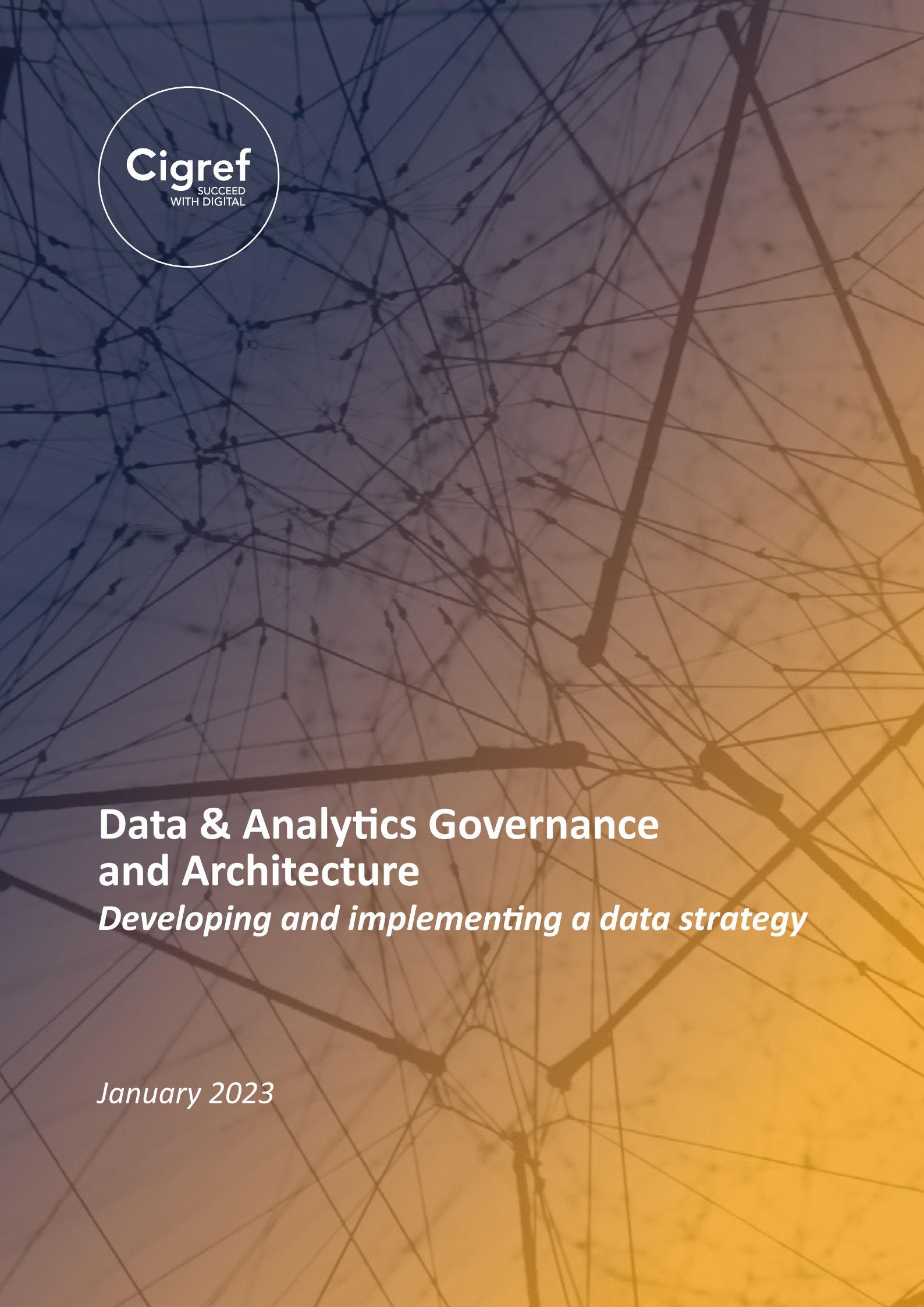

Data & Analytics Governance and Architecture

Developing and implementing a data strategy

January 2023

Cigref

Data & Analytics Governance and Architecture

Developing and implementing a data strategy

January 2023

OVERVIEW

Companies and public administrations believe in the value of data. Data is an asset for the entire organisation, not just for one department or business unit. This is why companies and public administrations are transforming themselves to make better use of this data, investing in building and implementing a data strategy that can serve their overall strategy.

For this, organisations identify their motivations and what they expect from this approach and estimate the potential gains, such as better time to market, shares in new markets, operational excellence, improvements and adjustments in their responses to customers and users, and ensuring their solutions and products are compliant. Each organisation has its own environment in which they build their data strategy, allowing them to place it within an overall strategy that suits its mission and structural and conjunctural needs, such as the group's organisation or a healthcare crisis.

Building the data strategy then follows a few major steps, such as the following:

1. Assess the maturity of the company and build the data knowledge of its employees,
2. Define specific business objectives,
3. Identify the relevant data to be used,
4. Prepare the architecture and technical bricks.

This guide to building a data strategy is a synthesis of the various approaches of the contributors to the "Governance and Data Architecture" working group. Of course, it can be amended and adapted to suit the needs of each organisation.

Once the data strategy has been defined, it must be implemented at the levels of technical foundations, governance, organisation and processes. The architecture of the IT system, its history, technological dependencies and the maturity of the data culture are all elements that impact this implementation.

A data strategy requires the company to have dedicated data governance to tackle challenges such as knowledge of the data, quality control, harmonisation of reference data and regulatory compliance. Since all the players must work together to prepare the data to be used, it is important to coordinate the data strategy's governance with the other governance systems in place, such as the IT strategy, business strategy and the digital transformation, which all contribute to the company's general strategy.

To implement the data strategy, companies and public administrations must also work on its organisation, i.e., how tasks are distributed among employees and how each person's role in rolling out this strategy is defined. This issue is increasingly being addressed within the larger issue of the Agile enterprise.

Finally, once the data strategy has been defined and put in place, the company or public administration still must manage and monitor its implementation and evaluate its benefits in various areas. The report shares the indicators and metrics that the participants in the Governance and Data Architecture working group use for these purposes.

ACKNOWLEDGEMENTS

We would like to thank **Alice Guéhennec**, Chief Digital & Information Officer at SAUR Group, and **Patrick Mahu**, Enterprise Architect at Pôle Emploi, who steered this study, as well as all those who participated and contributed to this Cigref working group (in alphabetical order):

Mourad ABBAS – RENAULT
Olivier ACCOLAS – GROUPE EGIS
Emmanuel BATT – GROUPE SEB
Sindbad BELLILI – SYSTÈME U
Frédéric BERNARD-PAYEN – AIRBUS
Frédéric BERQUEZ – MATMUT
David BIREBENT – LABORATOIRES PIERRE FABRE
Mohamed BOUTAYBI – BANQUE DE FRANCE
David BRETON – FAYAT
Léa CAPGEN – SCOR
Laurent CASSASSOLLES – RENAULT
Xavier CASTA – POLE EMPLOI
Evagelos CHAIDACIS – TOTALENERGIES
Florence CHIRAT – GROUPE SAVENCIA
Raphaël COMMARET – VIRBAC
Nicolas CONNAULT – NAVAL GROUP
Frédéric Damez – ESSILOR
Aurélie DAVIAUD – GROUPE BPCE
Xavier DELAGE – POLE EMPLOI
Stéphanie DELAUNAY – ENEDIS
Fabrice DESCHAMPS – COVEA
Gregory DENIS – SAUR
Hassan DRISS – MINISTRY OF THE INTERIOR
Bertrand DUBRULLE – CNAV
Youssef EL MOUDDEN – SCOR
Stéphane GIRARD – TDF
Alice GUEHENNEC – SAUR
Stéphane – GUILLOTVALLOUREC
Jocelyn HABEILLON-WOLF – HAGER GROUP
François HERLENT – MAIF
Alexandre HUDELOT – SCOR
Cédric JUBLOT – EIFFAGE
Christophe JUILLET – EDF
David KRIEFF – GROUPE ADP
Andreea LACHAPELLE – EDENRED
Raynald LASOTA – FRANCE TELEVISIONS
Rudy LEROUX – BANQUE DE FRANCE
Patrick MAHU – PÔLE EMPLOI
Roland MANFREDI – GROUPE ADP
Philippe MORILLON – KEOLIS
Lionel NAESSENS – MSA
Sandra NAUD – SYSTÈME U
Valérie PENAUD – POLE EMPLOI
Pierre RAMEL – BNP PARIBAS
Franck RICOTTIER – AIR FRANCE KLM
Hugues ROUMEZIN – MAIF
David SCEMAMA – MINISTRY OF THE INTERIOR
Ronan TIGREAT – NAVAL GROUP
Gilles VAN SCHENDEL – GROUPE SAVENCIA
Claude VARY – GROUPE ADP
Florent VERRIERE – EDF
Dominique VIENNE – ENEDIS
Frédéric WALKER – MSA

