Improving the utilisation of AIOps analyses by improving the quality of incident logging data

Bachelor thesis Industrial Engineering and Management

Author:	N. B. Idema (Nils)
Student number:	s2370956
Educational Institution:	University of Twente
Programme:	Industrial Engineering and Management
Company:	Rabobank
Date:	31-08-2022
Internal supervisor:	MSc A. Baris (Arie)
First supervisor (UT):	Dr. Rer. Nat. D. Braun (Daniel)
Second supervisor (UT):	ir. R.L.A. Harmelink (Rogier)

Acknowledgements

Before you lies the thesis that is the result of my research at Rabobank, which concludes my Bachelor's study. During the period that this study encompassed, I have received support from a variety of people, whom I would like to show my gratitude.

First of all, I would like to thank Arie Baris, my supervisor at Rabobank. His perspective and critical thinking have provided me with incredibly valuable input on how to improve the research. Combined with his enthusiasm and energy he made for an excellent supervisor.

I would also like to thank my first supervisor at the university, Daniel Braun, who has given great guidance on the methodological aspects of this study, and who has always been willing to make time for me.

Apart from these, I would like to thank the employees of Rabobank that I worked with and who have made this study possible, either via organising it, proofreading it, or being interviewed. Specifically, I wish to express my gratitude towards the AIOps team, who have always been willing to share their expertise with me, and who have made my time at Rabobank a very enjoyable time.

I sincerely hope that you enjoy reading my thesis.

Nils Idema
Apeldoorn, 31 August 2022

Management Summary

Problem

This study focuses on improving the data quality of the incident registration process at Rabobank, so the Artificial Intelligence (AI) recovery of incidents in Rabobank's Information Technology (IT) systems can be improved. The goal of the improved data quality is that a solution recommender can be built, which recommends solutions to operators, meaning that the time it takes an operator to solve a problem will be decreased. This in turn reduces the impact of the incident, meaning that the action problem of having more impactful incidents than preferred is remediated through the core problem:

The quality of data of the incident registration process is insufficient to optimise AI recovery of incidents.

This lack of data quality is expressed in two manners. First, measurements show that field 'solution' is too often filled in with a non-English language, or not filled in at all. Second, the user ratings required for a recommender system cannot be collected. The goal of this study is to find a way to collect the required user ratings from Artificial Intelligence for Operations (AIOps) using teams and to find a way to improve the quality of data filled in by all teams. The AIOps-using teams can test the non-AIOps-specific solutions.

Solution approach

Using an approach based on the Managerial Problem Solving Method (MPSM), this study presents four deliverables: a descriptive model of the incident registration process, an evaluation of several solution concepts, a recommendation on how to improve the data quality, and an examination of further research possibilities that were encountered at some point in this research. To achieve this, the conducted research is done using literature studies, but mostly via interviews with different stakeholders at Rabobank. As a result of such interviews, the difference between the current process definition and the preferred process definition was encountered. This difference entails a lack of information between the AIOps team and the operators, this mainly being the lack of required user ratings. In order to solve this, various solution concepts have been constructed:

1. Only using data mining techniques such as Nature Language Processing (NLP) on the currently existing data
2. Automatically sending Microsoft Forms to operators after they have solved an incident
3. Adding editable work items on teams' Azure DevOps boards to collect data in
4. Using the already existing Splunk dashboard to gain the user input from operators
5. Completely building a new web-based application using crowd knowledge

Using a questionnaire, these AIOps-specific solution concepts have been evaluated and conclusions about the quality of the concepts have been drawn. The ease of implementation, effectiveness and efficiency, and operator usage have been taken into account when evaluating the various solution concepts. Key is the conclusion that data mining presents possibilities, but poor quality data might result in poor quality analyses. Microsoft Forms is a flexible channel, but the notifications have a high perceived annoyance. Next, Azure DevOps boards although intuitive, present difficulty in the restrictiveness of the work items. Last, crowd knowledge is unnecessary and having all the dashboarding in one place is desirable, but Splunk is not made for the collection of user input.

Besides these concepts to collect the required data for the solution recommender, various non-AIOps-specific solution concepts have been constructed. The goal of these concepts is to improve the overall data quality that is being registered about incidents in all IT systems. The main premise of these solutions is to be implemented in the registration software Service Manager, which can be tested using the AIOps-specific solutions if these solutions are to be implemented. The solutions include:

1. Delinking the customer message and the solution field
2. Giving warning of poor data quality whenever a text is too short, is in the wrong language etc.
3. Giving a closed solution and description field with generalised terms to identify keywords
4. Applying templates to guide the operator to enter information in a uniform manner
5. Improved link with Knowledge Management and Problem Management so different knowledge items can more easily be related to each other.

The evaluation of non-AIOps-specific concepts, done with a questionnaire, shows the potential value of the templates and poor quality warnings as easy to implement and effective solutions. It is key to test these solutions in the Form before implementing them company-wide on how well-received and effective they are in practice. Apart from this, it is worth exploring to what degree the generalisation of the solution and description field can be reasonably achieved to a satisfying degree with various groups of people. This step is key if the high potential of generalisation is to be utilised.

Recommendation

Based on the evaluation of the AIOps-specific solutions, a recommendation can be made on how to improve the problem of poor data quality. It is recommended to implement Microsoft Forms as the way to collect the required data for the recommender system. However, to minimise the perceived annoyance at the side of the operator, instead of using Microsoft Teams, it is recommended to add a link to Forms on work items in the Azure DevOps boards. This way the flexibility of Forms is combined with the intuitiveness of Azure DevOps boards, without the drawbacks that these concepts individually possess.

The Forms can also run as a testing ground for the non-AIOps-specific solutions, the modifications to Service Manager. This way, before implementing the changes companywide, they can be properly evaluated and accordingly adjusted. Regarding the implementation of the solution concepts, it is recommended to start with a team that is experienced with the anomaly detector, such as Instant Payments. It is key to explain to the team why the data is required of them, and especially how the data benefits them. Based on the non-AIOps-specific solutions that are implemented in the Forms, it can be assessed whether the data quality has been sufficiently improved before further implementation, or if alterations are necessary.

Apart from this, during the study, several possibilities for future research have been indicated. Self-healing applications are a natural follow-up of the solution recommender, and although research can be done, implementation of self-healing must await the implementation of the recommender system. The use of AI for security and Know Your Customer (KYC) are worthy to investigate, but do not have the same potential value for Rabobank as the use of AI for impact determination. Such impact determination, and linking incidents from different applications together into payment chains can provide possibilities for impact prevention, especially for the crucial Major Incidents. It is recommended that this subject is researched further.

Contents

1 Introduction

This section of the report intends to provide information about Rabobank, the context of this problem, and the problem identification using the first phases of the MPSM (Heerkens & van Winden, 2021). This problem identification consists of determining the action problem, constructing a problem cluster, determining the core problem and measuring the gap between norm and reality.

1.1 Rabobank's strategy and mission

Rabobank is one of the biggest banks in the Netherlands (Corporate Finance Institute, 2021). As a bank that started as a collaboration between Dutch farmers and horticulturists, one of the most important spearheads of the bank is the cooperative Rabobank (Rabobank, 2022b). From its beginning until now, Rabobank places great value on a mindset focused on finding solutions that benefit everyone. This also encompasses working together with others to achieve this goal (Rabobank, 2022b).

This is reflected in Rabobank's mission statement: "Growing a better world together". This is the goal that they aim to achieve by being client-driven, professional, and considerate (Rabobank, 2022a). As a part of this mission statement, Rabobank focuses on the transition towards sustainable energy supply and consumption and the food transition.

Regarding its strategy, Rabobank is a versatile bank. This means that they offer customers products and services such as loans, payments, and savings, but also strategic advice concerning treasury, mergers and acquisitions, and mortgages. Their four strategic pillars are *Excellent Customer Focus, Meaningful Cooperative, Rock-Solid Bank, and Empowered Employees.* These are used to create focus and measure their success on output, so value can be created (Rabobank, 2022a).

1.2 Contextual information

To see how different processes fit in the bigger picture, it is important to first determine how different departments work together at Rabobank. The bank implemented a team-based agile working style, just like another Dutch bank did (Barton, 2018). These teams consist of *tribes, squads,* and *chapters* as shown in Figure 1Figure 1 Overview of . Tribes are centred around key client needs, a few examples are Accounts and Payments Factory, and Business to Business (Rabobank, 2021a). Within these tribes there are multiple areas, which have the goal of uniting the interdisciplinary squads into a common goal or function. This study is grounded in the Artificial Intelligence Operations (AIOps) squad, in the tribe Accounts and Payments Factory. However, this squad regularly crosses areas, tribes, and domains (overarching departments) to cooperate with other teams. As such, so does this study. One of the major benefits of an agile approach, as opposed to a 'traditional' hierarchical structure is the autonomy of various departments. Apart from this, especially considering the software development and IT operations (DevOps) environment that this study takes place in, the relatively unclear final goal can be reached iteratively and incrementally (Project Management Institute, 2017).

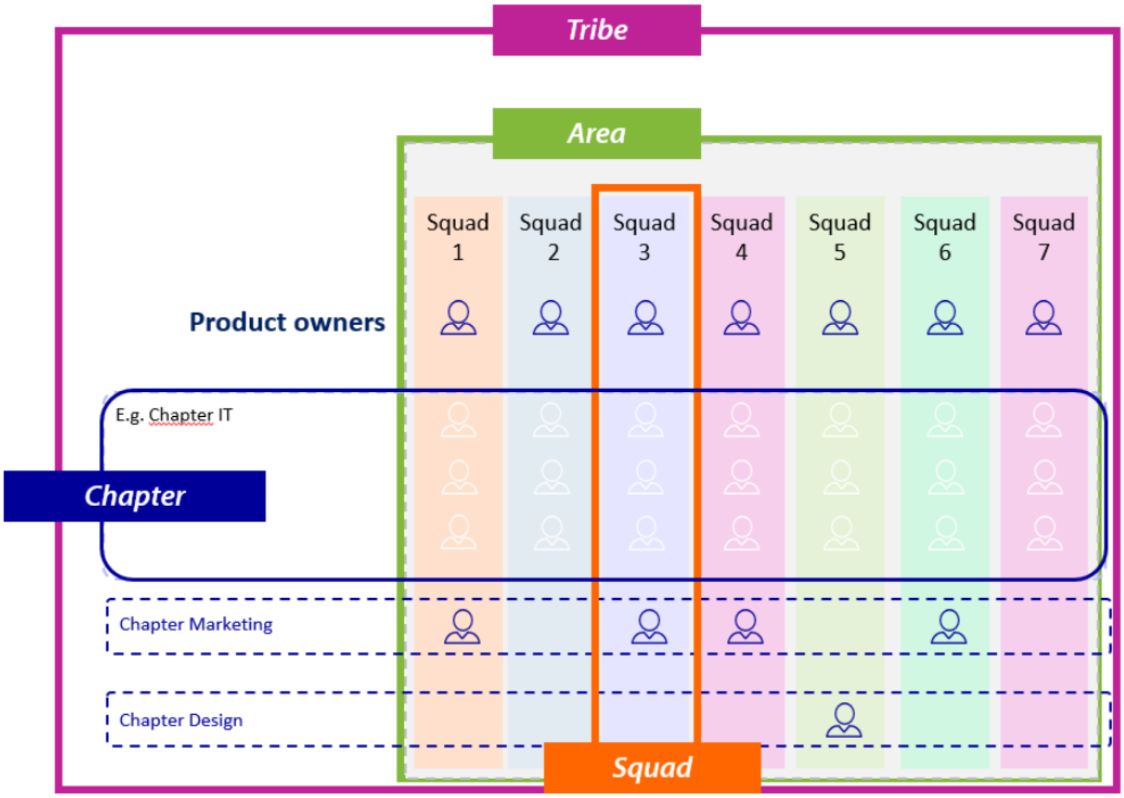

Figure 1 Overview of agile work teams in Rabobank

First, one must establish more about the current state of AIOps. This squad within the tribe Accounts and Payments Factory has the goal of applying Artificial Intelligence and machine learning to the monitoring data of the relevant applications. The squad does this using anomaly detection. Using machine learning to monitor gives the advantage of learned behaviour that is relatively unbiased from human intervention and of analyses both faster and better than humans could do (Gupta & Mangla, 2020). As Mehrotra et al. (2018) state "Anomalies or outliers are substantial variations from the norm." Anomaly detection approaches are based on models and predictions from past data. The underlying processes that lead to a certain behaviour of a system are assumed to not have changed, which can be observed through the data (Gupta & Mangla, 2020). Large variations from this normal state, the anomalies, can also be observed. Although data drift (the gradual change of data over time) is an issue that must be taken into account here, the model can be retrained if the degree of false positive anomalies (detected anomalies that are actually normal system behaviour) is too high (Mallick et al., 2022).

