Data-Driven Transformation to Green Steel

The Role of Data Governance in Sustainability-Related Compliance Reporting for Accelerating the Decarbonization of Steel Manufacturing

Pim Weijland
4965299

Graduation Committee
First Supervisor: Dr.ir. A.M.G. Zuiderwijk
Second Supervisor: Dr.ir. E.J.L. Chappin
External Supervisor: Dennis Opdam

In partial fulfilment of the requirements for the degree of:
MSc Complex Systems Engineering & Management
at the Delft University of Technology, to be defended publicly on Friday, September 5, 2025, at 1:00 PM.

August 2025

TU Delft

(This page was left blank intentionally)

Preface and acknowledgements

Dear reader,

This thesis marks the completion of my Master's degree in Complex Systems Engineering and Management at Delft University of Technology. The research was carried out in collaboration with Tata Steel Nederland and focuses on the role of data governance in supporting the company's transformation toward green steel production.

Writing this thesis has been both intellectually challenging and personally meaningful. It allowed me to combine academic rigor with real-world complexity, particularly at the intersection of sustainability, industrial transition, and digital infrastructure. The process taught me not only how to navigate technical and organizational systems but also how to listen carefully, ask the right questions, and translate complex realities into actionable insights.

I am deeply grateful to my graduation committee. First, a big thank you to Anneke Zuiderwijk for your sharp academic guidance, continuous encouragement, and attention to both structure and nuance. Emile Chappin, thank you for your thoughtful feedback and critical perspective at the key stages.

I also wish to thank Dennis Opdam from Tata Steel for making this research possible. Your ability to open doors within the organization was instrumental. I am particularly grateful to the professionals I interviewed for generously sharing their time and perspectives.

Pim Weijland
Amsterdam, July 2025

Executive summary

Steel manufacturers are under increasing pressure to reduce emissions and operate more sustainably. This transition requires strategic decisions based on reliable, consistent, and traceable data. This thesis investigates how large industrial companies like Tata Steel Nederland (TSN) can improve data governance to support that goal. The study focuses on improving reporting and decision-making processes that rely on shared data across departments.

The thesis begins with a literature review that highlights key gaps in current research. Eight core data governance criteria were derived from the challenges to structure the findings: data lifecycle management, interoperability, data classification, data security & compliance, data unit responsibility, version control, data storage, and performance monitoring. These criteria were drawn from academic literature and refined based on the steel industry context. While the steel industry has used data to optimize processes and monitor performance for decades, little attention has been paid to how this data is governed. Studies often focus on real-time production data but overlook the governance of supporting documents for sustainability disclosures, regulatory reports, and investment proposals. The role of data governance in aligning technical, regulatory, and operational reporting remains underexplored. Especially with new regulations like the Corporate Sustainability Reporting Directive, the need for structured, transparent data governance has become urgent.

To explore this, the study uses a qualitative case study at TSN's Transformation Office. This team coordinates strategic reporting, including the Green Book, a report used to secure investment for the transition to green steel. Data was collected through interviews, document analysis, and observations during a five-month internship. The analysis showed that governance practices at TSN are often inconsistent, with unclear document versions, missing responsibilities, and scattered file storage. Most issues are not technical, but organizational.

To address these challenges, the thesis presents practical recommendations for each of the eight criteria. These include using consistent versioning, labeling files with classification levels, assigning responsibility for key data elements, and creating shared templates. The recommendations are designed to work within TSN's existing systems and routines. They aim to enhance clarity, coordination, and accountability across teams without requiring the implementation of major new technologies. Importantly, the recommendations emphasize that lasting change depends on three key factors: top-down prioritization, practical bottom-up training, and alignment among the eight data governance criteria.

To test the recommendations, a focus group was held at the end of the study. One recommendation per criterion was discussed. Most were confirmed as useful, especially when aligned with team behavior and leadership support. Six out of eight recommendations were seen as transferable to other industries with similar data challenges. However, data unit responsibility and performance monitoring were not confirmed as broadly applicable.

This thesis contributes to the growing conversation about how data governance can support sustainability in heavy industry. It shows that better governance improves data quality, which in turn strengthens strategic reporting and decision-making. While the recommendations are tailored to TSN, many of the insights also apply to other data-intensive organizations. The findings highlight that successful change depends not only on tools and frameworks but on leadership, shared responsibilities, and practical skills.

Contents

List of Figures

List of Tables

Keywords

Data Governance; Green Steel; Sustainability; Strategic Reporting; Regulatory compliance

List of abbreviations

CBAM	Carbon Border Adjustment Mechanism
CCS	Carbon Capture & Storage
CMMI	Capability Maturity Model Integration
CSRD	Corporate Sustainability Reporting Directive
DG	Data Governance
ECP	European Commission Prenotification
ETS	Emissions Trading System
GB	Green Book
IF	Innovation Fund
IM	Information Memorandum
TO	Transformation Office
TSL	Tata Steel Limited
TSN	Tata Steel Nederland

1 Introduction

This chapter introduces the context, scope, and positioning of the research. First, *Section 1.1* outlines the background of the study. *Section 1.2* explains the organizational context and clarifies the scope of the study. *Section 1.3* identifies the knowledge gap this thesis aims to address. *Section 1.4* discusses the societal and academic relevance, and *Section 1.5* highlights the relevance for the CoSEM programme. Finally, *Section 1.6* provides an overview of the report's structure.

1.1 Background

The steel industry generates between seven and nine percent of the total carbon dioxide (CO2) emissions (Muslemani et al., 2021, p. 1), primarily (75 percent) from the use of fossil fuels (Forum, 2023). Growing efforts to combat climate change have increased demand for green steel, steel produced with at least 30 percent lower CO2 emissions than the European average (Muslemani et al., 2021; Tata Steel Nederland, 2022). Transitioning to green steel requires a significant transition, in which areas such as data analytics and governance are essential.

This shift toward lower-emission steel production is strongly driven by European and national climate policy. At the European level, the Green Deal outlines a legally binding pathway toward net-zero greenhouse gas emissions by 2050, placing the industrial sector under structural pressure to decarbonize (European Commission, 2019). Regulatory instruments such as the EU Emissions Trading System (ETS) require firms to accurately track, report, and pay for their verified emissions (European Commission, 2023b), and the Carbon Border Adjustment Mechanism (CBAM) puts a premium on the ability to demonstrate carbon performance across the supply chain (European Commission, 2023a). Another example of such a report is the Corporate Sustainability Reporting Directive (CSRD), also by the European Commission. The CSRD requires companies of 1000 employees or more to disclose information on the influence their activities have on people and the environment (European Commission, 2022). These policies increase the need for traceable, auditable data to support compliance. At the national level, ministries such as the Dutch Ministry of Climate and Green Growth translate these goals into concrete subsidy structures, investment conditions like the European Innovation Fund (IF), and transition strategies. In this environment, steel manufacturing firms are expected not only to reduce emissions but also to produce transparent and auditable data to support their progress with evidence. As a result, sustainability claims and capital investment decisions increasingly rely on internal reporting processes and robust data governance.

To meet these growing policy demands, steel manufacturing companies must now back up their emissions reduction plans with reliable, traceable data. This includes reporting on energy use, production inputs, lifecycle impact, and process efficiency. These types of internal reports often require input from multiple departments that use different tools and definitions. As a result, it becomes increasingly difficult to bring all this information together clearly and consistently. Data governance refers to this exercise of authority and control over the management of data (Abraham et al., 2019). The importance of strong data governance lies in its ability to support consistent and well-informed decision-making. Strategic reporting and operational choices should be based on accurate and consistent data. When inconsistencies arise, they can lead to regulatory and, consequently, sustainability risks.

Data governance is especially crucial for CO2 emissions-intensive industries like steel manufacturing to mitigate climate change (Di Foggia & Beccarello, 2024). It contributes to sustainability by en-

hancing information efficiency, which enables organizations to better identify areas for improvement throughout the product lifecycle (Chen et al., 2020). This allows manufacturing organizations to improve decision-making strategies to bridge the gap between their current state and the desired state of data maturity (De Carolis et al., 2025). Despite our time being the fastest digital innovation, steel producers often face problems in improving data governance (Gajdzik & Wolniak, 2021).

To summarize, steel manufacturing organizations face growing pressure to manage data better, especially in internal reporting. They must meet stricter demands for accuracy, traceability, and coordination under increasing pressure. This increases the need for clear and practical data governance. The next section explains the type of environment this thesis focuses on and defines its scope.

1.2 Organizational context and scope

The transition to green steel not only depends on new technologies but also on how steel manufacturing organizations govern their data. While the steel manufacturing sector has a long tradition of using data to monitor processes and emissions (Mishra et al., 2023), the use of that data for cross-departmental decision-making and sustainability-related compliance reporting is often complex and difficult to coordinate. Specifically, when internal reports are used to apply for subsidies, attract investments, or comply with new regulations, the ability to combine information from different departments becomes critical. Yet in many steel manufacturing organizations, data is managed by separate teams using different data governance systems. This makes consistent data governance a practical challenge, especially when reports require cross-departmental input.

This raises the question of how data governance can help organizations coordinate internal reporting in a way that supports both transparency and decision-making. Existing frameworks such as DAMA-DMBOK or CMMI provide general guidance, but may not reflect the specific needs, pressures, and reporting workflows found in the steel manufacturing industry. There is a growing need to understand how data governance can be implemented in practice, not just to optimize technical processes, but to enable organizations to deliver credible, aligned, and timely reports in a complex, multi-stakeholder environment to accelerate their transition to green steel production.

This study operates within the scope of an internal reporting process to justify the investment in transitioning to green steel production. For example, a new blast furnace is designed to run on natural gas instead of coal. The proposed furnace, intended to be operational by 2030, aims to reduce CO2 emissions by at least 30 percent through a combination of fuel switching and Carbon Capture and Storage (CCS) (Tata Steel Nederland, 2023a). To build a credible business case, various departments must contribute data to one integrated report, including engineering inputs, emissions calculations, financial scenarios, and regulatory alignment. These contributions rely on different teams, systems, and timelines, making coordination across the organization both essential and difficult.

The success of such a report depends not only on the technical accuracy of each component but also on the organization's ability to govern how data is structured, validated, and shared across teams. Without shared standards or clear responsibilities, delays, inconsistencies, and miscommunication are likely to occur, jeopardizing the quality of the report and weakening decision-making. This reporting process defines the organizational setting in which this research is embedded and highlights the importance of practical, organization-wide data governance to support the green transition.

The coordination of internal reports for green investment involves a network of actors with different interests and levels of influence, ranging from global to organization-level players. First, the actors are listed individually, then *Figure 1* presents the dynamics between these actors in a PI grid, after which the figure is explained. Globally, the World Steel Association (WSA) is an organization for steel producers, industry associations, and research institutes. It shares worldwide production figures, tracks sustainability progress, and spreads best practice guidelines, giving steelmakers a common reference point for measuring and improving their environmental performance. On the European level, the European Commission (EC) plays a role by setting emission deadlines and managing instruments such as the Innovation Fund (IF), which supports industrial decarbonization projects. Additionally, the European Environment Agency (EEA) is the EU's main body for tracking and analyzing environmental data across Europe. It reports on issues like air quality, emissions, and industrial pollution, sharing independent reports that influence decisions made by the EU and national governments affecting the steel industry. On a national level, the Dutch Ministry of Climate and Green Growth has the power to allow for subsidies, financially enabling large projects towards greener steel manufacturing. Large institutional investors and banks, in turn, assess whether public support is in place before committing capital. Their priority is to minimize financial risk, so they expect projects to be backed by governments.

Inside steel manufacturing organizations, several departments must collaborate to produce reports that meet these expectations. Sustainability teams frame the environmental benefits of projects, engineering teams contribute technical and emissions data, finance departments translate plans into cost scenarios, and IT provides access to large data warehouses. These inputs are brought together under management supervision, often coordinated by a strategy or transformation team. However, misalignment between internal departments, such as unclear data ownership or conflicting timelines, can lead to delays or credibility issues. These tensions underscore the need for data governance to improve compliance reporting.

To better understand these dynamics, *Figure 1* maps this study's key actors involved in sustainability-related compliance reporting in the steel manufacturing industry onto a power-interest grid, based on the model originally developed by Mendelow (1991). In this paragraph, Mendelow's model is explained, and the relevant actors' positions on the grid are explained after the figure. There are four categories of actors on the PI grid. Actors with high power and high interest need close coordination, as they strongly influence both the process and the outcome. Those with high power but low interest should be kept satisfied, since they can still affect results if their priorities shift. Low-power but high-interest actors should be kept informed and involved. Actors with low power and low interest need only minimal attention. This helps decide how to engage each group effectively.

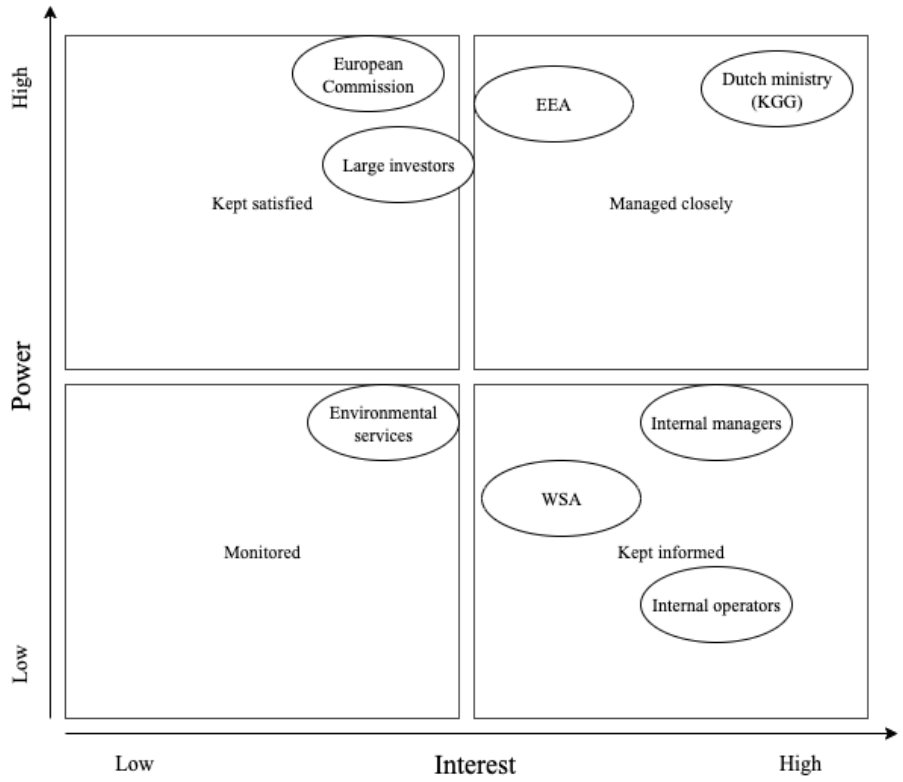


Figure 1: This figure maps actors, as originally defined by Mendelow (1991), based on their influence over, and interest in, sustainability-related compliance reporting in the steel manufacturing industry.

The first group, high power and high interest, are the actors that need the closest coordination. At the European level, the European Environment Agency (EEA) gathers and analyses environmental data from across all EU countries. It publishes reports on air quality, emissions, and pollution. While it doesn't focus only on steel, its findings can lead to new rules or stricter targets that affect how steel manufacturers must report. In the Netherlands, the Ministry of Climate and Green Growth (KGG) decides who receives state aid and keeps a close eye on whether projects match national climate goals and protect the health of local residents near steel plants. It requires companies to first apply to the Innovation Fund, run by the European Commission, before they can qualify for national subsidies. This means the Ministry's funding decisions depend directly on the Commission's approvals.

The second group, high power but lower interest, are actors who can strongly influence reporting outcomes but are less involved in the day-to-day process. At the European level, the European Commission has great power because it sets climate policies and runs the Innovation Fund. This fund gives large grants to projects that make industry more sustainable and is critical for state aid, as mentioned above. Large investors and banks have the power to approve or block projects financially. They watch for clear signs of government and EU support before committing money, and they want information that reduces financial risks. This shows another cause-and-effect link between actors in this power landscape.

The third group, low power but high interest, includes actors who are deeply involved in reporting but have less influence over big decisions. Globally, the World Steel Association (WSA) represents steel producers, industry groups, and research bodies. It shares production data, promotes sustainability measures, and spreads best practices that influence how steelmakers can present their performance, but it does not control compliance rules. Inside steel manufacturing organizations, managers and operators are also in this group. Operators work in areas like sustainability, engineering, and IT. They collect and prepare the technical and environmental data for reports, making sure it is accurate and traceable. Managers oversee this process, decide what goes into the reports, and coordinate between departments. Both have a high interest in the reports, as the process affects their work. This group's significance is further discussed in *Chapter 2.3*.

The last group has little power and little interest, needing the least attention, but should nevertheless be monitored. In this system, the actor is the Environmental Services (Omgevingsdienst). When air quality is good, they stay in the background and are only loosely involved in compliance reporting. But if air quality drops below legal limits, they can quickly step in, increase checks, and play a much bigger role. Seen together, these actors form a landscape where decisions at one level can influence all others: a decision in European Commission funding can limit or encourage national subsidies, which in turn can affect investor confidence and slow projects down.

This research was conducted within a steel manufacturing organization undergoing a large-scale transformation to align with European climate policy and Dutch national sustainability targets. Like many actors in high-emission sectors, the organization is under pressure to reduce its environmental footprint while remaining economically viable and eligible for public funding. Achieving this transition requires extensive internal reporting to justify green investments and demonstrate policy alignment. These reporting requirements introduce new dependencies between departments and place significant demands on internal coordination and data governance.

To support this effort, a dedicated coordination team was established in 2024 to assist departments in producing strategic documents and investment reports. This team does not hold ownership over the underlying data, nor does it deliver operational outputs. Instead, its role is to synthesize inputs from across the organization, bringing together technical, financial, and environmental data into integrated reports. The thesis is written from within this team.

This thesis focuses on one high-impact report used to attract long-term investment planning for the steel manufacturer's sustainability transition, mentioned earlier in this section. This internal document brings together emissions data, cost projections, technical feasibility assessments, and alignment with national and European policy instruments. It serves multiple purposes: guiding internal decision-making, informing negotiations with external stakeholders, and demonstrating eligibility for public and private funding. While this report is central to the analysis, it is embedded within a bigger reporting system that includes documents for funding and compliance.

This report does not aim to evaluate a company's entire data infrastructure or conduct a full assessment of data maturity across an organization. Instead, it focuses on how data governance challenges emerge during the coordination and delivery of complex internal reports. Particular attention is paid to how departments align their data contributions and how reporting processes are shaped by organizational structure and external accountability demands. By narrowing the focus to a single coordination team and its reporting process, this study enables a detailed, embedded analysis of

data governance practices in context. At the same time, the challenges are not unique to this organization. They reflect broader issues faced by other industrial companies navigating the transition to sustainability under increasing compliance demands. As such, the insights generated here may apply to organizations in other large, regulated, and data-intensive industrial environments.

1.3 Knowledge gap

The steel manufacturing industry plays a central role in Europe's decarbonization efforts, but faces significant challenges in adapting to rising compliance reporting and data governance demands. While academic literature has studied the technological and process-based aspects of this transition, such as improving energy efficiency (Prashar, 2017), integrating digital tools like AI (Branca et al., 2020; Gajdzik & Wolniak, 2021), and deploying emissions monitoring systems (Branca et al., 2020), less attention has been paid to how data is coordinated, validated, and structured across departments to support sustainability-related compliance reporting. As steel manufacturers face rising demands to produce internal reports that meet regulatory, financial, and sustainability standards, it becomes increasingly important to understand how data governance can support the coordination and credibility of these reporting processes.

Existing research on data governance in the steel industry has largely concentrated on improving production processes through technical innovation. Studies have examined the use of real-time sensors to monitor operations and detect faults (Gajdzik & Wolniak, 2021; Mishra et al., 2023), as well as broader digitalization efforts for reducing emissions and increasing efficiency (C. Zhang et al., 2024). Broader frameworks such as CMMI have been referenced in relation to industrial data governance (Hajoary, 2023), but their use in practice is often limited to finding operational areas of improvement rather than compliance reporting for sustainability ends. What remains underexplored is how steel manufacturers can govern and align data across departments to produce strategic internal reports, particularly those required for sustainability compliance, funding applications, or long-term investment decisions. Few studies address how data is validated, version-controlled, or synthesized to meet the steel industry's demands for data traceability and accountability.

This gap has become increasingly urgent as regulatory requirements and funding conditions raise the stakes for sustainability-related compliance reporting. Under instruments such as the CBAM, Corporate Sustainability Reporting Directive (CSRD), ETS, and the EU Innovation Fund, organizations must provide traceable, auditable, and policy-aligned data to qualify for funding and demonstrate compliance. As described in *Section 1.2*, these reports are crucial to securing major investments, critical for sustainable transitions, and require input from multiple departments across the organization. While these external pressures have grown, academic research has not yet addressed how internal data governance practices support such reports in practice.

This thesis contributes by examining how data governance supports the coordination of sustainability-related compliance reporting in an industrial organization under pressure from government bodies. It focuses on how departments align data, validate inputs, and meet external demands for credible, policy-aligned reports. The research offers practical insights into data governance practices that are often overlooked in current literature. To guide this effort, the study is driven by the following research question:

> *How can steel manufacturing organizations leverage data governance to support the transition to green steel production?*

A detailed description of the methodology is provided in *Chapter 2*. The next section outlines this study's societal and academic relevance.

1.4 Societal and academic relevance

This study aims to contribute both to academic understanding and to the practical challenges faced by steel manufacturing organizations during their sustainability transition.

- *Societal relevance* Industrial organizations are under growing pressure to reduce emissions while meeting strict reporting requirements from governments, investors, and funding bodies. This study offers practical insight into how internal data governance can improve the credibility, traceability, and alignment of sustainability-related reporting. More effective coordination between departments helps organizations qualify for subsidies and green investments more quickly, accelerating the transition to lower-emission production methods. This has broader societal benefits, particularly for communities near industrial sites, where emission reductions directly impact air quality and public health. In addition, the findings support the implementation of climate policies by showing how reporting requirements can be operationalized within large organizations. Although the case focuses on the steel industry, the data governance challenges and solutions addressed here are relevant to other heavy industries facing similar pressures to decarbonize, report transparently, and act in line with climate goals.

- *Conceptual contribution* This thesis contributes to the academic understanding of data governance by addressing a largely overlooked aspect: how governance practices support internal coordination for sustainability-related compliance reporting in steel manufacturing organizations. Most existing research focuses on infrastructure, digital tools, or operational efficiency. This study shifts the focus to the data governance criteria that shape the credibility of internal reports. It brings conceptual clarity to how data governance functions under real-world reporting pressure from external bodies, especially in large industrial organizations responding to regulatory and investor demands.

- *Methodological contribution* This thesis combines literature research, an embedded case study, and a focus group to study how data governance supports sustainability-related compliance reporting in a complex industrial setting. While much academic research focuses on high-level frameworks or technical infrastructure, this study offers a rare inside view of how data governance practices function across departments under real reporting pressure. Through triangulated data collection, including interviews, internal documents, and field observations, the study captures how governance criteria found in the literature research are applied in practice. Unlike studies that describe data governance as a technical framework or set of rules, this research shows that it is a coordination process shaped by how departments interact, who takes responsibility for data, and how organizations respond to external reporting demands. This approach contributes to academic research by aiming to show how data governance can be studied in practice. It highlights that successful sustainability reporting depends not only on digital systems but on clear roles, collaboration, and coordination within the organization.

1.5 Relevance to CoSEM

This thesis fits closely with the goals of the TU Delft's MSc program Complex Systems Engineering & Management. It deals with a complex sustainability challenge that combines technical systems, organizational coordination, and institutional pressure. The study is rooted in a real socio-technical

environment, where data governance affects not only technical infrastructure, but also how multiple different actors coordinate to achieve emission reduction efforts.

The research applies a CoSEM-style problem analysis by identifying barriers to sustainability reporting in a large industrial organization. By focusing on internal reporting for decarbonization investments, the thesis contributes to the management of multi-actor systems inside an industrial setting, where engineers, finance teams, sustainability officers, ministries, and EU bodies all have different goals and timelines.

This work is directly connected to the Energy track within CoSEM, as it explores how data governance supports large-scale energy-related investments like new blast furnaces and carbon reduction strategies. Through its embedded case study, the research gives insight into how real coordination happens in industrial practice. By combining technical, organizational, and policy perspectives, the study reflects the interdisciplinary approach at the core of CoSEM.

1.6 Structure

This thesis continues with *Chapter 2*, which outlines the research methodology. It explains the choices behind the three-phase research design: literature research, an embedded case study, and a focus group. The chapter also describes how data was collected, and governance criteria that guide the thesis were extracted. Because the literature review was designed and conducted as part of the methodology itself, rather than as a separate theoretical chapter, the report's structure begins with the methodology. *Chapter 3* presents the literature research, including the search strategy, the eight extracted data governance criteria, and the theoretical propositions that guide the analysis. These criteria form the foundation for both the results and the recommendations. *Chapter 4* presents the results, organized by the eight governance criteria and based on a triangulation of interviews, internal documents, and observations. *Chapter 5* builds on these findings by offering targeted recommendations and assessing their relevance beyond the case context. Finally, *Chapter 6* presents the study's conclusions and discussion, and reflects on its academic and societal contributions, limitations, and suggestions for future research.

The findings of this study help steel manufacturing organizations improve how they coordinate compliance reporting during the transition to more sustainable production. This is essential for meeting policy targets, qualifying for funding, and reducing emissions in the steel and related heavy industries. The next chapter will explain this thesis's methodology.

2 Methodology

As mentioned in the last chapter, this research aims to bridge the knowledge gap by answering the main research question:

> *"How can steel manufacturing organizations leverage data governance to support the transition to green steel production?"*

This research question focuses on how data governance can support sustainability-related compliance reporting. Steel manufacturing organizations are increasingly required to produce traceable and reliable data to apply for subsidies, demonstrate emissions reductions, and meet regulatory standards. This requires input from multiple departments and alignment across teams, which can be difficult without clear data governance practices. Rather than designing a new framework, this study investigates how governance criteria, found in the literature research, can improve the credibility and coordination of those reports. In doing so, the research shows how data governance plays a key role in enabling the transition to green steel production.

To understand the practicalities of leveraging data governance to accelerate the achievement of an organization's sustainability goals, this study adopts a threefold qualitative approach, consisting of literature research, a case study of Tata Steel Nederland (TSN), following the definition outlined by Yin (2018). Insights from Eisenhardt (1989) and Swanborn (2010) are used to enforce the choices made during the case study; this is further explained in *Section 2.2*. The case study makes the impact of data governance on an organization's sustainability goals tangible by applying theoretical concepts to a real-life challenge. TSN's current data governance practices are the core of the case study. Lastly, a focus group is conducted for expert validation of this thesis's results and recommendations. To find the most comprehensive answer to the main research question, it is divided into three sub-questions:

Sub-question 1: *Which governance criteria support sustainability-related compliance reporting by addressing key challenges in steel manufacturing organizations?*
Research method: *Literature review*
Literature research comprises "sequential steps to collect, know, comprehend, apply, analyze, synthesize, and evaluate quality literature to provide a firm foundation to a topic and research method (Levy & Ellis, 2006, p. 182)." This sub-question is answered by first identifying key challenges in data governance through existing literature, after which eight data governance criteria are extracted from the literature to operationalize key challenges, making it possible to apply them in a real-world case to answer sub-question 2. The search for literature concerning data governance centers findings around concepts like challenges and criteria, following Levy and Ellis (2006, p. 184) saying: "An effective and quality literature review is based upon a concept-centric approach rather than a chronological or author-centric approach (Webster & Watson, 2002)."

Sub-question 2: *"How can steel manufacturing organizations apply data governance criteria to improve sustainability-related compliance reporting?"*
Research method: *Case study*
Where the theoretical data governance criteria were found to answer the first research question, the second requires this methodology's case study section to assess them practically on Tata Steel Nederland (TSN). Doing a case study is relevant for answering this question because this thesis wants "to understand a real-world case and assume that such an understanding is likely to involve

important contextual conditions pertinent to your case" (Yin & Davis, 2007, p. 45). This question is answered by triangulating data from internal documents, interviews, and field observations, allowing the study to assess how each governance criterion plays out in practice and how it can be tailored to address the case's specific needs for data governance alignment.

Sub-question 3: *"To what extent can the recommendations of this study be applied to other steel manufacturing organizations and organizations beyond the steel industry?"*
Research method: *Focus group*
According to Kitzinger (1994, p. 1), "Focus groups are group discussions organized to explore a specific set of issues, such as people's views and experiences of contraception. (Barker & Rich, 1992)" She adds that "the group is 'focused' in the sense that it involves some kind of collective activity, such as debating a particular set of questions." In the case of this study, the focus group debates this thesis's recommendations and their broader application. For scientific purposes, this thesis tests broader applicability to industries beyond the steel manufacturing industry. Many related sectors, like automotive and other heavy manufacturing industries, have needs similar to the steel industry. This question consults experts from the Transformation Office team at Tata Steel Nederland (TSN) through a focus group to evaluate the recommendations and discuss further applications. This question validates prior phases, as well as identifies limitations for other industries.

By giving answers to these research questions, this thesis contributes to theoretical and practical research in data governance within the steel manufacturing industry and beyond. This thesis also offers recommendations and actionable insights to Tata Steel Nederland (TSN) as the acting case study to leverage data governance more effectively in its organization. Green steel initiatives are growing in importance with a lot of pressure from government emission deadlines, which require steel manufacturing organizations to ensure data traceability, accessibility, and security.

This chapter will explain the three elements of the thesis's qualitative research approach, starting with the literature research in *Section 2.1*. Then, *Section 2.2* justifies the selection of TSN as the case in this study. Next, *Section 2.3* will outline the case study by defining the case, revisiting the scope of the study, explaining the timeframe, its embedded elements, and the narrative to which this case study is designed. In *Section 2.4*, the data collection methods will be elaborated on, after which the interview protocol will be discussed in *Section 2.5*. Next, the data analysis plan for *Chapter 4* is highlighted, and the focus group is introduced in *Section 2.6*. Lastly, in *Sections 2.7 & 2.8*, following Yin's way of conducting a case study, the study's validity, reliability, and ethics will be highlighted. The remainder of this thesis will reference back to this chapter's methodological details where relevant.

2.1 The literature research's search strategy

This section will first explain the goal of *Chapter 3*, the literature research. Secondly, the literature research's search strategy is explained. *Figure 2* visualizes how this literature research informs the core of the thesis: challenges identified in the literature are distilled into eight data governance criteria, which are then applied to structure the methodology and evaluate practices in the case study. *Figure 2* is further explained below. The next section highlights the research strategy.

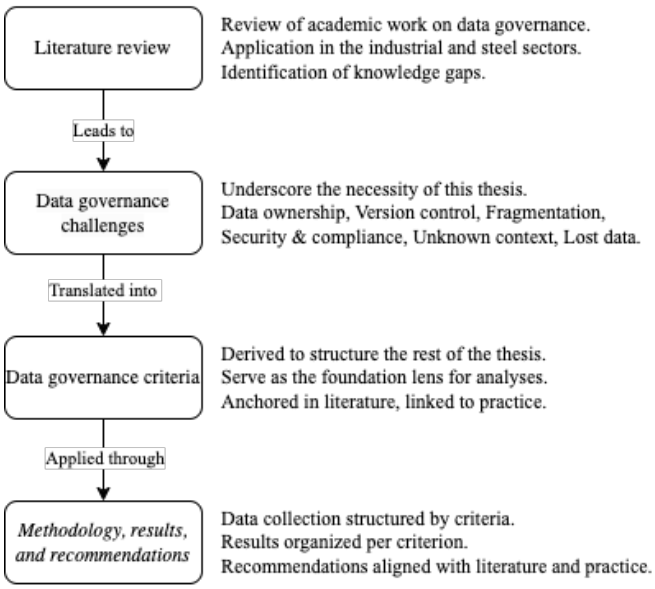


Figure 2: This flowchart illustrates the conceptual logic of the literature research. It begins with a review of academic literature on data governance, which reveals a set of recurring challenges. These challenges are distilled into eight data governance criteria to ensure usability in the rest of the thesis. These criteria serve as the analytical foundation for the case study, guiding the structure of data collection and the development of recommendations. The diagram demonstrates how academic insights are translated into evaluative tools for real-world application.

