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

State of Affair in Terms of Big Data Utilization in Complex System Engineering Organizations: A Case Study in the Context of Norwegian Industry

Fahim A. Salim
Dept. of Science and Industry Systems
University of South-Eastern Norway
email: fahim.a.salim@usn.no

Haytham B. Ali
Dept. of Science and Industry Systems
University of South-Eastern Norway
email: haytham.ali@usn.no

Tommy Langen
Dept. of Science and Industry Systems
University of South-Eastern Norway
email: tommy.langen@usn.no

Pragna Labony Deb
Dept. of Science and Industry Systems
University of South-Eastern Norway
email: pragnalabony.sust@gmail.com

Andreas Wettre
Centre for Design Research
Oslo School of Architecture and Design
email: andreas.wettre@aho.no

Gerrit Muller
Dept. of Science and Industry Systems
University of South-Eastern Norway
email: gerrit.muller@gmail.com

Kristin Falk
Dept. of Science and Industry Systems
University of South-Eastern Norway
email: kristin.falk@usn.no

***Abstract*—Effective utilization of big data is still an open question for most organizations. In the presented case study, we attempted to get a nuanced understanding of the state of affairs regarding big data utilization in Norwegian high-tech industries. This case study uses research methods like questionnaires, semi-structured interviews, observations, and co-creation sessions. These methods explore the data utilization processes at partner organizations of the H-SEIF2 consortium or lack thereof to systematically utilize big data in their projects from the perspective of employee perception. The presented case study provided insights into the case study organizations. For example, organizations still heavily rely on inconsistent manual data logging and our survey found that the Project Managers have a more optimistic perception of their usage of big data. In contrast, upper management has a more modest opinion of their current state. The presented case study also provided a more in-depth analysis of challenges that hinders data utilization and identified opportunities to enhance the value of ongoing and potential digitalization initiatives at the organizations.**

***Index Terms*—Questionnaire; Big data; Early Phase Decisions.**

I. Introduction

The paper is an extended version of the article presented at the Modern Systems Conference 2022, aiming to get a nuanced understanding of big data usage in the context of Norwegian Industry [1].

Big data analytics (BDA) and digitalization is a trending topic and is expected to be a big asset and an engine for innovation that has the potential to propel new technological revolutions [2]. The benefits of using big data in enhancing operations in general and in project life cycle management are well understood by organizations of all types and sizes [3]–[6]. However, how to do so effectively is still an open questions for most organizations [4] [7] [8]. For example, a study by Qlik and Accenture states that over 74% of employees feel anxiety with when working with data [9]. Similarly, a recent study by Rackspace Technology [3] reported that organizations perceived a rise in difficulty in terms of utilizing Big Data Analytics (BDA) and specifically Artificial Intelligence (AI) and Machine Learning (ML).

Researchers have identified multiple challenges that limit organizations' ability to enhance big data utilization. These challenges include a lack of common language among engineers, ineffective knowledge sharing, and difficulty in finding system information, to name of few [10]–[13]. The presented paper focused on the users of big data, specifically the internal users such as employees working on projects.

The presented case study in the context of the Norwegian high-tech industry is part of the H-SEIF2 [14] project. The case study and the analyses are being performed in close collaboration with the H-SEIF2 consortium industry partners to provide a nuanced understanding of the state of affairs at Norwegian high-tech companies in terms of big data utilization. The case study is designed using a combination of techniques such as industry as a laboratory [15], Co-creation [16], questionnaires and semi-structured interviews [17]–[19].

The remainder of the paper is structured as follows. The following section (section II) describes the related work. We describe the design of the case study in section III. This is followed by results from analysis and observations (section

This research is part of the H-SEIF 2 project and is funded by the Research Council of Norway through the innovation project (IPN project no.: 317862).

IV) and concluding remarks (section V).

II. Related Work

Related work describes the notion of big data as described in systems engineering. It further discusses related surveys conducted by other researchers.

A. Big Data and Complex System Engineering

Regarding the notion of big data utilization in complex systems environments, there are many definitions for data, also called big data. Many authors and practitioners have defined big data as the notion of (Vs). Some authors [20]–[22] have defined big data in terms of the 3Vs: Volume, Velocity, and Variety. Others [23]–[25] have extended the definition by adding value as the fourth V (4Vs).

In addition, M. White [26] suggested adding Veracity as the fifth V (5Vs). In this context, "Volume" refers to the vast amount of data. "Velocity" refers to the speed at which new data is generated, whereas "Variety" represents different data types. The fourth V, "Value," refers to how we can benefit from big data by turning it into value. "Veracity" includes biases in the data and strives to encompass the level of sufficiency or insufficiency of the data [27].

The elements in our world are connected and dependent on each other. The complexity is higher than ever and is continuing to be more complex. Products are linked in an intertwined network of dependencies, and services rapidly develop to satisfy demanding customers. Fig. 1 shows an example of an intertwined network of dependencies.

It visualizes an elaboration of the characteristics of what Schätzet et al. call CPS (Cyber-Physical Systems) [28]. CPS is closely related to several concepts such as (Big) data and its analysis, the Internet of Things (IoT), Systems of Systems (SoS), Mechatronics, and Embedded Systems. We also add the sociotechnical aspects within the other three circles: Human Social Organizational, Innovation Ecosystems, and Political aspect.

Systems are also adapting in a dynamic behavior to technical and social factors. Thus, there is a need for companies to explore new ways to maintain competitiveness. It is crucial for these companies to use the most effective methods and that these are sufficiently founded. A solution is to utilize data early in product and service development.

Organizations and their employees recognize that big data analytics will be a significant source of competitive advantage in the future. Still, several impediments inhibit them from fully utilizing big data's benefits. Most technologies were not developed to satisfy the expanding demands of big data analytics [29]. Employees who want to exploit the power of big data may run into significant issues due to data complexity and inherent messiness. Moreover, digital data is often stored in various forms, such as unstructured databases and discrete different text files [30].

Lack of data analytics skills among current employees may increase data entry mistakes, resulting in misinterpretation and loss of important information and ultimately reducing the value of the data [31]. Ethical considerations like privacy and cultural barriers are also hurdles regarding big data usage. For example, some businesses know how big data might help them improve their operations. Still, cultural or technological limitations prevent employees from using big data in production.

B. Survey Questionnaires Regarding Big Data Utilization

Questionnaires and surveys are widely used for different purposes, e.g., comparing two products or services or both [8] [32]. They are often used to understand the needs of perspective users of future products and services, or as part of user-centric design [19] [17], or to collect data on customer, employee, or student satisfaction, [33]–[35].

Regarding the utilization of big data in the operations of organizations, Qlik and Accenture [9] conducted a survey to understand big data utilization in enterprises. They found that 60 to 70% of the collected data in an enterprise is never used and a vast majority of about 74% of employees feel overwhelmed or simply unhappy working with data. They also found that only 37% of employees trust their decisions more when they are based on data, while 48% preferred gut feeling over data-driven decision-making [9].

Focusing on high-level decision makers, a survey by Rackspace Technology [3] found that employees perceived difficulty with ML and big data has been increasing. The survey [3] also found that employees considered data dispersed across many different systems to be one of the most significant barriers to drawing insights from it. Lack of skillset and talented employees is perceived as a considerable concern and limitation in fully utilizing big data in enterprise decision making [3] [9] [36].

Raguseo [7] focused on the CIOs of French medium and large enterprises to understand differences across industries and the size of organizations. Analysis of the questionnaire found that the organization's size influences investments in technologies like ML software tools. The author did not find any statistical differences across different industries.

