

In 9 steps to

DATA INTEGRATION AND A DATA-DRIVEN ORGANIZATION

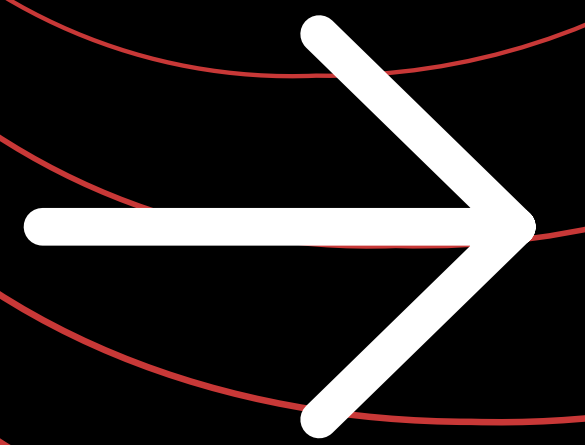


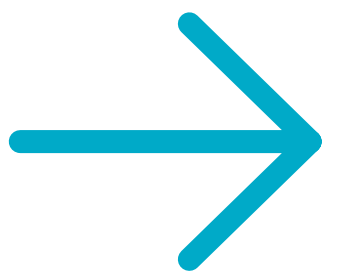
Table of contents

Introduction	page 3
1 What is the added value of a data-driven approach?	page 4
2 Why do data projects fail so often?	page 9
3 How do I go about a data project?	page 13
4 How can Conclusion help?	page 16

"This is how to extract value from data through proper integration"

As organizations increasingly seek to become data-driven, their ambition to collect data in the hope that it will lead to valuable insights grows.

With the right approach, collecting more data can indeed result in deeper and more relevant insights. That's why it's so crucial to also look at the quality of the data. By being strategic with your data and focusing on what really matters, you can avoid drowning in your data lake. This whitepaper offers valuable insights and practical tips to get the most out of your data and actually create added value.





1

WHAT IS THE ADDED VALUE *of a data-driven approach?*

Let's start by providing a definition of a data-driven approach, because this concept is not always clear. Applying a data-driven approach means that you make decisions and manage affairs based on facts in order to achieve your strategic goals. The result is that the organization as a whole works more efficiently and effectively: the quality of the work improves (the chance of making errors decreases), the lead time for completing processes is shortened and you can save costs on your resources. So there are plenty of reasons to get started with this concept.

1 What is the added value of a data-driven approach?

"A data-driven approach requires an integrated method that includes business, processes, organization and technology. Effective data integration and flexible data and analytics capabilities support a modern, digital way of doing business. They ensure that you can work efficiently and effectively and that you are compliant, with full control over risks and the ability to create real value."

Growth path of traditional analytics and a data-driven approach

Applying a data-driven approach as a concept is something of recent years, but has in fact been around for a very long time. It has been a growth path from traditional BI, using retrospective reports to look back to the advanced predictive modelling and real-time decisioning used today.

In that growth path, more and more sources were gradually unlocked. The structured data warehouse was replaced by a data lakehouse, which also allows for the storage of unstructured data. Moreover, the data is no longer always stored centrally (where you make a copy of the data), but is increasingly only made centrally accessible. If data *is* still stored centrally, this no longer always involves a daily dump, but more and more streaming, thereby allowing for real-time decision-making. Developments on the static model side are moving fast as well, for example the arrival of generative AI. In short, technology has developed considerably over the years. This at the same time enables new applications.

Eneco

How can we respond quickly to peaks in supply and demand using sustainable resources?

A major challenge on the energy grid today is balancing supply and demand. Eneco therefore constantly asks itself the following question: how can we respond quickly to peaks in supply and demand using sustainable resources? The variables you can control are mainly the supply of energy from solar and wind and, on the demand side, large industrial processes such as e-boilers and batteries. In the 'Grid Smartening' programme, Conclusion helped Eneco build a data platform in the Azure cloud that monitors the status of assets in real time (per second). The platform also provides data on weather conditions and forecasts. Based on IoT and other data, Eneco uses smart algorithms to monitor and control wind and solar parks, e-boilers and heating networks in great detail, and also plan maintenance more meticulously (on cloudy days with little wind).

1 What is the added value of a data-driven approach?

Importance of master data management

In many articles you read on this subject, the emphasis is on the technology; on a data lake in the cloud; on technologies to unlock data from different sources; on artificial intelligence (AI) and machine learning (ML).

The real challenge, however, is not technical in nature, but rather semantic. When we're talking about sales price, are we talking about the price including or excluding VAT? And is that including or excluding shipping? And what currency is used? When we say customer, are we talking about everyone in the CRM system, those who placed an order, or just those who also paid? The right definition is key in working with data. This challenge is also referred to as master data management.