To build a strong foundation for this thesis, a structured literature search was conducted to identify relevant academic work on data governance, as well as references, particularly within industrial and sustainability contexts. This section outlines the search strategy, including the databases used, search strings, and exclusion criteria. By transparently describing this process, the section clarifies how the selected literature informs the thesis. It ensures the theoretical grounding of the eight data governance criteria and supports the study's focus on improving data governance practices at steel manufacturers to accelerate their sustainability goals. The final selection consists of 66 academic sources, which together shape the conceptual and practical lens of this research. The search strategy is visually shown using a PRISMA diagram in *Figure 3*.

Scopus and ScienceDirect(SD) were chosen as the two scientific search engines for this research. The search strings are listed below. The terms and Boolean operators were chosen carefully to ensure the extraction of the most relevant literature in the knowledge area.

- "Steel industry" AND "data governance" AND ("sustainability" OR "emissions reduction")
- "Sustainable production" AND "data governance"
- "Data maturity model" AND "sustainability"
- "Data lake" AND "steel industry"
- "Digital transformation" AND "Steel industry" AND "data governance"

- "Manufacturing industry" AND "digital transformation" AND "data governance" AND "framework"
- "CMMI" AND "maturity framework" AND "steel industry"
- "Emissions" AND "steel industry" AND "sustainability" AND "data governance"
- "Capability maturity model integration" AND "heavy industry"
- "Data governance" AND "steel"
- "Sustainable transformation" AND ("steel" OR "industry") AND "governance"

These search strings ensure a holistic collection of literature to cover the entire scope of this thesis study. By incorporating "steel industry", "data governance", and "sustainability" with keywords digging up related content, the selected literature comprises all topics needed to give a comprehensive answer to the research question mentioned in the *Introduction*. The search strategy's next part involves the exclusion criteria based on which the most relevant literature is selected. The three exclusion criteria are:

- **Year of publication**: To only consider the most relevant literature in this rapidly expanding field of study, articles written before 2010 are excluded from this literature research.
- **Literature type**: Only scientific articles were considered except for a book from Gajdzik et al. (2024). Reviews and conference papers were excluded from the selection. This ensures the inclusion of original research and empirical studies.
- **Specific relevance**: To include only the most relevant literature in the research, the search string must appear in the title, abstract, or author-specified keywords.

These exclusion criteria brought back the literature yield from an initial 3499 to 96 articles. After applying the exclusion criteria, a manual selection strategy was used to ensure the most relevant final selection. During this phase, snowballing was applied, a method where the researcher extracts relevant literature from existing articles' reference lists. *Figure 3* shows the PRISMA Flow Diagram, which visualizes the search and selection process for the search strings listed above.

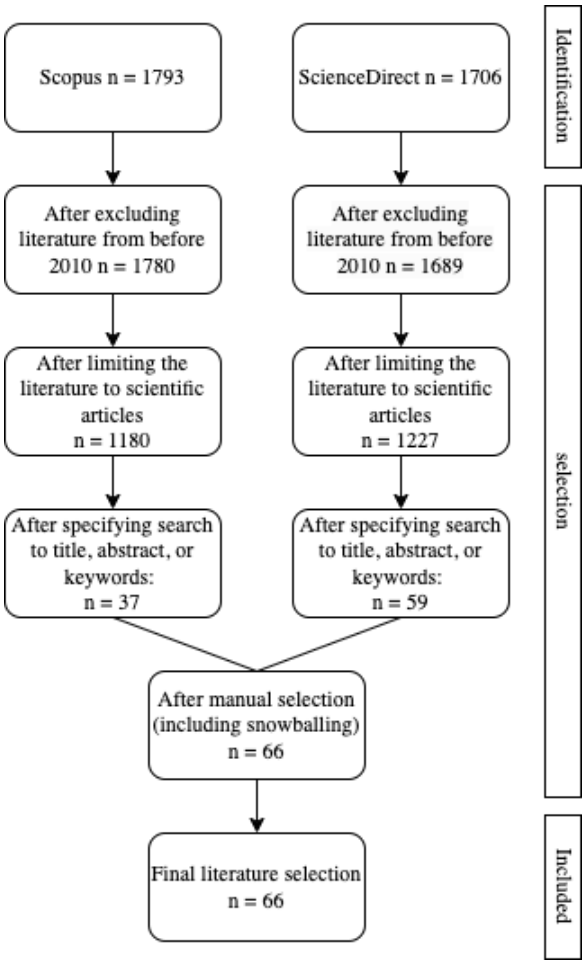


Figure 3: PRISMA Flow Diagram for literature selection

The initial database queries generated 3499 articles across Scopus and ScienceDirect. Filtering by publication date removed only a small number of sources, reflecting the fact that most academic work in this domain was published after 2010, showing that industrial data governance is still an emerging field. Next, the limitation of the literature to scientific articles narrowed the selection further. After applying all filters and snowballing to cover as many relevant works as possible, a set of 66 academic articles remained. These sources form the base of the data governance criteria developed in *Chapter 3*. The next section justifies the case selection for the case study in *Chapter 4*, applying the findings of the literature research that was conducted based on this search strategy.

2.2 Case selection and justification

After the literature research has derived eight governance criteria, a case study is conducted on Tata Steel Nederland (TSN) to assess and improve data governance in a real-life case. This case selection is justified in this section. The case study functions not only as a methodological approach but also as the primary source of empirical insight in this thesis. Yin (2018, p. 35) emphasizes that case

studies are powerful tools for understanding complex social and organizational phenomena, enabling holistic, in-depth research. Case studies offer the opportunity to analyze operational processes over time through documents, interviews, and observations that are central to this methodology. Furthermore, the use of a case study allows the researcher to examine theoretical propositions derived from the *Literature research* and assess how they apply within a real-world organization. Swanborn (2010) adds that a case study should be understood as the in-depth investigation of a bounded system, in which multiple data types are systematically analyzed within their organizational environment. Eisenhardt (1989) highlights the value of conducting case studies for theory-building, emphasizing the importance of incorporating insights grounded in real-life empirical studies. The data governance criteria are assessed using data collected in this chapter. Choosing the right case is described by Yin as a process, listing several rationales upon which to base the choice of a case. TSN fits the critical rationale for choosing a single-case study, as it is an organization in which the case study can "help to refocus future investigations in an entire field (Yin, 2018, p. 85)." In this case, it helps to set an example for improving data governance within the steel manufacturing industry and beyond.

This thesis chooses an embedded single-case study, as the scope comprises several different teams, each practicing their data management differently. The smallest subunits are participants in interviews. Different units of analysis together form the case study on TSN. Adopting an embedded single-case design as defined by Yin (2018) is appropriate when the research seeks to understand a contemporary phenomenon within its real-world context. According to Yin, case studies are the preferred method when addressing "how" or "why" questions about "contemporary events over which the researcher has little control." This study fulfills all three of Yin's criteria for conducting a case study: it asks a "how" question, seeking to answer this study's second research question: *"How can steel manufacturing organizations apply data governance criteria to improve sustainability-related compliance reporting?"* TSN deals with a contemporary issue and does not involve manipulation of behavioral events. As stated above, it also fits the rationale of a critical case: TSN is not only a representative steel manufacturer but also one whose trajectory can influence future directions in both industry and academia.

Tata Steel Netherlands (TSN) occupies a critical position within both the Dutch industrial landscape and the broader European steel sector. As one of the largest carbon emitters in the Netherlands, historically accounting for approximately 8% of the national CO_2 emissions, (Tata Steel Nederland, 2023b) TSN's environmental footprint alone makes it a significant subject for academic and policy-oriented inquiry. What elevates its relevance further is the company's ambitious Green Steel plan, or project Apollo, which outlines a phased decarbonization strategy aimed at achieving carbon neutrality by 2045 (Tata Steel Nederland, 2023a). This transition includes large-scale technological shifts, such as replacing blast furnaces with hydrogen-based direct reduction facilities, and restructuring operational systems to meet evolving regulatory, financial, and stakeholder demands. The great industrial complexity of this project makes it a prototypical example of the challenges faced by heavy manufacturers. Moreover, at the time of research, TSN was operating under considerable external and internal pressure. Regulatory developments such as the CSRD and the temporary suspension of public subsidies coincided with significant financial losses and the announcement of approximately 1,600 layoffs. These developments shaped the organizational climate in which this study was conducted. Given these factors, TSN serves as a justifiable and strategically important case study for exploring how data governance can enable sustainable transformation in industrial environments.

Several selection criteria were designed to determine TSN's fit as the case for this study. The

following criteria were used to find the best-suited case for this thesis; they are further explained after the list:

- Operational in the steel industry
- Accessible for research and observations
- Open for interviews
- Relevant illustrative project
- Representative of heavy industry organizations for broader applicability

In selecting the case for this study, several practical selection criteria were applied. First, the case organization must operate within the steel industry, a sector that is both emissions-intensive and undergoing a significant sustainability transformation, making it highly relevant for studying the role of data governance in green transitions. Second, the researcher has secured access to internal documents and data governance structures, including process documentation and the Green Book project, which is critical for an in-depth case study. Third, there is confirmed interview willingness from key stakeholders across various departments and management levels, ensuring a diverse and representative data set. Fourth, the case is working on a relevant illustrative project for a transformation towards more sustainable processes, making it most relevant to answer the main research question. Finally, the company's size and complexity must make it representative of large industrial organizations facing similar governance and sustainability challenges, which increases the broader applicability of the study's findings. Together, these criteria confirm Tata Steel Nederland (TSN) as a strategically relevant and most suitable case for this thesis. The next section will give a more in-depth description of TSN as a case for this study. It will also highlight the case study's scope, as was explained in *Chapter 1*, this study doesn't aim to evaluate the company's entire data infrastructure or conduct a full assessment of data maturity across the organization. Instead, it focuses on how data governance challenges emerge during the coordination and delivery of complex internal reports.

2.3 Case definition

To define the unit of analysis, this study focuses on TSN's internal data governance practices and strategy. Subunits of analysis include different departments and management layers, whose roles and responsibilities in data management vary. The regarding departments are the Transformation Office and the Data Governance Team. The teams' functions are explained after *Figure 4*. Within these teams, respondents from different management layers are interviewed. This allows for a richer understanding of how governance operates across the organization. The case study's timespan is the current state of data governance practices, with forward-looking ambitions to be carbon neutral in 2045. A visual overview of this study's scope is shown in *Figure 4*.

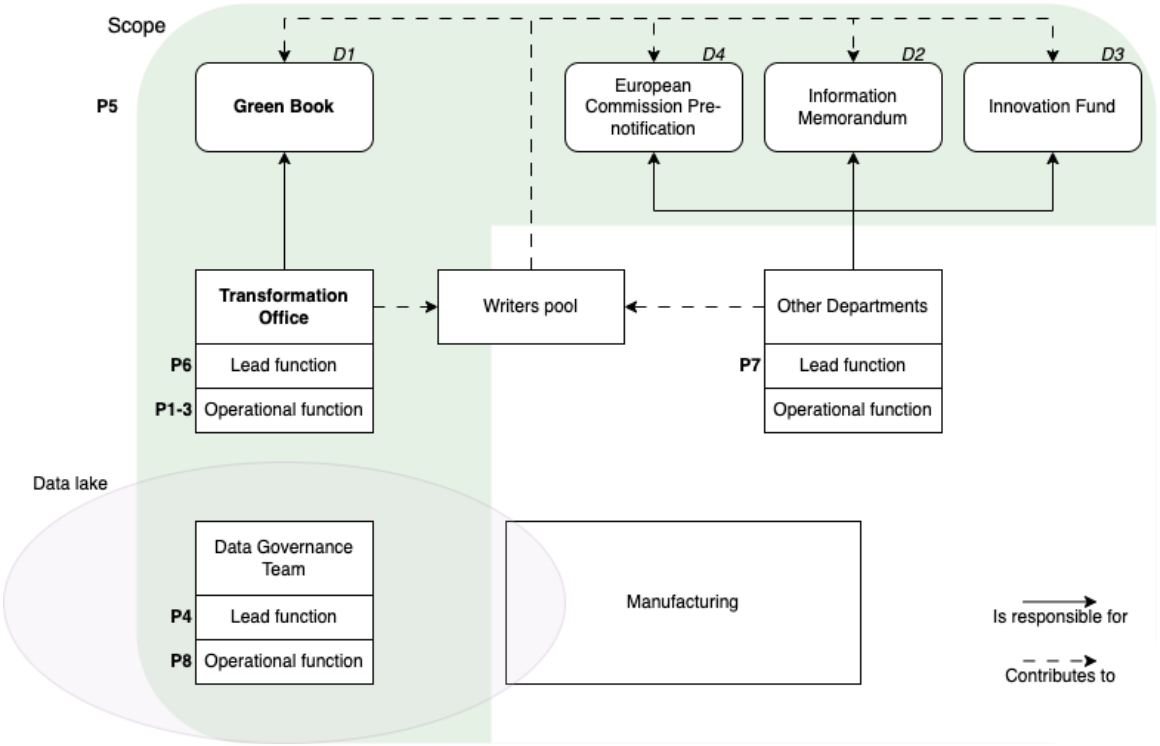


Figure 4: The organizational overview offers insights into where within Tata Steel Nederland this study's scope lies. The figure's outline represents TSN as a whole, the boxes represent this study's (semi)relevant departments and documents. The green area is the study's scope. The P's stand for the different interview participants. The D's represent documents 1-4, found in *Table 2*.

Figure 4 shows the organizational context and scope of this thesis within Tata Steel Nederland (TSN). The thesis is conducted from within the Transformation Office, which holds primary responsibility for the development and delivery of the Green Book (coded D1). This report is one of several key deliverables within TSN's broader transition strategy, alongside the European Commission Pre-notification (ECP, coded D4), Information Memorandum (IM, coded D2), and Innovation Fund documentation (IF, coded D3). While all reports share overlapping content and contributors, they differ in purpose, audience, and departmental ownership. *Tables 1 and 2* elaborate on the data sources and their relevance to this study.

The Green Book is supported by inputs from a writers' pool composed of contributors across various teams, including members of the Transformation Office and other departments. The Transformation Office leads the Green Book effort but works closely with other distributed contributors, many of whom also support the other three reports. Despite this collaboration, the Transformation Office retains the central coordinating role for the Green Book.

The Data Governance Team is organizationally distinct but relevant to this study. It is responsible for maintaining the data lake, a mature infrastructure used primarily for manufacturing data. Notably, this Data Lake is not used as a source system for the four reports, including the Green Book, highlighting a key governance disconnect examined in this thesis.

All involved teams, the Transformation Office, Data Governance, and other contributing departments, operate across multiple management levels, lead functions, and operational functions. These different layers influence how governance is implemented and how willing various actors are to adopt new practices. Interview participants from both these levels (P1–P8) have contributed to this research through interviews and internal discussions; they are marked in the figure. The scope of this thesis is shaded in green and includes not only the Transformation Office and Green Book process, but also the interdepartmental dynamics surrounding the production of related reports and the governance systems, or lack thereof, that support them.

The Green Book serves as a narrative for the case study. The Green Book is a report for TSN's mother company, Tata Steel Limited (TSL) in India, aiming to bring in the required investments to enable the transformation to greener steel and carbon neutrality in 2030. This report comprises 32 chapters and involves 15 responsible writers from different teams. The complexity of this project and the challenges it brings regarding data governance make it a good illustrative example for data governance practices within TSN. This report entails financing, technological, and environmental plans, among others, necessary to secure big investments. Therefore, this book relies on consistent data usage, requiring version control for all supporting documents, clear ownership, and strong security, while preventing data loss, fragmentation across multiple writers, and fading context. Its success is therefore contingent upon streamlined data governance structures and coherent data practices. Successfully handing over the Green Book results in a higher likelihood of timely approval for the investments needed for the transformation to greener steel, and therefore directly accelerates the achievement of sustainability goals. The next section will highlight the data collection methods for the case study on TSN.

2.4 Data collection methods

The data that was collected to complete this thesis originates from different types of sources, listed in *Table 1.* For this thesis, a total of 8 interviews, 21 documents, and regular team meetings and lunch conversations were used as the basis for the results. This ensures a holistic exploration for the answer to the main research question, and extensive knowledge of TSN and data governance practices. The different sources are coded into P's (participants), D's (documents), and O's (observations). 3 types, comprising 18 different sources, count up to 31 different data units on which this study is based. After the table, the sources' content is elaborated on.

Table 1: Data sources and description. The findings of this thesis's case study are based on this data. P stands for participant, D for document, and O for observation. *Chapter 4* will analyze the data further. They are discussed below.

No.	Coded Name	Job title / Source name
Source: Interviews		
1	P1	Transformation Officer
2	P2	Transformation Officer
3	P3	Transformation Officer
4	P4	DG (Data Governance) Lead
5	P5	Ex CIO (Chief Information Officer) of the Dutch National Government
6	P6	Transformation Office Lead

Table 1 (continued)

No.	Coded name	Description
7	P7	Long-term Assets Lead
8	P8	Elite Data Engineer
Source: Document analyses		
9	D1	Internal project documents Green book
10	D2	Internal project documents Information Memorandum
11	D3	Internal project documents Innovation Fund
12	D4	Internal project documents EC Prenotification
13	D5	11 Green book writers' workfiles
14	D6	Document control register
15	D7	4 Data management Powerpoints made by P1
16	D8	Datapoint sheet
Source: Observations		
17	O1	Participation in team meetings
18	O2	Lunch conversations
Total data sources:		
Interviews: 8		
Documents: 21		
Observations: 2		

This thesis uses a qualitative data collection strategy, combining both primary and secondary sources listed in *Table 1* to gain an understanding of data governance at Tata Steel Nederland (TSN). The primary data consists of internal documents and semi-structured interviews conducted with key internal stakeholders, including transformation officers, a lead from the DG (Data Governance) team, a former Chief Information Officer of the Dutch National Government, and an elite data engineer. These respondents were selected through purposive sampling to reflect a cross-section of hierarchical levels and departmental perspectives, ensuring that the insights gathered represent the diversity of data-related roles and experiences within TSN. As Yin (2018) advises, selecting interviewees should be strategically aligned with the research objectives, ensuring that participants are positioned to provide information relevant to the propositions, sub-questions, and literature.

In addition to interview data, the study uses data from internal documents, including TSN's Green Book, Information Memorandum, Innovation Fund, and EC Prenotification reports. Where the Green book aims to secure investments from TSN's mother company in India (Tata Steel Limited), with the business side of their current practices as well as future operations. The EC Prenotification is the start of an application for State aid from the Dutch State, which highlights the transition's impact on people living in the surrounding areas. The Information Memorandum (IM) is meant to reel in investments from banks, focusing on the business model and revenue streams. Lastly, the Innovation Fund (IF) is a European award for innovative companies, rewarding green initiatives with up to €250 million. The Dutch Ministry of Climate and Green Growth also uses participation in the IF as a requirement to apply for State aid. Together, these documents provide crucial information for understanding the Green Book's context, data needs, and reporting regarding the transition to green steel. Further, the data sources include the document control register, which helps map the formal structure of data ownership and governance. Workfiles from Green Book contributors and internal PowerPoint materials on data management are also reviewed to understand how data-related decisions are communicated and shared across teams. Lastly, the datapoint sheet is an important source of data. In the datapoint sheet, all numerical datapoints from costs to deadline years are

kept in a sheet made by the researcher, covering the four internal reports. This sheet is then used to detect inconsistencies within the four reports (D1-D4). If figures or numbers differ across reports, stories are not straight, and questions about certain datapoints may not be answered correctly. Inconsistent documentation leads to a loss of credibility, potentially postponing investments.

Complementing the interviews and documents as a secondary data source are field observations. These include participation in internal meetings and informal lunch conversations with the Transformation Office members P1-3 and P6, which offered unfiltered insights into TSN's data culture, routines, and collaboration between teams. Together, these data sources provide a holistic empirical foundation to triangulate findings and support the case study's validity, ensuring meaningful insights into how data governance is currently implemented and how it might be improved at TSN to support its transformation to green steel production. The next section will explain the interview protocol to collect hands-on data for the case study.

2.5 Interview protocol

The goal of conducting interviews in this thesis is to gain in-depth insights into TSN's current data governance practices from the perspectives of different management levels and departments. While the literature review established general challenges and criteria for evaluating data governance at steel manufacturers, the interviews are designed to translate theory into concrete, observable practices within the organization. The interviews aim to assess how data governance is currently structured and what major barriers exist. In addition, interviewees are asked to evaluate the relevance and feasibility of different framework features. These conversations not only provide narrative depth for the case study but also help to learn how different roles look at data governance at TSN.

To gain a well-rounded understanding of data governance practices at Tata Steel Netherlands (TSN), this thesis conducts semi-structured interviews tailored to different organizational levels and roles. For example, interviewees included multiple Transformation Officers (P1–P3) and a Transformation Office Lead (P6), whose roles often focused on governance vision, cross-departmental coordination, and high-level compliance reports. Transformation Officers could focus more on the broad structure, while the Transformation Office Lead might bring more tactical insights about how governance principles are or are not applied in daily workflows.

The interview questions were grounded in the data governance criteria for effective governance as synthesized in *Section 3.3* of the literature research, such as data unit responsibility, version control, interoperability, and security. For example, respondents were asked to describe "how is version control technically handled in your system?" (P8) and "Does it happen that different teams work with different versions of data?" (P2), aligning directly with literature on data challenges like Abraham et al. (2019) and Gierend et al. (2023). Other questions, like "What are the biggest obstacles in implementing a better data governance model and where does resistance lie within the team?" (P6), extract structural barriers cited by Al-Ruithe et al. (2018). The concept of fragmented systems, a key theme from Janssen et al. (2020), was explored through questions such as "What are the risks of data fragmentation?" (P8) and "What data-related processes or systems should be better aligned?" (P6). These helped assess how governance, or the lack thereof, affects coordination and interoperability. Several questions highlighted the link between governance and sustainability performance, for example: "Do you see a direct link between better governance and achieving sustainability goals?" (P8), which were inspired by studies such as Branca et al. (2020), Di Foggia and Beccarello (2024), and Otto (2011), who emphasize the importance of governance for emissions tracking and reporting

reliability. Lastly, questions like "How does the process of data ingestion work for you?" (P8) and "How is the Green Book put together?" (P2) focus on the process of data intake, reflecting criteria tied to data classification and stewardship, both highlighted in frameworks like DAMA-DMBOK and CMMI. By aligning interview themes closely with criteria found in the literature and real-world pain points identified from documents and observations at TSN, the protocol ensured depth and direct relevance to the results discussed in *Chapter 4*.

The interview protocol also engaged with roles from other teams, such as a Lead from the Data Governance (MDC) team (P4) and an Elite Data Engineer (P8). These respondents were asked practical and system-level questions like: "What is the process of data ingestion like for you?" (P8) and "Can you explain what the Data Lake at Tata Steel is and how it is used?" (P4). These questions highlighted technical issues, documentation inconsistencies, and the lack of clear ownership that mirror governance challenges found in the literature, highlighted by Bernardo et al. (2024) and Su et al. (2022). The technical nature of these interviews was particularly useful for evaluating how data governance challenges like data ownership and lost data affect day-to-day business operations.

A recurring theme across all interview levels was the willingness to change. Both strategic and operational interviewees were asked to reflect on the organization's adaptability, with questions like: "In which layer of management do you think change readiness is lowest and highest?" (P1) and "What do you think should be the first concrete step to improve data governance within the team and who should be involved?" (P6) and "What do you think are the biggest challenges in data governance implementation within Tata Steel?" (P8). Responses suggested that while leadership should recognize the need for structured governance to successfully implement, operational roles often face day-to-day obstacles that make data governance feel like an "overhead" rather than an enabler. The insights from these interviews informed the weighting of framework criteria.

In total, eight participants were interviewed, with each interview covering between 10 and 17 questions, depending on the participant's role and availability. The interviews lasted between 15 and 30 minutes; this was due to both scheduling constraints and the technical limitations of the transcription tool used, Turboscribe.ai, which supports audio files up to 30 minutes long. While the structure of the interviews was consistent, the tone and level of engagement varied between participants. Some respondents showed a high level of interest and openness, allowing for more conversational and in-depth discussions, while others were more reserved or had less time, which occasionally limited the number of questions asked. These elements sometimes influenced the interviews and will be further explained in *Chapter 6*. Despite this, all data governance criteria identified in *Section 3.3* were successfully addressed across the interviews. The interviews yielded numerous insights that contribute meaningfully to the empirical foundation of this thesis.

Overall, this interview protocol is designed to allow for targeted questions for each interviewee's role within TSN. The interviews will be triangulated with internal documentation and observations, and are intended to serve as a primary input for the results presented in *Chapter 4*. This approach guarantees that the collected perspectives will holistically uncover the link between the literature, TSN's data governance practices, and the organization's sustainability goals.

After the literature review and interviews are conducted, the data analysis plan is executed, the results are constructed, and actionable recommendations for TSN are presented; a focus group is held as defined by Kitzinger (1994) to validate the recommendations and debate the broader application of this study's findings to increase the study's academic contribution. Kitzinger (1994) suggests

that focus groups work best as interactive group conversations, where participants respond to each other and build on each other's ideas. This interaction helps uncover shared beliefs, disagreements, and assumptions that might not surface in individual interviews. The researcher should guide the discussion without controlling it, using open questions to keep the conversation going. A good mix of participants, those with shared experience but different perspectives, can lead to richer discussion. In this thesis, the focus group will be used for two main goals: first, to validate the recommendations based on the case study findings; and second, to explore whether the recommendations could also be applied in industries beyond steel. Following Kitzinger's method, this group setting is expected to offer not only practical feedback but also valuable academic insights into how the recommendations for improving data governance can be adapted and applied in other contexts. Participants P1-3 and P6 will form the group, as they are all part of the Transformation Office, where the researcher operates. The conversation will start by presenting the results of this study. In this way, the researcher also has information input, which stimulates the participants to put in information themselves. In the next section, the data analysis plan for the results in *Chapter 4* will be explained.

2.6 Data analysis plan

This section outlines the structured approach used to analyze and interpret the qualitative data collected through the two primary research methods: the literature review and the case study of Tata Steel Nederland (TSN). The central objective of this analysis is to evaluate the extent to which TSN's current data governance practices support its sustainability transition and to form actionable recommendations that best fit the organization's needs.

As explained in this chapter, this study follows a qualitative research design. It combines insights from internal documents, interviews, and observations. These data sources will subsequently build on each other per criterion in *Chapter 4*. The analysis captures contextual and experiential insights from various stakeholders within TSN, whose roles are listed in *Annex IIA*. This method ensures that the results are grounded in the real-world conditions of the organization.

The data analysis is carried out in four sequential steps: (1) outlining the problem using internal documents, (2) analyzing interview transcripts based on the data governance criteria, (3) triangulate findings across all data sources, and lastly, (4) comparing the results to the theoretical propositions discussed in *Section 3.5* through pattern matching.

The data analysis plan of this thesis draws on a combination of methodological principles from Eisenhardt (1989), Swanborn (2010), and Yin (2018), offering a comprehensive approach to understanding complex organizational phenomena. Yin (2018) emphasizes the role of theoretical propositions and pattern matching in linking qualitative data to the research questions. This study adapts Yin's general strategy of relying on theoretical propositions and supports it by analyzing results using pattern matching. Swanborn (2010) holds up this study's analysis plan by highlighting the importance of using analysis of data from three sources: documents, interviews, and observations in that order. Swanborn (2010) complements this by stressing the importance of designing case studies within the boundaries of a social system, such as a company. By combining these authors' perspectives, this study ensures robust construct validity, analytical depth, and a strong connection between evidence and the resulting recommendations for Tata Steel Nederland. Next, the data analysis plan's five steps are further explained.

2.6.1 Step 1: Problem outline based on documents

Swanborn (2010) repeatedly emphasizes the order in data analysis for case study research. According to his book, the process should start with document analysis, followed by interviews for pattern matching, and finally, the use of observations to support the findings. In line with this structure, this study begins by conducting a document analysis to outline the context of the problem at TSN. As indicated in *Table 1*, access was gained to eight document types, offering a wide-ranging view of TSN's current data governance structures and challenges. The documents and their contents are further explained in *Table 2*. One of the most critical documents is the Green Book, which serves as the narrative backbone of the case study, detailing TSN's strategic transformation plans and the role of data in supporting that transition. Other documents, such as the Information Memorandum, internal PowerPoint decks, and the document control register (DCR), provide supporting evidence about how data is managed across different parts of the organization. Together, these documents make it possible to build a detailed and factual overview of TSN's current data governance practices. They reveal important shortcomings in areas such as formal governance structures, the traceability of data throughout the organization, and the coordination between different departments. In the next steps, the findings from interviews will be systematically compared to this baseline to either confirm or challenge the initial understanding of TSN's data governance situation.

Table 2 gives an overview of the 21 analyzed documents and their functions. Together, they represent the most relevant sources for understanding how data governance is practiced in strategic planning, sustainability reporting, and operational coordination at TSN. The output comprises an evidence-based narrative of how TSN currently applies data governance, and a summary of key bottlenecks, including a table of findings per criterion in *Section 4.9*. These findings are systematically recorded using a document coding matrix with four columns:

- Document type (e.g., Green Book, control register)
- Page or section (specific reference)
- Identified example (what is seen or missing)
- Link to challenges/criteria from *Chapter 3*

This coding framework ensures that every document example is traceable to a specific data governance challenge or criterion and allows direct comparison with the same themes identified in the interviews. To conduct the document analysis systematically, all findings were first recorded in an Excel sheet with four columns as described above, and linked to a relevant data governance criterion. This process led to a total of 70 identified examples across the document set, ranging from inconsistencies in data reporting to examples of unclear ownership, fragmented documentation, and gaps in version control. Below, the complete document set will be defined and described. The complete list of observations can be found in *Annex III*.

Table 2: Overview of documents (D1-D8) used in *Chapter 4* for analysis. The table shows the document code, name, and description of their functions and relevance for this study.

Code	**Name**	**Description**
D1	Green Book	This internal strategy document outlines TSN's vision for becoming a sustainable steel producer. It is structured into 32 chapters across 11 workfiles, each written by a different expert, and presents the business case, roadmap, and key enablers needed to secure long-term investment from the parent company, Tata Steel Limited, in India. It is the central narrative document of TSN's green transformation.
D2	Information Memorandum (IM)	The IM is a finance-oriented document intended for potential external investors, especially banks. It outlines TSN's future business model, expected returns, and risk mitigation strategies, making the case for financial backing beyond internal stakeholders.
D3	Innovation Fund (IF)	This document is a formal application for a major European sustainability grant. It demonstrates TSN's innovative capabilities and climate impact in an effort to secure up to €250 million in funding to support green initiatives and technologies. It's also a condition for applying to the Dutch State's aid.
D4	ECP	This report is part of TSN's formal request for state aid from the Dutch government. Its focus lies more on societal and environmental impact, especially for communities around the plant, and explains how the transition will meet both national and European sustainability policy objectives.
D5	Workfiles GB	These are 11 preparatory documents created by the individual authors of the Green Book. They include working notes, intermediate calculations, data sources, and communication with co-authors, like comments. These files help trace how information was gathered, interpreted, and transformed into the final Green Book, making them highly relevant for assessing traceability, ownership, and version control in TSN's current data practices.
D6	Document control register	The Document Control Register is an internal Excel file that tracks the development, ownership, and status of the Green Book, Blue Book, and Innovation Fund (IF). It lists every chapter or workfile, its responsible owners, and includes a status dashboard that monitors progress. This register is essential for understanding how TSN structures and coordinates documents across departments. It also provides insight into traceability, version control, and responsibility within the organization's data governance practices.

Table 2 (continued)

Code	Name	Description
D7	Powerpoints by P1	This document comprises 4 PowerPoint decks about the current and desired data governance practices at TSN and its shortcomings, including a proposed framework, vision, and roadmap. The documents specifically focus on classification and security, making them most relevant for this thesis.
D8	Datapoint sheet	The Datapoint Sheet is a structured Excel file created by the researcher to extract and organize quantitative data from key internal documents, specifically the Green Book (D1), Information Memorandum (D2), Innovation Fund application (D3), and the EC Prenotification (D4). Each row captures a unique data point, with columns specifying the source document, index reference, thematic category (e.g., CO2 emissions, hydrogen use), numerical value, unit of measurement, date, source reference, and a brief description. This tool supports consistency and traceability in document analysis and helps identify inconsistencies, missing data, or gaps across TSN's strategic reporting.

