How To Rule Your Data World

Authors:

Jacqueline Bloemen
Timm Grosser

BARC

A TEKNOWLGY GROUP COMPANY

MarkLogic®

Inhalt

How To Rule Your Data World

Jacqueline Bloemen, Senior Analyst BI

Timm Grosser, Head of Consulting BI & Senior Analyst BI

November 2018

RESEARCH STUDY

A new core discipline for data-driven companies

Data is everywhere. With the growing interconnectedness of people, companies and devices, we are now accumulating increasing amounts of data from an ever-widening variety of channels. New data (or combinations of data) enable innovative use cases and assist in optimizing internal processes. As a result, data is set to have a significant effect on tomorrow's business and will undoubtedly become a value-adding factor.

However, effectively using data needs to be learned. Companies using multiple systems for various purposes tend to have complex technology landscapes with highly specialized silo solutions. From operational systems to support „smart processes", to the data warehouse for enterprise management, to exploring new use cases through advanced analytics, all of these environments incorporate disparate systems, each containing data fragments optimized for their own specific task. Data and data processes are everywhere in the organization so there is a growing need for a comprehensive view of business objects and data. It is therefore vital that data is subject to some form of overarching control, which should be guided by a data strategy. This is where data governance comes in.

*"Data Governance refers to the **individuals, processes and technology** required to **manage and protect enterprise data assets**. Its goal is to ensure **interpretability, correctness, completeness, trustworthiness, security, accessibility and traceability** of enterprise data in an efficient and effective manner."*

Data governance is a widely discussed trend at the moment. After all, it is important to network the data competencies in any company in order to be able to implement new types of application and to fulfill compliance requirements such as GDPR. Best practices are still all too rare. There is a lack of available knowledge and orientation at the current time.

This study examines the current status and actions of companies that have implemented or are planning to implement data governance initiatives. It reveals interesting insights into a discipline that will become indispensable in the future for anyone wanting to work seriously with data.

Enjoy reading!

Jacqueline Bloemen and Timm Grosser

Management Summary

Data governance is a trending topic predominantly focused on BI and data warehousing, and mainly driven by compliance

Hot Spot 1

Only 4 percent of participants regard data strategy and data governance as an inconceivable approach for their business now and in the future. A majority focus their current governance activities on BI and the data warehouse. Best-in-class companies have realized that it is important to cover all data environments and establish a feedback loop from data usage in BI and analytics to drive operational data improvements. While compliance is the major driver for data governance, it bears the risk of reducing it to a very restrictive procedure.

Data quality is the top challenge when it comes to using data, closely followed by multi-faceted organizational issues

Hot Spot 2

Inadequate data quality remains the foremost challenge users face when using data. This has been shown repeatedly by market research and data-centered projects for many years.[1] Yet it appears that the reasons for the apparent inability of businesses to cope with this problem in order to achieve continuous improvement are largely organizational. In spite of pressing data quality issues, there is a lack of acceptance and priority for data governance at executive level and in lines of business.

Practitioners and planners have different views on how to approach data governance

Hot Spot 3

Best practices in data governance are still rare. However, there is widespread agreement on one point: technology is not the limiting factor. Businesses currently planning their data governance endeavor tend to focus on administrative tasks: they favor the development of a data catalog as their top-rated measure, followed by roles and processes. Practitioners concentrate more on practical execution, such as data quality monitoring and training. This way of working helps generate business demand. Challenging and developing users in data governance issues is a promising approach as it addresses the most widely identified challenges today and in the future.

---

1   i.e., BARC Survey "Advanced & Predictive Analytics 2017"

The most striking benefit of data governance is the effective use of data based on a unified understanding

Hot Spot 4

Targeted data governance creates strategic and operational added value. It is a mechanism to become more effective and efficient in the use of data. Creating a unified understanding of data can raise this effectiveness to a higher, overall strategic enterprise level and help a company along its path to digitalization. However, if these actions are predominantly targeted at the data warehouse, their achievements will be of limited usefulness. Business value is ultimately generated in core business processes.

Data governance success depends heavily on adequate resources which cannot be provided by a central unit alone

Hot Spot 5

There is clear agreement that data governance can only be achieved by interdisciplinary teams with strong participation from the line of business. At the same time, the number one challenge is a lack of resources. Tasks are primarily delegated to business analysts and key users who already own other line duties. It appears that efforts, for example for data stewardship, are not sufficiently acknowledged by executive and divisional management. Due to the siloed history most enterprises share, convincing these stakeholders of the need for – and value of – a holistic data strategy is presumably the most difficult challenge they face. Data strategy and governance must be closely aligned with the enterprise and digitalization strategy, and with an overarching view of business processes.

Technology is not a limiting factor for implementing governance

Hot Spot 6

Technology is not a limiting factor for implementing governance. Compared to other challenges, 'lack of tool support or wrong tools in use' is of relatively minor significance to our panel. Unsurprisingly, data cleansing (58 percent), data quality management (53 percent) and data integration (52 percent) are identified as the most needed functions to support data governance. Due to the complexity of the enterprise landscape and the often limited availability of human resources, companies need to find innovative ways to boost efficiency in governance tasks. Awareness is growing that machine learning and AI can play a central role in this area.

Data governance, a trending topic

Data governance is indeed a trending topic. Our survey starts out with an evaluation of its current relevance in the market. Only 4 percent of participants were not convinced that data governance will become relevant in their company.
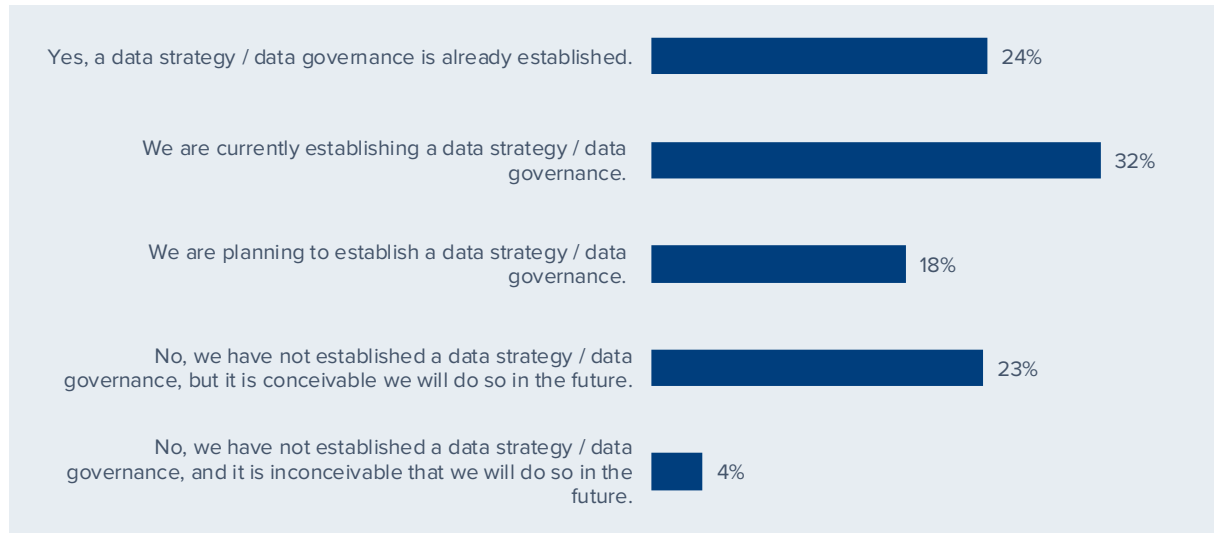
Figure 1: Do you have data governance / a data strategy established within your business or do you plan to do so? (n=364)

Our analysis shows North America as the leading region when it comes to established (30 percent) or planned (50 percent) data governance, with a total of 80 percent. Only 3 percent currently see the topic as irrelevant, and 16 percent consider it viable in the future.

Europe is catching up. 21 percent of participants state that they have data governance in place and 48 percent are in the planning phase. The share of participants currently not addressing governance is 31 percent. This is an interesting outcome in the context of the European General Data Protection Regulation (GDPR), which has been in force since 25th May 2018. However other regulatory requirements (such as BASEL II/III and Solvency II) become apparent when looking, for example, at the banking, finance and insurance industry. Here, the proportion of businesses with data governance already in place is relatively high at 39 percent.

It seems to be obvious that the percentage of companies that have established data governance is relatively high among the best-in-class group (87 percent), whereas only 13 percent of laggards have adopted such an approach.

> Europe is catching up with North America, with 48% planning data governance.

Data governance drivers

As the adoption of data governance varies throughout the market, it is interesting to look at the drivers for data governance initiatives. This should help us to understand to what extent data governance can be a relevant approach for companies that have not yet established a data strategy.

Driver	Percentage
Compliance with legal requirements	56%
Handling data more efficiently	51%
Increasing number of analytics use cases	50%
Growing amount of business-relevant internal / external data	50%
Need for functional harmonization of data used	44%
Higher expectations within lines of business	43%
Gaining an overview of / information on data available	38%
Transformation towards a digital business	32%
The increasing amount of data processing technology needs to be managed more effectively to prevent uncontrolled growth	30%
The need to remain competitive	28%
Other	2%
No drivers	2%

Figure 2: What is driving the application of data governance in your company? (n=351)

Compliance with legal requirements is the most frequently cited driver for data governance, although this varies somewhat between regions (Europe: 64 percent, North America: 48 percent, Asia-Pacific: 30 percent). In contrast to the finding in the previous chapter, the high relevance of compliance as a driver for data governance in Europe may be an indicator that GDPR does indeed drive adoption.

However, setting up data governance exclusively to comply with regulations reduces it to a very restrictive procedure. It underestimates the value data governance can have with regards to facilitating data quality. Data quality in turn is the prerequisite for utilizing data and analytics to foster innovation. A well-defined data governance process can help make the perspectives of data collectors and data consumers more compatible and contribute to better overall data

Setting up data governance exclusively to comply with regulations reduces it to a very restrictive procedure.

quality. Businesses frequently underestimate the true potential of data governance beyond regulation and compliance.

Handling data more efficiently, the second highest rated driver, is an indicator that working with data has become a central and highly prioritized task in the era of digitalization. This is pointed out by 51 percent of participants, which can be connected to the increasing number of analytics use cases (50 percent) and the growing amount of internal and external data (50 percent).

While the need for functional harmonization (44 percent) and higher expectations within the line of business (43 percent) do not lead the list of drivers, their shares are significant enough to conclude that the data pools currently available frequently fail to meet business needs. In general, there appear to be regional differences in governance drivers. Our analysis shows that Europe, Asia-Pacific and North America have varying priorities:

	Europe	Asia and Pacific	North America	Total
Compliance with legal requirements	64%	30%	48%	56%
Handling data more efficiently	54%	23%	58%	51%
Increasing number of analytics use cases	45%	53%	58%	50%
Growing amount of business-relevant internal / external data	54%	40%	46%	50%
Need for functional harmonization of data used	46%	30%	46%	44%
Higher expectations within lines of business	47%	33%	38%	43%
Gaining an overview of / information on data available	38%	27%	47%	38%
Transformation towards a digital business	35%	27%	28%	32%
The increasing amount of data processing technology needs to be managed more effectively	29%	17%	34%	30%
The need to remain competitive	28%	37%	26%	28%
No drivers	0%	13%	1%	2%
Other	2%	3%	3%	2%

Figure 3: Data governance drivers by region (top four drivers per region marked blue) (n=351)

Best-in-class companies are more driven by the opportunity to analyze internal and external data (63 percent) than laggards (41 percent). This is also true with regard to an increasing number of analytics use cases, which indicates that best-in-class companies have reached a higher state of awareness of the importance of data and analytics as drivers for business efficiency and innovation.

