International Journal on Advances in Intelligent Systems, vol 15 no 3 & 4, year 2022, http://www.iariajournals.org/intelligent_systems/

Supporting Systems Thinking Application by Data Analysis

A Case Study: An Automated Parking System

Haytham B. Ali
University of South-Eastern Norway
Kongsberg, Norway
Haytham.ali@usn.no

Gerrit Muller
University of South-Eastern Norway
Kongsberg, Norway
gerrit.muller@usn.no

Mo Mansouri
University of South-Eastern Norway
Kongsberg, Norway
mo.manouri@usn.no

Fahim A. Salim
University of South-Eastern Norway
Kongsberg, Norway
fahim.a.salim@usn.no

Kristin Falk
University of South-Eastern Norway
Kongsberg, Norway
Kristin.falk@usn.no

***Abstract*— This study applies Systems Thinking (ST) and its tools to define and validate a case study in the early phase of a complex socio-technical research project. We used ST and other tools, including a stakeholder interest map, context diagram, and Customers, Actors, Transformation, Worldview, Owner, and Environment (CATWOE) analysis. These tools are the foundation of ST systemigrams, which is a top-down approach. Further, we support ST with data analysis, which is a bottom-up approach. In this context, we collected and analyzed failure data. We applied machine learning in terms of Natural Language Processing (NLP), Frequent Pattern Growth Algorithm (FBGL) for association rule mining, and the Gensim model to cluster the failure data. The case study indicates that both approaches complement each other as we apply them in an iterative and recursive manner. Data analysis supports ST, and ST guides the data analysis. Furthermore, ST implementation facilitates understanding, communication, and decision-making regarding the case study and its multiple units of analysis. Moreover, we adapt Nonaka and Takeuchi's model to articulate the tacit knowledge within the Company using Systemigrams, canvas in the form of A3s, and post-its. We adopted the Systems Engineering methodology to construct the canvas we used in the workshops and interviews.**

Keywords- Case study; Systems Thinking; systemigram; failure data; data analysis; machine learning.

I. Introduction

This paper is an extended version of the article presented at the Modern Systems Conference 2022. The conference paper aims to apply Systems Thinking (ST) to validate a case study definition in the early phase of a complex socio-technical research project, where the case is an Automated Parking System (APS) [1]. ST is a process focusing on understanding the problem and its aspects and relations among these aspects as a whole [2]. In this context, Barry Richmond, one of the pioneers of ST, mentioned that it is essential for a systems thinker to look at the tree and forest simultaneously [3]. In other words, ST focuses on synthesis by analyzing the whole system, its parts, and its dynamic behavior. Synthesis is one of the foundations of ST, emphasizing parts, things, or aspects to understand them through the context of their relationships. On the other hand, the analysis focuses on dealing with one part, thing, or aspect as a system using reductionism. Reductionism is a process of breaking down a system into its elemental part, then describing the whole as a sum of its elementary parts of the system [4].

However, the system as a whole is more significant than the sum of its parts, things, or aspects [5]. In this context, a system can be complex, a problem definition for a case study, etc. The contribution of this extended version is applying both ST and analysis in a complementary manner. ST, which is a top-down approach, guides failure data analysis. Failure data analysis, which is a bottom-up approach, supports ST. In other words, ST and data analysis support and guide each other in an iterative and recursive manner. In addition, the extended version addresses tacit knowledge articulation from the Company's key persons in terms of data and visualization using ST and Systems Engineering (SE) methodologies. The International Council on SE (INCOSE) defines SE as "*Systems Engineering is a transdisciplinary and integrative approach to enable the successful realization, use, and retirement of engineered systems, using systems principles and concepts, and scientific, technological, and management methods*." [6]

We use systemigrams, an ST tool, for this articulation, where we use SE methodology to construct a virtual board using tools such as a canvas in the form of A3s and post-its described in section IV. This implementation enhances the understanding of the case study in breadth through ST and SE and in depth through data analysis.

This research uses a case study within a complex socio-technical project called H-SEIF 2 [7]. H-SEIF stands for "Harvesting value from big data and digitalization through the Human Systems Engineering Innovation Framework." H-SEIF 1, the first research project, focused on developing new human-centered SE methods to cope with the rapid increase of socio-technical complexity within the systems and market needs [8]. H-SEIF 2, the following research project, focuses on digitalization, enabling data-supported early decisions and capturing value from big data. The H-SEIF 2 research project seeks to enhance the companies' product development process through relevant industry cases. The research initiative uses an industry-as-laboratory method and includes two universities and nine companies from various fields. Defining and early validating the case study is essential to ensure the research project's success as it facilitates a Company's active participation and sharing of all needed data. Ali et al. [9] are also one of the foundations of the H-SEIF 2. That research [9] showed the value of analyzing data to close the feedback loop to the early phase of the product development process.

The Company's case study delivers Automated Parking Systems (APS), mainly in metropolitan cities. Metropolitan cities can benefit from Automated Parking Systems (APS) since there is a shortage of available land and a need for more parking spaces [10]. However, APSs frequently fail for two reasons: when used often and when the end-user is unfamiliar with the APS. Additionally, some mechanical issues cause the APS to malfunction [2][3]. An APS has a higher failure rate than a conventional system. Consequently, there is a requirement to raise APS reliability [13].

This research applies ST and its tools to define and validate a case study early using ST and other tools, including systemigrams. Systemigrams also aided in articulating the Company's tacit knowledge in terms of data and visualization. In addition, we used a canvas in the form of A3s and post-its for this articulation using a virtual board. We used SE methodology to construct the canvas in the form of A3s, and we used post-its in the digital workshops. This articulation is crucial to maintain the knowledge management process in the Company, which is vital for a competitive advantage and survival.

Moreover, we supported the ST application with data analysis. In this context, we collected and analyzed failure data using machine learning. We used machine learning in terms of Natural Language Processing (NLP) [14] and further association rule mining to discover patterns and co-occurrence among the failure data, as the data are mainly in a text-free format that is logged by the maintenance personnel. We used the Frequent Pattern Growth (FPG) algorithm for association rule mining. In addition, we clustered the failure data using machine learning using the Gensim model.

A. Introduction to the Company Where We Conducted the Research

The Company is a small and medium-sized enterprise that delivers APS in Scandinavia, mainly Norway and Denmark. The Company has more than 36 parking installations. The Company primarily provides semi-automated parking systems, the System-of-Interest (SOI) for this study, which we refer to as APS in the rest of the paper. The Company is transitioning to manufacturing its parts instead of getting the parts from suppliers. In this transition, the Company's vision is to install sensors to develop a Condition-Based Maintenance (CBM) system. The Company believes that this vision will give the ability to increase the Company's market share for the whole of Europe. In addition, the Company believes that being a first mover in this direction will allow the Company to sell the CBM as a service to other industries. Further, they see it as a build-up toward a digital twin.

The Company uses an Excel file to log the failures called failure data for each parking installation. The challenge is to investigate the value of this data. Due to the complexity of the data, it can be called big data [9]. The use of (big) data will enhance the APS's reliability. This enhancement adapts to earlier data-driven decisions in product development and maintenance [15].

The research questions for this study are as follows:

RQ1: How can we define and validate a case study early, including its multiple units of analysis using Systems Thinking and its tools?

RQ2: How can we articulate tacit knowledge?

RQ3: How can we support Systems Thinking with data analysis?

Many authors suggest using Soft System Methodology (SSM) to develop a conceptual model. ST and its tools, mainly systemigrams, use SSM as a methodology [16]–[18]. Thus, ST implementation and its tools, primarily a systemigram, is a conceptual model.

The structure of the paper after the introduction section is as follows: (II) Background section with an informal literature review, (III) Research methodology section explaining case study research and Checkland's Soft Systems Methodology, (IV) results and analysis section that includes: (a) articulating tacit knowledge (b) Systems Thinking implementation (c) data analysis results, (V) a thorough discussion in the relation of the research questions, and ultimately (VI) conclusion. The paper has a detailed data analysis methodology in Appendix A.

II. Background

The background section provides informal literature, starting with defining ST and its tools, mainly the systemigram, then explaining the data, and information, ending by illustrating knowledge taxonomy and Nonaka and Takeuchi's model of knowledge creation regarding tacit knowledge articulation. The informal literature review illustrates the two disciplines, i.e., ST and data analysis. Many authors addressed a lack of empirical research to apply the two disciplines [19][20]. This research addresses this gap by

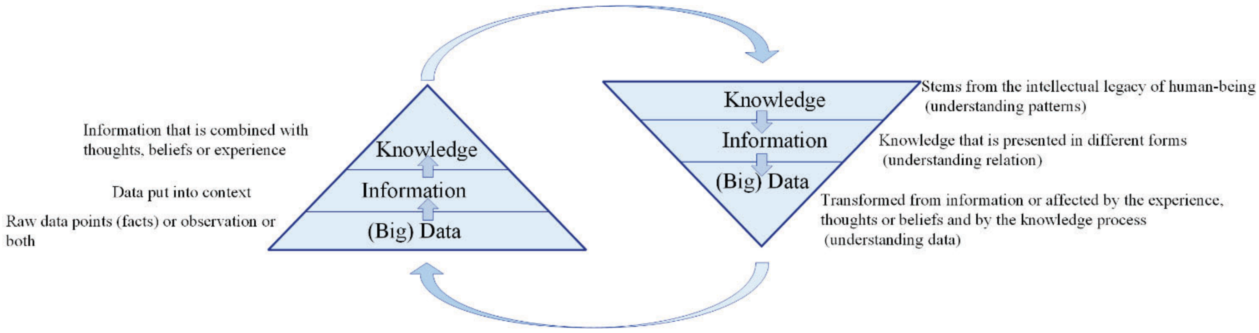

Figure 1. Data, information, and knowledge transformation and vice versa.

implementing both disciplines. In addition, we address tacit knowledge articulation in terms of data and visualization using ST and SE methodologies.

A. Systems Thinking (ST)

In 1994, Barry Richmond introduced ST as a term. He defined ST as the "*art and science of making reliable inferences about behavior by developing an increasingly deep understanding of underlying structure*" [3]. Barry Richmond emphasized one essential attribute of a systems thinker in this context: the ability to look simultaneously at the tree and the forest. However, there are many definitions of ST. Arnold and Wade proposed an ST definition based on a literature review and ST implementation [21]. We adopted this proposed definition in this paper:

"*Systems thinking is a set of synergistic analytic skills used to improve the capability of identifying and understanding systems, predicting their behaviors, and devising modifications to them in order to produce desired effects. These skills work together as a system*".

Furthermore, Sauser et al. [2] used ST methodology and its tool, i.e., systemigram, to define a problem. The authors chose the ST methodology, as ST gives an understanding of how things affect each other as a whole. The authors visualize this understanding using a systematic diagram technique (ST's tool) called Systemigram.

1) Systemigram

A systemigram, also known as a systemic diagram, is an ST tool introduced by Boardman [22]. A systemigram is a graphical visualization using storytelling. This visual presentation consists of nodes and links. The nodes are names, and the links are the verbs between those names. A Systemigram aids in communicating through visualization a problem with its aspects and their relations. These aspects can also be related to defining a problem, solution, or both within its context. Gharajedaghi [23] claims that problems or solutions always have a specific context.

A systemigram is usually based on another ST's analysis tools, such as context diagram and Customers, Actors, Transformation, Worldview, Owner, and Environment (CATWOE) analysis [1][18]. In addition to other tools, such as a stakeholder interest map to perform a stakeholder analysis.

Implementing ST and its tool, including systemigram, needs data and data analysis support. Data and data analysis increases understanding of the several aspects within the problem or solution domain by investigating patterns among data [19]. Supporting ST and its tools with data analysis improves verification and validation [25]. In addition, ST facilitates an overview. Further, ST enhances understanding, communication, and decision-making, while data go more in-depth, and we can transform it into information through data analysis. This information can be further transformed into knowledge using ST. In other words, data and its analysis complement ST and its tools.

B. Data, information, and knowledge

Data, information, and knowledge have different meanings in knowledge management [21][22]. However, they are interconnected and related to each other. Figure 1 shows that we can transform data into information and further knowledge and vice versa.

One of the leading models that describes this transformation and differences is called the Data, Information, Knowledge, and Wisdom (DIKW) hierarchy [21][23]–[25]. Wisdom presents the future, vision, design, and implementation. Wisdom mainly means understanding the principles in order to implement them [24][25].

Data are discrete and non-significant facts, i.e., raw data points. Data can be transformed into information through, for example, analyzing the data and further visualizing these results. We derive information from the data. We may use knowledge and thoughts in the deriving process [21][24][25].

Information is a significant fact and can be further transformed into knowledge. Information is data that gets a defined meaning through an information model. The various pieces of information have relations that increase the meaning of the information further [21][24][25].

Knowledge is information that is combined with thoughts, beliefs, or experiences. In other words, knowledge is a product of data and information combined with thoughts, beliefs, or experiences. Knowledge implies understanding the patterns in which ST aids in this understanding. On the other hand, we can transform knowledge in terms of data and visualization. Knowledge stems from the intellectual legacy of human beings. Knowledge also leads to information that identifies data [21][24][25].

1) (Big) Data

There are several definitions of (big) data. Based on a literature review and articles from the industry, De Mauro et al. [31] define (big) data as follows: "Big Data is the Information asset characterized by such a High **Volume**, **Velocity,** and **Variety** to require specific Technology and Analytical Methods for its transformation into **Value**." This definition includes the " V " notion many authors have used to define (big) data. [32]–[34] used what is referred to as the "3VS" to define (big) data: Volume, Velocity, and Variety. [35]–[37] added the fourth "V" value. [38] appended the 5th V to the (big) data definition, which is Veracity.