In the case of AIOps, it tries to find anomalies in the data retrieved from the IT systems that it services to. These anomalies are based on the normal behaviour that has been learned from historical data (Mehrotra et al., 2018). Examples of such anomalies are too much memory usage or an application becoming slower, but this varies on what information is fed to the detector and what the important features of the application are (Luo et al., 2014). Subsequently, these anomalies are plotted on a dashboard, showing where the potential anomalies occur in the application. If the situation is dire enough, a ticket is created and assigned to an operator who is then notified of this. Obviously, this is a simplification of the intricate tool that is used, but the scope of this study lies elsewhere.

1.3 Problem identification of action problem

As a bank that focuses on the importance of customer experience and has *Rock-solid bank* as one of its strategic pillars, the reliability of Rabobank's IT systems must at all times be kept in mind (Rabobank, 2022b). This is especially true for digital payments. This is one of the key services that a bank provides for its customers. Any downtime of a payment chain is time that a customer either has to wait for their money or cannot make payments at all. For a customer-driven company such as Rabobank, it is important to keep the number of incidents in the payment chains as low as possible. Concerning the payment chains, every incident is one too many.

This presents the action problem of this study (Heerkens & van Winden, 2021):

There are more impactful incidents in the IT landscape of the payment chains at Rabobank than preferred.

Rabobank defines incidents as follows: "An incident is a malfunction of a single infrastructure or software component of a solution that processes information. Examples of failures are memory failure, network component failure, etc." (Rabobank, personal communication, 2022). These incidents put the status of the entire payment chain at risk and waste valuable resources such as time, money, and data storage. In this case, impactful refers to the time it takes to resolve an incident. This is to keep the scope of the study in mind. Impactful can also refer to financial or personal consequences, but measuring this, especially in a company as big as Rabobank, is beyond the scope of this research.

1.4 Finding the core problem

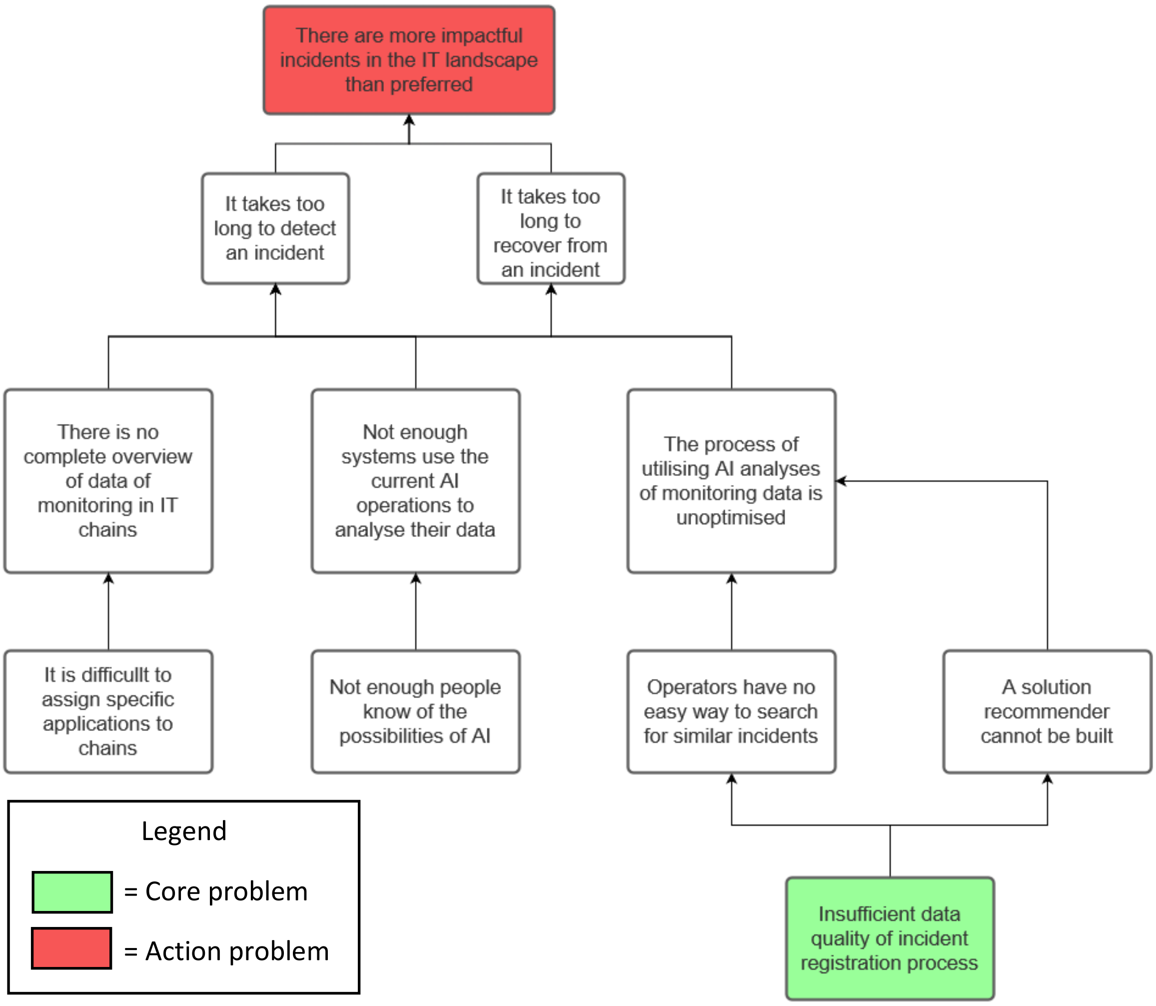

Figure 2 Problem cluster

When logically looking at the problem cluster in Figure 2 following the causal relationships, one can distinguish two main reasons behind the observation that there are more impactful incidents than preferred (Heerkens & van Winden, 2021). The first is that it takes too long to detect an issue, and the second is that it takes too long to recover from an incident. Multiple reasons can be distinguished for this to occur, but to keep the problem cluster orderly problems such as 'lack of personnel' have been omitted. In conversations with relevant stakeholders, three major relevant problems were identified as reasons that the time to detect and time to recover are longer than preferred. The one that is looked at is the fact that the utilisation of AI analyses is unoptimised. The reason behind this is also the core problem of this study (Heerkens & van Winden, 2021):

The quality of data of the incident registration process is insufficient to optimise AI recovery of incidents.

It is important to distinguish two different processes here. The first process is the aforementioned AIOps process. This is the process of detecting possible incidents before they occur, or when they occur as soon as possible. For these incidents, a ticket is created and an operator is assigned to solve the problem. Then the second processes is started.

This is the process called the 'incident process'. In this process, an assigned operator looks at the problem, analyses it, and solves or tries to solve it. While doing this, they register certain information in the tool Service Manager (Micro Focus, 2022). Examples are root cause, activities they undertook, time it took them to take certain actions, solution, or priority. Ideally, this information is then given back towards AIOps, who can utilise this data to create smarter systems, but this is not yet the case.

The utilisation of AI analyses, such as AIOps can be improved, for example when an engineer can easily look into a database to see how similar problems were solved, or this can even be automated using a recommender engine. Such a solution recommender however presents requirements that the data must correspond to. In Appendix A, a detailed indication of such a solution recommender is given, and it shows that the required data must at least contain user ratings of actions to undertake in order to solve the incident.

Another example is that further analyses can be done using the improved data, to assess the average time needed for certain stages. Good data quality, however, is key (Olson, 2003). Right now the quality of the data is insufficient. Important fields are sometimes not filled in at all, or in Dutch or Portuguese, or the amount of text is too long to be utilised. This prevents projects like AIOps from taking the next step in the utilisation of their analyses. Measurements of this lack of quality are discussed in section 1.5.

As Figure 2 indicates, apart from this problem there are two other potential core problems. The reason that these were not chosen as the core problem of this study is a combination of scope and current activities. There are already ongoing projects to improve the situation of the other two problems. Apart from this, the scope of the chosen core problem better fits the time frame. It must be noted that the perspective of the research is from an AIOps point of view. This means, that although the other types of IT incidents are taken into account, the main focus is the anomaly detection incidents.

1.5 Measurement of norm and reality

To judge the current situation, three indicators are chosen to assess the quality of the data. These indicators do not cover the entire spectrum of data quality as defined by Batini and Scannapieco (2006), however, they serve to give the reader a reference on how the situation is currently. The values of the indicators are calculated from the data of one month's total incidents. A more precise definition of data quality can be found in section 3.1 where the theoretical framework is defined.

Appendix C elaborates on the measurements that were taken to calculate the numbers in table 1, the data cleaning and what contributes to the calculation. The values have been calculated based on the 'Solution' fields of one month of incident data. The 'Description' field is not measured due to often (partially) consisting of machine-generated data.

Table 1 Measurements of data quality

Dimension	Definition: the extent to which…	Calculated as	Value
Completeness	data is of sufficient depth, breadth, and scope for the task at hand	% of fields filled in	78.2%
Interpretability	data is in appropriate language and unit and the data definitions are clear	% of fields in English	83.3%
Relevancy	data is applicable and useful for the task at hand	Number of terms used for the concept 'false positive'	~10

The calculation of relevancy needs an explanation. This measurement is not necessarily a measurement representative of the entire dataset, but rather an exemplary measurement that shows how the data can be of poor quality keeping in mind the goal of building a solution recommender. A more precise measurement would be to do some form of grouping based on the entered data and to see whether tickets that should be grouped together actually are. However, considering the time frame of this study, this is deemed out of scope.

There are no strict rules or norms on what the values of the above indicators should be. However as one can see, the current values are unsatisfactory. At least 21.8% of the data is unusable, 16.7% requires translation or will also be unusable, and there are approximately 30 terms used for the same, relatively generic concept.

There is some overlap in these groups (for example, "*opgelost"* is in both non-English and empty fields), but this is considered insignificant (498 entries in total, which is about 1% for both completeness and interpretability). In a preferred situation, completeness is at least 90%, the percentage of fields in English is at least 95%, and the number of terms used for 'false positives' would be 5.

Important to notice here is that the required fields for a solution recommender, those being the user rating for actions to solve the incident (see Appendix A) are not included as a measurement. It is known that this is currently not included in the incident registration, and one of the key challenges of this research is finding a way to do this.

2 Problem-solving approach

In this section, the approach towards solving the core problem is defined, based on the given main research questions. Besides this, knowledge problems of different stages of this approach are determined, as well as the deliverables that this report contains.

2.1 Main research questions

To solve the core problem mentioned in Section 1.4, two main research questions are defined. These questions highlight the different aspects of the core problem. These main research questions are not the only research questions that are answered (see section 2.3). The two main research questions are highlighted because they clearly distinguish two aspects of the core problem, which the other knowledge problems do not. The first research question highlights the need for data quality improvement in order to build a solution recommender.

