

Data Democratization

Discover and Realize the
Value of Your Data Assets
With the Analytics Engineer
and the Modern Data Stack

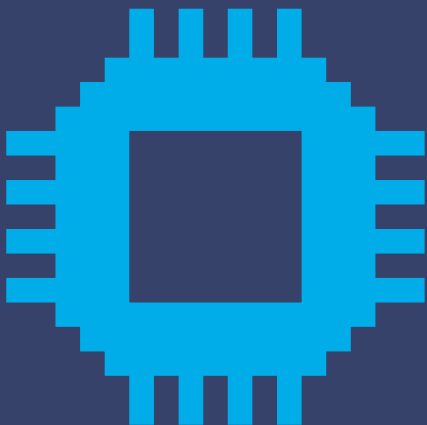
```
elif _operation == "MIRROR":  
    mirror_mod.use_x = False  
    mirror_mod.use_y = False  
    mirror_mod.use_z = True  
  
#selection at the end  
mirror_ob.select=1  
modifier_ob.select=1  
bpy.context.scene.objects.  
obj["Selected"] = str(mod)
```

Whitepaper

GO
DATA
DRIVEN

_ TABLE OF CONTENTS

Introduction	3
What is Data Democratization?	4
» 2.1_The Data Assets Journey	4
» 2.2_The Data Supply Chain	4
» 2.3_Introducing Self-Service Analytics	4
» 2.4_Increasing Availability of Data and Information	5
» 2.5_Levels of Data Democratization	5
» 2.6_Benefits and Risks of Data Democratization	6
Enablers of Data Democratization	8
» 3.1_Availability Enablers	8
3.1.1_Modern Data Stack	8
3.1.2_Data as a Product	10
3.1.3_Data Governance	11
» 3.2_Self-Service Enablers	12
3.2.1_Data Literacy	14
3.2.2_Support Structure	
The Analytics Engineer	15
The Road to Data Democratization	17
Meet the Experts	19



_ WHY DATA DEMOCRATIZATION IS IMPORTANT

Introduction

A growing number of organizations recognize data as a critical asset. By containing and leveraging their data, organizations can expedite business decision-making and obtain a competitive advantage in their market. Embracing analytics throughout an organization can fuel innovation, business agility, and growth, making data integral to many modern business strategies.¹

When organizations leverage data, it becomes an asset that allows them to create better customer experiences, improve operational efficiencies, and develop new innovative products, as Table 1 shows. In addition, as organizations accrue data and increase its useability, they can turn it into an asset by offering it as a service, for example, benchmark studies or product catalogs.

Table 1
Value of Data Assets

Improved Products & Services	New Product & Service	Data is the Product & Service
<ul style="list-style-type: none"> » Assess process bottlenecks » Improve operations » Make better forecasts » Support decisions 	<ul style="list-style-type: none"> » Understand customers better » Personalized offers » Expand product portfolio » Introduce customer self-service » Exploit external data sources 	<ul style="list-style-type: none"> » Anonymized market insights and benchmarks » Product catalogs

Note. Potential areas where the value of data assets can be unlocked.

While the potential of data is widely recognized, many organizations struggle with using their data assets efficiently. Some of the challenges include:

- » **Limited Availability.** Data models and reports are often fully governed in the IT or data domain, and access to data is limited. Corporate business intelligence (BI) solutions and their reporting tools provide little to no flexibility to address non-standard analyses. This lack of capability limits how much information a business can extract and the reusability of data (models) across use cases.
- » **Lack of Data Literacy and Awareness.** As the access to and usage of data assets is limited, practice and application do not reach the average business end-user. Not all employees are aware of the potential and possibilities of working with data. Consequently, the business domains do not acquire the necessary experience, skills, and literacy to use data in their day-to-day work.
- » **No Data Management or Quality Assurance.** Visibility in data handling is limited within the organization. It is unclear who owns data sources or assures data quality, and existing tooling does not enforce data quality, access management, or confidentiality regulations (e.g., GDPR).
- » **Trust.** Data quality issues compromise trust in data assets. In addition, managers and IT departments do not trust business users in using data while adhering to security standards. Trust is specifically an issue if the data and related reports support critical business decision-making.

These challenges all converge into one overarching obstacle: data and information do not reach the right people at the right time to enable data-driven decision-making. Leading organizations remediate this by democratizing their data assets. Data democratization is the process of improving data and information availability through modern cloud-based technologies and enabling employees to retrieve information from that data through self-service analytics.

This whitepaper discusses the process of data democratization in-depth and shares our experiences of working with business leaders and organizations across industries and sizes. In addition, it provides a practical guide that will help your organization leverage and monetize its data assets by improving data and information availability and enabling self-service analytics. In other words, it's a complete guide to democratizing data.

¹ <https://godatadriven.com/topic/data-ai-survey-2021>

What is Data Democratization?

To become data-driven, an organization must make data readily available and enable business users within the organization to leverage it. We refer to this process as *data democratization*. Data democratization puts the right information into the right people's hands at the right time, empowering them to extract information, make decisions, and uncover opportunities directly.

2.1_The Data Assets Journey

As Figure 1 illustrates, data assets go through various stages from capture to consumption as information. Real-world data is captured as it enters an organization's IT domain. Data professionals pick up the operational data stored in IT systems and prepare it to drive business intelligence (BI) or artificial intelligence (AI) initiatives. Both internal and external end users can then draw insights from the information found in these data assets to create business value.

2.2_The Data Supply Chain

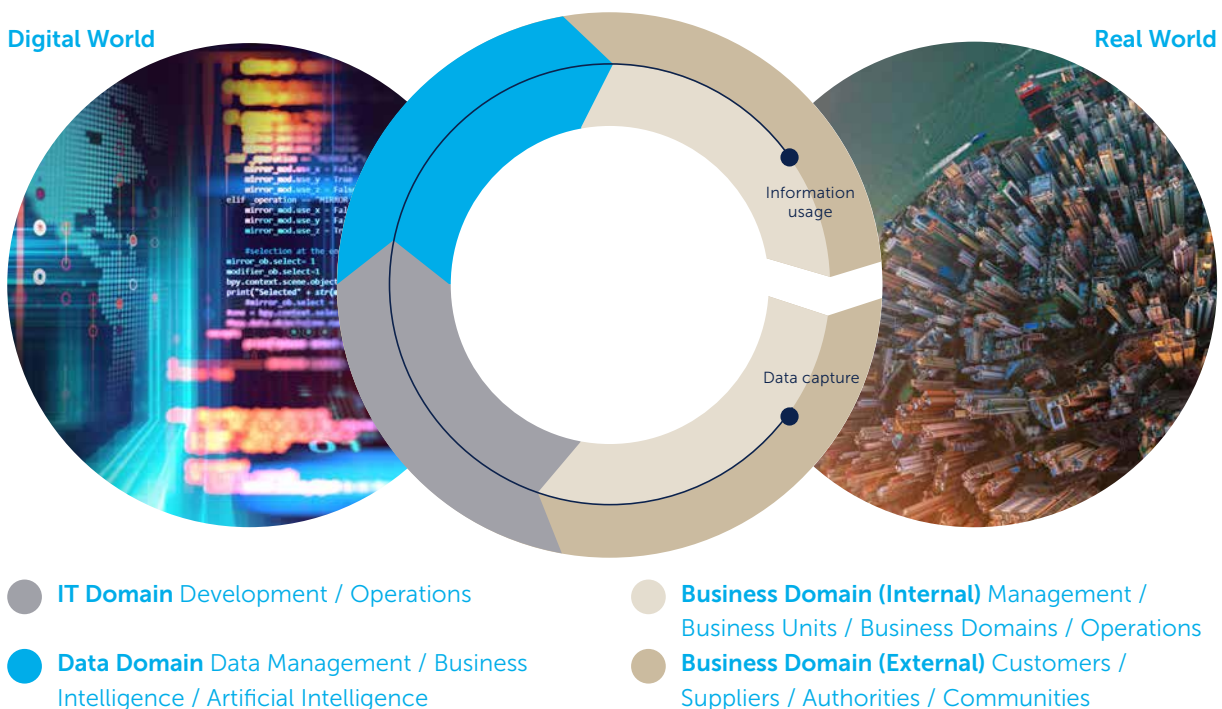
The data assets journey is formalized as the *data supply chain*, shown in Figure 2. Data assets move through IT, data, and business domains, from data capture to information usage. Along the data supply