Best-in-class companies have reached a higher state of awareness of the importance of data and analytics as drivers.

Is limiting data governance to specific environments common practice?

To obtain a better understanding of how data governance is actually carried out, participants were asked which data environments are actually affected by governance measures.
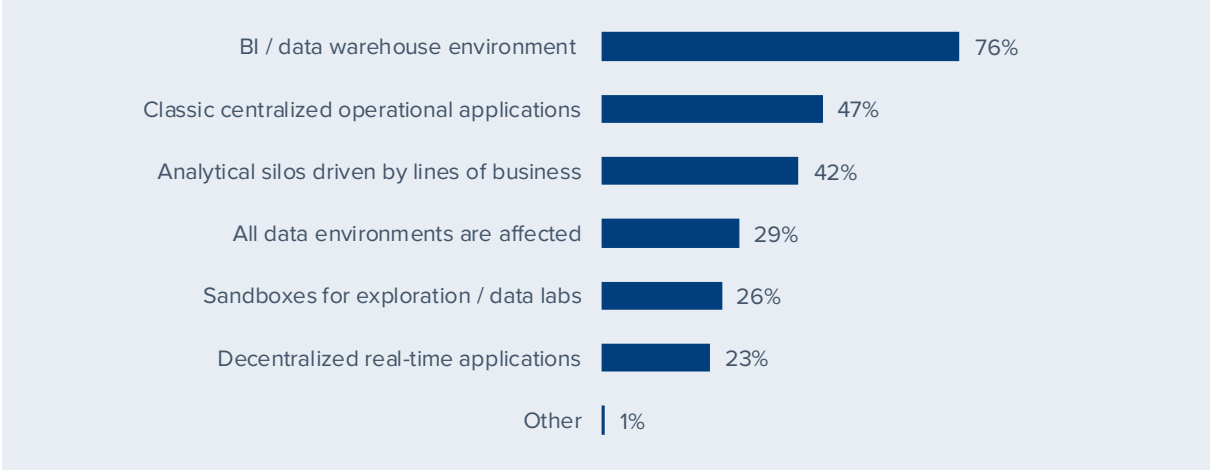

Figure 4: Which data environments are (or would be) affected by your data governance? (n=348)

The majority (76 percent) of participants see their BI and data warehouse environment affected by data governance. This is not surprising, as the enterprise data warehouse has for decades been defined as the "single point of truth" in a landscape of siloed operational applications and data. The data warehouse has been seen as the only feasible vehicle to establish a harmonized view of enterprise data and is therefore the first natural target for organized governance of data and how data is used.

However, this approach has clear limitations. The data warehouse depends on data delivered by the original source systems, where the data is actually generated. The state of this source data is mostly inadequate for further data consumption, both from a quality and a structure point of view. This has led to immense efforts when transferring data to the warehouse and has made its development and maintenance very costly. Data governance can therefore only be fully effective when applied to source systems. 47 percent of participants appear to have realized this and apply data governance to classic, centralized operational applications.

According to the analysis, sandboxes and data labs are largely unaffected by governance. Only 26 percent include these environments in their initiatives. On the other hand, an increasing number of analytics use cases is the third most popular driver for data governance. This might indicate that the intensive utilization of data for analytics is creating a feedback loop between analytics projects and data governance initiatives, even though the data lab environment itself may not be the target of operational governance activities. This

Data governance can only be fully effective when applied to source systems. 47 percent of participants appear to have realized this.

Establishing a feedback loop between analytics projects and data governance initiatives may help boost the overall usability of data.

would make sense as only the actual intensive and creative use of fine granular data seen in analytics can make clear which requirements of source data need to be considered in, for example, central operational applications.
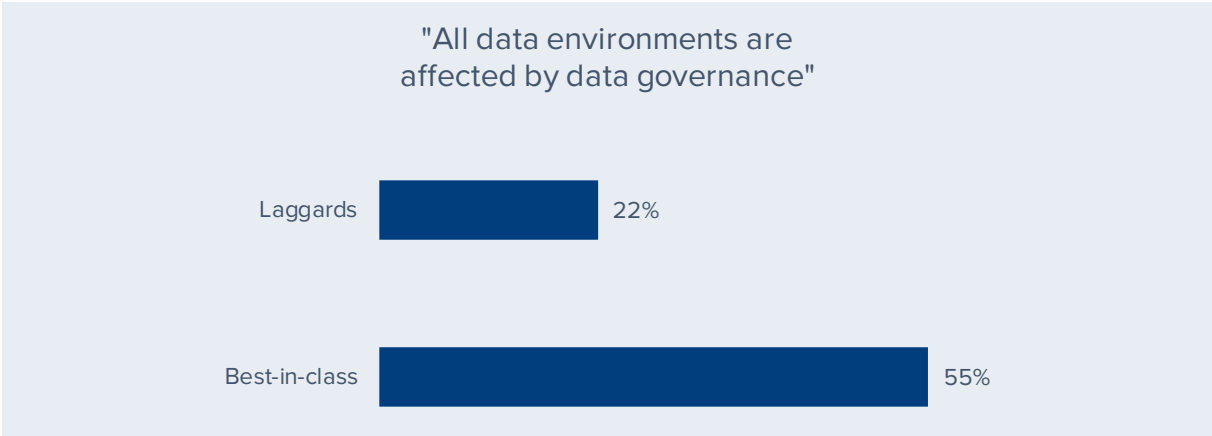
Figure 5: Best-in-class versus laggards seeing all data environments affected by data governance (n=102)

In a digitalized and data-driven world, data is an asset that is needed throughout the value creation chain in varying ways. Therefore, companies will ultimately need to embrace all data environments in their data governance initiatives. Overall, only 29 percent of participants have taken on this challenge. For those who have already established data governance, the share lies at 43 percent. Among best-in-class companies, the share is even higher at 55 percent. It is also remarkable that 42 percent of respondents from the public sector focus their governance on all data environments.

Companies will ultimately need to embrace all data environments in their data governance initiatives.

Current challenges when using data

We asked participants to reflect on their main data challenges today and in the future. It seems there is still a long way to go. Data quality is perceived to be the top challenge and will only be surpassed by the lack of skills to cope with the increasing complexity of systems and data in the future.

Challenge	Today	In the future
Inadequate data quality	48%	36%
Ineffective organization	42%	34%
Insufficient cooperation	38%	30%
Lack of information on data	38%	30%
Centralized and/or decentralized, unsynchronized data silos	34%	28%
Relevant data cannot be found / lack of knowledge of existing data	34%	20%
Lack of agility in terms of data sourcing / data provision	32%	26%
Data availability / lack of data access	31%	20%
Lack of skills to cope with increasing complexity of systems and data landscapes	29%	38%
Lack of suitable tools	22%	17%
Corporate policy / decisions	22%	26%
Other	3%	1%

Figure 6: What are your company's biggest challenges when it comes to using data? (n=348)

Several organizational issues appear to be relevant here. Ineffective organization and insufficient cooperation are the second and third highest ranked challenges respectively. Laggards in particular appear to struggle with organizational issues: 64 percent complain of ineffective organization (best-in-class: 21 percent) and 53 percent perceive missing collaboration (best-in-class: 21 percent) among their main challenges. From a wider perspective, this indicates that fixing organizational issues can help in addressing some of the other issues listed in Figure 6. Best-in-class companies struggle less with finding relevant data than laggards do – only 18 percent of best-in-class companies have an issue with this, compared to 50 percent of laggards. Additionally, the number one challenge – data quality – seems to apply to best-in-class companies to a much lesser degree (21 percent) than laggards (53 percent).

Laggards in particular appear to struggle with organizational issues. This indicates that fixing these can be a help when addressing other issues.

Regional analysis also reveals some differences. Data quality appears to be a larger issue in Europe (56 percent) and Asia-Pacific (53 percent) than in North America (36 percent). Asia-Pacific struggles much more with a lack of suitable tools (37 percent) than the survey sample as a whole.

Surprisingly, data silos and finding data was only rated as a challenge by 34 percent. It would appear reasonable that these issues might bear the same rating as the topics are correlated. However, the outcome is different from a regional perspective. Both Asia-Pacific (53 percent) and Europe (40 percent) see silos as a more significant issue than North Americans do (36 percent). In terms of finding data, North America (41 percent) and Asia-Pacific (40 percent) are almost level, whereas Europe lies at 29 percent. In general, this does not match our experience in projects. Based on the numerous interviews we conduct, we would have expected the shares to be higher.

Besides the lack of skills, corporate policy is also expected to grow in relevance in the future. However, participants expect all the other challenges to become less important in the future. This applies especially to data quality and finding relevant data, which each show a decrease of 14 percent. This indicates that companies feel confident these challenges can be tackled within a reasonable time period. Considering the significance of the problem and the complexity of the data landscape of most businesses, this could prove to be an optimistic assumption.

Companies are confident they will be able to solve the issues of data quality and finding data in the future.

Measures to develop data governance

There are a number of challenges to tackle with regards to data governance. However, implementing data governance can in itself involve some difficult tasks. We asked participants to share which measures they are pursuing to deploy governance in the enterprise. In order to obtain a more detailed view of the measures taken, we also distinguish between those participants who already have data governance in place (practitioners) and those who are currently implementing or planning to implement data governance (planners).