The next step after analyzing the 70 observations of data governance practices at TSN is comparing them with findings from the interviews conducted with eight sector specialists. How this was done is explained in the next step.

2.6.2 Step 2: Criteria-based analysis of interview data

The second step involves systematically analyzing the semi-structured interviews conducted with eight stakeholders from across TSN's organizational structure and beyond. These interviews were designed to give insights into the current state of data governance, pain points in implementation, and willingness to change. Respondents included transformation officers, data engineers, and team leads involved in sustainability reporting and strategic decision-making.

The analysis follows a criteria-based coding approach. This aligns with Eisenhardt (1989, p. 540), who warns for smothering in piles of data using Pettigrew (1988)'s quote: there is an ever-present danger of "death by data asphyxiation." Categorizing the data helps get an overview and for the pattern matching in the next step of the data analysis plan. The criteria are derived from *Chapter 3*, emphasizing the data governance challenges and criteria and the willingness to change.

Results will be developed by carefully reviewing the interview transcripts and identifying patterns, ideas, or phrases that were not included in the initial set of predefined criteria. These could include, for instance, references to the Green Book being used as a governance tool, signs of tension between engineering and governance teams, or examples of informal practices that bypass formal data processes. Swanborn (2010) supports this method by highlighting that organizations should be viewed as complex social systems where different ways of working and interpreting rules coexist. This approach ensures that the coding remains closely tied to TSN's specific context, while also capturing how different parts of the organization relate to each other within a larger system.

The process will be conducted using a structured approach, in which each interview is mapped to the relevant criteria identified in *Chapter 3*. Responses from different stakeholders will be compared across roles to identify how perceptions of data governance vary between strategic and operational levels. For instance, strategic-level respondents (e.g., P4, P6, P7) may focus on policy gaps and long-term integration, while operational-level respondents (e.g., P1-3, P8) may highlight tool limitations, inconsistent documentation, or insufficient operability, which directly influences their ability to work. The data will directly inform the remaining analytical steps and ultimately, the formation of actionable recommendations for TSN's data governance practices to accelerate their sustainability goals.

2.6.3 Step 3: Triangulation of interviews, documents, and observational data

The next step of the data analysis is triangulation, which means comparing different types of data to make the findings more reliable and grounded. Rather than relying only on what interviewees said, this step checks those insights against internal documents and real-world observations. Yin (2018, p. 170) highlights that triangulation strengthens the credibility of case study research by ensuring that the results are based on more than one kind of evidence, building on the first principle of data collection: use multiple sources of evidence. Swanborn (2010) adds that combining documents, interviews, and observations in that order is especially useful in complex organizations, where formal structures and informal practices often operate side by side. In the same way, Eisenhardt (1989, p. 538) supports this by stating that triangulation "provides stronger substantiation of constructs and hypotheses."

In this study, triangulation is applied by comparing what was learned from the interviews to what is shown in internal documents and what was observed in daily work situations. The documents analyzed include internal project reports such as the Green Book, Information Memorandum, and process-related files as listed in *Table 2*. These sources help to clarify how data governance is documented and communicated. For example, if an interviewee claims that data flows are not standardized across teams, the researcher checks for signs of this in the document control register or by identifying inconsistencies in the Green Book. These inconsistencies are also noted in the researcher's datapoint sheet, which tracks such issues systematically across the four key documents.

Observations from meetings and informal conversations provide an additional layer of insight. These include how teams talk about data issues, how responsibilities are shared, and whether there is openness or resistance to change. Often, these cultural and behavioral elements are not mentioned in interviews or written down in official documents. Swanborn (2010) emphasizes that such informal interactions are crucial for fully understanding how things work in practice.

Altogether, triangulation supports the earlier steps in the analysis by checking whether key themes listed in step 2 show up across multiple data sources. It also helps identify where there may be differences between what people say, what is written down, and what happens. This layered approach allows for a more realistic and critical view of TSN's data governance and ensures that the pattern matching in the next step reflects both theory and reality.

2.6.4 Step 4: Pattern matching with theoretical propositions

The final step of the analysis involves pattern matching, a central analytic technique in case study research as described by Yin (2018). This method compares empirically observed patterns from the interview data to a set of theoretical propositions formulated in *Section 3.5*. These propositions

serve as conceptual anchors that help structure the analysis and determine whether the empirical case at TSN supports or challenges the expectations derived from the literature.

Both of the propositions will be translated into testable indicators, as recommended by Eisenhardt (1989, p. 548), who stresses that "a strong theory-building study yields good theory (that is, e.g, testable)." Any (partially) confirming, unexpected, or contradictory patterns will be noted and used in the conclusion in *Chapter 6*. Pattern matching in this study is not only a test of theory but also a bridge between conceptual assumptions and organizational practice, supporting the iterative logic of case-based research. The next section will highlight this study's validity and reliability measures.

2.7 Validity and reliability

To ensure the methodological rigor of this study, extra attention is paid to Yin (2018, p. 58), where he describes four key dimensions of research quality: (1) construct validity, (2) internal validity, (3) external validity, and (4) reliability. Construct validity is supported by the use of multiple data sources, including semi-structured interviews, internal documents, and field observations, which are systematically cross-referenced through triangulation and listed in *Table 1*. Internal validity is strengthened by applying a combination of explanation-building and pattern-matching techniques, whereby empirical observations are compared with theoretical propositions developed in *Chapter 3*. This strategy enables the study to identify causal links between data governance practices and sustainability outcomes within TSN. External validity is addressed during a focus group in *Chapter 5* to validate this research's findings and justify the case's broader relevance: as a major steel producer undergoing a complex green transition, TSN represents a typical case for heavy industrial organizations facing similar pressures to decarbonize and improve data governance structures. While findings are not statistically generalizable, they offer analytic generalizability to comparable contexts, such as other European manufacturers with legacy infrastructure and evolving sustainability mandates. Finally, reliability is achieved through the use of consistent research procedures, including documented interview protocols, structured coding instruments, and a maintained research log across the entire thesis period. These measures ensure that the study can be replicated with the same procedures and would produce similar findings if conducted by another researcher.

2.8 Ethics

To conclude *Chapter 2* with a quote from Yin (2018, p. 125): "A good case study researcher, like any other social scientist, will strive for the highest ethical standards while doing research." Ethical considerations have been integrated into every stage of this research to ensure responsible data handling. All interviewees were fully informed about the nature and purpose of the study and provided explicit consent to participate according to the Human Research Ethics Committee (HREC) standards (Delft University of Technology, n.d.). Their identities have been anonymized in this thesis to protect their privacy and ensure that no personal or role-specific information can be traced back to individual respondents. Given the sensitivity of the data involved, such as internal reports, working documents from the Green Book project, and email correspondence, careful attention was paid to secure data storage and handling. These materials were accessed and analyzed following established research ethics and data protection standards by the HREC. In addition, the study strictly adheres to the non-disclosure agreements signed with the researcher, Tata Steel Nederland, and the TU Delft, ensuring that no confidential or proprietary information is disclosed. With these ethical safeguards in place, the analysis presented in the remainder of this thesis can confidently build on a foundation of academic integrity and participant trust. The next chapter presents the first phase of this thesis's

methodology, literature research to identify key challenges in data governance and extract relevant governance criteria.

3 Literature research

The goal of this chapter is to identify the academic and practical knowledge gaps in the field of data governance, with a specific focus on large industrial organizations transitioning toward sustainability. To address this, the chapter conducts a structured review of the literature on data governance, its associated challenges, and its role in the steel manufacturing sector. This chapter aims to answer sub-question 1 and uses a method as outlined in *Section 2.1*. This review forms the basis for extracting eight data governance criteria, which are used in the remainder of this thesis to evaluate current practices and develop targeted recommendations.

The chapter begins by defining key concepts such as data governance and green steel to establish a consistent vocabulary with definitions that remain valid throughout the thesis. In *Section 3.2*, recurring challenges in the implementation of data governance are discussed to understand what obstructs effective practice. These insights are then applied to the steel manufacturing industry in *Section 3.3*, and eight governance criteria are derived from the challenges to ensure practical usability. These criteria structure the rest of the thesis, where they are used to evaluate current governance practices through a case study in *Chapter 4*. Next, *Section 3.4* identifies knowledge gaps in the literature. These include a lack of research on cross-departmental data governance practices, limited attention to data governance for compliance reporting and sustainability disclosures, and insufficient focus on behavioral and organizational factors that influence the successful implementation of data governance in steel manufacturing organizations. These gaps explain the academic contribution and direction of this thesis. Finally, *Section 3.5* compiles two theoretical propositions to guide the case study's results and root the remainder of this thesis in existing literature.

3.1 Key concepts and definitions

This subsection ensures clarity and consistency in the use of key phrases throughout the thesis report. In this thesis's field of study, a few terms are used frequently. It is most important to ensure the consistent use of terminology without synonyms and to ensure one clear definition of the phrases that stays valid throughout the entire report. The definitions are substantiated by existing literature and form the terminological foundation for the rest of the thesis.

3.1.1 Green steel

Green Steel is a term describing a sustainable initiative to produce steel without the use of fossil fuels (Forum, 2023). The definition this report uses is slightly different. According to TSN's current steel-making practices, the blast furnace uses iron ore and cokes (coal) to produce pig iron. In a further process, a great amount of oxygen is blown through the pig iron to lower the number of connections between the cokes and the iron, resulting in higher-quality steel. Instead of carbon bonding with iron, the carbon bonds with the oxygen (O2), resulting in CO2 emissions. In the new blast furnace, running from 2035, natural gas is used as the fuel for melting iron ore. With the same process, this results in more hydrogen bonding with oxygen and less carbon bonding with oxygen. This, together with CCS, Carbon Capture & Storage technique, steel should be produced with a minimum of 30% less CO2 emissions (Tata Steel Nederland, 2023b). The next step is developing hydrogen blast furnaces, potentially achieving a carbon-neutral steel production by 2045 (Tata Steel Nederland, 2023b).

3.1.2 Data

According to the National Institute of Standards and Technology (NIST) (2025), data is a "representation of facts, concepts, or instructions in a manner suitable for communication, interpretation, or processing by humans or by automatic means." This encompasses all types of data and is a very generic definition. In this study, data has two forms. The first form entails documents used to substantiate reports for decision-makers and investors like the European Commission or large banks. These reports are crucial for obtaining approval for big decisions and policies for the transition to green steel, containing estimates or sensor data on emissions, economic feasibility, and technological requirements. The second form is numerical and related to manufacturing data, used for process optimization. With the transition to a more sustainable business model, such as using natural gas or hydrogen instead of cokes, the company's economic and environmental parameters differ significantly. In many cases, assumptions are made about the amount of a specific component needed for the organization's production after transitioning. These assumptions may differ across teams, and for many assumptions, no number is baselined. For example, the amount of iron ore needed for steel is crucial for calculating operational costs, product output, and CO2 emissions. These numbers fluctuate due to ongoing reorganizations within organizations and ongoing speculations on the transition to a more sustainable business model. As a result, constant revision of numbers is needed for the currently needed resources, investments, and expected production.

3.1.3 Data governance

Data governance is a specialized area within information architecture that manages data and its associated metadata as an asset of an enterprise in a comprehensive manner, involving data management, data quality, data stewardship, data policies, and data security (Luisi, 2014, p. 206). "The main objective of data governance is to enhance a company's capability, effectiveness, and sustainability by supporting its decision-making processes with robust, high-quality data, leading to more precise and informed choices (Bernardo et al., 2024, p. 13)." In the context of this study, data governance is controlling the management of data, because the growing volumes of data from different sources cause inconsistencies in the data. This is visualized by Ladley (2012) in his 'Governance V' depicted in *Figure 5*. Data governance enables the identification and addressing of these inconsistencies before decisions are made based on incorrect data (Abraham et al., 2019). In the steel industry, steel quality plays a big role, and it directly benefits from real-time monitoring and data collection (Gajdzik & Wolniak, 2021). Furthermore, according to the Paris Climate Agreement, the steel industry must cut 45% of its carbon emissions by 2050, and data governance contributes to making this goal achievable (Gajdzik et al., 2024).

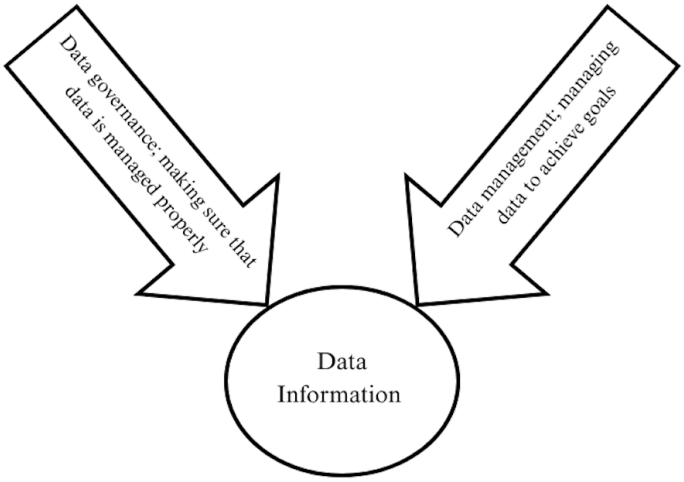


Figure 5: Governance V model by Ladley (2012)

3.1.4 Data governance framework

Organizations use data governance frameworks to implement data governance in a structured way. This ensures the alignment of technological advancements with economic and environmental goals (Shirwa et al., 2025). Enterprises use 30% of their time on practices that do not add value to their organization due to low-quality and unavailable data (Petzold et al., 2020). A large part is due to poor data governance (Q. Zhang et al., 2022). Improving data governance by structuring data into frameworks is of great importance to the productivity and sustainability of an organization. A few examples of Data Governance Frameworks are DAMA-DMBOK, CMMI for data management, and ISO 50001 (Prashar, 2017). In *Table 3*, the relevant data governance frameworks are listed. In *Chapter 5*, this thesis's recommendations, the different frameworks will be linked to data governance criteria-based recommendations for the steel manufacturing industry.

Table 3: Available data governance frameworks for the manufacturing industry, these will be scored on weighted criteria in *Section 4*.

Framework	**Publisher**	**Focus**	**Reference**
DAMA-DMBOK	DAMA International	Best practices for enterprise-wide data governance, quality, and security.	(DAMA International, 2017)
DSMS	(Nahshon et al., 2024)	Stakeholder communication following FAIR principles	(Nahshon et al., 2024)
ISO 38505	ISO	High-level enterprise data governance principles and policies.	(Zorrilla & Yebenes, 2022)

Table 3 (Continued)

Framework	**Publisher**	**Focus**	**Reference**
CMMI for Data Management	CMMI Institute	Maturity model for data management and analytics.	(Thomas et al., 2019)
DCAM	EDM Council	Data governance maturity assessment for enterprises.	(Liu, 2023)
FAIR Principles	(Wilkinson et al., 2016)	Ensuring data is Findable, Accessible, Interoperable, and Reusable.	(Wilkinson et al., 2016)
DGI Framework	DGI	Focuses on policies, processes, roles, and responsibilities.	(Al-Ruithe et al., 2018)
DMM model	DMM	Evaluating data management practices at an organization.	(Thomas et al., 2019)
ISO 8000	ISO	Data quality, interoperability, and master data management.	(Al-Badi et al., 2018)
Industrial Data Space (IDS)	Fraunhofer Institute	Secure and sovereign industrial data sharing.	(Alexopoulos et al., 2023)
MES Data Governance	Industry-Specific	Governance for a manufacturing executed system's generated operational data.	(Costa et al., 2024)
GRI (Global Reporting Initiative)	GRI	Standardized sustainability and ESG reporting.	(Singh et al., 2007)
NIST Data Governance	NIST	Best practices for big data governance and security.	(National Institute of Standards and Technology (NIST), 2025)
GAIA-X Industrial Data Governance	GAIA-X	European cloud data governance for industry.	(Alexopoulos et al., 2023)
ISO 5001	ISO	the international energy management standard	(Prashar, 2017)
Git	Git	captures the full state of a collection of files as part of a project and captures changes to that state as the content of files is altered.	(McCaffrey, 2020)
Data-Centric	Various publishers	separates data from applications and technology platforms, which avoids the appearance of data silos and enables data to be shared and used by the entire organization	(Zorrilla & Yebenes, 2022)

Table 3 presents a selection of data governance frameworks that are relevant to industrial and manufacturing contexts. These frameworks differ in scope and emphasis: some focus on organizational governance and maturity models (such as DAMA-DMBOK, CMMI, and DCAM), while others address technical principles (like FAIR and ISO 8000), industry-specific use cases (such as MES governance and Industrial Data Space), or sustainability-related standards (such as GRI and ISO 5001). Together, they illustrate the diversity of approaches available to structure data governance practices. Although this thesis does not adopt any single framework as its foundation, its components serve as useful references when identifying criteria and formulating practical recommendations. The next section identifies data governance challenges that exist within the industrial environment.

3.2 Challenges in industrial data governance

This section explores the main challenges associated with data governance and extracts recurring bottlenecks from academic literature that form the foundation for the data governance criteria. Understanding these obstacles is essential for tailoring governance solutions to the realities of complex organizations, especially during the case study. By identifying the most pressing challenges, this section defines the practical needs that any governance framework must address to support reliable data use and guide an organization toward higher data maturity. This section serves as input for defining the data governance criteria in *Section 3.3*, which also offers a visual overview of the literature per data governance challenge.

3.2.1 Data ownership

Data owners are the people in an organization responsible for a certain unit of data (Abraham et al., 2019). We lack knowledge of how an organization identifies its data owner, for example, by the location where a unit of data is stored. This makes determining data owners a difficult task (Abraham et al., 2019) and thus a challenge in data governance. First, it is important to determine at what management level an organization should appoint a data owner. Vilminko-Heikkinen and Pekkola (2019) say that data ownership is often regarded as the task of any CIO, but as data spreads out over every business branch of an organization, it should be determined at the office level. Secondly, many data conflicts are related to the data owner's responsibilities. The owner's responsibility should extend to the control of the data's content, functionality, and quality (Vilminko-Heikkinen & Pekkola, 2019). Therefore, a well-designed data governance framework should be able to allocate its data owners, along with their responsibilities (Zorrilla & Yebenes, 2022).

3.2.2 Version control

Version control is the practice of capturing changes made to a unit of data and labeling it as the last modified version, creating a history for each file. An important element for reproducible research (Gierend et al., 2023). This is essential to manage large engineering projects of any type and makes change management much easier because personal memory fades quickly, and team communication can become complicated (McCaffrey, 2020). To ensure this, version-control methods must be applied for data that changes over time, and data owners must be informed of those changes (Muenzen et al., 2022). To do so manually, cross-teams in an organization is a big challenge, but there are tools like Git, mentioned in *Table 3*, that organize data and track updates of the data (McCaffrey, 2020, p. 286), making sure version-control is managed, contributing to overall data governance.

3.2.3 Lost data

The definition of data loss in the context of this study is inaccessible data and is fully captured in the term "data silo". Data silos appear when managed stacks of data, or repositories, are inaccessible to others within the same team or organization. This leads to 'lost' data for the people who can't access the data and to inefficiencies, hindering collaborations between those teams (Bernardo et al., 2024). In this way, data silos cause duplications of data, unclear responsibilities, and missing control over data. This is a common challenge within data governance. Two reasons for this happening are the ad-hoc handling of low-level data without procedures, processes, and data infrastructure (Janssen et al., 2020), and a lack of cross-team collaboration (Abraham et al., 2019). This challenge, too, is countered by implementing an effective data governance framework.

Without effective data governance, different teams and disciplines may repeatedly perform the same basic data collection. This leads to unnecessary duplication of effort because different teams choose inconsistent parameters when working on the same task because of inaccessible data, preventing uniformity and comparability in research outcomes (Su et al., 2022). Critical to disabling collaboration between parts of an organization is the use of traditional tools like Excel for handling data instead of cloud and data government models. Data mismanagement can easily lead to a lack of traceability and consistency (Corallo et al., 2023). Because of this inconsistency, the authenticity and integrity of the data are questioned. To overcome this challenge, data governance needs to support the traceability of data (Su et al., 2022).

To address silos and to enhance collaboration, data governance may be applied to accelerate the process of aligning data within a business (Q. Zhang et al., 2022). Thus, data governance models can avoid the appearance of data silos and enable data to be shared and used by the entire organization (Zorrilla & Yebenes, 2022). Therefore, further research should be done on how data governance frameworks can foster collaboration and avoid data silos (Abraham et al., 2019) to counter the challenge of lost data.

3.2.4 Fragmentation

Fragmentation is when data is without a unified governance structure and leads to untraceable data and other inefficiencies (Abraham et al., 2019). Like lost data, data silos caused by a lack of cross-team organization can result in fragmented data. Data is often fragmented across organizations that implement different data policies within their organization. This can result in unclear responsibility, diffused accountability, and unknown data quality (Janssen et al., 2020, p. 2).

Organizations lack uniform systems or data governance frameworks for consolidating data, which leads to a fragmented data landscape where separate systems exist in different departments. The situation results in challenges in data exchanges and interoperability in the organization (Suprun et al., 2024, p. 9). Because of fragmented data sources, businesses have challenges in accessing relevant data, because they come in inconsistent formats (Kavak & Rusu, 2025). Without structured data governance, past attempts to counter this challenge have failed. The reason for this is specific unstructured repositories without the wider support of the organization (Al-Ruithe et al., 2018). The result is a strong need for a data government framework.

Storing data in different places also raises security risks for any personal information, as the oversight over fragmented data is lost and data might not be consistent with each other and the states of the real world (Janssen et al., 2020, p. 6). Data governance plays a crucial role in countering these

challenges by actively targeting fragmented data and silos (Vilminko-Heikkinen & Pekkola, 2019), enabling data sharing and accessibility across an organization.

3.2.5 Data security

Data security is another issue that organizations often face. Building on data fragmentation, spreading data over multiple places within an organization, and losing oversight, brings security challenges. In the article from Bernardo et al. (2024, p. 15), Johnston (2016) identifies three fundamental topics for managing and maintaining a solid data governance strategy. Besides the first two pillars, data visibility (lost data) and federated data (fragmentation), the author emphasizes data security as one of the pillars of data governance. "Deleting low-value data that has little to no business value, such as outdated and duplicated data (version control), while organizing valuable data by storing it in controlled repositories, enforcing rigorous data security procedures and mechanisms to ensure data privacy and the protection of financial, personal, and other types of data." This problem particularly happens in cloud environments, since 94% of businesses use data computing (Edge Delta, 2024). Research indicates that governance-related issues account for 41% of cloud security problems, emphasizing the need for strong data governance that incorporates classification mechanisms to protect sensitive information (Al-Ruithe et al., 2018).

Therefore, data security is a crucial data governance challenge in organizations. A framework can ensure the quality and proper use of data, but it requires a mechanism for data security (Janssen et al., 2020). Implementing this not only assures security in an organization but also fosters innovation by ensuring that organizations can access and utilize their data effectively (Gillan, 2021). However, while there is an understanding that sharing data is necessary, the mechanisms for protecting against potential harm while maximizing usefulness are not well-defined. These two aims are often in tension and lend themselves to diverse data governance strategies (Muenzen et al., 2022, p. 11). Therefore, it should be further discovered how data governance can effectively tackle the challenge of data security.

3.2.6 Unknown context

The last challenge addressed in this section is the unknown context of data. Data itself has no meaning without the context of usage (Vilminko-Heikkinen & Pekkola, 2019). A cause may be mechanics, like data lakes, which can be both structured and unstructured, often with limited contextual information (Bernardo et al., 2024, p. 15). A suitable data governance framework should therefore encompass knowledge and data protection, increasing visibility (Bernardo et al., 2024). Similarly, Rivett (2015) describes that the context of data is crucial for data visibility. Visibility as a component of data governance allows organizations to recognize data in its context, and identify ownership (Bernardo et al., 2024). Rivett (2015) even says that an organization can't guarantee proper data governance without ensuring complete visibility of data, otherwise massive amounts of data are "thrown" into the lake, with little contextual information. This creates a data swamp.

Rivett (2015) states six questions that can be answered with sufficient context of data: 'What information is out there and what can I use? Who do I contact if the information looks wrong or if I need more?' 'Who owns/supports what?' 'How up-to-date is the information, and who is using it?' 'Does the data have consistent meaning?' And lastly: 'What are the technology dependencies and risks?'

Together, these challenges reflect the literature's most frequently cited obstacles to implementing effective data governance. They highlight structural issues such as poor version control, unclear roles, inconsistent classification, and limited context and visibility. *Table 4* visualizes the link between the derived data governance criteria and summarizes the challenges identified in this section across the reviewed literature. The next section applies the challenges identified in this section to the context of the steel manufacturing industry. It explores how data governance can support organizations in improving reporting and decision-making processes. The challenges discussed in this section form the foundation for the governance criteria used throughout this study. These criteria are used to assess current practices, guide interviews, and structure the final recommendations for steel manufacturers and other data-intensive industries aiming to accelerate their sustainability goals.

3.3 Data governance criteria for the steel manufacturing industry

After identifying the challenges of data governance, this section focuses on how those challenges play out in the steel manufacturing industry. Steel production increasingly depends on digital systems that generate and use large amounts of data. Examples include defect detection, material property predictions, and quality control for optimizing processes (Branca et al., 2020). Managing this data effectively is not only crucial for efficiency and process improvement but also for meeting sustainability targets and complying with regulations. This section first explains the specific role of data governance in the steel industry and then discusses how data governance is being adopted in practice. Finally, it connects the challenges identified in *Section 3.2* to a set of eight criteria that guide this thesis. These criteria are used to analyze current practices during the case study in *Chapter 4* and shape the recommendations developed in *Chapter 5*.

The steel manufacturing industry uses data capturing to obtain information from different processes. But understanding them in-depth to enhance decision-making is a second. After capturing, processing, and analyzing digital data, storing, structuring, and sharing must be applied. This allows better forecasting of process behavior as well as smarter, easier, and faster decision-making (Branca et al., 2020, p. 13). However, storing, processing, prioritizing, and analyzing vast amounts of information have been challenging for steel manufacturing organizations, while these issues are crucial for effective decision-making for sustainability matters among others (Singh et al., 2007).

The application of digital systems in various steel manufacturing production processes, such as energy, wastewater management, and emission tracking, enhances quality, costs, and environmental performance (Branca et al., 2020). The raw data that is obtained from these systems at the manufacturing level of a steel-making organization is the base for supporting documents of reports. An example of such a report is the Corporate Sustainability Reporting Directive (CSRD) by the European Commission. The CSRD requires companies of 1000 employees or more to disclose information on the influence their activities have on people and the environment (European Commission, 2022). The collected data can provide an extensive overview of the current sustainability practices of an organization, and even more importantly, it is critical to the way forward in decarbonization by forming the base of predictive models and the support for regulatory compliance reports like the CSRD. However, only if proper data governance is applied to counter the processing challenges discussed by Singh et al. (2007). Apart from the CSRD, supporting data is used for similar reports for local governments, banks, and overseas investors, as explained in *Chapters 1 & 2*. Consistent data accounting and reporting are critical for approval, accelerating the investment in green initiatives in steel manufacturing organizations. To achieve this, Bernardo et al. (2024, p. 13) says "data governance frameworks can be applied to ensure robust, high-quality data, leading to more precise

and informed choices (Sifter, 2017)."

Data governance emerges as a critical enabler for the transition to a more sustainable steel manufacturing industry. This can be used to address the barriers to the sustainable transition (Chinnathai & Alkan, 2023). The key to this is the implementation of a data governance framework to ensure homogeneous data within reports, tackling the challenges found in the previous section. A robust data governance framework should comply with strong criteria to counter those challenges. During the exploration of data governance frameworks, Al-Badi et al. (2018) found components of such a framework that counter most of the challenges. *Figure 6* shows the criteria for an effective data governance framework to implement in a steel manufacturing organization based on insights from this study, insights from Al-Badi et al. (2018), and existing literature. The figure is explained further below.

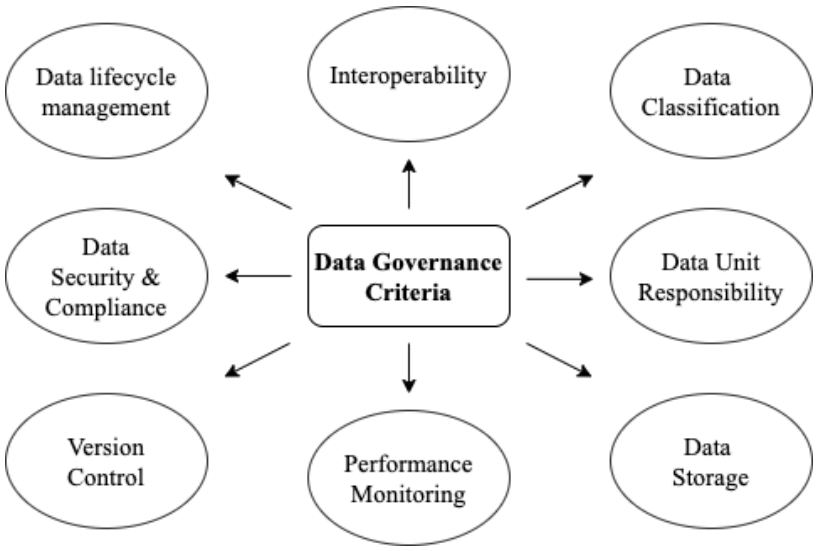


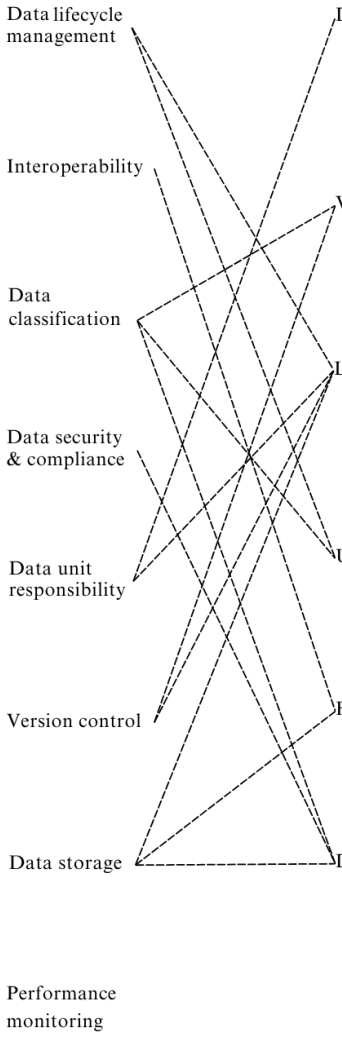
Figure 6: Criteria, derived from the challenges found in existing literature, for a data governance framework based on the challenges from *Section 3.2* and formulated by this study. A visual overview of the connection between the criteria and the challenges they address can be found in *Table 4*.