We would also like to thank all of the participants whose input significantly contributed to the thinking in our working group (in alphabetical order):

- Sindbad BELLILI, Head of the EDH team, Système U,
- Frédéric BERNARD-PAYEN, Data Governance Manager, Airbus,
- Gregory DENIS, Chief Data Officer, SAUR,
- Sylvain GIRARD, Chief Digital & Information Officer, Hager Group,
- Alice GUÉHENNEC, Chief Digital & Information Officer, SAUR,
- Jocelyn HABEILLON-WOLF, Data Management Specialist, Hager Group,
- Christophe JUILLET, Head of Enterprise Architecture, EDF,
- Frédéric LAU, Mission Director, Cigref,
- Marie-Emmanuelle LECLERCQ, Data Governance Manager, SAUR,
- Patrick MAHU, Company Architect, Pôle Emploi,
- Lionel NAESSANS, Data Architect, iMSA,
- Hugues ROUMEZIN, Head of Data Governance and Data Quality, MAIF,

This document was designed and written by **Marine de SURY**, Cigref Mission Director, with the support of **Elena SILVERA**, Cigref Mission Officer.

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF INSERTS

1 INTRODUCTION

The tidal wave of data from various sources like sensors, machines, complex systems, human experts and the environment raises the question of how to make the most of this data using methods, approaches and processes in a very diverse group of contexts, including management data, customer data, scientific and industrial data as well as open data,[1] which can be supplemented by metadata for characterisation.

In 2016, the Cigref working group on "Adding value to data in large corporations"[2] looked at best practices in the valuable use of data and developed a self-assessment tool[3] of an organisation's maturity in this area. This work led to the understanding that data must be recognised and managed as an **asset of the business or public administration in its own right**, the ownership of which is not exclusive to a particular group, **as all the actors are required to work together to prepare the data for certain use cases**.

While it is often stated that "data is everything in business", the transition from top management's aspirations to an implementation of a data strategy is less clear. The aim of this report is to provide IT and digital technology departments with the keys to meeting the high expectations of general management. It is aimed at all employees who are called upon to become stakeholders in the data strategy.

Companies and public administrations are transforming themselves to make better use of data, and this constitutes an opportunity to serve the company's strategy through the data strategy. The first part of the report aims to answer two questions: "why implement a data strategy?" on the one hand, and, on the other hand, "for what? For what purpose(s)?"). It should be pointed out at this stage that the use of data cannot be conceived as a purely technological matter. This is because **data strategy is intrinsically linked to the mission** of the company or public administration, to which it must respond. This section also suggests some major steps for developing and building the data strategy within the company or public administration.

The data strategy touches on the very image of the organisation, its values, its aspirations and projects for the future: it really is an issue of transforming the business. The report offers best practices for its implementation. The right approach is one that helps reach the objectives set out in the strategy on aspects such as how the transformation programme is structured in the company, the technology platform, governance, project, product and service portfolio management, and industrialisation.

It became clear companies need to take special care to clearly identify the objectives of its data strategy so that it can measure results and compare them to the expected targets. Also, a relevant measurement of results makes it easier to organise this strategy, plan out the steps, mobilise resources in a useful manner, and carry out the work in the right order. The "Data Governance and Architecture" working group is aware of the issues that underlie data governance and architecture and set out to respond to these two main issues, measuring the benefits of good data policy management.

[1] Open data is data that is freely available and can be used and shared as desired.
[2] https://www.cigref.fr/un-rapport-cigref-sur-la-valorisation-des-donnees-en-entreprise
[3] https://www.cigref.fr/un-rapport-cigref-sur-la-valorisation-des-donnees-en-entreprise

2 BUILDING THE DATA STRATEGY

Companies and public administrations believe in the value of data. They are an **asset of the whole company**, not just a department or business unit. It is therefore important that data governance is organised at the **company level** with a person **in charge** who ensures its quality, manages its life cycle, secures it, makes it available according to specified rules and is aware of all the ways it is used internally or within its ecosystem (suppliers, partners, customers, vendors, etc.).

Everyone agrees that data is today's black gold, although it is a "non-rival good". Indeed, it has a value that can be broken down in three ways: a use value linked to its use (the more the published data is used, the more its value increases), a market value linked to its scarcity, and a protection value linked to its sensitivity.

Companies and public administrations vary greatly in their maturity in the valuable use of data. Some have already defined their data strategy and are in the process of deploying it. Data serves their strategic objectives. Above all, **data strategy is a business strategy**. However, **the data must be well documented, identified, up to date and usable**. Others are at the awareness stage and are starting to think about how to make valuable use of the data. Still others are seeking to adopt a more "data-centric" vision within their organisation, i.e., to integrate data into decision-making. The question of the benefit of a data strategy therefore arises for several of them.

After answering the question of *why* to develop a data strategy, this chapter looks at *how* to build it and suggests some best practices for this.

2.1 WHY DEVELOP A STRATEGY?

The challenge in building a data strategy for the company or public administration is to **create value** for the business departments and for all teams more generally. Several objectives are pursued through the valuable use of data, and the development of a data strategy is generally driven by one of these, listed below:

- Increasing market share
    - Improved customer service,
    - Sale of services,
    - Risk management.
- Defending against the risk of disruptive innovation, external to the company or public administration, using data made available (open data).
- For public services, improving the quality and relevance of responses to customers/users
    - Appropriate responses to a wide range of clients
    - Encouraging users to make full use of their rights and fighting fraud in public services
- Improving operational excellence
    - Improving operational efficiency,
    - Provision, use and control by the end customer,

- Better matching of responses to public service users.

- Compliance / Regulation / Quality
  - Consideration of environmental aspects and sustainable development,
  - Social responsibility,
  - Compliance with the General Data Protection Regulation (GDPR).

2.1.1 INCREASING MARKET SHARE

In many cases, companies and public administrations already have projects to make use of data. **Bringing these initiatives together** can sometimes lead to the development of new services that offer business opportunities that can supplement the company's traditional activities. For example, the sale of collected or generated data or the sale of services created by using them.

Several companies or public administrations sought to assess the value chain of their customer/user journey so they could identify ways to improve their services internally and/or within the ecosystem.

Others began building their data strategy to get to know their customers or users better, for example by determining and monitoring their customer profile and related consumption behaviours. This more precise targeting helps them optimise sales and/or highlights leads for new business opportunities. However, the data strategy must be easy to implement in a timeframe that fits the time to market and comes with an acceptable investment.

For some, the development of a data strategy starts with the desire to digitalise services or to redesign them so as to bring more value to customers and/or increase their monetisation.

Finally, some companies have built their data strategy after the acquisition of new entities in order to integrate the new business units and reap the benefits quickly.

2.1.2 DEFENDING AGAINST THE EXTERNAL RISK OF DISRUPTIVE INNOVATION

A data strategy can be put in place to respond to an expressed external or transversal threat to the company. Indeed, public company data (for example, the nutritional value of food products sold), can be used by a third party to create a new service (for example, Yuka, if we continue with the case of nutritional value). The use of this data can then cause the destabilisation of an entire ecosystem by constituting a disruptive innovation that changes customer expectations. The core business is being disrupted by the emergence of new players on the periphery with a business model based on using data from the original business. The question for companies is how to react. The data was already present in the company but not used for the same purpose. They were used in the supply chain to obtain the formulas of products sold or to be shown on labels for regulatory purposes and have become a subject for R&D. The specificity of this external or transversal threat is that it changes the perception of customers, which forces the company to adapt to better meet their expectations. Another example: Trainline. Trainline is an independent aggregator that offers tickets from 270 train companies in forty-five European countries and allows users to compare competitors such as SNCF and Trenitalia between Paris and Lyon. This British company, which entered the French market in 2016 by acquiring French start-up Captain Train, is shaking up SNCF Connect's offer.

Several startups are thus relying on open data, i.e., data made available by companies or public administrations, to offer new services.

In "platform businesses", businesses are offered from the entry point that is the platform (see the Cigref report New business platform strategies[4] published in 2019). The platform answers a customer's problem in an ecosystem. Companies in the eco-system can find themselves turning into sub-contractors, even if they have their own website or platform, since the market share of the platform or comparison engine makes it virtually inescapable. For example, most hotels must be present on the leading Booking.com website, even if they have their own online booking platform. In this example, Booking.com's intermediation results in lost margins for the hospitality companies and more profits for the platform.

