



Talan<sup>★</sup>

# DATAOPS

**Building better data strategies  
for organisations**

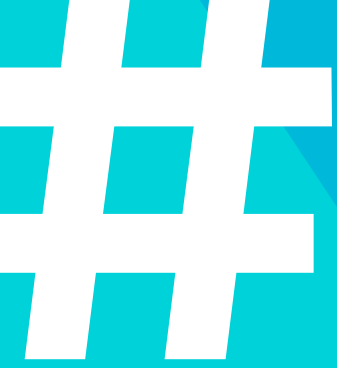
## EDITORIAL

Gartner defines DataOps as “a collaborative data management practice focused on improving the communication, integration and automation of data flows between data managers (data engineers, data architects and data stewards) and data consumers (data scientists, business analysts, business lines or other departments) across an organisation.”

DataOps uses technology to deploy and manage data delivery automatically, effectively orchestrating your project.

Consequently, DataOps is guaranteed to improve the quality of data projects and optimise their delivery cycle, ensuring a faster analysis and prediction time-to-market.

Find out how the Talan approach to DataOps can help you with your company's data projects.



# A brief idea of DataOps

## A brief idea of DataOps

Today, every organisation, regardless of its business and size, knows it is **vital to industrialise and optimise its data use**.

There are several “types” of data:

- Data that already belongs to the organisation. This asset is often underutilised, requiring the necessary knowledge to fully exploit it.
- Data produced through its commercial, production or service activities.
- Data exchanged with its partners and customers.
- All the data available to it on the Internet.

An organisation’s ability to benefit from all its valuable information resources by transforming them into added value is a decisive factor in its digital transformation strategy. **Control over its data therefore gives the organisation a competitive advantage**. It can develop new services, products and methods as well as fields for improvement, development, diversification or innovation. A wide range of possibilities, in fact!

**“99.5% of all data collected is never used”**

## Data in 2020

Data is being created worldwide at unprecedented, ever-increasing speed. Imagine — more data was generated during the last two years than ever before in history. In 2020, every person on the planet will create enough data to fill a 256 GB hard disk — every two days!

Amazingly, although Big Data has created nearly 15 million jobs worldwide since 2012, **99.5% of all data collected is never used or analysed**.

And yet, this data holds such great potential that analysts estimate that Fortune 1000 companies could generate an additional 65 million dollars in net income simply by having access to 10% more of it.

## From DevOps to DataOps

Let us be clear: DataOps is not simply a matter of transposing DevOps (a familiar aspect of IT development) into data operations. Firstly, it **improves both the data’s quality and its analytical and operating cycles** by setting up new tools and methodologies, and secondly, it **completely rethinks the optimisation, working structure and communication of everyone involved**. This is where parallels can be drawn with DevOps.

Since the emergence of Data Analytics, data experts have designed models for extracting usable, useful information from diverse big data. However, this puts pressure on IT Operations because **companies increasingly depend on this analytical data and need it in production rapidly**.

At the same time, they must **provide guaranteed quality, relevant and highly trusted data** while also complying with constantly changing legal and ethical **regulatory requirements**.

Market volatility, world unpredictability—who could have foreseen that a virus would disrupt the world economy in 2020?—, responsiveness of competitors and client requirements: all these factors oblige companies to adapt to the market.

Adopting all DataOps practices is essential for companies to improve coordination between their Data and Operations teams. The challenge is simply stated: **to reduce the time taken from identifying and analysing the data to producing usable models that can be used to create value and make decisions**.

**“It is not only a question of producing the next data analysis. It is also a matter of building a system that can produce more analyses more quickly.”**

Consequently, DataOps is based on the idea that in order to create value rapidly, **the data must be rapidly and continuously detected, integrated, cleaned, stripped of duplicates, classified, aggregated, indexed, referenced, modelled, shared and governed** so that teams throughout the company can extract valuable information from it. This information will then be used either to make decisions or as the basis for action.