In recent years, we are also witnessing increasingly strict requirements being imposed on data governance. Different ways of using data simply place different demands on data quality. How do you keep control over this? In the next chapter we will delve deeper into the challenges that all this brings.

Zorg van de Zaak wants to reduce sickness absence rates.

Medical examiners play a crucial role in this, but sky-high workloads prevent them from monitoring and guiding everyone calling in sick equally closely. That's why Conclusion helped this company develop a predictive model: for which employees does guidance offer added value and for which not? For employees who are on the verge of emotional exhaustion, timely guidance makes the difference between sustainable employability or being out of work for a year due to burnout. In collaboration with the medical examiners, Conclusion determined which data are predictive in this regard. Subsequently, a model was developed that sends an alert to medical examiners when an employee who is at risk of long-term sickness absence calls in sick. As a result, medical examiners now know better understand which employees they should pay what attention to, resulting in lower sickness absence rates.

1 What is the added value of a data-driven approach?

Importance of data integration

Data integration is often overlooked when it comes to business improvement by applying a data-driven approach.

Projects often focus on collecting (high-quality) data and building reports and algorithms to draw conclusions. These conclusions almost always lead to an action. And you preferably want to carry out that action automatically. You want to automatically transfer an order from the CRM system to the ERP package. Eneco wants an adjusted weather forecast to automatically lead to different decisions about scaling up or down wind or solar parks.

In practice, we can see that many organizations still update information from dashboards or calculation results manually or through uploads. While real-time data integration can contribute to business improvement in various ways:

- **Effectiveness:** integration ensures automatic updating of data in the different systems, so you always have real-time insight;
- **Ease of use:** data integration means that everyone always looks at the same data, making the data extremely useful;
- **Compliance:** data integration can provide you with instant insight into the extent to which your organization complies with laws and regulations;

- **Efficiency:** you no longer have to retype data, which saves a lot of time and prevents the risk of errors. This at the same time prevents any time-consuming rework;
- **Risk reduction:** you always have the right information at hand when you need to make decisions. Too often, decisions are still made on the basis of data that is not up to date or incorrect, or simply on the basis of a gut feeling because the right data is missing;
- **Innovation power:** having an integrated view of your organization's current data strengthens your ability to develop new business concepts.

"Data integration is often overlooked when it comes to business improvement by applying a data-driven approach."

1 What is the added value of a data-driven approach?

Three forms of insight from data

We distinguish three ways of fact-based management:

- **Descriptive:** What happened? This involves you describing the current situation and looking back. The output is often generated in the form of a report or dashboard. Financial reporting is a common form of this type of data-driven working.
- **Predictive:** what can happen? This involves predicting the future based on data. For example, if you know how many orders you need to process next week, you can adjust your staff planning and the deployment of other assets accordingly. Predictive maintenance is a good example of this form of data-driven working.
- **Prescriptive:** what can I do to tackle a specific problem or improve performance? This is the most difficult form, because this involves you generating insights (real-time if desired) that enable you to make better decisions. Scenarios provide insight into what happens when you change a certain variable you can control. You can use this form to solve an acute problem: 'A machine has stopped unexpectedly. What do I do to get it up and running again as soon as possible?' But you can also use prescriptive applications for your strategic planning: 'Our warehouse is reaching its limits. Is it wise to build new premises at the current site or would it be better to open a second warehouse at a completely different location and save on transport costs?'



2

Why do

DATA PROJECTS FAIL SO OFTEN?

A study by Conclusion shows that more than a third of the data that organizations collect is not being used, i.e. gathering dust somewhere. Not only does this involve unnecessary costs, it also often means that this data obscures the data that is actually important for decision-making. As many as 42 percent of respondents in this study indicate that they do not always gain the necessary insights due to the large amount of data they have at their disposal. And 27 percent say they don't always have access to the data they need for their business challenge.

2 Why do data projects fail so often?

It is therefore not surprising that only 39 percent of respondents say that the contribution to business objectives is higher than the costs. A total of 40 percent say that the costs and benefits can be offset against each other. And 21 percent even indicate that costs exceed the benefits.

Various causes

All this leads to the question: how come organizations achieve so little returns from their data-driven projects? The causes are diverse.

- A lot of unused data is stored 'because it might come in handy in the future'. Nice idea, but in practice it almost always turns out that the data in its stored form is not useful for any applications that are added later. If you collect data without a purpose, you cannot determine which measurement frequency is required, what level of detail the data should have, etc. Our advice is therefore: first determine the goal and only then start collecting data.
- Organizations want quick insight. This leads to projects that are too simplistic. It's important to very carefully unravel the need, i.e. to define the question behind the question. Only then can you determine which data is needed and in what form in order to create an answer.
- Data quality is often poor and visibility into data quality is lacking. A customer being entered into the CRM system three times, with three

different records, is not uncommon. So which one is correct? And what to think of IoT data, which does not always come through properly due to a poor connection. Or typing errors that can have major consequences (e.g. a comma in the wrong place in a number).