chain, data gets processed into different *data layers* containing different characteristics (e.g., *cleaned data* or *prepared data*).

2.3_Introducing Self-Service Analytics

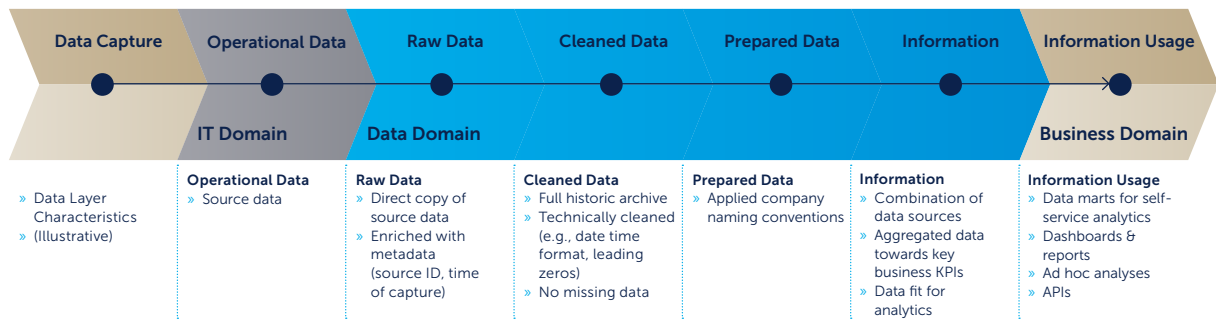
As companies strive to become more data-driven, the demand for information grows, potentially overloading IT, BI, or data engineering and creating a bottleneck. To overcome these challenges, organizations can expand the skillsets of information users, enabling them to use the data directly. As Figure 3 shows, self-service analytics shifts the data-business boundary in the data supply chain. Improving data literacy across an organization and establishing a dedicated support structure can enable self-service analytics (section 3.2).

Figure 1
The Journey of Data Assets



Note. The circular journey of data assets—captured in the real world, processed in the digital world, then used again in the real world to drive business value.

Figure 2
The Data Supply Chain



Note. Example data layers with data characteristics are company-specific and shown for illustrative purposes only.

2.4_Increasing Availability of Data and Information

As the demand for data and information grows, organizations need to facilitate its discovery, usage, and distribution in a scalable, secure way to increase availability to users, as shown in Figure 4. A modern data stack, creating data as a product, and a data governance program can promote company-wide availability and usage of data assets. These topics are covered more thoroughly in section 3.1.

Data Democratization, therefore, only occurs when an organization invests in both self-service analytics and data and information availability. Figure 5 shows the different states companies experience in their journey to democratize their data. Not all states are relevant to every organization. A data strategy determines the exact as-is and desired to-be states. A data and AI maturity scan can help your organization establish its data strategy and assess where it is currently and where it's headed.

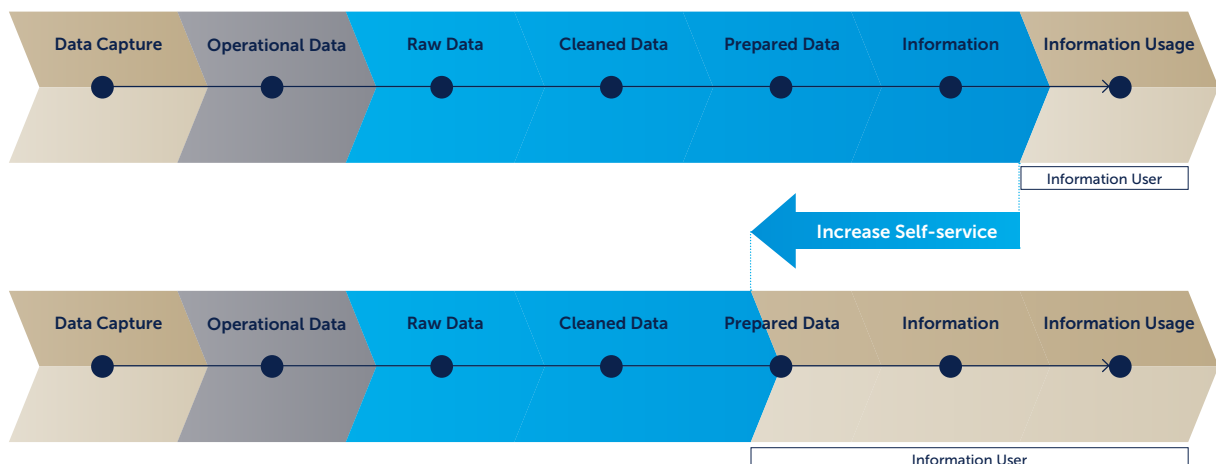
2.5_Levels of Data Democratization

The two main components of data democratization are self-service analytics and the availability of data and information. Companies that want to democratize their data assets need to mature in both areas. There is no return on investment if all data is made available across the organization, yet no one can extract information and create business value. The same holds if everyone in the organization is trained and skilled in data and analytics—if the data is not made available to information users, there can be no return on investment.

...
“Data democratization increases the usage of data assets, uncovers new opportunities with data, and improves existing data operations. All this expedites decision making and facilitates the creation of business value.”