While employees often appear in the discourse and are considered a crucial element in utilizing big data systems, they are often the ones most neglected [37]. Moreover, most of the research mentioned above focused on high-level decision makers, not a cross-section of employees, departments, and job roles.

The presented paper is part of the H-SEIF2 [14] research project that aims to develop a human-centered framework for utilizing big data during early phase decision-making in the product development process. By focusing on Norwegian industry partners, the presented work extends the related work by focusing on a cross-section of employees to understand better the differences among different departments, employee profiles, and across various industries.

III. Case Study Design

The primary goal of the case study is to understand the current state of affairs in terms of big data utilization at the partner organizations in the H-SEIF2 [14] project. The H-SEIF

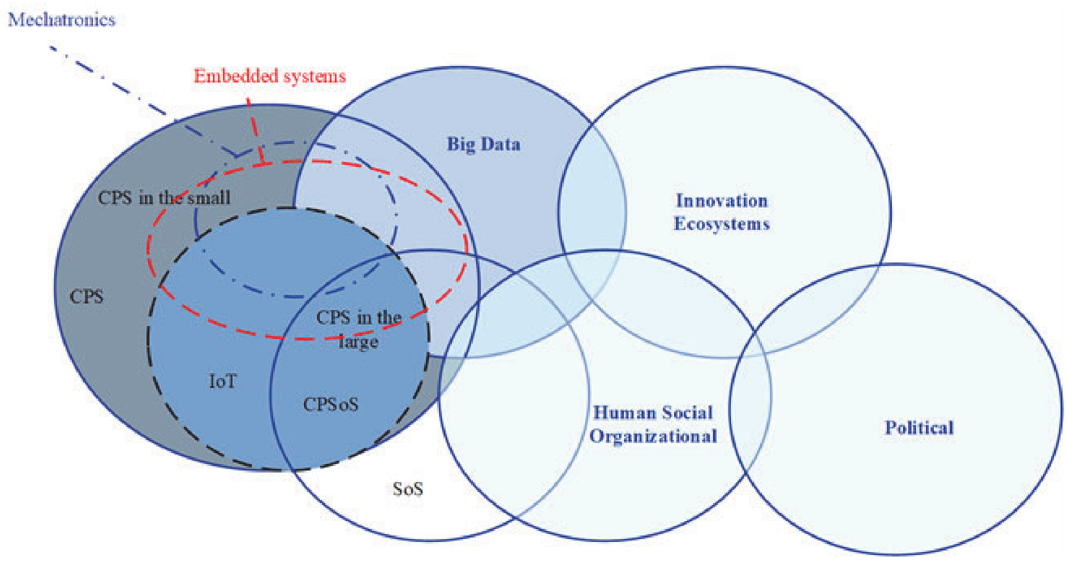

Fig. 1. Complex Interlinks

2 project aims to harvest the value of big- data to enhance the experience of stakeholders during complex system engineering projects by collaborating with industry partners to improve their digitalization efforts. The goal is to design data-driven frameworks and methodologies to allow the industry partners in data-supported early-phase product development decisions.

The presented case study, following the aims of the H-SEIF2 project of harvesting the value of big data for industry partners; evaluates the question: *How are the different industry partners utilizing big data in their operations*?

1) What are the gaps, opportunities, and barriers to enhancing the utilization?
2) Are there any differences in different organizations and departments within an organization, and what can the partners learn from each other's experiences?
3) What can partner organizations do to maximize the value of ongoing or potential digitization initiatives?

A. Methodology

We used a combination of workshops, interviews, surveys, on-site observations, and subject-matter expert feedback from the industry and academia to design the questions for the survey. An adapted version of the Applied Research Framework was used as the research method to design this survey [38], [39]. The framework consists of the following steps:

Step 1: Shape the line-of-reasoning. In this step, the line of reasoning is expressed by following the structure of *Problem-goal-solution-rationale*. Additionally, the research questions we formulated more specifically based on the broad problem statement we expressed within the line of reasoning. The main research question was establishing a baseline for the H-SEIF 2 research project consortium.

Step 2: Explore literature. In this step related studies from literature to aid in designing the questionnaire (see section II-B for details).

Step 3: Elicit expert opinions. Expert's views, in this context, refer to the domain expert among the scholars within the research methods. Semi-formal discussions with different experts were conducted. For example, two of the co-authors have decades of consultancy experience. Several workshops were conducted with experts from academia and industry to gather their feedback regarding good, best, and emergent practices regarding the survey's design and questions.

Step 4: Determine the research design. Notes were kept from the workshops and related literature using a shared platform. Furthermore, steps 2 and 3 were performed iteratively.

A total of 6 companies participated in the presented study.

In the first stage, a survey was conducted with a cross-section of employees at different industry partners. A total of 5 companies from the H-SEIF2 consortium participated in the survey. This included a technology consultancy, an autonomous transportation solutions provider, an oil and gas company, a Data Agency and an Industrial conglomerate (3 divisions/subsidiaries). Further, we conducted semi-structured interviews at the technology consultancy and co-creation and observation sessions with different partners (Automated Parking System (APS), the oil and gas company, and an industrial conglomerate's defense division) to better understand the state of affairs at the individual organization.

The partner organizations vary in size and they work in different sectors; in the second phase, we had more direct interactions with some of the individual partners to better understand the state of affairs at the company. We customized the approach per partner organization based on factors such as the size and time commitment of the organization, the type and duration of projects organizations are involved in, the industry sector to which the organization belongs to and the level of access granted to researchers by each organization. For example, some co-authors work closely with partner organizations, allowing them more opportunities to observe and hold co-creation sessions. While at some partners, semi-structured interviews were conducted as co-creation sessions were not feasible due to the limited availability of the involved personnel.

B. Survey Questionnaire

Based on the methodology described above (section III-A) the finalized version of the survey questionnaire consisted of 24 questions in 5 categories. We incrementally developed the survey's language, style, and structure with continuous feedback from participating subject-matter experts and practitioners from industry and academia.

We wanted to cover many aspects in our survey. We formed questions about the data **availability**; if the data is challenging to find, it is also difficult to utilize. This category also goes to data ownership, as external organizations own the data, and there might be obstacles to utilize.

Then if you have the data, the data needs to be **usable**, contextualized so that it is possible to transform to value and use the analyses in the decision-making process focusing on early phase product development.

We wanted to know about data integrity. Suppose users do not **trust** the data, thinking it is extracted from sources or through a process that reduces the reliability or presented in ways that make you doubt what the data is saying. In that case, it will most likely not be utilized.

Then even if the datasets are reasonable, there might be **processes, politics, habits, time** or a lack of competence that prevents the organization from using the data for decision making. Sometimes, the essential decisions are made on gut feelings, emotions or based on older experiences from similar projects.

We asked the participants to rate the extent to which they agreed or disagreed with the statements on a 5-point Likert scale. In the initial phase, we collected 40 responses from employees at partner organizations.

C. Semi-structure Interviews

The survey was followed by semi-structured interviews with employees working in the technology consultancy. The consultancy works on different project types with variable duration and activities they performed on those projects. For this research, we focused on the personnel working on one project the consultancy did for one of its clients.

The interviews focused on how the consultancy acquired insight about how they utilized big data in their previous and current projects considering the project team as an example. The in-depth interviews allowed us to achieve a broader understanding of the point of view of the engineers and managers at the consultancy, which facilitated a qualitative analysis [40]. We designed the semi-structured interviews for eliciting information about specific topics [41], [42]. The interview guide consisted of 20 main open-ended questions that reflected the stakeholder needs while using big data. Fig. 9 shows the interview guide.