How can a data channel be designed in order to collect data for a solution recommender, keeping in mind the ease of implementation and user-friendliness?

The second question highlights the need for an overall improvement of the quality. This is useful for operators as this allows the search engine within Service Manager to be more useful than it currently is.

How can the overall data quality of the incident registration in Service Manager be improved?

2.2 Approach

To solve the core problem of low data quality systematically, the MPSM is applied as a basis (Heerkens and Winden, 2021). The MPSM distinguishes 7 phases:

1. Defining the problem
2. Formulating the approach
3. Analysing the problem
4. Formulating solutions
5. Choosing a solution
6. Implementing the solution
7. Evaluating the solution

To keep in mind the scope and goal of this study, as described in the intended deliverables in Section 0, this method is slightly adjusted. A new approach (Figure 3) has been made based on the MPSM. In Figure 3 the various stages of the approach are highlighted as well as the corresponding chapter in which they are found in brackets.

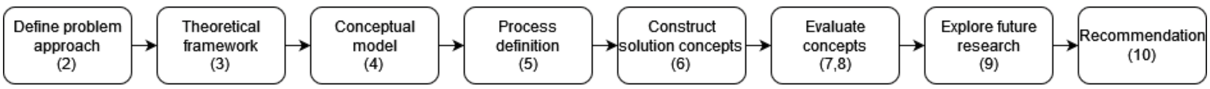

Figure 3 Problem approach

2.3 Knowledge problems

During these phases, different knowledge problems have to be answered. Under each phase of the problem approach are questions that must be answered in order to successfully arrive at a recommendation.

1. Defining problem approach
    1.1. What knowledge problems occur in the various phases of the approach?
    1.2. What deliverables must be delivered?
    1.3. What is the research design for the various deliverables?
2. Theoretical framework
    2.1. What is the best way to model the business process?
    2.2. How can data be generalised?
    2.3. How can the knowledge transferring be improved?
    2.4. How can the concept of 'data quality' be delineated?
    2.5. How can automation be used to improve the process?
3. Conceptual model
    3.1. Based on the theoretical framework, how can data quality be improved?
    3.2. How do the various concepts influence each other?
    3.3. Which steps should be automated?
    3.4. How are degrees of freedom and usability of data fields related to each other?
        3.4.1. How much can the concept of 'solution' be generalised?
4. Process definition
    4.1. What does an operator do when an incident occurs?
    4.2. How does an operator interact with the anomaly detector?
    4.3. Where would the solution recommender be in a preferred process?
    4.4. In what stages is information exchanged?
    4.5. What are the assumptions of the model?
5. Constructing solution concepts
    5.1. What are current state-of-the-art solutions to similar problems?
    5.2. How do the concepts influence the operator?
    5.3. How can the currently used software be used to improve data quality?
    5.4. What is possible within Service Manager?
6. Evaluate solution concepts
    6.1. What are criteria to evaluate the concepts?
    6.2. How can the concepts be evaluated as objectively as possible?
    6.3. What are the strengths and weaknesses of the solution concepts?
7. Explore future research
    7.1. What other potential does AI have within Rabobank?
    7.2. What are challenges to the opportunities?
    7.3. How can challenges be mitigated?
8. Recommendation
    8.1. Based on the evaluation, what is the best solution to the core problem?
    8.2. What steps can be taken to implement the improved process?
    8.3. What should be the direct next step for future research?

2.4 Deliverables

Four key deliverables are presented as a result of this study. These deliverables were obtained in different stages of the problem approach and serve different goals. In this section of the report, the four deliverables and the research designs used to create the deliverables are defined.

2.4.1 Process description

The first deliverable is **a description of the current process regarding the incident and incident registration processes**. This is necessary to adequately analyse what can be improved and how improvements should be implemented. This also emphasises the usability of the study from a managerial perspective. Overall, an overview of the different processes is a great way to give a clear overview of what is currently being done and how it should preferably be done.

This deliverable has the goal of defining the current processes. This means that the research is of the qualitative type as the research “seeks to develop understanding through detailed description” (Cooper et al., 2014, p. 146). Subjects of this research are the managers that have a clear overview of the incident process, as well as the operators that have to work with the registration process on a daily basis.

Table 2 Research design process definition design

Research question	**What does the current incident (registration) process look like?**
Type of research	Descriptive research with the goal of defining the current process
Research population	Rabobank’s engineering employees,
Subjects	Process owners, operators, and AIOps team members
Research strategy	Qualitative
Choice of data gathering	Interviews, existing data; interviews are necessary since interactivity is required to get all the data.
Choice of data analysis	Collected data is summarised in a process definition, according to BPMN.
Limitations of design	Reliant on willingness and knowledge of interviewees.
Validity, reliability	Considering that the process should be uniform, the reliability of this research should be high. Validity is highly dependent on the subjects of the research. If the subjects have sufficient knowledge about the process, as is expected then the validity is high. Since the model will always be an interpretation of some sort, the validity and reliability of the models is increased by relaying them to the interviewees and asking whether they feel the model is accurate, readable, and complete.

The interviews are semi-structured. There is a clearly defined goal and based on this goal, several questions are being kept in mind. However, the interviewee must have the freedom to ask questions as well, as this increases the accuracy of the received information. Besides, it is likely that the process is not immediately clear, meaning that follow-up questions need to be asked. A semi-structured interview allows this freedom, while still ensuring that the goal is reached. To efficiently collect information in a relatively short time, an interview protocol has been constructed. This protocol and its justification are to be found in Appendix B.

2.4.2 Solution concepts

The second deliverable consists of **different concepts on how the data quality can be improved**. The purpose of this deliverable is to show stakeholders where progress can be made on the data quality and what technology or methodology can be used to achieve this, keeping in mind the goal of taking the next step in the utilisation of AI analyses.

The goal of the second design is to ideate: getting different ideas on how to improve the incident logging process, taking into account the goal of utilising incident data to improve the recovery process. Investigating a wide array of case studies is used in order to gain multiple ideas on how the process can be improved.

Table 3 Research design solution concepts

Research question	**What solutions can be constructed to improve the data quality?**
Type of research	Descriptive with the goal of finding solution concepts to the core problem
Research population	Companies
Subjects	If possible, in the finance field
Research strategy	Qualitative
Choice of data gathering	Literature study; to see case studies, related studies and to obtain ideas to answer the research question. Field study is not doable within the timeframe and scope of research
Choice of data analysis	Qualitative, data is analysed to determine requirements, opportunities and relationships.
Limitations of design	It is possible to get too focussed on the ideas already proposed by existing literature and in doing so, reject good ideas preliminary.
Validity, reliability	Apart from this, as explained, there is the danger of rejecting good ideas preliminarily due to already having found some ideas in the literature research. Another possibility is having some sort of sunk cost fallacy, where time has been invested in finding a solution and therefore, it must be considered whether it is a good solution or not (Haita-Falah, 2017).

2.4.3 Evaluation of the various solution concepts.

To address the (dis)advantages of the various solution concepts, a design must be made to ensure that the concepts are evaluated in a method that is as objective as possible. Considering the scope and time frame of this study, a truly objective design in the form of experiments is not possible. The time to implement the solutions, to get operators to properly use them, and to ensure that enough relevant data is gathered is considered too large. Therefore, the choice is made to apply questionnaires.

Table 4 Research design evaluation

Research question	**What are (dis)advantages of the various solution concepts?**
Type of research	Descriptive, with the goal of evaluating solutions
Research population	Rabobank's engineering employees in the Payments domain
Subjects	1. The AIOps team, those who have to implement the AIOps-specific solution 2. The SWIFT team, a team of operators that does not use AIOps 3. The IP team, a team of operators that does use AIOps 4. A group of solution architects
Research strategy	Quantitative
Choice of data gathering	Questionnaires are used to collect qualitative data in a time-efficient manner that produces qualitative data to analyse. To ensure that the relevant context is known to those filling in the questionnaires a presentation is given beforehand.
Choice of data analysis	Quantitative, data is analysed to determine perceived quality of the solution concepts.
Limitations of design	The main limitation of this design is that a recommendation is given based only on the perceived quality as judged by humans, as opposed to experiments which produce more objective data. However, the choice for questionnaires is made due to the difficulties that come with experiments regarding authorisation, producing enough data in a limited timeframe, and the technical aspects of implementation.
Validity, reliability	In an effort to increase validity and reduces biases, different groups of people have filled in the questionnaire.

The questionnaire is filled in by four groups that can be expected to have the required knowledge to properly assess the concepts. The AIOps team is chosen, since they have to build and implement the solution recommender and are responsible for the logic behind it. Besides this, two teams of operators are chosen. The first is the SWIFT team, who do not use AIOps and therefore are not (yet) impacted by the AIOps-specific solutions, as opposed to the non-AIOps-specific solutions which they are affected by. The IP team do use AIOps and therefore is affected by both types of solutions. The group of solution architects have the required knowledge, but will not directly work with either solutions and are therefore an important group of people due to a supposed lack of bias.

2.4.4 Future research exploration

The third deliverable is **an exploration for future research**. This deliverable aims to shed light on different possibilities of the improved data, what steps to take to utilise it, and where future research can be done to explore more opportunities. This deliverable entails different opportunities, their benefits, and what challenges have to be taken into account. To keep scope in mind, this deliverable only concerns AI-related projects.

This deliverable can be seen as a by-product of the other deliverables. Considering that the AIOps projects (and similar projects) of Rabobank are heavily rooted in innovation and research, it is important to explore where such innovation can take place as well. Especially in the interviews that were a part of the problem approach and the process definition, a lot of different problems or opportunities were encountered. This deliverable aims to shed a light on these opportunities, and conducts a literature study on them.

Table 5 Research design future research

Research question	**How can Rabobank use AI to improve processes?**
Type of research	Descriptive, with the goal of finding information about opportunities that are encountered in previous stages.
Research population	Companies
Subjects	If possible, in the finance field
Research strategy	Qualitative
Choice of data gathering	Literature study is used. Specifically, literature reviews of other research can efficiently give information on the respective subject in a structured manner.
Choice of data analysis	Qualitative, data is analysed to determine opportunities and challenges.
Limitations of design	This design is limited in that literature does not necessarily reflect the status within Rabobank and it will therefore not give the entire picture. However, concerning the explorative nature of this deliverable, and the timeframe of the study, this is deemed acceptable.
Validity, reliability	Especially when a literature study by other researchers is used, there is the risk of being biased due to the biases of the original researcher. Therefore, different studies need to be compared. Besides, the biases of interviewees will have some influence on the highlighted opportunities.

3 Theoretical framework

To assess the problem statement one needs a theoretical background (Cooper et al., 2014). In this segment, various theories are discussed which help define different concepts or propose theories to solve the problem. A summarisation of the theoretical framework and its key takeaways is to be found in Appendix D Summary of the theoretical framework.

3.1 Data quality

First of all, the variable *data quality* is too vague and broad of a concept, both to measure and to understand how to influence it. Hassenstein and Vanella (2022) note how few uniform definitions or established frameworks there are. To delineate this concept, a literature review is performed. The knowledge problem that this literature review aims to answer is: *"How can the variable 'data quality' be delineated?"*.

The literature review brings us to the following answer to the knowledge problem. Various theories regarding data quality exist. Batini and Scannapieco (2006) discuss a multitude of these theories. To divide the collective term data quality into smaller pieces the empirical approach, proposed by Wang and Strong (1996) is used. They discuss four categories which are further divided into several dimensions, as opposed to the intuitive approach proposed by Redman (1996), whose approach only contains one level of classification. Table 2 shows the category, the relevant dimensions, and its definitions. The reason this approach is used is due to its two-level classification, giving a clear overview of different divisions and subdivisions, and its use in multiple other studies (Melkas & Harmaakorpi, 2008).