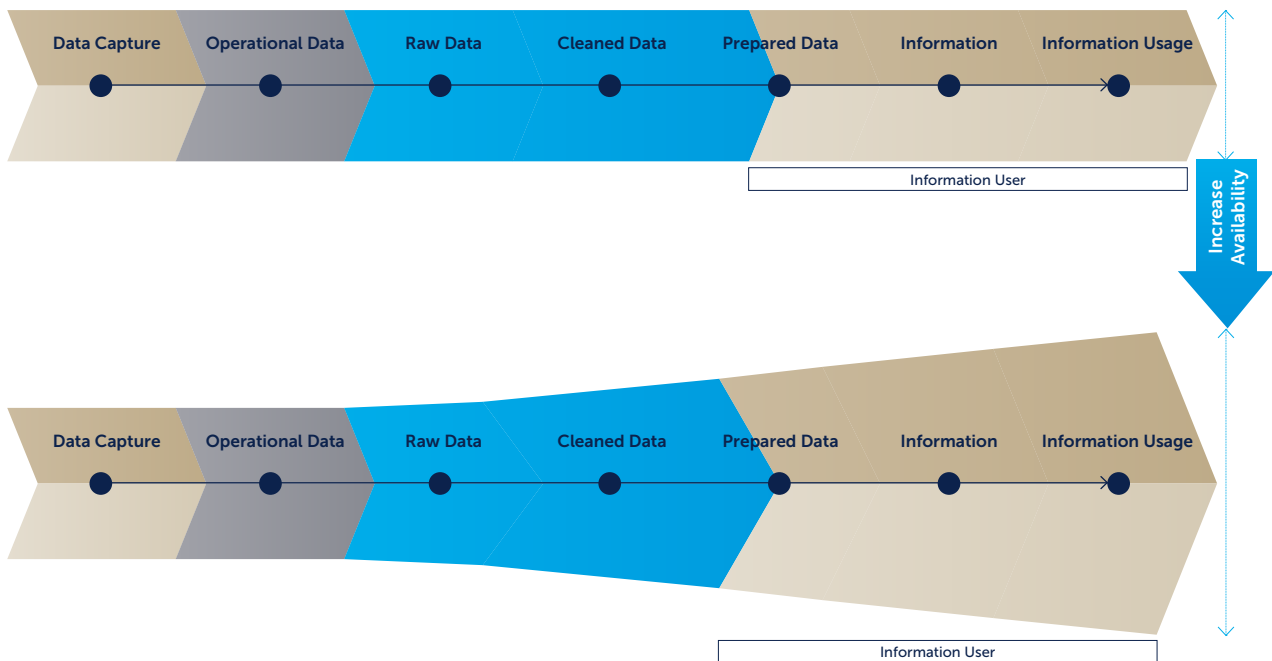
— Bram Ochsendorf, lead data scientist, GoDataDriven

Figure 3
Introducing Self-Service Analytics



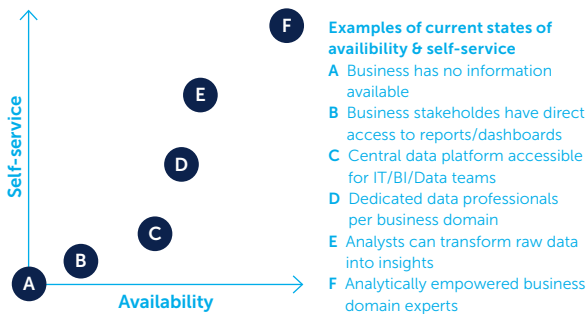
Note. Introducing self-service analytics enables (business) users to extract information from data.

Figure 4
Increasing Data and Information Availability



Note. Organizations can increase the availability of data and information by facilitating the discovery, usage, and distribution of data.

Figure 5
States of Data and Information Availability and Self-Service Analytics



Note. Different states companies may experience in their journey to democratize their data. Not all states are relevant to every organization.

2.6_Benefits and Risks of Data Democratization

Up until now, we have focused on the benefits of data democratization, as summarized in Table 2. Data democratization also comes with risks when certain key enablers are not in place. For example, business users who lack knowledge or experience can misinterpret data and create erroneous information.

Additionally, people in democratized business domains may not have the technical skillset to use and prepare data for analytics, leading to

suboptimal solutions or even a stand-still in efforts to create value from data. Organizations that invest in data literacy programs, data learning journeys, and dedicated support structures can ensure more effective steering. It's also important to note that data democratization can lead to chaotic situations.

Table 2
Benefits and Risks of Data Democratization

Benefits	Risks
Utilize Data Assets <ul style="list-style-type: none"> » More informed decisions » More business domain understanding 	Lacking knowledge <ul style="list-style-type: none"> » Data misinterpretation » Unconscious incompetence results in "garbage out"
New Data Opportunities <ul style="list-style-type: none"> » Increase literacy, creativity, and opportunity awareness » Reduce AI hype and impractical data endeavors 	Privacy & Security <ul style="list-style-type: none"> » No right to access the data » No right to use data for specific purpose » Data breaches
Improved Data Operations <ul style="list-style-type: none"> » Reduce ad hoc questions to data analysts » More focus on higher-impact projects » Improve time-to-market » Retain data talent 	Chaos <ul style="list-style-type: none"> » Decentralized, diverging definitions and metrics implementation » Work redundancy

These situations might involve uncoordinated data initiatives and diverging definitions of business metrics and key performance indicators. Data governance programs define the policies and regulations that must be in place to control access and enable secure sharing of information to internal and external users.

So, although data democratization has significant potential benefits, the initiative can be counterproductive when the right enablers are not in place. Section 3 covers the enablers that can set your data democratization initiative up for success in more detail.

...

“We’ve transformed the way we work at Funda, shifting our focus more towards understanding our customers to drive product and business innovations. It would have been impossible to do that without a reset in how we technically collect, share, and explore data and how we make it accessible for analysis.

In just a few years, we’ve moved away from traditional centralized and closed business intelligence to an open, explorative, and collaborative approach to data analysis and customer research. It has positively influenced how we present arguments and make decisions across the entire organization.”

— Eduardo Neves, CTO, Funda



Enablers of Data Democratization

This section covers the enablers of data democratization, divided into two main components: availability and self-service.

3.1_Availability Enablers

There are three main enablers of data and information availability: a *modern data stack*, using *data as a product*, and a *data governance* program.

3.1.1_Modern Data Stack

Data platforms and advances in (cloud-based) technology have revolutionized the ways organizations can make data widely available to support their data-driven initiatives.

Data Platform

As shown in Figure 6, a data platform makes data accessible within the organization and empowers its users to create insights and value for the organization quickly. The general architecture of a data platform covers the following functions:

- » **Ingestion.** The process of loading data from separate data sources into the storage solution.
- » **Storing Data.** Data is stored in a data platform to be accessed in a scalable way.
- » **Transforming Data.** Data is transformed by a distributed processing stack, which scales with the load of the processing job.
- » **Analysis Stack.** Supports complex analysis by users to create insights for the organization.

...

“The data platform is the engine room for our democratization. It not only centralizes technical resources but also consolidates innovation.”

— Kevin Duisters, commercial analytics manager, Interfood Group

Leveraging Technology

Recent advances in technology have changed the way data platforms are designed and provisioned, building on the following principles:

- » **Simple and Straightforward Ingestion.**
The introduction of cloud-based data warehouses has created a shift from extract-transform-load (ETL) to extract-load-transform (ELT) processes. This means minimal transformations are applied to data before it arrives in the data warehouse, which dramatically simplifies the design of a data platform. For example, there is no need for custom data ingestion pipelines anymore, which means that organizations can use standardized ingestion solutions. This results in less development and maintenance effort.
- » **Scalable Data Warehouse Infrastructure.**
In traditional on-premise data stacks, storage and compute are not separated, requiring an expensive stack of computing resources to store data. With the introduction of the cloud, a separation is created between storage and compute, making storage much more cost-effective. These storage solutions are often zero-ops such that no operation capacity is needed for maintaining the underlying infrastructure. Modern data warehouse solutions in the cloud are built on top of these technologies, resulting in a solution that can scale and support the growing demand for data and information with minimal development effort.
- » **Self-Service Analytics Tools.** This tooling is accessible, scalable, and designed to reduce the cognitive load and required skills to prepare raw data and extract its information.

...

“A modern data platform should be simple, scalable, and not limit the end user’s analytics needs. When it follows these principles, it can be an essential tool for value creation within any business.”