We conducted all the interviews using Microsoft Teams, a teleconference tool widely used for video conferences and taking interviews. We recorded each interview using the built-in mobile recording feature, and transcripts of these recordings were generated by Office Dictation, powered by Microsoft speech services, and embedded in Microsoft Word. We analyzed the data later by following thematic analysis, a commonly used approach for qualitative studies.

The details about the thematic analysis and its outcomes are described in section IV-B.

D. Co-creation and Observations

We conducted multiple co-creation and observation sessions with industry partners at the Automated Parking System (APS), the oil and gas company, and an industrial conglomerate's defense division.

The APS provider is a medium-sized company. It delivers APSs and provides maintenance services in operation. The company has around 35 parking installations throughout Norway.

The energy services (oil and gas) company is a multinational corporation that provides life cycle services for the energy industry. We conducted multiple workshops and observation sessions to understand the real-life context of the company involved in complex engineering projects.

The industrial conglomerate have divisions in areas such as Shipping, Defence, and financing, to name a few. We focused on the defense and aerospace division's employees for the presented paper.

IV. Results and Observations

This section details results and observations from the survey analysis.

A. Analysis of survey responses

The questionnaire results (see Figs. 2, 3, 4, 5 and 6) show that internal stakeholders (employees) feel dissatisfied by the utilization of big data in their projects, especially in early phase decision making as the Net Promoter Score (NPS) is negative across the board.

Regarding data availability, the respondents either agree or are partial to the availability of data they need (see Fig. 2 Q1 to Q3) although they do think there is room for improvement as the NPS is negative. However, the availability of the right tools to explore and process the data is a more significant issue for them; as they mostly disagree or have a neutral response to the questions regarding the availability of such tools (see Fig. 2 Q4 to 6). One notable surprise, however, is question# 7 (see 2), which asks the respondents if data is being held back from them for confidentiality reasons, to which the respondents disagreed. In this case, it is a positive outcome and runs counter to our earlier assumptions [43].

The respondents expressed more dissatisfaction with the usability of data compared to its availability (see Fig. 3). For example, respondents largely disagree with whether they spend sufficient time analyzing past data in the beginning

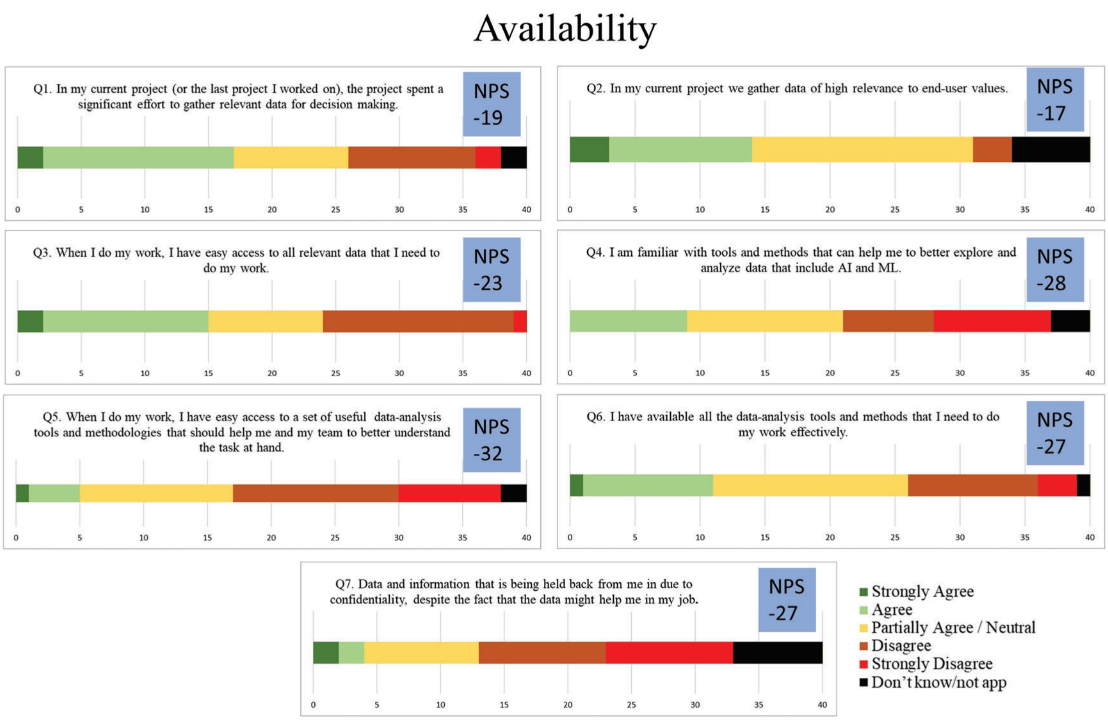

Fig. 2. Responses to Questions about Data Availability. NPS stands for Net Promoter Score

phases of their projects (Q# 8). They also think the procedures for sharing data at their organizations are insufficient (Q# 13). The respondents also have negative or neutral sentiments regarding utilizing past (historical) data as lessons learned for new projects (Q#14).

In cases when data is available to them, the respondents expressed trust in the integrity and correctness as the responses to Q# 15 are positive or neutral (see Fig. 4).

The respondents are somewhat divided on the competence of using (big) data. Respondents avoided this category of questions more often than other questions, and responses have a relatively even split (see Fig. 5).

Respondents believe their organizations have a long way to go when utilizing big data in their operations and project development. Respondents mostly disagreed with the question related to organization behavior (see Fig. 6).

For the most part, the survey results are not surprising, as it is not only the finding of our initial conversations with industry partners, but other surveys reached the same conclusion [3], [9]. However, there are some interesting findings as well.

For example, one interesting outcome is that engineers and personnel involved with the technical aspect of projects gave lower scores than project managers and upper management. Project managers seem to have a rosier perception than others (see Fig. 7). There is a need for more outstanding communication among project managers and other non-technical stakeholders, engineers, and technical personnel. While the more positive responses are somewhat in line with [9], there are notable differences compared to [9]. For example, in our survey, the upper management seemed less optimistic than the project managers.

Another notable exception is the "Competency" section of the questionnaire. For example, the report [9] stated that business leaders overestimate the capabilities of their workforce. In contrast, our survey showed that engineers and project managers gave higher responses than upper management (see Fig. 7).

In terms of age groups, employees in younger and older age groups overall gave higher scores compared to the middle (35-44) age group, while the middle age group reported the most confidence in their competency compared to the others (see Fig. 8). Also, regarding the organization behavior category, younger and senior employees express greater optimism than the 35-44 age group. Question# 21 was an exception, which asks about taking full advantage of operational data in early phase decision making, to which the 35-44 age group gave a higher score than the others.

B. Thematic analysis of semi-structured interviews

To extract themes from the interviews, we used a technique known as thematic analysis [44] for finding, analyzing, organizing, summarizing, and reporting the outcomes from the data collected [45]–[47]. Fig. 9 outlines the utilized approach.

