating artefacts than adding value. It therefore got a reputation as being the remit of a particular team, who were more likely to pester you for a definition than to give you anything that would help get your day-to-day work done. As the needs around governance have changed and the reach of it has far expanded, many companies are rebranding it as something bigger, e.g.:

- Data Excellence
- Data Intelligence
- Data Strategy
- Data Empowerment

This change must be more than just a change of name, connected, integrated views of different disciplines are required to actualize the business goals governance links to.

Solution: Extensible Model

So how to combat this view of data governance? Well changing the name isn't a bad idea but changing the impact is a better one. Data governance (or Data Excellence or Data Empowerment or whatever you would prefer to call it) should be about executing on enterprise understanding, instead of just exposing it and it should provide your business leaders with information that they can trust. If they can trust the data, then they can trust the analysis and decisions made using that data. The ultimate goal should not be the ability to update definitions periodically but to affect real change, to bring together different disciplines to enable your organization to execute Enterprise Data Management.

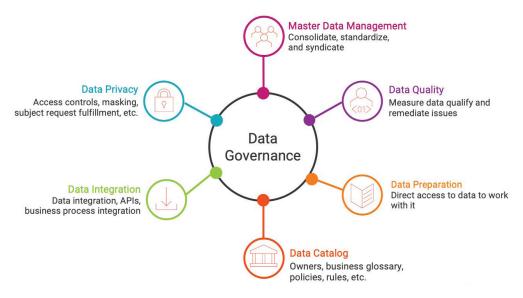


Figure 5 - Intelligence Data Governance

What Is the Same After the Revolution?

A lot has changed since the Intelligence Revolution for data. The way we run things and operationalize processes is of course affected, because we are now able to farm out so much of the most mundane work to machines. However, there is a lot that has also remained the same, such as the basic principles and framework that we build our governance programs around. Let's have a look at what has stayed consistent before we dive into the differences.

The Basic Principles of Data Governance Are the Same

• Engagement & Adoption Above All Else

Any data governance program ultimately succeeds or fails based on whether or not people adopt its practices. You can build the greatest documentation system of all time, but if people are not using it, then it is automatically useless. Furthermore, not everything can be automated, so if you do not have regular human contributors, your documentation will not stay 'the best' for long. This has not changed with the intelligent data revolution because human adoption and use is still critical.

· Think Big, Start Small

Data governance is, as we have discussed, a big undertaking, but that does not mean you have to do it all at once. It makes sense to tackle a project or specific area and prove value in order to drive the mission forward. However, while starting small, it is necessary to keep your end goals in mind. Just because you are focusing on a particular project to start with does not mean the stakeholders of that project should run riot—if the end game is an enterprise solution, it's wise to keep things broad enough to appeal to more than the immediate group.

· Be Guided by Practical Usage

There is no need to make all the decisions before you get going on your governance journey. 'Analysis paralysis' can hold you back for months; if you are set on making sure things are perfect before 'starting', you may never start. Allow people to build up a knowledge graph together, enriching only that which needs to be enriched instead of governing everything for the sake of it! If something is wrong and being used you will hear about it, if something is wrong and not being used how much does it matter? Allow practicalities to guide priorities.

· Breadth Over Depth

While you do not need to do everything on day 1, it is a good idea to build some broad structure into your efforts. If you begin with creating some outlines, people who specialize in those areas can fill in the blanks if and when they need to. It helps to give people some structure and something to enrich or collaborate on, rather than going to the nth degree of detail from the get-go.

Do Not Model

Data Modelling is a useful activity (in its area) but trying to create a strict data model that will neatly fit everything for an entire enterprise is a losing battle. Be flexible in your governance efforts, allow people to document things in the way that makes sense and is understandable to 99% of your audience. Do not let strict rules get in the way of practical progress.

Driving Cultural Change Is Still Critical

As engagement and adoption of data governance is still a primary concern, driving cultural change within your organization is still a critical activity. In some ways this should be easier than ever with the support of automation, as it is no longer necessary to ask contributors to painstakingly document everything they know.