Figure 6 describes the eight criteria of an effective data governance framework for a steel manufacturing organization. *Table 4* shows a schematic overview of which criteria cover which challenges in data governance. The eight criteria are extracted from the challenges, grounded in the existing literature found during the thesis's literature research. The first criterion is data lifecycle management. It covers the challenges of lost data and unknown context. According to Vilminko-Heikkinen and Pekkola (2019), improving data governance comes with changing data responsibilities, roles, and ownership. Harnessing this change can be captured in data lifecycle management. Losing track of these changes can cause organizations to lose access to the context, purpose, or access to data units generally over time, hence the linked challenges. Data lifecycle management is described as one of the core values of data governance by Otto (2011), ensuring traceability across all stages, and eventually controlling the disposal of excess data. This too is crucial because, in the past years, the amount of created, consumed, and stored data has grown exponentially (Castro et al., 2024). To avoid data fragmentation, interoperability is extracted as one of the criteria for an effective

framework. As Janssen et al. (2020, p. 2) said, data is often fragmented across organizations that implement different data policies within their organization. Interoperability ensures the tuning of policies across teams, avoiding data fragmentation. Next is data classification to clarify the context of the data and data security. As Vilminko-Heikkinen and Pekkola (2019) mentioned, data has no meaning without context. Classification partly tackles this problem by labeling the data to make it traceable throughout its history. If an organization wants to access supporting data for any report, the framework should ensure that the data is easy to find for business or process topics. Another example of classification is whether data is publicly available across an organization or strictly confidential and accessible for a small group of people. This overlaps with the next criterion, data security & compliance. This criterion is selected because of the consequences unauthorized access may have. Data breaches and the exposure of personal information of employees are risks posed by insufficient data security. Data ownership is next in the data governance challenge and is countered by data unit responsibility. The framework should be able to give an organization's data an owner who, in the end, is responsible for the sharing, storing, and disposing of the data unit. As Al-Badi et al. (2018) mentions, data stewardship is a crucial component of data governance. The next criterion is version control, tackling the big challenge that is essential for managing large engineering projects of any type (McCaffrey, 2020, p. 286). Version control strongly correlates with data lifecycle management as it classifies data into metadata, making it easy to determine which units are outdated and ready for disposal, without the risk of destroying valuable data. It also plays a big role in consistency throughout an organization's reporting by making sure each report is supported by the same historical version of data. This is especially crucial in the steel manufacturing industry, where real-time data is updated frequently and forms the base for big process optimizations critical for the transition to a more sustainable steel-making process. Data storage is a fundamental criterion for effective data governance, tackling lost data and data fragmentation by securely storing data in a centralized way. Effective data storage is crucial for making data traceable, retrievable, and for regulatory compliance. Losing client information, for example, or valuable emission data, can cause drastic regulatory and reputational harm (United Nations Environment Programme Finance Initiative, 2023). It also stimulates interoperability by acting as one central storage space across an organization. Performance monitoring is a criterion that ensures the quality of the framework over time. Effective data governance involves the use of performance metrics to monitor data quality and system resilience (Lebaea et al., 2024, p. 1). The criterion focuses on the framework itself, and not on the steel manufacturer's data governance practices, which is why performance monitoring does not apply to one of the challenges. An organization's data governance practices should continuously monitor itself and the performance of every criterion by defining key performance indicators (KPIs) for data traceability and security.

Together, these criteria comply with Wilkinson et al. (2016) and his FAIR Guidance principles, findability, accessibility, interoperability, and reusability. This ensures a most suitable data governance framework for implementation in a steel manufacturing organization to make the next step towards a greener steel-making process. *Table 4* visually connects each of the eight data governance criteria to the challenges from *Section 3.2* they address. In addition, the table offers a literature overview by listing the academic sources that support each challenge.

Table 4: A schematic overview linking the eight data governance criteria to six key challenges derived from literature. Connecting lines indicate which criteria address which challenges. Supporting academic references are listed for each challenge to ground them in literature.



This section has translated the challenges found in data governance into eight practical criteria tailored to the steel manufacturing context, listed in *Table 4*. These criteria form the analytical foundation for this thesis. They are used to assess current practices at Tata Steel Nederland (TSN) to align the results in *Chapter 4*, and to structure the recommendations developed in *Chapter 5*. Together, these criteria summarize how the challenges identified in the literature can be addressed in practice. However, the process of defining these criteria also reveals important limitations in existing research. The next section highlights key knowledge gaps in the academic literature and explains how this thesis aims to contribute to the academic field of data governance in industrial sustainability transitions.

3.4 Identified gaps

To ensure this study makes a relevant academic and practical contribution, it is essential to understand what current literature has not yet addressed. This section identifies the main gaps in existing research on data governance in the steel manufacturing sector. By doing so, it clarifies why this thesis focuses on reporting practices, data maturity, and sustainability alignment, rather than purely technical or operational optimization. Identifying these gaps helps define the scope and purpose of the case study conducted at Tata Steel Nederland (TSN).

The steel manufacturing industry has been collecting, storing, and analyzing manufacturing data for decades (Mishra et al., 2023, p. 2) to use data-driven approaches to optimize steel-making processes and monitor quality and emission numbers. Despite decades of data processing, there has been a lack of research on countering challenges in data governance, specifically on the implementation of a data governance framework to manage this data effectively. Studies on data governance in the steel manufacturing industry have found insights into sensors that produce real-time data on production processes or asset defects, but they lack insights into a structured approach to governing data, specifically including data as supporting documents for regulatory, sustainability, and investor reports. The steel manufacturing industry's shift to a greener business model is not without pressure from local and international governing bodies. An example is the Corporate Sustainability Reporting Directive (CSRD), along with strict demands from international investors for transparent sustainability data, as explained in *Section 3.1*. These developments make structured and effective data governance increasingly important for supporting the green transition. To align data with regulatory reporting, strong data traceability, accessibility, and security are essential.

While existing data governance tools and frameworks like DAMA-DMBOK, CMMI for data management, and data lakes are already implemented, steel manufacturing organizations often fail to implement digital tools because of the lack of focus on steel manufacturing organizations (Hajoary, 2023). Data governance frameworks have not been tailored to the needs of the steel industry. "The slow pace of change in the industry and lack of stakeholder pressure for using the latest technologies for decarbonization is the foremost hurdle (Mishra et al., 2023, p. 23)." Existing studies with a focus on an industrial environment almost exclusively discuss how data governance can optimize production processes. A gap has proven its existence for steel manufacturing organizations to enhance their sustainability-related compliance reporting through data governance to support the transition to greener steel-making. This is underexposed in existing literature.

These gaps highlight the need for research that bridges operational data systems and strategic reporting within the steel industry. This thesis addresses that need by examining how practical governance criteria can improve decision-making, sustainability reporting, and cross-departmental

data alignment.

To guide this thesis further, the final section of this chapter formulates a set of theoretical propositions. These propositions are expectations grounded in the literature of this chapter and serve as the analytical lens for the case study at Tata Steel Nederland (TSN) in *Chapter 4*.

3.5 Theoretical propositions

This study uses a case study approach and follows the guidance of Yin (2018) by developing theoretical propositions to guide the design and analysis. These are informed expectations based on the literature, not statistical hypotheses. They help structure the data collection and interpretation in a focused and theory-informed way. The propositions link insights from academic literature to practical challenges in the steel manufacturing sector, particularly regarding the role of data governance in sustainability-related reporting. They shape the interview questions, inform the document analysis, and underpin the pattern matching in *Chapter 4*. Based on the literature review in *Sections 3.1–3.4*, the following propositions were formulated:

1. **Data governance must address operational fragmentation to enable reliable reporting.**
   Research shows that in the steel manufacturing industry, reliable sustainability data depends on good coordination between departments and legacy systems that are often not well connected (Gajdzik & Wolniak, 2021). When systems and teams work in isolation, it becomes harder to trace data, align information, and hold people accountable. In addition, sustainability reporting requires structured and consistent indicators across different areas, such as environmental and operational performance (Singh et al., 2007). This proposition will be assessed through findings on data lifecycle management, interoperability, and data storage through case study findings in *Chapter 4* and validated by the focus group in *Chapter 5*.

2. **The successful implementation of data governance depends on both structural measures and cultural readiness.**
   Abraham et al. (2019) emphasizes that effective data governance requires more than just formal procedures; it depends on organizational awareness and cultural support. Otto (2011) adds that individual ownership and willingness to change are critical success factors for implementing governance practices. This proposition connects to findings on performance monitoring and data unit responsibility, results from field observations, and insights from the interviews. This proposition is examined through case study findings in *Chapter 4* and validated by the focus group in *Chapter 5*.

The two propositions help focus on important structural and cultural patterns in the case. The validation of these propositions is revisited through the focus group in *Chapter 5*. While not every data governance criterion is directly linked to a proposition, all eight are examined in detail in the next chapter, the results, using evidence from interviews, documents, and observations.

4 Results from an embedded case study at TSN

This chapter presents the results of the findings from the literature review in *Chapter 3* and the case study conducted at Tata Steel Nederland (TSN). This chapter aims to help answer sub-question 2 by forming the base for the recommendations in *Chapter 5* and uses a method as outlined in *Chapter 2*. The goal is to provide a holistic and evidence-based understanding of TSN's current data governance practices to support the acceleration of TSN's sustainability goals, forming the empirical foundation for answering the main research question:

> *"How can steel manufacturing organizations leverage data governance to support the transition to green steel production?"*

The *sections 4.1–4.8* correspond to one of the eight criteria derived from the *Section 3.3*. The findings reflect how each criterion is currently addressed within the Transformation Office and its surrounding teams. All sections are based on the 21 documents, eight semi-structured interviews, and field observations made during the internship at TSN's Transformation Office. After the three analyses (documents, interviews, and observations) per criterion, triangulation was applied to extract patterns across the data sources to strengthen the findings. Lastly, each paragraph is concluded with whether the criterion's findings confirm their corresponding theoretical proposition formulated in *Section 3.5*. *Section 4.9* concludes the chapter with an overview of the most important findings that emerged during the analysis of each criterion. Together, these sections build toward the recommendation chapter and evaluation of broader applicability in *Chapter 5*.

4.1 Data lifecycle management

Data lifecycle management describes how data moves from creation to revision, approval, and eventual archiving. At Tata Steel Nederland (TSN), this process is present but not consistently formalized. This makes it difficult for teams to know which version of a document is current, complete, or approved.

Across interviews, documents, and observations, a clear pattern emerged: lifecycle stages are informally managed and inconsistently tracked. Interviewees often relied on personal judgment rather than shared systems: "I just know I have the latest version. But that's not structurally arranged" (P1). P6 noted that changes are made without explanation: "We change a number from 3 to 4 based on feedback, but no one records why or what the impact is." This informal handling was reflected in the documents, where version numbers and status labels were frequently absent or unclear, especially in high-stakes files like the Green Book, IM sheet, and Datapoint sheet. Observations during Green Book development reinforced this issue: in a meeting on April 25th, slides and Excel files were circulated without clear ownership or status, prompting concern from P6 about submitting unchecked material to senior management. One exception came from P8, who described how the Data Lake team requires a data owner and classification before anything is accepted, showing that formal lifecycle management is possible but not yet widespread.

Together, these sources reveal that while lifecycle management is conceptually understood, it lacks the structural support needed for reliable implementation. This undermines data traceability, slows collaboration, and increases the risk of errors in strategic reporting. This supports theoretical proposition 1: Data governance must address operational fragmentation to enable reliable reporting. Without clear agreements and shared systems, teams work in isolation, and important information can get lost or misaligned.

4.2 Interoperability

Interoperability concerns the ability to share and reuse data across teams, documents, and systems. At TSN, this includes aligning inputs across departments, coordinating related deliverables, and ensuring that data and documents are compatible across teams. This section draws on document analysis, interviews, and observations to examine how well interoperability is supported in current practice.

Interoperability issues at TSN primarily arise from fragmented authorship, inconsistent access, and unaligned data references. Interviewees described frequent collaboration without supporting coordination tools. One participant noted, "Everyone started working in the document at the same time. Then it became chaos" (P3). Another pointed out structural divides between departments: "Data governance works well for manufacturing, but HR, finance, and commercial work differently" (P4). These comments reflect the absence of shared standards or visibility across teams. The documents support this, showing misaligned sections, unclear references, and a lack of source traceability. In one case, progress could not be tracked because contributors used informal codes like "no update?" or "no one?" instead of structured status entries. In another, references were made to other reports, before being numbered, complicating integration. These coordination gaps were visible in practice. Teams sometimes worked on overlapping indicators without realizing it, and access to shared models like the CSM tool, a tool for documenting datapoints on long-term assets, was requested but not granted or followed up. As a result, efforts were duplicated and reports included inconsistencies that could have been avoided with clearer structures for reuse and alignment.

Together, these findings indicate that while collaboration is common, it is not systematically supported. Without shared access points, clear responsibilities, or aligned formats, interoperability remains a structural challenge. This supports theoretical proposition 1. When each department uses its own tools and standards, it becomes difficult to align their input and produce a consistent report.

4.3 Data classification

Data classification refers to how meaning and structure are assigned to data through categories, labels, formats, or file naming conventions. Effective classification helps improve clarity, traceability, and consistency across deliverables. This section examines how classification is currently applied at TSN, based on internal documents, interviews, and observations.

Across documents and interviews, inconsistent classification practices emerged as a common issue. Documents often showed classification systems being introduced but not maintained. In one case, a color-coded status table was presented at the start of a report but ignored in later sections. In another, an index in the Document Control Register did not match the structure of the Green Book itself, making cross-referencing difficult. Interviewees confirmed the lack of standardization. P2 noted, "There is no standard structure for how we name or store data," while P3 suggested that a predefined fact list per topic would simplify interpretation: "A standard fact list per topic. With AI tools, you could fill that automatically." In contrast, P4 highlighted the more formal structure within the Data Lake: "The data owner determines that. We, as the Data Lake team, set the process, but we do not determine the classification."

The absence of observational data on this topic suggests that classification decisions are embedded in digital workflows rather than visible in team behavior. Still, the consistency between documents and

interviews provides a clear picture: classification exists in parts of the organization but is applied unevenly, often depending on the habits of individual teams. This inconsistency leads to delays in interpretation, confusion over file purpose, and greater room for error when documents are reused.

4.4 Data security & compliance

Data security and compliance refer to the protection of sensitive information and the enforcement of access controls, especially where regulatory, strategic, or commercially sensitive data is involved. At TSN, ensuring that data is stored, handled, and shared in a secure and compliant way is important for internal governance and external accountability, and trust.

Inconsistencies in key figures across high-level reports suggest weak control over sensitive content. Strategic documents such as the Green Book, ECP, and IM sheet contained conflicting data points, raising concerns about whether updates are properly verified before distribution. Interviews confirmed that security and access are managed unevenly. P2 noted, "In the worst case, the investment doesn't happen because someone uploaded an outdated version," while P4 described secure environments where access must be formally requested: "For restricted data, we use buckets. Access has to be requested and assigned. At the same time, P6 acknowledged that some reports include market-sensitive content not always treated accordingly: "Some documents contain market-sensitive information that shouldn't be freely accessible." These reflections are supported by field observations. There was no process to review data security practices during a meeting. In the meeting, it was determined that P3 would be responsible for sharing maps with people outside of the Transformation Office, but no structured approach was provided. The absence of such a structured approach suggests a lack of data security embedded in everyday work practices.

Together, the evidence points to a partial security structure: formal controls exist in some systems, but strategic documents often move between teams without verification or access governance. This gap risks undermining both internal trust and external credibility.

4.5 Data unit responsibility

Data unit responsibility refers to the clear assignment of ownership over specific data, documents, or deliverables. Ownership ensures that data is updated, verified, and aligned across teams. At TSN, ownership is sometimes assigned, but not always maintained or communicated. This section explores how responsibility is recorded in documentation, how it is experienced in daily work, and how it affects coordination.

Across sources, ownership appears inconsistently applied and rarely enforced in shared reporting processes. Document analysis revealed outdated owner lists, such as in one case where documents were still attributed to people no longer active within TSN. In other files, multiple owners were listed or none at all, making it unclear who held final responsibility. Some comments tagged individuals by name across teams, reducing traceability. Content overlaps across files further complicates accountability, increasing the risk of conflicting updates. Interviews confirmed that responsibility is often assumed but not structurally documented. P2 noted, "The underlying documents are made public by the responsible leads for the Transformation Office," suggesting delegated responsibility without systematic oversight. P3 described a lack of continuity in reporting: "A new document is requested, someone makes a status list, and starts collecting data from scratch." In contrast, P4 pointed to stricter rules in technical environments: "Before data may enter the landing zone, there

must be an owner, a classification, and a steward." P6 highlighted the gap between expectation and reality: "In a mature organization, I expect we know where data is stored, that version control is traceable, and that there is an owner responsible for it." Observations reinforced these points. In one instance, multiple people were listed as contributors to a file, but no one was assigned to finalize it. In another, team members asked each other informally who was responsible for updating a key compliance document. These examples show how unclear ownership leads to reliance on hallway conversations, slowing work and increasing risk.

Together, these findings show that responsibility is often assumed rather than managed. While technical systems enforce ownership protocols, shared reporting processes rely on informal delegation or implicit roles. This weakens accountability, reduces document continuity, and complicates coordination across departments. These findings confirm theoretical proposition 2: The successful implementation of data governance depends on both structural measures and cultural readiness. Although structural rules for ownership exist in parts of the organization, they are not consistently followed or embedded in reporting routines. The lack of shared habits around responsibility shows that, without cultural support, formal rules fail to enable reliable reporting.

4.6 Version control

Version control refers to the ability to manage, trace, and differentiate between document or dataset versions over time. It supports consistency across teams, reduces duplication, and helps prevent errors in shared outputs. At TSN, version control practices are present in some areas but are not systematically enforced. This section evaluates how version control is handled across documentation, how it is perceived in practice, and how it affects daily collaboration.

Across all sources, version control emerges as a known but inconsistently managed issue. Interviewees described relying on personal judgment rather than shared systems. P1 stated, "I just know I have the latest version. But that's not structurally arranged," and P2 noted, "Different teams work with different versions." Under time pressure, coordination breaks down, as P3 observed: "People work in parallel, which leads to different versions." P6 framed the problem as one of maturity: "The process by which we log those changes is not mature." These reflections highlight that teams are aware of the risks but lack common procedures. Document analysis supported this with repeated examples of missing version indicators, misaligned updates, and conflicting file structures. One file incorporated newer data without marking it as such; another lacked validation entirely, raising concerns about uncontrolled edits in critical reporting flows. Observations confirmed the consequences: team members were unsure which document version to use, key files were overwritten without backups, and uncoordinated parallel editing caused confusion.

These findings consistently point to the same core issue: version control is understood in theory but not embedded in daily workflows. Informal methods such as filename tagging are used, but they lack the reliability of structured systems. In high-stakes reporting environments, this opens the door to data inconsistencies, duplicated work, and loss of trust in outputs.

4.7 Data storage

Data storage refers to how information is stored, accessed, and maintained across systems and teams. Reliable storage systems support version control, consistent data use, and long-term traceability. At TSN, data is stored in various environments, but centralization and coordination remain challenges.

This section evaluates how storage is managed in documentation, daily practices, and team perceptions.

Across sources, the lack of centralized storage emerges as a major barrier to consistency and coordination. Interviewees repeatedly pointed to the absence of shared infrastructure. P2 stated, "We should be storing it in a central repository like P1's document repository," and added, "Yes, because people don't use central storage." P3 noted that reused content often moves between deliverables without structure: "We use texts and figures from other deliverables, but that process isn't well structured yet." P8 emphasized that "if everyone uses the same source, you reduce inconsistency. You get centralized and consistent access." P6 stressed the importance of storage systems that go beyond basic tools: "In a mature organization, I expect that we store data somewhere, that versioning is traceable, and someone is responsible. These comments suggest a clear recognition of the problem but no established solution. Documents partially reflect this. While one example showed structured tracking of document status, most cases revealed inconsistencies between key figures across the Green Book, the ECP, and the IM report. Slight variations in percentages, units, and framing suggest that the same data circulates through different documents without a reliable reference point. This pattern was also visible in daily work. In one case, a slide deck was recreated and sent again without any memory of the previous version's content or outcome. In another, identical input sheets were found across three Teams folders, with no indication of which version was correct. One team member could not retrieve critical data simply because it was never centrally stored.

Together, the evidence shows that while useful data is produced across the organization, the absence of a coordinated storage approach undermines its long-term value. Teams waste time searching for the latest files, duplicate efforts, or rely on outdated figures. Although some isolated systems function well, they are not institutionally embedded. These findings support theoretical proposition 1. Without shared storage, each team keeps its own version of the truth, making it harder to work together and deliver consistent information.

4.8 Performance monitoring

Performance monitoring refers to the ability to track progress, measure data maturity, and evaluate the impact of governance practices. It is essential for identifying gaps, improving workflows, and ensuring that governance frameworks are not only implemented but also maintained over time. At TSN, performance monitoring mechanisms are still developing. This section assesses how monitoring is addressed in internal documentation, team experiences, and daily routines.

Across all sources, performance monitoring is acknowledged but not yet embedded in practice. Documents refer to future ambitions. One internal strategy slide proposed "establishing mechanisms for ongoing monitoring," while another suggested "training and support to employees to ensure adoption." However, none of these plans are backed by evidence that structured monitoring is already in place. Interview participants confirmed this gap. P2 said, "We think about improving, but not about what happens if we're called back on our data," indicating that accountability is not enforced. P4 added, "We are currently not looking at data maturity. That could be a next step." P6 noted, "We're doing this on heroic effort. There is no structure in place yet," and P8 observed, "What's missing are in-depth workshops where people really learn how to work with governance. Right now it's based on awareness." Together, these quotes show that performance monitoring is viewed as important but not yet operational. Observations supported this. In one meeting, follow-up actions on Green Book contributions were discussed informally, but no one was assigned to track progress.

In another case, the researcher had to manually create a data overview due to the absence of a dashboard or feedback tool.

Taken together, these findings show that performance monitoring is a shared ambition but lacks formal tools, assigned roles, or consistent practices. It currently depends on individual awareness and effort rather than structured evaluation. This confirms theoretical proposition 2. Performance monitoring requires formal systems, such as dashboards and routines. But also a culture in which teams see ongoing evaluation as part of their work. At TSN, both elements are currently missing: the structures are not yet built, and the habit of continuous evaluation is not yet embedded.

4.9 Findings

This section synthesizes the findings from the eight data governance criteria to identify shared patterns and structural limitations in how data is managed at TSN. While each criterion, ranging from lifecycle management to performance monitoring, was analyzed individually, the collective results provide a more comprehensive understanding of TSN's data governance practices. The goal is not only to highlight individual strengths or weaknesses, but to draw connections between recurring challenges such as unclear ownership, inconsistent version control, fragmented storage, and the absence of system-wide monitoring. These integrated insights form the foundation for the recommendations in the next chapter.

Across the eight governance criteria, several cross patterns emerge that help explain the broader understanding of data governance at TSN. One of the most frequent challenges is the lack of consistently defined or updated responsibility for specific data and documents, particularly in collaborative environments such as the Green Book. In many cases, ownership is assumed rather than formalized, which can limit accountability and make coordination more difficult. While structured governance practices are clearly in place in technical domains such as the data lake, where roles, classifications, and validation steps are embedded, these practices are not yet adopted across all teams or use cases. Instead, much of the collaboration around data and documentation depends on informal routines, such as sharing files through Teams folders, linking Excel sheets across silos, or aligning verbally during meetings, or across the office. These methods provide flexibility but reduce traceability and standardization. Many deliverables are developed in parallel without shared formats, source references, or clarity on which version is final. This can lead to inconsistencies, especially when documents are reused or updated across teams.

Despite these operational challenges, there is a strong organizational awareness of the importance of data governance. Strategic documents and interviews suggest that teams are engaged with the topic and motivated to improve. However, monitoring, measurement, and enforcement mechanisms remain limited. As a result, data governance often depends on the effort and experience of individual writers, rather than on organization-wide processes. This combination of ambition and awareness reflects a transitional phase in TSN's data governance. A phase where data governance is underway, but not yet integrated across the organization.

These cross-insights reveal that many governance challenges are not isolated to individual criteria but have their roots in structural issues such as unclear responsibility and informal interoperability. To further prioritize these findings, *Table 5* summarizes how each data governance criterion performed in terms of practical relevance and strength of evidence. The overview includes a synthesis of key findings per criterion. A short explanation of each summary statement follows the table.

Table 5: Overview of findings per data governance criterion, based on document analysis, interview insights, and field observations. The table synthesizes the findings of the criteria and provides a foundation for the recommendations in *Chapter 5*.

Criterion	**Main findings**
Data lifecycle management	Documents revealed that lifecycle stages, such as draft, final, or outdated, are often not clearly marked, particularly in the Green Book and related workfiles. Interviews confirmed this, with participants noting that status updates are handled inconsistently or not recorded at all. Observations supported this by highlighting moments where colleagues were unsure which file version was active or up-to-date.
Interoperability	The document analysis showed that overlapping content, mismatched structures, and inconsistent terminology across deliverables hinder alignment between teams. Interviewees strongly confirmed this, especially concerning the Green Book and IM, describing frequent confusion over inputs. Observations further supported this, noting duplicated efforts and inaccessible shared models across departments.
Data classification	Documents showed that classification practices, such as titling, labels, and use of status colours, are inconsistently applied. Interview data confirmed this unevenness, but also suggested that there are technical domains, like the data lake, that are more structured. As there were no specific observations linked to this criterion, conclusions rely mainly on documents and interview feedback.
Data security & compliance	Inconsistencies in sensitive data between the Green Book, ECP, and IM raised concerns about secure versioning and compliance. Interviews validated these risks, particularly around market-sensitive documents. However, no supporting observations were recorded, limiting insight into how these risks manifest in day-to-day work.
Data unit responsibility	Documents frequently lacked clear author attribution or assigned ownership. In some cases, multiple contributors were listed without a final reviewer. Interviews confirmed this fragmentation, while observations showed that responsibility is often informally clarified through verbal communication. Together, these sources highlight a structural gap in accountability.
Version control	The analysis showed that version numbers were missing, outdated, or inconsistently applied in many key files. Interviews strongly reinforced this, with participants describing how different teams work in parallel without clarity on the final version. Observations confirmed these issues in real-time, with teams unknowingly using different versions or overwriting one another's work.

Table 5 (continued)

Criterion	**Main findings**
Data storage	Some documents, like the Document Control Register, showed good storage practice. However, most content was stored across disconnected Teams folders, resulting in inconsistencies. Interviews confirmed that storage systems are fragmented and often localized. Observations supported this, revealing repeated data retrieval issues and siloed storage of critical data.
Performance monitoring	Documents referenced strategic goals like continuous improvement and change management, but lacked concrete tracking. Interviews confirmed that no maturity models or KPIs are in place, and that governance is driven more by awareness than evaluation. Observations revealed that follow-up actions are rarely tracked, and consistency must often be manually assessed.

Taken together, the findings from this chapter highlight that TSN is in an important phase of developing its data governance maturity. While certain systems and roles are in place, particularly within teams, many governance practices remain informal and inconsistent across different teams in the organization, or dependent on individual initiative. The variation in maturity across criteria points to a need for more integrated, organization-wide approaches. *Chapter 5* builds on these insights by proposing targeted, actionable recommendations that address both foundational challenges and strategic opportunities. The recommendations aim to strengthen TSN's data governance capabilities and align with the results section as well as with ongoing transformation goals.

5 Recommendations for data governance at TSN's Transformation Office

This chapter translates the research findings from the embedded case study in *Chapter 4* into practical recommendations to improve data governance at Tata Steel Nederland (TSN). It focuses on how the Transformation Office, and therefore TSN, can take concrete steps to address the challenges identified in *Chapter 3*. In doing so, the chapter also considers what organizational conditions are needed to make these changes possible, and at what management level the recommendations need to be fit in.

This chapter has two goals. First, this chapter provides specific recommendations for each of the eight data governance criteria. These recommendations are based on the results of interviews, document analysis, and observations from *Chapter 4*, and are tailored to TSN's current situation. Second, the chapter explores whether these recommendations can be applied beyond TSN. A focus group was conducted to reflect on the findings and to assess their broader applicability to other industries. By combining clear recommendations with an analysis of change readiness and expert feedback, this chapter helps answer the second and third sub-questions of the thesis and uses a method as outlined in *Chapter 2*. It aims to support both practical improvements at TSN and broader learning for other organizations facing similar challenges, supporting the academic aim of this thesis.

Section 5.1 presents an overview of the main barriers and enablers of change within TSN, based on the interviews. Next, *Section 5.2* gives recommendations per data governance criterion. *Section 5.3* highlights cross-cutting themes. Lastly, *Section 5.4* discusses the focus group, broader applicability, and validates the recommendations and the theoretical propositions.

5.1 Willingness to change within TSN's organization

Improving data governance at Tata Steel Nederland (TSN) is not just about setting new rules or introducing better tools; it also requires people in the organization to change the way they work. The findings in earlier chapters show that many governance problems come from unclear responsibilities, inconsistent working methods, and a lack of shared understanding. This section looks at what needs to happen inside the organization to make real change possible. It uses interview findings from both managers and operational staff to describe how ready TSN is for change and what still stands in the way.

The need for change from an operational perspective

People working in the Transformation Office see very clearly that change is needed. They also know that the conditions to make that change are far from ideal. One Transformation Officer (P3) explained the scale of the shift: "The steel industry is in a unique situation. For a hundred years, the same production process has been followed with small improvements. Now, a radical shift is needed." This shift is taking place in an environment that is often hectic and unstructured. For example, when talking about working on internal reports like the Green Book, the same participant said: "That went well for a while, but in the last few days, everyone worked on the document at the same time. Then it became chaos."

This breakdown under time pressure is not unique. Another Transformation Officer (P1) added: "The biggest bottleneck is that everything constantly changes." These quotes show how difficult it

is to introduce structure when the organization itself feels unstable. At the same time, people are frustrated by the lack of clear ways of working. As P2 said, "The biggest bottleneck? Behaviour." This highlights that structured ways of working are still developing within TSN, and that teams benefit from guidance and support on a managerial level to adopt and maintain new data governance practices.

Managerial perspective

Managing roles also recognize that the current approach is not working. A lead in the Transformation Office (P6) said, "I don't think we even meet level 1. We definitely don't meet level 2," referring to how far TSN still has to go in terms of governance maturity. The same manager also described a clear example of a broken process: "We all know it has to improve. We feel uncomfortable editing Excel files late at night, which go to the Ministry the next day. But we keep doing it."

This shows that awareness alone is not enough. Many managers understand what is wrong, but that doesn't always lead to action. In the data governance team, P4 explained why this happens: "If managers don't see that the current way is unsustainable, no time or money will be invested." Without strong leadership support, it is hard to make time for training or introduce better tools. P4 also noted that not all teams approach governance in the same way: "Data governance works well for manufacturing, but HR, finance, and commercial work differently." This kind of inconsistency makes it difficult to implement shared standards.

Another lead, P7, made a similar point: "Improvement is mainly in the discipline of delivering updates. That is more a management question than a technical one." In other words, technical advancements like version control or interoperability cannot succeed unless leaders make governance a visible and supported priority.

A shared perspective

Even though operational staff and managers often focus on different things, they agree on one key point: improving data governance requires a shift in culture, not just systems. P6 put it clearly: "You can implement a data lake, but if people don't understand why it matters, the problem remains." The same point came up in the operational interviews. According to P8, a data engineer, "People see data governance as extra work. You see, often, people don't yet see the benefit."

At the same time, P8 stressed that change won't work unless people at all levels are involved: "It helps if it's on the agenda at management level... but only from above is not enough. You also need support from below." This confirms the literature, where Vilminko-Heikkinen and Pekkola (2019) say that data ownership is often regarded as the task of any CIO, but as data spreads out over every business branch of an organization, it should be determined at the office level. P8 also explained what is missing today: "What we're missing are in-depth workshops where people learn how to work with governance, Data Lakes and Warehouses." There are some general awareness sessions already, but those are not enough. Without time and training to really learn new methods, people will keep doing things the way they always have. As P8 warned: "Without definitions, version control, provenance and validation, you can't do reliable analyses... you're fooling yourself."

Conclusion

TSN is in a moment of transition, where the need for improved data governance is widely recognized across the organization. Both operational staff and managers see the potential for better ways of working and understand that cultural alignment is just as important as technical solutions. While practices and responsibilities are still taking shape, there is clear momentum and a shared willingness to improve.

These conditions provide a promising starting point for change. With the right support, clear roles, and a stronger connection between strategic direction and day-to-day work, TSN can move toward more consistent and effective data governance. To make governance practices an operational priority, they must first be adopted at the managerial level, both in words and in action. The next section offers targeted recommendations to guide this process, translating the findings from *Chapter 4* into concrete steps for implementation.

5.2 Criterion-based recommendations

This section presents concrete recommendations to improve data governance at Tata Steel Nederland (TSN), organized around the eight criteria defined in *Section 3.3*. Each criterion is addressed in a separate subsection, following a fixed structure that moves from insight to action. First, a summary of findings highlights the most important issues identified through interviews, document analysis, and observations. This is followed by an assessment of how they impact decision-making, efficiency, or compliance. Based on these findings, one or more targeted improvements are proposed. Where relevant, the recommendations are connected to established governance frameworks, such as DAMA-DMBOK, FAIR principles, or CMMI for Data Management found in *Section 3.1.4*. This helps ground the recommendations in the literature for academic credibility and offers an implementation guide according to existing methods, which shows that these gaps are not unique. Finally, each subsection includes practical guidance on implementation. This covers who should take the lead, at the operational or managerial level, what conditions must be in place for change to succeed, and what forms of support or resistance may be expected based on earlier analysis.