Interviews transcripts comprised of 26 pages, and we added them as input for the NVivo, a qualitative data analysis software. In the first step, we read the transcripts to get a

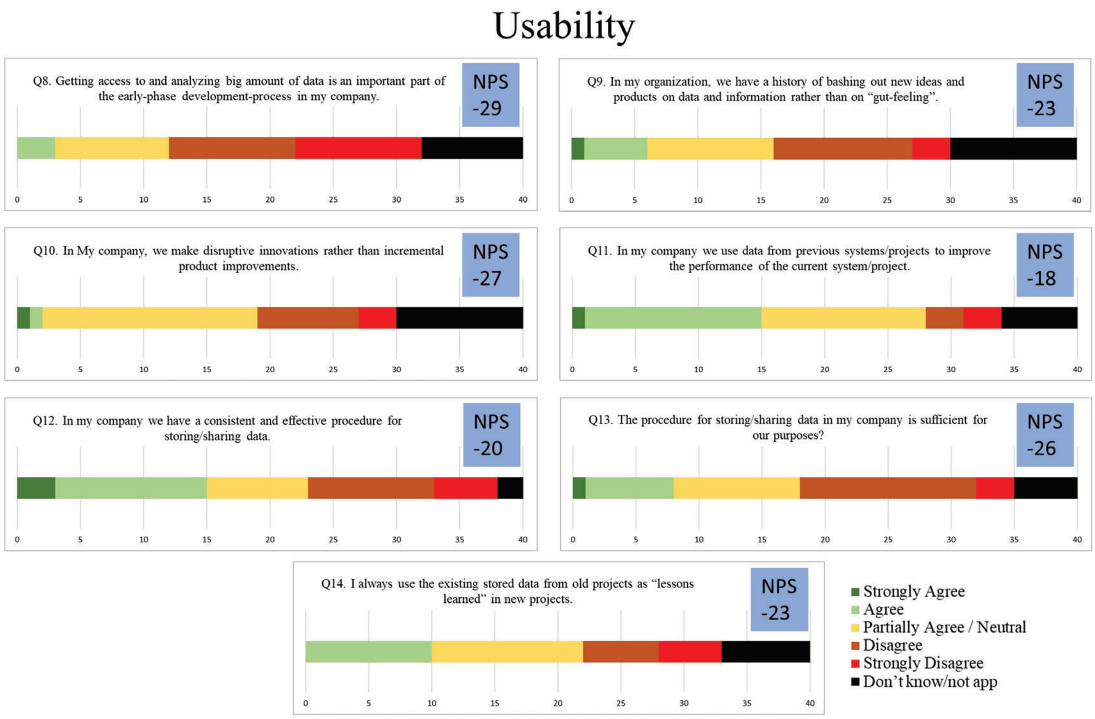

Fig. 3. Responses to Questions about Data Usability. NPS stands for Net Promoter Score

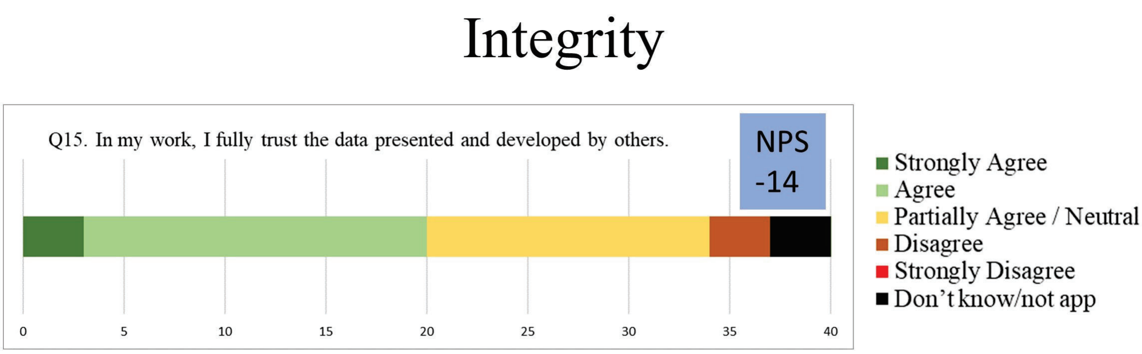

Fig. 4. Responses to Questions about Data Integrity. NPS stands for Net Promoter Score

broad picture of the collected data and familiarize ourselves with the text. Important phrases and words were identified that were repeatedly used by participants and assigned codes. Then, in order to facilitate analysis, the codes were grouped together into larger categories based on their shared qualities such as personal beliefs or professional progress [48]. In the second stage, we used NVivo software to create a hierarchical category system based on the linkages or ties between codes and categories. We searched and identified patterns across the data. These patterns were considered themes. We identified and iterated several times over potential themes that we identified from the codes and categories. These iterations ensure that we included all relevant data from the interviews.

In qualitative research a code refers to a word or phrase that captures the meaning or essence of a piece of data. Codes can be organized in to categories to further get a nuance understanding of the data in this case the interview transcripts. Using a tool such as NVivo, certain themes can be extracted from that. In NVivo, a theme is a topic that is found within the data.

We spotted the crucial statements based on the themes, including codes and categories with descriptions, using the

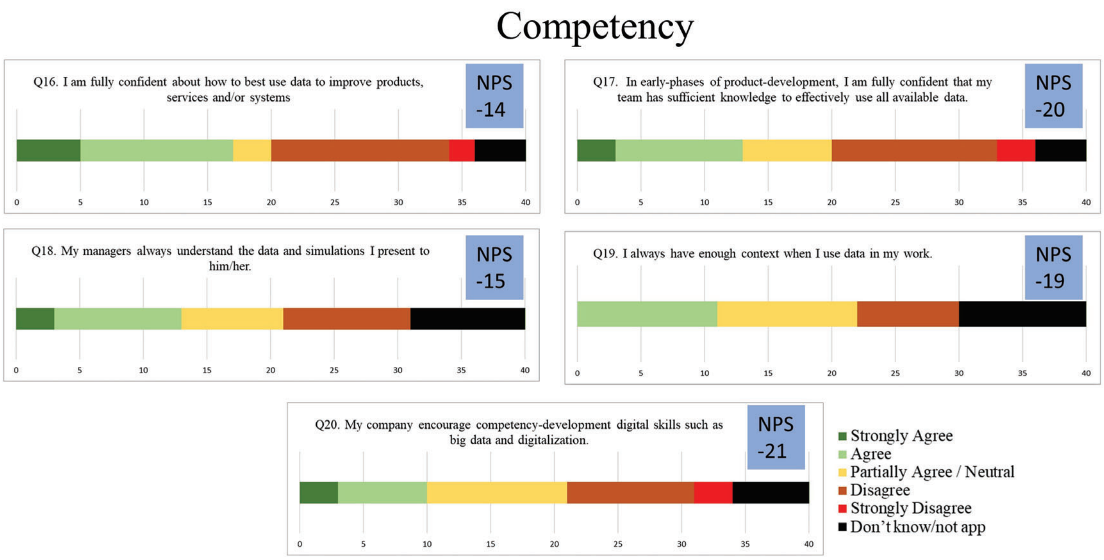

Fig. 5. Responses to Questions about Competency. NPS stands for Net Promoter Score