Instead, you are just asking people to enrich what has been discovered by technology with the information that is not stored on any database—the day-to-day knowledge that lives within their minds. To do this, you need to embed governance practices into their daily lives, driving cultural change throughout the organization in 3 steps:

1. Mind: Getting Support

- · Identify engagement drivers/blockers
- · Appeal to people across the spectrum
- · Offer incentives for a change in behavior
- Meet the challenges head on

2. Body: Making the Change

- · Micro changes breed macro changes
- · Layer repetition until changes becomes the norm
- · Start simply; challenging activities can come later
- · Support initiatives then lead initiatives; become self-perpetuating

3. Soul: Making it Stick

- Maintain faith through the inevitable dip
- · Socialize broadly so that progress is visible
- Embed into users' everyday practice with continued incentives
- · Build communities to build resilience
- · Revisit and revalidate for lasting sponsorship

☐ Which part of the business is it relevant to?

The Questions are the Same

25 Things you STILL want to know about your data:				
	What does it mean?		What end-goals does it enable?	
	How is it structured?		Who has access to it in the company?	
	Where can I find it?		Who has access to it outside the company?	
	Why do I have it?		Where was it collected?	
	Who is responsible for it?		Is it suitably protected?	
	How should it be used?		Does keeping it pose any compliance risk to	
	How is it being used?		the company?	
	How has it been used in the past?		What regulations is it subject to?	
	Has it been transformed since its		What key processes depend upon it?	
	creation point?		What would be the downstream impact of	
	How can I have it changed/updated?		changing it?	
	Who is overseeing it?		Who would need to be informed before	
	Is it of a good quality?	_	changing it?	
	How is it moving between locations?	Ш	Is storing it adding value to my business?	

As you can see, the goals are very much the same. We are still all looking to create enterprise-wide understanding and trust of our data landscape and manage how that maps to our business. The thing that has changed however, is how we get there. There are many automated techniques we can use now to ease the burden of governance efforts and the next chapter will consider how best to exploit these, minimizing the work that has to be done by your data stewards and maximizing the amount of time they have to spend on infinitely more valuable work.

The Governance Framework Is Consistent

The final thing to consider that has not changed is the building blocks which we use to govern data in the first place. Just because the way we gather the information has transformed in some areas does not mean that the information we are gathering is not consistent.



Figure 6 - Governance Framework

What Do We Need to Do Differently? How to Operationalize Intelligent Data Governance

Automated Tagging of Metadata

Before: The Manual Way

- Who did the work? System subject matter experts
- How? Trawling through databases, viewing column titles and the data within them to classify what kind of data a specific column held
- Why was this an issue? This classification of data took a great deal of time and as SMEs could not stop doing their day jobs it often went on in the background, taking years. It was also very dull repetitive work that was likely to disengage the SMEs.

After: The Automated Way

- Who does the work? The bulk of it is done by the algorithms in the technology, with some curation and specialist input from the SMEs as needed
- How? Intelligent governance solutions use name matching, predefined rules and similarity propagation techniques to auto-tag scanned data to business definitions
- Why is this transformative? A great deal of matching can be done based on standard rules without manual input. SME curation will enrich that further, taking up a fraction of the time they would have spent and yielding richer results.

How will this matching work?

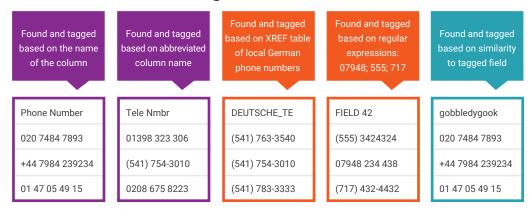


Figure 7 - Matching

Automated Application of Quality

Before: The Manual Way

- · Who did the work? Data quality technicians
- · How? Creating a new rule for every single DQ check that needs measuring
- Why was this an issue? Even if a company only wanted to check 1000 data points, with 5 rules on each, that is still creating 5000 rules. In reality, these numbers are MUCH higher, and the manual effort involved is unmanageable.

After: The Automated Way

- Who does the work? The business decides what they want to measure at scale.
- How? Intelligent quality solutions use the aforementioned tagged data to apply the appropriate
 quality checks everywhere a certain data concept is found.
- Why is this transformative? If the effort to do baseline quality is next to nothing, you can create an amazing foundation for understanding how good your data is. Removing the upfront ask of manual effort from governance makes initiatives far more likely to succeed, as stakeholders have something significant to gain when they engage.

Track KDEs at Scale

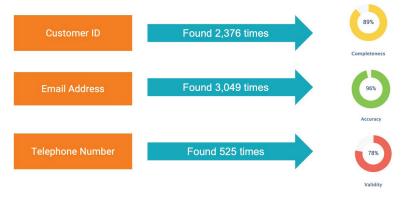


Figure 8 - Track KDEs at Scale

Automated Change Notifications

Before: The Manual Way

- Who did the work? System stewards (if anyone)
- How? Comms to downstream consumers when a field was added or changed
- Why was this an issue? Changes made locally often lacked any documentation updates for downstream systems. If documentation updates were made, this meant manual updates for all lineage diagrams.