Table 6 data quality model proposed by Wang and Strong (1996)

Category	**Dimension**	**Definition: the extent to which…**
Intrinsic	Accuracy	data is correct, reliable, and certified free of error
	Believability	data is accepted or regarded as true, real, and credible
	Objectivity	data is unbiased and impartial
	Reputation	data is trusted or highly regarded in terms of its source and content
Contextual	Completeness	data is of sufficient depth, breadth, and scope for the task at hand
	Value-added	data is beneficial and provides advantages for its use
	Relevancy	data is applicable and useful for the task at hand
	Timeliness	the age of the data is appropriate for the task at hand
	Appropriate amount of data	the quantity or volume of available data is appropriate
Representational	Concise representation	data is compactly represented without being overwhelmed
	Representational consistency	data is always presented in the same format and is compatible with the previous data
	Ease of understanding	data is clear without ambiguity and easily comprehended
	Interpretability	data is in appropriate language and unit and the data definitions are clear
Accessibility	Access security	access to data can be restricted and hence kept safe
	Accessibility	data is available or easily and quickly retrieved

Intrinsic data quality captures the quality that data has on its own. Challenges that might occur here are data from a questionable source, incorrect data, or heterogeneous data (Cho et al., 2021). Contextual data quality considers the context in which data is used. Considering the background of this study, this is a key category, as the context of anomaly detection and Artificial Intelligence is a complicated one. For the same reason, representational data quality, the degree to which the data can be read, understood, and presented, is a key category. A human can fairly easily filter out unnecessary information, but for AI this is much more difficult. Accessibility is about who has access to data and how easy and quick this access is.

For the scope of this study, the different categories and dimensions differ not necessarily in their importance, but rather to what degree they must be taken into account. This can be illustrated using the category of accessibility. Security is already at a very high level at Rabobank and any solution to the core problem may not diminish this security but does not need to improve the security either. Next to this, the data is coming from a reliable source: the operators that are employed at Rabobank and that are trusted to resolve the incidents. This means that believability, objectivity, and reputation become less relevant dimensions. The most important categories in this study are the contextual and representational quality. This literature research has given us a framework on how to both measure and improve data quality.

3.2 Knowledge management

As Machado and Davim (2021) state "knowledge management is understood as the process of creating, sharing and managing the information and knowledge of an organisation." They claim that knowledge is a critical tool that contributes positively to performing jobs efficiently and effectively. Such is also the case in this study. The transfer of the knowledge of the operators about the incidents must be properly documented and made accessible to the AIOps team. They build a model, and the output of the model gives new information back towards the user, which can increase the effectiveness and efficiency of their work.

Key to successful knowledge management is the creation of a context of understanding between the transfer partners. (Dove, 1996). Darr and Kurtzberg (2000) describe how the strategic similarity between two different domains positively influences the capability of transferring knowledge. Fortunately in this study, the domains between which information must be exchanged have similar strategic goals, since both teams want to reduce the number of impactful incidents. It is therefore important to stress this similarity when implementing an improved process. This also alleviates the fear of change that people often have. Informing users and giving them a voice is of major importance when one requires their cooperation (Umiker, 1997).

3.3 Business process modelling

Since the processes will have to be defined, a certain notation must be applied to express the model. There are various options of which to choose from, such as the Petri-net-based modelling languages and the UML activity diagrams (Booch et al., 2005; Girault & Valk, 2003). Both of these methods provide advantages and disadvantages in their use (Weske, 2019). For this study however, the Business Process Modelling Notation (BPMN) is used (Object Management Group, 2006). One of the reasons behind this is that the BPMN has the aim of providing different levels of abstraction, meaning that both the technical level and the business level can be represented and understood, which is useful in this study due to the combination of information flow and business processes. (Weske, 2019). It must be noted that the BPMN 2.0 has been developed, which allows the inclusion of the Business Process Definition Metamodel (Object Management Group, 2010). For the purpose of this study, there is no distinction between the two languages, due to their similarity, unless explicitly mentioned.

Apart from this, the BPMN has a simple set of core elements, which can be expanded on if the modeller prefers to. This makes it an intuitive tool to use, which focuses on the applied value of the modelling language, not on the restrictions of a language. The BPMN also provides the utility of swimming pools and lanes (Object Management Group, 2006). These give the modeller the option of easily differentiating between different parties. To summarise, the BPMN provides a modelling language that is flexible rather than restrictive, providing great usability to the modeller.

3.4 Data generalisation

Data generalisation is the concept of summarising data by changing relatively low-level values into relatively high-level concepts (Han et al., 2012). An example would be to change a time value from the exact time of day an incident happened into the day the incident happened. There are different purposes and methodologies for data generalisation (Petry & Zhao, 2009). Typically data generalisation is used in data mining, however, for the purpose of this study, one can apply data generalisation for data acquisition. For the scope of this study, data generalisation can be useful to group different actions or processes that are undertaken into more general and more useful values, which can more easily be analysed.

One of the key challenges with regard to data generalisation is the balance between data becoming useful through summarisation and the loss of specificity (Han et al., 2012). Although there is not one specific solution to this problem, multiple approaches have been proposed with the utilisation of concept hierarchies (Yager & Petry, 2006). An example of such a concept hierarchy would be the generalisation of postcode into city, province, country, and continent respectively. This makes the balance between specificity and generalisation more navigable.

3.5 Lean management

Lean is the concept of streamlining an operation or process in such a manner that the amount of waste is minimised (Slack et al., 2016). This philosophy, although originating from the manufacturing industry has expanded towards different industries, including the finance and IT industries (Durham & Michel, 2021).

With regard to this study, one must take lean management into account when configuring solutions to the problem. Specifically, the process must be automated where this is deemed appropriate. This ensures an adequate speed of the process, but more importantly, reduces the barrier for the users who must provide the input data. The easier the action becomes for them, the more they are inclined to perform it (Umiker, 1997).

It is also interesting to discuss lean management in the context of the solution recommender. Assuming that an effective solution recommender can be built, it logically follows that this will increase the degree to which the process is lean since the increased efficiency in resolving incidents means that fewer resources are wasted in the incident process.

4 Conceptual model

This section of the report discusses the conceptual model of relationships that is used when designing solution concepts, based on the theoretical framework defined in chapter 3 (see Appendix D for a summarisation).

As has been discussed, data generalisation can increase the relevancy of the data, as defined by Wang and Strong (1996), but at the cost of decreased accuracy (Han et al., 2012). When designing solution concepts, this must be taken into account. There must be some generalisation to make the data useful, however, the accuracy of the data must not be compromised too much.

Logically, unifying the language increases the interpretability of the data quality. Since a relatively large part of the data is currently in non-English, as is described in section 1.5, large parts of the data are mostly useless in the current context. The increased interpretability, if the language were unified, increases the data quality of the incident logging process.

Next to this, the degree of automation is a significant influencing factor. The first reason is that automation both increases the efficiency and therefore the leanness of the process (Slack et al., 2016). Apart from this, automation makes the registration process easier, increasing users' willingness, meaning that the quality of data will be increased. Besides, automation means that data will be more uniform, making it easier to analyse.

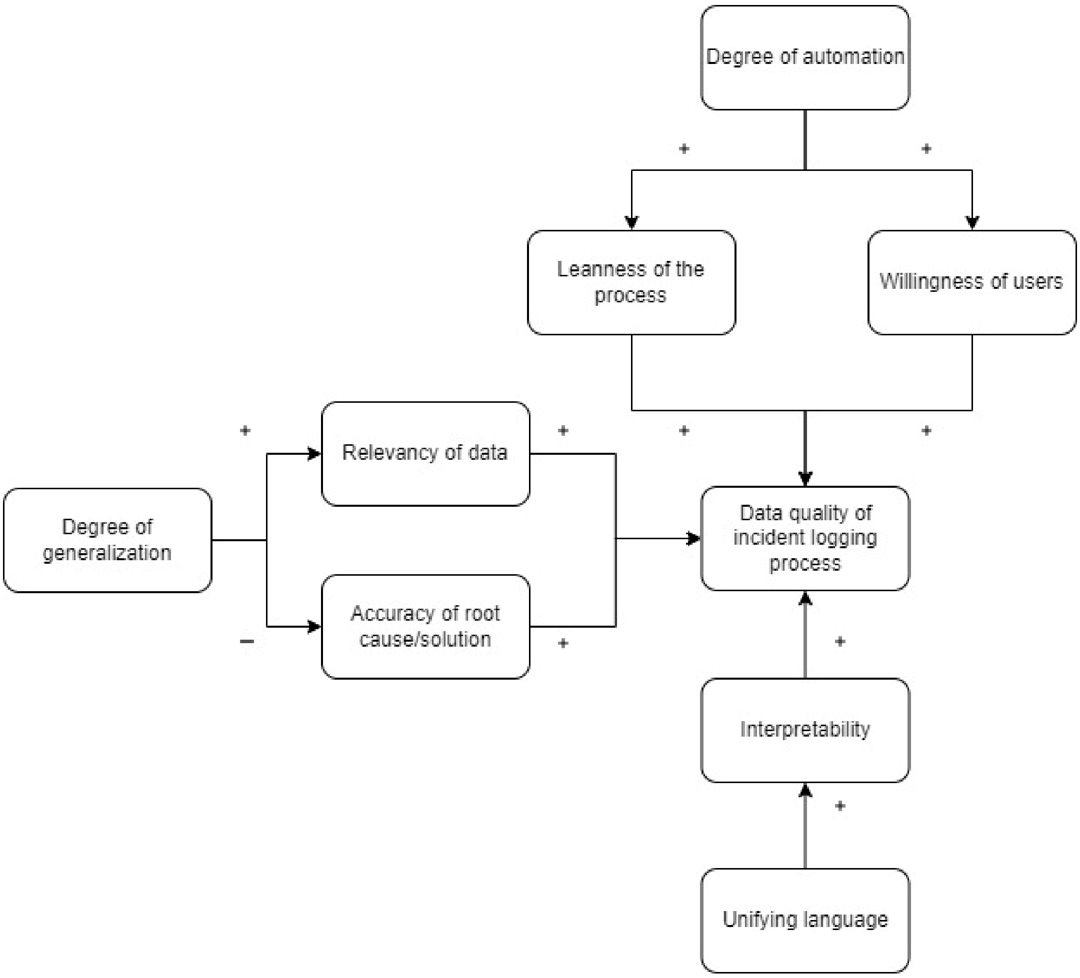

Figure 4 Conceptual model

5 Process definition

To assess the core problem of low data quality, it must first be determined where and when data is gathered, where the solution recommender would be implemented and what this implies for the work of an operator. To do so, two process descriptions have been made, based on interviews (see section 2.4.1) with, but not limited to, process owners, operators, and the AIOps team. As a result of these interviews, the preferred model is defined. This model is an interpretation of the information gathered during the interviews, taking into account the scope and goal of this study.

In Figure 5 and Figure 6, respectively the current and preferred process definitions have been noted using the BPMN as a result of interviews with different groups within Rabobank (Object Management Group, 2006). These models have two purposes: first, to provide a background on the context and stakeholders of this project. Second, the models are used to identify where in the process the necessary information flow would be. This is mainly to illustrate the missing information stream that is necessary for the solution recommender.