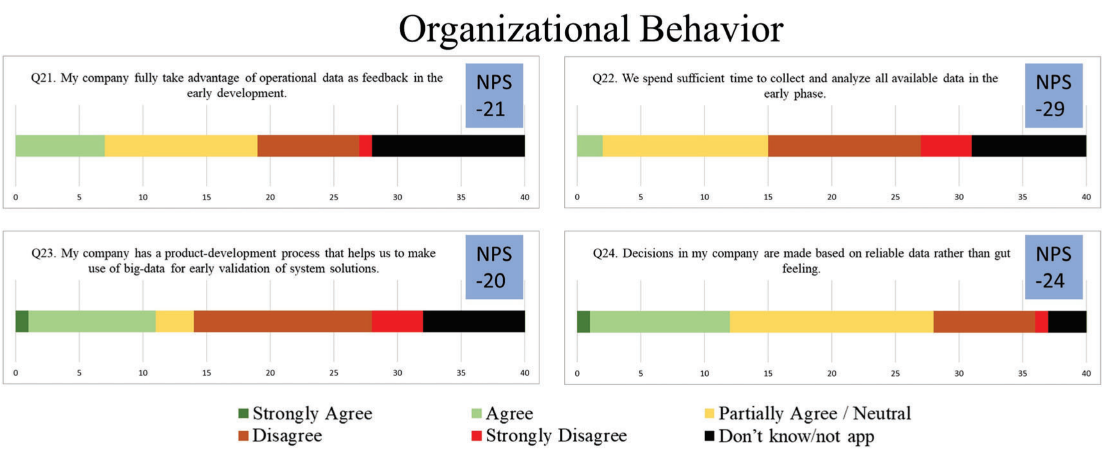

Fig. 6. Responses to Questions about Organization Behavior. NPS stands for Net Promoter Score

NVivo software tool. Furthermore, we calculated the frequency of the categories and sub-categories. Ultimately, we visualized the results to highlight the most frequently referred categories that emerged from the interview data.

We identified 11 codes in four categories. These categories (and codes) further belong to three themes. Fig. 10 visualizes the generated code-book with descriptions. Fig. 12 depicts the codes, categories, and themes and their relation.

Calculation of relative frequencies of the identified codes and categories demonstrated that the most cited categories are data handling in projects (49.1%), data types and tools (17.6%), and professional development (17.6%). Focusing solely on the codes, we can see that access to past data is the most cited, with 18.5%, followed by data storage (14.8%), presence of analysis tools (11.1%), detect certain data (10.2%); those four sub-categories totaled 54.6% of all occurrences. The other 7 subcategories total 45.4% (Fig. 11).

We identified three main themes from the categories (Fig. 12):

- Reflection on big data usage.
- Utilization of big data in projects.
- Approaches to establishing the data-driven culture.

Theme 1: Reflection on Big Data is a significant theme that reflects the experiences and feelings of the employees on big data usage at work. The theme portrays the whole story of the stakeholders' viewpoints on this project and is depicted in

Fig. 7. Average responses per job roles.

Fig. 8. Average responses per age group.

codes and categories. The most common words in this category were important, useful, better, and right (Fig. 13).

Theme 2: Utilization of Big Data in Projects depicts how the employees use big data and what issues they face in the organization. Codes and categories also illustrate data tools and the data usability process. For instance, when we asked about access to the relevant data from past projects, the respondents revealed they don't have access to past (historical) data due to a lack of a central storage system and privacy. In the final analysis, we used this theme to identify employees' current main issues. However, one respondent noted it differently:

"*The case is also that we don't have some that much data from the other projects. We have not been good enough at collecting data and storing the data, so even if it were relevant, we wouldn't have access to a lot of it. But if I want to answer the question, it could be relevant if we spend more time analyzing and understanding how we can use data between projects*."

The replies from the employees suggested that the consultancy lacks essential data access and storage facilities and the ability to comprehend and devote time to data analysis.

Theme 3: Approaches to Establish the Data-Driven Culture conveys the approaches the organization is taking to create a data-driven environment and the competency level of the company's employees. Employees feel the need to develop their skills in data analysis more, although the management appears to be interested in such affairs of the organization. We know from in-depth interviews that the users have knowledge about data analysis, but the organization also need to emphasize more workshop or courses regarding on big data or digitalization skills to improve their data literacy.

C. *Observations from co-creation sessions*

This section details the observations from co-creation sessions at the APS provider, the Oil and Gas company and an Industrial Conglomerate.

1) Observations at the Automated Parking Systems (APSs) Provider: The co-creation and observation sessions revealed that the APS provider has mostly unstructured data with many variations. One primary source of this data is maintenance log data, also called failure data. Maintenance personnel are logging failure data manually using an Excel file. Failure data includes a description of the failure events, its possible cause, and a possible implemented solution to the failure called the reason parameter (column) in the company's failure data. In addition, the failure data include, among others, the following parameters (columns): date (for a maintenance/failure event), time, telephone number (for the maintenance personnel who investigated the failure event), place number (for which parking lot the failure event occurred), invoiced yes/no (if the failure event is invoiced as it is not included within the maintenance agreement with the company, or not).

However, the company also has some in-system data. This in-system data is logging data that the System of Interest (SOI), i.e., APS stores automatically. This logging data includes the status of each subsystem, its position, and the date and time for this status. The in-system data also includes alarm log data. The alarm log data register only the abnormal situation of subsystems, with its position, date, and time.

This abnormal status or operation can be a gate not closed, or a motor may have stopped during the operation. The company has different installations for the APSs and storing mechanisms for each system; some are similar, and others differ. The APS Company cooperates with a third party for their in-system data, as this third party is responsible for this data. Unfortunately, this data can be saved daily or for a few days from only one system. However, the company is investigating if this data can be extracted from some of its systems for a more extended period with the third party. The company has more than 36 installations. The company uses a sheet in the excel file for each system or installation to store their failure events (data). The template for each sheet differs between sheets. This difference includes rearranging the parameters (columns) and describing the same issue using different terminology.

2) Observations at the Oil and Gas Company: : Observations, interviews, and co-creation sessions at the oil and gas company also revealed that they also have unstructured data with some variations of structured data. The unstructured data

In-depth Interview Guide

A. Introduction

1. Consent of Recording

B. Personal Information

1. Name and Experience
2. Work tasks and responsibilities

C. General Questions about Big Data Needs

1. What could be the definition of internal stakeholders? what's your opinion?

2. What could be the internal stakeholders needs while they are using data? According to your experience, is it important to identify internal stakeholder needs?

3. Is the usage of data equally important for all team members? If it is important, then How would be it?

4. What is big data? What kind of data you are using in this project? are you considering this part of big data or just data?

5. Do you think, using data helps to improve products and minimize the production costs, risks? Why do you think this?

6. Would your product or development be better if you had more data?

7. Do you think, using data helps to improve products and minimize the production costs, risks? Why do you think this?

8. Do you have easy access to a set of useful data-analysis tools and methodologies that can help a team to better understand the project task? What types of tools you have used, or you are using?

9. Did you have the similar project as FlexLink project you have now?

10. In your present project, do you have any relevant data from past projects that you think you can use in this project? what kind of relevant data there are? how useful is the relevant data from earlier projects?

11. Do you have access of past data from previous projects? If there is any data, then what kinds of data? How are you (or team) using past data in your current project?

12. If you do not have access to certain data, where could you find it?

13. Can you spend sufficient time to collect and analyse all available data in the early phase? Why you cant spend more time on it?

14. You are doing mobile robot for FlexLink. May be after 10 or 15 years, someone from your group or your company will get the chance to build similar kind of robot. Would be efficient for him if he could have the access of this data and can use this data in his project? How could you ensure this process-do you have any comment?

15. Why is big data important for early-phase development-process? Is there any importance of big data in your project?

16. Is there any procedure for storing previous and current data that you or your team can easily use in your current project? What is the procedure? Why you (don't) have this storing procedure?

17. Sometimes data and information are being held back from team member because of confidentiality. Have you ever faced this issue in your this project or in your previous project? How will/did you solve this?