2.1.3 BETTER ADJUSTMENT OF RESPONSES TO CUSTOMERS/USERS OF PUBLIC SERVICES

Public services are seeking to improve their service bundles and better tailor their responses to customers/users. By using carefully identified data, they can better detect fraud and make it easier for more digitally reluctant users to use their services.

Public administrations also seek to improve take-up of eligible rights and social services, which is one of the strategic axes of poverty prevention and reduction policies. Digital technology can be a vehicle for easier access to public services or a means to help remedy the situation. This initiative relies on a variety of data, including public statistics, surveys and specific feedback, that helps to identify, describe and evaluate the phenomenon.

2.1.4 IMPROVING OPERATIONAL EXCELLENCE

Encouraging foresight, providing tangible elements to guide teams in their decision-making or making interactions within the company's ecosystem with its customers, employees, suppliers and partners more effective by using data are all factors that can contribute to a data strategy.

Companies are always looking to continuously improve their solutions, services and associated processes, increasing business control, reducing costs, improving internal processes and operational efficiency and delivering more value to their customers, and sometimes to their internal customers as well, especially in organisations with shared service centres. Metrics, indicators and all that constitutes "data" offer the possibility of refining forecasts, speeding up decisions and better targeting customer expectations, thus strengthening the capacity for decision-making or strategic orientation; after all, "knowledge is power". Teams rely on targeted measurement data to validate or confirm perceptions or customer feedback and avoid developing solutions/modules with misperceptions. Brand image is a key issue for the company. The data used to evaluate brand image allows it to analyse customers' and detractors' perceptions to identify what action to take to reach its objectives. This is how data can help shape the brand image.

Finally, the search for efficient interactions within the company's ecosystem, with its customers, employees, suppliers and partners, is for some companies and public administrations one of the triggers for the development of a data strategy. This setup requires users in business units to be more independent in using data to increase interactions with their ecosystem and offer better quality in

[4] https://www.cigref.fr/new-business-platforms-strategy-design-deployment

their relationships. Internally, the information must be efficiently shared between the company's business lines and available through a catalogue of data accessible to all (see paragraph 3.1.2.).

2.1.5 COMPLIANCE, REGULATION, QUALITY

New regulations, market or quality requirements – related to RGPD compliance, medical confidentiality and cybersecurity, for example – are catalysts for developing a data strategy in some business areas. These constraints are then used as a driver. Traceable data leads to better control throughout their life cycle. For some companies and public administrations, the strategic orientation of data responds to climate or environmental issues such as the decarbonisation of IT strategy or the greening of data.

2.2 DEVELOPING A DATA STRATEGY

A data strategy is part of the overall strategy of the company or public organisation, whether in terms of operational performance or growth. The approach to developing a data strategy differs between entities, although certain commonalities are emerging. It emerged from the "Governance and Data Architecture" working group's workshops that data strategies are built in different stages as presented below, based on several business, organisational, regulatory and sometimes financial stakes.

- Stage 1: Assess the company's maturity and develop its employees' data culture,
- Stage 2: Define specific business objectives,
- Stage 3: Identify the relevant data to be used,
- Stage 4: Prepare the architecture and technical bricks.

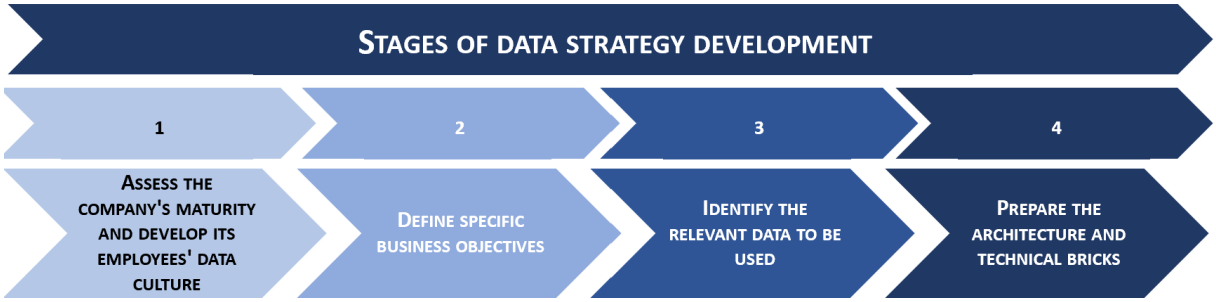

Figure 1 Stages of data strategy development - Cigref Source

2.2.1 STAGE 1: ASSESS THE COMPANY'S MATURITY AND DEVELOP ITS EMPLOYEES' DATA CULTURE

2.2.1.1 Assessing the company's maturity on data

Several contributors to the "Governance and Data Architecture" working group recommend that the data strategy should take into account the company's collective data maturity. Find out where the company is starting from, assess its maturity and share this assessment with employees. This helps to determine what steps remain to reach the target level. You can use Cigref's valuable use of data self-

assessment tool, published in 2016 with the report "Adding value to data in large corporations". The maturity evaluation grid uses the following axes:

1. *Business Models*
2. Knowledge (customers, ecosystem)
3. Transversality and operational efficiency
4. Strategy and governance
5. Ethics and compliance
6. Tools and solutions
7. Digital sobriety

"Digital sobriety" was added as an axis following the Cigref / Telecom Paris Tech - Master 1 project which was carried out by students of the "Management of Information Systems and Digital Transformation" course as part of the "Governance and Data Architecture" working group.

The maturity assessment may reveal that there are differences of appreciation between different teams in the same company. The questions in the self-assessment tool ensure that the company or public administration covers the most important topics. It is a good way of exchanging information internally and helps teams to compare how they each see the data strategy's issues and challenges.

2.2.1.2 Data awareness, training and communication

This sub-section shares the different ways the working group's participants tackle the challenge of transforming mentalities.

Awareness

Companies are not necessarily aware of what they can gain from sharing and exploiting data. These require employees to be aware of and trained in concrete examples of the value that can be created for the business units. The cost of qualifying, storing, consolidating and even destroying data must also be addressed. Data must be seen as a **tool** (enabler).

To get the business units on board, it may be helpful to show the impact of the data by using concrete use cases, sharing feedback, or using data that is already available (datasets, KPIs and dashboards) to respond to issues raised by management on existing projects.

Some employees may see data as just an Excel file. This is why it is important to explain the variety and wealth of simple and complex data once it has been analysed or processed by artificial intelligence. It is essential to **raise awareness among employees of the quality of data, its life cycle,** as well as the **legal constraints and local or global regulations** to which personal data (RGPD) and industrial data (*Data Act*) are subject. To be able to deliver a trusted service to customers, data must be characterised transparently, taking into account the ethical or legal risks associated with its use. For example, these could be related to their origin, whether they come from the ecosystem (external data) or from the company itself, or to their traceability and compliance, particularly when they are integrated into a consolidated indicator.

Data is a shared asset, and companies and public administrations must make it easy for data to be shared smoothly between the organisation's employees. Some companies have set up a cross-functional team of legitimate business actors to break down functional silos and promote and facilitate data exchange between teams.

In some organisations, the business units are invited to reflect on **the value that the data produced for their benefit is likely to bring to the other business units within the company**, including by crossing it with public data. Therefore, we should **encourage thinking about the company as a whole or even taking a systemic view of the ecosystem, not just focusing on the business unit.**

Some companies have chosen to make data sharing systematic to encourage exploring how it could be used and develop new use cases.

Feedback from EDF: Awareness and training

Deploying the data strategy within EDF requires raising awareness among as many employees as possible and training certain people with specific needs to increase their data skills.

The company thus decided to roll out an awareness-raising programme for everyone, including managers and executives. Managers should be ambassadors for the data transformation. Raising awareness involves sharing content that explains the main principles of this transformation but also demystifies key data concepts. It is also an opportunity to publicise the data use cases that have already been implemented and their benefits, while highlighting the players in the data sector who are not always well known to operational staff. It should be noted that this awareness-raising process takes time and can be boosted by people and networks already active in the digital transformation.

As far as the training process is concerned, it is handled within a digital academy which has created a training catalogue dedicated to data. The target populations are both beginners and those in the data sector who have more specific training needs. Training can be provided by external organisations or by internal staff, in this case experts from the company's data entities. There are special training courses for people who want to start a new career. This responds to the problem of internal sourcing of data skills that are still rare in the entities. This data training catalogue can be freely consulted by all and is updated by the Group's IT department.