On the other hand, the Method for an Integrated Knowledge Environment (MIKE2.0) [39] project defines (big) data emphasizing the complexity and not the size of the data: "Big Data can be very small and not all large datasets are big." Intel's definition also emphasizes the complexity aspect: "*Big data has the characteristics of being complex, unstructured, or having high volume*" [40]. Table 1 lists the most cited definitions within the literature.

TABLE I MOST CITED BIG DATA DEFINITIONS FROM THE LITERATURE

Source	Definition
(Beyer, 2012) [41]	Big data is defined using the **3Vs: Volume, Velocity, and Variety**. These characteristics require a cost-effective **information-processing model** to **improve insights** and **decision-making.**
(Dijcks, 2012) [36]	Big data is defined using **4Vs: Volume, Velocity, Variety, and Value.**
(Intel, 2012) [40]	Big data has the characteristics of being **complex, unstructured, or having a high volume.**
(Schroeck et al., 2012) [42]	Big data means a mixture of **Volume, Variety, Velocity, and Veracity**. Big data leads to a **competitive advantage** within recently **digitized** industries.
(NIST Big Data Public working Group, 2014) [43]	"Big data consist of extensive datasets, primarily in the characteristics of **volume, velocity and/or variety**, that require a scalable **architecture** for efficient storage, manipulation, and **analysis**."
(Microsoft Research, 2013) [44]	"Big data is the term increasingly used to describe the process of applying serious **computing power**—the latest in **machine learning and artificial intelligence**—to seriously massive and often highly complex sets of information."
(Boyd & Crawford, 2012)[45]	Big data is datasets with a size greater than **traditional software tools' capacity** to capture, store, manage, and **analyze**.
(Boyd & Crawford, 2012) [45]	Big is defined as a **cultural**, **technological**, and **scholarly phenomenon** that rests on the **interplay** of, **Analysis** and **Mythology**.

Based on these definitions mentioned above, including Table I, data, or big data, is a structured and unstructured data collection. In this study, we collected unstructured data from the Company; thus, MIKE2.0 and Intel's definitions apply to our case study. In other words, we adapted these two definitions to define data or (big) data in our research. However, we may expand our adaption if we can obtain or collect more structured data, i.e., in-system (sensor) data.

2) Data Sensemaking

Data sensemaking is the process of understanding the data within its context. This process occurs iteratively and recursively using data analysis and the visualization of the data analysis results. Klein et al. [46] emphasize the importance of the frame or perspective around the data as it affects the data collection, analysis, and interpretation process. Data, i.e., using data, can also affect the existing or preexisting frame regarding change or reinforcement. This effect is one of the sensemaking aspects [47].

Weick defines data sensemaking as a two-way process to fit data into a frame and frame around data. Weick emphasizes that this process must be iterative until data and frames (mental models) unite. These iterations also aid in avoiding oversimplifications [48]. Klein et al. stated that data sensemaking includes several functions, such as problem identification and detection, forming an explanation, and seeing relations or correlations [46].

C. Knowledge Taxonomy

Many authors classify knowledge into two main types: tacit and explicit knowledge [49]–[52].

Explicit knowledge is the knowledge we can codify, transfer and articulate to natural language or symbols [53]. Harry Collins calls such articulation or transformation a string. A string is a physical object with recorded patterns. For instance, a figure is a string of ink recorded on paper, and a pixel is a string of recorded patterns on a screen [44][45].

Tacit knowledge is that knowledge that is generated through a person's thoughts, beliefs, and experiences. In other words, tacit knowledge is deeply rooted in individuals (e.g., mental models, know-how, personal skills, etc.). Tacit knowledge is embedded in the action, commitment, and involvement within a particular context, which makes it challenging to articulate [44][46][47]. However, there are many ways to articulate this knowledge to some extent using SE and ST [49]. One of the models that illustrate this articulation process is Nonaka and Takeuchi's model [51].

1) Nonaka and Takeuchi's model of knowledge creation

Figure 2 visualizes Nonaka and Takeuchi's model of knowledge creation, which is generated due to social and intellectual processes [44][46][49]. This model is widely used in knowledge management for knowledge transformation and includes the following four modes:

- Socialization. This mode is about transforming tacit to tacit knowledge. This tacit knowledge is mainly generated through interactions between individuals within a particular group. Thus, learning occurs through observation, imitation, and sharing experiences.
- Combination. This mode is about transforming explicit to explicit knowledge. This explicit knowledge is the knowledge that has already been captured. We can transform it into more evident explicit knowledge, i.e., form it better through deduction or induction of previously restructured items.

- Internalization. This mode is about transforming explicit to tacit knowledge. This transformation occurs through the "learning by doing" process, for example, following a written manual.
- Externalization. This mode is about transforming tacit to explicit knowledge. This transformation contains an explanation of practices and beliefs.

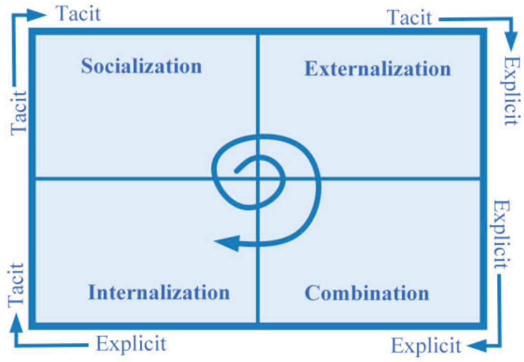

Figure 2. Nonaka and Takeuchi's model of knowledge creation.

III. Research Method

A. *Case Study Research*

In this research, we use case study research as part of industry-as-laboratory research during the research project [55]–[57]. Case study research consists of the following three steps: (1) defining the case study well, (2) selecting the design, and (3) using theory in design work [56]. This paper focuses on the first step in implementing ST. A case study often involves multiple units of analysis.

We collected primarily qualitative data. The qualitative data included direct observations, participant observations, open-ended (unstructured) interviews, and physical artifacts. The direct and participant observations resulted from the primary author being involved in a real-life context by participating in the events and meetings within the Company. We also conducted open-ended interviews separately and as part of the observation and part of the workshops, we performed with the Company.

In total, we conducted nine interviews with Company management, maintenance personnel, the head of the maintenance department, sales personnel, project leaders, and developers. In addition, we conducted three workshops with company management, maintenance personnel, and project leaders. The main author also participated in the weekly development team meetings for six months. In addition, the primary author also participated in the maintenance activities conducted by the maintenance personnel for two parking installations. The principal author also had an office in the Company and conducted several informal interviews while working from the Company's borrowed office.

Through observation, interviews, and workshops, we identified and collected stored data within the Company, also called physical artifacts. These data were downloaded by the Company's employees and provided to the primary author of this paper. These data are particularly failure data. The principal author analyzed these data, which are primarily failure data. Observations, interviews, and physical artifacts and their analysis are different sources that allow the collection of evidence from these sources. Thus, we can investigate the consistency of the findings from these various sources of evidence. Further, we can also converge these pieces of evidence, a process known as data triangulation, to increase the robustness of the results [58].

B. *Checkland's Soft Systems Methodology*

Applying ST in a case study within the industry-as-laboratory enables Soft Systems Methodology (SSM) and

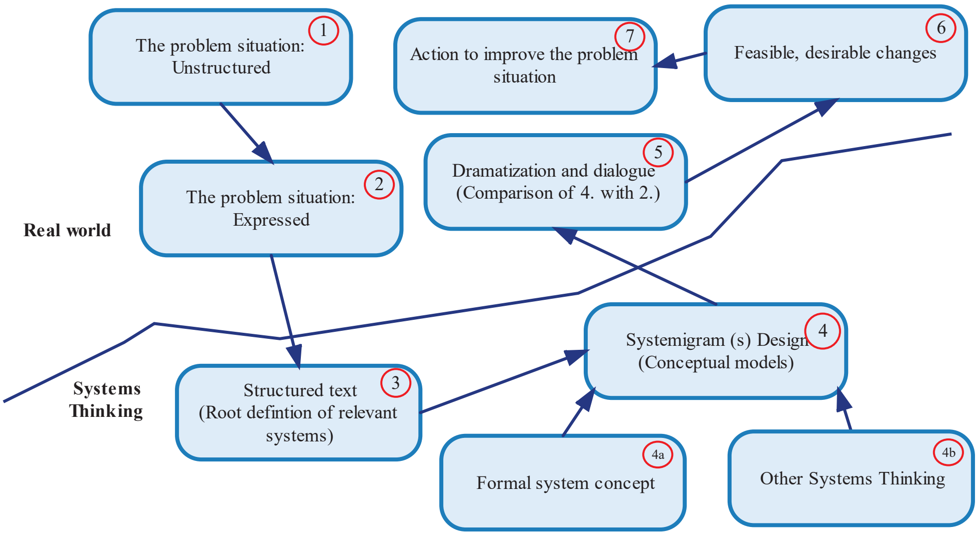

Figure 3. Checkland's soft systems methodology (SSM) is based on [11][15][54][56].

supports SE. Figure 3 depicts Checkland's SSM. We modified the method to be iterative and emphasized that there was no one right path. Further, we use the systemigram as a conceptual model, structured text as the root definition of relevant systems, and dramatization and dialogue as a comparison of steps 2 and 4. This modification was inspired by Sauser et al. (2011) [2], who called the SSM, which includes those modifications by Boardman's SSM (BSSM).

SSM allows for the inclusion of different perspectives and desirable outcomes of the case study. In addition, SSM bridges the real world and ST [11][15][54]. The SSM consists of the following seven steps, as Figure 3 visualizes:

Step 1: The Problem Situation: *Unstructured*. We have observed that the Company's request, as a part of the case study definition, is not wholly defined and validated early within the early phase of a complex socio-technical research project. The case can evolve and change based on many variables. These variables include the different perspectives and desirable outcomes of the various individuals involved in defining the case study.

Moreover, academia and industry emphasize different interests. For instance, the industry is more oriented toward maximizing profit and understanding how academia is more interested in research and the why and what [60]. The problem is defining the case study from different perspectives and including desirable outcomes from the various actors. Thus, early validation of the Company's request within the case study definition is a need within the early phase of the research project.

Further, the Company's request may touch the surface of the problem definition and not express the real need or problem definition that affects the case study within the research project.

In this step, the researchers conducted open-ended interviews with the Company, interactive workshops, and participant and direct observation. In addition, the primary author collected physical artifacts, which are failure data documents. The Company expressed through the interactive workshop a request. The Company's request is the start of the problem evolution. This request demonstrated the Company's vision through the visualization of a dashboard. This dashboard shows the condition of the Company's system through a traffic light code color. This dashboard has been called a Condition-Based Maintenance (CBM) system, which is the starting step toward developing a predictive digital twin system. In the workshops, we used Canvas in the form of A3s and pos-its to describe the problem situation and its aspects.

Step 2: The Problem Situation: *Expressed.* Alongside reviewing the literature, the researchers have created a text document on one page describing the Company's request within the case study, followed by meetings to verify the description.

Step 3: Structured text, known as the root definition of relevant systems. We moved to the ST domain by conceptualizing step 2. We used the CATWOE analysis in this step to understand the root definition of the APS, case study, and analyze it.

Step 4: Systemigram(s) design is a conceptual model. We developed two systemigrams models based on the output from step 3. In addition, we used other ST tools as input for the systemigrams. These tools include a context diagram, also known as an openness diagram, and a stakeholder interest map. Systemigrams capture the essence of the conceptual thinking for the Company's request within their case study.

Step 5: Dramatization and dialog by comparing steps 4 and 2. We moved back to the real world by comparing the systemigrams from step 4 with the Company's request, representing part of the case study description in step 2. Further, the authors dramatized the systemigrams via storyboarding in a Company workshop. The systemigrams as conceptual models were the basis for dialogue and discussion stimulation and for comparing the models with reality.

Step 6: Feasible, Desirable Changes We identified the feasible and desirable changes from Step 5 in a way that makes sense. These changes are translated into proposed leverage points and evaluated regarding the technical and cultural feasibility of the Company's request within their case study.

Step 7: Action to improve the problem situation. We called attention to the findings and applied them to our future case study work as part of the research project.

We repeated steps one to seven until the individuals involved in the case study reached a consensus. In other words, the process, including the steps, is repeated until the Company's need as part of the case study definition is verified and validated by experts from the industry (Company) and academia (scholars involved in the research project).

IV. Results and analysis

This section lists the case study results and analysis. The section starts with the tacit articulation and then the Systems Thinking implementation with its tools before it ends with the data analysis results.

A. *Articulating Tacit Knowledge*

We used SE and ST methodologies to articulate the Company's tacit knowledge in terms of data and visualization [49]. In the next subsection, we describe ST implementation and its tools, including the systemigrams from Company management and maintenance personnel perspectives (ref. III, B). These systemigrams are the data and visualization we used for the tacit knowledge articulation from the Company's key persons regarding the case study, focusing on Company management and maintenance personnel. On the other hand, Figure 4 visualizes the SE methodology implementation. We used this methodology implementation to structure a virtual board, i.e., the Miro board, for the workshops and interviews we conducted with the Company [61]. Figure 4 also represents the tacit knowledge articulation.