18. If data is breached, what (technical) solutions can be taken to ensure privacy?

19. When you work in the early-phases of product-development, are you fully confident that your company (and project team) has sufficient knowledge to effectively use all available data? If no, then which causes behind these problems?

20. How do your company encourage competency-development and training on digital skills such as big data and digitalization? If you do any courses, are these paid and endorsed by the company? Does the company encourage the employees to take these courses?

Fig. 9. Interview Guide.

Code	Description
Views of Stakeholders of the Project	The thoughts of team members about data needs in the project
Internal stakeholders' data needs	Importance of identifying the needs of employees when they use data in early-phase development-process
Product improvement	Using data in products improvement and production costs, risks minimization
Data types and tools	The data types and tools are used by participants
Presence of analysis tools	Utilization of data-analysis tools and methodologies to better understand the project
Data Wishlist	Data types which are wished for
Data in projects	Usage of the relevant data from the past projects in ongoing projects
Spending time	Sufficient time to collect and analyse all available data in the early phase
Detect certain data	Certain data if do not have the access or are being held back from team member for confidentiality
Data storage	Procedure for storing previous and current data and usage of lessons learned
Access of past data	Obtaining the similar data from past projects
Data accessibility	Establishment of the data usability and accessibility in future
Training	Training on digital skills is emphasized by company
Ensure privacy	Solutions for data breaching
Data confidence	Sufficient knowledge to effectively use all available data

Fig. 10. Thematic Categories

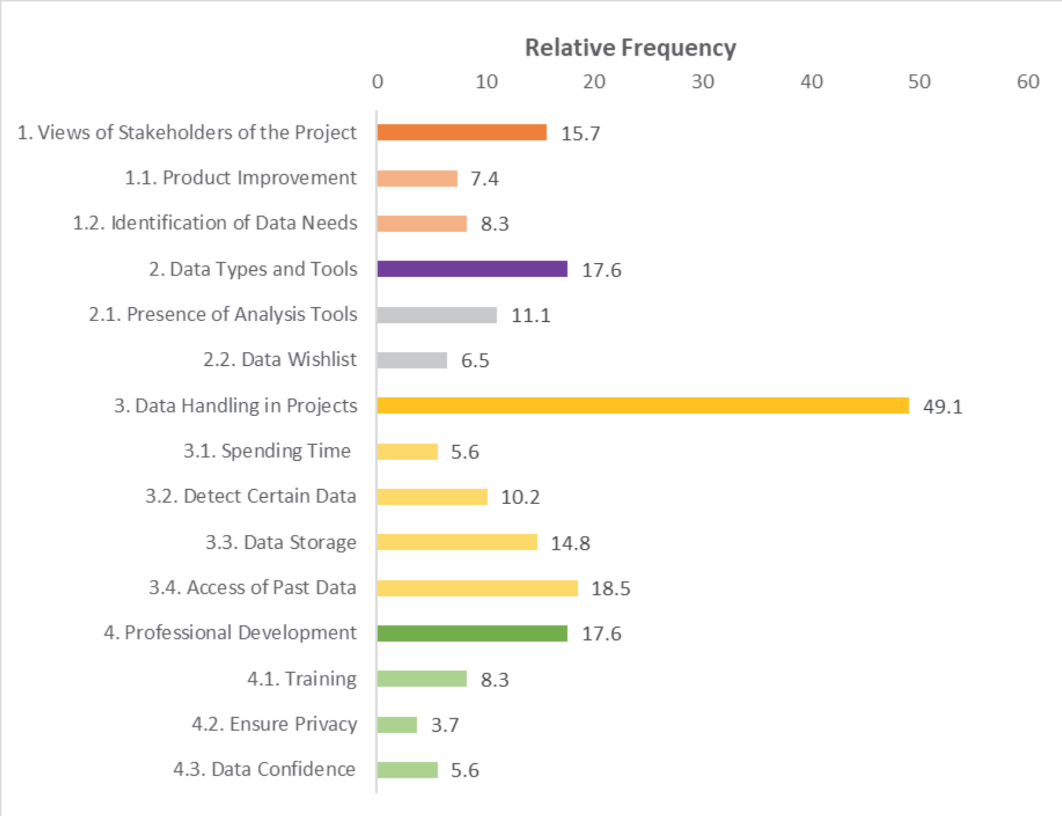

Fig. 11. Visual representations of the categories identified from the interviews. The relative frequencies of the categories and subcategories are also shown in the illustration.

is mainly in event log. The test personnel are logging manually using Excel. The central part of this log is a description of the unexpected events/issues/problems (also called emergent behavior) during the test process, such as the System Integration Test (SIT).

The event log also contains other parameters (columns) such as data about the equipment and its serial number, information about the project, e.g., work package and product responsible department or supplier, and other project-specific information. The company uses an Excel file for each project. However, the template varies for each project. This variation can be rearranging some columns (parameters) and having some rows in the middle of the Excel file. In some Excel files, different terms describe the same issue. There are some errors in the titles of some parameters.

3) Observations at Industrial Conglomerate (Defence Division): We have observed that the defense division struggles with adopting suitable methods for harvesting and utilizing data relevant to Human Systems Integration in unmanned vehicles. A particular area is generating enough appropriate data for supporting designers in exploring and testing various Human Machine Interactions (HMI) solutions. Their decision-making is primarily based on customer requirements, standards, in-house expertise, and subjective opinions from multiple developers. However, they wish to put more emphasis and weight on objective data in their decision-making process. Best practices, such as the Design Thinking process, highlights the need for gaining grounded understanding from relevant users, obtained through prototyping and testing.

The data they need are sourced from the end-users, the components of the SOI, and the system's use. The data can be accessed from the technical system, which is being developed and tested in-house. The defense division also has access to general information on how the system should be used. However, the procedure of how the system should be used is not necessarily the way the system will be used in the field. The organization cannot generate enough data for making objective data based decisions as they have an incomplete picture of the system context, specifically the data from the end-users. As they lack access to all system context information, they can only measure and analyze their data, not necessarily the data they should analyze.

V. CONCLUDING REMARKS

The case study gave us a deeper understanding of the state of affairs at some of the industry partners. It provided a glimpse at the disparity of different organizations regarding their utilization of big data in their operations and decision-making process, focusing on the early phase product development process. Overall, the case study concluded that employ-