A useful property of the BPMN are pools and swimming lanes (Weske, 2019). These give a clear distinction between different groups of people. In this model are two pools, those being Rabobank and external stakeholders. The purpose of the external stakeholders here is mainly to stress the communication of Major Incidents, and as such the internal processes of different stakeholders are not modelled. Rabobank's relevant parties are modelled in three lanes, those being the anomaly detector, the team of operators and the Major Incident Management Service. It is important to note here that the anomaly detector is actually used by the operators, yet, according to agile principles, is owned by AIOps and maintenance and improvements are conducted by AIOps (Barton, 2018). As such, it cannot be said that AIOps does the tasks in this model, but rather that the anomaly detector does these tasks. The last swim lane is the Major Incident Management Service, which is reached whenever an incident is deemed to be of high priority. Such incidents have a higher impact on customers or other systems, leading to a financial or reputational loss and/or a higher urgency because the system cannot be offline for a long time (Rabobank, personal communication, 2022). These situations are direr, but also more unique, both in their similarities and in occurrence. This is why the incidents with priorities 1 or 2 are not taken into account for the solution recommender.

As is adamant from the differences between Figure 5 and Figure 6, there is a clear lack of desired communication. First, this means that the anomaly detector has no proper feedback on its detected anomalies. If it fires a false positive and is not notified that it did, then it might fire for the same false positive again. Apart from this the necessary knowledge transfer between the anomaly detector (in this case AIOps) and the operating team is modelled in the preferred model, as this link is missing in the current model.

The preferred process model is made using the following assumptions.

1. Determining that there is no actual problem (i.e. a false positive) still counts as 'Resolving' the incident, for the sake of simplicity.
2. Workarounds are not taken into consideration, as the actual solution is of greater importance for this study.

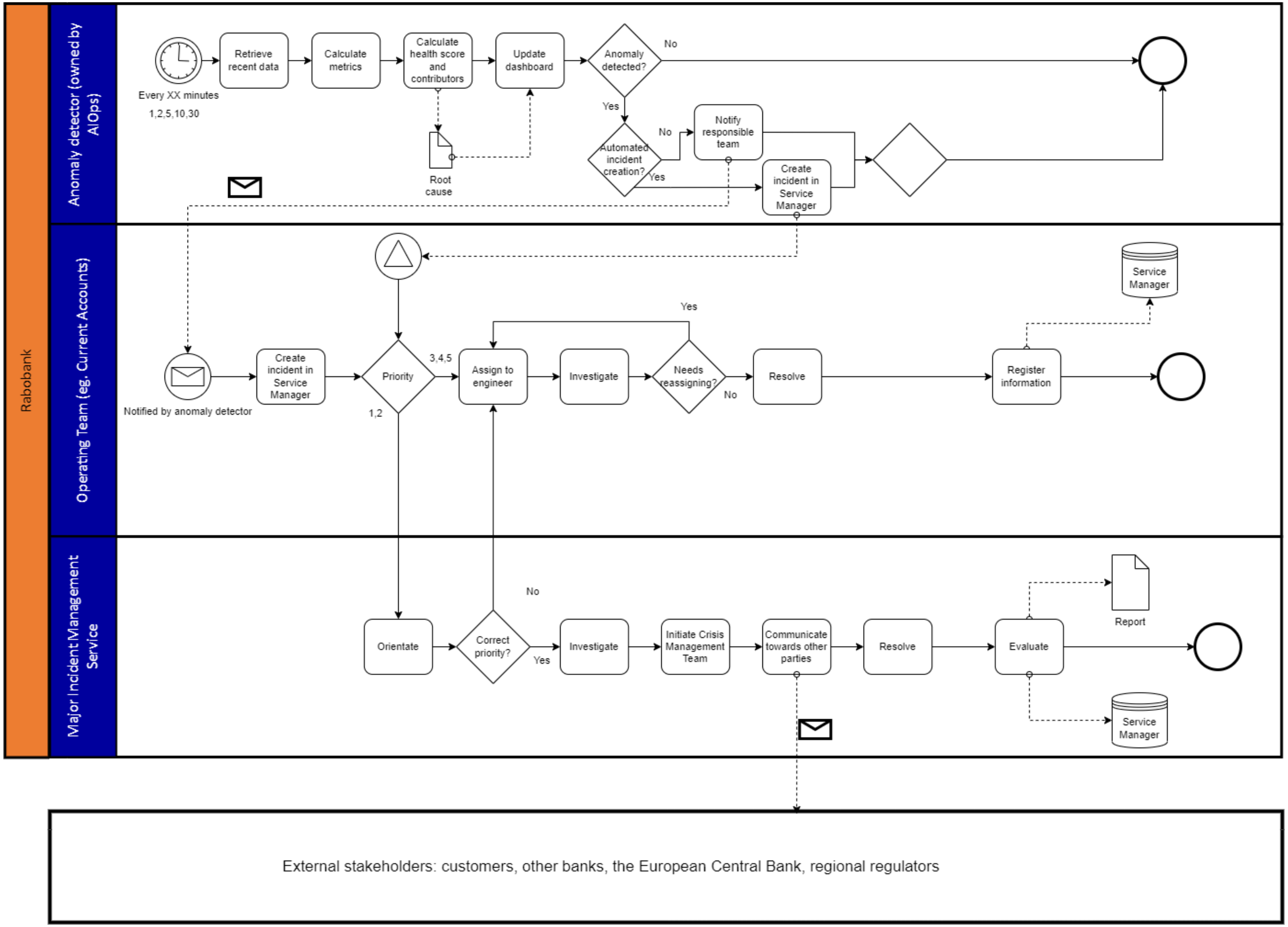

Figure 5 Current process definition

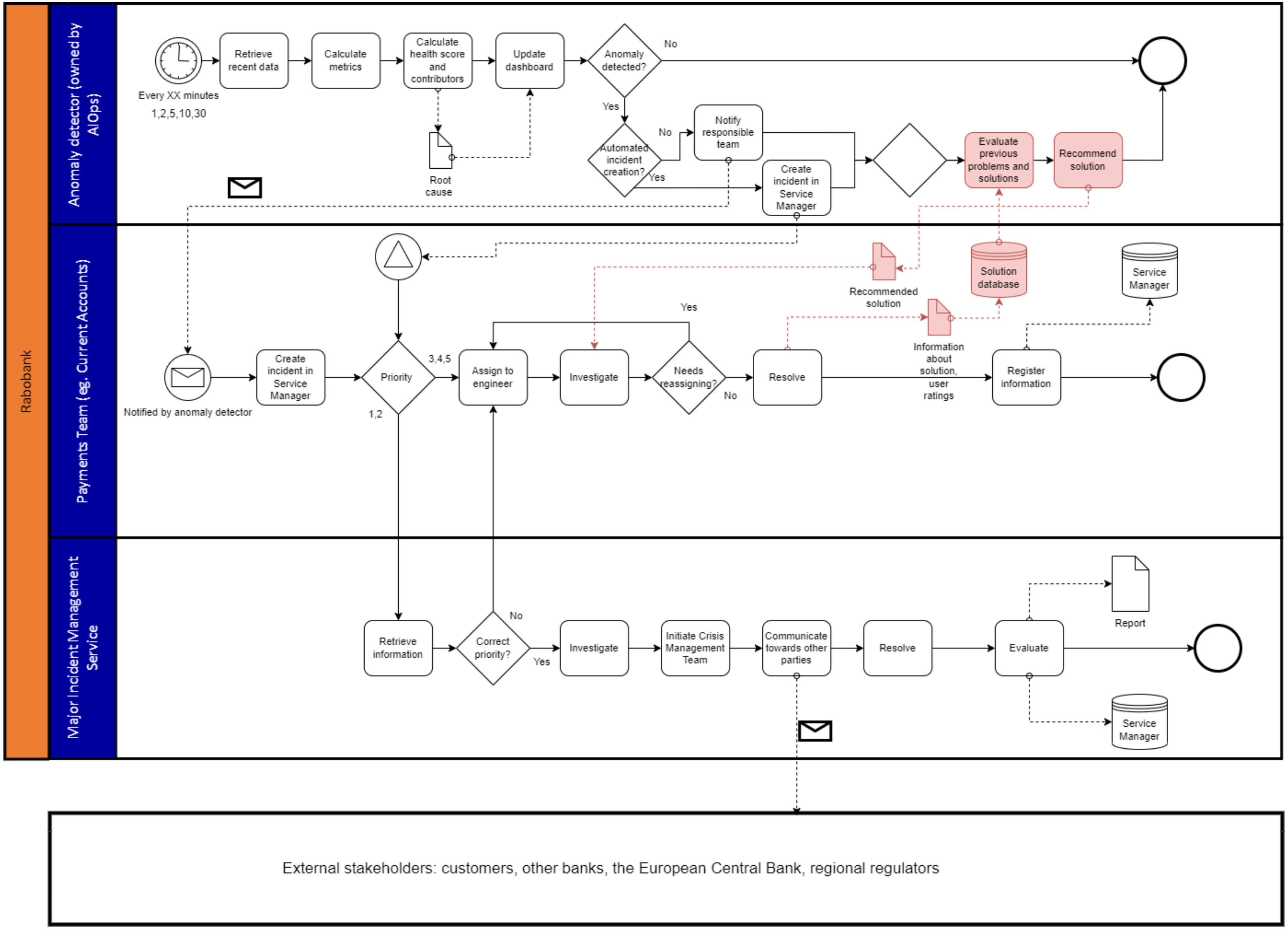

Figure 6 Preferred process definition

6 Solution concepts

To tackle the core problem of insufficient data quality, two different levels of solutions are considered, in which different concepts are presented as a result of the previously conducted interviews for the process definition and literature study (see section 2.4.1 and 2.4.2). These levels can simultaneously be seen as a distinction based on scope and scale of the solutions. Solutions can be distinguished between AIOps-specific solutions and non-AIOps-specific solutions, where the first group can only be used by teams making use of AIOps, and the second group can be used by all teams. It is necessary to consider all types of solutions, even though during implementation, one group may be skipped entirely if it is deemed inefficient to implement it.

Apart from this, the different dimensions can be distinguished based on goal. The goal of the first solutions is to deliver a way for operators to rate different actions on how effective they were in solving the problem, so a recommender system can be built. Specifics on how the recommender system would function are to be found in Appendix A. The goal of the second dimension is to improve the overall data quality. This can increase the utility of the search engine already in place in Service Manager.

All solution concepts are the result of the aforementioned literature study and interviews. Key while designing the solution concepts was to keep solutions relatively straightforward to use, especially for operators. This means that the considered software is mostly software that is already used by operators. Apart from this, the proposed improvements to Service Manager (non-AI specific solutions) are kept relatively straightforward, as Service Manager is a relatively restrictive software.

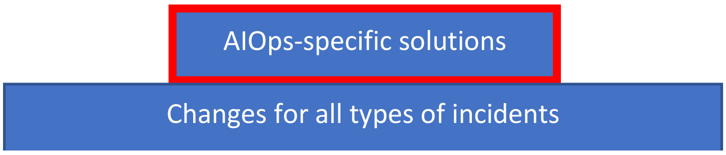

Figure 7 Visualisation of dimensions of the solution concepts

6.1 AIOps-specific solutions

As Figure 7 indicates, in this section of the report, the smallest scale is first considered: designing a solution such that the AIOps-team can start building the recommender engine as soon as possible without necessarily having to wait for changes on the larger scale, while simultaneously testing how suitable the implementations are for when a solution recommender is implemented companywide.

As became clear when analysing the difference between the current process and desired process (see chapter 5), there is a lack of information flow between the different departments of Rabobank. The solutions discussed in this section aim to solve this problem by providing a way to deliver the required data for a recommender system (Appendix A).

6.1.1 Pure data mining

The first solution is pure data mining. The idea behind this solution is not to change the incident registration process at all, but instead to work with the data that is currently possessed, and to mine the solution logs to gain valuable information using Natural Language Processing (NLP), without gaining user input through recommendation ratings. There have been written several studies that use this approach. For example, Zhou et al. (2016) build on the k-nearest-neighbour algorithm, which is also used by Tang et al. (2013), using other algorithms and similarity measures to relate events to historic events, under the assumption that if the events are similar, the solution will be as well. They analyse the text fields using the Jaccard index for the bag of words model, that being the proportion of shared words in two texts (Manning et al., 2008). Recent development in this field is the multi-view similarity measure framework proposed by Xu et al. (2020) that can be applied to clustering algorithms of solution recommenders.