— Diederik Greveling, CTO, GoDataDriven Products

Enter the Modern Data Stack

The recent development of tools and services that leverage technological advances further democratizes the ingestion, storage, transformation, and self-service of data assets. As such, organizations can reduce the complexity of setting up and tuning a data platform by choosing a set of tools that together cover the main functions of a data platform. In this way, they can create a fit-for-purpose solution that is easy to implement, maintain, and leverage. We refer to this set of tools as the modern data stack.

Lowering Technical Barriers

Some key elements that distinguish a modern data stack from a more traditional legacy data stack:

- » **Managed and Serverless.** The modern data stack aims to be fully managed and serverless. There is no need to manage custom solution deployments, and computing resources are abstracted away. It’s a preferred solution for companies with small data teams. If they want to reduce dependencies on highly specialized SRE/Cloud/DevOps teams, this solution allows them to manage their own data platform end-to-end.
- » **Low-Code Configuration.** Tools of the modern data stack are straightforward to configure. For example, an ingestion pipeline can be created by simply indicating the type of connector and destination. Low code configuration removes the need to use, e.g., infrastructure as code—a skillset that can be difficult to find.
- » **Out-of-the-Box Integrations.** One of the key features of the modern data stack is the ability to integrate different parts of a data platform by performing a few clicks. For ingestion, tools like Fivetran or Airbyte have built-in connectors to the most popular data sources, which eliminates complex tasks like building and maintaining custom ingestion pipelines. For transformation, dbt integrates with the most popular data warehouses on the market. Finally, self-service tools retrieve data from data warehouses for reporting or data science use cases.

- » **SQL First.** Over the last years, big data processing has been using programming languages such as Java, Scala, or Python, which are skills that are available to highly skilled technical individuals. Conversely, SQL is a language familiar to many data users and has proven itself an easy, versatile language to manipulate data. This gives SQL a special ability: a universal language across data users that democratizes the preparation and accessibility of data. Tools like dbt allow analysts to create, document, and version control new datasets through SQL. As such, analysts can start to own a larger part of the data supply chain, from data extraction to generating the prepared datasets from which information will be derived.

Observe and Monitor Your Data Assets with Little to No Effort

Easy integration allows organizations to effectively map data assets in a data platform, resulting in more products that aim to improve the observability and monitoring of data assets. These products focus on two areas:

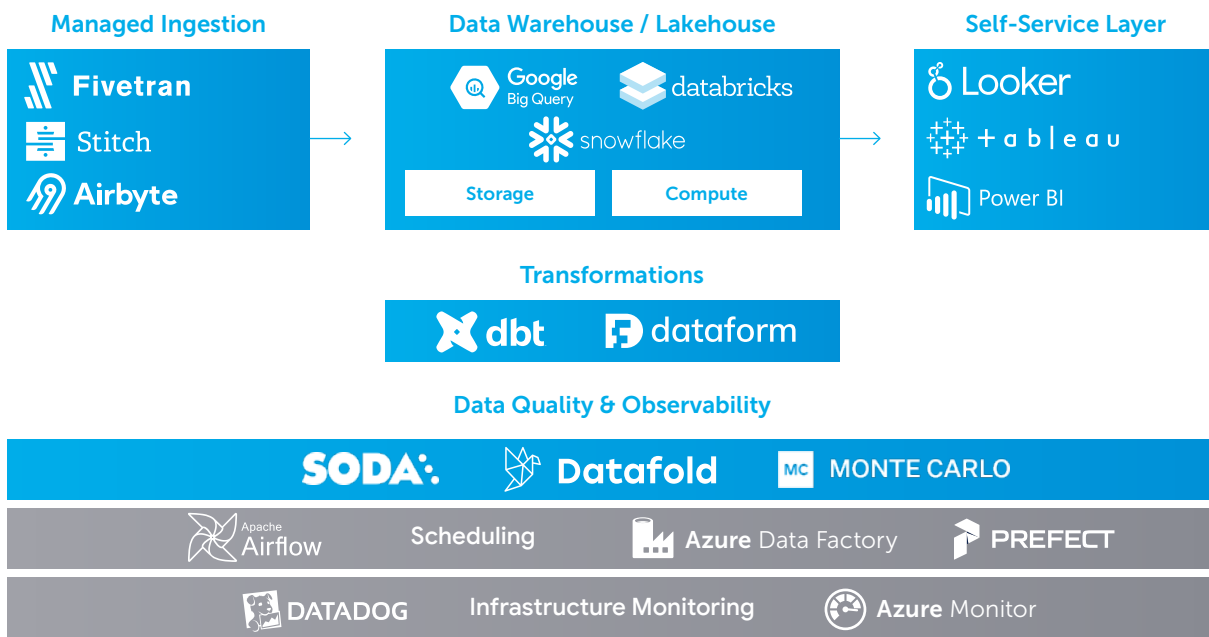
- » A meta data-oriented approach that catalogs data assets, establishes dependencies between them (i.e., data lineage), and indicates data ownership.
- » Monitoring the health of datasets by calculating metrics and setting threshold values for them. Anomaly detection algorithms can also spot unusual behavior in key business metrics, monitor schema drift, or enforce data testing to check primary keys or accepted values.

...

“Scaling up the benefits of data and analytics capabilities through the modern data stack also shifts new responsibilities to the platform team. Where previously data consumers with domain knowledge ensured data quality on demand through manual wrangling, today’s automated pipelines require savvy monitoring and new business processes for governance.”

— Kevin Duisters, commercial analytics manager, Interfood Group

Figure 6
General Architecture of a (Cloud-Based) Data Platform



Note. Shows the primary functions of a data platform: ingestion, storage and compute within the data warehouse, and the self-service area where data and information are consumed. Includes examples of modern data stack tools. Organizations can select and bundle tools together to create a data platform that is easy to implement, maintain and leverage, thereby democratizing its data and information.

When to Choose a Modern Data Stack

Modern data stack tools allow organizations to create a data platform by selecting a bundle of services. The benefits are clear: lowering technical barriers saves engineering time, enabling data professionals to pursue higher-value analytical projects. In addition, a modern data stack reduces the required skillset to maintain data infrastructure and warehousing, making it easier to hire qualified talent. Finally, the modern data stack components are built with end-users in mind, allowing users from all backgrounds to self-service across the data supply chain.

In some cases, a platform may have specific security requirements or needs to provision resources in a particular way. Perhaps your use case requires a specific level of performance to provide value. These could be edge cases where adopting the modern data stack may not cover your needs. In these situations, you may need to provision your own platform by writing custom applications to bundle cloud services together. It's essential to weigh the benefits and drawbacks of entering a field of higher complexity in making these decisions. After all, the skilled professionals you need could be scarce in the market.

3.1.2_Data as a Product

There is a big difference between raw data and high-quality data ready for analytics. Failing to create high-quality, relevant datasets will result in low relevance and value of data products. Still, many organizations delegate the preparation of trusted datasets to a central, relatively small team of data professionals while business domain knowledge is spread across the entire company. The disparity between data and business expertise can lead to high costs in discovering, creating, trusting, and eventually using high-quality data. Moreover, since making sense of raw data and preparing it for usage requires collaboration between both, data teams can become a bottleneck when organizations want to scale the development of data products.

Organizations can democratize the production of prepared data sets by decentralizing the effort. For example, instead of treating data as a by-product of operational applications, product teams can serve up well-structured, documented, monitored data suitable for analytics. These high-quality datasets would be treated as products that the rest of the company could then use.