Sabine Macia, EDF

Training

To encourage buy-in to the data strategy, companies must offer **robust training and up-skilling** for operational staff as well as management and top executives. Several companies and public administrations have chosen to develop a programme to raise awareness among all employees, including executives, management, and field staff, and accelerate cross-functionality and cross-fertilisation of the data development sector. However, care must be taken to ensure that the upgrading is targeted at the interested and intended profiles and that it provides the desired level in a sustainable manner. Indeed, one company mentioned that it had set up a training course whose content was highly rated by the participants. However, those same participants took a test of knowledge three months later, and the results failed to meet expectations. This training module has therefore been completely overhauled.

For IT employees, the training provided must be more advanced, especially in the technical aspects such as data acquisition infrastructure, management, exposure, exchange and enhancement, backup, archiving, and purging and must include design rules and business information models such as a

glossary and canonical data mobilisation. Providing visibility on the scope of missions and the different skills of the team, such as data management, the IT data unit and the business units, allows employees to better understand the various stakeholders and how the skills of the IT department, the business units and external experts work together.

Some companies have chosen to make an organizational chart of internal staff who work with data to help identify their skills. Communication is adapted to the targeted roles and backgrounds, and training courses are better adjusted.

For the data strategy to be successfully implemented, it must be based on a framework that clearly explains how users can use the data. Therefore, it is very important that employees are trained in the new tools and processes so that they can take ownership of them and use them correctly. Some companies and public administrations are setting up transversal data communities to raise awareness and share best practices.

Feedback from Cnav: Data quality awareness

CNAV manages and is responsible for data for its own needs as a general pension scheme, but also for needs coming from other pension schemes and/or organisations in the social sphere (e.g., NIR - *Numéro d'Inscription au Répertoire*, better known as the national insurance number).

Cnav's IT department has created a data management department which is responsible for data governance. This department offers a bundle of services on Master Data Management, exchange management, data quality management and data capture for the whole of the IT department and the business units that use data. As such, it provides awareness-raising and training on these various themes. **The goal is to move towards informed, accurate and high value-added knowledge of the data for which Cnav is responsible on behalf of the social sphere.**

Bertrand Dubrulle, Cnav

Communicate

A number of companies and public administrations state the importance of sharing and **communicating** widely on the data strategy, its **purpose and its expectations** in order to get all employees on board and raise awareness**.** This company-wide sharing helps these different entities to align. It also helps to promote a data culture. The language used plays an important role: avoid overly technical terms and use simple messages that are clear to all stakeholders in order to get everyone on board**. Data strategy is not just a technical issue, it is part of business transformation.**

Some data is shared within organisations as open data, i.e. it is freely available and can be used and shared freely. By definition, open data must be accessible, reusable and redistributable without restriction by any user, which requires ad hoc communication for proper use.

The company or public administration also communicates on its data strategy externally to its ecosystem, but also more widely through hackathons[5] or challenges to schools, through partnerships

[5] A hackathon is a gathering of IT people for several days and at least one night (often organised over a weekend) to collaborate on cutting-edge and innovative computer programming topics.

with startups, during talks at universities or colleges or during internships. The media is also a good way to communicate about the company's data strategy, new services or the **availability** of data catalogues. In some entities, a data ambassador is appointed in the communication team. Finally, external competitions such as IT Night[6] also highlight work on data strategy in the media.

Management support

As the data strategy leads to a transformation of the company, sponsorship from management is a real asset, whether it acts as a leader or as a stakeholder. This makes it possible to include all the entities in building the data strategy and in studying and prioritising the essential value chains at the corporate level. It also facilitates internal communication, contributing to the rapid implementation of the valuable use of the data. Finally, this essential sponsorship smoothes out the difficulties encountered when developing data governance in the IT, digital and business teams and then implementing it.

2.2.2 STAGE 2: SET SPECIFIC BUSINESS OBJECTIVES

Companies and public organisations must identify the business, organisational, regulatory and sometimes financial stakes and objectives around which to build its data strategy. These elements will allow it to determine the measures, key performance indicators (KPIs) and relevant metrics for monitoring the impact of the data strategy and steering its deployment. It also helps to identify data sources for curation[7] and analysis.

2.2.2.1 Business stakes

Some companies have built their data strategy by thinking about the use cases that tackle their challenges or pain points, the needs of the business, or the broader objectives of the company with a systemic approach to their ecosystem (customers, suppliers, vendors and partners). Others wondered how data could contribute to the strategic objectives set by management.

> **Feedback from Hager Group: Levers for building the data strategy**
>
> One of Hager Group's ambitions is to become the trusted digital advisor to its customers by leveraging the systems that manage and capitalise on the data in its electrical and energy management installations. Its value-adding strategy aims to improve energy management, which is one of its key areas of development.
>
> Data no longer only supports the operational functions of the company; **it now generates business as its systems become connected or embed artificial intelligence in its products**.
>
> **Sylvain Girard, Hager Group**

One of the participants said that he used the business value analysis of his data use cases to assess the business benefits he could gain from a structured data strategy.

[6] During this evening, IT professionals celebrate the most beautiful innovations.

[7] Content curation (etymologically from the Latin curare: to take care of) is the practice of selecting, editing and sharing the most relevant data and content within the ecosystem for a given query or topic.

It is important to translate the data strategy into small steps that each bring value, in other words, quick wins.

Another approach shared by several participants in the "Governance and Data Architecture" working group consists in identifying the data at the heart of the use cases and initiatives and then assessing the payback period and this data's contribution to strategic issues and priorities with regard to customer programmes and new services.

A bottom-up approach is also sometimes used: it consists in identifying the data investments already made in order to coordinate them at the level of the company or public administration. In this case, the participants recommend that the bottom-up and top-down approaches should be conducted in parallel with meeting points so that the strategy and the field remain connected.

Finally, some organisations **begin with their mission statement** to develop their data strategy.

> **Feedback from the SAUR Group: Example of a data strategy trigger**
>
> SAUR's core business is water management. The pressure on water resources is accelerating. All regions are affected in France, which was not the case before. SAUR used its business, environmental and socio-demographic stakes as a basis for determining its data strategy.
>
> SAUR's ambition is to protect water through digital technology with the installation of the necessary facilities to remotely monitor the entire network and process the data reported by the plants, equipment and network. This is why SAUR Group's digital technology system is organised to put data at the heart of the information system. Data circulates between users, customers, territories, equipment, facilities and the various water information systems.
>
> **Alice Guéhennec, SAUR Group**

2.2.2.2 Organisational issues

In some cases, the data strategy is built to respond to organisational issues. Thus, the data strategy serves the businesses to help them to carry out their tasks more easily.

> **Feedback from Pôle Emploi: Example of a data strategy trigger**
>
> The launch of Pôle Emploi's data strategy was the result of the merger between ANPE (the agency in charge of returning job seekers to employment) and ASSEDIC (the agency in charge of paying benefits to job seekers). Pôle Emploi sought to provide autonomy to jobseekers who were familiar with the use of digital technology, so that advisers could focus on providing custom support to less experienced jobseekers. Digital technology, with its use of data, offers this possibility.
>
> **Patrick Mahu, Pôle Emploi**

Companies or public administrations are seeking to include data in their decision-making, in other words to set up a "data-driven" organisation. Once the annual business objectives can be measured

with clearly defined indicators and data, they are organised so that the business units can better drive initiatives by the value they deliver.

2.2.2.3 Other issues

Many companies have assessed the entire value chain of their products, solutions and services to identify other benefits of having a data strategy that ensures data availability and quality and that integrates data processing modules from its production to its use internally and within its external ecosystem.

In some cases, companies and public organisations need traceability and security on their services and products. This requires the ability to audit data access. This type of audit is especially required when data is shared with entities outside the company. Clear rules for APIs, cataloguing and data circulation must be put in place throughout the company or public administration.

Certain regulatory issues require that so-called sensitive data be specifically identified (e.g. personal data, confidential defence data, medical confidentiality, etc.) to manage certain constraints related to their storage, retention, access and sharing.

Finally, there may be financial issues where using the data allows for optimisation or estimates (taxation, investments, costs, etc.).