## Three main principles

Of course, DataOps is not an exact science and so no single method exists. The DataOps projects that will be used to support your company’s strategy may adopt different approaches. However, all these approaches are equally **vigilant regarding the necessary governance**. Many companies design and adopt “Centre of Excellence” (CoE) approaches for their data strategy, which combines business, strategic, technological and ethical requirements.

That said, the following essential principles must be applied to produce quantifiable results for the company.

### Principle 1: Democratisation

All recent studies tend to show that business users within the company say they need greater access to a wide range of data. According to Experian Data Quality<sup>1</sup>, this is confirmed by almost all (96%) Chief Data Officers (CDOs). Despite the weight of these demands, however, 50% of them say that their companies do not make sufficient use of the data to enable them to unleash the full potential value of this information. The CDOs say that their main priority is to **provide greater access to the data while guaranteeing its quality and governance through a secure infrastructure geared towards self-service access**. As it is not enough merely to unlock access to the data, the governance processes must be maintained throughout their life cycle.

<sup>1</sup> [The Chief Data Officer, Powering business opportunities with data. Source: Experian](#)

This enables the business lines to trust the value of the data available to them and implement their digital transformation programmes responsibly and effectively.

Any factors hindering data access, along with data confidentiality and more generally, GDPR compliance, are likely to limit any digital transformation projects.

### Principle 2: Automation

To exploit the data more quickly and effectively in projects requiring big data, it is essential to **automate any time-consuming, tiresome and repetitive steps** that occupy the teams unnecessarily. For example, data management professionals worldwide spend between 60% and 80% of their time in formatting, cleaning and sometimes requalifying data. At the end of the processing chain, **the data must meet stringent validity, consistency, integrity, accuracy, relevance, accessibility, life-cycle and up-to-dateness criteria**. Its use must also comply with the company's data format or standardisation criteria and data ownership and confidentiality guidelines.

The **use of Robotic Process Automation (RPA), Virtual Assistant** and, more generally, **Artificial Intelligence technologies** enable the company's teams to concentrate on high added-value tasks.

### Principle 3: Data platform

Market forces oblige business lines to implement data-driven projects. These companies hire CDOs who participate in their company's digital transformation through actions that include conducting a true data strategy to achieve the corporate objectives.

The implementation of DataOps, with its continuous integration and continuous deployment (CI/CD) chain, requires a **technical platform open to the data environment** with tools such as GitHub, Docker and Jenkins. This platform must provide guaranteed data quality and security and be geared towards self-service access. It must be able to perform processing or calculations that involve big data as well as containerised processes.

The data platform must never be used to store data solely reserved for a few data scientists or data engineers. **Data is a potentially highly valuable corporate asset**. Consequently, the data platform must instead be used to **share information available to all users** for the company's greatest possible benefit.

**“The objective is not only to answer your customer’s question but also to be able to quickly answer all the questions that follow it.”**

## DataOps and effectiveness

Companies do not set up technology “for technology’s sake”, but to meet a precise objective instead... This reduces technology to a tool serving their strategy. The implementation of DataOps must form part of an approach intended to achieve this goal. But how can we measure its effectiveness?

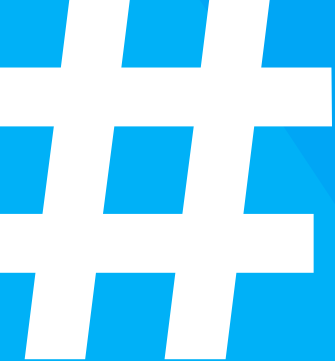
Its effectiveness is measured by means of **concrete indicators** mainly relating to team collaboration, error rates, increased productivity, deployment speed, successful testing and the meeting of deadlines.

Team work appears within virtual collaborative workspaces, enabling the level of collaboration to be monitored. Production data errors are also monitored and must constantly decrease until they reach a minimum limit. **This monitoring is based on data quality systems and data visualisation and analysis tools.** Conversely, team productivity is monitored to ensure that production deployment is optimised. All these factors help to ensure that SLA objectives are met.