Zeng et al. (2017) use a hierarchical multi-label classification to sort events in different parts of a hierarchy using a variety of methods. Among other uses like problem diagnosis and ticket assigning, this model can be used to recommend a solution based on the idea that the tickets can have a solution that addresses different hierarchical levels. Heras et al. (2008) also propose a multi-domain module for customer support tickets using case-based reasoning.

Apart from this, Ali Zaidi et al. (2021) developed an end-to-end framework to suggest resolution actions, as opposed to recommending free-form resolution text, which they observed outperformed the similarity search-based methods. They use resolution actions, specific actionable steps to undertake, extracted out of the resolution text, achieved by mining phrases based on a model proposed by Wu et al. (2020). This means that instead of one single resolution text, the recommendation consists of a multisequence set of actions. This is done using semantic role labelling (assigning roles such as *verb*, *agent*, and *recipient* to parts of a sentence), followed by filtering the resulting set of actions (Màrquez et al. 2008). Out of these actions, recommendations for action sequences are made.

Zhou et al. (2017), who propose a System for Ticket Analysis and Resolution (STAR), give another insight. They propose a model that integrates the difference in quality between different ticket resolutions, specifically giving the example of "resolved" as a low-value resolution text. They found that for a typical ticket 33 features can be categorised into four groups, those being: character-level (ratio of exclamations, colons, punctuation etc.), entity-level (number of dates, file paths, percentages, etc.), semantic-level (ratio of verbs, adjectives, etc.), and attribute-level (length of resolution, Jaccard similarity between summary and resolution) features (Manning et al., 2008; Zhou et al., 2016). They then evaluated three popular regression models on their test data to evaluate the features (Alpaydin, 2010). In this evaluation, they found that the semantic features have the advantage over the other features, as well as that the length of the resolution is positively related to the quality of the text. A quantifier is used to assign a quality measure to the ticket which is then used in the solution recommender system.

Last, a study by Güven et al. (2016) explores a causal link between changes of the system configuration and incidents. Since 80% of incidents that present client outages are reportedly caused by changes, this causality implies that one can both predict incidents and propose a solution to the incident (Scott, 2005). Especially when the client is affected, simply reversing a change means that although the problem is not completely solved, the client impact is mitigated at least. In tests of Güven et al. (2016), their system was able to predict the change responsible for an incident within 5 shown results 75% of the time. Apart from this, 58 % of the time, the top predicted change is the actual responsible change.

If the presented studies are applied to this study then it follows that the STAR model proposed by Zhou et al. (2017) fits the problem. The inclusion of the quantification of the quality can help alleviate the problems presented by the poor data quality, without disturbing the current registration process, on the side of the operator. The relevant data will be gathered from Service Manager, where the registration takes place, and the tickets will be analysed according to similarity measures, for example, a variety of the k-nearest-neighbour (Zhou et al., 2016). Recommendations are made based on the textual information in the description of the incident, and the similarity of a problem to others, without user input. However, the feature extraction from Ali Zaidi et al. (2021) and the linking of changes proposed by Güven et al. (2016) possibly provide value, and can be combined in a single system.

This solution concept presents the advantage of not having to change any behaviour of the operators, or of the registration system. This does however come at a cost. One problem with this solution concept is that it does not inherently solve the problem of poor data quality. Rather, it gets poor quality input, accepts this, and accommodates for this. It brushes up the problem, but does not solve it. This goes hand in hand with high-quality data being a core requirement of good information systems (Olson, 2003). The lack of proper data quality means that the potential of the recommendations is decreased from the start.

6.1.2 Automatically send Forms

Another solution concept is to automatically send teams a form whenever an anomaly occurs. This can mostly be done using the infrastructure that is already in place. There is already a channel that sends a notification to the team whenever an anomaly occurs. Since Rabobank already uses the Microsoft package for teams' daily operations (Word, Teams, Outlook, SharePoint, etc.), Microsoft Forms can be used in combination with Power Automate and Azure DevOps, the latter of which is the specific tool used by DevOps teams at Rabobank (Microsoft, 2022a; Rabobank, personal communication, 2022). Harinarayanan (2021) describes the uses of (among others) Power Automate, and the value that it provides to businesses.

Advantages of this solution concept are mainly in two components. First of which is that this solution concept exclusively makes use of Microsoft products, meaning that a lot of the workflows can be automated easily to fit better into the natural workflow of the AIOps team and the operators' teams, without having to radically transform the data. The other benefit is the versatility and flexibility of the Forms. Forms are easy to change and provide a range of options for different types of fields to collect the data, such as user ratings.

The greatest downside to this problem lies on the side of the operators. This comes from two directions. First, this solution will probably be perceived as annoying or wasteful, since it disrupts operators' workflow to some degree and they must register information twice: once in the Forms, and once in Service Manager. Even though this information is not necessarily the same, the perceived inefficiency might be harmful to the implementation. Apart from this, if an operator is uninterested, forgetful or in a hurry, Forms are easily ignored. This can be remediated by emphasising the value of Forms to the operators themselves, or 'hiding' the recommendation in the Forms to encourage the operator to at least open it, but these are just coping measures and the latter might even increase unwillingness if done uncarefully.

In this solution, the data collection is in Forms, which can easily extract information to Excel (if even necessary) after which the data can be exported in CSV, XML or event TXT format.

6.1.3 Azure DevOps boards

In order to work conform the agile working style (see section 1.1), Rabobank uses Microsoft Azure to (among other things) keep track of teams' DevOps progress using scrum boards (Barton, 2018; Microsoft, 2022b). This study will not digress too far in the entirety of scrum, but it is important to know that a team has a board, on which can be a variety of work items (majorly user stories) and several columns to keep track of work items, as has been shown in Figure 8. Teams use these boards daily to update their progress on several tasks. Azure DevOps has been widely adopted to suit this purpose (Santos Júnior et al., 2021).

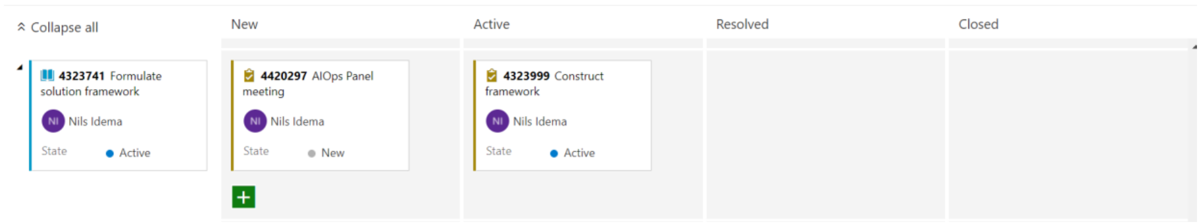

Figure 8 Example of Azure DevOps board

What is interesting about both the boards and the user stories is that they are customisable (Microsoft, 2022c.). For this study specifically, work items can be customised to fit the form needed: where the title contains information about the anomaly, the top contributors and solution recommendation are on the card and there are fields for the actual solution, undertaken actions, and user ratings. Work items can easily be exported to CSV files, which are easily exported to other applications or tools.

Apart from this, Azure DevOps uses the feature of OneClick Actions, where actions (e.g. set field value, link to existing item) are automatically undertaken whenever a trigger fires (e.g. a click, field changed, upon creation). This helps streamline the process and thus makes it more lean. Once again, Azure is a product of Microsoft, meaning that the Power Automate can be used to automate certain steps. This can be useful if work items are to be created automatically, or the data is to be sent immediately when closing the incident.

There are two variations to this solution concept. The first is to create a single board for all the teams to use. The advantage of this is that the customisation of the board allows for greater versatility and flexibility for the AIOps team. This makes automated rules easier and gives fewer problems with authorisation on the boards of other teams. This does come at the cost of the operator teams suddenly having to use two boards, one of which is not in their daily operations. Apart from this teams will likely be able to edit each other's work items, which complicates the authorisation issue further.

The other variation is to utilise the already existing boards of the operator teams. This restricts the versatility and flexibility since the team must still be able to use their 'normal' board. Apart from this, since most boards are inherited from a template board, a new type of board must be created, possibly disrupting the boards currently used by the teams. Authorisation is once again an issue, since the AIOps team is not the owner of these boards, permission must be given to them to edit these boards. It does become easier for operators however since they can keep using their own board.

In this solution, the data collection is in Azure DevOps, which can extract information to Excel or even directly to CSV files, after which automated processes can export the data to Azure, or another cloud-based database.

It must be noted that Rabobank has expressed that certain boards will be merging in the near future, meaning that certain projects will have their boards combined. This is something that must be taken into account if this concept were to be implemented.

6.1.4 Splunk dashboard

The main idea behind this concept is to use the already existing Splunk dashboard to gain the necessary user input. The greatest advantage of this solution is its intuitiveness. Whenever an operator is notified of an anomaly, they are expected to go to this dashboard to see the top contributors to the anomaly, which is to serve as a root cause identification. If this solution were to be implemented, then all of the relevant information will be available to the operator in one place.

When they have solved the incident, they can document the solution and the recommendation rating in input fields on the dashboard.

The difficulty of this solution is whether it is feasible or not. Splunk is a tool that is mainly used to process, analyse, and visualise machine-generated real-time data (Splunk, 2022). The only instances where user input comes into the picture is for dynamic dashboarding, that is, to present different dashboards based on the user input. Storing the user input and analysing it, is not a heavily discussed topic, and it is unsure whether this solution is possible at all. If it is possible, it is unlikely to be in a straightforward manner. Based on how this is evaluated, it must be determined whether this idea is worth exploring further or not.

It is possible to use a URL to a Forms in the Splunk Dashboard, this results in a situation very similar to the automatic Forms solution. The main difference is that the teams do not get an (extra) notification, meaning that there should be less annoyance on the part of the operators. On the other hand, the same problems with ignoring Forms apply.

6.1.5 Web-based application

In several studies, completely new tools have been made. An example of this is presented by Dayarathna et al. (2017), who have built a solution recommender built inside the WSO2 Data Analytics Server. The recommendations are presented in the dashboard, where the user can add a solution and rate a recommendation. Based on this rating, and other, unmentioned metrics, a confidence score is presented.

Apart from this, Amintabar et al. (2015) present ExceptionTracer, a completely web-based tool that helps developers with programming issues in an Integrated Development Environment by mining software systems as well as listing discussions from Stack Overflow. This means that the type of problems, and thus the solutions, differ from this study. Amintabar et al. (2015) however, provide an interesting perspective on what is possible with web-based tools. Similar to this study is the Crowd Knowledge Answer Generator (CROKAGE), a tool takes delivers solutions for a programming task based on the description of said task (da Silva et al., 2020). As the name suggests, websites such as Stack Overflow are used to retrieve possible solutions to the programming task. Similarly, Babu et al. (2021) analyse error messages of web applications using NLP and apply algorithms to find solutions from across different internet sources.

These studies are not necessarily a solution to the problem of this specific study, given the difference in the types of problems that these solution recommenders try to solve. However, they do provide an interesting perspective on how completely new tools can be used to present an all-encompassing recommendation dashboard for its users, and how the internet can play a role in this.

6.2 Companywide adaptation

In this section of the report, a consideration is made on how suitable the implemented solution concepts discussed in section 7.1 are when extended companywide, or whether they purely serve as a testing ground of the proposed changes to the Service Manager.

Pure data mining is considered suitable for companywide adaption as no addition must be made to the behaviour or applications used by operators.