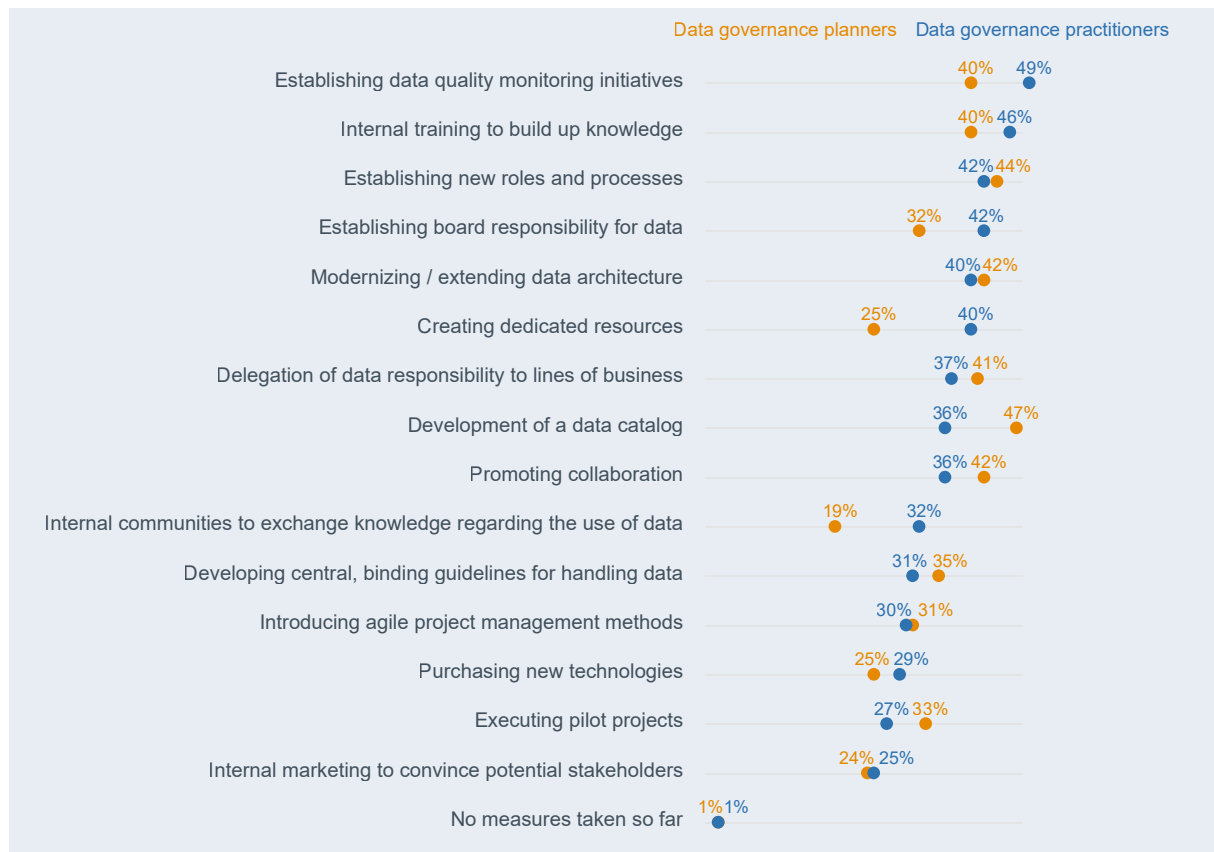| Measure	Data governance planners	Data governance practitioners
Establishing data quality monitoring initiatives	40%	49%
Internal training to build up knowledge	40%	46%
Establishing new roles and processes	42%	44%
Establishing board responsibility for data	32%	42%
Modernizing / extending data architecture	40%	42%
Creating dedicated resources	25%	40%
Delegation of data responsibility to lines of business	37%	41%
Development of a data catalog	36%	47%
Promoting collaboration	36%	42%
Internal communities to exchange knowledge regarding the use of data	19%	32%
Developing central, binding guidelines for handling data	31%	35%
Introducing agile project management methods	30%	31%
Purchasing new technologies	25%	29%
Executing pilot projects	27%	33%
Internal marketing to convince potential stakeholders	24%	25%
No measures taken so far	1%	1%

Figure 7: Which measures have you taken (or do you plan to take) to develop your data governance?
(n-practitioners=84; n-planners=179)

Developing a data catalog, establishing new roles and processes, and data quality monitoring initiatives are the most frequently named measures participants rely on to drive data governance (43 percent on average each). Internal marketing and promoting internal communities are the lowest prioritized measures.

Companies that have already established data governance have a different focus to those who are still planning to do so. Practitioners focus mostly on data quality initiatives (49 percent), which is further underpinned by internal training (46 percent) and new roles to actually do the job (42 percent). Interestingly, planners find monitoring far less important, even though inadequate data quality is rated as

> Internal marketing and promoting internal communities are the lowest prioritized measures. This is surprising. Convincing stakeholders and motivating departments should be one of the core tasks.

the most significant challenge when using data. However there is a general consensus between practitioners and planners on organizational measures.

Developing a data catalog is only a priority for 36 percent of practitioners, perhaps indicating that in practice this measure requires some groundwork in terms of management support, skills and dedicated resources to be successful. From a regional perspective, North America (44 percent) appears to be more focused on lines of business taking an active role than Europe (40 percent) in terms of delegation of data responsibility and promoting collaboration. In general, enterprises with over 10,000 employees have the strongest focus on establishing new roles and processes (54 percent).

It is notable that although a lack of suitable tools when using data is not regarded as a major challenge today (22 percent) or in the future (17 percent), 26 percent of participants do see a need to purchase new technologies for data governance.

North America appears to be more focused than Europe on lines of business taking an active role.

The benefits of data governance are undisputed

Those who have already established data governance were asked to share their experiences concerning the outcome of their initiatives so far.

Enhanced decision-making support through better, more effective use of data — 53%

Unified understanding of data — 53%

Creating conditions for data-driven work / becoming a digital company — 47%

Creating a harmonized view of all relevant data — 45%

Reducing effort with more efficient processes — 42%

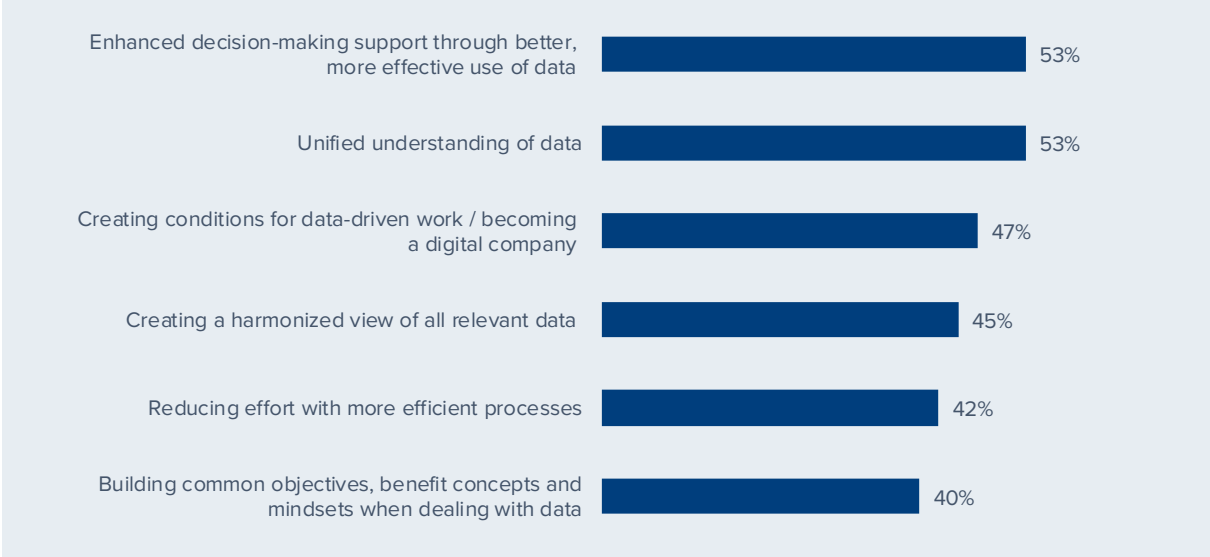Building common objectives, benefit concepts and mindsets when dealing with data — 40%

Figure 8: Top six benefits achieved by implementing measures to develop data governance (n=83)

A majority (53 percent) say they have enhanced their decision-making and accomplished a unified understanding of their data. Governance measures have also helped to create the conditions for data-driven work and becoming a digital company (47 percent).

It might seem contradictory that 45 percent claim to have created a harmonized view of all relevant data while only 40 percent have achieved common objectives when dealing with data. In particular, North America (52 percent) and very large enterprises (63 percent) seem to have successfully implemented measures in this area, in contrast to only 36 percent of respondents from Europe.

Creating a harmonized view and building common objectives on enterprise data are high value benefits that businesses have been pursuing for countless years with limited success to date. With regards to the panel of this survey, we recommend interpreting their views with caution. The majority of our participants are BI experts (53 percent) and as such are predominantly focused on business intelligence and data warehouse projects. Most companies have been able to achieve a decent level of data harmonization in the data warehouse, which has become a common practice. However, provisioning such a view of data for business intelligence applications does not necessarily mean that it has become the common ground for the frequently siloed lines of business in everyday work. As the data warehouse is in principle a silo of its own that is separate from other business systems and desktop applications, it does not guarantee the building of overarching common mindsets when dealing with data.

Data governance helps businesses to become more data-driven

Creating a harmonized view and building common objectives on enterprise data are high value benefits that businesses have been pursuing for countless years with limited success to date.

Challenges to implementing data governance

You reap what you sow: Insight into implementation challenges can help set the right priorities and order when setting up your data governance deployment model.

Challenge	Percent
Lack of resources	46%
Convincing potential stakeholders: Making the added value of data governance visible	44%
Building up knowledge	41%
Management support	41%
Acceptance of business users and future stakeholders	39%
Lack of structure to start data governance initiatives	35%
Lack of budget	31%
Lack of tool support or wrong tools in use	17%
Generating demand	10%
Other	3%

Figure 9: In your opinion, what are the main challenges to implementing data governance measures in order to master data? (n=350)

Our participants opted for lack of resources as the top challenge (44 percent), followed by convincing potential stakeholders of the value add of data governance. In fact, these and all other challenges listed are connected in a causal chain. Ensuring availability of resources entails management and stakeholder support, as well as overall acceptance of the endeavor. Budget, resources and skill building will naturally follow. Especially with rising company size, stakeholder support appears to be an exceptional difficulty with 52 percent of large enterprises with more than 10,000 employees citing it compared to 39 percent on average.

On the other hand, technology appears not to be a significant limiting factor for implementing governance. Although 29 percent of practitioners purchased new technologies for these tasks, only 17 percent rate a lack of tool support to be a challenging factor.

Generating demand is perhaps an underestimated factor. This need might be clear to the spearheads of data governance, who frequently come from business intelligence teams and have experienced the questionable state of enterprise data when transferred into the

Generating demand is perhaps an underestimated factor.

data warehouse. Aside from those selected business analysts, who struggle with data preparation in the course of their own, mostly self-serviced analysis tasks, business unit managers and executive management rarely have a realistic view of the relevant data issues. This means that, regardless of the factual demand apparent to data experts, it is essential to generate executive demand at decision-maker level.

Looking back at the top measures defined by those who have already established data governance, a remarkable match with the top challenges mentioned above can be observed. Priorities 2 to 5 on the practitioners' measures list address their top challenges: internal training, new roles and processes, establishing board responsibility for data and creating dedicated resources. But above all of these, practitioners put data quality monitoring initiatives into practice. This is likely the most feasible and promising way to generate demand for data governance and subsequently secure management and business unit support as well as resources to tackle governance tasks in a collaborative fashion.

Above all, practitioners put data quality monitoring initiatives into practice. This is likely the most feasible and promising way to generate demand for data governance.

Key data governance tasks to manage data

Now that we have looked at the current status of data governance and the specific measures and challenges that businesses face, the following chapters will examine the organizational and technological approaches currently being taken to set up data governance. But first of all, what tasks are involved in data governance?