The most usual graphical representation of DataOps is in the form of an infinite loop that we will describe later: Envision, Define, Plan, Design, Develop, Test, Deploy, Monitor, etc. The **principles of agility and continuous improvement**, as well as the very essence of data projects, underlie the iterative concept.

The expertise of the teams that will implement the methodology and data platform on which your DataOps project will be based will enable you to fully benefit from your project.

In the second part of this document, we will present the Talan approach by referring to DataOps projects that we have undertaken alongside our clients.



# DataOps through the lens of the Talan approach



## DataOps through the lens of the Talan approach

At Talan, a team in our Data Intelligence Solutions division has acquired considerable DataOps project management expertise. One of the lessons we have learned from managing and implementing these projects is undoubtedly that most of our clients are now fully aware of the **need to ensure that the functional specifications are consistent with what the business lines say they need**. They fully understand that the use of agile methodologies encourages and maintains this consistency.

Today, we have identified a number of **points likely to complicate the adoption of their solutions** within the company if the client's teams overlook them. These points include the following:

- Regressions of a known, stable perimeter
- Comprehensiveness and traceability of tests

- Technical delivery security
- Compliance with development and architecture norms and standards

**Talan is convinced that the basic concept of DataOps provides a concrete, operational solution to these points.**

As we saw in the first part of this document, DataOps addresses the issue of coordination between the data managers in general — and data scientists and data engineers in particular — and the data systems and administration teams.

It reduces the solution time-to-market to **meet business line demands as rapidly as possible** and especially, effectively cover the critical aspects of projects, including the following:

- Short development cycles
- Guaranteed continuous integration process
- Project exploitation in continuous deployment

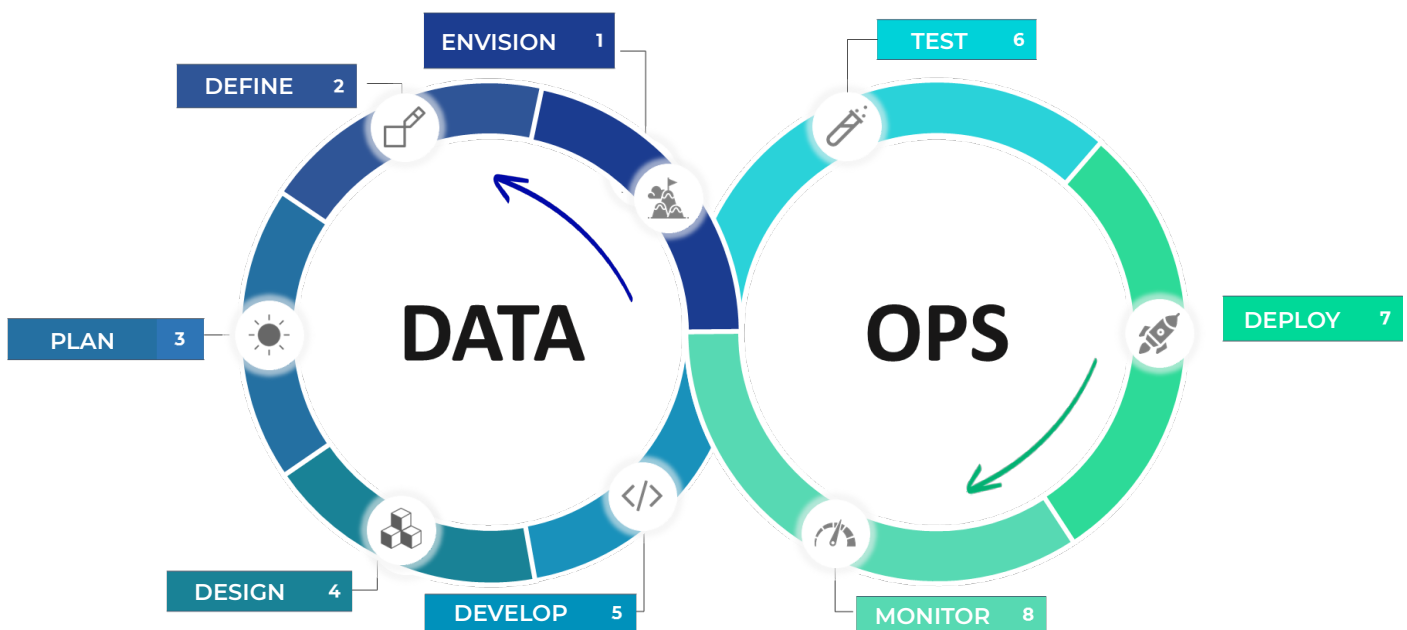
*“In the field, I am increasingly asked to integrate the Agile methodology into my projects, which are sometimes extremely large, focusing on data delivery cycle optimisation. DataOps is becoming essential in order to remain competitive and rapidly deliver robust, high-quality projects based on automation combined with integration, development and continuous research. Talan’s pool of experts in these fields addresses these issues, brings innovation and shares their considerable experience.”*