...

“By bringing data producers and consumers closer to each other, we accelerate the momentum of the data flywheel.”

— Sacha Roggeveen, head of data and analytics, bol.com

Creating data products closer to where business knowledge lives decreases time-to-market and increases data and information availability overall. The modern data stack enables business domains to own their data sets. It does this by lowering technical barriers and offering observability through data discovery, lineage, monitoring, alerting, and quality metrics. A central data infrastructure and decentralized data ownership are key principles of a data mesh, a new domain-oriented data platform architecture. Its purpose is to democratize and scale data and information initiatives to match growing data demands from organizations.²

3.1.3 Data Governance

As data is made available to the organization, a certain level of access control is needed to enable the secure sharing of information to internal and external users. In addition, specific data quality levels must be met for the data to become meaningful and valuable. Hence the need for data governance, which describes the policies, regulations, and standards to make decisions affecting both your company’s and customers’ data.

Balancing Promotion and Protection of Data Assets

Data governance is about creating a balance between data promotion and data protection. On the one hand, data governance promotes data usage to improve and expedite day-to-day decision-making. On the other hand, a data governance program protects the data and ensures that the promotion of data assets is well-governed.

It also aims to protect trust in data by ensuring data integrity, quality, and security. A successful data governance program correctly balances the promotion and protection of data assets. Your data strategy should clearly describe the desired balance, as illustrated in Figure 7. However, using data should always be the primary goal. Data governance initiatives that overly focus on protection by using a command-and-control approach will limit innovation and ultimately meet increasing resistance from within the organization.

...

“The biggest bottleneck used to be data availability. Now it’s the quality and useability of data.”

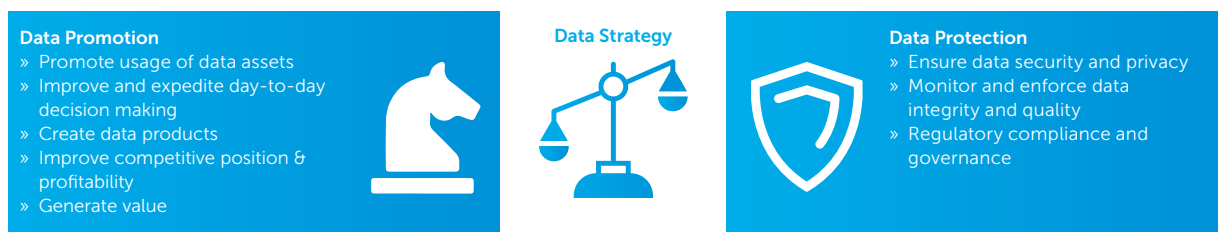
— Sacha Roggeveen, head of data and analytics, bol.com

Key Components of Data Governance

There is no one-size-fits-all approach that creates a successful data governance program. However, we have identified several key components that need to be in place for any data governance initiative to be successful. For a detailed overview, see Table 3.

- » **Data Ownership.** Making roles and responsibilities clear across the organization for the care of data assets.
- » **Data Observability.** Ensuring users can find, understand, and know the quality of data. Data observability enables you to monitor and assure the quality and integrity of your data assets.
- » **Data Privacy and Security.** Knowing and agreeing to the purpose and use of your data assets. Data privacy and security ensure trust in using data assets to their fullest potential.

Figure 7
Data Governance



Note. Data governance enables you to promote the usage of your data assets in a secure manner.

² <https://martinfowler.com/articles/data-mesh-principles.html>

Table 3
Key Components of Data Governance

<p>Data Ownership</p> <ul style="list-style-type: none"> » These aspects include granting permission, maintaining and sharing knowledge, ensuring long term data integrity, resolving data quality issues, etc. » Responsibilities for certain aspects will be divided over different roles » Organize data ownership by agreeing on process flows, roles and responsibilities. 	<p>Data Observability</p> <ul style="list-style-type: none"> » A data catalog contains an overview of all available data assets. The documentation should include the origins and lineage of the data asset, and a detailed description of its history, usage, limitations, pitfalls, meaning, definitions, scope, stakeholders, maintainers, users, dependencies, etc. » Tools are available to monitor data quality and catch quality issues as soon as they arise. 	<p>Data Security & Privacy</p> <ul style="list-style-type: none"> » Data security allows you to be in control of your data assets by controlling access. » Data privacy formalizes whether the intended use of data complies with regulations, company policies, and customer consent. This makes sure data is used according to the intended purpose of it at the time of data collection.
---	--	--

Note. Ownership, observability, and security and privacy contribute to the success of any data governance program.

An Effective Approach to Data Governance

Many data governance initiatives are ineffective, rarely making it up to company-wide adoption. For example, data governance programs are commonly executed by IT, which lacks the mandate and business context to gain support and drive change. A more effective approach involves data governance driven by executives or C-suite members who champion its potential, enforced by automated processes to ensure data quality and integrity. Table 4 describes other impediments to data governance progress and how these challenges can be mitigated.

3.2_Self-Service Enablers

The two main enablers of self-service analytics are *data literacy* and a dedicated *support structure*.

3.2.1_Data Literacy

The emergence and prevalence of data affect many business domains and roles. Today, 75% of the global workforce needs data in their daily work, yet only 21% are fully confident in their data literacy skills.³ This means that there is a significant knowledge gap hindering people’s use of data for efficient decision-making.

Data Literacy

Data literacy is the ability to read, write, and communicate data in context. Organizations can create data literacy programs that focus on bringing analytical capabilities to users from all business functions to close the knowledge gap.

Table 4
Common Causes of Data Governance Failures

Root Cause	Description	Recommended Approach
Frameworks Over Intentions	Implementing comprehensive data governance frameworks tends to become an objective, overshadowing original intentions.	There is no one-size-fits-all approach to data governance. Instead of sticking to predefined frameworks that often feel overengineered, one should create a practical, fit-for-purpose, solution-oriented approach. Prioritize and scope the initiative with all relevant stakeholders to ensure value creation.
Restrictive Image of Data Governance	Although it may feel like it only offers a tight set of restrictions, data governance should enable full autonomy and expanded opportunities within clear boundaries.	Limit the program to restrictions, and you will lose attention and pace. Instead, build upon earlier work, communicate benefits and original intentions clearly and often to get commitment within the organization and accelerate implementation.
Complexity and Scope	Data governance programs touch every business unit, domain, and team, making progress difficult. Implementing data governance with a company-wide scope may be too complex.	Start small. Prioritize data domains based on the value that they can deliver. Focus on a single domain, introducing data governance end-to-end, from data ingestion to information usage, empowering analytical use cases. Learn, adapt the approach, and continue with the next domain.

Note. Common pitfalls and recommended remedies in data governance projects.

³ https://www.accenture.com/_acnmedia/PDF-115/Accenture-Human-Impact-Data-Literacy-Latest.pdf

It is important to appoint people who will push for increased data literacy levels and lead the cultural change within the organization. *Data ambassadors* who passionately ignite change and support others to start using data in their day-to-day work can make a significant difference in success or failure.