- Little attention is paid to master data management. The term 'price' has different meanings depending on the system, but that is not always noticed when that data is merged into a data lake.
- The timeliness of data does not always match the purpose for which you want to use the data. If a planner schedules service technicians based on their certifications, but HR has a backlog of recently obtained or expired certifications, the wrong people will be sent to the job.

Nutreco *Animal feed producer Nutreco initiated a data strategy with the aim of extracting more value from internal and external data sources.*

An important wish was to respond more quickly by not only receiving reports after the end of a month, but through dashboards refreshed daily. Conclusion helped design a data platform and data products that use this platform. A data market has been created that is used by many Nutreco employees in descriptive, predictive and prescriptive applications.

2 Why do data projects fail so often?

- Not all data in the systems is representative or contributes to a reliable analysis or prediction. Take the coronavirus crisis. In one industry, sales went through the roof (for example, the DIY sector), whereas in another, demand completely disappeared (hospitality and retail). This doesn't mean that the data from this period is unusable, but it does mean that this data doesn't say much about averages. You need domain experts who can indicate which data is useful in your models and in what form.
- Domain expertise is also needed to identify cause and effect. If you use AI to look for patterns in data, there is a good chance that you will find connections. But the AI model cannot indicate what is cause and what is effect, only that there is a connection. The connection found may not even be causal. If data specialists and business experts work well together with domain experts, you avoid drawing conclusions such as: we sell more ice cream, so the weather will probably be nice.
- Bias can easily creep into the models. Bias is an (often unconscious) bias that can lead to AI systems making decisions based on prejudices. Bias can arise from the way AI models are trained, the way algorithms are programmed or the context in which they are applied.

A woman with dark curly hair, wearing blue scrubs and a lanyard, is smiling and looking at a tablet computer. The background is dark with warm, out-of-focus bokeh lights.

An academic hospital which specializes in complex abdominal operations is developing a predictive model that indicates when patients can be safely discharged from the ICU.

If this model is deployed without further adjustments in a general hospital, with a different patient population and a different type of surgery, it will most likely not perform as well, because the different patient population and the different type of surgery will introduce bias into the model.

2 Why do data projects fail so often?

- More and more companies are including external data in their analyses in addition to internal data. They have no control over the data quality when using external sources. You just have to assume that the data is correct. Moreover, you don't know whether changes are taking place that will give the data a different meaning. For example, a different measurement method that is used, which means you have to interpret the data differently.
- You will need to continuously monitor the performance of your models, as changing circumstances can easily introduce bias into the models. Therefore, when setting up governance, make sure that you also pay attention to changes that could affect the predictive value of your models.
- Different data sources often speak different languages. In one field the degree of standardization is high (e.g. in the financial world), while in another there is still room for improvement (e.g. healthcare or transport & logistics). In worlds where many proprietary data formats are still in use, it's extremely difficult to combine data from different sources. The only solution is standardization. Fortunately, the European Data Spaces initiative has been set up for this purpose: EU legislation that will encourage software suppliers to work with open standards. However, we have not yet reached the point where organizations can already reap the benefits of these laws and regulations.

- Applying a data-driven approach must become part of the corporate culture. And that is not yet the case in all cases. Fortunately, the study by Conclusion referred to above shows that things are moving in the right direction. As many as 81 percent of larger organizations have secured data ownership and data policy at board and/or management level.

Water boards were traditionally accustomed to managing based on averages.

In the models controlling pumps and other assets, extremes were classified as exceptions and ignored. Now that extremes have become normal due to climate change, models need to be adjusted. After all, you don't want to drain water from a downpour during a dry period as quickly as possible, as we used to, now you want to retain it.

Pension provider APG has secured data policy and ownership through a data shop.

The shop contains a data catalogue that APG employees can use to request a data product. The request is sent to the data owner, who can accept the request, ask additional questions or reject the request.



3

How do I go about **A DATA PROJECT?**

As stated previously, gaining insights from data allows you to work more effectively and efficiently. You can make this as big or as small as you want, but our advice is to start small and expand the solution over time.

3 How do I go about a data project?

■ **Step 1.** ***Definition of a business issue***

Define the business issue and preferably be as specific as possible. For example, if your production capacity is reaching its limits, but you don't yet have a permit to build a new factory, you will want to increase the occupancy rate of your machines. In that case, you first analyse what the main reason is for machine downtime: is the changeover time too long? Is there a lot of unplanned maintenance needed due to breakdowns? Are there too few operators who can operate the machines? Start your project by tackling the root cause of your business issue.