**David Jayaseelan, Head of Visual Analytics, Talan**

## The DataOps loop

Talan's Data Intelligence Solutions teams have been working for a leading French energy provider for several years. They have been tasked with **building a Data Factory that consolidates, transforms, retrieves and improves the quality of data from various business units.**

Our teams are involved in all data phases, ranging from collecting and refining the requirements to data retrieval and including corporate Data Warehouse modelling, integration, transformation and data enhancement.



The first five steps are inspired by the **Scrum method**.

This method firstly **makes it easier to visualise the workflows and their location** in all processes. It then **reduces the amount of work in progress** by setting the maximum number of simultaneous tasks in each state of a workflow. Lastly,

Accordingly, our data experts have **set up a DataOps methodology that divides the data processing flow into eight different phases**. It consists of two main cycles:

- An initial **data-oriented cycle** called "DATA";
- An **administration/operations cycle** called "OPS".

These cycles are indicated in blue and green, respectively, in the following diagram:

it **measures and optimises process execution lead times** to make them as short and predictable as possible.

The method is mainly intended to enable **timely decision-making**, set up a **continuous improvement process** and **minimise anything that will not generate value** for the client.

Most of our proposals concentrate on the last three steps, those in the green cycle. They are intended to **improve project quality by automating tasks whenever possible** and monitoring the environments and flows.

The elements of these eight steps are shown below and should be described in detail in a separate article.

1. **ENVISION:** Identify and challenge the different Agile epics
2. **DEFINE:** State the requirements – Subdivide the project – Detect and define the sources
3. **PLAN:** Finalise the study of the requirements – Define the development phases – Prioritise the user stories
4. **DESIGN:** Build the data model – Transform the sketched model into a graphic model
5. **DEVELOP:** Develop the user stories – Write all test cases – Prepare for deployment – Write documentation
6. **TEST:** Check that the test cases cover all situations – Integrate the new tests and check their impacts in the project testing repository
7. **DEPLOY:** Create a new release – Deploy the new version of the product
8. **MONITOR:** Execute the testing repository – Analyse the impact upon the servers

## Quality and community: two fundamental pillars in DataOps projects

This entire loop is based on two fundamentals on which Talan places

particular importance in DataOps projects: quality and the community (or promotion) built around the solutions.

### 1. The user experience builds project quality

When we examine what improves the quality of data projects, it seems that the user experience produced by the deliverables plays a significant role.

Consequently, it is essential to **ensure that all teams are aware of good practice in data visualisation techniques** to help to improve the quality of data projects. The two unavoidable steps in the data retrieval user story design phases must also be included: **sketching and modelling**. These two steps are essential when creating a new report or a new table, but also when designing an application.

**Sketching serves as the basis for the prototyping phases** as well as being used to validate the structure of the data, the routing of required analyses and the suitability of charts relative to the requirements and think about navigation within the application, among other things.

Similarly, **modelling** (using PowerPoint or a web-based tool) **is used to include the interactive aspect**. For example, links can be set up to show the navigation available between different pages.

This can be used to introduce the users to the elements on which they can comment.

The aim is therefore to **make it easier to change the interface proposed initially in order to correct or approve the choices** (technical or functional) before the project's development begins.