We first conducted interviews and workshops virtually due to Covid pandemic restrictions. Later, we conducted physical workshops and interviews. Due to confidentiality, we removed the most sensitive information from Figure 4. The point in Figure 4 shows how we constructed the digital board artifact, i.e., the canvas using SE methodology. Before the workshops and meetings, we sent a document, also called an artifact, that included a description of the research project, questions we wondered about, included among other, the aim of the workshops and meeting, and who we desired to meet, e.g., the Company's key persons. These key persons include Company management, system developers, and maintenance personnel. The initial workshops and meetings aimed to define the case study, including all its aspects. These aspects are also called embedded units of analysis [56].

Figure 4 includes the structure of the workshops, which contains six parts using a canvas in the form of an A3 and post-its. We portray these parts using yellow circles with a number. Part 1 includes the H-SEIF 2 research project descriptions, goals, and research questions. The Company we use as a case study in this research is a partner in the research project. Part 2 contains the Company introduction, description, visions, and so forth. Part 3 includes the Company's suggested aspects descriptions for the case study, the main challenges for these aspects, and expectations.

Furthermore, parts 4, 5, and 6 support the central part, i.e., part 3. Part 4 illustrates what, where, when, and how for the most significant case study aspect, e.g., data collation, identification, and analysis, to utilize data to enhance early-phase product development decisions. Part 5 discusses the critical stakeholders of the case study. Ultimately, part 6 illustrates the two approaches to utilize data toward digitalization, i.e., top-down and bottom-up, where we discussed the acting balance and iterations using the two approaches to achieve the intended goal(s) for the Company's case study. For part 6, the top-down approach starts by defining the questions we need to answer and then finding the appropriate data for these questions. The bottom-up approach begins by using data to exploit its hidden value.

In the workshops, the Company's key persons and scholars from academia from two universities that are partners in the research project have access to write and talk simultaneously. This access using post-its with different colors ensures that all participants, despite their personality (i.e., introvert or extrovert), express their ideas, beliefs, and thoughts based on their experience. This expression is essential to articulate the tacit knowledge within the Company's key persons toward the case study, all its aspects, and the academic scholar's tacit knowledge. This articulation is vital for defining the case study well.

Before the workshops, the academic scholars agreed on a workshop facilitator. The workshop's facilitator also ensured that all participants participated and had enough time to write, talk, or both. The facilitator also provided a warm environment during the workshop. After the workshops, we sent a one-page document to the participants, including the case study's definition and all its suggested aspects. We modified this document through several iterations until workshops participant, mainly the Company, verified and validated the document [62].

Figure 4. Canvas in the form of A3 and post-its we used to articulate the tacit knowledge where the Canvas's construction follows Systems Engineering methodology.

B. *Implementing Systems Thinking in A Case Study*

This subsection first illustrates the case study context by describing the SOI and systems boundaries. We demonstrate the system boundary using a context diagram. Further, this subsection shows the application of ST and other tools: a stakeholder interest map and CATWOE analysis. These tools are the foundation for systemigrams. The subsection ends by listing the possible leverage points from applying these tools after describing the systemigrams.

1) *The Case Study Context*

To understand the case study context, we first illustrate the SOI to understand the case study context. Further, we define the context for the SOI (i.e., the system boundaries through the context diagram). We categorize the variables for the system context using three categories: operation, development, and both contexts.

a) *System-of-Interest (SOI)*

As mentioned, the SOI for the case study is a semi-automated parking system (APS). The APSs are parking structures that store cars vertically to save place. The APS's design permits transporting vehicles from the entrance to the parking lot without the car owner being present. The degree of assistance from the car owners to the APS is the criterion used to distinguish between the fully and semi-automated parking system. The fully doesn't need any parking attendant, whereas the semi requires an attendant to drive or direct the car into the system [63]. The APS is a complex system because of its multitude of hardware and user interactions [13].

Figure 5 visualizes the SOI and its main parts. The main parts are: the gate, control unit, platform, and wedges. The end-users use the gate to open or close it to enter or retrieve their cars. The control unit controls the SOI and car owner, and maintenance personnel can use it as a panel to open or close the gate. The wedges are movable and help the end-users to park their cars in the correct position. Figure 5 also shows the SOI in 3D with a drive-in indication. This drive-in can be inclined or straightforward, depending on the building's architecture.

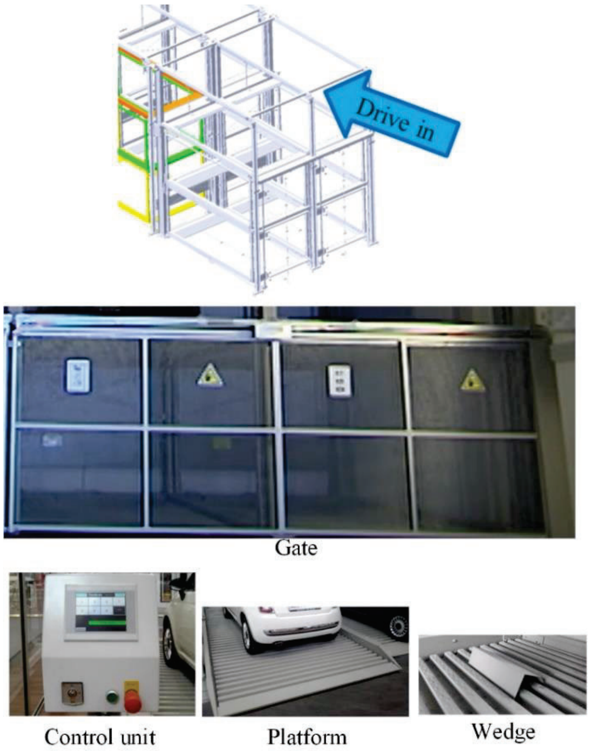

Figure 5. The System-of-Interest (SOI) and its main parts.

b) Context diagram: The system boundaries

We adopted the Gharajedaghis definition to define the SOI's boundaries [23]. Gharajedaghis defines system boundaries as "a subjective construct defined by the interest and level of influence and/or authority of the participating actors. The system, therefore, consists of all variables that could be sufficiently influenced by the participating actors". These boundaries aid in understanding the SOI in the context of their environment. Further, Gharajedaghis defines environment as follows: "the environment in which the system must remain viable consists of all those variables that, although affecting the system's behavior, could not be directly influenced or controlled by the participating designers."

Figure 6 portrays the system boundaries, also known as the context diagram or openness diagram, for the SOI, which includes three variables. We also categorize these variables according to operation, development, or both contexts. We used color text for this categorization, i.e., the green-color text refers to the variables that belong to the operational context. In contrast, the yellow-color text indicates variables that belong to the development context. The orange-colored text refers to variables that belong to both contexts, i.e., operational and development. The three variables are as follows:

Controllable Variables. The controllable variables are those that we can control. In this context, we can control the SOI (i.e., APS). We allocated the SOI in the innermost circle, which indicates that it is essential to act sufficiently to achieve the desired outcome.

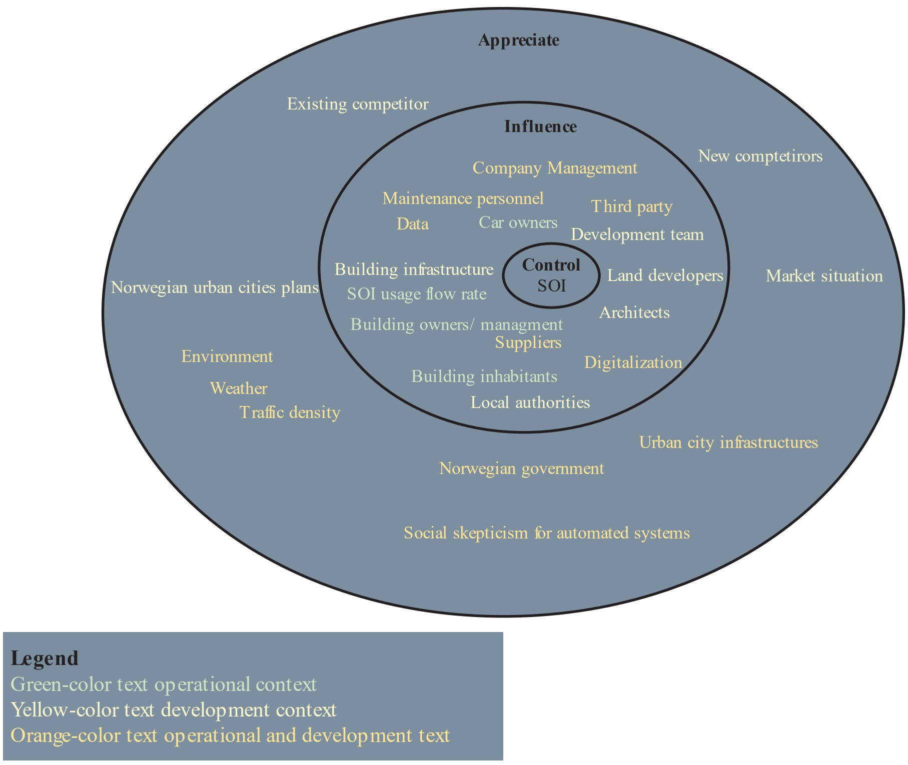

Figure 6. Context diagram of the System-of-Interest (SOI) with its variables.

Influencing Variables. Influencing variables are uncontrollable variables, but we can influence them. These variables include variables from operational, development, and both contexts.

The influencing variables that belong to the operational context are: car owners using the SOI, SOI usage flow rates, building owners or management, and building inhabitants, as they can affect the SOI operations. Further, we considered the following influencing variables in a development context: the development team, building infrastructure, local authorities, land developers, and architects.

Moreover, the variables that we included in both operational and development contexts are: *Company management*, as they are the decision makers for operation and development processes, *the third-party* as Company hires a third consultant party to develop or take part in the operation process, *maintenance personnel* as they maintain the SOI and take part in the development process due to their tacit knowledge, i.e., experience, thoughts, or beliefs, *suppliers*, as they supply the SOI with its part in the operational and development context, *digitalization* as we can digitalize processes or steps in a process that belong to the operational, development, or both contexts. For instance, a CBM system can be considered in the operational context. However, the analysis and prediction results from the CBM can also be used in the development context, that is, the next development cycle or version of the SOI.

In addition to the variables mentioned above, we included data as a variable in both operational and development contexts. Data depend on the data we identify, collect, and analyze. Data can belong to operation, development, or both contexts, depending on which data we identify, collect, and analyze. For instance, we can collect maintenance record data, also called failure data, from an operational context. We can use these data as feedback in the early product development process, which belongs to the development context.

Appreciating Variables. Appreciating variables are uncontrollable variables that we cannot influence. Thus, we must appreciate them. The appreciating variables that we considered to belong to the development context are: existing competitors, new competitors, market situation, and Norwegian urban city plans. On the other hand, the appreciating variables that we considered to belong to the operation and development contexts are: urban city infrastructure, social skepticism for the automated systems, the Norwegian government, and the environment, including weather and traffic density.

Even though we attempted to classify the three variables into operational, development, or both contexts, we noticed that it is not easy to have a clear distinction between those two contexts. One of the main reasons is that many variables, such as the data example we mentioned, can be used in both contexts. The same applies to other variables, such as the environment, that also affect or belong to both contexts.

2) Stakeholder Interest Map

Figure 7 visualizes the stakeholder interest map. Figure 7 also shows the SOI in the middle. Furthermore, we depicted the stakeholders with lines. These lines connect the stakeholders with each other and with the SOI. The lines are

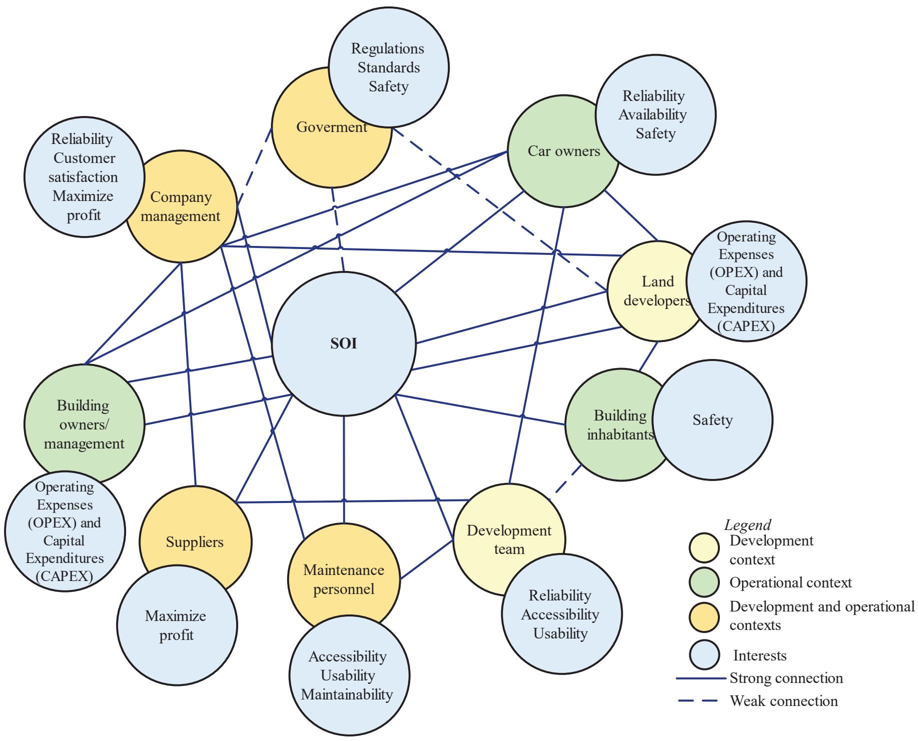

Figure 7. Stakeholder Interest Map

two types, i.e., solid and dashed lines. The solid line represents a strong connection, whereas the dashed line refers to a weak connection. This connection includes an interest in or influence on the SOI.