5.2.1 Data lifecycle management

At TSN, data lifecycle management, the process of guiding data from creation to approval and archiving, is present in practice but not consistently formalized. Across interviews, documents, and observations, frequent issues were identified: unclear versioning, inconsistent updates across related files, and missing status labels. While some technical teams, such as those working with the data lake, apply stricter procedures, these practices are not adopted more broadly. The recommendations, therefore, include adopting data lifecycle methods from the data governance team.

This lack of structure affects day-to-day efficiency and traceability. Teams often spend time checking whether documents are current, and inconsistencies across files create confusion and risk. These issues become especially visible under deadline pressure, when informal coordination proves insufficient. The absence of a shared approach reduces confidence in collaborative work and weakens the reliability of key reports. To address this, TSN should adopt a simple but mandatory data lifecycle protocol, starting with the strategic reporting teams. The following steps are recommended:

- Use consistent version numbering and status labels (e.g. *draft*, *in review*, *final*, *frozen*).
- Track changes in a dedicated change log, briefly describing what was changed and why.

- Require both versioning and change tracking before any file is added to a shared repository or submitted for review.
- Centralize document storage to reduce duplication and improve accessibility.
- Learn data lifecycle practices from structural trainings by the data governance team to develop skills from the bottom up.

These recommendations are in line with the CMMI for Data Management framework, which emphasises traceable lifecycle stages and controlled versioning as core elements of data maturity.

Implementation should begin within the Transformation Office and be supported by management. A limited pilot, for example, within the Green Book reporting process, can demonstrate value and build experience before wider application. Simple tools like Excel-based logs or folder conventions can be used initially, with room to scale into digital systems as needed.

5.2.2 Interoperability

Interoperability refers to the ability of systems and teams to exchange data smoothly and consistently (Prasad, 2024, p. 122). At TSN, the lack of interoperability is a recurring challenge. Teams often maintain separate datasets, templates, and spreadsheets, which makes it difficult to align figures, track updates, or maintain a shared understanding of current data. Interviews and document analysis showed repeated inconsistencies between the Green Book, the Datapoint Sheet, and the Information Memorandum or other compliance reports.

This fragmentation results in duplicate work, misalignment, and uncertainty about which data is correct. During periods of intense collaboration, such as the preparation of investment reports, this slows down decision-making and increases the risk of errors. When interoperability is lacking, trust in shared outputs declines, and teams tend to revert to working in isolation.
To improve interoperability, TSN should focus on building shared structures for collaboration and alignment. The following steps are recommended:

- Develop and implement a single, shared reporting template for strategic documents (e.g. Green Book, IM, IF) with linked data fields.
- Introduce a central validation workflow to ensure consistency across key reporting documents.
- Organize cross-team review sessions at fixed project stages to catch inconsistencies early.

These recommendations are well aligned with the FAIR principles, which emphasize interoperability as a key condition for consistent and reusable data across organizations and systems.
For these improvements to succeed, they must be made a clear priority at the managerial level. Without strategic commitment to breaking silos and standardizing practices, operational efforts will not be sustained. At the same time, interoperability must be supported from the bottom up. The data governance team should lead training sessions to show how shared standards improve data quality and reduce duplication. This combined top-down and bottom-up approach is essential for embedding interoperable practices into TSN's everyday way of working.

5.2.3 Data classification

Data classification refers to the process of labeling and organizing data based on its content, sensitivity, and purpose. At TSN, classification practices are limited and inconsistently applied. While some data environments, such as the data governance team's data lake, require classification before data can be loaded, this level of structure is not applied in other areas, particularly in strategic reporting processes. Interviews indicated that while some teams are aware of the need for data classification, there is no shared approach or sufficient labeling system in place. As a result, the status, intended audience, or sensitivity of a document is often unclear, especially when files are reused across multiple reports.

The absence of classification standards weakens data traceability and increases the risk of miscommunication. When it is unclear what a figure means, who it is for, or whether it is final, teams spend time seeking clarification or risk using data incorrectly. This slows down collaboration and increases the likelihood of inconsistent reporting across critical documents such as the Green Book. To strengthen data classification, TSN should introduce simple, consistent rules that apply across teams and reporting types. The following steps are recommended:

- Introduce classification labels (e.g. *confidential*, *internal draft*, *final – public*) for use in all strategic reports.
- Require all input files and spreadsheets to include a classification tag before being uploaded to shared folders or submitted.
- Create a simple template for labeling data sheets and supporting documents.
- Offer targeted training sessions by the data governance team on how to apply classification consistently and why it matters for quality and compliance.

These steps align directly with the DAMA-DMBOK framework, which identifies classification as a foundational component of data quality, access control, and governance.

As with previous recommendations, success depends on prioritization at the managerial level. Teams will not change habits unless classification is expected and enforced by leadership. At the same time, training delivered by the data governance team is key to building operational understanding and showing how classification supports cleaner, safer, and more reusable data. The data governance team has already made it clear that it is open to expanding its training and is eager to improve operational skills bottom-up. When adopted in both strategy and practice, classification becomes a key enabler of reliable and transparent decision-making.

5.2.4 Data security & compliance

Data security & compliance concerns the protection of sensitive information and alignment with internal and external requirements. At TSN, these practices vary across teams. While technical environments like the data lake apply access controls, strategic reporting documents often lack similar safeguards. Files are sometimes shared without clear labeling, and it is not always clear who has access or editing rights.

This inconsistency introduces avoidable risks. Without defined access levels or compliance checks, sensitive drafts may be misused or distributed too widely. For strategic reports like the Green Book,

this can reduce credibility and raise regulatory concerns.
To reduce these risks, TSN should adopt basic security and compliance measures for reporting processes:

- Use visible labels to mark all documents by access level (e.g., internal, restricted, public).
- Store sensitive files in shared folders with managed permissions.
- Appoint responsible people for allowing access to sensitive files.

These measures are supported by the NIST Data Governance framework, which emphasizes visibility, access control, and practical compliance steps. Clear expectations must come from management to ensure that the secure handling of information becomes a shared responsibility.

5.2.5 Data unit responsibility

Data unit responsibility refers to assigning clear ownership over specific sets of data, so that someone is accountable for accuracy, completeness, and maintenance. At TSN, this responsibility is often assumed but rarely made explicit. Interviews revealed that contributors rely on their own judgment to decide which version of data to use, and that shared reports usually lack a named owner who validates or maintains the figures. While some technical environments, such as the data governance team's data lake, assign data owners, this is not standard in the broader reporting process.

The absence of assigned responsibility creates confusion about who is accountable for key numbers, especially when documents are updated across teams. Without ownership, errors go unnoticed, misalignment continues between documents, and collaboration becomes more difficult.
To address this, TSN should define data unit responsibility clearly in all strategic reporting streams. The following actions are recommended:

- Assign a responsible person for each key data element or section within shared reports.
- Have one central person for assigning those responsibilities.
- Include the name and contact information of this responsible person in report templates and data sheets.
- Set clear expectations that this role includes reviewing inputs, validating updates, and coordinating changes across teams, for example, through training.

This approach aligns with the DGI Framework, which emphasizes clearly defined roles and responsibilities as central to effective data governance.
Managerial support is essential to embed this practice. This is because responsibility is rooted in top-down management. Ownership will only be taken seriously if it is backed by expectations from leadership and built into planning processes. At the same time, operational teams need clear guidelines and expectations. Training from the data governance team can help staff understand what responsible data stewardship looks like in practice.

5.2.6 Version control

Version control ensures that changes to data or documents are tracked and that teams are working with the correct and most recent version. At TSN, version control is applied inconsistently. In interviews, participants noted that while they often know which version they are working with, this knowledge is not documented or supported by a shared system, making it difficult for others to verify or follow up. Document analysis showed missing version numbers and inconsistent updates across related files. This leads to confusion, duplicated work, and sometimes the use of outdated data in critical reports.

Without proper version control, collaboration becomes risky and inefficient. Especially in time-sensitive projects like the Green Book, unclear versioning results in last-minute misalignment and extra effort. It also undermines confidence in reporting, both internally and with external stakeholders.
To improve version control, TSN should adopt basic yet mandatory practices that can be applied consistently across teams. The following actions are recommended:

- Apply a standard versioning system to all strategic documents (e.g., v0.1, v1.0, v1.1);
- Require a short changelog to be maintained with each version, summarizing edits and their purpose.
- Enforce that no file is added to a shared folder or sent externally without proper versioning and a changelog.
- Provide templates with version fields and changelog sections.

These practices reflect the Git methodology, a version control model widely used in technical environments that offers structured tracking and reproducibility. Git's underlying principles are directly transferable to day-to-day work.

Management must set the expectation that versioning is a top priority for collaborative work environments. When leaders consistently use and require versioned files, the habit will take hold. From the operational side, training by the data governance team will help staff understand how version control reduces errors, saves time, and improves accountability in everyday work.

5.2.7 Data storage

Data storage refers to how and where data is saved, accessed, and maintained. At TSN, storage practices vary widely across teams. Documents are frequently kept in local folders, personal drives, or email threads, making them hard to trace or align across departments. While the data lake offers a centralized solution for technical and manufacturing data, it is not used for strategic reporting streams like the Green Book. Observations and interviews showed that during collaborative efforts, teams frequently asked who had the latest file or where key data was stored.

The lack of centralized storage leads to fragmented data, lost time, and an increased risk of version conflicts. It also weakens continuity. If someone leaves the team, their files may be difficult to recover, or it may lose its context. This fragmentation reduces confidence in the data and makes coordinated reporting more difficult.
To improve data storage, TSN should adopt shared and structured storage practices for all strategic reporting activities. The following actions are recommended:

- Use centralized shared drives or platforms for each reporting stream, with clear folder structures and access permissions. P1 was on his way to doing this.
- Standardize where key inputs, templates, and final files are stored to ensure they are always accessible to the right teams.
- Define simple naming and archiving rules to avoid duplication and make retrieval easier. This can be achieved through training.
- Provide on-boarding materials that explain where documents are stored and how to use the structure, also to be achieved by training.
- Include data storage practices in applied training sessions from the data governance team, using real case examples.

These practices are consistent with the ISO 8000 standard, which emphasizes accessible, high-quality storage and documentation practices to support data integrity and reusability.

Leadership must clearly define and support a consistent storage strategy. Without this top-level alignment, teams will continue to create workarounds based on preference or habit. At the same time, day-to-day application depends on operational awareness. Through training, the data governance team can help teams understand how structured storage improves reliability, reduces risk, and enables better cross-team collaboration.

5.2.8 Performance monitoring

Performance monitoring involves tracking how consistently data governance practices are applied and where improvement is needed. At TSN, formal monitoring processes have not yet been introduced, which is understandable given that many data governance practices are still being established.

While not a current priority, performance monitoring will become increasingly valuable as other improvements, such as version control, classification, and ownership, are adopted. Monitoring helps identify progress, highlight areas where teams need support, and demonstrate the impact of governance efforts.
To build toward this, TSN is advised to take the following steps:

- Define a small, practical set of data governance KPIs.
- Assign the data governance team to coordinate tracking and provide periodic updates.
- Use results in quarterly management reviews to support learning and recognize improvement.

These steps align with the CMMI for Data Management framework, which identifies performance monitoring as essential for reaching higher levels of data maturity.

By integrating light-touch monitoring early, TSN can gradually develop a culture of continuous improvement, ensuring that good practices remain consistent and scalable over time.

5.3 Cross-cutting themes

The recommendations in this chapter address the eight key criteria for improving data governance at TSN. While each criterion targets a specific issue, three shared conditions consistently emerged as essential for effective implementation: visible prioritization at the management level, structured learning through bottom-up training, and the integration of governance practices into a cohesive system. This section brings these elements together and outlines what TSN must do to ensure the recommendations are not only introduced but sustained.

Across all interviews, it became clear that structural improvement will only occur if governance is treated as a managerial priority. Teams respond to what leadership reinforces. When practices like version control, classification, or secure storage are inconsistently supported or modeled by management, they often lose momentum in daily work. "It will be seen as a hassle." In contrast, when governance is expected, planned for, and resourced from above, teams are more likely to adopt and maintain it. For TSN, this means governance should not remain a background task or a project-specific concern. It should be embedded into regular planning activities, such as periodic stand-ups, project checkpoints, and review moments, where data governance expectations and progress are discussed.

At the same time, operational staff need practical tools and guidance to put governance into action. The interviews revealed a consistent desire for more in-depth, applied learning. "If people do not understand it, it won't be seen as a problem." Training sessions should move beyond general awareness and instead focus on how governance is applied in real scenarios, such as aligning Green Book inputs, tagging versions, or classifying/sharing sensitive data. The data governance team is well-positioned to lead this effort. Their experience with the data lake provides a strong benchmark for how structured governance can work in practice. They can help teams gain the skills needed to apply governance practices effectively.

A final precondition for improving data governance at TSN is the recognition that governance practices must work together. Improvements in one criterion, such as centralized storage, reinforce others, like version control and interoperability. If practices are introduced in isolation or used inconsistently, they are less likely to produce a lasting impact. TSN should focus on aligning tools, templates, and procedures so that they support one another and fit into a broader, usable framework. The framework mentioned after each criterion's recommendations can function as a guideline. Even small steps toward consistency can significantly improve collaboration and reduce rework.

In conclusion, the success of the recommendations in this chapter depends on more than technical issues, as expected by the second theoretical proposition from *Section 3.5*. It relies on three mutually reinforcing efforts: prioritizing data governance from the top, skill-building from the operational level, and alignment across criteria. Together, these form the foundation for sustainable data governance at TSN.

5.4 Focus group analysis

To validate the practical relevance of the developed data governance recommendations and explore their potential applicability beyond Tata Steel Nederland (TSN), a structured focus group was conducted. This session, lasting 45 minutes, involved three professionals from TSN's Transformation Office, P2, P3, and P6. All three participants have experience in different industrial sectors: the offshore, automotive, and financial sectors, respectively. The focus group served two primary purposes:

first, to validate the perceived usefulness and feasibility of each recommendation within TSN's operational context, and second, to discuss the extent to which these recommendations could be applied in other industrial sectors, as all three participants have professional histories in different sectors.

The focus group methodology was chosen for its ability to capture not only individual perspectives but also the dynamic interplay of ideas among participants. As Kitzinger (1994) emphasizes, focus groups are particularly effective in revealing how knowledge and ideas are developed and shared within a cultural context, allowing the researcher to see how participants build shared meaning by reacting to each other's ideas. It helped uncover not only what they agreed or disagreed on, but also how their views on the recommendations developed during the discussion.

The insights gathered from this focus group are presented in four subsections. *Section 5.4.1* discusses the validation of the recommendations, *Section 5.4.2* examines their broader applicability across different industrial contexts. *Section 5.4.3* validates the identified cross-cutting themes, and lastly, *Section 5.4.4* validates the theoretical propositions from *Chapter 3*.

5.4.1 Validating the recommendations

This section presents the outcomes of the recommendation validation exercise. For each of the eight data governance criteria identified in this study, a single recommendation was discussed with participants. This was due to the limited time for the focus group. Their responses were assessed using three categories: yes (clearly validated), partial (conditionally supported), and no (not validated). *Table 6* provides an overview of each recommendation, its validation status, and illustrative quotes that support the classification.

The findings show that the majority of the recommendations were either validated or partially validated by the participants. Only one criterion, data unit responsibility, received a negative validation. *Table 6* is further explained after the table.

Table 6: Validation of recommendations per criterion based on the focus group discussion.

Criterion	Validation status	Recommendation	Quotes
Data lifecycle management	Partial	Use consistent status labels (such as "draft," "final"). Keep track of changes in a log and mandate these steps before files are shared or saved.	P2: I think those status labels are a minimal step, but a good start - provided you clearly define what you mean... P3: In theory it would work, it depends on the discipline of the teams working with it. It's helpful to manage the information well and label it clearly. P6: I do recognize it as a useful recommendation for TSN.

Table 6: (continued)

Criterion	Validation status	Recommendation	Quotes
Interoperability	Partial	Develop and use one shared template for strategic reports (such as green book, IM, IF), in which data fields are linked. Link fixed validation times to this and involve multiple teams in joint checks. So introduce a central validation workflow to ensure consistency across key reporting documents.	P6: What makes implementing this successful? It's not to create a template, because it's already there. It's just a consequence-free business. P3: That behavior has to be influenced in the right way, so that you are indeed going to use it.
Data classification	Yes	Introduce classification labels (such as "internal draft," "confidential," "public," "final" or "draft") and require that each file include a label before it is uploaded or shared.	P2: There's no standard language on classification. everyone just does something slightly different.
Data security & compliance	Partial	Visibly mark documents with their access level (such as "internal," "restricted," "public"). Store sensitive files in folders with managed access rights, and designate one person responsible for access permission per folder or file.	P6: So what you want to solve is that people are very much aware of the classification of what they have in their hands and what they act upon afterwards. So applying classification is one thing, but you have to start acting on it.
Data unit responsibility	No	Assign a responsible person for each key data element or section within shared reports.	P2: What I miss here... is a central data point... that is deliberately managed to set this kind of thing up. P3: I think you don't solve that with a designated department, but rather that it should just be part of the task assignment that people get.
Version control	Yes	Use a standard system for version numbers (such as v0.1, v1.0 or date) and require a short changelog summarizing modifications. No file should be in a shared folder or shared externally without a version and a changelog.	P6: For me everything before the point is a draft, and the definitive versions are the whole numbers... P2: No, no one really feels the pain of it.

Table 6: (continued)

Criterion	Validation status	Recommendation	Quotes
Data storage	Yes	Use centralized shared drives with clear folder structure and access rights. Document this in onboarding materials.	P6: There are simply no agreements. If you decide to create your own folder structure tomorrow, no one will stop you.
Performance monitoring	Yes	Establish a small and achievable number of data governance KPIs. Have the data governance team track these and share the results quarterly in the management meeting to encourage learning and improvement.	P6: Yes, I think this is a good one. I actually like this. P2: This one is quite useful, because we are already working with KPIs here.

Data lifecycle management was partially validated. Participants agreed with the principle of using status labels and change logs, but pointed out that success depends on clear definitions and team discipline. P2 remarked, "I think those status labels are a minimal step, but a good start, provided you clearly define what you mean by 'draft' or 'final', for example." P3 noted that the system would work in theory, but only with consistent team behavior. P6 confirmed it was a useful recommendation for TSN. Interoperability also received a partial validation. While the idea of shared templates and validation moments was supported, participants emphasized that behavior, not tooling, is the main obstacle. P6 said, "What makes implementing this successful? It's not to create a template, because it's already there. It's because it's just a consequence-free business." P3 added that behavioral incentives must be in place to ensure use. Data classification was validated. Participants recognized the absence of a shared classification standard and the resulting inconsistencies. P2 stated, "There's no standard language on classification. Everyone just does something slightly different," indicating that a consistent approach is both needed and achievable. Data security & compliance was partially validated. The importance of classifying access levels was acknowledged, but participants pointed out that merely applying labels is not sufficient. As P6 explained, "Applying classification is one thing, but you have to start acting on it." This highlights the gap between technical practices and behavioral follow-through. Data unit responsibility was not discussed during the focus group and was therefore marked as not validated. This was due to an oversight during the session and does not reflect a rejection of the recommendation itself. Version control received full support. Participants described mature practices from other industries and contrasted them with the current absence of structure at TSN. P6 said, "For me, everything before the point is a draft, and the definitive versions are the whole numbers," showing how versioning can also signal document status. P2 added, "No, no one really feels the pain of it," underscoring the need for stricter enforcement. Data storage was also fully validated. The current absence of structure was seen as a barrier to effective governance. P6 remarked, "There are simply no agreements. If you decide to create your folder structure tomorrow, no one will stop you." This quote points to the urgency of centralized file management and onboarding guidance. Performance monitoring was validated as well. Participants saw governance-related KPIs as a helpful mechanism. P6 said, "Yes, I think this is a good one. I actually like this," and P2 noted, "This one is quite useful, because we are already working with KPIs here."

In summary, the validation results show that the recommendations match the actual challenges

and needs at TSN. Most of them were either fully supported or seen as useful with some conditions. The discussion made clear that the success of each recommendation depends not only on the tools or systems used but also on team behavior and the work culture. These findings confirm that improving data governance requires more than technical fixes; it also needs support from management and active involvement from the people who work with data every day.

5.4.2 Broader applicability

This section presents the findings from the focus group discussion on whether the data governance recommendations could also be useful in organizations outside of Tata Steel Nederland. The goal was to add to this thesis's academic rigour. For each criterion, participants were asked if the recommendation would be relevant in other industries. P2 had experience in the offshore industry, P3 in the automotive industry, and P6 in the financial industry. Based on their responses, each criterion was labeled as yes (broadly applicable), partial (applicable in some cases or with adjustments), or no (not discussed or not confirmed). *Table 7* summarizes these outcomes, along with quotes that support each assessment.

Six of the eight recommendations were confirmed as broadly or partially applicable to other sectors. Only two, data unit responsibility and performance monitoring, were not validated as transferable, either due to a lack of discussion or doubts about their general relevance. *Table 7* is further explained after the table.

Table 7: Broader applicability of recommendations per criterion based on the focus group discussion.

Criterion	Broader applicability	Quotes
Data lifecycle management	Yes	P2: I think this is true of any company that is producing anything of data-like stuff. P6: What you mention here is generic. But what's interesting: so why doesn't it happen here?
Interoperability	Partial	P2: In other industries, you often see that the ERP system or the data system is leading, so these things are automatically enforced. Researcher: Does TSN specifically lag behind? P6: Yes, this is specifically for TSN.
Data classification	Yes	P6: I think a lot of companies, even the small start-ups, are already doing this.
Data security & compliance	Partial	P6: That's generic. In the last 10 years, IT companies have gotten all big on it.

Table 7: (continued)

Criterion	Broader applicability	Quotes
Data unit responsibility	No	There is no literal quote in the focus group confirming cross-sector applicability of the Data Unit Responsibility recommendation.
Version control	Yes	P2: We at Boskalis had a client who demanded that we include all the different revision steps and review moments completely embedded. P3: At automotive, it was just expected that you paid attention to that.
Data storage	Yes	P6: This is generic; it applies to any company.
Performance monitoring	No	P2: It is definitely applicable in many companies. Researcher: Isn't this TSN specific, as there is a need for a data governance team?

Data lifecycle management was marked as broadly applicable. P2 noted, "I think this is true of any company that is producing anything of data-like stuff," and P6 added, "What you mention here is generic." These comments show that versioning, status labels, and change tracking are seen as relevant in most data-driven environments. Interoperability received a partial score. While participants acknowledged that other industries often enforce structure through tools like Enterprise Resource Planning (ERP) systems, they also pointed out that TSN specifically lacks this type of enforcement. P2 stated, "In other industries, you often see that the ERP system or the data system is already leading," but P6 confirmed the problem is more prominent at TSN. Data classification was confirmed as broadly applicable. P6 said, "I think a lot of companies, even the small start-ups, are already doing this," showing that labeling and classifying data is a common practice across sectors. Data security and compliance were rated as partially applicable. While participants recognized its importance, they framed it as a more established practice in IT environments. P6 noted, "That's generic. In the last 10 years, IT companies have gotten all big on it," suggesting that its relevance may depend on sector maturity. Data unit responsibility was not discussed during the focus group and was therefore marked as not validated. This was due to an oversight during the session and does not reflect a rejection of the recommendation itself. Version control was seen as clearly transferable. P2 shared an example from Boskalis where clients required detailed revision tracking, and P3 described how similar practices were standard in automotive. These examples show that structured version control is a common expectation in many technical and industrial contexts. Data storage was also validated. P6 called it a "generic" challenge, and the researcher's observation, that fragmented file structures often get copied across projects, was recognized as a problem in other companies, too. Performance monitoring was labeled as not broadly applicable. Although participants saw the recommendation as useful within TSN, its relevance depends on having a dedicated data governance team to coordinate and track KPIs. P2 noted that similar metrics are already in use, but it remained unclear whether such an approach would work in organizations without comparable roles or structures.

In short, most of the recommendations were seen as useful beyond TSN, especially in data-heavy or technical organizations. The focus group shows that many of the same challenges, such as unclear file storage, version control, or inconsistent classification, also occur in other industries. This suggests that the recommendations developed in this study could be helpful in other settings as well. However, they should always be adjusted to fit the specific structure and needs of each organization.

5.4.3 Validating cross-cutting themes

The three cross-cutting themes introduced in the recommendations chapter, top-down prioritization, bottom-up skill development, and interdependence between criteria, were explicitly tested in the focus group and received strong confirmation from participants. Rather than emerging for the first time during the group discussion, these themes served as a framework for reflection and validation of the governance criteria and the recommendations given in this thesis.

Top-down prioritization was universally acknowledged as a requirement for sustainable change. As summarized by the researcher during the session, "TSN operates without consequences unless priorities are set from the top." Participants agreed that data governance cannot become part of daily routines unless it is endorsed, resourced, and expected by senior leadership. As one participant noted, "That comes back to the need for top-down prioritization. Only then do training and skills matter." Or "That would only work if it comes from management." Bottom-up training formed the second theme. While some participants were skeptical of traditional training approaches, the broader point that data governance should be embedded in day-to-day responsibilities received broad support. One participant emphasized, "It should just be part of your job," suggesting that skill development becomes effective when it is aligned with what people are already accountable for. The third theme, interdependence between criteria, was highlighted through both internal logic and practical comparison. The researcher argued that version control is more effective with centralized storage and that classification depends on clear responsibility. This was supported by participant experiences at Boskalis and in the automotive sector, where versioning, validation, and traceability were seen as systemically enforced. As one participant remarked, "These tools are just instruments. The real question is, why isn't it happening?"

Together, these themes highlight that successful data governance at TSN requires more than isolated process improvements. The recommendations in this thesis are only likely to be effective if they are supported from the top, reinforced through bottom-up learning, and implemented in a coordinated and coherent manner. The focus group thus served not only to test individual recommendations but also to validate the overarching approach proposed in the recommendations chapter. These validated insights form the basis for the concluding reflections in the next section.

5.4.4 Validating the theoretical propositions

This section evaluates whether the theoretical propositions formulated in *Chapter 3* were confirmed based on the focus group. While the case study already provided empirical support, the focus group allowed participants to reflect on their practical validity.

The focus group confirmed both theoretical propositions, but not without further notable insights. First theoretical proposition 1, that *data governance must address operational fragmentation to enable reliable reporting*, was validated. Participants repeatedly stressed the lack of alignment between teams and systems at TSN, highlighting that "one team creates a folder structure, and then later the project moves and it gets a new structure elsewhere", and that "there are no rules. If you start

your own folder structure tomorrow, no one will stop you". These examples reflect how fragmented data storage and inconsistent practices limit reliable reporting. The same holds for interoperability: "everyone still uses their own." As such, the focus group confirmed that addressing fragmentation, especially in data lifecycle management, interoperability, and data storage, is essential to improve reliability. What stood out though, is that this is not just a technical problem. People know that fragmentation makes reporting unreliable, but it keeps happening because there are no clear rules or strong incentives to work in the new, desired way. In other industries, central systems such as ERP tools enable alignment, but at TSN this is missing. This finding confirms the proposition: reliable reporting depends on fixing fragmentation. At the same time, it also shows why change is so hard, awareness is there, but without structure and discipline, daily practices do not improve.

Secondly, the group also affirmed theoretical proposition 2, that the *successful implementation of data governance depends on both structural measures and cultural readiness.* Participants repeatedly emphasized the importance of culture within teams: "It really depends on the discipline of the team using it." As P6 explained, TSN currently functions as "a consequence-free company." Cultural readiness was seen as critical. This shows that structural measures like guidelines and templates are not enough on their own. They only make a difference if teams take them seriously, apply them consistently, and feel responsible for doing so. The validation of this proposition highlights a recurring theme, why change is difficult: people see the need for stronger governance and often agree with the principles, but without cultural readiness and a sense of accountability, the impact remains limited. The next chapter highlights the thesis's conclusions and discussion, marking the end of this report.

6 Conclusion & discussion

This chapter brings together the main findings of the study to answer the research questions and reflect on their relevance for both academic research and real-world practice. Firstly, *Sections 6.1 & 6.2* address the research questions by summarizing the main empirical insights. Then *Section 6.3* revisits the theoretical propositions introduced in the literature research. Next, *Sections 6.4 & 6.5* outline the scientific and societal contributions of the study. A reflection on the study's relevance to the CoSEM program is provided in *Section 6.6*. The chapter concludes with a discussion of the study's limitations and suggestions for future research in *Sections 6.7 & 6.8*.

6.1 Answering the sub-questions

Before answering the main research question, three sub-questions were addressed to guide the study. This section summarizes the answers to each of them.

> **Sub-question 1:** *"Which governance criteria support sustainability-related compliance reporting by addressing key challenges in steel manufacturing organizations?"*

This study identified eight data governance criteria as particularly relevant: data lifecycle management, interoperability, data classification, data security & compliance, data unit responsibility, version control, data storage, and performance monitoring. These were derived from existing literature on data governance and sustainability reporting in industrial settings. Each criterion addresses common themes in data governance in steel manufacturing found in the literature. Data lifecycle management, version control, and data storage are essential for maintaining consistency and ensuring traceability. Interoperability and classification help align data across teams and systems. Data unit responsibility and performance monitoring support organizational readiness and continuous improvement. Security & compliance is essential for managing risks and meeting regulations from external bodies, as explained in the introduction. These criteria formed the basis for answering the next sub-questions.

> **Sub-question 2:** *"How can steel manufacturing organizations apply data governance criteria to improve sustainability-related compliance reporting?"*

The case study shows that both structural and cultural factors influence how well data governance criteria are applied. On the structural side, many teams work without clear ownership of data, consistent version control, or shared storage systems. This makes it difficult to manage data in a reliable and standardized way. Often, files are shared informally, and changes are made without clear documentation. This leads to confusion about which data is correct or final, especially when different teams work together.

Cultural and organizational factors also play a role. While many employees are aware that data governance is important, they are not always supported by clear processes, training, or follow-up. As a result, improvements often depend on heroic effort rather than shared routines. Some teams, like those managing technical systems, like the data governance team, already use more structured data governance practices. But in strategic areas, such as reporting for sustainability goals, practices are less developed.

The focus group discussions showed that several governance criteria were not fully validated in practice, which underlines the difficulty of applying them effectively. For data lifecycle management, participants agreed that tools like status labels and change logs are useful, but only if terms such as "draft" or "final" are clearly defined and consistently applied. In the case of interoperability, templates and validation procedures already exist, but participants emphasized that these tools are often ignored unless managers create the right behavioral incentives. For data classification, participants recognized the value of labeling access levels, yet pointed out that classification only improves governance when teams also act on those labels in daily practice.

Taken together, these findings show that applying data governance criteria is not just a matter of having the right tools or frameworks. Success depends on how people in the organization use them. Without clear management priorities, consistent team discipline, and practical support at the operational level, data governance efforts quickly become fragmented. Some teams might use good practices, but unless these are applied across management, sustainability reporting remains unreliable. This means that improving compliance reporting requires both strong top-down prioritization and active bottom-up engagement, with criteria implemented as part of everyday work.

Sub-question 3: *"To what extent can the recommendations of this study be applied to other steel manufacturing organizations and organizations beyond the steel industry?"*

The findings of this study were discussed in a focus group with participants from the Transformation Office team at TSN to assess the results' relevance and potential application beyond the steel manufacturing industry. During the session, participants confirmed that data lifecycle management, data classification, version control, and data storage were not only relevant within TSN but also applicable to other data-intensive organizations facing similar reporting and compliance demands. The recommendations for data unit responsibility were not covered during the focus group and therefore remain unvalidated.

At the same time, the focus group made clear that not all recommendations travel equally well. Interoperability was only partly validated. Participants explained that in other sectors, systems like Enterprise Resource Planning (ERP) already enforce structure and alignment, which reduces the challenges seen at TSN. This means the recommendation is especially relevant for organizations without strong system enforcement, but less useful in industries where validation workflows are already embedded.