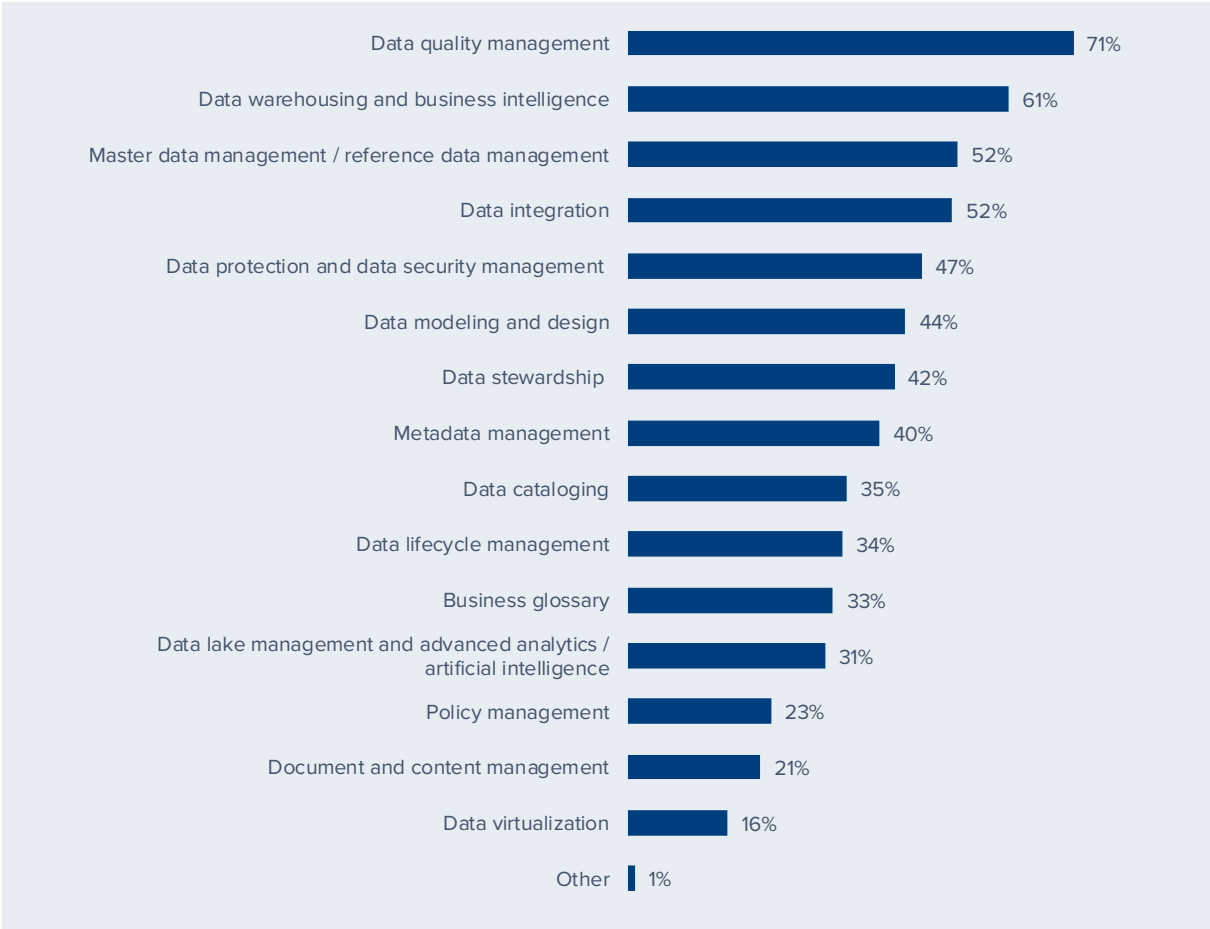| Task	Percentage
Data quality management	71%
Data warehousing and business intelligence	61%
Master data management / reference data management	52%
Data integration	52%
Data protection and data security management	47%
Data modeling and design	44%
Data stewardship	42%
Metadata management	40%
Data cataloging	35%
Data lifecycle management	34%
Business glossary	33%
Data lake management and advanced analytics / artificial intelligence	31%
Policy management	23%
Document and content management	21%
Data virtualization	16%
Other	1%

Figure 10: What do you think are the key tasks in data governance that need to be carried out in order to manage data and make it usable? (n=347)

Data quality management is perceived as a key task by the majority of participants (71 percent). This finding goes hand in hand with data quality being one of the biggest challenges today and in the future. We would have expected master data management to be rated higher than 52 percent, especially since master data often provides the framework and context for broader data usage and is therefore closely related to data quality.

Data warehousing and business intelligence has been rated as the second key task (61 percent). The data warehouse is mostly positioned as the central data source for enterprise KPIs and recurring information needs. It connects and integrates disparate systems in order to service corporate management and is therefore a high

> Data quality is one of the biggest challenges today and in the future.

> We would have expected master data management to be rated higher.

priority in data governance. In addition, insight gathered by the data integration processes that provision the warehouse is an immensely important input for raising awareness regarding the state of source data originating from the enterprise's operational applications. Data lake management, however, does not have such a high priority (31 percent). We expect the importance of the data lake in the context of data governance to grow in the future. As the application of advanced analytics evolves, working with granular, raw data will generate additional requirements concerning the state of operational data.

Data cataloging is another task that was not highly rated (35 percent), even though it was seen as the most important measure to develop data governance. This may owe to the fact that building a data catalog in itself is a tedious and time-intensive task that requires business and technologically oriented data skills that are rarely available.

Data stewardship shows a midfield rating with 42 percent. This does not quite reflect the importance of business user participation mentioned earlier in this study and the highly rated tasks of data quality and master data management, which are fundamental data stewardship responsibilities. Business glossary and policy management should also be mentioned here. Their lack of importance in this list may be attributed to the challenges of missing resources and management support.

Lastly, best-in-class businesses share their top priorities in this list with the overall sample, although master data management and data integration swap places. It is noteworthy that many of the remaining tasks reach a higher than average rating, indicating that these companies have a wider overall focus within their data governance initiatives.

The data warehouse connects and integrates disparate systems in order to service corporate management and is therefore a high priority in data governance.

Building a data catalog in itself is a tedious and time-intensive task. This may be the reason for its fairly low rating as a measure.

The rather reserved rating for data stewardship does not quite reflect the importance of business user participation.

Organizational set-up

When examining which roles are needed in the context of data governance, we compare the views of participants who already have data governance in place (practitioners) and those who are currently implementing or planning to implement data governance (planners).
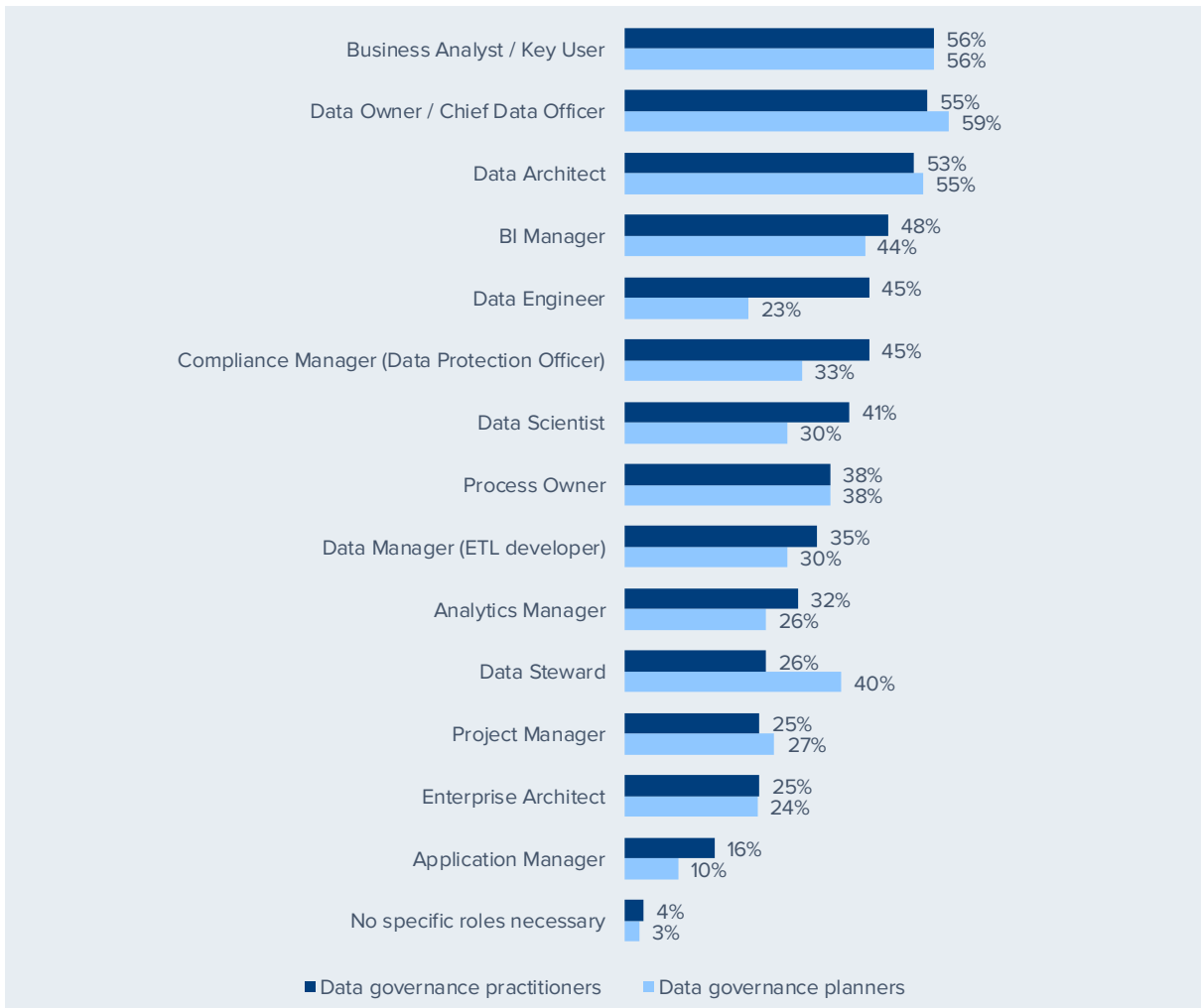

Figure 11: Which roles are established / needed in order to implement data governance successfully?
(n-practitioners=85; n-planners=255)

The perception of practitioners and planners looks quite similar when it comes to the top three roles: In both groups, 56 percent highlight the role of a business analyst / key user. These are not typical data governance roles but are obviously the first address in the line of business for data issues. At the same time, it can be expected that these individuals are not sufficiently prepared for the tasks at hand. Data governance is likely not (yet) part of their job description and not considered in the overall resource planning. These must be seen as significant risks for data governance initiatives. At the same time, only 26 percent of practitioners have implemented the role of data steward. This role actually articulates the added responsibilities a business analyst/key user takes on when taking part in data governance. This circumstance mirrors the perhaps major current chal-

If business analysts/key users are to take on new data governance responsibilities, they need to be explicitly assigned with specific roles and granted appropriate time and training measures to fulfill their duties.

lenge in data governance implementation, namely that while business participation is a key success factor, the organizational and resource set-up needed is not in place to support it. This is also reflected in the consensus of 40 percent of planners that the role of data steward is important. It is interesting to note that practitioners in North America report a relatively high adoption rate of the data steward role (40 percent), whereas it is only 12 percent in Europe.