Moreover, we categorized the stakeholders into three categories, similar to the context diagram. These three categories are: stakeholders that are more involved in the development, operation, or both contexts. In addition, we listed the interests of these stakeholders in Figure 7.

Company management, maintenance personnel, government, and suppliers belong to operation and development contexts. The Company management strongly connects to the SOI, suppliers, and maintenance personnel. The interest of the Company management is maximizing profit, reliable SOI, and customer satisfaction. The Company management's customers are the land developer and building owners/ management. The customer of the customer, i.e., the end-user for the Company management, is the car owner. The interests of the maintenance personnel are: accessibility, usability, and maintainability. The suppliers' interests are: winning the contract to maximize the profit or Return on Investment (ROI).

On the other hand, the government has a weak connection to SOI and Company management. The government states the standards and regulations to ensure SOI's safety. These standards and restrictions affect the development and operation processes. However, the government is not involved in the details of SOI's development and operation to the same degree as Company management and maintenance personnel.

We categorized the land developers, development, and team within development contexts. The development team includes a third consultant party and key Company employees. The interests of the land developer are: Operating Expenses (OPEX) and Capital Expenditures (CAPEX) to maximize the Return on Investment (ROI), whereas the interests of the development team are: ensuring project success in terms of developing the SOI as reliable, accessible, and usable for the car.

Ultimately, the stakeholders that are more involved in the operation context are: building management or owners, building inhabitants, and car owners. The land developers often sell the SOI, including the building or estate, to building owners or individuals with shared ownership and select management for the estate. The Company is involved from the beginning, i.e., before the building is built. The building also includes inhabitants who either have no car or use traditional conventional parking, as some facilities include APS and traditional parking. The interests of the building management (owners of the building) are Operating Expenses (OPEX) and Capital Expenditures (CAPEX), whereas the interests of the building inhabitants are safety. The care owners' interests are reliable, safe, and available systems each time they use or park their cars.

3) CATWOE Analysis

To include the different key stakeholder perspectives, we used Customers, Actors, Transformation, Worldview, Owner, and Environment, called CATWOE analysis. This analysis is part of the SSM methodology. CATWOE analysis facilitates understanding the root definition of the system and analyzing it. The system can be a problem definition for a case study [24]. Figure 8 illustrates the CATWOE analysis. In this context, we apply CATWOE analysis to understand the purpose, need, or opportunity for the Company's SOI as a part of the case study from the two main stakeholder perspectives, i.e., Company management and maintenance personnel. Tables II and III show the results of implementing the CATWOE analysis regarding the Company management and maintenance personnel, respectively [1].

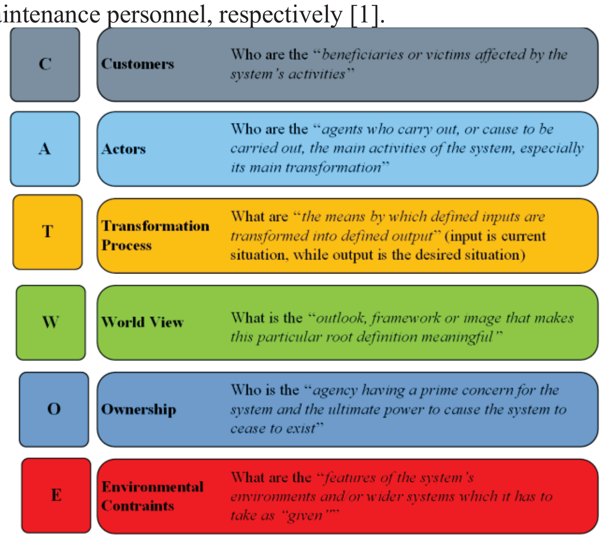

Figure 8. Customers, Actors, Transformation, Worldview, Owner, and Environment (CATWOE) Analysis.

TABLE I. CATWOE: COMPANY MANAGEMENT

Aspect	Description
Customers	Company management
Actors	Partners, suppliers, maintenance personnel
Transformation	Increase the reliability of the SOI
Worldview	Maximize profit
Owner	Company management
Environment	Urban cities

TABLE II. CATWOE: MAINTENANCE PERSONNEL

Aspect	Description
Customers	Maintenance personnel
Actors	Suppliers, Company management, car owners
Transformation	Maintenance process and method
Worldview	Increase reliability and availability of the SOI
Owner	Department heads of service and maintenance
Environment	The Automated Parking System (APS), buildings, cars, traffic density, weather, city infrastructure

4) Systemigram

A systemigram, also called a systemic diagram, is a graphical representation in the form of storytelling. The systemigram starts from the upper left with the SOI and ends on the bottom right. The SOI can be Company management, organization, or APS. The storytelling in the systemigram comprises visualized nodes and links. The nodes are names, and the links connecting these nodes are usually verbs. The main story of the systemigram is called the mainstay. The mainstay is usually a diagonal line in the systemigram.

Figure 9 represents the Systemigram from the Company management perspective. We also categorized the nodes in this systemigram into two colors: dark grey blue and light blue. The dark grey color refers to the mainstay, whereas the light blue indicates sources for (big) data. These sources include internal and external data sources. These sources include available stored data and possible sources to generate the needed data.

The mainstay for the first systemigram from the Company management perspective, i.e., Figure 9, is: "*Company management owns the SOI that in transition or manufacturing own parts that can include sensor(s), which permits anomalies observation remotely that aids mechanical failure detection and prediction in real-time that further allows continuous monitoring of the SOI that maximizes business viability that leads to maximizing profit.*" Business viability includes many other nodes: increase SOI's reliability, availability, and customer satisfaction.

Company management is transitioning to manufacturing the SOI's parts using a team consisting of the Company's key persons and a third consultant party. This manufacturing development process includes installing sensor(s) that allow continuous monitoring towards predictive maintenance through a digital twin. This digital twin includes monitors showing the real-time condition of the SOI and its parts and predicting possible failures. The Company management believes that this vision reduces maintenance costs and maximizes business viability and profit.

The Company's key persons that are part of the development team include the project leader and maintenance personnel. The maintenance personnel own tacit knowledge about the SOI. The researchers can interact with the maintenance personnel to articulate the tacit knowledge in terms of visualization and data. The Company can also take part in this articulation. The maintenance personnel manually log the failure using an Excel file that includes a sheet for each paring installation. This file is called service-log or failure data. We consider these data as feedback data into early phase

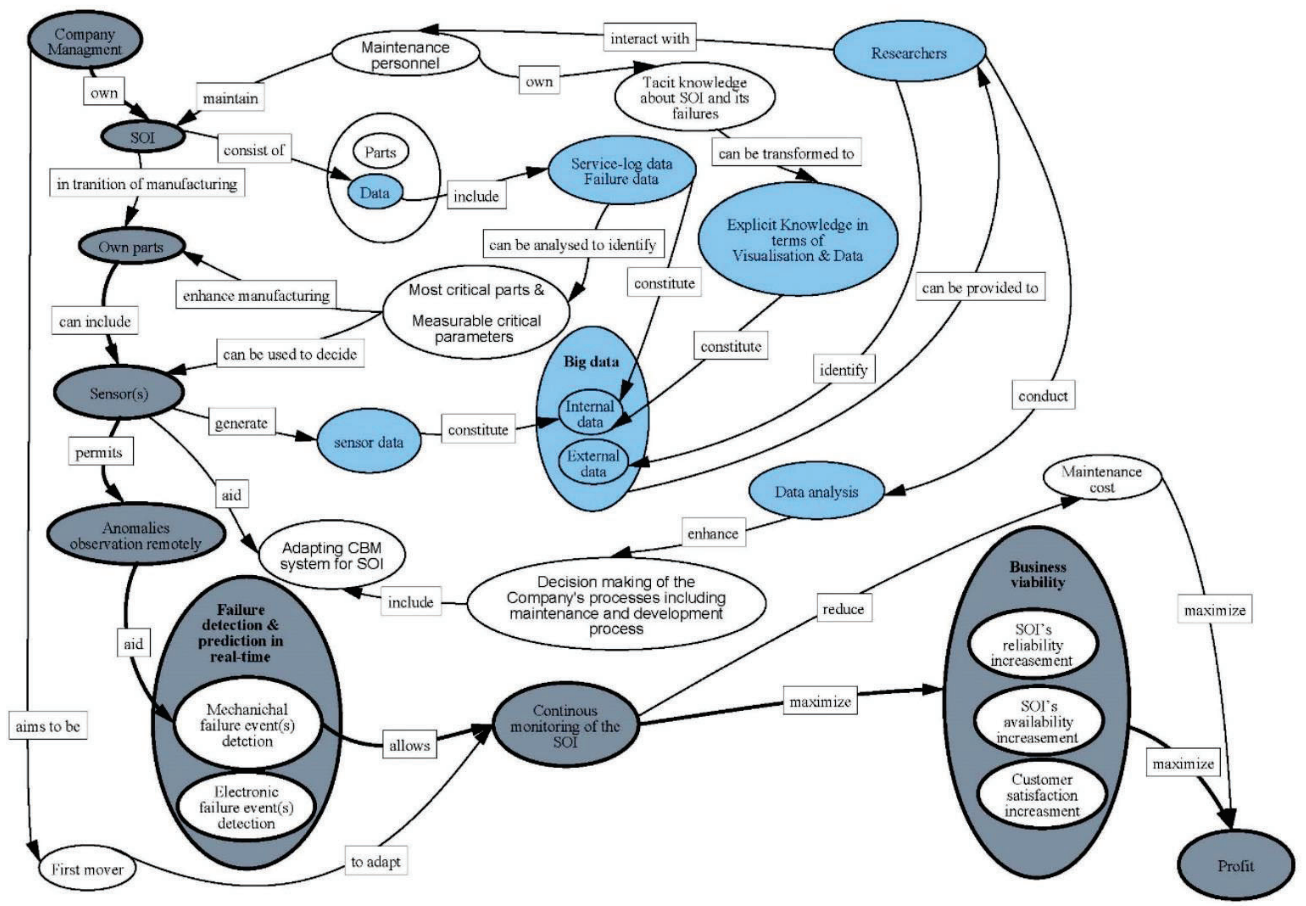

Figure 9. Systemigram from the Company management perspective.

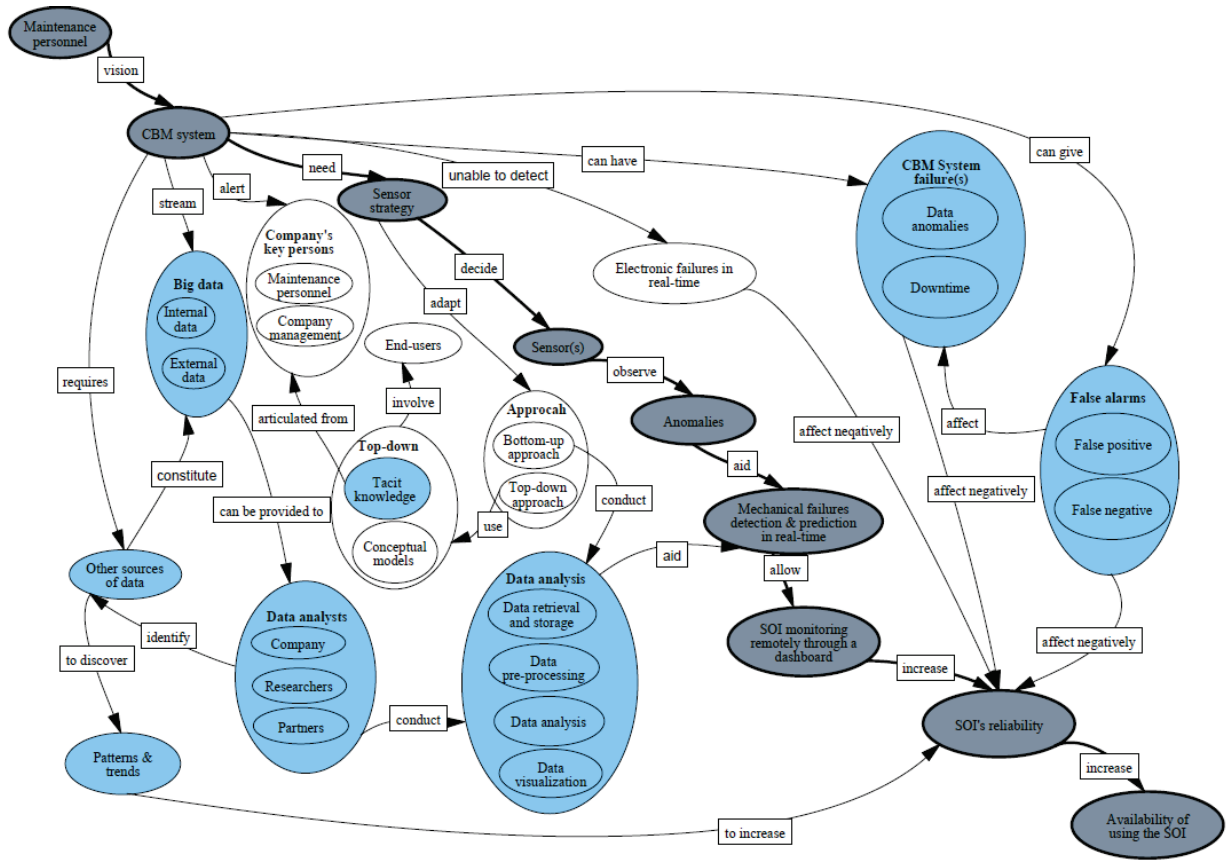

Figure 10. Systemigram from the maintenance personnel perspective.

product development as it is generated from the operation phase.