Data security & compliance was also seen as only partly transferable. Participants agreed that classifying data access levels is important everywhere, but they stressed that maturity differs across sectors. Effectiveness depends on factors such as organizational size, culture, and digital infrastructure. In IT-central industries, access controls and classification are already standard practice, while in more traditional sectors, like the steel manufacturing industry, these measures may still lag. This suggests that while the criterion applies broadly, its impact depends on sector maturity.

Performance monitoring, finally, was judged not transferable. The recommendation relied on TSN's data governance team to coordinate KPIs, a structure that many organizations simply do not have. While the idea of monitoring governance performance made sense to the participants, its practical relevance depends heavily on organization-specific setups.

In summary, the findings of this study are relevant beyond the steel industry, particularly for large organizations with fragmented reporting practices. However, some depend on specific organizational structures, levels of digital maturity, or cultural readiness. As such, the governance criteria and recommendations developed in this study can serve as a useful starting point for improving data governance practices to support sustainability-related compliance reporting.

6.2 Answering the main research question

> *"How can steel manufacturing organizations leverage data governance to support the transition to green steel production?"*

Steel manufacturing organizations can support their transition to green steel production by strengthening data governance. This study shows that data governance is the bridge between sustainability-related compliance reporting and decarbonization of steel production. Without consistent reporting, steel companies cannot prove their progress to governments, regulators, or investors. This means subsidies are delayed, investments are harder to secure, and credibility is lost. With strong data governance in place, reporting becomes trustworthy, which unlocks funding and policy support. That funding makes large-scale decarbonization projects, such as hydrogen-based steel production with new blast furnaces, both financially and policy-wise, possible.

The findings from the literature research revealed that the most relevant governance criteria for supporting such reporting include data lifecycle management, interoperability, data classification, data security & compliance, data unit responsibility, version control, data storage, and performance monitoring. These criteria help ensure that sustainability data is traceable, comparable, and aligned across departments. Together, they form the foundation for more coordinated and credible reporting practices.

However, the case study showed that the criteria that were only partially validated are the ones that reveal why change is difficult. Employees see the need for better data governance and often express willingness to change, but these intentions do not translate into practice. Data lifecycle management tools like status labels and change logs are supported in principle, but they only work if definitions are clear and teams consistently apply them. Interoperability goes beyond creating shared templates; it requires managerial incentives for teams to use them. And data classification only has meaning if teams act on the labels they assign. This illustrated a broader pattern: organizations want to move forward, but progress requires top-down prioritization from management together with active bottom-up engagement from employees, with every data governance criterion consistently applied as part of daily work.

Lastly, the challenges and solutions identified in this thesis are not unique to steel manufacturing. Other large organizations with complex compliance responsibilities face similar issues. The focus group showed that most criteria, data lifecycle management, data classification, version control, and data storage are broadly applicable. However, some were only partially validated or not at all. Interoperability was only partly validated. In other industries, mature systems enforce structure, but at TSN, this is missing, which makes success depend more on behavior than on tools. This means the recommendation is relevant elsewhere, but its impact depends on an organization's digital maturity. Data security & compliance was also seen as partly transferable. The idea of managing access and protecting sensitive data is universal, but in IT-heavy sectors, it is already well established, while

in more traditional industries like steel manufacturing, it is still developing. Performance monitoring, however, was not considered broadly applicable. At TSN, it works because a dedicated data governance team tracks progress, but many organizations lack such a structure. Without it, the recommendation loses much of its relevance. This showed that while the recommendations offer a useful starting point beyond TSN, they must be adapted to the governance structures, maturity, and culture of each organization.

In conclusion, this study showed that data governance is not just a technical matter but an organizational one that directly links sustainability-related compliance reporting to decarbonization of steel manufacturing. The validated criteria demonstrate that progress is possible, but the partially or unvalidated ones highlight the barriers that keep change from happening. This makes clear that strengthening data governance requires more than tools: it depends on top-down prioritization from management, active bottom-up engagement from operational teams, and even integration of all eight data governance criteria. When organizations align these elements, they can leverage data governance to accelerate their transition to green steel production.

6.3 Revisiting the theoretical propositions

This subsection evaluates the two theoretical propositions that were formulated in *Chapter 3* to guide the thesis. By comparing these expectations with the findings from Tata Steel Nederland (TSN), this section assesses whether the propositions were supported by the evidence. In doing so, it connects the literature to the case study findings and recommendations developed in this thesis.

Proposition 1: Data governance must address operational fragmentation to enable reliable reporting.

This proposition is supported by the case study findings. Across multiple criteria, including interoperability, version control, and data responsibility, employees at TSN from different teams and management layers reported issues caused by fragmented data practices. The Green Book made this especially clear, as those issues weakened its reliability as a strategic document.

The literature already warned that steel manufacturers need strong sustainability reporting and coordination between departments (Gajdzik & Wolniak, 2021; Singh et al., 2007). The TSN case confirms that. But one finding was especially noticeable: employees at every level of different teams recognized these problems and said they wanted change, yet little actually happened. The focus group echoed this: "there are no rules. If you start your own folder structure tomorrow, no one will stop you."

In short, Proposition 1 was confirmed: fragmentation between teams and systems clearly disables reliable reporting, just as was expected from the literature. What was less expected, however, was how widely this problem was recognized across TSN without leading to change. Employees at every level acknowledged the issue, but with no rules in place, fragmentation continued.

Proposition 2: The successful implementation of data governance depends on both structural measures and cultural readiness.

This proposition is supported by the case study findings. At TSN, many employees acknowledged the importance of better data governance practices. However, the study revealed that the successful

application of these practices often relied on heroic effort rather than shared routines or institutionalized support. In areas where governance was more mature, such as the technical domains managed by the data governance team, ownership, structure, and classification were clearly embedded. In contrast, in strategic reporting environments like the Transformation Office, data governance practices were inconsistently applied and highly dependent on the effort and expertise of individual contributors.

These findings align with the literature. Abraham et al. (2019) highlighted that governance frameworks must be culturally supported and embedded in day-to-day operations, not just formally defined. Otto (2011) similarly stressed that without ownership and behavioral change, data governance initiatives risk being ineffective. The TSN case confirms this: Technical tools alone will not lead to consistent practice. Managerial prioritization, training, and reinforcement mechanisms are just as important for enabling change.

Together, this research confirmed that the transition to stronger data governance requires both structural measures and cultural adoption, as expected. What stood out, however, was how strongly this proposition highlighted the same theme as proposition 1: at TSN there is a clear desire for change, but it does not happen on its own. This insight proved critical for answering the main research question. The next section turns to the scientific contributions of this thesis.

6.4 Scientific contribution

This section explains how the findings of this thesis contribute to academic knowledge on leveraging data governance in sustainability-related compliance reporting in complex industrial settings. As outlined in the introduction, the study aims to clarify how governance practices support internal coordination under real-world reporting pressure. *Section 6.4.1* highlights the conceptual contribution, while *Section 6.4.2* highlights the methodological contribution.

6.4.1 Conceptual contribution

A first contribution lies in validating the core challenges of data governance identified in the literature. These challenges are extensively documented in recent research on digitalization in the steel sector and the need for integrated data management approaches across departments and systems by Branca et al. (2020) and Gajdzik and Wolniak (2021). The case study confirms that these issues remain persistent even in organizations that have invested significantly in data infrastructure, like the steel manufacturing industry has done for decades (Mishra et al., 2023).

Second, the thesis contributes by operationalizing these challenges through eight concrete data governance criteria. These criteria act as a structured lens to evaluate data governance maturity in a sustainability-related context. For example, interoperability and version control are highlighted as essential yet underdeveloped mechanisms for aligning sustainability-related reporting across departments, as discussed by Janssen et al. (2020) and Q. Zhang et al. (2022). By including the criteria data unit responsibility and performance monitoring, this study expands on existing literature by highlighting the importance of accountability and continuous improvement, two aspects underrepresented in data governance frameworks, as stressed by Vilminko-Heikkinen and Pekkola (2019).

Third, this study adds to the conceptual understanding of what makes data governance effective in decarbonization contexts. While most prior work on the digital transition in the steel industry

focuses on technological or economic factors (Di Foggia & Beccarello, 2024), this thesis shows that data governance structures, such as assigning responsibility and offering training, are just as critical for enabling credible and consistent sustainability reporting. This insight builds on existing work that highlights the importance of organizational elements in data governance implementation (Otto, 2011). In particular, the findings confirm that formal rules alone are not enough: willingness to change and practical support are key to embedding data governance practices into daily work routines.

Finally, the case study contributes to knowledge by applying these conceptual insights to the steel industry. While existing frameworks often remain general, this research tailors its recommendations to a sector characterized by legacy systems, technical interdependencies, and compliance pressure. By demonstrating how a data governance lens can help navigate these complexities, the thesis offers an analytical bridge between high-level academic frameworks and practical reform in emissions-intensive industries. Next, this thesis's methodological contribution is highlighted.

6.4.2 Methodological contribution

This thesis contributes methodologically by developing and applying a criteria-based evaluation of data governance practices in a complex industrial setting. While previous studies often remain high-level or focus on technical systems, this research breaks down data governance into eight concrete criteria derived from the literature and systematically applies them to a real-world case. This structure enabled a targeted diagnosis of strengths and weaknesses in sustainability-related compliance reporting practices, as well as actionable recommendations.

The triangulation of interviews, document analysis, and observations strengthens the validity of the findings and allowed for cross-checking between stated practices and actual behavior. Rather than evaluating data maturity at a general level, the study found how each criterion plays out across departments and reporting layers. This approach offers a replicable structure for other researchers or organizations facing similar data governance challenges. Next, this study's societal relevance is discussed.

6.5 Societal relevance

This study is based on a case study conducted at Tata Steel Nederland (TSN), which granted full access to internal documents, interviewees, and observational data. While the insights are grounded in TSN's specific organizational context, the findings are not unique to this case, as validated during a focus group. They reflect broader issues faced by large industrial companies across Europe as they prepare for stricter sustainability-related compliance reporting obligations. European legislation like the Corporate Sustainability Reporting Directive (CSRD) and the Carbon Border Adjustment Mechanism (CBAM) applies to all large companies with more than 1000 employees within EU jurisdiction, requiring them to report transparently and credibly on their emissions and sustainability performance. These regulations increase the need for robust data governance, not just for compliance, but also to secure State support, attract green investments, and maintain a license to operate in a carbon-constrained economy. By showing how data governance can support trustworthy and coordinated sustainability reporting, this thesis contributes to a more realistic understanding of what organizational readiness for the green transition looks like in practice.

These findings have clear implications for both public and private stakeholders. Public institu-

tions should move beyond reactive compliance checks and actively support data governance through training, clear requirements, and incentives. Reliable sustainability reporting requires stronger internal data governance, not just stricter rules. For companies, data governance must be seen as a strategic asset, not an administrative task. It enables credibility, trust, and long-term sustainable transformation. Without it, strong green ambitions risk delay or rejection. Middle managers and operational leads play a key role. As shown in the case study, without clear responsibilities and routines, reporting remains fragmented.

The findings from this thesis extend beyond the steel industry. The focus group confirmed that issues such as data lifecycle management, data classification, version control, and data storage are not unique to Tata Steel Nederland (TSN) but also affect other emissions-intensive sectors like the automotive, offshore, and financial industries. As public pressure intensifies and as companies across Europe prepare for CSRD and CBAM obligations, solid data governance is no longer optional; it is a requirement for credibility and sustainable growth. The triangulated method used in this thesis offers a practical tool for identifying weak spots in internal data governance and designing targeted improvements. In this way, the study responds not only to academic gaps but to urgent societal demands for transparency, accountability, and the acceleration of sustainability transitions.

As industries work toward greener production, this thesis shows that improving data governance is not just an internal task; it supports a broader societal goal. Clear roles, consistent data, and better interoperability help organizations meet growing demands under public policies. These improvements enable faster, more credible sustainability-related compliance reporting, which in turn contributes to society's larger ambition: accelerating decarbonization.

6.6 Relevance to CoSEM programme

This thesis reflects the interdisciplinary and systems-oriented approach central to the CoSEM programme. The study explored a complex, multi-actor problem at the intersection of data management, sustainability regulation, and organizational change. It combined document analysis, stakeholder interviews, and observations to examine how internal data governance can enable credible sustainability-related compliance reporting under pressure from EU and national policies.

Through this methodology, the thesis integrated technical and institutional perspectives, linking data practices to broader challenges in transition management and regulatory alignment. By identifying concrete barriers within a large industrial organization, operationalizing them through data governance criteria, and proposing actionable improvements at both strategic and operational levels, the thesis applied core CoSEM skills, including stakeholder analysis, systems thinking, and implementation-focused problem solving. This approach exemplified how systems engineers can contribute to governance and decision-making in real-world sustainability transitions.

6.7 The study's limitations

This study has several limitations. Organizational context shaped the feasibility of the research, as mentioned in the case description in *Section 2.3*. TSN is currently undergoing major restructuring, including the announcement of 1,600 layoffs following public subsidy pauses and ongoing financial losses. This situation understandably lowered engagement: interviews were difficult to schedule, participation was uneven, and some priorities shifted during the research. As a result, data availability and participation had their limitations. One interview had to be excluded due to a lack of

consent, slightly reducing the diversity of perspectives. Nevertheless, this underscores the thesis's importance in avoiding these public subsidy pauses.

Research positioning and strategy also influenced the scope and outcomes of this study. The research was conducted from within the Transformation Office, which enabled close access to internal governance efforts but limited the focus to this specific unit. Broader organizational practices, like in production and finance, were not included. Additionally, literature research was intentionally selective. To ensure relevance in a rapidly evolving field, only scientific articles published after 2010 were considered. Conference papers and reviews were excluded, except for two academic books. Furthermore, articles were included only if the search terms appeared in the title, abstract, or keywords, which helped maintain focus but may have excluded relevant but less explicitly labeled studies.

Lastly, generalizability is limited. As the findings showed, especially for performance monitoring. While the triangulated method and findings offer transferable insights, the context of TSN, like its organizational structure, regulatory pressures, and internal dynamics, may differ significantly from other companies. This should be considered during future research, discussed in the next section.

6.8 Directions for future research

This thesis has shown that improving data governance can strengthen the credibility and coordination of sustainability-related compliance reporting in large industrial settings to accelerate the transformation to lower-emission steel production. While the findings are based on a single case study at Tata Steel Nederland, the identified challenges may be present in other organizations as well. To build on these results and test their broader relevance, further research is needed in several directions.

First, future research should examine the behavioral and organizational factors that affect how data governance practices are adopted. This thesis found that the willingness to change and the engagement of both managers and operational roles play an important role. Studies should further investigate how leadership signals, role clarity, and resistance to change influence adoption. Long-term studies can track how these factors develop after training or the introduction of new routines.

Second, there is a need for sector-specific tools to evaluate the maturity and impact of data governance in other industries. While this thesis used a qualitative approach grounded in internal documents and actor interviews, future research could develop maturity models or structured assessments tailored to industrial contexts. These tools would help organizations measure how well governance criteria are implemented and support continuous improvement through benchmarking and feedback.

Third, to explore broader applicability, future studies should investigate whether data governance challenges in partially or unvalidated criteria recommendations appear in other emissions-intensive or regulated sectors. Industries such as automotive and off-shore also face increasing demands from carbon emission regulations. Comparative case studies could better test which solutions are transferable and identify sector-specific needs.

Fourth, as EU sustainability policies evolve and tighten, organizations must adjust their internal processes. Future research should examine how data governance structures respond to changing regulatory requirements and how organizations align their practices with shifting expectations. Collaboration between researchers and policymakers at a national or European level could help ensure

that organizations' data governance capabilities keep up with external demands.

Together, these directions point to the need for research that is both practical and context-specific. By studying how data governance adoption works in practice, developing evaluation tools, and conducting studies across sectors, future research can help organizations strengthen their data governance practices to support the transformation to a sustainable industrial future.

References

Abraham, R., Schneider, J., & vom Brocke, J. (2019). Data governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management*, *49*, 424–438. https://doi.org/10.1016/j.ijinfomgt.2019.07.008

Al-Badi, A., Tarhini, A., & Khan, A. I. (2018). Exploring big data governance frameworks. *Procedia Computer Science*, *141*, 271–277. https://doi.org/10.1016/j.procs.2018.10.181

Alexopoulos, K., Weber, M., Trautner, T., Manns, M., Nikolakis, N., Weigold, M., & Engel, B. (2023). An industrial data-spaces framework for resilient manufacturing value chains. *Procedia CIRP*, *116*, 299–304. https://doi.org/10.1016/j.procir.2023.02.051

Al-Ruithe, M., Benkhelifa, E., & Hameed, K. (2018). A systematic literature review of data governance and cloud data governance. *Personal and Ubiquitous Computing*, *23*(5), 839–859. https://doi.org/10.1007/s00779-017-1104-3

Barker, R. T., & Rich, S. R. (1992). The use of the focus group method in evaluating communication training. *Communication Education*, *41*(3), 242–247. https://doi.org/10.1080/03634529209378894

Bernardo, B. M. V., Mamede, H. S., Barroso, J. M. P., & Santos, V. M. P. D. (2024). Data governance quality management—innovation and breakthroughs across different fields. *Journal of Innovation Knowledge*, *9*(4), 100598. https://doi.org/10.1016/j.jik.2024.100598

Branca, T. A., Fornai, B., Colla, V., Murri, M. M., Streppa, E., & Schröder, A. J. (2020). The challenge of digitalization in the steel sector. *Metals*, *10*(2), 288. https://doi.org/10.3390/met10020288

Castro, V., Georgiou, M., Jackson, T., Hodgkinson, I. R., Jackson, L., & Lockwood, S. (2024). Digital data demand and renewable energy limits: Forecasting the impacts on global electricity supply and sustainability. *Energy Policy*, *195*, 114404. https://doi.org/10.1016/j.enpol.2024.114404

Chen, X., Despeisse, M., & Johansson, B. (2020). Environmental sustainability of digitalization in manufacturing: A review. *Sustainability*, *12*(24), 10298. https://doi.org/10.3390/su122410298

Chinnathai, M. K., & Alkan, B. (2023). A digital life-cycle management framework for sustainable smart manufacturing in energy-intensive industries. *Journal of Cleaner Production*, *419*, 138259. https://doi.org/10.1016/j.jclepro.2023.138259

Corallo, A., Crespino, A. M., Del Vecchio, V., Gervasi, M., Lazoi, M., & Marra, M. (2023). Evaluating maturity level of big data management and analytics in industrial companies. *Technological Forecasting and Social Change*, *196*, 122826. https://doi.org/10.1016/j.techfore.2023.122826

Costa, L., Almeida, A., & Reis, L. (2024). Methodology for implementing a manufacturing execution system in the machinery and equipment industry. *Procedia Computer Science*, *232*, 2028–2037. https://doi.org/10.1016/j.procs.2024.02.025

DAMA International. (2017). *Dama-dmbok: Data management body of knowledge*. Technics Publications, LLC.

De Carolis, A., Sassanelli, C., Acerbi, F., Macchi, M., Terzi, S., & Taisch, M. (2025). The digital readiness assessment maturity (dreamy) framework to guide manufacturing companies towards a digitalisation roadmap. *International Journal of Production Research*. https://doi.org/10.1080/00207543.2025.2455476

Delft University of Technology. (n.d.). Human research ethics [Retrieved April 29, 2025]. https://www.tudelft.nl/over-tu-delft/strategie/integriteitsbeleid/human-research-ethics

Di Foggia, G., & Beccarello, M. (2024). Decarbonization in the european steel industry: Strategies, risks, and commitments. *Environmental Challenges*, *16*, 100988. https://doi.org/10.1016/j.envc.2024.100988

Edge Delta. (2024). How many companies use cloud computing in 2024? [Retrieved February 27, 2025]. https://edgedelta.com/company/blog/how-many-companies-use-cloud-computing-in-2024

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review*, *14*(4), 532–550. https://doi.org/10.5465/amr.1989.4308385

European Commission. (2019). European green deal [Retrieved April 29, 2025]. https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal_en

European Commission. (2022). Corporate sustainability reporting [Retrieved February 27, 2025]. https://finance.ec.europa.eu/capital-markets-union-and-financial-markets/company-reporting-and-auditing/company-reporting/corporate-sustainability-reporting_en

European Commission. (2023a). Carbon border adjustment mechanism (cbam) [Retrieved April 29, 2025]. https://taxation-customs.ec.europa.eu/carbon-border-adjustment-mechanism_en

European Commission. (2023b). Eu emissions trading system (eu ets) [Retrieved April 29, 2025]. https://climate.ec.europa.eu/eu-action/eu-emissions-trading-system-eu-ets_en

Forum, W. E. (2023). Steel industry net-zero tracker [Retrieved June 16, 2024, from https://www.weforum.org/publications/net-zero-industry-tracker-2023].

Gajdzik, B., & Wolniak, R. (2021). Transitioning of steel producers to the steelworks 4.0—literature review with case studies. *Energies*, *14*(14), 4109. https://doi.org/10.3390/en14144109

Gajdzik, B., Wolniak, R., Grebski, W., Szymszal, J., & Grebski, M. E. (2024). *Sustainability and smart manufacturing: The transformation of the steelwork industry* (1st). Routledge. https://doi.org/10.4324/9781032697611

Gierend, K., Freiesleben, S., Kadioglu, D., Siegel, F., Ganslandt, T., & Waltemath, D. (2023). The status of data management practices across german medical data integration centers: Mixed methods study. *Journal of Medical Internet Research*, *25*. https://doi.org/10.2196/48809

Gillan, A. (2021). Governance: The key driver for data-driven innovation. *Computer Fraud Security*, *2021*(4), 10–13. https://doi.org/10.1016/S1361-3723(21)00041-5

Hajoary, P. K. (2023). Industry 4.0 maturity and readiness: A case of a steel manufacturing organization. *Procedia Computer Science*, *217*, 614–619. https://doi.org/10.1016/j.procs.2022.12.257

Janssen, M., Brous, P., Estevez, E., Barbosa, L. S., & Janowski, T. (2020). Data governance: Organizing data for trustworthy artificial intelligence. *Government Information Quarterly*, *37*(3), 101493. https://doi.org/10.1016/j.giq.2020.101493

Johnston, M. (2016). Are you prepared for your next data disaster. *NetworkWorld Asia*, *13*(2), 52–52.

Kavak, M., & Rusu, L. (2025). Challenges and opportunities of artificial intelligence in digital transformation: A systematic literature review. *Procedia Computer Science*, *256*, 369–377. https://doi.org/10.1016/j.procs.2025.02.132

Kitzinger, J. (1994). The methodology of focus groups: The importance of interaction between research participants. *Sociology of Health & Illness*, *16*(1), 103–121. https://doi.org/10.1111/1467-9566.ep11347023

Ladley, J. (2012). Definitions and concepts. In J. Ladley (Ed.), *Data governance* (pp. 7–20). Morgan Kaufmann. https://doi.org/10.1016/B978-0-12-415829-0.00002-2

Lebaea, R., Roshe, Y., Ntontela, S., & Thango-Mabizela, B. (2024). The role of data governance in ensuring system success and long-term it performance: A systematic review [Preprint]. https://doi.org/10.20944/preprints202410.1841.v1

Levy, Y., & Ellis, T. J. (2006). A systems approach to conduct an effective literature review in support of information systems research. *Informing Science: The International Journal of an Emerging Transdiscipline*, *9*, 181–212. https://doi.org/10.28945/479

Liu, K. (2023). Research on capability maturity evaluation model of power grid data management. *Procedia Computer Science*, *228*, 1030–1037. https://doi.org/10.1016/j.procs.2023.11.135

Luisi, J. V. (2014). Part iv - information architecture. In J. V. Luisi (Ed.), *Pragmatic enterprise architecture* (pp. 189–261). Morgan Kaufmann. https://doi.org/10.1016/B978-0-12-800205-6.00004-4

McCaffrey, P. (2020). Analysis best practices. In P. McCaffrey (Ed.), *An introduction to healthcare informatics* (pp. 279–289). Academic Press. https://doi.org/10.1016/B978-0-12-814915-7.00019-3

Mendelow, A. L. (1991). Environmental scanning: The impact of the stakeholder concept. *Proceedings of the Second International Conference on Information Systems*, 407–418.

Mishra, R., Singh, R. K., & Gunasekaran, A. (2023). Adoption of industry 4.0 technologies for decarbonisation in the steel industry: Self-assessment framework with case illustration. *Annals of Operations Research*. https://doi.org/10.1007/s10479-023-05440-0

Muenzen, K. D., Amendola, L. M., Kauffman, T. L., Mittendorf, K. F., Bensen, J. T., Chen, F., Green, R., Powell, B. C., Kvale, M., Angelo, F., Farnan, L., Fullerton, S. M., Robinson, J. O., Li, T., Murali, P., Lawlor, J. M., Ou, J., Hindorff, L. A., Jarvik, G. P., & Crosslin, D. R. (2022). Lessons learned and recommendations for data coordination in collaborative research: The cser consortium experience. *Human Genetics and Genomics Advances*, *3*(3), 100120. https://doi.org/10.1016/j.xhgg.2022.100120

Muslemani, H., Liang, X., Kaesehage, K., Ascui, F., & Wilson, J. (2021). Opportunities and challenges for decarbonizing steel production by creating markets for 'green steel' products. *Journal of Cleaner Production*, *315*, 128127. https://doi.org/10.1016/j.jclepro.2021.128127

Nahshon, Y., Morand, L., Büschelberger, M., Helm, D., Kumaraswamy, K., Zierep, P., Weber, M., & de Andrés, P. (2024). Semantic orchestration and exploitation of material data: A dataspace solution demonstrated on steel and copper applications. https://arxiv.org/abs/2406.19509

National Institute of Standards and Technology (NIST). (2025). Data [Retrieved February 27, 2025]. https://csrc.nist.gov/glossary/term/data

Otto, B. (2011). Data governance. *Business & Information Systems Engineering*, *3*(4), 241–244. https://doi.org/10.1007/s12599-011-0162-8

Pettigrew, A. (1988). Longitudinal field research on change: Theory and practice [Paper presented]. *National Science Foundation Conference on Longitudinal Research Methods in Organizations*.

Petzold, B., Roggendorf, M., Rowshankish, K., & Sporleder, C. (2020). Designing data governance that delivers value [Accessed: 2024-06-16]. https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/designing-data-governance-that-delivers-value

Prasad, A. N. (2024). *Introduction to data governance for machine learning systems: Fundamental principles, critical practices, and future trends*. Springer. https://doi.org/10.1007/979-8-8688-1023-7

Prashar, A. (2017). Energy efficiency maturity (eem) assessment framework for energy-intensive smes: Proposal and evaluation. *Journal of Cleaner Production*, *166*, 1187–1201. https://doi.org/10.1016/j.jclepro.2017.08.116

Rivett, P. (2015). Are you creating a data swamp. *KM World*, *24*(3), S5–S5.

Shirwa, A. M., Hassan, A. M., Hassan, A. Q., & Kilinc, M. (2025). A cooperative governance framework for sustainable digital transformation in construction: The role of digital enablement

and digital strategy. *Results in Engineering*, *25*, 104139. https://doi.org/10.1016/j.rineng.2025.104139

Sifter, C. J. (2017). Establishing a data governance center of excellence within your bank. *New Jersey Banker*, 24–25.

Singh, R. K., Murty, H. R., Gupta, S. K., & Dikshit, A. K. (2007). Development of composite sustainability performance index for steel industry. *Ecological Indicators*, *7*(3), 565–588. https://doi.org/10.1016/j.ecolind.2006.06.004

Su, J., Yao, S., & Liu, H. (2022). Data governance facilitates digital transformation of oil and gas industry. *Frontiers in Earth Science*, *10*, Article 861091. https://doi.org/10.3389/feart.2022.861091

Suprun, E., Roumi, S., Mohamed, S., & Stewart, R. A. (2024). Digital transformation maturity assessment framework for large infrastructure asset owners. *Digital Engineering*, *1*, 100003. https://doi.org/10.1016/j.dte.2024.100003

Swanborn, P. (2010). *Case study research: What, why and how?* SAGE Publications.

Tata Steel Nederland. (2022). Tata steel netherlands accelerates plans to significantly reduce dust and odour emissions from ijmuiden steelworks [Retrieved from https://products.tatasteelnederland.com/corporate/news/tata-steel-netherlands-accelerates-plans-to-significantly-reduce-dust-and-odour-emissions-from-ijmuiden-steelworks].

Tata Steel Nederland. (2023a). Green steel plan. https://www.tatasteelnederland.com/en/sustainability/green-steel-plan

Tata Steel Nederland. (2023b). Sustainability report summary 2022-2023. https://www.tatasteelnederland.com/sites/default/files/sustainability-report-2022-2023-summary.pdf

Thomas, M. A., Cipolla, J., Lambert, B., & Carter, L. (2019). Data management maturity assessment of public sector agencies. *Government Information Quarterly*, *36*(4), 101401. https://doi.org/10.1016/j.giq.2019.101401

United Nations Environment Programme Finance Initiative. (2023). *Climate risks in the industrials sector*. UNEP FI. https://www.unepfi.org/wordpress/wp-content/uploads/2023/04/Climate-Risks-in-the-Industrials-Sector.pdf

Vilminko-Heikkinen, R., & Pekkola, S. (2019). Changes in roles, responsibilities and ownership in organizing master data management. *International Journal of Information Management*, *47*, 76–87. https://doi.org/10.1016/j.ijinfomgt.2018.12.017

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, *26*(2), xiii–xxiii.

Wilkinson, M., Dumontier, M., Aalbersberg, I., et al. (2016). The fair guiding principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018. https://doi.org/10.1038/sdata.2016.18

Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th). SAGE Publications.

Yin, R. K., & Davis, D. (2007). Adding new dimensions to case study evaluations: The case of evaluating comprehensive reforms. In G. Julnes & D. J. Rog (Eds.), *Social policy research and evaluation* (pp. 75–93). Jossey-Bass.

Zhang, C., Fang, J., Ge, S., & Sun, G. (2024). Research on the impact of enterprise digital transformation on carbon emissions in the manufacturing industry. *International Review of Economics & Finance*, *92*, 211–227. https://doi.org/10.1016/j.iref.2024.02.009

Zhang, Q., Sun, X., & Zhang, M. (2022). Data matters: A strategic action framework for data governance. *Information & Management*, *59*(4), 103642. https://doi.org/10.1016/j.im.2022.103642

Zorrilla, M., & Yebenes, J. (2022). A reference framework for the implementation of data governance systems for industry 4.0. *Computer Standards & Interfaces*, *81*, 103595. https://doi.org/10.1016/j.csi.2021.103595

7 Annex

7.1 Annex I: AI statement

This appendix documents how AI tools were used during the development of this thesis. It follows a format for AI transparency by recording reflections on how the tools ChatGPT, Turboscribe AI, DeepL, and Grammarly contributed to different sections of the thesis. After the AI statement, three reflective questions look back on the use of AI during this study, following the TU Delft's format for transparent AI statements. For all AI tool statements goes: I read every piece of text produced by AI tools to make sure I stand behind everything in this thesis.

7.1.1 ChatGPT

Throughout the thesis, ChatGPT was generally used to help make this thesis's writing more concise and to suggest clear formulations when it was difficult to put ideas into words. It was also used for other specific tasks, listed in this paragraph per chapter. In chapter 2, ChatGPT was used to help turn an outline for the data analysis plan into a step-by-step plan that matched subsections §2.6.1–2.6.4. For the interview protocol, it was used to clean up question phrasing. ChatGPT was also used to draft short bridges from one section to the next. An example prompt is: "This part needs to be rewritten in clearer concise language and needs a bridge to the validity and reliability section." For chapter 3, ChatGPT was used to synthesize cross-cutting themes across the literature and to extract relevant text about each data governance criterion from the literature. It was also used to outline the knowledge gap, after which it was rewritten to cover exactly what the writer meant. An example prompt was: "Identify the specific knowledge gaps that this literature research addresses." For theoretical propositions, it helped phrasing new propositions more clearly after the first draft was advised against by the thesis supervisor. In chapter 4, ChatGPT was used to process coded notes and transcript excerpts and to list findings per criterion. This was a summarizing step; it helped to see which quotes aligned with which criterion and where patterns recurred. Then, data was selected manually to keep only literal, traceable quotes, and final findings were put in own wording. In chapter 6, ChatGPT was used to translate findings into cross-cutting themes. It was also helpful for structuring the focus group so the session would run smoothly. An example prompt here was: "Draft a focus group outline of a 45-minute focus group, following Kitzinger's definition." Lastly, ChatGPT was used in chapter 6 to help list the main findings and to propose clean, academic phrasing for answers to the sub-questions and the main question. It also suggested limitations to this study and future research directions, in addition to the writer's notes on both parts made throughout the thesis project.