That the data owner / chief data officer role was rated almost evenly between 'established' and 'needed' (55 and 59 percent) implies data responsibility. Data owners are already frequently defined in the enterprise with specific data domains (e.g., customer data, product data, etc.). Their focus is primarily on operational data and frequently a rather siloed point of view. In contrast, the chief data officer (CDO) takes responsibility for all enterprise data and plays an integrative role covering all data domains. The CDO particularly deals with conflicts of interest that may arise among the different data owners and data consumers. The CDO is quite a new role and has not yet been established on a broad level. From a regional perspective, North America shows an establishment rate of 47 percent, whereas Europe shows 63 percent.

The data architect is the third role established or planned by more than half of our participants (53 and 55 percent). Data architects are not easy to train. They need to have a broad understanding of company data and processes. Therefore, a highly developed skillset in business, conceptual and application-specific data modeling and data storage and access is required.

The need for a compliance manager from a planners point of view appears to be highest in Europe (34 percent). Practitioners see a greater need for this role, which is fairly consistent across all regions. So, while the higher compliance awareness among planners in Europe is perhaps due to GDPR, all regions appear to be affected by compliance in the practical application of data governance.

Lastly, it is interesting to note that Asia-Pacific has a relatively high rating for "no specific roles necessary" (21 percent).

The Chief Data Officer carries responsibility for all enterprise data. This role is quite new and has not yet been established on a broad level.

Data governance organization

Assuming that data ownership for specific data domains is allocated in different business units and spread throughout the organizational structure, the question of how data governance organization can be implemented in an efficient and effective manner takes on greater importance.

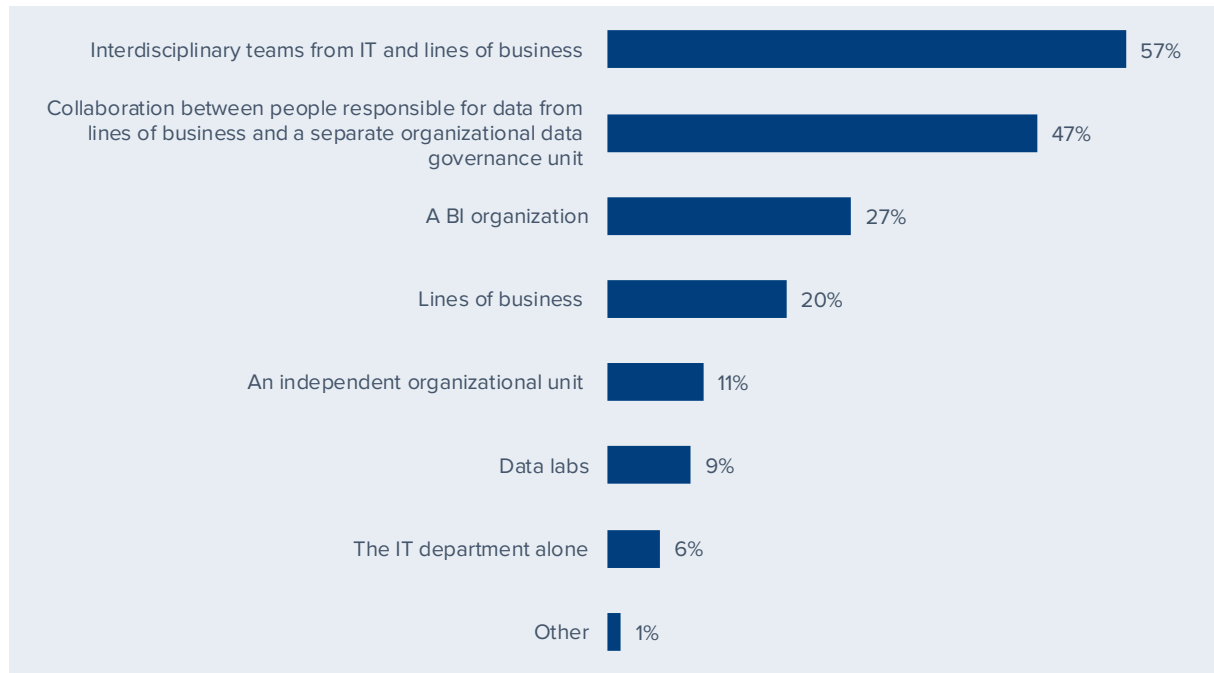|	
Interdisciplinary teams from IT and lines of business	57%
Collaboration between people responsible for data from lines of business and a separate organizational data governance unit	47%
A BI organization	27%
Lines of business	20%
An independent organizational unit	11%
Data labs	9%
The IT department alone	6%
Other	1%

Figure 12: Who should organize data governance? (n=351)

There is a clear consensus about the necessity to collaborate. 57 percent agree that data governance should be organized by interdisciplinary teams including representatives from lines of business and IT. 47 percent believe a separate data governance unit that collaborates with the responsible data representatives from the lines of business is also a good option.

Only a minority of participants is convinced that data governance can be handled by enclosed initiatives for example in the lines of business (20 percent) or IT departments (6 percent). Establishment within the BI organization is the most popular centralized approach (27 percent). This judgement probably follows the assumption that a BI team has a natural and strong data competency. This option is nevertheless rather difficult to implement in practice due to the missing focus on – and lack of authority for – operational source data.

Due to the complexity and interdisciplinary nature of data governance tasks and the necessity to view requirements from an overarching process perspective, the favored virtual organizational approach is self-evident. At the same time, it is important to realize that virtual organizations frequently lack power. Data governance issues are

> There is a clear consensus about the necessity to collaborate.

> Implementing data governance in the BI organization is difficult in practice due to the missing focus on – and lack of authority for – operational source data.

often of a cross-functional nature and therefore prone to conflicts. Dissolving these conflicts may bring a virtual governance organization to its limits.

Lastly, an interesting regional observation is that Asia-Pacific appears to favor centralized forms of organization (BI, line of business, IT) more than Europe and North America.

Technology support and benefits

While a lack of suitable tools is not a predominantly limiting factor, technology can play an important role in mastering data governance. We asked participants which functions they need to support their data governance tasks.

Technology can play an important role in mastering data governance.

Data cleansing — 58%
Quality measurement and monitoring — 53%
Data integration — 52%
Data cataloging — 47%
Data modeling — 44%
Data protection and data security — 43%
Stewardship — 38%
Business glossary — 38%
Data profiling — 33%
Collaboration — 33%
Data enrichment — 32%
Data lineage and impact analysis, documentation — 31%
Data search and identification — 30%
Description of data correlations — 29%
Policy management — 29%
Data as a service — 24%
Functions for automation / machine learning / AI — 20%
Data virtualization — 16%
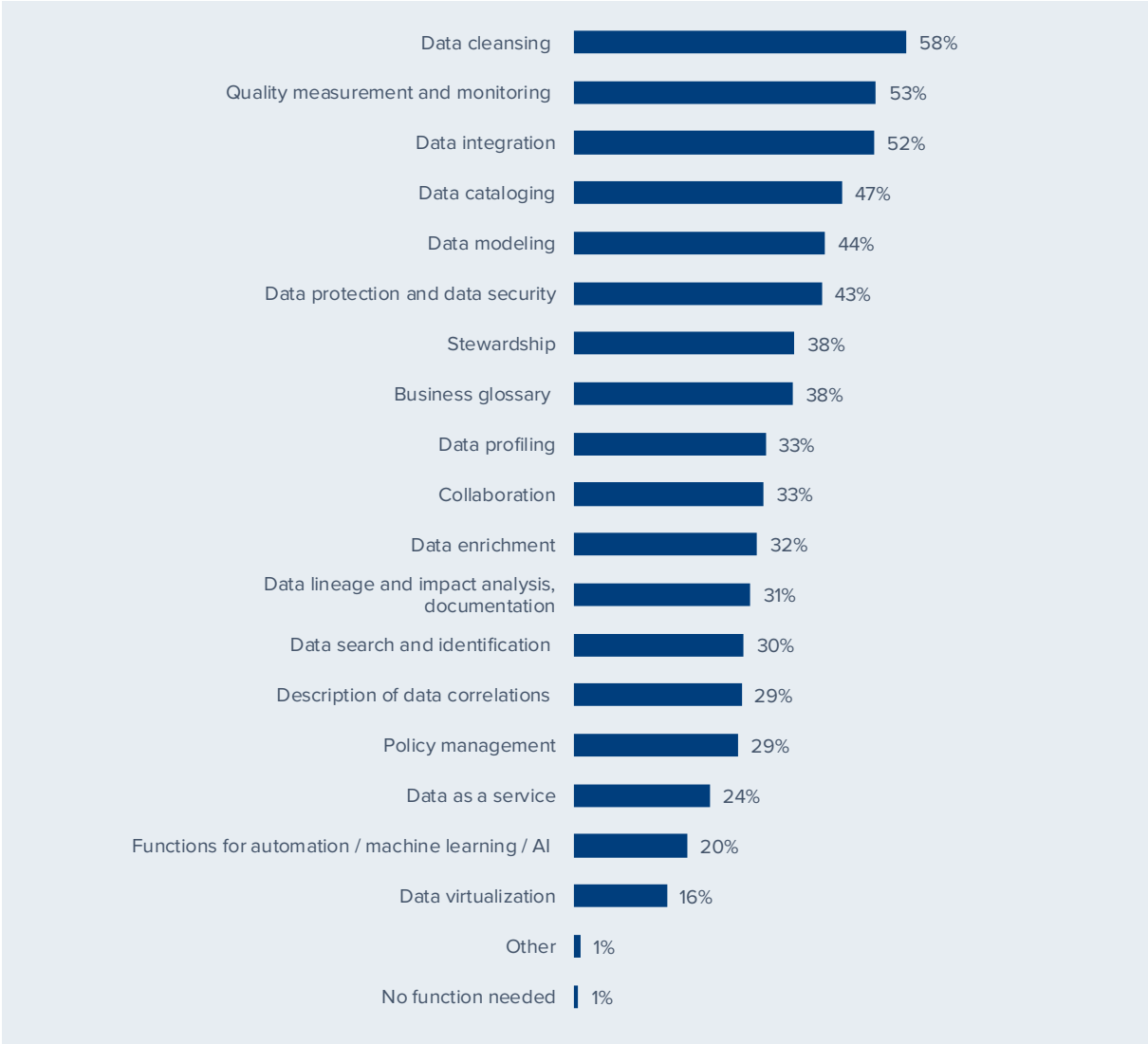Other — 1%
No function needed — 1%

Figure 13: Which of the following functions are needed to support data governance tasks? (n=338)

Unsurprisingly, data cleansing (58 percent) and data quality management (53 percent) were identified as the most needed functions. Both address the highest rated data governance challenge: insufficient data quality. Data integration (52 percent) follows as the third most important function. This is an interesting outcome, as data integration is not per se a data governance activity but rather a specific task needed when data from various siloed source systems is transferred to a target system. However, considering that the majority of participants focus their data governance initiatives on the BI and data warehouse environment, the importance of data integration becomes more apparent.