The tacit knowledge articulation and the failure data analysis can aid the development process, including manufacturing the SOI's parts. With tacit knowledge articulation and the data analysis results, the Company can identify the most critical parts and measurable critical parameters. The Company's decision-makers can use this identification to decide which sensors to install for which parts using a bottom-up sensor strategy.

The researchers can analyze and participate in articulating the maintenance personnel's tacit knowledge. The Company provides the data to the researchers and aids the interaction between researchers and maintenance personnel. The researchers identify the needed external data to investigate patterns and correlations between external and internal data. Internal data includes failure data and sensor or in-system data.

The internal and external data constitute the (big) data for the Company's case study. Analyzing the data enhances decision-making, including maintenance and development processes. This enhancement is moving toward more data-driven decision-making instead of a gut feeling. This data-driven methodology reduces maintenance costs and enhances the SOI's reliability, which maximizes profit.

Figure 10 visualizes the systemigram from the maintenance personnel perspective. Like the first systemigram, i.e., Figure 10, we categorized the second one into two categories. The first category, which has a dark grey color, represents the mainstay, while the other category has a light blue color that refers to the data in the Company's case study.

The Company's key person's vision, particularly maintenance personnel, is to have a Condition-Based Maintenance (CBM) system that can alert the personnel when the system detects failures. In addition, the CBM system should show the health condition in real-time of the SOI and its part through monitors. We can read the mainstay of the second systemigram as follows: *"Maintenance personnel vision CBM system. CBM system needs a sensor strategy that decides on the sensor(s) that observe anomalies, which aid mechanical failure detection and prediction in real-time that further allows SOI monitoring remotely through a dashboard, which increases the SOI's reliability and further increases the availability of using the SOI."*

The CBM system, which Company can invest in developing toward a digital twin, alerts the Company's key person, i.e., particularly the maintenance personnel. The strategy for the CBM system includes both top-down and bottom-up approaches. The bottom-up approach involves conducting data analysis, including failure data, sensor data, and other external data, such as environmental data.

The top-down approach consists of conceptual models and tacit knowledge articulation of the Company's key persons and end-user involvement. The end-user involvement gives the needed feedback in terms of notification when the end-users hear something not expected within the SOI operation, such as a sound of parts that need lubrication or are on their way to fatigue. This notification end-users can notify the Company in many ways, such as an app, picture, mail, or phone call. The Company, researchers, and partners can conduct conceptual models in many ways, such as using SSM methodology and ST tools such as systemigram, as we attempt in this research. In addition to other models, such as stakeholder interest map and workflow analysis using swimming lanes, we present the workflow analysis later (ref. Section IV, C).

In other words, the CBM strategy is an act of balance between the top-down and bottom-up approaches. This strategy aids in deciding which sensors to install for which parts to develop the CBM system. These sensors observe data point anomalies that assist in detecting and further predicting mechanical failure in real-time. The CBM system requires other data sources to verify and supplement the sensor's anomalies. These other data sources include environmental data, such as weather and traffic density.

The CBM system is also streaming data that constitute (big) data for the Company together with the other sources of data. The data analysts

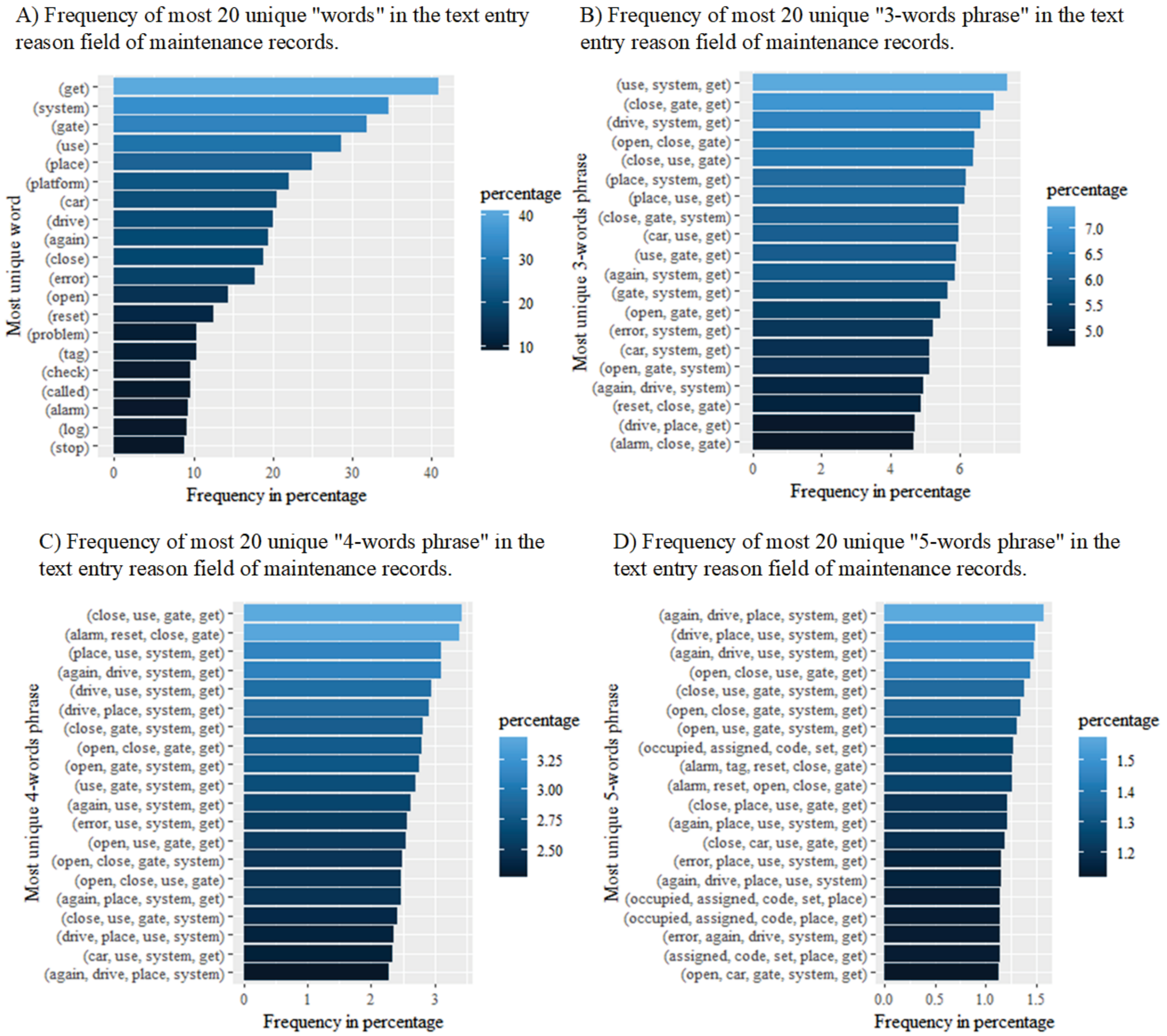

Figure 11. Most repeated unique words.

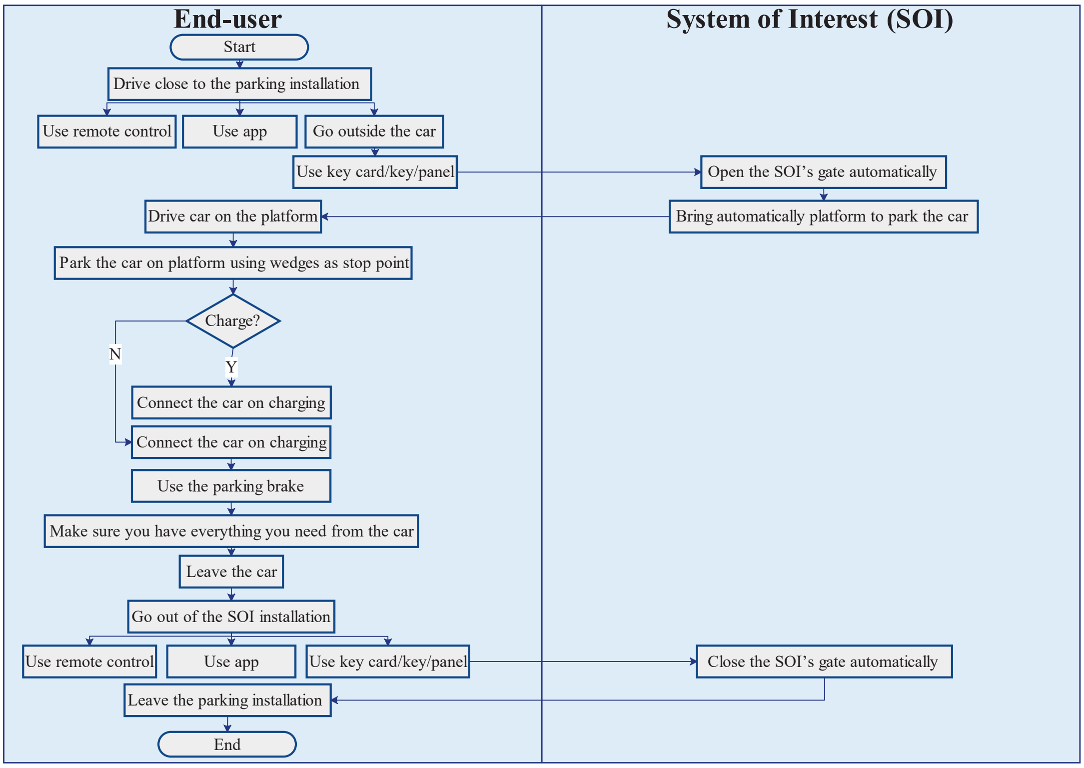

Figure 12. Workflow analysis uding swimming lanes for car entry.

include: the Company, researchers, and partners. The data analysis can identify the needed other data sources. Other data sources aid in discovering patterns and trends that increase SOI reliability and availability.

However, the CBM system cannot detect electronic failures like the Programmable Logic Controller (PLC). The CBM system can also give false alarms in terms of false positives and false negatives. False positive means the system provides an alarm or notification where there is no failure. On the other hand, a false negative means the system gives no

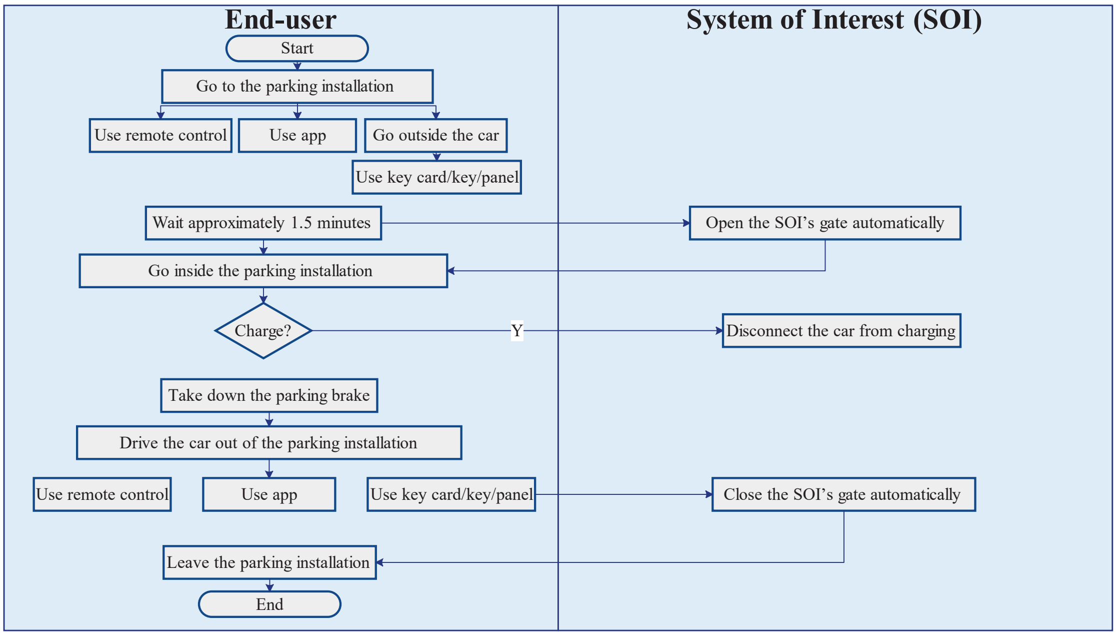

Figure 13. Workflow analysis using swimming lanes for car retrieval.