2.2.3 STAGE 3: IDENTIFY THE RELEVANT DATA TO BE USED

A data strategy requires identifying and capturing the right data and then sharing their definitions in a data catalogue to reach the set objectives while integrating the principles of sobriety. Companies and public administrations collect data, but their sources need to be "cleaned" and "certified" so that users can use data that is appropriate for their various use cases and purposes while being confident that it comes from reliable and accurate sources that comply with existing standards. Hence the importance of taking into account the concerns of legal, financial, HR, IT and compliance departments. This is why companies and public organisations are concerned with the knowledge of data assets and their quality. Once the data has been captured, it must be made reliable and then integrated into the data dictionary. Dashboards are set up to measure the quality of the data, although this is not always a prerequisite. On the other hand, it seems essential to know the sensitivity of the data in order to determine the level of quality control to be implemented and to be able to assess its degree of quality in a proven manner. Companies and public administrations are setting up enterprise data repositories and Master Data Management (MDM), which is a central element of the data strategy. MDM uses a set of tools to store, manage and disseminate master data within an organisation. It ensures that they are correctly identified, safe to use, error-free and of good quality.

In order to better manage the data strategy by measuring the progress of its rollout and its benefits, companies and public administrations implement metrics and KPIs.

2.2.4 STAGE 4: PREPARING THE ARCHITECTURE AND TECHNICAL BUILDING BLOCKS

Another prerequisite to easily implement a data strategy is to define a target data architecture, which can be reached in several intermediate steps. The architecture implemented for the data must be

compatible with the IT architecture standards for data transport (messages, files) or data access (API or Web Services) so as not to impact the source systems of these data. In other words, the architecture must facilitate access to data. The most complicated part consists in securing the information or data in a simple and effective way while guaranteeing that it can be shared. Next, all the functional bricks needed to implement the data strategy must be prepared. For example, to develop online sales, it is important to know how to manage stocks in real time.

Companies and public administrations are connecting to the sources of their data and taking into account all the ways in which their data can be enriched. This sometimes requires a move away from the traditional, one-piece data warehouse toward a multi-pronged approach that allows large amounts of data to be handled quickly and agilely.

3 IMPLEMENTING THE DATA STRATEGY

Once the company's data strategy has been determined, it still needs to be implemented, whether at the level of the technical foundations, governance, organisation or processes. Its implementation must also be monitored alongside the benefits with clearly identified measures and indicators.

3.1 TECHNICAL FOUNDATIONS

The participants in the "Governance and Data Architecture" working group noted common principles in the foundations of the architecture of companies to implement their data strategy, which are based on several types of building blocks: **data capture and storage**, **data governance, data exploration, data exposure and exploitation** (including calculation and analytics) including business intelligence building blocks, and finally **the software and infrastructure strategy**.

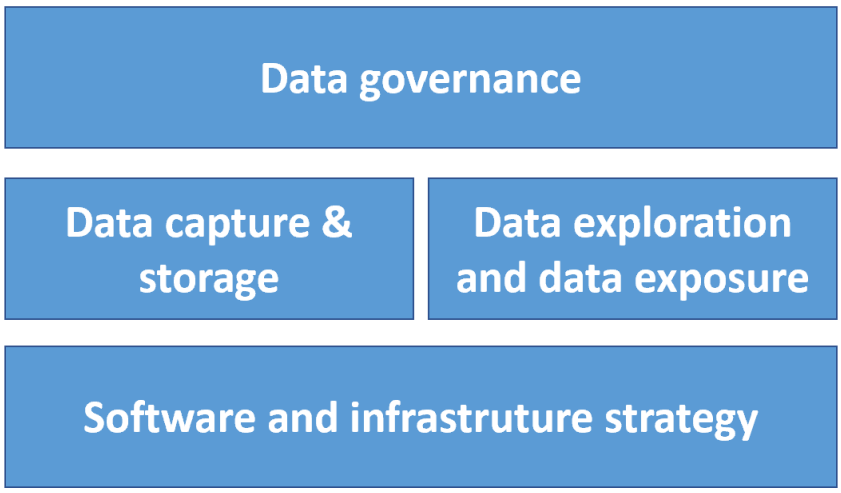

Figure 2 Common bricks in the architecture's foundation for deploying a data strategy – Cigref source

Improving the quality of data is an essential step before storing it.

There are a variety of technological choices when implementing a company's data strategy. For each of the building blocks of the data architecture, companies and public administrations use the most suitable solutions, which diversifies the technologies and solutions used and creates problems in terms of ownership costs, technological debt, and maintenance of skills on a complex software landscape. This is why many of them are now trying to find a good compromise between rich functionality, the use of new technologies and interoperability between solutions to arrive at a global solution.

Companies and public administrations are analysing the risk involved in the choice of tools and are looking to become less dependent on these technological solutions and their software providers. The ability to quickly change software providers is desirable, as is the ability to export all content. On the one hand, companies lack visibility on how tools' costs will evolve over time, and on the other hand, software providers' business models can change very quickly. Open-source solutions can be preferred to avoid becoming captive. It is a matter of "choosing your dependencies".

Finally, several companies are looking at data meshes, which offer a set of practices and organisational principles for dealing with the new challenges posed by large-scale data management. Indeed, data meshes are based on decentralised management and an exposure of data as a product in respect of the following fundamentals:

- Empowerment by allowing local data management,
- Grow the system by easily adding nodes in a decentralised manner,
- Maintain a strong coherence of the data platform (technical and organisational).

3.1.1 DATA CAPTURE & STORAGE

Companies and public administrations collect data via sensors like meters, probes and input HMIs as well as applications that input into these storage systems (ELT, ETL, messaging, scripts, etc.) and use storage systems like data lakes and data warehouses. Data hubs act as mediators and data sharing points for them. Data warehouses, data lakes and data hubs are not substitutes for each other, they are complementary and, when combined, they allow the company to implement its data strategy.

3.1.1.1 Data lake

A data lake is a single space where all the data from the company's systems, and even from outside the company, is gathered together. It contains a controlled and qualified set of operational data, i.e. the **raw data**, ingested and stored in its original format (from the source systems), for example, for traceability and possible reconstitution in case of a problem. There is no hierarchy or organisation between the different data elements. Data models and catalogues are necessary for the operation and organisation of data lakes.

The data lake can be queried for relevant data, and this smaller set of data can then be analysed to help answer business problems.

In general, data lakes are a solid basis for data preparation, reporting, visualisation, deep analysis, data science and machine learning.

One participant in the "Governance and Data Architecture" working group, whose organisation has many internal business units, resulting in disparities in terms of data and data processing, chose to move towards a centralised data lake at the institutional level of its organisation for all its entities. This choice of a centralised data lake allows for a recentralisation of the scope of data, which facilitates the meshing of data and adds value.

3.1.1.2 Data warehouse

A data warehouse is a central repository of **structured**, **modelled** and **filtered data** that has already been transformed for a specific purpose, in other words, it is a relational database hosted on a server in a data centre or in the cloud. First, the company or public administration must decide whether it wishes to keep its resources in-house or outsource some or all of them. This goes beyond data strategy alone and must be decided in the company's strategy.

The data warehouse collects data from a varied and heterogeneous set of sources with the main purpose of supporting analysis and facilitating the decision making process. The data warehouse can be fed by the ETL process (Extract, Transform, Load), but also by other processes such as messaging and search engines, which allows data to be loaded from the various applications of the operational IT.

In order to make it more robust, some companies and public administrations have chosen to support their data lake and data warehouse on several instances in different geographical data centre sites, and others use different technologies to operate their data warehouse.

3.1.1.3 Data hub

A data hub is an architectural model to support the collection, meshing and valuable use of data from multiple sources to deliver operational business use cases. The aim of the data hub is to converge data into a single repository irrespective of its source and business unit (this is called the data standardisation process). The data that passes through the hub is enhanced: catalogued, indexed, tagged, secured and harmonised. Technical performance is guaranteed so as to make the best use of them, and the business value is enhanced. Data hubs only offer services to provide data that it has obtained elsewhere (via dataset/files or open data/API).

The siloing inherent in the company's activity because of the different business processes (the purchasing process, the sales process, the operations process, the human resources process and many others) makes this task of transversality very complicated. Hence the importance of the data hub.

Some companies choose to centralise data storage in order to facilitate data de-siloing and for some companies, this technological component is directed either towards cloud solutions, which are a means of ensuring centralisation, or towards on-premises solutions.

3.1.2 DATA GOVERNANCE BRICKS

Data governance is the set of processes, skills, resources, practices, methods and tools implemented to promote the efficiency of business uses and optimise the use and quality of data to meet business stakes. Companies and public administrations are unanimous in stressing the importance of setting up a common data governance approach throughout the organisation. Starting with a global approach helps to encourage choosing inter-compatible solutions used throughout the region or product. Since the data strategy is inherently transversal to the whole company, it is important to **ensure a certain consistency in the processing of data at the global level**. Similarly, consistency in the storage and processing of data at the company level encourages their use in offers and services and helps to make the service more relevant and identify potential fraud, for example.