Professional development is critical to attracting, nurturing, and retaining talent so they are not lost to competitors.⁴

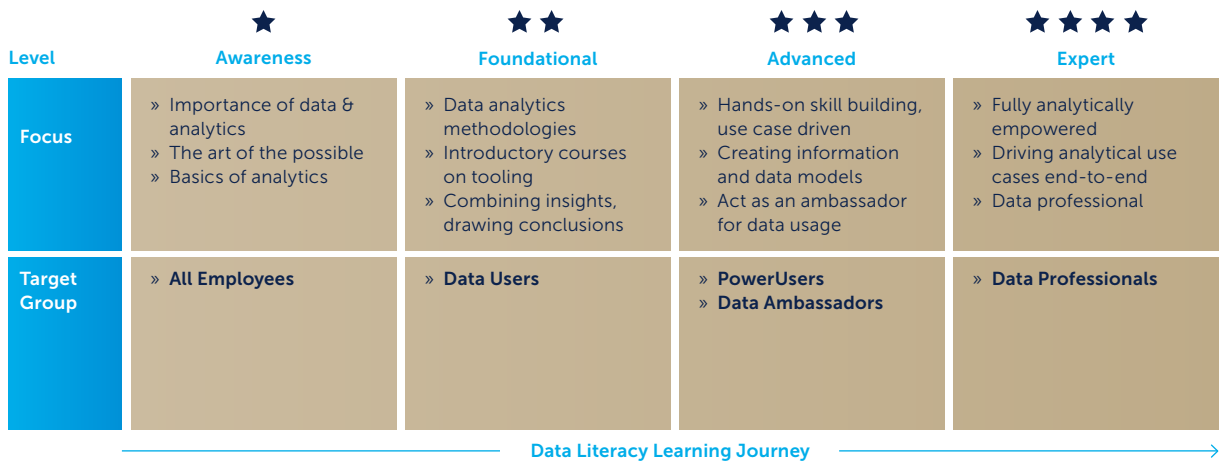
Improving Data Literacy with Learning Journeys

A proven way to improve data literacy is to promote data asset usage through dedicated learning journeys. These allow users to choose the start and endpoint of their learning path based on their background, existing skillset, and current or future responsibilities within the organization. In addition, it provides a framework for analytics leaders to track and stimulate professional development.

...
“Data literacy creates synergies between the data and business domains. It enables professionals to contribute to analytics use cases and embeds data analytics in the business processes.”

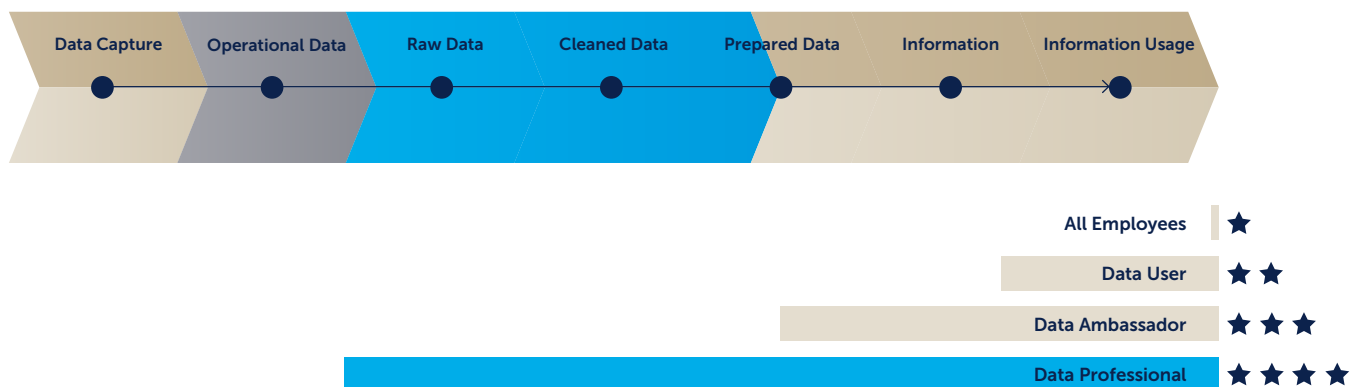
— Bram Ochsendorf, lead data scientist, GoDataDriven

Figure 8
Data Literacy Learning Journey



Note. A data literacy leaning journey allows an organization to match data literacy levels to specific personas based on ambitions and day-to-day work responsibilities.

Figure 9
Data Literacy – Levels and Supply



Note. Directing the correct data towards people with the right data literacy levels in the data supply chain.

⁴ <https://godatadriven.com/topic/data-ai-survey-2021>

Figure 8 illustrates an example learning journey. Here, the awareness level focuses on creating a shared understanding and appreciation of data capabilities and the opportunities (and potential risks) that it brings. This track can be valuable for all employees. A foundational level then introduces modern self-service tooling and analytical methodologies. A potential target group for this track are data users, i.e., people that are currently using data or want to start using data in their day-to-day work. Finally, an advanced/expert track focuses on advanced knowledge of tooling that supports the creation and sharing of insights across the organization. This level can be tailored to data ambassadors or data professionals (i.e., business intelligence, data science, or IT teams) to hone their analytical skillset.

Deciding Data Literacy Levels for All Users

In creating a dedicated learning journey for your employees, it is important to know where people are when they start and where they should land afterward. It should include learning goals for each role in the company touched by the program and enable professionals to identify use cases and leverage data through self-service analytics. As such, learning journeys allow you to decide the necessary data literacy levels for all users in the organization, as illustrated in Figure 9.

Data Literacy and the Data Supply Chain

Companies can link data literacy levels to access management to make data available to appropriately skilled people. For example, an organization could require consumers of reports and dashboards to complete awareness training. However, to change or create information that is displayed in self-service tools, a user may be required to have followed either a foundational or advanced training track. The general idea is that with increased data literacy levels, people can interact with increased parts of the data supply chain. They are encouraged to self-serve data and information. Combining data literacy levels with access management minimizes the risk of unskilled users interacting with your data assets.

3.2.2_Support Structure

More data users mean more questions and requests for support. This extra support tends to come in addition to the responsibilities of IT, BI, or data engineering teams. These teams may already be overstretched with existing backlogs and requests from all business domains. Therefore, mature companies set up *communities of practices* that interact on an ongoing basis. These are groups of people who share experiences and best practices around data-related topics to deepen their knowledge and expertise.

Community-Driven Support

Communities play an essential role in fostering a data-driven culture and ensuring business adoption. They can drive awareness, engagement, and act as a forum where data users and data ambassadors define, refine, and guide analytical use cases towards value.

Communities can offer ad hoc support on data-related questions, such as the availability and quality of data sources and how to interpret elements of data. In addition, technical questions on self-service tools and data literacy learning journeys can be discussed within these communities. In this way, communities act as a first line of support as data users and data ambassadors work together self-sufficiently.

_ PREREQUISITE FOR DATA DEMOCRATIZATION

The Analytics Engineer

As discussed in depth in section 3, the recent developments in technology, the rise of the modern data stack, and the prevalence of self-service analytics have disrupted the field of data and analytics. Out of these game-changing events, a new professional role has emerged in the data field: the analytics engineer.

Analytics Engineering as a Prerequisite for Data Democratization

The shift to ELT-based data warehousing means that data enters a warehouse before it is transformed. This means analysts who have both business domain knowledge and technical skills can now work directly in a data warehouse to transform raw data into trusted datasets to enable analytics across the company. This is where the analytics engineer comes in.

Analytics engineering bridges data engineering and data analysis by bringing engineering best practices towards analytical workflows. As shown in Figure 10, an analytics engineer's responsibilities include:

- » Producing high-quality datasets for reporting, machine learning modeling, and operational workflows.
- » Automation, version control, monitoring, and testing of data pipelines.