It is noteworthy that data cataloging, the highest rated data governance measure, was considered necessary by only 47 percent of respondents. In general, functional priorities appear to have a stronger focus on administrative than on shaping and restructuring functions. Functions such as stewardship (38 percent), business glossary (38 percent), collaboration (33 percent), data search and identification (30 percent) and policy management (29 percent) that could help boost efficiency and effectiveness in working with data only reached midfield ratings. In North America, data cataloging, data stewardship and business glossary are rated higher than the global average.

Priorities differ somewhat between best-in-class companies and laggards. Best-in-class companies appear to have a stronger focus on monitoring and overseeing, whereas laggards concentrate more on operational activities.

Interestingly, 20 percent of participants found they needed functions for automation, machine learning and artificial intelligence to support data governance. Due to the complexity of the enterprise landscape and, at the same time, the limited availability of human resources, companies need to find innovative ways to boost efficiency in governance tasks. Awareness is clearly growing that machine learning and AI can play a central role in this area, for example, to support search and discovery tasks or in the creation and maintenance of business glossaries and data catalogs.

The functional priorities appear to have a stronger focus on administrative than on shaping and restructuring functions.

Best-in-class companies appear to have a stronger focus on monitoring and overseeing, whereas laggards concentrate more on operational activities.
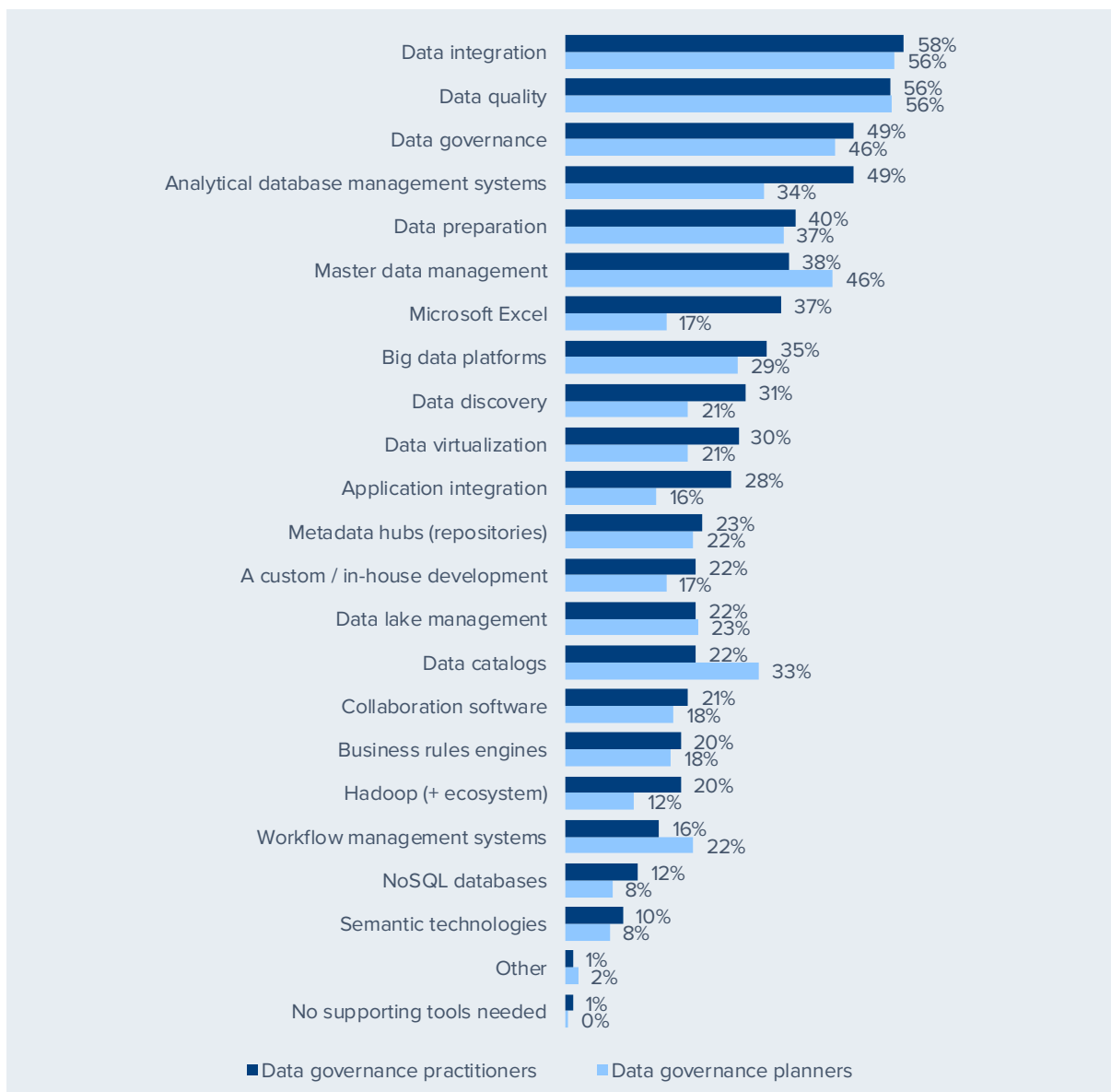
Figure 14: With which tools do you implement / plan to implement the functions in the previous question?
(n-practitioners=81; n-planners=238)

The two most favored tool categories, data integration and data quality, correspond to the highest rated functions needed for data governance. Data integration tools (58 percent for practitioners and 56 percent for planners) were rated somewhat higher than data quality (56 percent for both). This may be due to the fact that many data integration platforms include integrated functions for data quality.

Data governance tools appears to be a popular category (49 percent for practitioners, 46 percent for planners). However, it is not clear how this tool category is exactly defined. Interestingly, analytical databases received a high ranking (49 percent) from practitioners.

Microsoft Excel also appears to be a widely used tool for practitioners (37 percent). While not overly surprising, depending on the nature of the use of Excel this may be rather worrying from a governance and, in particular, from a compliance perspective.

In general, the views of practitioners and planners do not vary significantly. The frequent mention of data catalogs by planners, which reflects the higher importance this measure has for this group, is notable. The need for data catalogs is frequently driven out of data labs projects where they help to boost the usability of raw data collected in a data lake approach.

The wide range of different technologies participants voted for indicates how complex the technology set-up for data governance can be. It seems there is no clear preference as to the most important tools required for data governance. It also appears that the measures identified and the tools being used are not necessarily aligned.

At any rate, participants are fairly satisfied with the technology they use for data governance. Only 3 percent claimed to be not very satisfied with the tools they used.
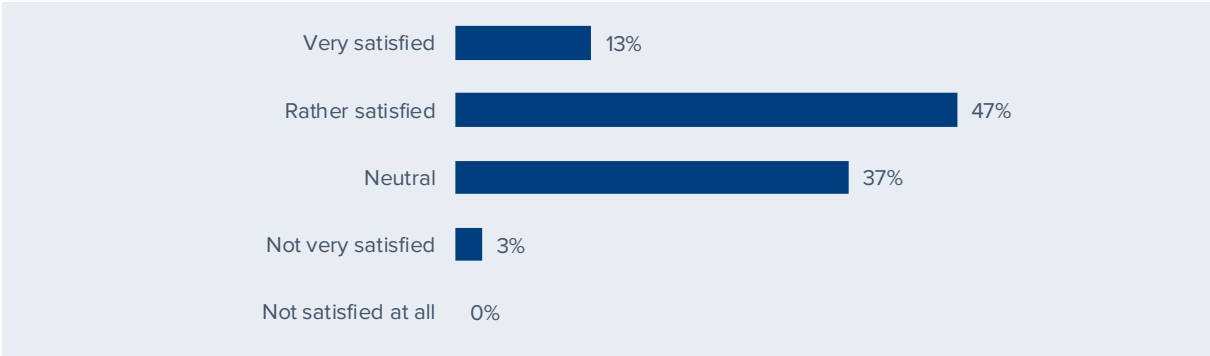
Figure 15: How satisfied are you with the introduction / use of these tools? (n=78)

Overall, almost half of the survey participants are rather satisfied with their tools (47 percent) although 37 percent describe their level of satisfaction as 'neutral'. This leaves room for improvement. Only 13 percent are very satisfied with their tools for data governance. Conversely, finding suitable tools is not perceived as a big challenge for most participants (see Figure 6). This finding is consistent with the challenges identified during implementation, where only 17 percent pointed out a lack of tools support or that the wrong tools were being used (see Figure 9).

Detailed analysis shows that participants tend to be more satisfied with some technologies than with others. In particular, we found relatively high levels of satisfaction with in-house developments and metadata hubs (78 percent). Over 70 percent of users are satisfied with their data virtualization tools, data lake management, big data platforms and Hadoop ecosystems.

In particular, the latter three technologies are likely to reveal that participants from BI and advanced analytics roles are looking at

The widespread use of Microsoft Excel for data tasks is rather worrying from a governance and in particular from a compliance perspective.

It seems there is no clear preference as to the most important tools required for data governance.

There is clearly room for improvement when it comes to tool usage.

data governance. Hadoop systems, big data platforms and data lake management systems are not very likely to help to view, control and monitor data in classic operational systems.

Interestingly, semantic technologies achieved the highest satisfaction rate (88 percent). This outcome must be interpreted with some care, since the number of participants who actually use this kind of technology is rather small. However, it will be interesting to observe how market adoption of this technology evolves.