The rules of governance must be precisely established. However, as they are subject to change for different reasons (e.g. a new organisation, an acquisition, etc.), organisations are looking for flexible data management solutions.

The tools and architecture principles are chosen according to the governance that is to be put in place.

With mergers and acquisitions and a variety of solutions or technologies, information systems are increasingly heterogeneous, and data comes from a growing number of sources. The data is mapped

according to its type, its uses, and its location (production and storage). Defining principles and rules helps to avoid duplication and to identify the source application.

Data governance requires the use of tools such as data catalogues and data dictionaries.

- A **data dictionary** gathers information from the company's or public administration's databases and IT systems. It also contains definitions and detailed descriptions of the data sets (grouped in tables) and their values (fields). The data dictionary is primarily intended for IT teams but is increasingly used by the business units as part of the democratisation of access to data, underpinned by the data strategy. It helps to standardise information on its data: each data source is described in a data dictionary according to the chosen standard: naming convention, source, data typology, size, format, description, example, etc. It is the single entry point to all the information from the various data sources.
- The **data catalogue** is a computer program (developed internally or purchased) that provides a unified view of all the information used to describe data in the company (type of file, origin, date, etc.). It exposes the description (metadata) of the data assets hosted in the IT system and allows the information to be centralised and standardised. It allows you to manage access rights to databases, to know who is the "Owner" or "Steward" of the data. This tool is therefore also very useful to make the data actors in organisations accountable and know who to contact (e.g. who is supposed to update the information). Depending on local translations and the level of information, the fields are not always used correctly, which impairs processing. Some participants recommend setting and monitoring objectives for each field to ensure a quality data catalogue.

The data catalogue is aimed more at business-unit users and project owners. The rollout is gradual and sometimes follows difficult negotiations with top management when the data strategy is bottom up.

Many companies have implemented data strategy governance, starting with Master Data Management, but find that they get lost in the complexity. For them, they first need to homogenise the tools they use because their ecosystem has varying levels of knowledge and maturity. Governance can be based on sophisticated analysis. However, it is important to avoid the pitfall of increasingly costly solutions. Furthermore, we should add that a transformation of this type takes time because of its human aspect.

3.1.3 EXPLORATION, OPERATIONS AND BUSINESS INTELLIGENCE

Participants stressed the importance of making users as self-sufficient as possible through two modes, managed and self-service, with clearly defined and shared rules and principles. However, sharing a common definition of data within the company is sometimes not as simple as it seems. Two different pieces of data might have the same name or, conversely, two different names may refer to the same piece of data.

There is also a data quality issue given manual processes. The data is sometimes enriched by the business units, hence the importance of making employees aware of the importance of data quality.

With the arrival of cloud computing and its SaaS solutions, businesses are tempted to acquire new tools. Faced with a growing number of tools, many companies are trying to streamline their sets of solutions. Young graduates trained in recent tools (R, Python, etc.) are arriving on the market and getting hired by companies and administrations, setting up a challenge for the tools already in place.

Business Intelligence (BI) refers to the technologies, applications and practices of information collection, integration, analysis and presentation. BI's objective is to analyse data and support better decision-making in business areas like sales, marketing, finance, and HR.

When BI's data warehouses are outdated, the priority is often to modernise the entire data platform by updating the BI bricks. Several companies are looking to foster a self-BI culture, with data access rights being approved beforehand.

Some companies use data virtualisation tools. These tools allow access to data with context through APIs, while hiding technical aspects such as location, structure or language of access. However, these organisations are keen to streamline the number of visualisation tools to have data integrated into a single interface that makes decision-making easier.

There is also a growing need for data as artificial intelligence solutions such as machine learning and deep learning become more important. These new needs reinforce the interest and necessity of a solid data strategy.

3.1.4 SOFTWARE STRATEGY AND INFRASTRUCTURE

A single company often has multiple instances of data repositories located in many places, and the number increases with acquisitions. This creates new challenges for data access, especially in a large organisation. It is clear that technology alone does not solve data problems. Hence the issue for some organisations to design a functional unified data architecture, or UDA. This is not a simple problem to solve, and there are several approaches companies can take to address it, each with its own strengths and weaknesses. There is no single solution.

Companies and public administrations vary in their definitions of the notion of security for cloud computing solutions. This notion is comprised of several aspects: confidentiality (encryption of data directly on the storage units with on-site keys), integrity (non-degradation of the data during its handling and life cycle, and its controlled transformation), and availability (resilience in the event of a physical access breakdown).

Companies point to the real benefits of cloud computing services but also the problem of vendor dependency. There are several non-exclusive schools of thought which are reflected in the hybrid data storage models:

1. Storage on the company or public administration site (on-premises). Companies may choose to keep their data in on-site data centres due to a policy that considers that operational activities should remain in-house, to a need for privacy, security and compliance with the GDPR, or to remain immune to extraterritorial laws. Indeed, companies and public organisations agree that, on average, between 10 and 35% of their data, with great variability depending on their sector of activity, cannot be **migrated to the public cloud since it does not meet their required level of trust**. They want to **guarantee the security of their most strategic data and processing and expect industrial offers adapted to this need**.[8] Cigref's "Trusted Cloud" reference document [9] translates Cigref members' expressions of trust in cloud computing services and solutions into functional and objective requirements. If they have not

[8] https://www.cigref.fr/german-and-french-it-business-user-associations-position-on-the-european-certification-scheme-for-cloud-services

[9] https://www.cigref.fr/cigref-publishes-its-second-version-of-the-trusted-cloud-reference-document

been able to find a solution on the market that meets their needs, some companies and public administrations are building their own, internal cloud computing system.

2. Storage on a public cloud. Members mentioned cloud computing's flexibility as an advantage several times along with the guarantee of security that it provides, which the companies themselves cannot do to the same high level of requirements in-house. Some have also opted for hybrid private/public and semi-private cloud computing solutions. Several companies are currently in the process of migrating, having begun their journey towards SaaS and PaaS solutions. This gradual migration makes it possible to define governance rules according to the nature of the data, the type of flow, the hosting location, etc. Participating organisations report outdated data ecosystems made up of different platforms with different technologies. One of them, where about 70-80% of the data is in the ERP, uses strong, though decentralised and heterogeneous, historical platforms. However, it has chosen a US public cloud computing system to meet its present and future challenges. Others have chosen multi-cloud solutions to meet their service needs.[10]

3.2 GOVERNANCE

Governance is a prerequisite for the successful implementation of a data strategy, and it tackles several challenges: knowing the data, controlling data quality, harmonising reference data, ensuring its protection and confidentiality, managing access and, finally, educating teams about data. Discussions within the "Governance and Data Architecture" working group revealed the importance of linking the governance of the data strategy with those of other strategies implemented within the company or public administration, such as the IT strategy, business strategy and the digital transformation, which all contribute to the company's overall strategy. In addition, the governance of the data strategy must be able to be articulated at the level of the world, a region of the world, or the market.

Governance confirms the business's alignment with the data strategy and makes it easier to prioritise the services to be implemented. Given the issues that governance raises, the governance committee must have weight at the company level. More generally, several companies emphasised the importance of mobilising sponsors in the executive committee and therefore of finding arguments to convince them to support the introduction of data governance in the company. Depending on the company, data governance is handled by the CIO (Chief Information Officer), the CDO (Chief Data Officer) or the business units.

Data compliance and quality require rigour, integrity and an understanding of the semantics of the data. The issue of data and semantics is all the more important today in the age of enterprise digitalisation since many different systems are in place. There is always an increasing number of new systems that are being built. Problems related to semantics have always been present. But if you deal with them as you go along, which is no longer the case today, you get overwhelmed. In the past, standards were set, and systems were interconnected. Today, we can see a completely different procedure: first, the interconnection is set up, and then the standards are put in place.

Standard naming rules help unify the facets used in different countries or even in different regions of the same country. Standardisation of the business unit glossary is also a key issue. Should the same piece of data not have the same definition between distribution and operation, it would pose a real

[10] To go further on the subject, we invite you to consult the Cigref report "Strategies for migrating IT to cloud computing: A strategic venture for enterprise", November 2021. Further work on this document is underway and will be published at the end of 2022.

problem in the daily life of the company (length, width, height – a common definition is needed to sell its products, fill the trucks that transport them, etc.). It is thus imperative to standardise the data in the information system.