- » Improving data observability and maintaining data definitions and documentation.
- » Enabling analysts and business users to leverage tooling for self-service analytics.

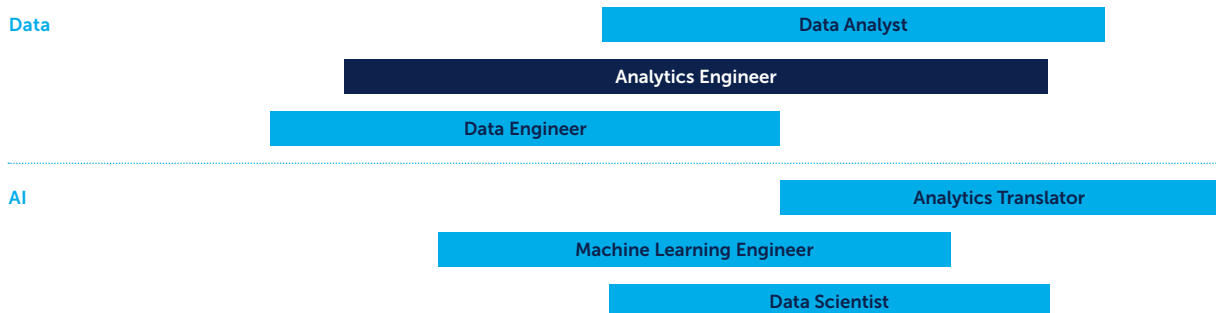
...
"We empower our data analysts with engineering capabilities. Our goal is to enable the power of data democratization by automating the most time-consuming parts of the analytics workflow. The introduction of the analytics engineer is a crucial step."

— Sacha Roggeveen, head of data and analytics, bol.com

Figure 10
 Role of the Data & AI Professional



Data & AI Professionals



Note. Several roles of data professionals mapped onto the data supply chain. The horizontal bars indicate where the role's responsibilities lie, which can differ between organizations. The analytics engineer is the new role in modern data teams that brings high-quality datasets towards an organization while enabling those within the organization to use these datasets for business purposes.

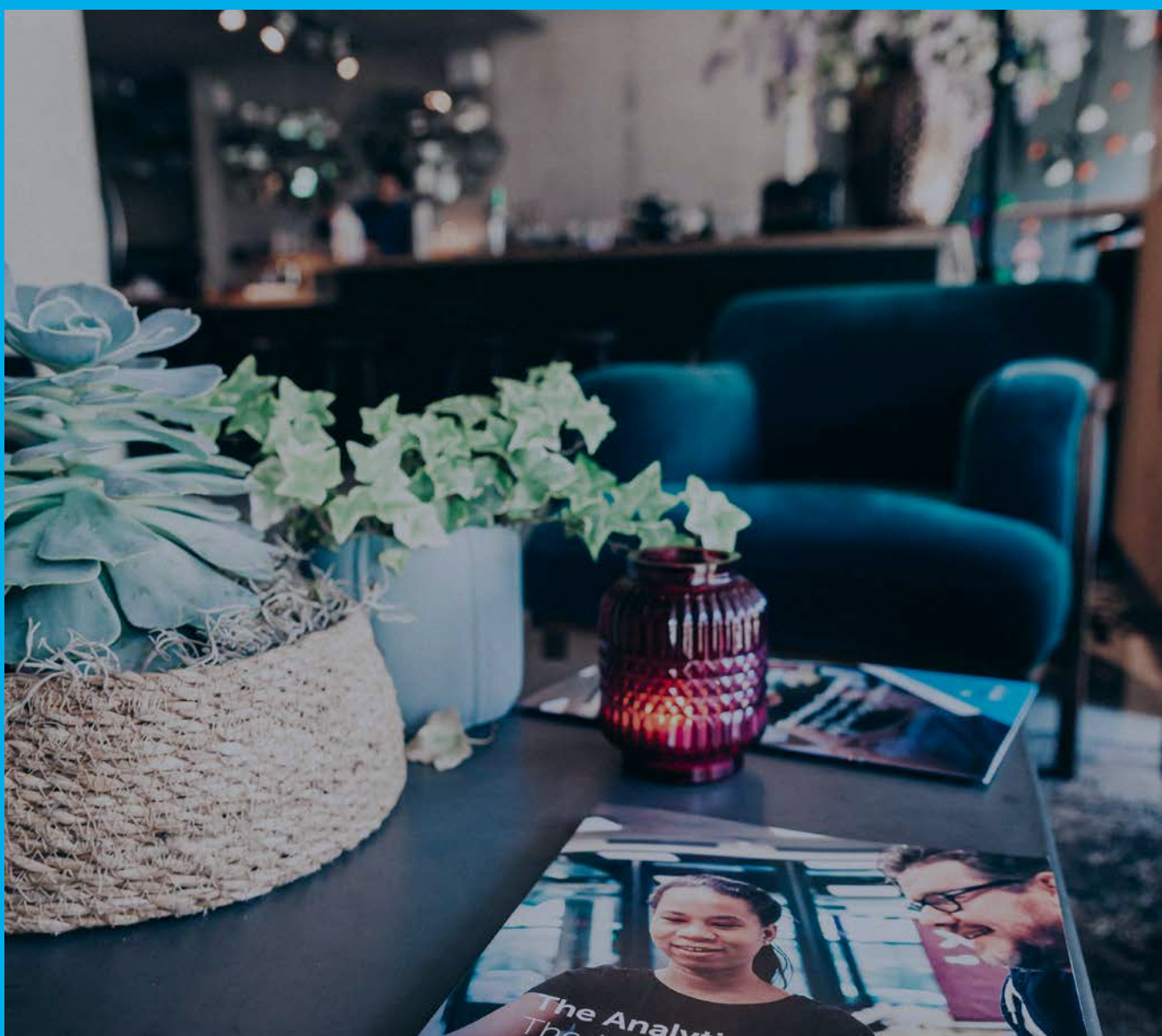
Leveraging modern data stack tools allows the analytics engineer to oversee a large part of the data supply chain. Furthermore, the analytics engineer introduces data observability to improve data discovery and increase awareness of opportunities within the organization. Data observability makes data quality measurable and the root causes for any quality issues visible. The analytics engineer can then implement processes to improve data quality and enforce data governance programs. Finally, the analytics engineer trains analysts and other information users to produce high-quality, trusted data and information to answer their own questions.

By accelerating data and information availability and facilitating self-service analytics across the data supply chain, the analytics engineer plays a crucial role in helping organizations discover and realize the value of their data assets through data democratization.

...

“The analytics engineer makes it possible for people to use data to generate value. Whether consumers need new tooling to access new datasets or require training on how to self-service from data consumption points, the analytics engineer meets their needs.”

– Guillermo Sanchez, analytics engineer,
GoDataDriven



_ REALIZE THE VALUE OF YOUR DATA TODAY

The Road to Data Democratization

So far, we've covered the benefits, risks, and key enablers of data democratization, as well as the role of the analytics engineer. This section explains how to discover and realize the value of your data, starting today.

The road to data democratization begins with defining your company's current maturity and future ambitions. Executive support is crucial to determining the organization's vision. They must make sure the necessary funding is available, appoint roles and responsibilities to facilitate the organizational change and clear roadblocks. Learn more on how to ensure business adoption of data and AI in our AI Maturity Journey Whitepaper.⁵

A data strategy provides a concrete roadmap for building and establishing the right enablers for data democratization and outlines the plan to raise data literacy and awareness across business domains. At GoDataDriven, we help our clients design and execute their data strategies along three tracks, as illustrated in Figure 11: organize, build, and train.