7.1.2 TurboScribe AI

TurboScribe AI was used to generate automated transcripts of all recorded interviews and supervisory meetings. These transcripts were used as working documents for coding, thematic analysis, and extracting relevant quotes. I manually reviewed each transcript while listening to the audio to correct errors, clarify inaudible sections, and ensure that the final text accurately reflected the spoken content.

7.1.3 DeepL

DeepL was used for translation from Dutch to English. This included translating Dutch brainstorm notes in the methodology chapter, literal interview quotes in the results chapter, and entire interview

transcripts in the annexes, which I transcribed using Turboscribe AI. In each case, I manually checked the translations for accuracy, corrected any technical or contextual errors, and adjusted the wording to match the academic tone and style of the thesis.

7.1.4 Grammarly

Grammarly was used to check grammar, spelling, and sentence structure in all chapters. I accepted suggestions that improved clarity, readability, and flow, but only if they did not change the meaning or alter the tone. In some cases, I overrode Grammarly's suggestions to preserve my original phrasing where it was more precise or context-appropriate.

Reflection Questions

1. What was the most helpful AI suggestion and why?
ChatGPT's help in summarizing and restructuring the interview responses per criterion was most helpful. It allowed me to present complex participant input in a clear, academic format while maintaining traceability to the source.

2. Did you notice any bias or error from the AI? What did you do about it?
Yes. ChatGPT sometimes over-interpreted interview quotes or introduced summaries that weren't grounded in the transcript. I countered this by manually checking all paraphrases for accuracy and removing any untraceable interpretations.

3. Next time you use AI in an assignment, what will you do differently and why?
I would more clearly separate tasks between AI tools (e.g., only use ChatGPT for structure, DeepL for translation, Grammarly for polish) and document AI use in real time to make reflection easier later.

7.2 Annex II: Interviews

7.2.1 Annex IIA: Participant roles

The table below provides an overview of the roles of the participants interviewed during this study. This context helps interpret their perspectives on data governance practices at Tata Steel Nederland. To preserve anonymity while maintaining analytical clarity, participants are referred to by code (P1–P8) throughout the thesis. The table indicates whether each participant holds a managerial or operational role and identifies their primary area of expertise or involvement in the organization.

Interview participants and Their roles

No.	Coded Name	Job Title / Role
1	P1	Transformation Officer
2	P2	Transformation Officer
3	P3	Transformation Officer
4	P4	DG (Data Governance) team Lead
5	P5	Former CIO of the Dutch National Government
6	P6	Transformation Lead

Table 8: Interview Participants and Their Roles (continued)

No.	Coded Name	Job Title / Role
7	P7	Long-term Assets lead
8	P8	Elite Data Engineer

7.2.2 Annex IIB: Interview transcripts

This annex contains the transcripts from the interviews conducted as part of this research. These interviews formed a core source of qualitative data, supporting the analysis of current data governance practices at Tata Steel Nederland. All but one participant gave informed consent, and identifying details have been removed to protect confidentiality. The content is organized by participant code (P1–P8) and structured around the key data governance criteria used in the study. The interviews were conducted in Dutch and translated into English.

Interview with P1 – Transformation Officer

P1: And what have you learned at TU Delft? I'm not sure if you had lectures on this?

Interviewer: In the way I'm now going to work with data governance. That's the term I'll use in my research. It includes actively managing policies around how to deal with data. I'm not that familiar with it yet myself. It hasn't been covered in depth in my program, since my focus was more on energy and industry. My supervisor, Anneke, is from I&C, which is more focused on information and data. So I think the TU research will definitely be useful. It's also a new area for you, so it's good to share that with P6.

P1: I think he assumed you had a degree in data management. I already thought, well, that's not necessarily the case.

Interviewer: No, I do have experience with data analysis and working with large datasets. That's not new to me. But the perspective I'm taking now, treating it as a real specialization, is different.

P1: Okay, that's good. That was also my starting point two months ago.

Interviewer: Right, and it's interesting for me to run alongside that. My first goal is to have interviews scheduled by March 8. It could be a bit later, but if you don't immediately know people I could speak with, I'd like to plan that in time. That way, I keep momentum and make sure my literature review continues well. I'm mainly looking for people from various levels within Tata. I don't need to interview external people.

P1: No, no. Let me think—what would be smart? If I know someone, I can ask next week who might put together a short list... I would definitely schedule a half-hour with P3 on Monday morning. He does similar work to me within the Transformation Office. He could guide you a bit over the next two weeks. I already discussed this with him. Do you know how to schedule that in Teams or Outlook?

Interviewer: Yes, I can handle that.

P1: Great. And then—what would be wise... I'm thinking. What exactly would you like to ask?

Interviewer: It's mainly about the questions I'll also ask you in a minute. It concerns how data is currently structured at Tata Steel Netherlands. I want different perspectives on how people think the situation currently is. And whether, for example, a framework is already being used—both within your team and others. It's interesting to see whether there's a uniform approach or if everyone uses their own system.

P1: I think that would definitely be useful. We could start within Apollo... well, that's not the name anymore.

Interviewer: Why not?

P1: No idea. There's a new CPO who wants to put their own stamp on it. It's now called the Green Steel Business Transition.

Interviewer: And that might change again?

P1: Yes, it's now called the Green Steel Business Transition. You can definitely ask around within the different streams. That's what I've been doing these past few months. Everyone sees the need, but you get all kinds of different directions. I think X would be a good contact for you. You could call her. I also spoke recently with a document controller from PTC. PTC is Tata's engineering branch, and they work very differently than this department. I don't immediately know other internal departments.

Interviewer: Otherwise I'll just ask P3. I'm starting early, so it'll be fine. But how is data structured in your team? Is there a framework being used?

P1: Not really. We have a Teams environment called Apollo Transformation. Each branch has its own folder. But not everyone uses this. For example, Nina made a start, but I don't think she's continued. She just created some folders for her own reference. Every department has folders somewhere for internal use, but there's no real structure. What I'm trying to do now is create a sort of document repository. For instance, we have the Green Book, which is a collection of key documents. But those documents are scattered everywhere, and no one has a clear overview. I want to use the Green Book as a fixed reference point, so we know: this is the truth as it stands now.

Interviewer: And by "fixed," you mean it won't be changed anymore?

P1: No, it can still be updated, but it must be clear that there is a version of record. The Green Book is a major milestone, and from there we can continue to develop. What I want is that every supporting document has a simple cover sheet stating: created by, checked by, approved by, plus a changelog.

Interviewer: That's definitely smart. But there's no structured approach yet?

P1: No, not really.

Interviewer: And what do you think is the biggest challenge in making this work?

P1: The biggest bottleneck is that everything is constantly changing. As soon as you define something, it's already outdated. That doesn't apply to all documents, but to many. Everything in Tata's business case depends on other long-term assets, and everything is still shifting.

Interviewer: P6 mentioned a scale from 1 to 5 to measure data maturity. Is that something you've looked into?

P1: No, I haven't really explored that yet.

Interviewer: But it might be helpful to make it measurable. I'll ask P6 again. But when it comes to data governance, what role does it play in Tata's transition to Green Steel?

P1: The key function is traceability: being able to see which data led to which output. For example, this morning Koen asked which version of the emissions data was the latest. Nobody knew. I knew because I had been involved, but that's not a sustainable system.

Interviewer: Where do you store that?

P1: I just know I have the latest version. But it's not structurally arranged. What we need is a system where data is collected, validated, and given version numbers.

Interviewer: Has there ever been an attempt to improve data governance?

P1: Yes, P3 tried to set up something called the CSM Data House, but it didn't work out too well. Some departments create a data book every quarter with updates, but that's purely internal, and there's no overarching policy.

Interviewer: And do you think it's important to keep using Teams?

P1: Yes, for its ease of use. Eventually, I'd like to have a search function, like an internal chatbot that can search all documents.

Interviewer: Like an ERP system?

P1: More like a search assistant such as ChatGPT, but internal. So staff can easily find the right data.

Interviewer: Would that help encourage people to upload their documents?

P1: It might. If they know all documents in the system are validated and up to date, they might feel more urgency to use it actively.

Interviewer: Are there reports I can read to get more insight?

P1: No, there's no formal policy or central documentation. I do have some PowerPoints with my own thoughts—I'll send those to you.

Interviewer: Perfect, thank you! That will definitely help me.

Interview with P2 – Transformation Officer

Interviewer: It's very relaxed actually. Anyway, shall I just start with the questions?

P2: Yes, go ahead. My laptop is still starting up.

Interviewer: Alright. First, two fairly general questions. What is your role in the Green Book?

P2: P6 and X define the process. I'm the person who communicates that process to the Leadership Team (LT). So they design it, and I bring it to LT.

Interviewer: LT?

P2: Leadership Team – the middle block, including parts of Heracles and Finance. It covers Commercial, Technology Green Products, Health Emissions, Strategy, and Future Operations.

Interviewer: What exactly is the Green Book's function within Tata?

P2: The Green Book is for justifying the Business Transformation. You also have Asset Development – that includes Heracles, DRP, EAF...

Interviewer: I'm not familiar with those abbreviations yet.

P2: DRP is Direct Reduction Plant, EAF is Electric Arc Furnace, and Gas Holder is gas storage. That's all part of the Blue Book, which is the investment book we send to Tata Steel Limited (TSL).

Interviewer: TSL means Tata Steel Limited?

P2: Exactly. The Green Book focuses on Business Transformation towards TSL.

Interviewer: "Green" sounds like sustainability. Is that correct?

P2: Tata Steel needs to move from grey to green. You can:

- Start a small green business within Tata Steel;
- Or fully transform the entire company.

We've chosen full transformation to green steel.

Interviewer: Who writes the Green Book and for whom?

P2: It's written by many people, but it's for TSL. In short: Tata Steel Netherlands wants funding, and TSL wants a strong plan.

Interviewer: What decisions are based on it?

P2: The Green and Blue Books lead to a one-pager. Based on that:

- Does Tata Steel Netherlands receive funding?
- Is the transformation plan strong enough?

Interviewer: How is the Green Book assembled?

P2: You have themes like Health Emissions, New Energy, Tech Green Products. These are divided into teams. The transformation adds more layers. We had a triangle model with gates – each gate has specific deliverables and teams. The Green Book sits between the gates and connects teams, risks, and documents as evidence of the transformation's status.

Interviewer: Where do the supporting documents come from?

P2: Responsible leads publish them for the Transformation Office, often via Teams. But that's not ideal.

Interviewer: What would be better?

P2: A central repository like P1's document repository. That way, you can always trace back which data was used.

Interviewer: Is Tata Steel aware of the importance of data management?

P2: Not really. We focus more on being nice to each other, but we don't think about what happens if we get challenged on our data. If your data isn't properly stored, it's your problem.

Interviewer: What is the biggest bottleneck?

P2: Behavior. Tata Steel used to be dominant. Now, TSL and the government have more control. We're not used to being accountable for data.

Interviewer: What kind of data supports the Green Book?

P2: Tons of it: calculations, emissions, plans, manuals. But there's no consistent naming or storage structure.

Interviewer: That's exactly a data governance problem.

P2: Exactly. We do store things, but we don't have a document control register to track versions or changes.

Interviewer: So teams work with different versions?

P2: Yes, because there's no central storage.

Interviewer: How do you handle fluctuating parameters and versioning?

P2: By not putting too much pressure on people. During the EC pre-notification, there was a rush, and some data was processed too quickly, leading to errors.

Interviewer: What are the risks of sending inaccurate data to governments or investors?

P2: Worst case: the investment falls through because trust is lost.

Interviewer: Do you know industrial companies who manage this better?

P2: Boskalis and FrieslandCampina. They had clear document control, financial structure, and defined scope.

Interviewer: Their factories store their own data?

P2: Yes. Factories manage this better than sustainability teams because they're closer to the data source.

Interviewer: What is the biggest organizational challenge?

P2: Tata Steel is a chaotic business inside a structured project. The government adds more chaos than structure.

Interviewer: Thanks! This really helped me understand the Green Book and why data governance is so important.

P2: No problem! I hope it didn't sound too negative, but this needs to be said honestly.

Note: This interview was conducted on February 24, 2025. It has been translated from Dutch and lightly edited for clarity.

Interview with P3 – Transformation Officer

Interviewer: I'd like to dive into data governance soon, but first, can you explain your role? I've met so many teams and people in my first week.

P3: Sure. Have you already spoken with P6?

Interviewer: Yes, he's actually my official supervisor.

P3: He's the lead of the Transformation Office. I'm a Transformation Officer, basically the same role as P1 or X, but I focus specifically on planning. I have a bit more experience than the others, and I regularly work with P6, while coordinating a lot with P1 and others to align on direction and priorities.

The Transformation Office reports to Jeroen Klunter, have you met him?

Interviewer: Not yet.

P3: His scope changed recently. He used to oversee the entire green transition, including Heracles and CapEx assets like factory construction. Now, someone else handles that, and Jeroen focuses more on the business side.

Interviewer: That's also why the Apollo name disappeared, right?

P3: Exactly.

Interviewer: I understand these kinds of shifts happen regularly.

P3: They do. Scope changes often, especially because of questions from shareholders or government stakeholders. That shifts team focus often too.

Jeroen's department now falls under New Energy – Long Term Asset Strategy. That's about long-term development and how to structure it. Then there's Technology Transition, focused on product feasibility, product outlook, and the necessary testing.

René Kief leads that, definitely someone worth speaking to. There's a lot of data involved in his scope.

Then we have the User Organization for EAF and DRP. They make sure that once we get the keys to the new assets, everything is operationally ready.

There's also the PMO, Project Management Office. We were originally just the Transformation Office, but Tata is reorganizing. There's now also a Green Steel Transition Office at Tata Group level, which is why we renamed ourselves. We support information exchange between workstreams and make sure deliverables are submitted on time.

Interviewer: Is the Green Book the most important deliverable right now?

P3: Right now, yes. But there are others too. For example, an information document for banks, an application for the EU Innovation Fund, and proper communication with the government, led mostly by Karl.

We were also asked to provide insight into all contracts that must be finalized by 2030, when and how volumes should be secured.

These are new processes we're setting up. Sometimes it's unclear whether we should do it or another team.

Interviewer: With all these information flows and documents, I imagine data governance is essential. How do you see the current state?

P3: It's very mixed. Some topics are well developed, others not at all. We haven't yet found a structured way to implement step-by-step validated data governance.

We work iteratively. We take an approved version, build on that, and write a new version for revalidation. But the supporting documents and models are still scattered across teams.

Interviewer: How was this handled for the Green Book?

P3: Before Christmas we submitted a pre-notification to the EU. We're finalizing it now, due by the end of February, although that may shift due to government processes.

We're reusing that same data for a bank information document. But tone of voice differs, a document for government is different from one for banks. And there's no central data format yet.

Interviewer: What would a good solution look like?

P3: A standard fact sheet per topic. Using AI tools like ChatGPT, we could generate different versions of the same data quickly, one for a bank, one for shareholders.

We're already reusing content in places, but the process isn't well structured.

Interviewer: Do you see bottlenecks in data sharing across teams?

P3: Yes. For the EU pre-notification, we all worked in Teams and tracked which chapters were complete. That worked fine until the final days when everyone jumped in at once, it became chaos.

When we have time, it works well. But under time pressure, people overwrite each other's work and multiple versions emerge.

Interviewer: That's a vicious cycle, better version control saves time and reduces errors.

P3: Exactly. But right now, someone gets asked for a new document, creates a list, and starts gathering data, we start from scratch every time.

We should start with the last approved version and build from there. That's a governance issue, but solving it would save lots of time and errors.

Interviewer: If I want to analyze and improve data governance at Tata, where should I begin?

P3: Start with the Green Book and the pre-notification. Study how that process works and how it could be optimized.

Steel has had the same production process for 100 years, with small changes. Now we need radical transformation. That brings a lot of uncertainty and interdependencies.

Interviewer: So not just data, but assumptions and scenarios too?

P3: Yes. Sometimes data is unknown, we have to make estimates.

Interviewer: Do you know anyone else I should interview?

P3: Koen van der Loo. He's been heavily involved in the governance team and was key in collecting data for the EU pre-notification. He can explain how it went and the lessons learned.

Interviewer: Great, thanks!

Interview with P4 – Data Governance lead

Interviewer: I was told to reach out to you, since your team is literally called the Data Governance Team, and that's what my thesis is about. So that's really interesting. I already had a look online, but what is the exact role of the MDC team within Tata Steel Netherlands?

P4: You're mixing up two things there. Would you like a short intro first?

Interviewer: Yes, please.

P4: I studied mechanical and offshore engineering in Delft from 2010 to 2016, started here as a technical trainee, and got into data. I found it exciting because of the optimization potential. I left for a while and came back. Since this summer I'm Data Governance Lead, my predecessor left the company. Before this I worked on digital transformation, so I might not have all the answers, but I can refer you to someone on the team.

Interviewer: Great.

P4: The data governance process started when we began the data lake for manufacturing. We didn't want to throw everything in without rules, so we developed a framework around it.

Interviewer: How is that structured now? I read something about the TSM LE platform. And you mentioned I was mixing things up, like the MDC team?

P4: MDC originally stood for Manufacturing Data to the Cloud. It focused only on manufacturing. The team included data engineers handling ingestion, plus two governance staff. Before any data was uploaded to the lake, they ensured a data owner, a data steward, and a classification were assigned. I can share the documentation.

Depending on the classification, data marked as internal or public is accessible within the company. Restricted and highly sensitive data require special access control.

Interviewer: How does that access control work?

P4: Internal and public data are grouped in one schema in the lake, with 1000 to 2000 tables. Restricted and highly sensitive data are separated into buckets. Each bucket has a separate schema and may contain one or multiple tables. Access is requested and, once granted, gives access to all tables in that bucket.

Interviewer: What kind of data is in those tables?

P4: It varies, mostly financial, environmental, or strategic. We're working on ESG reporting, waste, water management, and emissions are often classified as restricted.

Interviewer: Who defines the classifications?

P4: The data owner decides. We as the governance team define the process, but don't classify. We give input, but the owner knows the data best.

Interviewer: So you manage the data lake?

P4: We ensure the data entering the lake follows a secure process. Our team sets policy and procedures. The Data Lake team, MDC, loads the data once governance is approved.

Interviewer: What are the biggest challenges in governance?

P4: We have a landing zone and a trusted zone. To enter the landing zone, you need an owner, classification, and steward. To move to trusted, the data must be documented. That's a big challenge. Without column descriptions, no one knows what the data means.

Interviewer: What happens if it's undocumented?

P4: Then data scientists can work with it to assess which columns matter. Before a model or dashboard goes live, those annotations must still be added.

Interviewer: How is data quality and consistency ensured?

P4: That's metadata management. Project teams define their own quality standards. The Lake team keeps pipelines running and data up to date. Often the source schema changes without notice, pipelines break, and that causes problems.

Interviewer: What's the biggest business challenge?

P4: Data lives in too many places. Governance works well for manufacturing. But HR, finance, and commercial teams use different methods. HR needs table-level access control, our bucket method doesn't fit that.

Interviewer: Why not use the manufacturing process company-wide?

P4: Commercial teams see it as a hassle. Finance and HR are open to joining, but commercial keeps their own warehouse and think that's fine. They already have their data and don't want more steps.

Interviewer: But the idea of a lake is that data enters only after governance?

P4: True, but commercial already has its data in warehouses. Switching would mean reclassifying 200 tables, extra work.

Interviewer: How does DAMA fit into your approach?

P4: We use DAMA as a reference. We were planning a team training in October. The framework helps define necessary elements. I wasn't deep into it yet, the process was already running. I focus now on optimization and reducing bureaucracy.

Interviewer: Could DAMA raise Tata's maturity level?

P4: We're not focusing on maturity now. That could be a next step.

Interviewer: And how does this help with sustainability?

P4: Better management reduces CO, by storing only useful data and showing if targets are met. If your data is organized, you can optimize better.

Interviewer: If the Green Book used the data lake instead of Excel, would that help?

P4: Not directly to hit targets, but yes, it would give faster insight into whether we're on track.

Interviewer: What's the biggest reason teams don't switch?

P4: Management. If leaders don't see that the current approach is unsustainable, there's no budget or time. Only then can we talk about execution.

Interviewer: This really helps me map the main governance challenges.

P4: Great! Let me know if you want to dive deeper.

Interviewer: Thanks so much!

Interview with P6 – Transformation Office lead

Interviewer: I'm currently mapping what ideal data governance should look like. It's quite a vague concept and I've identified some bottlenecks so far, such as version control and unclear ownership. What do you see as the biggest obstacles? What do you run into most when data governance isn't in place?

P6: We're currently in talks with a party capitalizing us, that includes banks, the parent company, and the government. This is all new. Nothing is being repeated, it's uncharted territory. We're operating on heroic effort, without a pre-existing structure. These conversations involve complex business cases full of assumptions and engineering-type calculations. Then we get feedback and iteratively adjust the numbers. I see situations where we change a value from 3 to 4 because it better fits the feedback, but no one records why we did that, what the rationale was, or how that affects other data.

Interviewer: That makes it hard to ensure consistency and traceability.

P6: Exactly. We have documents lying on different desks all saying different things. You could argue it's due to new insights, but if you can't explain the differences, it creates a reputational risk. If those four desks meet and compare, it's immediately clear there's no consistency, multiple 'versions of the truth'. Iterations should refine figures, but the process to document changes is immature.

Interviewer: I'm using CMMI in my research to assess maturity.

P6: I don't think we're even at level 1.

Interviewer: I haven't formally assessed it yet, but level 1 is 'poorly controlled and unpredictable', while level 2 means 'some tasks can be repeated'.

P6: We definitely don't meet level 2.

Interviewer: So the core problem is changes without version control or a shared, traceable repository?

P6: Yes. I don't care what tool or platform is used, in a mature organization, I expect data to be stored, versioning to be traceable, and a clear owner responsible. That allows accountability for figures and decisions based on them.

Interviewer: So ownership is crucial. Do you see data security as a related problem?

P6: Absolutely. For example, the Head of Terms in the Joint Letter of Intent contains stock-sensitive information. It's currently just on people's laptops or in JCPenney. We don't know how it's secured.

Interviewer: How does this impact the Green Book?

P6: Everything ultimately comes together in the Green Book, the overarching business document. The Blue Book contains the engineering plan. Together, they justify the investment.

Interviewer: So better data governance improves reliability and indirectly helps accelerate the sustainability transition?

P6: I don't know if speed is the point. The Green and Blue Books just need to be well-founded. But poorly managed data can make us seem untrustworthy. We already have a history of clumsy emissions publications, people label us as "fudging the numbers". No one does it deliberately, but human error and lack of structure create that impression.

Interviewer: What should be the first step to improve governance?

P6: The first step in CMMI is making processes repeatable. A repository helps, but we first need to classify our data, define what we need, what requires strict management, who owns it, and where it comes from.

Interviewer: The governance team told me that no data enters the lake without all metadata in place.

P6: That's a good rule. But the challenge is making people understand why it matters. You can build a data lake, but if people don't value it, it's useless. Raising awareness is step one.

Interviewer: So my thesis shouldn't just highlight what's missing, but also outline the first steps toward a higher maturity level?

P6: Yes. I'd find it useful if you mapped where we are now and gave a clear plan to move forward. And that's a moving target, what's sufficient now might not be in a few years.

Interviewer: Are there others I should interview?

P6: Definitely. P7 does economic analysis for the long-term assets and business cases. Sumit from Finance manages the business case. And Albert Cassis handles energy contracts and procurement, his data must be in order.

Interviewer: One thing I'm still unsure about is benchmarking. Do you think it adds value?

P6: Benchmarking happens, but does it drive change? If you tell me my neighbor lost weight, that doesn't help me do it. Accessing other companies would be time-consuming. You can generalize best practices instead.

Interviewer: So it's more useful to describe what companies at this scale typically do?

P6: Exactly. If you work on billion-euro projects, traceable processes are expected. That's your baseline.

Interviewer: That helps me sharpen the scope. One thing I still wonder: everyone knows governance needs to improve, so why hasn't it changed?

P6: That's the core issue. We all know it. We feel uncomfortable working late on Excel files that go to the Ministry of Climate the next morning, but we keep doing it. Why? That's a great question to explore.

Interviewer: Maybe a lack of pressure or shared ownership.

P6: Could be. That's something you might uncover in your research.

Interviewer: Thanks, this was very helpful. I'll reach out to P7 and move forward.

P6: Good luck, let's catch up again soon.

Interview with P7 – Long Term Asset lead

Interviewer: Could you briefly explain your role within Tata Steel Netherlands and the Green Book project?

P7: I'm P7. My role focuses on long-term asset strategy. I mainly work on answering the question: what must we do to first achieve 40% CO2 reduction, and eventually reach 100%? That means figuring out which factories should be phased out, which new ones we need, and how to restore a balanced site. The key is aligning the material and energy balances. It's challenging, removing a blast furnace and coking plant gives you different volumes than the new technology you add. It doesn't all fit nicely. So what's the best solution?

We also ensure the scenarios are economically substantiated. We calculate the cost per ton of steel under each scenario. That includes inputs like iron ore, scrap, coal, and natural gas, and we price those based on the route. This allows us to compare the current and future routes, together with the Finance department.

Interviewer: So long-term asset management includes both material flows and the business case?

P7: Exactly, and also engineering. They tell us a new plant has a specific capacity, uses so much electricity per ton, so much gas, etc. We ensure those figures are included in the site-wide material balance.

Interviewer: Where do these assumptions about the plants come from?

P7: From engineering, the Heracles team. They consult suppliers like Tenova and Danieli. Based on specifications (materials, inputs, scrap quality), they provide consumption data for the plant. That goes into our model. So the suppliers are an important source.

Interviewer: How dependent are those assumptions on internal Tata Steel data?

P7: Less so for new installations. That's like a new car, you get the specs. For existing plants, we use data we've known for years. We model on an annual level, using aggregated data from yearly plans. No sensor data. We work with averages, like those in annual reports.

Interviewer: Is that data stored centrally?

P7: We use the CSM model. It describes all plants with consumption figures and fixed costs. Those figures come from department data books, usually in Excel. Input files are sometimes transferred manually, sometimes loaded automatically. They're stored on the CSM site and managed by our department.

Interviewer: Do you load the data yourself? Does it come from a data lake?

P7: No. For example, plant capacities are projections we receive from other teams. The data lake contains historical data. We extrapolate into the future. So if a hot strip mill gets upgraded, we include that.

Interviewer: And you use the CSM model for the Green Book?

P7: Yes, it's the backbone of the entire business case. Inputs are traceable, with version control and change management. If a CapEx amount changes, the responsible department issues a new

Excel, and we adopt it using standard formats. It's not fully automated, there's a lot of manual work. That's how it was built, and there's no alternative.

Interviewer: Is there room for improvement?

P7: Definitely, especially in the discipline of submitting updates. Data transfer needs to be more reliable and secure. Copying numbers isn't the issue, that's quick. The main thing is verifying if the new numbers are correct and understanding why they differ from earlier versions. That's more of a management issue than a technical one.

Interviewer: Do other teams have access to the same data?

P7: Limited. We haven't structurally addressed that. Departments use their own data. Heracles, for instance, knows little about the power plants, understandably. The CSM model is complex. We don't want it floating around either, that causes misinterpretations and confusion. We want to keep it canonical.

Interviewer: Doesn't that lead to inconsistent figures?

P7: People write based on the data they own. But it's important to do sanity checks, especially where there's overlap. People need to check they're using the same numbers.

Interviewer: I've noticed many questions revolve around timelines and versions. The CSM model seems more mature than many other datasets.

P7: Yes. We typically have a "frozen" version and then developments. The current version is still internal. We have to be vigilant about that.

Interviewer: Do you work with others on the Green Book?

P7: I mainly deliver my own section. Sometimes I ask for support, but that's organizational, not content-related.

Interviewer: How does better data management contribute to decision-making on sustainable investments?

P7: It helps a lot. Especially if you think carefully about organizing data transfers and stick to procedures. There's definitely a gain, making sure information flows from department A to B effectively.

Interviewer: So you do see room for improvement in data governance?

P7: Absolutely. Especially in the departments involved in the Green Book.

Interviewer: Thank you for your time. Do you think I should speak to anyone else?

P7: Not right now. I don't know who you've already interviewed.

Interviewer: Thank you, and good luck with your work.

P7: Good luck with your thesis.

Interview with P8 – Data Engineer, Data governance team

Interviewer: What is your role within Tata Steel?

P8: My name is P8. I've been at Tata for almost six years now, five and a half internally, one year externally. I currently work as Lead Data Engineer in the Data Lake Team. I started as a Data Scientist, so more on the analytics side, and gradually shifted towards infrastructure: building the platform that Data Scientists can work on and that enables people to run analyses or reports. My job is mainly about ensuring data from various source systems comes together in one central place. I work closely with the Data Governance Team, I believe you've already spoken to P4?

Interviewer: Yes, I have.

P8: Then it's familiar territory.

Interviewer: So your work focuses on the Data Lake?

P8: Yes.

Interviewer: And that's mostly for manufacturing data, right?

P8: Correct. The organization is a bit split when it comes to data. On the commercial side, you have the DSS and CMA teams, they focus more on commercial projects and have their own smaller Data Lake and a Data Warehouse. On top of that is the Manufacturing Data Lake, where I work. It's broader; we gather data from factories like sensor data and installation settings, but also HR, HSE, and finance data. It spans the whole chain.

Interviewer: What's the difference between a Data Lake and a Data Warehouse?

P8: A Data Lake is where you load raw data, for example, sensor data straight from the plant. You make that data accessible with an authorization layer and a governance layer: who can see what, what does the data mean, where does it come from. A Data Warehouse is more tailored to specific use cases. For instance, you might turn millisecond data into daily averages for reporting. That adds a lot of context.

Interviewer: In my team (Transformation Office), we don't work with the Data Lake, and we experience issues like lack of ownership, version control, and loss of context. Is it true that those things must be arranged in advance in a Data Lake?

P8: Yes, at Tata we work with three zones: Landing Zone, Trusted Zone, and Consumption Zone. Before data can enter the Landing Zone, there must be a Data Owner and a classification: who is allowed to see it? That's essential for data security, you want to avoid unintended data breaches. So we require that upfront. If it's not in place, the data won't be loaded.

Interviewer: How about version control?

P8: That depends on the type of data. Sometimes you want to keep historical versions (e.g., for address data with time stamps), other times you only want the most recent version. It's a strategic choice: how important is history for your use case?

Interviewer: Do you know the Green Book?

P8: No.

Interviewer: It's a report with 32 chapters and 15 authors, sent to Tata Steel Limited in India for investment decisions on decarbonization. Each team delivers data. Would a Data Lake be useful for something like that?

P8: Depends on the data. But if you want consistency and centralized access, definitely. A lot of that data is probably already available. If everyone uses the same source, you get fewer inconsistencies. If their calculations are recurring or broadly reusable, you can put them in the Data Lake. If not, you can use a custom Data Warehouse linked to the Data Lake, then at least the source is uniform and centrally governed.

Interviewer: What's the bottleneck preventing teams from using a centralized governance system?

P8: Time and lack of awareness. People see data governance as extra work. If you currently have your own access and someone tells you to define ownership, classification, and column meanings first, it feels like overhead. People often don't see the benefit yet. But once teams experience the reusability and reliability, they change their minds.

Interviewer: Do you think implementation should be top-down?

P8: It helps if it's on management's agenda. They set the strategy and can create space for data topics. But top-down alone isn't enough. You also need buy-in from the people on the ground, show them why it's valuable and involve them in the process.

Interviewer: Should Tata offer more training?