3.2.1 KNOWING THE DATA ("KYD – KNOW YOUR DATA" PROCESS)

Participants stressed the importance of moving from the "idea of useful data", seen as a concept, to a concrete reality. This is why companies and public administrations are thinking about how to create use cases that really make the most valuable use of data sharing between teams. One of the priorities is therefore to identify the data that generates a business gain or a gain on the processes by calculating the ROI of the data while identifying the risks. To do this, employees in the company or in the company's ecosystem must know where the data is located and be sure that it is up to date.

This is why the participants in the "Governance and Data Architecture" working group recommend determining the leader of each part of the data and identifying who manages data sharing. Introducing data governance at the company level with the appropriate tools allows the responsibilities to be pinpointed throughout the data's lifecycles whenever there is an action to be taken. Governance serves multiple objectives: ensuring the quality of the data throughout its lifecycle, making it easy to use with up-to-date data asset documentation, tracking data consumption, ensuring it matches consumers' needs and that it is properly erased at its end of life. In order to share and use data, common ontologies and rules must be put in place within the business ecosystem. An ontology is a structured set of concepts that allows information to be made sense of with the primary objective of modelling a body of knowledge in a given domain. The business units that produce the data have a duty to make it easy to share and promote value creation for the company.

3.2.2 CONTROLLING DATA QUALITY

In order to use the data, employees must be assured of its quality. Each piece of data must be well catalogued, identified, defined (especially for indicators), up-to-date and usable. It is important to know how to measure the quality of the data in order to improve it until the reference value is reached. The business units are responsible for this analysis because the quality meets their needs. This quality measurement can often be based on Master Data Management (MDM), which allows the value of a nomenclature-type of data to be checked against its associated reference. Internal audits must assess and control the quality and maturity of data and data flows. This level of standards maximises what data can bring to the company's various tasks. This means that the business units, data analysts and data scientists can be self-sufficient in using it in a way that is relevant.

Several participants said that they made sure that data producers kept in mind that the **data is there to be used** when defining the rules of governance. **The rules are therefore aimed at the potential data consumers** and must **take into account the data's purpose**.

3.2.3 REGULATORY COMPLIANCE & PROTECTION

Sensitive or "valuable" data must be identified so it can be stored in a suitable place: internally, with partners or in the cloud. Access to data will thus be determined according to their sensitivity. The data

controller, often the DPO[11] or its bodies, integrates the associated regulatory constraints and shares them with the users of this sensitive data. Some confidential data requires the support of a security expert to make employees aware of security issues. The objective is to maximise data use with analytical processing while minimising the risks of non-conformity or ethics breaches that might arise from these uses.

3.3 ORGANISATION

When rolling out a data strategy, several participants recommend clearly defining and detailing how the missions of the various data players within the company or public administration (business units, IT and digital technology) will work together. However, there is a certain disparity in the names and descriptions of the different data professions and their missions. In 2023, the Cigref working group "IT Job Profiles Nomenclature" will update the descriptions of the rapidly changing data-related professions.[12] Some groups have appointed a Chief Data Officer[13] in each business division to implement the data policy within the division.

The organisation must ensure the coherence of the data assets, which is a real issue for the data strategy.

> *The overall organisation seeks to build a system that is collaborative and community-oriented. The aim is to be part of the company's Agile approach, respecting the functions of the business units and IT departments, and to get them to collaborate in the system's evolution.*
>
> **Lionel Naessans, iMSA**

Today, the companies and public administrations participating in the "Governance and Data Architecture" working group have a rather centralised organisation for implementing the data strategy, but they are aiming for decentralisation once they have become more mature. Even though data is taking on a major role in companies, IT generally takes the lead until the various business units take over. The centralised nature underscores that companies are very focused on the big data aspect.

3.3.1 FEDERATION RATHER THAN DECENTRALISATION

If you stay with a centralised model, there is a risk of limiting the footprint of data governance, as well as the take-up of best practices by all the actors involved in IT. Decentralisation increases the number of possible actions. A balance must be found between what is centralised and what is decentralised. Rather than talking about decentralisation, which implies the principle of subsidiarity, according to which a central authority can only carry out tasks that cannot be carried out at a lower level, some participants prefer to talk about "federation". The federation must allow everyone to access the data

[11] DPO = Data Protection Officer

[12] Remember, a generic job profile emerges after a process that takes approximately two to three years: initially, specific skills arrive in a disparate manner, appear in the existing profiles, and then a consensus is reached on the definition of these skills. When these skills reach a certain level of maturity, specific profiles appear and the description gradually gains a consensus. They can then be integrated into the HR nomenclature.

[13] See this job's description in Cigref's IT Job Profiles Nomenclature: https://www.cigref.fr/it-job-profiles-nomenclature-2022-version

even if it is in silos within the company. These companies seek to federate data stakeholders within a network while keeping the business units responsible for managing the data within their own scopes. Thus all stakeholders have access to the data. To facilitate federation in certain companies, IT offers transversal tooling programmes with centres of expertise.

3.3.2 DATA CONTROL AND ACCOUNTABILITY

In several companies and public organisations, the business units have greater control over data than before and are expressing a desire to take control of data governance. Indeed, control over governance ensures that the data will be of high quality and create value. This is why the desired target organisation for governance puts the business units in control. Several participants also consider that this control by business units comes naturally with data maturity and is a desirable development.

However, some data is transversal, and in this case the question of who owns the data arises. Taking responsibility for it comes with a cost, i.e., the resources and skills to manage it. Several companies have organised a liability model on data populations in this case.

In general, participants agreed that data should always be made available, unless there is a risk. In this case, the data controller, usually the Data Protection Officer (DPO), must present it to management. If the risk is high and the data is not very valuable, it is not passed on because the benefit/risk ratio is low. However, if the data is of great value and publishing it entails great risk, it is up to management to decide.

3.3.3 PROCESS

Several challenges have arisen regarding the identification and qualification of use cases in portfolio management and how they are prioritised. How can we prioritise use cases after they are initially assessed through a PoC[14] or MVP[15]: by the ability to scale up, by value, by a solid business case? Are transversal use cases likely to bring more value?

> *The effort to identify, qualify and validate the use cases must also take into account the level of maturity/confidence of the envisaged analyses. For example, some preliminary exploratory studies may conclude that there are no interesting/relevant analytical models and may be abandoned after initial exploration.*
>
> **Fabrice DESCHAMPS, COVEA**

[14] Proof of Concept.
[15] Minimum Viable Product

3.4 STEERING THE IMPLEMENTATION OF THE DATA STRATEGY AND MONITORING ITS BENEFITS

Companies and public administrations are looking to monitor the data strategy's implementation and identify and measure its benefits in the business units and the ecosystem. The participants in the "Governance and Data Architecture & Analytics" working group shared the indicators they used for this purpose.

3.4.1 STEERING THE DATA STRATEGY'S IMPLEMENTATION.

Progress in implementing the strategy is measured in various aspects, such as organisation, governance, processes, tools and capacity.

3.4.1.1 Organisation

Several indicators or measures were shared by the participating companies.

- The **user or customer satisfaction** indicator. It allows you to evaluate the satisfaction and loyalty of a customer at a given moment and to follow the change in the customer/brand relationship. These may be internal customers, business units, external customers at the national or international level, or employees. Depending on the result, the company can decide whether or not to implement an action plan.
- The **number of products under development or in use**. This is used to track the infrastructure's technical capacity when implementing the data strategy.
- A **chart modelling the resources with all the roles** in data governance, based on the company's processes. This helps visualise the suitability of appointments and FTEs (full-time equivalents). Once the chart has been calibrated, the company identifies the teams to be reinforced.
- An indicator of the **size of the data service centre**. One company recommends creating a data department to monitor all projects in which data is used. This indicator shows the changes in the number of projects involving data and to recruit employees accordingly.
- The **number of identified data controllers**. Ultimately, the company must be able to identify all data controllers. In one of the participating companies, once the data controller has been identified, an assignment letter is officially sent to them.

3.4.1.2 Governance

Several indicators or measures on governance were shared by the participating companies.

- The **number of data catalogue users**. This indicator is a metric measuring the number of consultations of the data catalogue, with the KPI (Key Performance Indicator) of the deployment,

- The **proportion of the IT documented in the data catalogue**, i.e., the coverage rate of the data catalogue,
- The **volume of data stored and processed**,
- The **number of data requests managed per month**, with a ratio between the different domains,
- The **proportion (percentage) of IT change projects** that implement best data governance practices,
- Some companies have set up a **data maturity score**. This is done, for example, with a request for feedback so that users of the data can give their opinion on its quality and availability,
- The **number of indicators used to qualify the quality of the data**. Companies start with those that are critical to the market, e.g., CRM (Customer Relationship Management) data, used for business and marketing. It is good that the email address and mobile phone fields are filled in. It is better if the email address and number are correct. Thus, several indicators[16] were shared:
    - **Completeness**,
    - **Validity**,
    - **Accuracy**,
    - **Consistency**,
    - **Timeliness**, which measures the time between when the data is expected and when it is available,
    - **Currency**, the fresher the data, the more it reflects the real state of the concept at a given moment,
    - **Accessibility**,
    - **Integrity**,
    - **Precision**. This indicator assesses the accuracy and gives the granularity of the data.
- **Lineage** provides information on the origin of the data and its transformation,
    - The **representation** of the data,
    - **Uniqueness**.