## 2. Promotion builds project adoption

We have always seen the following within our clients: **promoting the tools and the DataOps team's success fosters solution adoption.** The business lines' trust in the project and their appetite for the solution also depends on the communication efforts of the DataOps teams and their moves to share its value. It is crucial to **identify the project's sponsors and provide them with information to maintain their commitment** and encourage them to promote the solution.

### Adopting a data project integration approach

Before we end, here are a few words on our vision of the "OPS" part of the loop. **Talan does not recommend skipping the setting-up of a Testing Factory in projects.**

The aim of the DataOps approach is to enable improvements to be made during the cycles by introducing a framework of industrial tests including both purely technical and business test cases and feedback on acceptance testing results/anomalies. This increases and **maintains data reliability** over time, **enables proactive error detection, reduces the project execution cost** thanks to automation and especially, **increases and sustains the value of the data flows.**

Another important aspect of DataOps is the **continuous integration approach taken to project deployment.** This is crucial because it creates an environment in which the teams are encouraged to perform tests daily without the inclusion

of new functions adversely affecting what exists. **Continuous integration begins in the Testing Factory and continues during data retrieval non-regression testing.**

The final step in the loop's "OPS" cycle consists of monitoring. Once systems are industrialised, they include many applications available to run. These must meet stringent self-service, availability, performance and GDPR constraints, for example. **It is crucial to be able to identify any warnings before the project is industrialised.** These may involve resource consumption (disk space, memory or CPU), response times or the data life cycle, for example.

This **monitoring phase maintains the satisfaction level** of the application's users.

### Conclusion

If you still have any doubts, adopting DataOps practices brings many benefits in data projects.

These practices help companies to **improve the quality of their data projects, one of the first benefits seen being significant cost reductions.** Experience has also shown that they improve both data team productivity and the effectiveness of the company's administration/operations teams. **Today, data project optimisation is a crucial step in fully exploiting company projects** intended to generate more business and gain a competitive edge.

## WHO ARE WE?

For more than 15 years, Talan has advised and supported companies and public authorities and implemented their **transformation projects in France and abroad**.

To meet the challenges of an increasingly complex world, we help our clients embrace short transformation cycles, from design to implementation. We rely on both the technological lever and the strength of our DNA based on **collective intelligence, agility and entrepreneurial spirit**.

Collective intelligence consists of combining the diverse expertise of the group's workforce with that of our customers and partners. Agility allows us to capitalise on our experience and reinvent ourselves on each project with pragmatism and humility.

At Talan, we have firm convictions but no certainties. We are entrepreneurial, daring, bold and experimental when need be.

### Our Data Intelligence expertise

Companies are confronted with many new information channels, the complexity of the information system and the difficulty of fully exploiting their data. And yet, they must master these issues in order to:

- **Respond more rapidly in the competitive market**
- **Personalise the customer experience** to make it unique
- **Open up and develop data use** in the company so that its business lines communicate between themselves
- Have **global corporate visibility**

This fundamental digital transformation **places data at the heart of all business models and all actions to be managed. Effective, adaptive governance of this data is needed.**

At the technical level, we talk about **data visualisation, machine/reinforcement learning, data science, data warehousing as a service, data integration, data security, hybrid cloud (big/smart) data platforms, master data management and Robotic Process Automation (RPA)**.

Talan helps organisations to make this radical change successfully by providing the human resources, know-how and technological expertise required.

For more information, visit [our website](#).

# CONTACTS

## **Raphaël Goncalves**

Solution Manager, Talan Solutions  
[raphael.goncalves@talan.com](mailto:raphael.goncalves@talan.com)

## **Frederic Jacquet**

Head of Innovation, Talan Solutions  
[frederic.jacquet@talan.com](mailto:frederic.jacquet@talan.com)

## **David Jayaseelan**

Head of Visual Analytics, Talan Solutions  
[david.jayaseelan@talan.com](mailto:david.jayaseelan@talan.com)

Follow all the latest Talan news  
on [our website](#) and social networks:



**Talan**<sup>★</sup>