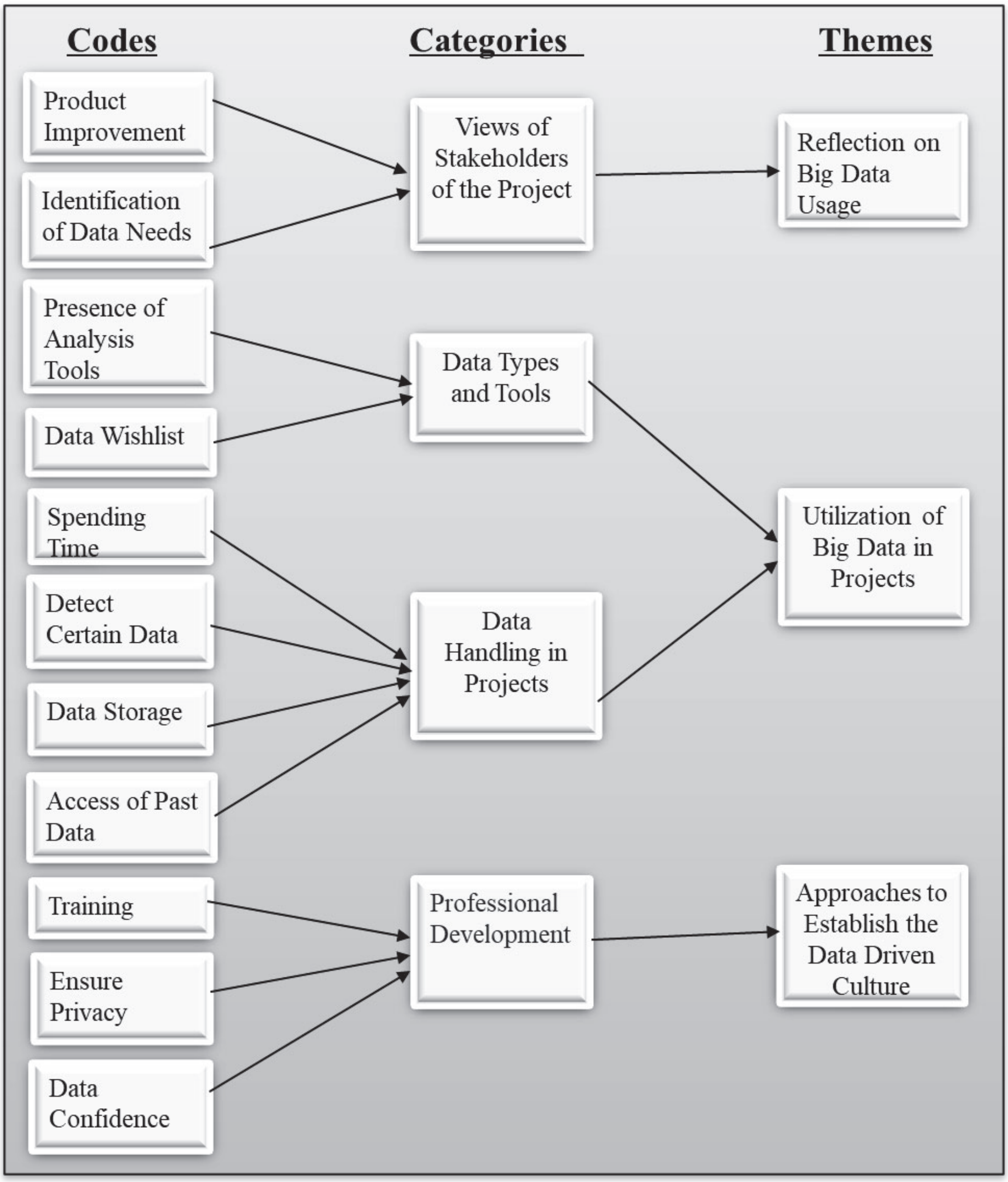

Fig. 12. Themes used for the thematic analysis of the conversation with employees.

Fig. 13. Word cloud showing the most common words appears in the “Reflection on Big Data Usage” theme.

ees at the H-SEIF2 industry partners understand the need to use big data in their projects to enhance their operations.

The employees at the technology consultancy reported that a considerable amount of data generated after every project is often stored only locally by the personnel instead of stored in an accessible database of the company. The company lacks sophisticated big data analysis tools to understand a complex project thoroughly. Currently, the company uses Python, Excel, and similar common tools for simple data analysis.

Similarly, in the observations at both Automated Parking System and oil and gas provider, we observed that both companies use Excel files that either maintenance or test personnel log manually. However, the template for the Excel file differs slightly for each system or project. This difference makes it cumbersome to automate the pre-processing in different degrees depending on the data, especially when gathering historical data for an extended period, e.g., five or ten years. Thus, manual pre-precessing is needed. In other words, we must pre-process the data manually by generating a template called frame around the data. This manual pre-processing consumes almost 80% of the analysis period.

However, manual pre-precessing is time-consuming and may result in some errors when doing it manually. Therefore, we recommend that the company unify the template and use a not-editing version. Also, only certain manager-level employees should change the template, not everyone in the test or maintenance department. In [49], we suggested a template that integrates the needed data and information on one platform using one tool.

For the defense division, the main limitation is a lack of end-user data. There are primarily two reasons for such lacking. Firstly, they have no access to the end-users usage in field operations due to security reasons. They are building simulators to combat limited access to end-users and end-user data. Secondly, they lack integrated and developed test procedures for simulator testing to generate and collect human factors data. The human factor data include physiological, mental, and operational data. These data can be collected through, for example, eye tracking, heart and galvanic skin response measurement data to measure the stress of the end-users, interviews and test questions, and performance data. These data aim is to understand the HMI influence on situational awareness during operations.

The presented case study provided the current state of big data usage scenarios using different research methods at Norwegian high-tech companies and identified issues hindering big data's full potential. The observations from the case study can be used for future research on similar business organizations in addressing the internal stakeholders' needs regarding big data usage.

REFERENCES

[1] F. A. Salim, H. B. Ali, T. Langen, A. Wettre, G. Muller, and K. Falk, "State of Affair in Terms of Big Data Utilization in Complex System Engineering Organizations," in *MODERN SYSTEMS 2022: International Conference of Modern Systems Engineering Solutions*, no. c, Nice, France, 2022, pp. 45–49.

[2] R. K. Perrons and J. W. Jensen, "Data as an asset: What the oil and gas sector can learn from other industries about "Big Data"," *Energy Policy*, vol. 81, pp. 117–121, 2015. [Online]. Available: http://dx.doi.org/10.1016/j.enpol.2015.02.020

[3] R. Technology, "AI / ML Annual Research Report 2022," 2022, (last accessed 05-Dec-2022). [Online]. Available: https://mikenashtech.com/2022/02/ai-ml-annual-research-report-2022-analysis/

[4] K. Falk, A. K. Kamara, E. P. Brathen, K. Helle, P. T. Moe, and S. Kokkula, "Digitizing the Maintenance Documentation; A System of Systems in Oil and Gas Industry," *SOSE 2020 - IEEE 15th International Conference of System of Systems Engineering, Proceedings*, pp. 493–499, 2020.

[5] H. B. Ali and F. A. Salim, "Transferring Tacit Knowledge into Explicit: A Case Study in a Fully (Semi) Automated Parking Garage," in *PROCEEDINGS OF The Society for Design and Process Science*, 2021.

[6] F. A. Salim, H. B. Ali, and K. Falk, "Towards a Data Processing Framework for Enhanced User Centric System Engineering," in *PROCEEDINGS OF The Society for Design and Process Science, Workshop on Smart Pervasive Computing*, 2021.

[7] E. Raguseo, "Big data technologies: An empirical investigation on their adoption, benefits and risks for companies," *International Journal of Information Management*, vol. 38, no. 1, pp. 187–195, 2018.

[8] F. A. Salim, F. Haider, S. Luz, and O. Conlan, "Automatic transformation of a video using multimodal information for an engaging exploration experience," *Applied Sciences (Switzerland)*, vol. 10, no. 9, 2020.

[9] Qlik and Accenture, "The Human Impact of Data Literacy," Tech. Rep., 2020, (last accessed 05-Dec-2022). [Online]. Available: https://thedataliteracyproject.org/humanimpact

[10] B. A. Delicado, A. Salado, and R. Mompó, "Conceptualization of a T-Shaped engineering competency model in collaborative organizational settings: Problem and status in the Spanish aircraft industry," *Systems Engineering*, vol. 21, no. 6, pp. 534–554, 2018.