Automatic Forms is a solution that can be realised companywide if the trigger of sending forms is changed to the Service Manager creation of the incident instead of the anomaly detection trigger. One must ask the question however, whether it is beneficial to have the forms as a solution as opposed to changing Service Manager and thus having all the work be done in one place. This is dependent on what the test indicates is required or desirable to improve the data quality and how well this is possible within Service Manager.

The Azure DevOps board is a solution that might be worth implementing companywide after the testing phase if this proves to be fruitful. Especially since the daily work of operators requires the use of the boards, this solution might provide a way to naturally move the incident registration in an operator's normal work, besides using Service Manager.

The Splunk dashboard is a solution that cannot be implemented companywide, because only certain AIOps use cases utilise the Splunk dashboard. As a result of this, this solution purely serves to test the proposed changes to Service Manager.

6.3 Non-AIOps-specific solutions

As Figure 9 indicates, this section of the report covers the non-AIOps-specific solution concepts. These solution concepts aim to improve the quality of the data gathered in Service Manager, so, for example, the search engine can be utilised better. These solutions are all changes to Service Manager itself, since any solution that needs to be implemented companywide and that is not in Service Manager, means extra work and complexity in the registration process.

It is important to note that these solutions can be implemented independently, but also simultaneously. For example, the templates and the solution generalisation can be implemented separately, but also be combined by giving a template based on the input of the closed solution type field.

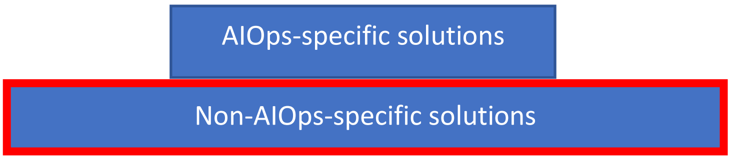

Figure 9 Visualisation of selected solutions

6.3.1 Delinking solution to customer notification

The first change that is necessary to the current Service Manager form is to decouple the solution from the customer notification. In the current system, the solution form has a box that can be ticked to send the solution to the customer. This is the only integrated way to send a message to the customer, and it includes the solution field. As a result of this, the solution field is often littered with entire emails to the customer. A separate customer field that allows operators to send a message to customers, with the possibility of copying and pasting the solution into the message, ensures that the data quality of the solution field will increase due to increased relevancy.

6.3.2 Warnings of poor quality information

Another possibility is to give warnings whenever a field has been filled in unsatisfactorily. An implementation of this would be to give a warning whenever a language that is not English is detected. Since this mostly concerns Dutch and Portuguese, in the most basic form, a warning can be given whenever text contains (one of multiple) words from a list of words often occurring in Dutch or Portuguese. This will nudge operators to use English. However, the more intelligent language detection is, the better.

This can be extended to detect whenever the solution field is filled in with too few words. Such input can even be forbidden, but often, triggering the intrinsic motivation of an operator by addressing the 'why' is more effective, without the drawback of forcing people to write long texts for the sake of writing long texts (Sansone & Harackiewicz, 2000). This would remediate the problem of the empty solution fields.

6.3.3 Description and solution generalisation

For both the solution and the description, there ideally would be closed entry fields combined with open text fields. This would allow for a combination of generalised information that can be properly analysed with a specific description that can guide other people in resolving the incident. As a bonus, this combination allows for a more guided form of NLP where the processor already has some idea of what to look for based on the input of the closed field.

The difficulty here lies in the fact that both description and solution can heavily vary between different departments within Rabobank. Some types can generally be applied to different departments, such as *'false alarm'*, *'reverse change'*, or *'coupled to greater problem'.* However, solutions with a nature such as '*File A was filling up system X. Copied file to system Y, where there was space*[1]*'* is more specific and not applicable to every department. To prevent a large list of problem and solution descriptors from showing up, the automatically filled in Configuration Item (CI) can be utilised. Based on the affected CI, and the already available information about the CI, a group of incident 'prototypes' and '*Not applicable'* can be presented to the operator, much like the hierarchical multi-label classification of Zeng et al. (2017). Creating these generalised terms requires intensive collaboration with the different departments of Rabobank before implementation, especially as these generalised fields are best employed when they are mandatory and inaccurate fields will lead to inaccurate data or an overabundance of '*Not applicable*'.

A lookup field of all the different options, although seemingly a good idea, is considered undesirable because natural language is used. This means that instead of a closed interval of options that always go by the same name (e.g. countries, cities), regular words are present. To give an example, if the action were to be a generic action like 'move file', an operator might first try 'reposition, transfer, relocate, or migrate' to search for the relevant action. This makes for an inefficient process.

6.3.4 Applying templates

Templates are already used in Service Manager, however there is more, unused utility to such templates. Specifically, these templates can be used to guide operators to enter free text that is more easily recognised by NLP. In accordance with Ali Zaidi et al. (2021) and Wu et al. (2019), a template can ease the semantic role labelling done using NLP to get actionable recommendations, solutions, or more concise descriptions. An example of a basic template in the open text solution is: "I have *(e.g. restarted, reversed, moved,…)* the *(server, database, file,…)*. The problem has been (*solved, decreased,…)*" The main idea of this solution is that by guiding operators to enter the data in a given format, the data quality will increase due to an increased ease of understanding and consistency. The templates can be applied with default values for fields, which are filled in upon creation and can then be edited.

6.3.5 Improved link with Knowledge Management and Problem Management

Within Service Manager, it is possible to link incidents with tickets from knowledge and problem management, meaning that an incident can be linked with a knowledge document that describes a workflow for given types of incidents. Apart from this, incident tickets can be linked with items of problem management, where trends of reoccurring incidents are analysed and thus valuable information is stored. What is seen now is that the field of incident solution and linked items often do not meet up. In some cases, the solution fields describe that the item is linked to a problem

[1] Not the actual names of the file or systems

management item, but this is not actually done. Other times, the reverse is true: a problem management item is linked, but there is no mention of this item in the solution description. Another interesting process which is relevant is change management, as mentioned by Güven et al. (2016). It is desired that whenever a change is the cause of an incident, and the incident can be solved by reversing the change, the responsible change is linked to the incident. This makes the analysis of causality between change and incident much easier for future incidents.

This problem can be remediated in two manners. First, a warning can be given or a button to link an incident ticket to another item whenever it is detected that a change, problem, or knowledge item has been mentioned in the description or solution field. Preferably, this is automated, where a message box is presented to the operator whenever another item is mentioned. Using this message box the operator can either allow or not allow Service Manager to automatically link the items together.

Reversely, it can be ensured that linking an incident to another item is automatically mentioned in the solution field. The benefit of mentioning the incident here as well is that whenever analysis is done purely on the solution field, the solution to the linked item can be copied in the field, or if the item is a knowledge item, the knowledge item itself is often the solution.

7 Evaluation

In order to assess the strengths and weaknesses of the various concepts, a questionnaire has been conducted in accordance with the research design described in section 2.4.3. This chapter aims to analyse the results of this questionnaire and to discuss the findings. The questionnaire itself is to be found in Appendix E, and the raw data (that is, the responses) are found in Appendix F. It must be noted that although the research design enables comparisons between different groups (e.g. an operating team using AIOps versus an operating team not using AIOps), the low response rate (9 in total) means that these comparisons are less meaningful as they would consider the opinions of about 2 people. Therefore, the choice has been made to analyse the responses globally, as if each participant had the same level of knowledge and bias.

7.1 AIOps-specific evaluation

The questionnaire has been divided into two parts. The first part concerns the AIOps-specific solutions. The concepts have been evaluated based on three main criteria which have been divided further into specific questions. The three criteria are *ease of implementation, effectiveness and efficiency,* and *operator usage*. All of these are divided into four questions, and to conclude a question to rank the concepts overall is presented. *Ease of implantation* discusses the knowledge, labour, money and time frame needed for implementation. *Effectiveness and efficiency* concerns the effectiveness, efficiency, flexibility and versatility of the concepts. *Operator usage* concerns the viewpoint of the operator regarding annoyance, ease of use, and work routine. Once again the exact questions are to be found in Appendix E.

When looking at the ratings given at the four questions of a single criterion, two main indicators are used. The first is the total average, where Equation 1 is used. 'Total' refers to the fact that the average is of the criteria, as opposed to a single question.
Given that *c* is the solution concept, $c \in \{1,2,3,4,5\}$
And *p* is the placement, $p \in \{1,2,3,4,5\}$
And *n* is the total number of times that placement *p* is given to concept *c* in the 4 questions,
$n \in \{0,1,\ldots,8,9\}$

Equation 1 Average

$$Total\ average\ (c) = \sum_{p=1}^{5} \frac{n * p}{9}$$

In order to address the spread of the values, the mean absolute deviation is used (MAD). This is an easy to use metric that shows the average distance to the total average. The following formula is used

Equation 2 MAD

$$MAD\ (c) = \sum_{p=1}^{5} \frac{n * |p - total\ weighted\ average\ (c)|}{36}$$

7.1.1 Ease of implementation

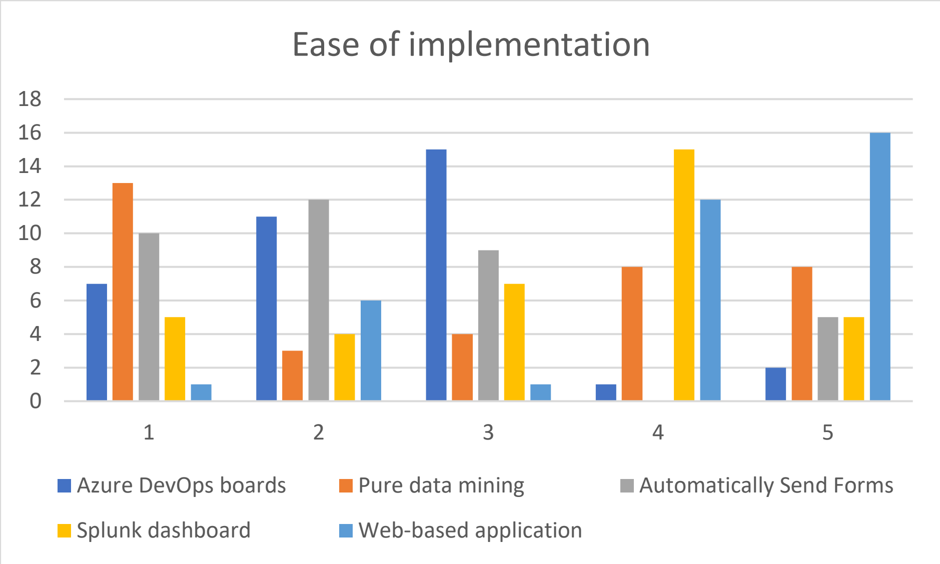

Figure 10 Ease of implementation AIOps-specific concepts

When looking at the total averages of the concepts in Table 7, two stronger concepts can be identified. Both *Azure DevOps boards* and *Automatically Send Forms* have lower scores indicating that they are identified as easy to implement. As can be seen in Figure 10 as well, *Pure data mining* is the concept with the widest spread, meaning that this is the concept with the most disagreement, which is something to keep in mind. Other than that, there are no values that stand out in this analysis. *Azure DevOps boards* and *Automatically Send Forms* are considered to be better, and the *Splunk dashboard* and the *Web-based application* are considered weaker. Among the individual questions, there are no results that significantly differ from this analysis.