Our survey participants find that tools in support of data governance deliver a number of benefits:
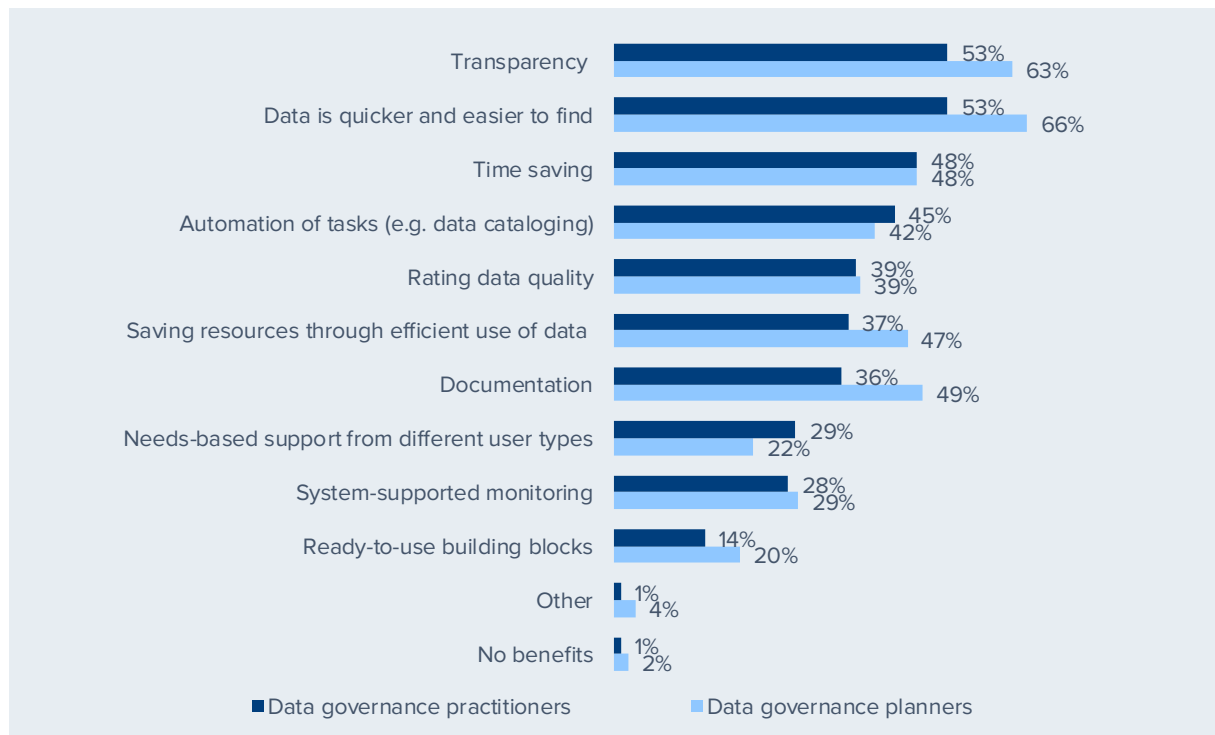

Figure 16: What do you see as the main benefits of the tools you use / plan to use? (n-practitioners=83; n-planners=253)

In established data governance environments, a number of benefits have been realized. However, for most of the benefits shown in Figure 16, the expectations of users still planning data governance initiatives exceed the actual benefits achieved by practitioners. Closer analysis does not clarify whether this is a technology issue or if it relates to the organizational challenges data governance initiatives face, which are described earlier in this study.

The realized benefits of practitioners are somewhat behind the expectations of data governance planners.

53 percent of practitioners regard transparency and better search efficiency as the primary benefits of the tools they use. Time saving (48 percent) and automation of tasks (45 percent) play an important role as well. Data seems to lose its 'black box' character and can be handled more efficiently. Little benefit is seen in utilizing ready-to-use building blocks and blueprints.

Success factors for data governance

Figure 17 shows that businesses are aware of important success factors when it comes to data governance initiatives.

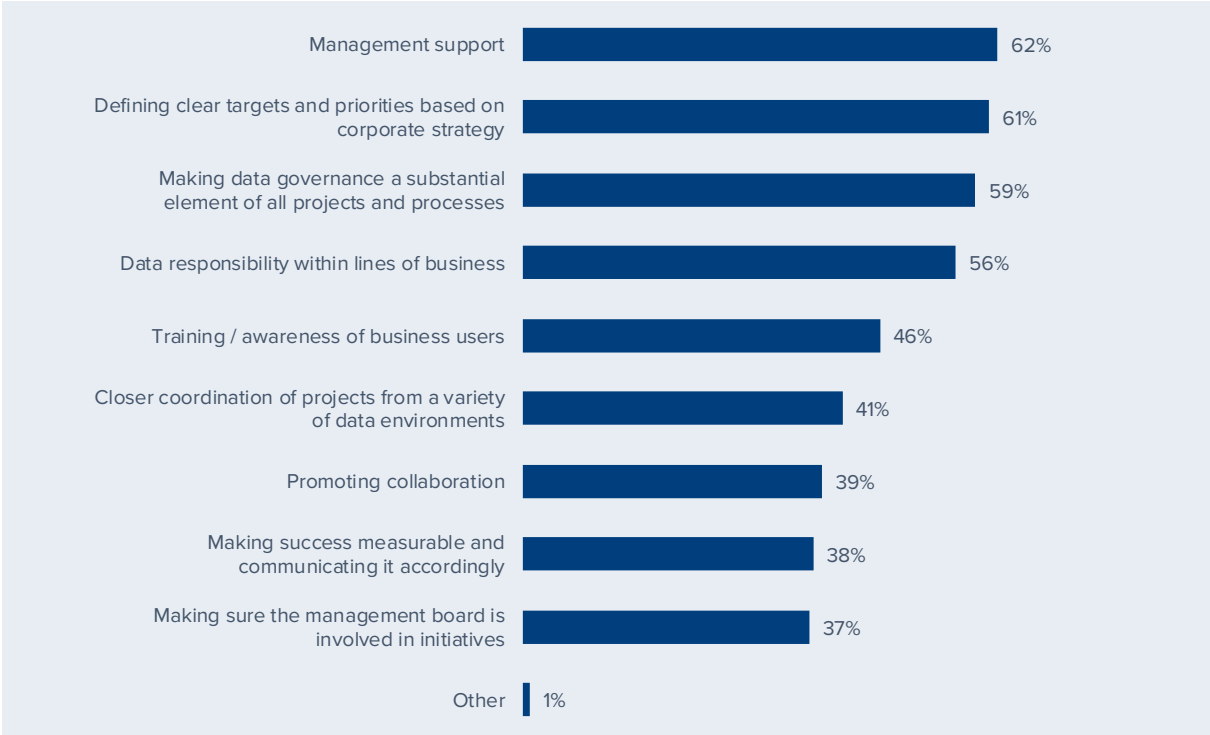| Factor	Percentage
Management support	62%
Defining clear targets and priorities based on corporate strategy	61%
Making data governance a substantial element of all projects and processes	59%
Data responsibility within lines of business	56%
Training / awareness of business users	46%
Closer coordination of projects from a variety of data environments	41%
Promoting collaboration	39%
Making success measurable and communicating it accordingly	38%
Making sure the management board is involved in initiatives	37%
Other	1%

Figure 17: In your opinion, which factors contribute to successful data governance initiatives? (n=347)

Participants previously identified convincing stakeholders and management support as major challenges. Consequently, management support and the identification of priorities based on corporate strategy are top of the list when it comes to critical success factors for data governance. The third priority – making data governance a substantial element of all projects and processes – appears difficult to realize in practice. The predominant focus of data governance initiatives on BI and data warehouse environments demonstrates that there is still a long way to go.

When these prioritized success factors are compared with current measures and established roles, it becomes apparent that data governance has not yet transferred from theory to practice.

For example, only 24 percent of participants believe that internal marketing of data governance is important to convince potential stakeholders from lines of business. However, this is a crucial measure to secure management support. Overall, the findings of this study suggest that much of the potential of data governance is yet to be discovered.

The predominant focus of data governance initiatives on BI and data warehouse environments demonstrates that there is still a long way to go.

It is apparent that data governance at this point has not yet transferred from theory to practice.

Conclusion

This study gives an overview of the current status of data strategy and governance initiatives. Companies find themselves facing great complexity and, in particular, organizational challenges. We would like to conclude with the central success factors we have learned from this study:

- Line of business involvement is essential for data governance establishment. Business users currently do not have the appropriate resources and skills available to master this task and must be empowered.

- Executive support and acceptance in business are indispensable prerequisites for pursuing data governance. Management can be convinced by obvious need. A good starting point is to monitor and visualize (lacking) data quality and create a feedback loop from analytical data usage back into core business processes.

- Data governance should be set up as an overarching enterprise initiative, not limited by boundaries set by specific data environments. Focusing data governance on the data warehouse maintains the data problem instead of solving it.

Appendix A: Methodology and demographics

The online user survey was conducted worldwide in August and September 2018. BARC promoted this survey through websites, in email newsletters and directly to the BARC research panel. A total of 378 people from a wide range of competencies (see figure 18), company sizes (see figure 19) and industries (see figure 21) participated in this survey. Most participants came from Europe (55 percent) while 28 percent were from North America. Asia-Pacific (9 percent), South America (5 percent) and Africa (2 percent) were represented by smaller proportions of participants (see figure 20). Deviations are possible due to the rounding of figures.

Position

Position	%
BI / analytics employee with management responsibility	22%
BI / analytics employee with no management responsibility	18%
IT employee with management responsibility	13%
Executive manager BI & Analytics	12%
External consultant	8%
Data governance delegates (data owner, data steward, data protection officer)	7%
Business managers, experts & analysts (in LoB)	7%
CxO (CEO, CFO, CMO etc.)	7%
IT employee with no management responsibility	5%

Figure 18: What position do you hold within the company? (n=378)

Company size

Company size	%
Less than 250	27%
250 - 999	18%
1,000 - 9,999	34%
10,000 or more	20%
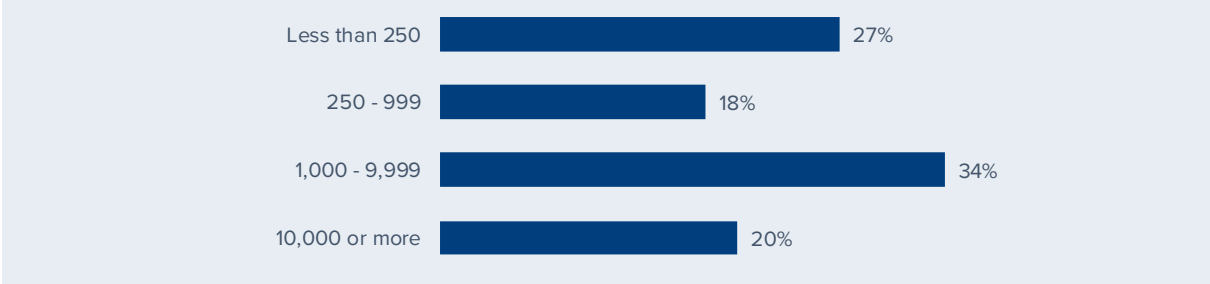
Figure 19: How many employees does your company have? (n=375)

Regions

Region	Percentage
Europe	55%
North America	28%
Asia and Pacific	9%
South America	5%
Africa	2%
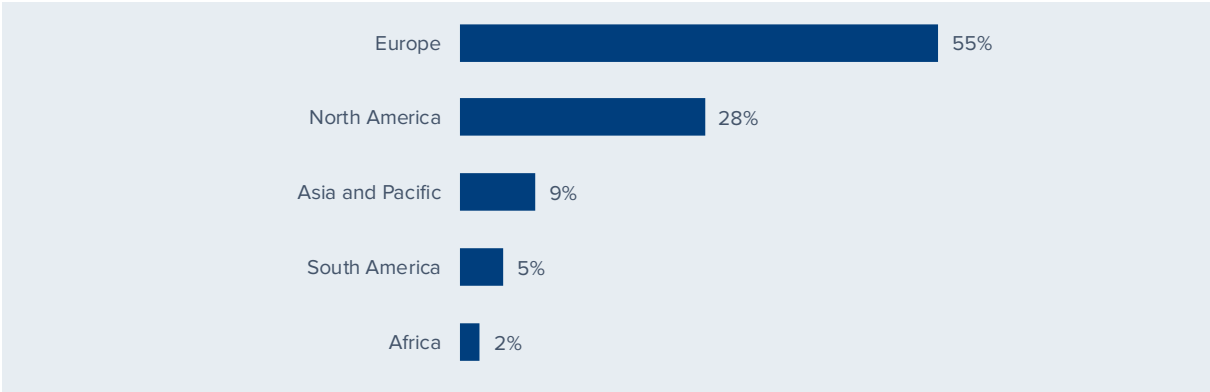
Figure 20: In which region are your located? (n=378)