■ **Step 2.** ***Determine what data you need and in what form***

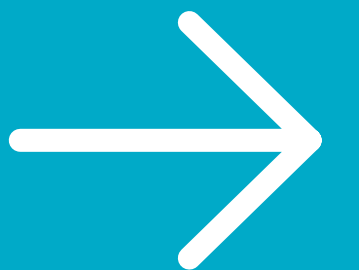
In the case of the factory, for example, this is a high number of breakdowns, which means that machines are down requiring unplanned maintenance. You then analyse: what data do I need to predict breakdowns and adapt my maintenance strategy accordingly (predictive maintenance)? In what form do I want to present that data? How up to date should the data be? Do we need to use streaming data or is refreshing the data once a day sufficient?

■ **Step 3.** ***Research the required data sources***

Only when you have answered all these questions does the question arise as to which sources contain the data relevant to this issue. For each data source you check: how is the data structured? Am I dealing with raw data, combined data or enriched data?

■ **Step 4.** ***Investigate the data quality***

The next question is: what is the quality of the data? It happens all too often that different systems present different 'facts'. So which system is right? It's important to always involve someone with domain knowledge when improving data quality.



3 How do I go about a data project?

■ Step 5.

Ensure real-time integration of current sources

Once you know which sources you want to extract data from and have insight into the quality of the data, you can start building integrations with those sources. After all, you don't just want to extract data from the sources, you also want to create a direct feedback loop to those systems. The aim is to make automated decisions, where possible.

■ Step 6.

Set up master data management

Together with your data quality research, you can make decisions in the field of master and reference data management. What definitions do you use for certain terms? What classification do you make? What is the hierarchy between data sources?

■ Step 7.

Organize data governance

Once this structure is in place, it is important to keep it in order. That's why you will need to set up data governance. Each data source needs

to have an assigned owner who oversees issues such as data quality and the correct use of master data. You will also need to set up processes to determine who has access to which data, which is especially important for applications in which the GDPR plays a role.

■ Step 8.

Build data products

The construction of the models is one of the final steps in the project. Starting with this earlier is tempting, but we cannot warn enough against this: make sure you first go through the previous steps properly, otherwise there is a high risk of you building a data product that has no bearings on the business issue.

■ Step 9.

Ensure adoption

You can build the most beautiful predictive models and come up with the most beautiful prescriptive applications, but they are worthless if employees do not trust the data. Therefore, make sure you start thinking about adoption from step 1.

NS

The NS (Dutch Railways) wanted to enable station managers at Utrecht Central Station to make automated decisions based on data.

Conclusion helped NS develop a digital twin of this station. The aim of this project is to increase safety, prevent defects in assets such as lifts and escalators (predictive maintenance) and improve passenger comfort.

To build the digital twin, Conclusion, together with the IT team of NS Stations, unlocked all data sources that provide information about the status of objects at the station. Examples include lifts, escalators, passenger flows and energy consumption. Based on this data, a 3D model was built that consists of no fewer than 20 billion intricate LIDAR data points and 540 high-resolution drone images. Together with the real-time streaming data, this has been processed into multiple models that are used to display the situation at the station and to simulate scenarios. This provides station managers with the tools to plan maintenance on lifts and escalators in such a way that the impact on passenger flows is minimized. Or to detect unwanted behaviour from travellers more quickly and intervene in panic situations. In addition, NS has more insight into energy consumption and waste and peak loads, so that CO₂ impact can be reduced.

4

How can

CONCLUSION HELP?

Conclusion is a dynamic ecosystem of over 25 expert companies working together on business transformation and IT services. Through close mutual cooperation, we offer clients a one-stop shop for all their issues. We approach every client request from the perspective of the ecosystem: which company or combination of companies can best address their needs? This approach gives clients the benefit of a one-stop shop combined with in-depth expertise on sub-topics.

Conclusion Intelligence

Conclusion Intelligence bundles the forces of six Conclusion companies that, with the help of data and advanced technologies such as AI and Gen AI, aim to achieve business improvement and innovation. Together these companies work every day to realize the ambitions of our customers.

Quick scan of the data value chain

Conclusion Intelligence has used its many years of broad experience in performance improvement and fact-based management to map out all crucial aspects for creating a data-driven organization. We call this the data value chain. By asking clear-cut questions about all aspects of the data value chain, we use a quick scan to show you where you stand and where the potential for improvement lies. Based on these insights, we support you in improving this chain.

LEAD THE CHANGE, BEFORE CHANGE LEADS YOU

Get in touch at
info@conclusion.nl

This white paper was made possible with the cooperation of the following experts in the Conclusion ecosystem:

Robbrecht van Amerongen

Head of Strategy AMIS Conclusion
LinkedIn: [Robbrecht van Amerongen](#)

David Buschman

Business Development Director Hot ITem Conclusion
LinkedIn: [David Buschman](#)