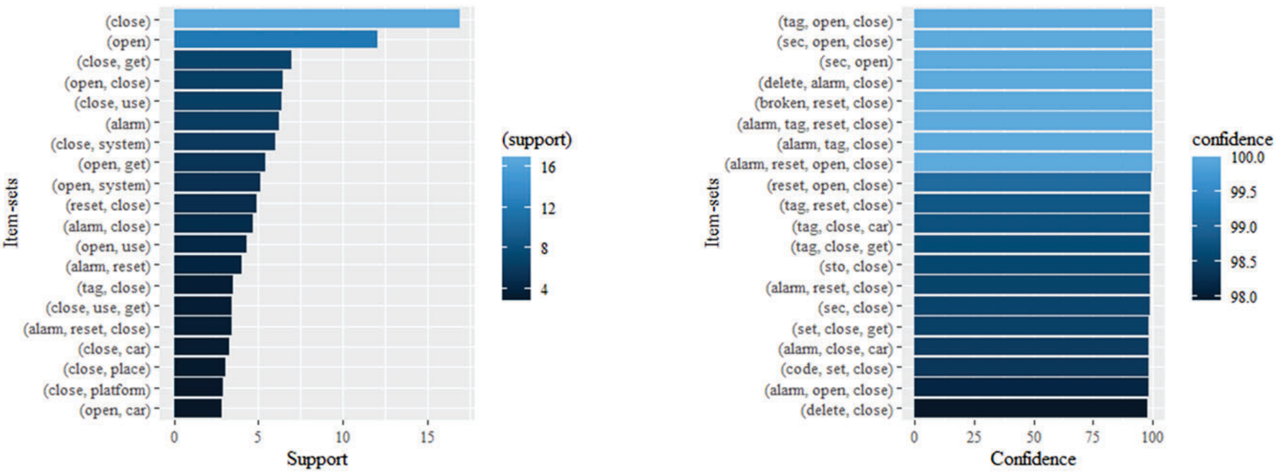

Figure 14. Most 20 unique frequent item-sets that appear with the gate in the failure events with its support and confidence values.

warning when there is a failure. These false alarms affect the CBM system in terms of failures. These failures include data anomalies because of false alarms. In addition, the CBM system failures include the system's downtime. The CBM system failures negatively affect the SOI's reliability.

5) Systems Thinking Possible Leverage Points

Implementing ST and its tools facilitates communication, understanding, and decision-making regarding a case study and its aspects. This case study is part of a complex socio-technical research project. We used the systemigrams to communicate with key stakeholders in academia and industry. The foundation for the systemigram is ST and other tools: stakeholder interest map, CATWOE analysis, and context diagram.

We developed two main systemigrams to visualize the two perspectives of the two main key stakeholders for this case study, i.e., Company management and maintenance personnel. The systemigrams facilitate decision-making regarding the development of a CBM system by visualizing all the significant aspects of this development. The systemigrams also visualized which approach suggested to the Company should adopt for the CBM's sensor strategy.

This approach is an act of balance between the bottom-up and top-down approaches. The top-down approach is tacit knowledge articulation and the use of conceptual models. The articulation involves the Company's key persons and end-users, while conceptual models include ST and its tools and other tools. The bottom-up approach uses data and data analysis.

C. Failure Data Analysis Results

We used the output of the NLP results as input for the association rule mining (i.e., FBGL), as mentioned above in the research methodology section. Figure 11 shows the results of using this mining. We used item-sets as one-word, three-word, four-word, and five-word phrases. Figure 11 depicts the 20 most unique frequent item-sets. Unique, in this context, refers to removing the duplication among failure events, i.e., if one item-set is repeated in one failure event (row), we remove its duplication.

Figure 11 shows that the most repeated unique words include: get, system, gate, use, place, platform, car, and so forth. The most frequent unique 3-word phrase, also called trigram, include: "use, system, get," "close, gate, get," "drive, system, get," and so forth. The most frequent unique 4-word phrases, also called 4-grams, include: "close, use, gate, get", "alarm, reset, close, gate", "place, use, system, gate". The most frequent unique 5-words, also called 5-grams, include: "again, drive, place, system, get", "drive, place, use, system, get", "again, drive, use, system, get", and so forth. We stopped the association rule mining using FBGL, also called n-gram analysis, as we got frequency results within one percent when we conducted the 5-gram analysis. We believe that the 3-gram analysis gives the needed information in this context.

This observation leads us to believe that the gate is the most critical subsystem within the SOI (i.e., the APS). Through interviews, workshops, and observations, we noticed that the maintenance personnel also used the system to mean a gate or segment or the whole system (i.e., APS). A segment is a collection of three gates, whereas the APS includes several segments depending on the building's architecture. Thus, even though the system comes before the gate, we conclude that the most critical subsystem is the gate. The next critical subsystem is the platform.

We also observe from Figure 11 that the open and close gates are the most frequent item-sets. In other words, "open" and "close" are the gate's most critical functions. In this context, we conducted a workflow analysis focusing on these gates' functions to investigate the end-users and the system's responsibility regarding these functions. These workflows for care entry and retrieval can be found in this publication [55]. We also developed the workflow analysis using swimming lanes to show the end-users and SOI responsibility more clearly regarding the gate's functions to open and close the gate with more details. Figure 12 and Figure 13 show these workflow analyses for car entry and retrieval, respectively. Further, we also notice that alarm, reset, and remote control are among the most repeated item sets or words.

Moreover, we dug deeper into the results showing the most frequent words or item-sets that come together with the "gate." Figure 14 portrays the 20 most unique frequent item-sets that appear with the gate in the failure events with its support and confidence values.

Support shows the percentage of occurrence of the item-sets, where the confidence indicates the percentage of the amount of time a given rule (if-then statement) is true among the dataset, i.e., failure events. This rule has two parts: an antecedent (if) and a consequent (then). The antecedent is the item-set we found among the datasets. The consequent is the item-sets found in combination with the antecedent [64].

Figure 14 (left) shows the support of antecedent item-sets, i.e., words in the y-axis where we determined the gate as consequent. We observe from Figure 14 (left) that if item-sets such as "close," "open," and "close, get," then we get the "gate" as a failure event, i.e., consequent.

Figure 14 (right) illustrates the confidence of the gate as a consequent. We notice that most item-sets, i.e., antecedents that co-occur with the gate as a consequent, include "tag, open, close", "sec, open, close", "sec, open", "delete, alarm, close", and "broken, reset, close."

The first antecedent, "tag, open, close," leads us to assume that issues with the tag the end-users use to open or close the gate trigger a gate failure event. The successive two antecedents, i.e., "sec, open, close," and "sec, open," lead us to assume that when the end-users use more time to open or close the gate in terms of sec, we get a failure event related to the gate. For the last two antecedents, i.e., "delete, alarm, close" and "broken, reset, close," let us assume that a broken signal to close the gate forces reset the whole system to close it again. In addition to one more assumption, an alarm must be deleted to close the gate. However, we can develop similar assumptions for more antecedents appearing in Figure 14 (right), but we believe we have mentioned the most significant ones.

1) Data Clustering

We used unsupervised machine learning to cluster the reason column in the failure data into three topics. We used the Gensim Python library for this topic modeling [65]. We got three topics with the most frequent words. The three topics included the following tokens (words) percentage: topic 1, 55% of the words, topic 2, 28.4%, and topic 3, 16.1% of the tokens (words). From the most 30 terms (words) included in the three topics, we assume that topic 1 indicates software issues. In contrast, topics 2 and 3 indicate human-machine issues (end-user failures) and mechanical issues (hardware failures), respectively. Figure 15 shows these clusters with their percentages.

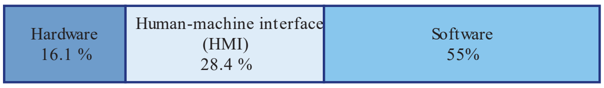

Figure 15. Failure data clustering into three clusters with their values.

D. Case Study Leverage Points

This subsection suggests recommendations for the Company based on applying the top-down and bottom-up approaches we implement in this case study. The top-down approach applies conceptual modeling via ST tools and other tools, whereas the bottom-up approach conducts failure data analysis. We visualize these recommendations in Figure 16.

The data analysis aids the decision-makers in the Company in reducing gut feelings. For instance, when we interviewed the maintenance personnel within the Company, different thoughts or beliefs were expressed regarding what is failing most, to which extent, and what causes the failure, e.g., users or the SOI. The data analysis answered such questions where the decision makers can base their decision on data-driven methodology reducing gut feelings. This feedback we got when we presented the analysis results in the Company

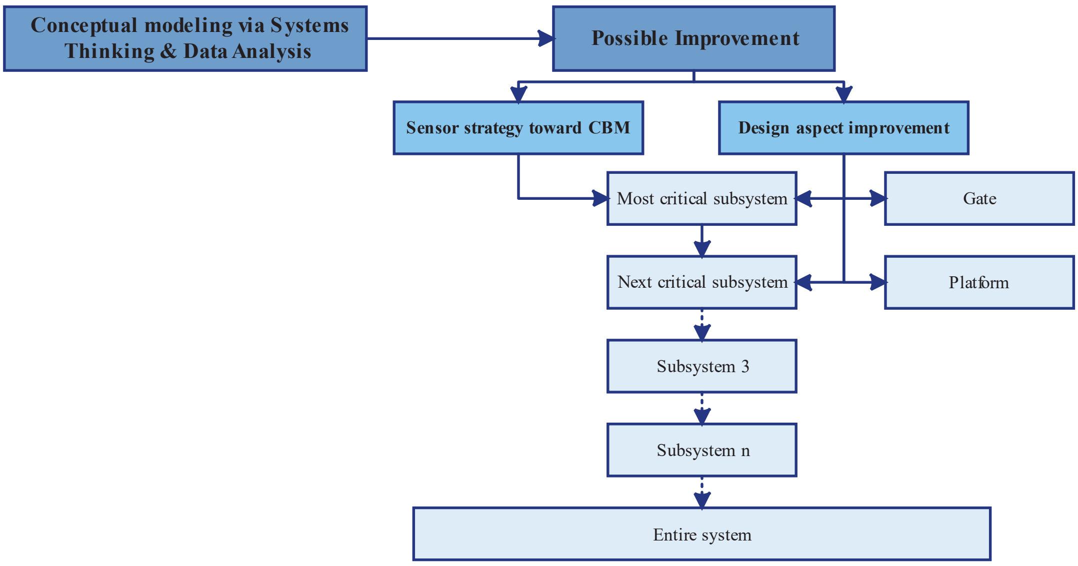

Figure 16. Case study leverage points.

during workshops. These workshops included Company management, maintenance personnel, and project leaders.

On the other hand, ST and other tools guide the data analysis. In addition, it facilitates communicating and understanding the problem domain, solution domain, and data analysis results. In this context, we developed systemigrams based on a stakeholder interest map, context diagram, and CATWOE analysis. Furthermore, ST triggers the development of other conceptual models, such as workflow analysis using swimming lanes, to understand and communicate the most failing critical part, which is the gate failures in this context, their natures, and their causes.

Based on applying both approaches, i.e., data analysis and ST, we suggest that the Company adopt a bottom-up sensor strategy to develop its vision regarding the CBM system and further digital twins. This bottom-up sensor strategy starts with the most critical subsystem, the next subsystem, and so forth. Data analysis aids in deciding the most critical failing subsystems instead of gut feelings. The failure data analysis results can also provide suggestions for design improvement. These improvements resulted from forming an explanation and seeing correlations and patterns among the failure data analysis results. However, we are collecting and analyzing other data sources, particularly weather data, to investigate the environmental factors and investigate correlations between these factors and failure events. This collection and analysis are still a work in progress.

V. Discussion

This section is divided into three subsections to discuss the article's research questions: RQ1, RQ2, and RQ3. The discussion section ends with limitations and further studies subsection.

A. *Defining and Early Validation of a Case Study*

The first research question, RQ1, is: "*How can we well-define and early validate a case study, including its multiple units of analysis using Systems Thinking and its tools?"*. We developed two systemigrams from two perspectives. These perspectives include the Company management and maintenance personnel perspectives. We used ST and other tools as a foundation for these two systemigrams: stakeholder interest map, context diagram, and CATWOE analysis.

We believe that these two systemigrams facilitate defining the case study well and aid in validating the case study early in a complex socio-technical research project. Defining the case study well means we have an overview and include all aspects of the case study, also called multiple units of analysis. Early validation refers to the industry (Company) and researchers (academia) validating the case study. We conducted this validation through workshops and interviews. The feedback indicated that applying ST and its tools enhanced the understanding, communication, and decision-making in the case study and its multiple units of analysis. The decision-making consists of prioritizing the most significant unit of analysis for the case study among industry (Company) and academia (researchers). In this study, we prioritized failure data analysis to support ST implementations and discover hidden values among these failure datasets. The hidden value aided the Company towards data-driven decisions that aided in understanding more about the size and reasons for the failure events.

B. *Tacit knowledge articulation*

The second research question, i.e., RQ2, is "*How can we articulate tacit knowledge?"*. We used ST and SE methodology to articulate the tacit knowledge from the Company's key persons. We used systemigrams based on ST and other tools, such as stakeholder analysis, context diagram, and CATWOE analysis.

We also used a canvas in the form of A3s and post-its in the virtual and physical workshops to articulate the Company's key persons' tacit knowledge. We used SE methodology to construct those A3 and post-its to ensure that all participants participated despite their personality, i.e., introvert or extrovert. However, we need to investigate further the effectiveness of the tacit knowledge articulation using the canvas in the form of A3s and post-its through a virtual platform.

This tacit knowledge articulation refers to the externalization mode in Nonaka and Takeuchi's model, which transfers tacit to explicit knowledge. This transformation contains an explanation of practices and beliefs. On the other hand, we observed that data analysis aids in creating knowledge, referred to as a combination mode by Nonaka and Takeuchi model [51]. The combination mode transfers explicit to explicit knowledge through restructuring the items already captured using deduction or induction.

C. *Support Systems Thinking with data analysis*

The third research question, RQ3, is "*How can we support Systems Thinking with data analysis?"* We collected and analyzed the Company's failure data to support the ST implementation for the Company's case study. ST and other tools, especially the systemigram, use a soft systems methodology (SSM). Thus these tools, including systemigrams, are considered conceptual models. In other words, ST and other tools cover the soft aspect, also called the top-down approach.