[11] S. Engen, K. Falk, and G. Muller, "The need for systems awareness to support early-phase decision-making—a study from the norwegian energy industry," *Systems*, vol. 9, no. 3, 2021.

[12] Juzgado P.D. Borches, "A3 Architecture overviews. A tool for effective communication in product evolution," Ph.D. dissertation, University of Twente, 2010.

[13] T. Tomiyama, V. D'Amelio, J. Urbanic, and W. Eimaraghy, "Complexity of multi-disciplinary design," *CIRP Annals - Manufacturing Technology*, vol. 56, no. 1, pp. 185–188, 2007.

[14] U. NISE, "H-SEIF 2 Project," (last accessed 05-Dec-2022). [Online]. Available: https://www.usn.no/hseif

[15] C. Potts, "Software-engineering research revisited," *IEEE Software*, vol. 10, pp. 19–28, 1993.

[16] M. Guntveit, M. Kjrstad, and B. Sevaldson, "Early Validation of Stakeholder Needs by Applying Co-creation Sessions," *INCOSE International Symposium*, vol. 30, no. 1, pp. 1398–1415, 2020.

[17] G. Gravier, M. Ragot, A. Laurent, R. Bois, G. Jadi, E. Jamet, and L. Monceaux, "Shaping-Up Multimedia Analytics: Needs and Expectations of Media Professionals," in *The 22nd International Conference on Multimedia Modelling*, vol. 9516, Miami, 2016, pp. 303–314.

[18] M. Haesen, J. Meskens, K. Luyten, K. Coninx, J. H. Becker, T. Tuytelaars, G. J. Poulisse, P. T. Pham, and M. F. Moens, "Finding a needle in a haystack: An interactive video archive explorer for professional video searchers," *Multimedia Tools and Applications*, vol. 63, no. 2, pp. 331–356, 2013.

[19] F. A. Salim, F. Haider, O. Conlan, and S. Luz, "An approach for exploring a video via multimodal feature extraction and user interactions," *Journal on Multimodal User Interfaces*, no. October 2017, 2018.

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

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

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

[23] 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, pp. 1–16, 2012.

[24] J. Dijcks, "Oracle: Big data for the enterprise," Oracle, Tech. Rep., 2012.

[25] S. Gogia, M. Barnes, B. Evelson, B. Hopkins, N. Yuhanna, D. Anderson, and H. Kisker, "The Big Deal About Big Data For Customer Engagement," Forrester, Tech. Rep., 2012, (last accessed

05-Dec-2022). [Online]. Available: https://www.forrester.com/report/The-Big-Deal-About-Big-Data-For-Customer-Engagement/RES72241

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

[27] H. B. Ali, F. H. Helgesen, and K. Falk, "Unlocking the power of big data within the early design phase of the new product development process," *INCOSE International Symposium*, vol. 31, no. 1, pp. 434–452, 2021.

[28] B. Schätz, M. Törngren, R. Passerone, S. Bensalem, A. Sangiovanni-Vincentelli, J. McDermid, H. Pfeifer, and M. Cengarle, "CyPhERS-cyber-physical European roadmap and strategy," fortiss GmbH, Munich, Tech. Rep., 2014.

[29] A. Alharthi, V. Krotov, and M. Bowman, "Addressing barriers to big data." *Business Horizons*, vol. 60, no. 3, pp. 285–292, 2017.

[30] M. Douglas, "Big data raises big questions," *Government Technology*, vol. 26, no. 4, pp. 12–16, 2013.

[31] S. Hoffman and A. Podgurski, "Big bad data: law, public health, and biomedical databases," *Journal of Law, Medicine Ethics*, vol. 41, no. 1, pp. 56–60, 2013.

[32] B. Laugwitz, T. Held, and M. Schrepp, "Construction and Evaluation of a User Experience Questionnaire," *HCI and Usability for Education and Work*, pp. 63–76, 2008.

[33] G. Nicolini and L. D. Valle, "Errors in Customer Satisfaction Surveys and Methods to Correct Self-Selection Bias," *Quality Technology Quantitative Management*, vol. 8, no. 2, pp. 167–181, 2011.

[34] B. Popovic, R. Maletic, and T. Paunovic, "Employee Satisfaction Survey in Function of Business Improvement," *Management - Journal for theory and practice of management*, vol. 20, no. 76, pp. 31–40, 2015.

[35] L. She, L. Ma, A. Jan, H. Sharif Nia, and P. Rahmatpour, "Online Learning Satisfaction During COVID-19 Pandemic Among Chinese University Students: The Serial Mediation Model," *Frontiers in Psychology*, vol. 12, no. October, 2021.

[36] B. Dudley, "Digital Transformation Initiative: Oil and Gas Industry In collaboration with Accenture," *World Economic Forum*, no. January, p. 32, 2017, (last accessed 05-Dec-2022). [Online]. Available: www.weforum.org

[37] D. H. Shin, "Demystifying big data: Anatomy of big data developmental process," *Telecommunications Policy*, vol. 40, no. 9, pp. 837–854, 2016.

[38] G. Muller, "Systems engineering research methods," *Procedia Computer Science*, vol. 16, pp. 1092–1101, 2013. [Online]. Available: http://dx.doi.org/10.1016/j.procs.2013.01.115

[39] ——, "Tutorial Architectural Reasoning Using Conceptual Modeling," University of South-Eastern Norway, Kongsberg, Tech. Rep., 2015. [Online]. Available: https://www.gaudisite.nl/TutorialARconceptualModelingSlides.pdf

[40] I. Seidman, *Interviewing as qualitative research: A guide for researchers in education and the social sciences:*, 3rd ed. Teachers college press, 2006.

[41] R. Berry, "Collecting data by in-depth interviewing," in *British Educational Research Association Annual Conference*, Brighton, 1999.

[42] M. Saunders, P. Lewis, and A. Thornhill, *Research methods for business students*. Pearson education, 2009.

[43] N. M. Sjøkvist and M. Kjørstad, "Eliciting Human Values by Applying Design Thinking Techniques in Systems Engineering," *INCOSE International Symposium*, vol. 29, no. 1, pp. 478–499, 2019.

[44] V. Braun and V. Clarke, "Using thematic analysis in psychology." *Qualitative research in psychology*, vol. 3, no. 2, pp. 77–101, 2006.

[45] J. Anuradha, "What is the Difference Between Thematic and Content Analysis," 2022, (last accessed 05-Dec-2022). [Online]. Available: https://pediaa.com/what-is-the-difference-between-thematic-and-content-analysis/

[46] A. Bogetz, E. Abramson, H. Haftel, and M. Klein, "Codes, concepts and categories, oh my! Building your skills in qualitative data analysis," Association of Pediatric Program Directors, Anaheim, Tech. Rep., 2017. [Online]. Available: https://appd.s3.amazonaws.com/docs/meetings/2017SpringPresentations/WS10Slides.pdf

[47] G. Gibbs, "Thematic coding and categorizing," in *Analyzing qualitative data*. SAGE Publications, Ltd, 2007, ch. 4, pp. 38–56.

[48] J. Corbin and A. Strauss, *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*, 4th ed. SAGE Publications, Ltd, 2014.

[49] F. A. Salim, H. B. Ali, G. Muller, and K. Falk, "User-centered Data Driven Approach to Enhance Information Exploration, Communication and Traceability in a Complex Systems Engineering Environment," in *MODERN SYSTEMS 2022: International Conference of Modern Systems Engineering Solutions*, no. c, 2022, pp. 37–42.