After: The Automated Way

- Who does the work? Intelligent scanning tools
- How? Intelligent scanners do not just run once and assume that the data landscape will
 remain stagnant, they run periodically and notify the appropriate stakeholders of the delta
 between scans.
- Why is this transformative? The governance activities related to columns being added or changed can be triggered without manual intervention, meaning the current state view of lineage does not become outdated/inaccurate over time.

Automated Triggering of Workflows

Before: The Manual Way

- · Who did the work? Data stewards
- How? Monitoring data governance issues and manually flagging issues
- Why was this an issue? Data stewards would have to manually review all the artefacts
 documented on a periodic basis to make sure they were up to scratch. This created a ton of
 unnecessary busy work for limited value.

After: The Automated Way

- Who does the work? Intelligent governance tools will trigger workflows automatically as required.
- **How?** You will decide what kind of changes require/do not require reviews and apply automation going forward.
- Why is this transformative? Data Stewards no longer have to trawl through artefacts or need
 to have unnecessary conversations with stakeholders about whether a definition has changed.
 The power now sits with the stakeholders, getting them more engaged as contributors but also
 ensuring that appropriate validations and controls are still met.

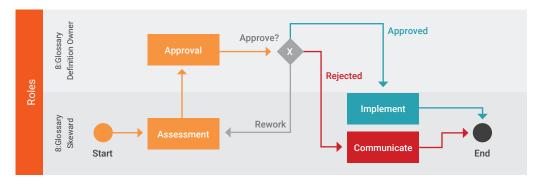


Figure 9 - Roles

Automated Provisioning of Data

Before: The Manual Way

- · Who did the work? Data Technicians
- How? For every data ad hoc data access request, a technical resource would have to gather the
 data and provision it to the requestor
- Why was this an issue? The sheer volume of data access requests has risen so much in recent
 years that this became unmanageable. It also became almost impossible to enforce good
 governance practices in a sea of ad hoc requests.

After: The Automated Way

- Who does the work? Data Owners
- How? They can choose to publish their best data assets for consumption and have the
 provisioning of this data automated in the right circumstances.
- Why is this transformative? Data access requests can now be properly centralized and tracked,
 while certain elements of their delivery can be automated. Due to security restrictions, this is
 not appropriate in all circumstances, but if the right conditions are met this process can be
 managed much more smoothly, freeing up data scientists to spend less time looking for data
 and more time using it.

What Are the Business Benefits of Intelligent Data Governance?

This paper has argued that no organization can afford to ignore the need for intelligent data governance. We have seen how the practice has developed at an increasingly rapid pace and while certain principles related to governance have survived, the practicalities of how we go about it have fundamentally shifted, due to the scope of what governance now addresses. However, it should not be seen as a stick, or a threat of something you must do, but instead an incredible opportunity which brings with it a powerful range of business benefits.

The **scalable, automated** approach to data governance allows you to **keep up** with the rate of change in the organization while keeping the **power in the hands of the business**. Rather than struggling to document things that have already happened and forever playing catch up, Informatica[®] automation allows you to represent the current state view in real time, from holistic views of data lineage to holistic views of data quality.

The **consistent** approach to data governance allows you to continue to **innovate and automate** over time. Any organization with a standardized framework can learn from their customers and continue to enhance their product because of the consistency of design and data gathered in relation to usage. Informatica can do the same, as we rely on an agile but robust data governance framework rather than encouraging endless customization, upon which you cannot automate.

The **extensible** approach to data governance allows you to create a **common centralized view** of your data, business, and technological landscapes, which can then be **deepened as necessary and when needed.** Informatica's platform goes far beyond Governance concerns, allowing you to first understand and then act upon what you have discovered, taking real action from improving the quality and privacy of your data to mastering it so that you can ultimately and consistently trust your data.

Conclusion

Intelligent Data Governance isn't something you can do alone. It involves bringing together a lot of capabilities and truly understanding how these capabilities work together. To achieve the level of automation and connectivity we have discussed here, you need to be looking at more than a point solution. Data governance depends on data cataloging, data quality and data privacy and while there is no one product that can solve all these problems for you, there is a platform that can support and empower teams across your organization—from the executive suite to marketing to manufacturing—to drive productivity, efficiency, and the effective use of data.