We complement the top-down approach with a bottom-up approach, also called the hard aspect. This aspect involves analyzing the failure data using machine learning. The feedback from the Company indicates that both approaches complement each other. Data analysis supports ST and reduces gut feeling. These gut feelings include the identification of the most critical subsystem that fails most, how often it's failing and why.

However, we need to iterate between these two approaches until the two approaches come into unity or until we accept the risk of moving to the next phase. We have started with data analysis and going back to the ST implementation in an iterative and recursive manner. This iterative and recursive manner aids in increasing the verification and validation of the case study's results in a complementary way. However, there is a need to investigate the number of details in both approaches, i.e., ST and data analysis. These details include investigating the number of nodes and links in the systemigrams. In addition, the number of figures showing the essential data analysis results and its way of visualization to ease communicating it to different stakeholders in industry and academia.

D. Research Limitations and Further Studies

One of the limitations of the presented paper is a lack of longitudinal research over multiple case studies. Any case study in complex socio-technical research would have several aspects that can affect the results. Therefore, we are starting to duplicate the same methodology with other industry partners, and the results seem promising.

Another limitation is analyzing additional data sources, such as sensor data from the Company for the same period of the failure data. Collecting these data when we conducted the case study was not technically possible. However, we are still in the process of further investigating the collection of these data with a third party. That data can affect the analysis and results and can be used for further research in this case study.

VI. Conclusion

Defining and validating a case study well and early at the beginning of a complex socio-technical research project is essential for the research project's success. This success ensures the Company's active participation and sharing of all needed data for the research. In this study, we used ST and its tools, including systemigrams, to define and validate the case study early. We used a stakeholder interest map, context diagram, and CATWOE analysis as a foundation for the systemigrams. We developed two systemigrams for the two main stakeholders for the case study, i.e., Company management and maintenance personnel. The Company's feedback indicates that using the systemigrams as conceptual models facilitates communication, understanding, and decision-making regarding the case study and its multiple units of analysis.

Moreover, using the systemigrams also aid in tacit knowledge articulation in terms of data and visualization. In addition, we used canvas in the form of an A3 and post-its for this articulation. We applied a SE methodology to construct this canvas. This articulation refers to the externalization mode in the Nonaka and Takeuchi model. According to the model, we refer to the data analysis as a combination mode. This mode creates knowledge through deduction or induction of already captured knowledge, which is the failure data analysis in this research.

We support ST and the implementation of its tools with data analysis. We collected and analyzed failure data using machine learning. We applied machine learning using the Frequent Pattern Growth Algorithm (FBGL) for association rule mining and the Gensim model to cluster the data. We considered the data analysis implementation the bottom-up approach, also called the hard aspect. Data analysis also reduces gut feelings and increases more data-driven early decisions. This data-driven methodology includes showing the failures, their size, and their causes.

In contrast, ST and applying its and other tools is the top-down approach or soft aspect. ST and other tools guide the data analysis. This guidance includes which strategy to adopt towards digitalization. The digitalization in this study refers to Condition-Base Maintenance (CBM) towards digital twins. Further, Systems Thinking aids in developing conceptual models to understand more and communicate the failures, their size, and their cause based on the data analysis results. In this study, we developed workflow analysis using swimming lanes for the most critical failing part, which is the gate, in this context, where data clustering shows these failures' natures, which are software, mechanical, and Human Machine Interface (HMI) issues. We conducted both approaches iteratively and recursively. The case study's results show that both approaches complement each other.

This study covers the lack of empirical research applying the two disciplines, i.e., ST and data analysis. ST provides the synthesis by investigating the case study, its aspects as a whole, and its relations among them. The data analysis goes in depth through reductionism, breaking the case study into more details. These two disciplines guide and support each other. Further, this research addresses tacit knowledge articulation in data and visualization using ST and SE methodologies in addition to the data analysis.

In further research, we plan to conduct longitudinal research applying the same methodology for multiple case studies. Furthermore, we plan to analyze additional data sources, such as in-system (sensor) data.

Acknowledgment

This research is part of a larger research project, the second iteration of the Human Systems Engineering Innovation Framework (HSEIF-2), funded by The Research Council of Norway (Project number 317862).

References

[1] H. Ali, M. Mansouri, and G. Muller, "Applying Systems Thinking for Early Validation of a Case Study Definition: An Automated Parking System," in Applying Systems Thinking for Early Validation of a Case Study Definition: An Automated Parking System, 2022. doi: https://www.thinkmind.org/index.php?view=article&articleid=modern_systems_2022_1_50_10020, https://doi.org/https://www.thinkmind.org/index.php?view=article&articleid=modern_systems_2022_1_50_10020.

[2] B. Sauser, M. Mansouri, and M. Omer, "Using systemigrams in problem definition: A case study in maritime resilience for homeland security," J. Homeland Security & Emergency Mang., vol. 8, no. 1, 2011.

[3] B. Richmond, "System dynamics/systems thinking: Let's just get on with it," System Dynamics Review, vol. 10, no. 2–3, pp. 135–157, 1994.

[4] "Systems Thinking | Systems Thinking Introduction," Si Network. https://www.systemsinnovation.network/courses/7297569 (accessed Dec. 05, 2022).

[5] W. Jaeger, "Aristotle's Metaphysics - Aristotle's Metaphysics. A revised text with introduction and commentary by W. D. Ross. Two vols. Oxford: Clarendon Press, 1924. Cloth, 48s. net," The Classical review, vol. 39, no. 7–8, pp. 176–180, 1925, doi: 10.1017/S0009840X00036660.

[6] "Systems Engineering Definition," INCOSE. https://www.incose.org/about-systems-engineering/system-

and-se-definition/systems-engineering-definition (accessed Dec. 07, 2022).

[7] "H-SEIF 2," Universitetet i Sørøst-Norge, Nov. 25, 2020. https://www.usn.no/english/research/our-research/technology/norwegian-industrial-systems-engineering-research-group/h-seif-2/ (accessed Aug. 15, 2021).

[8] M. Kjørstad, "Exploration and early validation in Systems Engineering: A study on combining systems and design practices in systems development towards innovations in Norwegian high-tech industries." University of South-Eastern Norway, 2022. Accessed: Dec. 05, 2022. [Online]. Available: https://hdl.handle.net/11250/2984753

[9] H. B. Ali, "Unlocking the power of big data within the early design phase of the new product development process: A case study in a Company," in INCOSE International Symposium, 2021. [Online]. Available: https://www.gaudisite.nl/INCOSE2021_AliEtAL.pdf

[10] M. Nourinejad, S. Bahrami, and M. J. Roorda, "Designing parking facilities for autonomous vehicles," Trans. Research Part B: Methodological, vol. 109, pp. 110–127, 2018.

[11] G. Cudney, "Parking Today | Articles - Automated Parking: Is It Right for You?" Available: https://www.parkingtoday.com/articledetails.php?id=181&t=automated-parking-is-it-right-for-you (accessed Jan. 21, 2022).

[12] F. Robles, "Road to Robotic Parking Is Littered With Faulty Projects," The New York Times, Nov. 28, 2015. Accessed: Jan. 21, 2022. [Online]. Available: https://www.nytimes.com/2015/11/28/us/road-to-robotic-parking-islittered-with-faulty-projects.html

[13] A. Mathijssen and A. J. Pretorius, "Verified design of an automated parking garage," in International Workshop on Parallel and Distributed Methods in Verification, 2006, pp. 165–180.

[14] C. Stenström, M. Al-Jumaili, and A. Parida, "Natural language processing of maintenance records data," Int. J. COMADEM, vol. 18, no. 2, pp. 33–37, 2015.

[15] "Defining and Clarifying Reliability." https://www.reliableplant.com/Read/98/defining-reliability (accessed Aug. 15, 2021).

[16] G. Muller, "Applying Roadmapping and Conceptual Modeling to the Energy Transition: A Local Case Study," Sustainability, vol. 13, no. 7, p. 3683, 2021.

[17] P. Checkland, "Systems thinking," Rethinking management information systems, pp. 45–56, 1999.

[18] J. A. B. Montevechi and J. D. Friend, "Using a soft systems methodology framework to guide the conceptual modeling process in discrete event simulation," in Proc. of the 2012 Winter Simulation Conf. (WSC), pp. 1–12.

[19] "Why systems thinkers and data scientists should work together to solve social challenges," nesta. https://www.nesta.org.uk/blog/why-systems-thinkers-and-data-scientists-should-work-together-solve-social-challenges/ (accessed Oct. 08, 2022).

[20] N. Shin et al., "A framework for supporting systems thinking and computational thinking through constructing models," Instructional Science, pp. 1–28, 2022.

[21] R. D. Arnold and J. P. Wade, "A definition of systems thinking: A systems approach," Procedia computer science, vol. 44, pp. 669–678, 2015.

[22] J. Boardman, B. Sauser, L. John, and R. Edson, "The conceptagon: A framework for systems thinking and systems practice," in 2009 IEEE International Conference on Systems, Man and Cybernetics, 2009, pp. 3299–3304.

[23] J. Gharajedaghi, Systems thinking: Managing chaos and complexity: A platform for designing business architecture. Elsevier, 2011.

[24] A. Basden and A. T. Wood-Harper, "A philosophical discussion of the root definition in soft systems thinking: an enrichment of CATWOE," Systems Research and Behavioral Science, vol. 23, no. 1, pp. 61–87, 2006, doi: 10.1002/sres.689.

[25] S. Armenia and F. Loia, "Integrating Big Data Analytics, Systems Thinking and Viable Systems Approach Towards a Shift from Individual to Collective Intelligence and Collective Knowledge Systems," puntOorg International Journal, vol. 7, no. 1, pp. 62–83, 2022.

[26] L. Fahey and L. Prusak, "The eleven deadliest sins of knowledge management," California management review, vol. 40, no. 3, pp. 265–276, 1998.

[27] A. B. Nassoura, "Analysis of Attitudes and Barriers to Knowledge Management among Undergraduate Learners in Higher Learning Institutions," مؤتمرات الآداب والعلوم الانسانية والطبيعية, 2020.

[28] B. Unhelkar, Big Data Strategies for Agile Business. CRC Press, 2017.

[29] G. Jifa, "Data, information, knowledge, wisdom and meta-synthesis of wisdom-comment on wisdom global and wisdom cities," Procedia Computer Science, vol. 17, pp. 713–719, 2013.

[30] G. Bellinger, D. Castro, and A. Mills, "Data, information, knowledge, and wisdom." 2004.

[31] A. De Mauro, M. Greco, and M. Grimaldi, "A formal definition of Big Data based on its essential features," Library review, 2016.

[32] O. Kwon and J. M. Sim, "Effects of data set features on the performances of classification algorithms," Expert Systems with Applications, vol. 40, no. 5, pp. 1847–1857, 2013.

[33] P. Russom, "Big data analytics," TDWI best practices report, fourth quarter, vol. 19, no. 4, pp. 1–34, 2011.

[34] D. Laney, "3D data management: Controlling data volume, velocity and variety," META group research note, vol. 6, no. 70, p. 1, 2001.

[35] J. Gantz and D. Reinsel, "The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east," IDC iView: IDC Analyze the future, vol. 2007, no. 2012, pp. 1–16, 2012.

[36] J.-P. Dijcks, "Oracle: Big data for the enterprise," Oracle white paper, vol. 16, 2012.

[37] S. Gogia et al., "The Big Deal About Big Data For Customer Engagement: Business Leaders Must Lead Big Data Initiatives To Derive Value [verkkodokumentti].[Viitattu 14.10. 2016] Saatavilla https://www. forrester. com/report/The+ Big+ Deal+ About+ Big+ Data+ For+ Customer+ Engagement/-," E-RES72241, 2012.

[38] M. White, "Digital workplaces: Vision and reality," Business information review, vol. 29, no. 4, pp. 205–214, 2012.

[39] "Big Data Definition - MIKE2.0, the open source methodology for Information Development." Mike2., 2021. Accessed: Nov. 14, 2021. [Online]. Available: http://mike2.openmethodology.org/wiki/Big_Data_Definition

[40] B. D. A. Intel, "Intel's IT Manager Survey on How Organizations Are Using Big Data," Intel report, 2012.

[41] M. A. Beyer and D. Laney, "The importance of'big data': A definition. Stamford, CT: Gartner. Retrieved June 22, 2014." 2012.

[42] M. Schroeck, R. Shockley, J. Smart, D. Romero-Morales, and P. Tufano, "Analytics: The real-world use of big data," IBM Global Business Services, vol. 12, no. 2012, pp. 1–20, 2012.

[43] NBD-PWG NIST, "DRAFT NIST Big Data Interoperability Framework: Volume 3, Use Cases and General Requirements," NIST Special Publication, vol. 1500, p. 3, 2014.

[44] Microsoft Research, "The Big Bang: How the Big Data explosion is changing the world," The Microsoft Enterprise Insight Blog.[Internet], 2013.

[45] D. Boyd and K. Crawford, "Critical Questions for Big Data in Information,,,Communication & Society"," Communication and Society, vol. 15, no. 5, pp. 662–679, 2012.

[46] G. Klein, J. K. Phillips, E. L. Rall, and D. A. Peluso, "A data–frame theory of sensemaking," in Expertise out of context, Psychology Press, 2007, pp. 118–160.

[47] L. Koesten, K. Gregory, P. Groth, and E. Simperl, "Talking datasets–understanding data sensemaking behaviours," International journal of human-computer studies, vol. 146, p. 102562, 2021.

[48] K. E. Weick, Sensemaking in organizations, vol. 3. Sage, 1995.