One company uses Lean 6 Sigma[17] in its data use cases. Thus, once the type of malfunction has been identified in a use case, the decision tree allows the problem to be solved with the measurement of various indicators to monitor the correction of the malfunction.

Several participants recommended centralising data-related investments in order to have a global and coherent vision at the company level.

3.4.1.3 Processes / tools / IT capacity

Participants shared several indicators or measures on the processes and tools used in their organisation:

[16] The indicators are described on this site: http://dimensionsofdataquality.com/content/list-underlying-concepts
[17] https://leansixsigmafrance.com/blog/la-difference-entre-le-lean-le-six-sigma-et-le-lean-six-sigma/

- The **level of CI/CD deployment for data**. The CI/CD (Continuous Integration/Continuous Delivery and/or Continuous Deployment) approach increases the frequency of application delivery by introducing automation into the application development stages. The CI/CD approach allows the application to be continuously monitored and automated throughout their lifecycles, from the testing and integration phase through to distribution and release.
- **The availability of tools to make data available** and the industrialisation of these tools,
- The **number of data products** or tools **available**,
- **The number of use cases**,
- The duration of the **process to qualify a data use case**, from the initial idea and the exploration phase to confirm it is relevant, until it is officially released. The establishment of a process for qualifying new use cases is one of the advances in the iterative approach to deploying the data strategy,
- The **level of automation or industrialisation of the end-to-end data processing chain** (or the levels of interoperability of the data "modules") - The tools in the data processing chain include: the data catalogue, data management repositories (Master Data Management), ETLs (Extract, Transform and Load),[18] APIs, BI (Business Intelligence) and AI (Artificial Intelligence).

Should you choose a complex tool with a multitude of modules that are complicated to implement and to have interoperate with the other tools in the data chain, or should you choose a tool with fewer functionalities but that is simple to interconnect in the chain? The business prefers the tools that promise the most functionality. However, participants agreed on the importance of preferring tools that process data in its entirety, i.e., from end to end, in order to extract maximum value (vertical coverage). To this must also be added the indicator measuring the capacity of a tool to handle use cases in parallel (horizontal coverage). Participants noted that the first use cases pose a lot of difficulties in terms of vertical and horizontal integration, but once the methodology is in place, the following ones are easier. One company uses the Lean 6 Sigma tool for this.

3.4.2 MEASURING THE BUSINESS BENEFITS AND VALUE OF USING DATA

One best practice is to label the indicators so that they can be reused in several entities of the company in an equivalent way. However, care must be taken with this type of data. An indicator is a piece of data which may initially seem common to many professions in the company but which, on closer inspection, reveals specificities depending on the role. Sharing it therefore requires prior alignment work, which is not always easy. This is why the objective is to label the indicators so that they can be reused within the company but based on the same calculation. **The benefit is that everyone speaks the same language**.

In order to raise awareness among business and executive staff of the data strategy and to get them on board, it is important to demonstrate the value of the strategy through concrete and convincing initial results.

[18] This integration process is equivalent to extracting raw data from a system, transforming it and sending it to a reference warehouse. It aims to achieve massive synchronisations between different systems. ETL tools are an excellent way to collect data from different sources and then use it. They represent leading Big Data and Business Intelligence solutions.

The "Governance and Data Architecture & Analytics" working group shared the indicators used to measure the benefits of the data strategy itself in the business units and the ecosystem. These benefits are based on time to market, performance and turnover.

3.4.2.1 Time to market

- The **number of PoCs**. For example, one company set a target for the number of data use case PoCs,
- The **cost and duration of PoCs**,
- The **speed of transition from PoC experimentation to production**: in one company, once the PoC was validated, it set itself the target of going into production in six months. It had to simplify the process that had become too cumbersome.
- The **time saved** by implementing a data use case.

3.4.2.2 Performance

- The **reduction in the number of FTEs** (Full Time Equivalents) observed following the implementation of the data use case, which helped avoid unnecessary interventions and reduce processing time, for example,
- **Reduction of identified risks**,
- The **number of APIs** that ensure rapid data retrieval with market analysis. A company is buying the ability to import reliable, open data from an external supplier in order to supplement its own CRM.

3.4.2.3 Turnover

- The **number of new services created**,
- The **Return on Investment (ROI).** A significant reduction in ROI following the implementation of a data use case helps to get the business unit on board with the company's data strategy,
- **Regulatory compliance** (e.g., RGPD or quality). The risk of non-conformity has financial consequences. It is therefore important to raise awareness of the "Risk-Incident-Consequence" triplet of non-conformity.

4 CONCLUSION

Most companies and public administrations generate a very large amount of data, whether it be government data (finance, HR, regulatory, etc.) or data related to their activity. A significant part of this data is currently used in an atomised way through initiatives taken outside the IT department. Some organisations are seeking to open up their IT system with a service approach addressed to its entire ecosystem. This ability to share data at scale and generate value helps make the organisation more competitive.

There is a common belief among companies and public administrations that data is strategic. However, this belief is not held to the same extent by all within the organisations, demonstrating the diversity of maturity within them. It is essential to take stock of how mature the "data culture" is within an organisation to set a reasonable and realistic path to a data strategy. The company's data strategy is not a technological issue but a **business transformation** issue in which **all teams have a role to play**. Depending on the chosen data strategy, the issue of technologies will be treated differently. Teams' data awareness is essential, and the IT department should hold key responsibilities that revolve around several actions:

- **Define and structure the process of deploying a data governance framework** on its own initiative, if the business departments are not sensitive to and do not drive the process,
- **Anticipate the necessary industrialisation of best practices** by providing tooling capabilities (e.g., data catalogue),
- **Facilitate the take-up of data assets** and **the emergence of use cases** by creating service bundles that encourage experimentation in "lab mode".

This increased use of data to serve the business units is an unavoidable trend, which is set to increase with the proliferation of new types of data and the widespread exchange of data between organisations (e.g. the 3DS law). In response to the limits shown by too much centralisation in data lakes and position efficient processing as close to the source as possible, some companies or public administrations are implementing a data mesh philosophy.

Furthermore, the data strategy, particularly in its governance, must take into account the increasing environmental impact. Indeed, data is both part of the problem and part of the answer. First, the exponential increase in the volume of data from the sensors and complex systems that had to be produced and recycled, but its storage, sharing and analytical processing have an impact on the carbon footprint of the company or public administration.[19] However, data itself is a precious tool to inform and education those in the organisation on reducing the carbon footprint in all its activities, helping to adopt a responsible approach to the environment with step-by-step improvements that follow set objectives.

[19] See the Cigref report in partnership with The Shift Project, "Digital sobriety: a responsible business approach" available for a free download: https://www.cigref.fr/digital-sobriety-a-responsible-corporate-approach.

Cigref
SUCCEED
WITH DIGITAL

Achieving digital success to help promote the economic growth and competitiveness of its members, who are major French corporations and public administrations, and users of digital solutions and services
Cigref is a network of major French corporations and public administrations set up with a view to developing its members' capability to acquire and master digital technology. It is a unifying player in the digital society, thanks to its high-quality thinking and the extent to which it represents its members. Cigref is a not-for-profit body in accordance with the French law of 1901, created in 1970.
To achieve its mission, Cigref counts on three business units, which make it unique.

Belonging
Cigref speaks with one voice on behalf of major French corporations and public administrations on the subject of digital technology. Its members share their experiences of the use of technology in working groups in order to elicit best practices.

Intelligence
Cigref takes part in group discussions of the economic and societal issues raised by information technologies. Founded nearly 50 years ago, making it one of the oldest digital associations in France, it draws its legitimacy from both its history and its understanding of technical topics, giving it a solid platform of skills and know-how, the foundation stones of digital technology.

Influence
Cigref ensures that its member companies' legitimate interests are known and respected. As an independent forum in which practitioners and actors can discuss and create, it is a benchmark recognised by its whole ecosystem.

www.cigref.fr
21 av. de Messine, 75008 Paris
+33 1 56 59 70 00
cigref@cigref.fr