Track 1. Organize

- » **Know where you are and want to go.** Our GoDataDriven Data & AI Maturity scan assesses your current data maturity and provides practical recommendations for advancement.
- » **Establish your data strategy.** With GoDataDriven Strategy, we guide your organization in balancing objectives, budget, and capabilities and translate your strategy into business solutions.

...

"A clear value creation plan for tech investments is critical to change traditional industries successfully. We rely on two principles to support our journey. First, we aggregate use cases to larger programs that share the costs and risks of embracing new technology. Second, our center of excellence leads the way to value in a targeted vertical before democratizing the full product. In practice, this means we nurture business change step-by-step while making big moves in our stack."

— Kevin Duisters, commercial analytics manager, Interfood Group

Track 2. Build

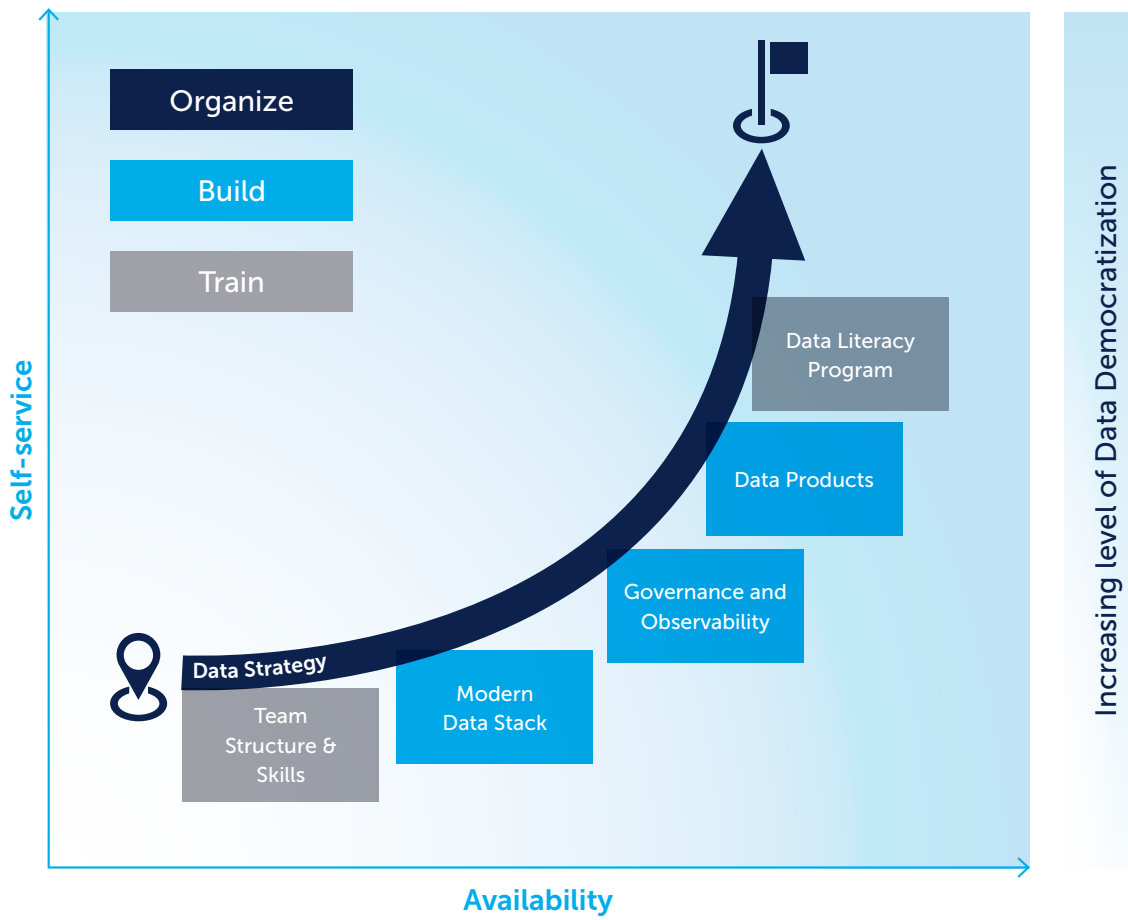
- » **Set up the technology to meet the demand.** Our data platform offering provides the proper cloud infrastructure, platform, and a modern data stack to serve your current and future data demand.
- » **Introduce the analytics engineer.** Add this new role to your data team to implement use cases, accelerate data and information availability and facilitate self-service analytics across the data supply chain.

Track 3. Train

- » **Find the right mix of people and skills.** Effective execution of your strategy depends on having the right people on board. Our analytics engineers will train in-house talent in making data available to the organization and leverage it for business purposes.
- » **Start leveling up your data literacy at scale.** The GoDataDriven Academy offers dedicated learning journeys for all business users to improve data literacy across your organization.

⁵ <https://godatadriven.com/topic/whitepaper-ai-maturity-journeys/>

Figure 11
 Three Tracks on the Road to Data Democratization



Note. Organizations can organize, build, and train to democratize data.



Interested in realizing business value through data democratization? Please contact Bram Ochsendorf, lead data scientist (signal@godatadriven.com).

Meet the Experts

_ **Bram Ochsendorf**

Lead Data Scientist _ GoDataDriven

Bram supports organizations in their road to data democratization. He organizes data teams, builds end-to-end data products, and enables businesses to realize value from data and AI.

_ **Sacha Roggeveen**

Head of Data and Analytics _ bol.com

Within bol.com, Sacha has founded the data science department and automated quality assurance. He combines an approach of short iterative sprints with proactively involving project leads, engineers, data scientists and senior leadership.

_ **Kevin Duisters**

Commercial Analytics Manager _ Interfood Group

As global data engineering and science practice lead, Kevin actively contributes to the formation and execution of data strategy, including democratization and change. He has a background in consulting and academia, mainly in Europe and the US.

_ **Eduardo Neves**

CTO _ Funda

For the past few years, Eduardo has shaped the structure, focus, and strategy of Funda's significant technical developments. With a background in software engineering and an appetite for continuous learning and knowledge exchange, he successfully guides his teams to reshape and strengthen Funda's platform and replace ineffective business-critical tools and processes.

_ **Diederik Greveling**

CTO _ GoDataDriven Products

Diederik delivers easy-to-maintain data and cloud-oriented products that quickly create value. He has an engineering background and focuses on distributed systems, cloud computing, and data architecture design.

_ **Guillermo Sanchez**

Analytics Engineer _ GoDataDriven

Guillermo has a pragmatic and business focused approach to data analytics. His skills range from designing modern data platforms, modelling data warehouses/lakes, and writing production-grade code for data pipelines.

_ **Arjan van den Heuvel**

Strategy Lead _ GoDataDriven

Arjan advises and supports companies in their data and AI Maturity Journey, which includes strategy design and execution, scaling data and AI capability, and generating more business value through the process. He brings over 15 years of experience in consulting and analytics to his role.



GO DATA
DRIVEN

PROUDLY PART
OF XEBIA

Wibautstraat 202
1091 GS Amsterdam
signal@GoDataDriven.com
+31 35 53 81 921

GoDataDriven.com