[49] H. Ali and F. A. Salim, "Transferring Tacit Knowledge into Explicit: A Case Study in a Fully (Semi) Automated Parking Garage," 2021, doi: https://www.sdpsnet.org/sdps/documents/sdps-2021/SDPS%202021%20Proceedings.pdf, https://doi.org/https://www.sdpsnet.org/sdps/documents/sdps-2021/SDPS%202021%20Proceedings.pdf.

[50] H. Collins, Tacit and Explicit Knowledge. Chicago: University of Chicago Press, 2010.

[51] I. Nonaka, "A Dynamic Theory of Organizational Knowledge Creation," Organization science (Providence, R.I.), vol. 5, no. 1, pp. 14–37, 1994, doi: 10.1287/orsc.5.1.14.

[52] M. Polanyi, The tacit dimension. Glouchester, Mass: Peter Smith, 1983.

[53] M. Alavi and D. E. Leidner, "Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues," MIS quarterly, vol. 25, no. 1, pp. 107–136, 2001, doi: 10.2307/3250961.

[54] W. Chergui, S. Zidat, and F. Marir, "An approach to the acquisition of tacit knowledge based on an ontological model," Journal of King Saud University-Computer and Information Sciences, vol. 32, no. 7, pp. 818–828, 2020.

[55] H. B. Ali, G. Muller, and Salim, Fahim, "Applying Conceptual Modeling and Failure Data Analysis for 'Actual Need' Exploration," in MODERN SYSTEMS 2022: International Conference of Modern Systems Engineering Solutions.

[56] R. K. Yin, Applications of case study research. sage, 2011.

[57] C. Potts, "Software-engineering research revisited," IEEE software, vol. 10, no. 5, pp. 19–28, 1993.

[58] M. Duneier and O. Carter, Sidewalk. Macmillan, 1999.

[59] J. Boardman and B. Sauser, Systems Thinking: Coping with 21st Century Problems. CRC Press, 2008.

[60] G. Muller, "1.1. 1 Industry and Academia: Why Practitioners and Researchers are Disconnected," in INCOSE International Symposium, 2005, vol. 15, no. 1, pp. 1–9.

[61] Miro, "The Visual Collaboration Platform for Every Team | Miro," https://miro.com/. https://miro.com/ (accessed Nov. 29, 2021).

[62] T. Langen, K. Falk, and M. Mansouri, "A Systems Thinking Approach to Data-Driven Product Development," Proceedings of the Design Society, vol. 2, pp. 1915–1924, 2022, doi: 10.1017/pds.2022.194.

[63] H. A. Research, "What Is an Automatic Parking System?," Car and Driver, Apr. 13, 2020. https://www.caranddriver.com/research/a31995865/automatic-parking-systems/ (accessed May 15, 2022).

[64] "What are Association Rules in Data Mining (Association Rule Mining)?," SearchBusinessAnalytics. https://www.techtarget.com/searchbusinessanalytics/definition/association-rules-in-data-mining (accessed Oct. 04, 2022).

[65] "Gensim: topic modelling for humans." https://radimrehurek.com/gensim/intro.html (accessed Sep. 17, 2022).

[66] M. Mittal, S. Pareek, and R. Agarwal, "Efficient ordering policy for imperfect quality items using association rule mining," in Encyclopedia of Information Science and Technology, Third Edition, IGI Global, 2015, pp. 773–786.

[67] J. Korstanje, "The FP Growth algorithm," Medium, Sep. 28, 2021. https://towardsdatascience.com/the-fp-growth-algorithm-1ffa20e839b8 (accessed Oct. 04, 2022).

[68] C. Borgelt, "An Implementation of the FP-growth Algorithm," in Proceedings of the 1st international workshop on open source data mining: frequent pattern mining implementations, 2005, pp. 1–5.

[69] "FP Growth Algorithm in Data Mining - Javatpoint," www.javatpoint.com. https://www.javatpoint.com/fp-growth-algorithm-in-data-mining (accessed Oct. 03, 2022).

[70] "Apache Spark - ArchWiki." https://wiki.archlinux.org/title/Apache_Spark (accessed Oct. 03, 2022).

APPENDIX A

Appendix A describes the data research methodology in detail. We collected failure data from the Company. The Company currently uses excel sheets to maintain logs about the failure events occurring in each semi-automated parking garage. Each excel sheets belong to a single installation. The Company installs semi-automated parking garages primarily for private buildings with fast-trained users.

The service-log data, also called maintenance records data, includes different parameters (columns) including, but not limited to: Date (for a maintenance event), time, telephone (for the maintenance personnel who investigated the failure event), place number (for which parking lot the failure event occurred), reason (possible reasons for the failure event), re-invoiced yes/no (if the failure event re-invoiced as it is not included within the maintenance agreement with the Company or not). These parameters construct the columns within the Excel sheets.

A. Natural Language Processing

Figure 17 visualizes a flowchart for the Natural Language Processing (NLP) methodology we conducted for the maintenance record data we collected from the Company. We use an introductory NLP analysis [14]. The NLP analysis we performed for the reason parameter (column) includes a description of the failure events. This description is a free text that is manually logged by the maintenance personnel. The NLP analysis included the following steps:

- Import data. We load all the data, which are failure events. These failure events are saved as an Excel file. Further, we import the text of interest, which is the reason parameter (column) in our case study.

- Tokenization. In this step, we divided the sentences into words, commas, etc. This step includes a sub-step that removes the numbers that appear directly after a word.
- Removing stop words. Stop words commonly occur in language, such as prepositions, pronouns, etc. These words do not add significant meaning to the sentences. We removed these words in this step.
- Stemming & lemmatization. In this step, we reduced the words to their origin, i.e., stem base or root form, also called a lemma. For instance, "engineering" or "engineers," to its base word, "engineer."
- Download the output file. We saved the results from the former steps into a Comma-Separated Values (CSV) file. Then, we downloaded the file for the next step, i.e., association rule mining.

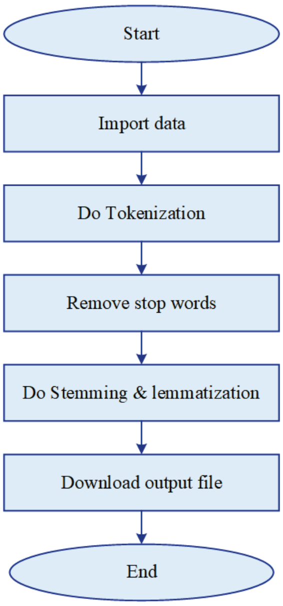

Figure 17. Flowchart for the Natural Language Processing analysis.

B. *Association rule mining*

Association rule mining involves machine learning models to analyze data. The association rule identifies frequent if-then associations, called association rules [64]. We used association rule mining to investigate patterns and co-occurrence among the words in the reason parameter, which is a text-free format [64][66]. We investigate these patterns by finding which words frequently co-occur after determining the most frequent word(s).

We used Frequent Pattern Growth Algorithm (FBGL). The foundation for this algorithm is the Apriori algorithm, where the FBGL is seen as Apriori's modern version as it is more efficient and faster, giving the same results [67] [68]. The FBGL uses an expanded prefix-tree structure called the Frequent-Pattern (FB) tree to store compressed and essential data about frequent patterns. It is effective and scalable for mining the entire set of frequent patterns by Fragment Pattern growth (FP-tree) [69].

Figure 18 illustrates the steps before, during, and after the association rule mining using FBGL [66]–[68]. The steps are following:

- Import data. We import and read data from the last step in the NLP analysis.
- Pre-process data. We pre-processed the data. The pre-processing included removing duplicate words. This duplication is the same words repeated in the same failure events. Each failure event is organized in one row in the Excel file.
- Cluster using Spark. We used an open-source cluster computing tool called Apache Spark [70]. This local clustering aims at using the association rule mining using FBGL faster, i.e., getting results in less time.
- Read each item. In this step, the FBGL reads each item. The item in this context refers to each word in the loaded file as input for this mining. We chose the item or item-set to be a one-word, 3-words phrase (trigram), 4-words phrase, and 5-words phrase.
- Counting the occurrences of each item. In this step, the algorithm determines the occurrence of each item or item-set, which is words in this context.
- Determine support for each item or item-set. The algorithm determines the support for each item. Support reveals the frequency, or percentage, of co-occurrence of the item-set [64][67]. We based our decision on the value we

called minimum support, which is 0.01 (1%) in this case study. In other words, the item sets shall co-occur at least 1% among the dataset, i.e., failure event data.

- Support >= minimum support decision gate. The algorithm has a decision gate to decide whether to include the items in the frequent item-set. If the items do not fulfill the minimum support, we remove them.
- Add items to frequent item-set. In this step, the FBGL includes the items to the frequent item-set if the support for these items is higher than or equal to the minimum support value.
- Determine confidence for each item-set. The FBGL determines the confidence value for the added item-set from the former step. Confidence reveals how frequently a rule is applied. The conditional probability of the right-hand side given the left-hand side is another way to put this. This rule can be, for instance, which words come to the right of a specific word. Another example is which items (words) come together among the dataset, i.e., failure event data. We set up the value for the confidence to be 0.6 (60%) [64][67].
- Confidence >= minimum confidence. The algorithm had another decision gate to decide to include the item-set. If the item-set does not fulfill the minimum value for the confidence, we remove it.
- Add item-set to the frequent item-sets. The FBGL adds the item-set that fulfills the minimum value for the confidence to the frequent item-set.
- Order the item-sets based on occurrences. The FBGL order the item-sets based on their occurrences.
- Create the tree for the item-sets. The FBGL creates the tree for the item-sets based on their ordered co-occurrences. Each item(s) (word(s)) is a node in the tree.
- Write results to a file. We added the results (item-set) from the former step to a CSV file.
- Visualize results. We visualized the results (items-set). This visualization includes all the selected items or item-sets, i.e., the most frequent unique word, most frequent unique 3-phrase words, 4-phrase words, and so forth.
- Translate results. After the visualization, we chose the most significant results based on feedback from the industry and academia practitioners and the Company's key persons. After this input, we translated the results from Norwegian to English. Ultimately, we visualized the most significant results we show in this study.

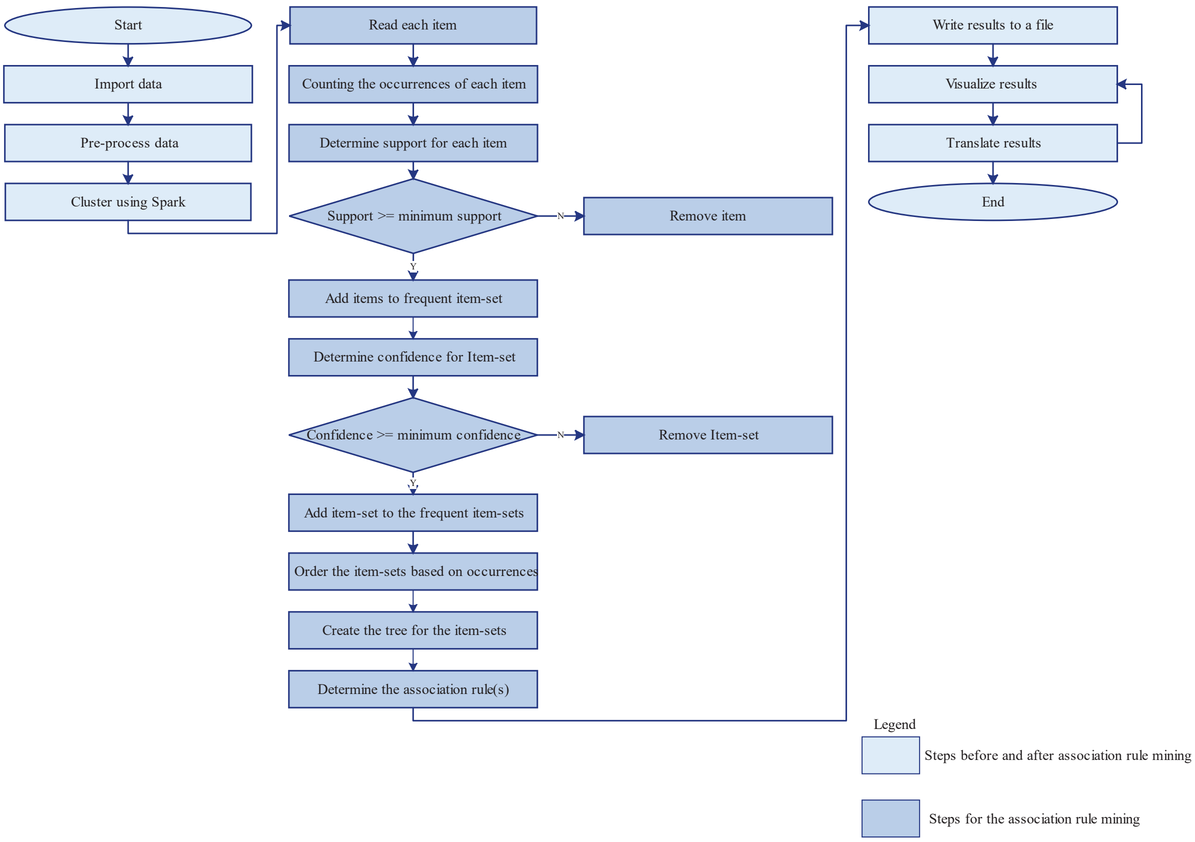

Figure 18. Flowchart showing the steps during, before, and after Association rule mining using Frequent Pattern Growth Algorithm (FBGL).