P8: Absolutely. We have the data awareness training from P4, where I've also been a trainer. But that's high-level. What we're missing are in-depth workshops where people really learn how to work with governance, Data Lakes, and Data Warehouses. Just doing, practicing, digging deeper.

Interviewer: What's the risk of data fragmentation?

P8: If you measure or interpret the same metric differently in different places, you can't build chain-wide insights. You lose comparability. That's fatal for sustainability or CO calculations. Fragmentation makes your analyses useless and blinds the organization.

Interviewer: Thinking of the Green Book, what's the biggest governance risk?

P8: Fifteen people using fifteen different methods is a recipe for inconsistency. Especially when reporting to investors or government, you need one clear story, based on shared definitions and datasets.

Interviewer: Do you believe data governance contributes to sustainability goals?

P8: Absolutely. Without governance, you don't even know if your figures are correct, let alone prove impact. You have to justify your numbers to governments, auditors, and internally: how were these numbers produced? That's impossible without governance.

Interviewer: Some say governance matters less than the data itself.

P8: Maybe for some things. It doesn't always matter whether person A or B owns the data. But without definitions, version control, provenance, and validation, you can't do reliable analysis. Then you don't know if your improvements are real, you're just fooling yourself.

Interviewer: Thank you so much.

P8: No problem.

7.3 Annex III: Document analysis results

This annex presents 70 structured observations derived from internal reports, spreadsheets, and workfiles reviewed during the case study at Tata Steel Nederland. Each observation was used to support the triangulation of findings in *Chapter 4*.

70 observations from the document analysis, displaying the index, referenced document, the observation, and the link to one or more criteria.

Index	Document	Observation	Link to criteria
1	D6/D1	The list of owners contains names of people who are currently inactive in TSN.	This links to data unit responsibility and the challenge of data ownership, for who is now responsible, and who will write the chapter?
2	D6/D1	The column 'status update' contains multiple cases of "no update?", as well as "no one".	This suggests a lack of clarity on the timeline of some of the chapters, and untransparent interoperability between teams. This links to interoperability as a criterion. This also proves that there are data silos.
3	D6/D1	Not every workfile has the version number added to its file name.	This observation links to the lack of version control and the same-named criterion.
4	D6/D1	Two of the 11 workfiles have more than one owner.	Multiple document owners can cause data ownership difficulties, linking to the data ownership challenge and data unit responsibility criterion.
5	D6/D4	Most IF deliverables have not been assigned a writer. Though it has been submitted	Not being assigned a writer, the documents' context tends to be lost over time, because people will only look back at their own registered work. This links to lost data, version control, and data unit responsibility.
6	D6/D2	The Blue Book has better data governance in the DCR.	This links to manufacturing data and practices versus sustainability reporting. Manufacturing has a long history of documenting its data very carefully, which shows in the document control register.

7	D6/D1	The Green Book index's first column WO5 has no clear description	The column has values either 'x' or cancelled, or blank. These values' meaning has not been specified, which jeopardizes their context. Linking it to the Data Lifecycle Management and Classification criteria.
8	D6/D1	The status report of the Green Book has been documented very well.	All statuses of documents per owner have been kept in a sheet. The sheet shows all status stages and how many documents per owner are at which stage, showing the overall progress of the Green Book. This shows good data storage.
9	D6/D1	The index of the GB in the DCR is not consistent with the GB table of contents.	Inconsistent index names may cause parts of the Green Book to be lost and fit in the wrong place in the Green Book. This is linked to Lost Data and the criteria classification.
10	D1	A status table with colors per stage has been added to the beginning of the document for each chapter, but is not kept for each chapter.	The idea of classifying chapter statuses is good, but it isn't followed up, therefore it is linked to the criterion classification.
11	D1	There has been made use of a standardized format at the beginning of each chapter, showing the proof point the chapter covers.	The proofpoints are listed in the Datapoint sheet, but are 100+ points classified per team/category, which is not specified in the GB. This lacks classification, version control, and ownership.
12	D1	Cross references to unknown chapters.	References are made to chapters that are yet to be numbered, so it's not known for each writer where other topics will be placed. This proves poor interoperability, causing fragmentation and lost data.
13	D1	Unlabeled figure.	The figure has not been titled or numbered, making the context of the data unclear, as well as classification and data storage.
14	D1/D7	Unclear ownership in comments	Though there are 11 workfiles which together form the Green Book, the integrated version has comments tagging people without mentions of teams. This stimulates unclear ownership.

15	D1	Unclear data support reference	Each chapter is written and owned by different people, but the reference to supporting documents, like the workfiles, is not present. This jeopardizes version control, data unit responsibility, classification, and interoperability.
16	D1	No version number added	Though it's verified that the integrated version is not up-to-date, and the workfiles are. The Green Book has been filled partly with content from the workfiles, and comments have been added. This results in the lack of version control and data lifecycle management.
17	D1	Standardized format in place	The appendix included a table where the text owner and writer are specified, and approval is tracked. This is clear data sanitizing and is only used in the appendix. Not including this in the rest of the document results in poor classification, version control, and data unit responsibility.
18	D1	No abbreviations are specified	Abbreviations across the document are not further explained or specified in the text, or on page 182. This results in context loss, thus linking to the criteria data classification and data lifecycle management.
19	D1/D2/D8	Inconsistent datapoints in the GB and ECP.	Where in the Green Book, 0.5 billion cubic meter biomethane per annum is converted into 19.1 PJ/a , the ECP says that 0.5 bcm (17 PJ/a) has to be sourced. This is an inconsistency that affects Data classification and data security. Interoperability because the pieces are written by different teams.
20	D1/D2/D8	Inconsistent datapoints in the GB and ECP.	Where the Green Book suggests that TSN's need for biomethane equals 40% of the Dutch production in 2030, the ECP says 30. This too affects Data classification and security, and compliance.

21	D1/D2/D8	Inconsistent datapoints in the GB and IM.	Where Green Book says that renewables will increase to a 84% of electricity generation share in 2035, the IM says 80%. This inconsistency also affects data classification and security.
22	D1/D2/D8	Inconsistent datapoints in the GB and IM.	The Green Book says one time that base case means 70% of electricity (2.4 TWh) is not generated on site, and one time 69%. IM says 70%. This small inconsistency proves difficulties in classification and security, and storage.
23	D2/D4/D8	Inconsistent datapoints in the ECP and IM.	External power sourcing increases to 9 PJ/a says the IM, where ECP says 8.6. Again, classification, security, and storage.
24	D8	Inconsistent use of units across documents.	Especially in EUR per annum or just EUR. Often, it is not specified whether a number refers to yearly numbers or not. This affects lifecycle management and classification.
25	D1/D2/D8	Inconsistent datapoints in the GB and IM.	The Green Book mentions a price of 30 EUR/MWh natural gas in 2030, and the IM uses 25-35 for the same number. This proves the lack of a centralized data storage where numbers can be referred to. Thus, it is linked to data storage, security, and lifecycle management.
26	D1/D8	Inconsistent datapoints in the GB.	Within the Green Book chapter, the onsite electricity generation share differs between 30 and 31%. This proves a lack of reference to a central data model. Links to classification, storage, and lifecycle management.
27	D1/D2/D8	Inconsistent datapoints in the GB and IM.	The timeline is not consistent in the two reports. Green Book says it wants to source biomethane by 2033, whereas the IM says from 2033.
28	D2/D4/D8	Inconsistent datapoints in the ECP and IM.	Natural gas consumption in Phase 1 of the transformation is inconsistent, 38.9 in the ECP and 39 in the IM. Data storage, lifecycle management, and security. Interoperability because the pieces are written by different teams.

29	D1/D2/D8	Inconsistent datapoints in the GB and IM.	Inconsistent timeline, 12.6 mton/a CO2 emission is the case. The Green Book uses this number for 2019 as a reference for reduction, the IM calls it the current numbers. Security, interoperability
30	D8	Inconsistent use of data across documents.	The CO2 emission reduction is set at 40% by 2030 compared to 2029. This is confirmed by the ECP and IM, but the Green Book uses 35-40%. This jeopardized security and compliance, as well as lifecycle management and storage.
31	D2/D4/D8	Inconsistent datapoints in the ECP and IM.	The needed sourcing for slabs to keep to rolling mills running during the transformation, which is normally acquired from traditional steel-making, is set at 1.1 mton/a by the IM, but at 1 mton/a by the ECP. This too affects classification and lifecycle management.
32	D8	Inconsistent use of data across documents.	The same goes for the reduction of CO2 in megatons per annum. This is 5 mton/a by 2030. This is confirmed by IM, but the GB and ECP say 4.5-5 mton/a. This inconsistency jeopardizes security and compliance, as well as storage, lifecycle management, and interoperability.
33	D7	The targeted criteria for the proposed data governance framework in the data management Apollo slides. (D9.1)	The slides aim to improve the data ownership, linking the first PowerPoint analysis to data unit responsibility. It also wants to verify data, linking it to security and compliance. The third criterion is a centralised data platform, which links to data storage.
34	D7	The Data security Apollo slides propose a framework for the classification of accessing certain files. (D9.2)	The data security slides stress the importance of classifying finance and planning files as classified, other than energy and technical documents. This links to data classification, security.
35	D7	The document owners are specified (D9.2)	For classified documents, the workstream lead should be responsible, according to the slides. This links to data unit responsibility.
36	D7	Adding a new classified data map requires approval. (D9.2)	TO's (P1-3) need to approve adding another classified map, proving good governance and linking to data unit responsibility, data storage, and data security.

37	D7	Rules are applied. (D9.2)	The security slides note the rule not to use ChatGPT for handling classified documents. This links to data security and compliance.
38	D7	Document management slides' problem definition of data governance (D9.3)	"Repeatedly collecting the same data or information for multiple reports, leading to redundancy and inefficiency. This links to data fragmentation, version control. The criteria are data lifecycle management, interoperability, data unit responsibility, and version control.
39	D7	Document management slides' problem definition of data governance (D9.3)	"Difficulty in locating relevant and latest data for analysis and reporting." This links to the criteria data classification, data storage.
40	D7	Document management slides' problem definition of data governance (D9.3)	"Lack of validation processes for ensuring the accuracy and reliability of the collected data." This links to data unit responsibility, version control, and data security."
41	D7	Document management slides' problem definition of data governance (D9.3)	"Absence of a standardized method for freezing and versioning data, leading to potential inconsistencies and confusion." This links to version control, data storage, and data classification.
42	D7	Document management slides' problem definition of data governance (D9.3)	"Difficulty in tracing the source of data used in reports, hindering transparency and accountability in reporting processes." This links to version control, data lifecycle management, and data unit responsibility.
43	D7	Document management slides ask whether TSN is mature enough to adopt a new way of working. (D9.3)	This links to all criteria, because it questions the overall capability of TSN to adopt a new data governance framework.
44	D7	Document management slides ' proposed solution to the data governance problem. (D9.3)	Data collection: create a repository that contains all the formalized data files, and all data owners are in place. This links to the criteria data storage, data unit responsibility, and data classification.

45	D7	Document management slides ' proposed solution to the data governance problem. (D9.3)	Data validation: Implement validation processes Mb/Cb/Ab for ensuring the accuracy and reliability of the collected data. This corresponds to data unit responsibility, version control, and data security.
46	D7	Document management slides ' proposed solution to the data governance problem. (D9.3)	Data versioning: Implement a standardized method for freezing and versioning data, leading to potential inconsistencies and confusion. This corresponds to version control, data storage, and data classification.
47	D7	Document management slides ' proposed solution to the data governance problem. (D9.3)	Data traceability: Implement a method for tracing the source of data used in reports, hindering transparency. This links to version control, data lifecycle management, and data unit responsibility.
48	D7	Document management slides ' proposed solution to the data governance problem. (D9.3)	A standardized format requiring the writer, owner, and date of data. This links to data lifecycle management and version control, as well as responsibility.
49	D7	Document management slides ' proposed solution to the data governance problem. (D9.3)	Proposing a "historic event overview that contains the details of the changes that were made and their respective timestamps." This proposal links to version control and data lifecycle management.
50	D7	The document management slides apply to the Green Book (D9,3)	There is a specific slide that applies the proposed data governance rules on the Green Book, saying that the reporting needs structural data governance. This links to data storage, interoperability.
51	D7	Data governance strategy includes change management (D9,3)	The strategy includes "training and support to employees to ensure adoption", which links to all criteria.
52	D7	Data governance strategy includes continuous imrovement (D9,3)	Links to performance monitoring. "Establishing mechanisms for ongoing monitoring"
53	D7	Data governance strategy includes defining data governance (D9,3)	This aims to establish clear policies and responsibilities for managing data. This links to data unit responsibility and version control.

54	D7	Data governance strategy includes data security (D9,3)	This links to data security and compliance, because the framework needs to "protect sensitive data from unauthorized access".
55	D7	Data governance strategy includes data lifecycle management (D9,3)	Links to data lifecycle management, because data needs to be tracked to its "archival or deletion".
56	D7	Example of a standardization format	The example includes version, writer, owner, and date, expressing the need for criteria version control, data unit responsibility, and data classification.
57	D7	Classification strategy for different levels.	This strategy proposes 4 levels of classification: internal use, restricted, classified, and public. This links to data classification.
58	D7	Rules include only sharing signed-off documents, using MS Teams, and SharePoint.	This links to data security and compliance, as well as to data lifecycle management.
59	D4	Confidentiality statement on each page	The line saying the document contains confidential information links to data security and compliance.
60	D5	Workfiles have overlap in GB chapters.	Several workfiles have overlap in chapter writing, resulting in potential difficulties concerning data unit responsibility and interoperability.
61	D5	Owners are sometimes not specified in the file name, writers are almost nowhere.	This lack of documenting results in interoperability and data unit responsibility issues.
62	D5/D7	Workfiles do not follow the PowerPoint's proposal.	The proposed standardized format is not followed in the workfiles, raising version control, data lifecycle management, data unit responsibility and security issues, and data classification issues.
63	D5	Changes are made based on comments in the workfiles.	Comments from members in other teams in the workfiles suggest changes, which raises interoperability issues.
64	D5	Workfile progress is significantly different across workfiles.	Some workfiles are practically done, while others have not been started. This links to interoperability, and data unit responsibility
65	D5	Not all figures and tables are numbered.	Unnumbered figures link to the lack of data classification.

66	D5	Comments look differently on the workfiles' functions	Some comments suggest a change, which is answered with: I think this goes deeper than is meant for the Green Book. This suggests a lack of interoperability.
67	D5	Overlapping content within workfiles	A comments suggest that the content is already covered in another chapter. This is due to issues in interoperability, data unit responsibility.
68	D5	Workfiles are inconsistent in using standardization formats.	A lack of interoperability results in that some workfiles do use the standardization format with writer, owner approval, and date, while others don't, and also classification.
69	D5	Inconsistent referencing across workfiles.	In some workfiles, in-text referencing has been used, while other files use footnotes. This links to data classification and data lifecycle management.
70	D5	Lack of explaining abbreviations	In the workfiles, no section explains all abbreviations. This links to data lifecycle management, to interoperability, to classification, and data storage.

7.4 Annex IV: Focus group transcript

This annex contains the transcript of the focus group held on the 4th of June 2025. The session was used to validate the recommendations and assess their broader applicability. Participants P2, P3, and P6 discussed each criterion and reflected on relevance, implementation, and transferability. The transcript has been translated into English and lightly edited for clarity.

Focus group with P2, P3, and P6

Researcher: I will briefly reintroduce this focus group by explaining what I've done so far and what we'll do during this discussion. Then we'll go through the recommendations, which are the final product of my research, one by one per criterion.
The meaning of each criterion will become clear as we go. This focus group is also formally intended to evaluate the broader applicability of the recommendations, an academic requirement that my TU Delft supervisor was eager for me to address. After that, we'll have a short discussion, and then we'll wrap up.
So first, a quick reintroduction to my work. Since mid-February, I've been conducting research here into data governance practices at Tata Steel Nederland. I started with a literature review, which I thankfully had time for, thanks to your support, and from that review, eight data governance criteria emerged. I then carried these into a case study at Tata Steel Nederland.
Feel free to interrupt me at any point, by the way.
In that case study, I used document analysis, think Green Book, work files, IM, JF, anything I

was allowed to access. I also gathered observations, interviews with the three of you, P1, and an external participant, as well as P4 and P8 from the data governance team. The observations were more of a secondary data source, mostly just things I saw around me during my time here.
So we're now going to go through those eight data governance criteria. For each one, I'll start with a short definition. I initially wanted to hand out a summary so we could skip that part, but instead I decided to prepare a short presentation, easier than printing something last minute during lunch.
The goal of this focus group is to validate the eight recommendations.
I interviewed two main groups: people in management roles and people in operational roles. I don't belong to either group myself. And while I found the research fascinating, your input will add significant depth to my findings. I'm particularly curious to hear from your perspective what you think is needed for successful implementation, both within and beyond TSN, because we'll also touch on broader applicability. As I mentioned, assessing whether these recommendations could work in other industries is part of the academic contribution.
We'll start with Data Lifecycle Management and spend a few minutes per criterion.
I'll read it out for clarity: Structuring the management of data across its entire lifecycle, from creation to storage, usage, updating, and finally archiving or deletion. This requires clear versioning, status labels, and insight into what was changed when. The recommendation here was to use consistent status labels like "draft," "final," or "definitive." Also: record changes in a logbook, and enforce mandatory steps before files can be shared or saved.
We'll now discuss three reflection questions for each of the eight criteria. The first question: Do you see consistent status labeling as a useful recommendation at TSN? Meaning: Have you seen this go wrong? Would this be helpful?
We can go around in turns, or anyone can just speak up. For example, P6, feel free to start. Also, if anything is unclear, please jump in.
P6: Go to P3.
P3: In theory, it would work, but it really depends on the discipline of the team using it. I do think it's helpful to manage information properly and clearly label it, so everyone knows where the latest version is and what quality level it has. Convenience plays a big role. But it's also about realism, it's easy to make a recommendation like this.
Researcher: But do you think that once I've made this recommendation, it could actually be implemented at TSN? Or would that require a lot of effort?
P3: It mainly depends on the group that's going to work with it, they have to carry it. You can design and set up a system like this well in principle, but it stands or falls with actual use by all the disciplines that supply this information. That's often where the difficulty lies. So I think it really needs to go hand in hand, the group that produces the data has to embrace the approach.
Researcher: Lastly, do you think this recommendation, or this concept of Data Lifecycle Management, is something specific to TSN based on my document, interview, and observation analysis?
P2: I think this applies to any company that produces anything data-related. You just need to be clear together on what you're looking at, whether it's validated, and what quality level you assign to that information. I do think the first step, using consistent status labels, is a minimal step, but a good start, as long as you clearly define what each label actually means.
Researcher: Okay, a good start. Thank you.
P6: I recognize it as a useful recommendation for TSN. What I find interesting, and I'm not sure if you described this, is what happens if you don't do this. Because the funny thing is, if you ask someone like X, X, or P7, they'll all say: "Yes, this is necessary, it's a smart idea." But still, they don't do it. So there's something cultural here that's a bit strange. Everyone acknowledges it, no

one finds it odd, but it still doesn't happen. So the real question becomes: what is needed for successful implementation? That might be something you're also being asked: why doesn't this work? Which elements are generic, and which are specific to TSN?
I actually think it's all generic. But what I find especially interesting is how the concept of "definitive" is handled. Because I know that documents labeled as definitive, like the pre-notification, are already outdated. So I believe those status labels also need to be managed on a managerial level. At a certain point, the content becomes obsolete. It may say "definitive," but maybe it should be "archived," "outdated," or "for reference", something that clearly signals it's no longer a source of truth. So, in your recommendations, I would find it helpful if you go beyond just "draft" and "definitive" as status labels.
Researcher: Alright. The next criterion is interoperability, the extent to which systems, documents, and teams can consistently exchange and reuse data. It helps prevent duplicate work, inconsistencies, and confusion in reporting.
The recommendation is to develop and use a single template for strategic reports, with linked data fields, fixed validation points, and joint checks by multiple teams. In short, a central validation workflow ensures consistency. This is something I personally noticed over the past few months. P2, would you like to start?
P2: Nice idea, but not realistic. For example, the CSM model isn't visible or accessible to most people. That's already a problem. Sure, we use similar templates, but the data itself isn't handled consistently. One team takes full responsibility and manually updates everything. A good first step would be building a system where people can view the data without being able to change it. I understand the idea of fixed validation points, but the incoming data isn't consistent, and deadlines vary between teams. Just like with the lifecycle topic, it depends heavily on team discipline and how things are managed.
Researcher: P6, what would be needed to implement this successfully, even if P2 says it's too ambitious for now?
P6: People need to understand the alternative. What goes wrong if we don't do this? What are the risks?
Researcher: So, the consequences.
P6: Not punishment, necessarily, but where will things start to hurt? Take the boardnote format, if you don't use that, your topic doesn't make it onto the agenda. That's one of the few templates actually enforced around here. We should introduce similar mechanisms. Not to enforce it for the sake of control, but to help people understand the benefit, or the downside of not using it.
P3: Otherwise, people will just ask, why should I change? That clarity has to come from the management level.
P6: Sometimes I just look at biology. Why does a dog come when you say "come"? Because it expects a treat. In a pack, it doesn't get a treat, it just drops a few ranks if it doesn't listen. That's how things work.
So, for successful implementation, it's not about designing a template. I've seen thousands of templates here, and everyone still uses their own. A central template would help, but the real issue is that this is a consequence-free company.
Researcher: And if we look beyond TSN, to other companies facing similar challenges? You have backgrounds in automotive, offshore, and finance. Does TSN lag behind?
P6: Yes. But the key is what it delivers. If you want to validate strategic reports, you need a shared template. And you have to steer based on that. One way to do that is by saying: if a report isn't in the right format, we won't take it seriously, we won't review it, we won't include it.
P3: Exactly. It's about behavior. People will only start using the structure if it affects outcomes.

Researcher: I agree. Let's move on for time's sake. The next criterion is data classification. This overlaps a bit with lifecycle management, but is not the same.
Data classification is about assigning a clear label to every document or dataset indicating how confidential it is, who it's intended for, and how it may be shared. This increases awareness and prevents unintended disclosure of sensitive information.
Unlike status labels, this includes labels like internal, confidential, public, but also draft or final. The recommendation is to make it mandatory to add such a label before uploading or sharing a file. I've noticed that this sometimes runs into practical issues.
P3: I've experienced that myself. The default label is "general." If you want to make something more secure, you have to change it. I once labeled something confidential because it was sensitive, but then half the people couldn't access it. So I just switched it back to general so they could actually open it. That defeats the whole purpose.
P2: The most important thing is that people are aware of what they're sharing and with whom. And the consequences of sharing confidential information need to be very clear. People need to weigh that themselves.
P3: Exactly. That's more important than the label itself. And labels suffer from inflation. I often see documents marked "confidential" that are so irrelevant that no one would care. The labeling is inconsistent and not taken seriously.
Researcher: How could this be implemented more successfully? Would it help to give people more access rights?
P2: I think it's more about attitude and awareness than system design. This really comes down to discipline. We already have a classification system. And people here are aware of data sensitivity. I've seen cases where major decisions were known internally for months but never leaked. So the behavior is actually quite solid, that matters more than a label.
Researcher: Can I move on? One more thing: would you say this recommendation is broadly applicable to other organizations as well?
P6: Yes.
Researcher: You once said that in the financial sector, you're held accountable even for small things. Here, you can share almost anything. So is this less relevant in that context?
P6: It's not quite like the example you gave, but I do believe many companies, even small startups, are already doing this. It's probably easier for them than it is here. For example, last Thursday, documents were submitted to KPMG for validating the enterprise model. I'm sure those were marked strictly confidential. And yet on Saturday, a journalist claims to have read them. So clearly, labeling alone is not enough, people must also act accordingly. Information security shouldn't stop at classification. The company should invest effort in finding out who leaked that file.
Researcher: That's a good point. Let's build on that. The next criterion is data security and compliance, protecting sensitive data and ensuring that all data use complies with laws and policies. This includes access control, classification visibility, and responsibility in sharing information.
The recommendation: mark documents with access levels, store sensitive files in restricted folders, and assign one person per file or folder to manage access. You mentioned earlier that you set something to general access. What if someone else had been responsible for that access level, would that have helped?
P3: That might help if you're unsure about access. But I had no doubts about the people I was sharing with. I was in a meeting with them where we discussed everything. The document just summarized what was already said. So I didn't feel I needed stricter controls.

Researcher: Do you think it even matters that much here? I was able to access a lot of things too. It didn't seem that risky.
P6: It definitely matters. Look at the pre-notification and business cases, many documents are not meant to be widely accessible. With some effort, you can get to them, but you're not supposed to.
P3: It's all fairly accessible. We can access a lot of documents, and we need that access. We're also very aware of what we're looking at. So this is something we already do to some extent, it's not new.
P6: Still, I think we could be much stricter. And that applies to all the criteria. We have templates, labels like confidential or strictly confidential, and we do security training. We use draft and final statuses. But the maturity level is still too low.
P2: What's missing here, compared to most companies, is a central data point, like a document control team that manages rights and access and is actively tasked with setting that up.
Researcher: So you're saying TSN doesn't have that, but other organizations do?
P3: We used to do it within this project. We were the ones keeping an eye on it.
P2: I think what's missing is a team within Tata Steel that takes full responsibility for this. A team that immediately gets involved when a new transformation program starts.
P3: I don't think you solve this with a team. It should just be part of everyone's job description, making sure access rights are set up correctly and being evaluated on that. That's more effective than having a central unit that never has full visibility anyway.
Researcher: Let's move to version control, systematically keeping track of versions and changes to avoid outdated or uncontrolled information being used. The recommendation is to always include a version number and a brief changelog before sharing or storing a file. Is that something useful or realistic here?
P6: It should be linked to the templates. If you use a reporting template, versioning should be part of it. For me, draft and definitive status already imply versioning. Anything before the decimal point is a draft, and the full version is final. Finance uses this: version 0.0 is a draft, 1.0 and up are approved, and 9.9 is obsolete. If you see 9.9.8, that means don't use it anymore.
Researcher: But we don't see that system here, do we?
P6: Not at all. I've seen files labeled "concept 3.0" or "final concept." Sometimes there's not even a number. A consistent system would really help, but again, what happens if no one follows it? Then there's no reason to do it.
Researcher: Right. So people don't feel the pain of not having it.
P2: Exactly. Nobody really notices the problem.
Researcher: Next is data storage, structured and centralized storage of files so everyone knows where to find inputs, templates, and results. This avoids duplicates and confusion. The recommendation is to use a shared Drive with structured folders and clear access rights. But right now it's team-specific, right?
P6: Yes. My team has already moved to OneDrive. But there are no rules. If you start your own folder structure tomorrow, no one will stop you. When we were still Apollo, we had some say in this. But once Heracles came in, it all fell apart. No offense, Pim, but these are safe recommendations, and yet they're not being implemented. They seem generic, but why aren't they working here?
Researcher: That's a good question.
P6: You're the researcher, right?
Researcher: I think what happens is that one team creates a folder structure, and then later the project moves and it gets a new structure elsewhere.

P6: Storage is a universal issue. It's been a big topic for IT companies for years. But the real question is, why doesn't it happen here? If this is proven methodology, why don't we apply it? We all know what's needed, but we don't do it.
Researcher: That's covered in the results chapter, here we're just reviewing the recommendations. I would've needed a whole afternoon with you to tackle the root causes.
P3: We would've done it.
P6: You're doing this for us. It's fine.
Researcher: I couldn't ask that much of you. Let's go to the final topic.
P6: Exciting.
Researcher: Performance monitoring, tracking the quality and progress of data governance so improvements are visible and issues are discussed. The recommendation is to define a small, realistic set of KPIs, have the data governance team track them, and share results quarterly in the management meeting.
P6: That's a good one. I like it.
Researcher: These KPIs could reflect the previous seven criteria. I see this as something that follows once the others are in place. Do you agree?
P2: Actually, I see the first seven criteria as initial KPIs. This is easy to implement, since we're already working with KPIs. The first seven can be your data hygiene indicators. How many documents are labeled? How many use the same label? If everything is labeled "general," you know nothing is really changing. But you need a data governance team to manage that.
Researcher: I've spoken to the data governance team. They're enthusiastic about training others, and some of them run webinars. They use a fairly mature data lake system and are aware that their way of working is different from most teams. I think it would be interesting to organize something with them when the timing is right.
P3: That brings us back to what I said earlier. This needs to be part of everyone's job. Good governance over data, everything we discussed here, has to be something you are expected to deliver. If not, no one will bother. It's hygiene work, and people skip that when they're busy. It gets left out. So I really think it needs to be included in performance reviews or job assignments. Otherwise, the governance team might give great webinars and reports, but no one uses them. There have to be consequences, or at least recognition that it matters. That's key.
Researcher: That would help. Are there elements that are specific to TSN, or is all of this broadly applicable?
P2: It definitely applies to other companies. But here at TSN, it really feels like we're starting from zero. Even the baseline KPIs are lower than average.
Researcher: Do you think it would be effective here to make small improvements visible and part of regular presentations, for example, led by the data governance team?
P2: That would only work if it comes from management. If a governance team shows up and tells someone they're doing a poor job, the reaction is usually: "Oh well, that's unfortunate for them." Then people move on, and nothing changes.
Researcher: That leads to the main point raised in the interviews: TSN operates without consequences unless priorities are set from the top. Without that, governance is just extra work, not something useful or necessary.
So, top-down commitment is essential. At the same time, bottom-up training, like that from the governance team, and recurring performance presentations can create momentum. A third point is how all the criteria reinforce one another. Version control improves with central storage. Classification becomes clearer with clear ownership. Everything is interconnected.
Do you recognize those first two? Or maybe even all three?

P6: I recognize all three points.
Researcher: Even the third one?
P6: Yes.
Researcher: Let's say you present the agreed KPIs in regular performance meetings, showing upward trends after training sessions. Would that be an effective way to implement change?
P3: I'm not sure. Maybe because I personally dislike bottom-up training. It should just be part of your job. That comes back to the need for top-down prioritization. Only then do training and skills matter.
P6: Earlier at lunch, you walked through the turnstile and held your hand to the card reader. Why did you move your hand away?
P3: Otherwise, my hand would get stuck in the door.
P6: And?
P3: That hurts.
P6: So?
P3: I wouldn't be able to enter the building.
P6: Was that a top-down rule?
P3: No.
P6: Were you trained for that?
P3: No.
P6: Is it connected to how you park your car?
P3: No.
P6: Yet you still do it. That's behavior driven by avoiding pain. That's the simplest form of behavioral change. You do something to avoid pain, protect something, or improve something. There are hundreds of management books about it, but it's that simple. These tools are just instruments. The real question is, why isn't it happening?
We all use data, we unpack our laptops and get to work. But how we handle data isn't seen as important. That's interesting. And also, do you recognize these patterns outside TSN?
Actually, I'd narrow it down further. Your scope has been Green Steel, basically a few former Apollo people. That doesn't reflect all of TSN. If you had given this presentation to Finance, they might say it's a waste of 30 minutes. That's not a critique of your energy, but it shows how low maturity is around data management in this part of the organization.
It would be interesting if, in your final weeks, you checked out a different project like Heracles or XYZ. You might find that all eight criteria are already being applied there, and that would be powerful. It's not that the ideas are missing, they're just not present in this environment.
Researcher: Got it.
P6: Did you record that?
Researcher: Definitely. One more question, though: Do you recognize these practices in other jobs?
P6: Absolutely. I just said that, and you'll hear it on the recording.
P3: I helped build the BI department in automotive, and this kind of structure was standard. Probably still is. People were expected to take this seriously.
P2: At Boskalis, our client required full integration of revision steps and reviews, they had scheduled checkpoints and expected full visibility of every versioned document.
P3: It's a more predictable process. The better your setup, the smoother it runs. Here, things swing around constantly. Internal or external shifts force things to pivot completely.
P2: At Boskalis, a ship cancellation forced a rewrite of the final version. We ended up with a C5 revision because the structure didn't fit. But the client insisted on another revision, even though

it was already marked approved. They wanted a traceable revision history. That's what made the process stay intact.
P6: It would be useful to compare more TSN teams. Each has its own view and dynamics.
Researcher: Definitely.
P6: Shall we wrap up?
Researcher: Yes. I want to thank you all again. I'll anonymize this in the final chapter so you know how it will be used. Thank you very much.