Industry

Industry	Percentage
Services	28%
Industry	18%
IT	17%
Banking and finance	16%
Public sector	10%
Retail / Wholesale / Trade	7%
Other	3%
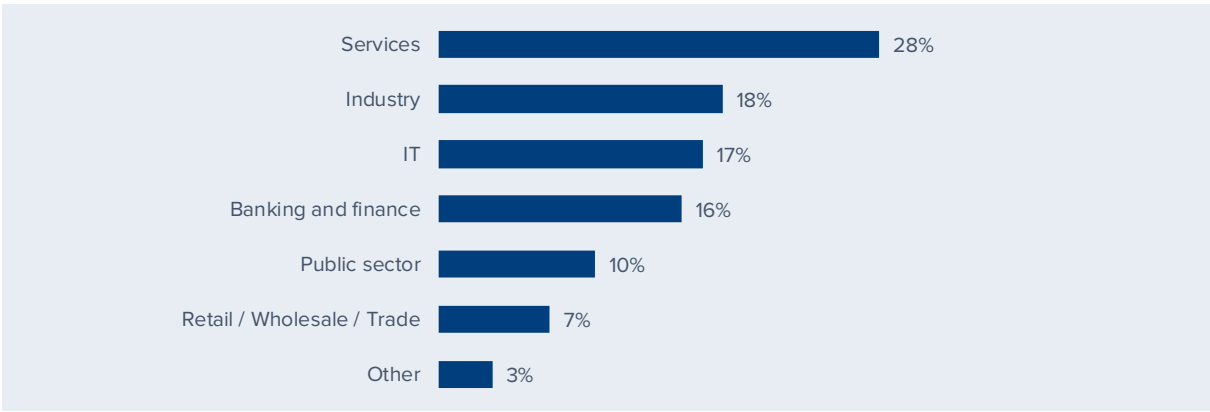
Figure 21: Which of the following best describes your organization's industry sector? (n=377)

Appendix B: Data governance practitioners and planners

What distinguishes companies that have already established data governance from those planning to establish it? In order to answer this question, we asked our participants whether they had established data governance within their business or whether they were planning to do so (see figure 22). For the purposes of this study, participants who stated that they had already established data governance are labeled "practitioners" (24 percent), while participants who are currently establishing or planning to establish data governance are referred to as "planners" (50 percent). Participants who have neither established data governance nor plan to do so were not considered at all in this aggregation.
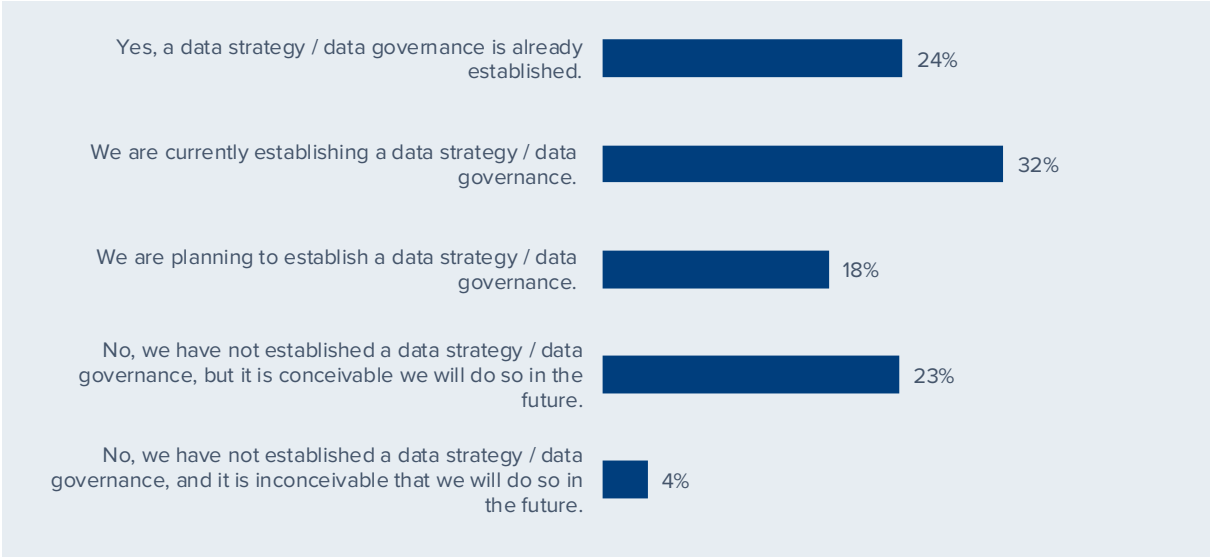

Figure 22: Do you have data governance / a data strategy established within your business or do you plan to do so? (n=364)

Appendix C: Best-in-class and laggards

Parallel to practitioners and planners, we also divided the sample into "best-in-class" and "laggards" in order to identify differences in terms of data governance drivers, data environments and challenges in data use. This division was made based on the question "How good do you consider your company to be at handling data compared to your competitors?" (see figure 23). Participants who considered themselves to be much better at handling data are referred to as "best-in-class" (12 percent) while those who perceived themselves as slightly worse or much worse are classed as "laggards" (20 percent).
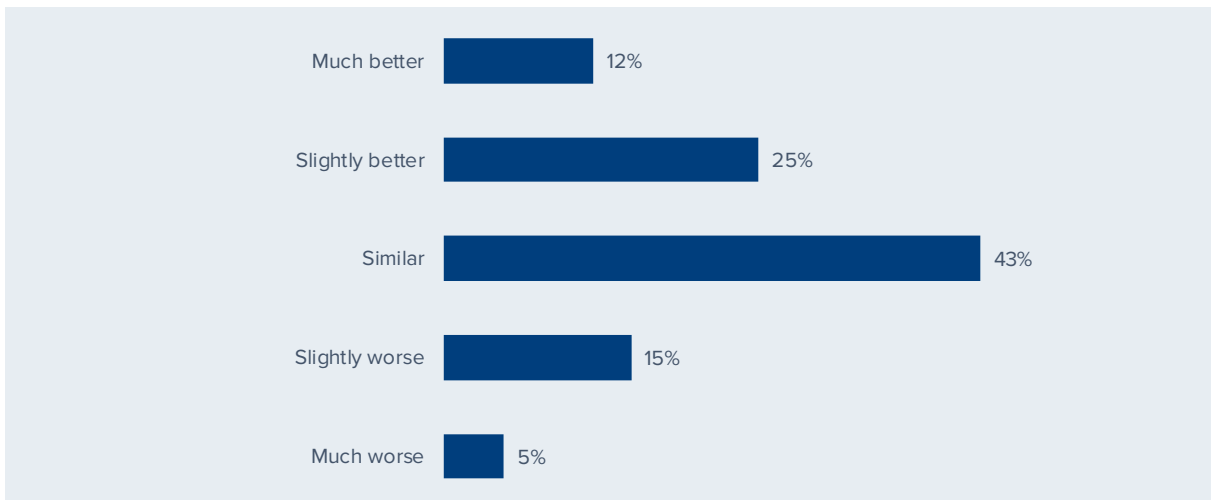
Much better — 12%
Slightly better — 25%
Similar — 43%
Slightly worse — 15%
Much worse — 5%

Figure 23: How good do you consider your company to be at handling data compared to your competitors? (n=333)

MarkLogic

www.marklogic.com

MarkLogic

Company profile

MarkLogic - the leading operational and transactional Enterprise NoSQL database provider

For over a decade, organizations around the world have come to rely on MarkLogic to power their innovative information applications. As the world's experts at integrating data from silos, MarkLogic's operational and transactional Enterprise NoSQL database platform empowers our customers to build modern applications on a unified, 360-degree view of their data. Headquartered in Silicon Valley, MarkLogic has offices throughout the U.S., Europe, Asia, and Australia.

Analytics and business intelligence in enterprises and government agencies can only be successful if there is a data-centric infrastructure in place that can consolidate data silos on a single agile platform. Designed to integrate data from silos better, faster, and with less cost, MarkLogic can help you integrate data and build your 360-degree view up to four times faster than if using a traditional database. And, you don't have to sacrifice any of the enterprise features required for storing and managing mission-critical data.

- **Easy to Get Data In – Ingest structured and unstructured data as is with a flexible data model that adapts to changing data. MarkLogic natively stores JSON, XML, text, geospatial, and semantic triples.**

- **Easy to Get Data Out** – With an "Ask Anything" Universal Index, you can run lightning fast searches across all of your data. We also provide APIs and other tools to enable fast application development and deployment.

- **100% Trusted** – MarkLogic is enterprise ready, having ACID transactions, scalability and elasticity, certified security, high availability and disaster recovery, and other enterprise features required to run your business.

Contact Details

MarkLogic

Theatinerstr. 11

80333 München

+49 89 71 04 22 151

germany@marklogic.com

www.marklogic.com

BARC – a CXP Group Company

BARC is a leading enterprise software industry analyst and consulting firm. Major companies, government agencies and financial institutions rely on BARC's expertise in software selection, consulting and IT strategy projects. BARC has specialized in core research areas including Data Management (DM), Business Intelligence (BI), Customer Relationship Management (CRM) and Enterprise Content Management (ECM).

Along with CXP and Pierre Audoin Consultants (PAC), BARC forms part of the CXP Group – the #1 independent European research and advisory firm in the field of digital, software and IT services.

CXP Group supports its customers in their decision making thanks to a unique 360° approach encompassing assessment capabilities covering global and local IT markets, products & technologies, services, and their business usage.

Every year the group supports more than 1,500 large and medium-sized business-user organizations in their sourcing decisions and investment projects, as well as their choice and optimization of software solutions, and offers consultancy services in many areas (BI, HR, ECM, IT management, finance, ERP, CRM, etc.).

Capitalizing on 40 years of experience, and with 155 seasoned professionals present in 8 countries, CXP Group supports more than 300 software vendors and IT service providers with market research products, strategic analysis, sales support & marketing services.

CXP Group has 3 subsidiaries: Le CXP, BARC (Business Application Research Center) and PAC (Pierre Audoin Consultants).

Germany

BARC GmbH

Berliner Platz 7

D-97080 Würzburg

+49 931 880651-0

www.barc.de

Austria

Meldemannstraße 18 / 01.14

A-1200 Wien

+43 660 6366870

Switzerland

BARC Schweiz GmbH

Täfernstraße 22a

CH-5405 Baden-Dättwil

+41 76 3403516

Rest of the World

+44 1536 772 451

www.barc-research.com

BARC

A TEKNOWLGY GROUP COMPANY