Table 7 Ease of implementation of AIOps-specific concepts values

Concept	**Total average**	**MAD**
Azure DevOps boards	2.44	0.83
Pure data mining	2.86	1.49
Automatically Send Forms	2.39	1.03
Splunk dashboard	3.31	1.05
Web-based application	4.00	0.89

7.1.2 Effectiveness and efficiency

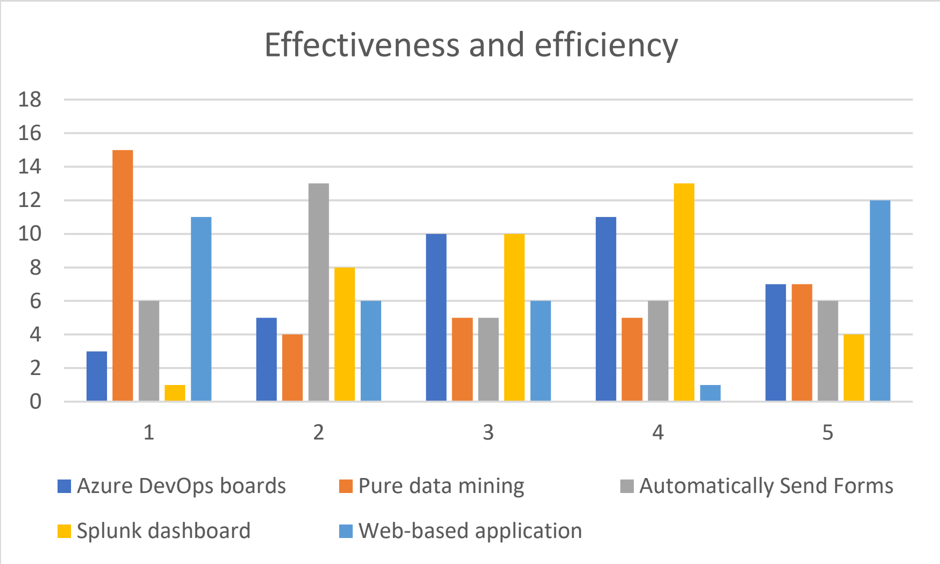

Figure 11 Effectiveness and efficiency AIOps-specific concepts

With regards to efficiency and effectiveness, *Pure data mining* scores the best, with a small margin to *Automatically Send Forms* and *Web-based application*. *Azure DevOps boards* and *Splunk dashboard* perform worse and as can be seen in both Figure 11 and in the MAD in Table 8, they are considered poor relatively consistently. The most interesting concepts are *Pure data mining*, which scores high but has a wide spread, and *Automatically Send Forms*, which scores a bit higher, but its spread is less wide.

It must be mentioned that although the *Web-based application* is considered reasonably effective, versatile and flexible, it is also considered to be over-engineered, which explains the wide spread of the responses. Other than that, *Pure data mining* is considered effective, but also inflexible.

Table 8 Effectiveness and efficiency of AIOps-specific concepts values

Concept	Total average	MAD
Azure DevOps boards	3.39	1.00
Pure data mining	2.58	1.45
Automatically Send Forms	2.81	1.18
Splunk dashboard	3.31	0.88
Web-based application	2.92	1.48

7.1.3 Operator usage

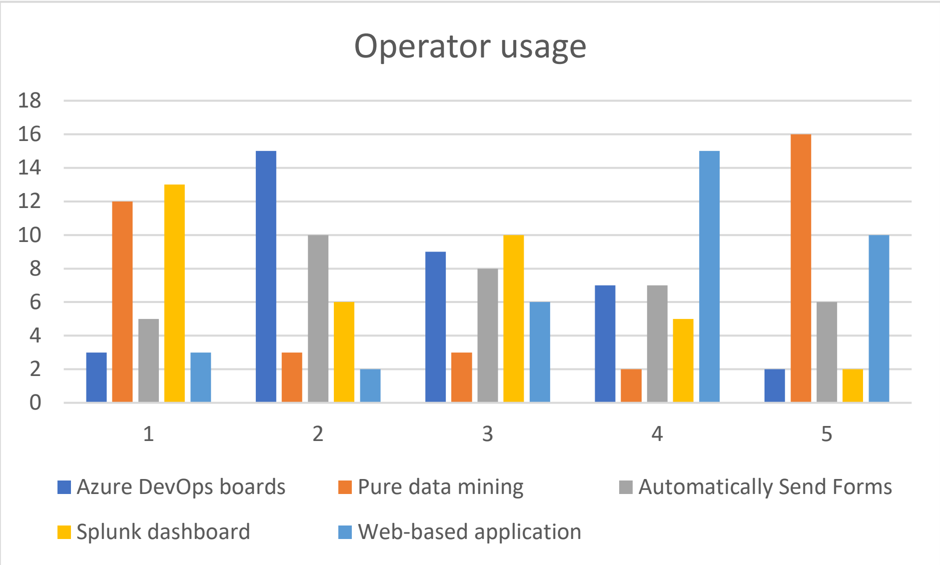

Figure 12 Operator usage AIOps-specific concepts

With regards to operator usage, it is clear to see that the *Splunk dashboard* is considered to be the most user-friendly. The *Web-based application* is considered to be the least user-friendly, by a significant margin. The *Azure DevOps boards* and *Automatically Send Forms* are considered as decent, where *Azure DevOps boards* peaks at the second ranking (as seen in Figure 12), but is rarely considered the best. *Automatically Send Forms* however is much more evenly spread, meaning that there is less consistency, which is also reflected in the MAD in Table 9. Most noteworthy here is *Pure data mining* which is a very dividing concept. It is considered either the best or the worst in this category, and rarely something in-between. This cannot be completely explained just by looking at the individual questions. Although some questions lean more towards positive and other questions lean more towards negative, there is still a high level of variation within those single questions. This points to a high level of disagreement.

Table 9 Operator usage of AIOps-specific concepts values

Concept	**Total average**	**MAD**
Azure DevOps boards	2.72	0.89
Pure data mining	3.19	1.69
Automatically Send Forms	2.97	1.09
Splunk dashboard	2.36	1.10
Web-based application	3.75	0.90

7.1.4 Overall

Lastly, the participants were asked to rate the concepts based on what they thought was the best concept overall. The average and MAD were taken just of this one question.

Table 10 Overall view of AIOps-specific concepts values

Concept	Total average	MAD
Azure DevOps boards	3.33	0.96
Pure data mining	2.33	1.48
Automatically Send Forms	2.89	0.99
Splunk dashboard	3.44	1.28
Web-based application	3.00	1.11

As can be seen in Table 10, *Pure data mining* is considered as the best overall concept, followed by *Automatically Send Forms*. However, it is noteworthy that while *Pure data mining* is rated higher, *Automatically Send Forms* is rated more consistently.

7.2 Non-AIOps-specific evaluation

This section is set up in the same manner as the AIOps-specific section, with the same calculations. One difference is that the questions are slightly altered to more accurately reflect the difference in this dimension of concept (Appendix E). The other difference is that the criteria *operator usage* now only has two questions as opposed to four.

7.2.1 Ease of implementation

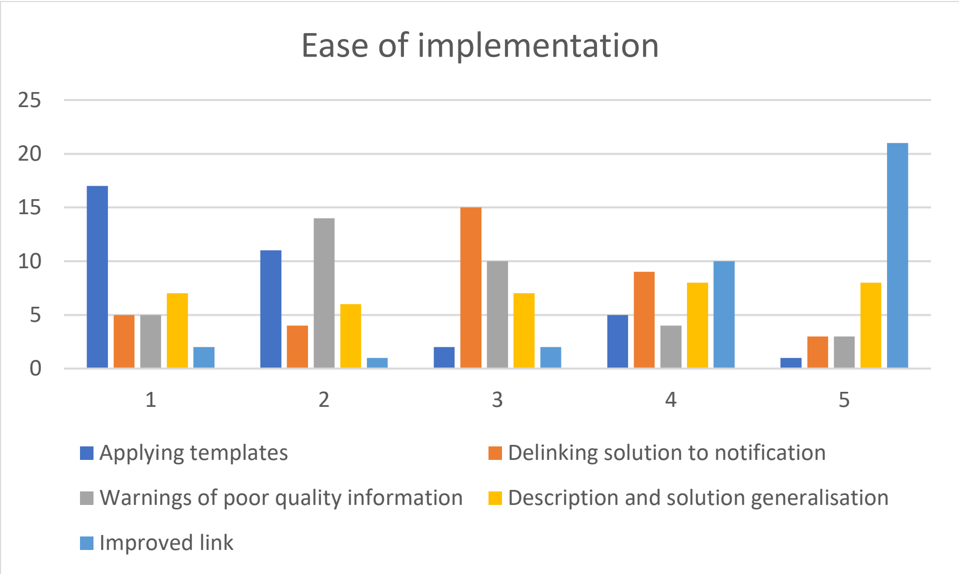

Figure 13 Ease of implementation non- AIOps-specific usage

As can be seen in both Figure 13 and Table 11, the templates are considered the easiest to implement by most people. It has a low average, with a relatively low MAD, meaning that it is consistently ranked high. Other than that, it is clear that the generalisation is a dividing concept, with a wide spread over the different ratings. Last, Improved link is considered to be the most difficult to implement, with consistency, as shown by the low MAD.

Table 11 Ease of implementation of non-AIOps-specific concepts values

Concept	Total average	MAD
Applying templates	1.94	0.89
Delinking solution to notification	3.03	0.81
Warnings of poor quality information	2.61	0.92
Description and solution generalisation	3.11	1.23
Improved link	4.31	0.81

7.2.2 Effectiveness and efficiency

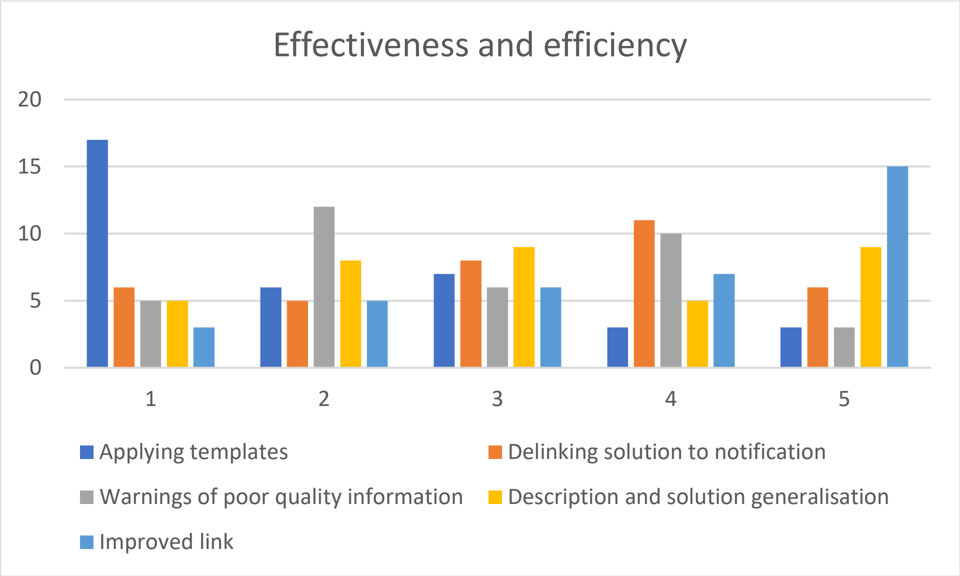

Figure 14 Effectiveness and efficiency non- AIOps-specific usage

Once again, *Applying templates* has the lowest average by a significant margin. It is noteworthy that the MAD of all concepts is around the same number. This would suggest a relative equal spread, but Figure 14 shows that this is not quite true. It is somewhat true for *Description and solution generalisation*, *Warnings of poor quality informa*tion, and *Delinking solution to notification*. However, *Applying templates* and *Improved link* are respectively skewed towards one and five, which is shown more clearly in the average in Table 12.

Table 12 Effectiveness and efficiency of non-AIOps-specific concepts values

Concept	Total average	MAD
Applying templates	2.14	1.12
Delinking solution to notification	3.17	1.12
Warnings of poor quality information	2.83	1.06
Description and solution generalisation	3.14	1.17
Improved link	3.72	1.17

7.2.3 Operator usage

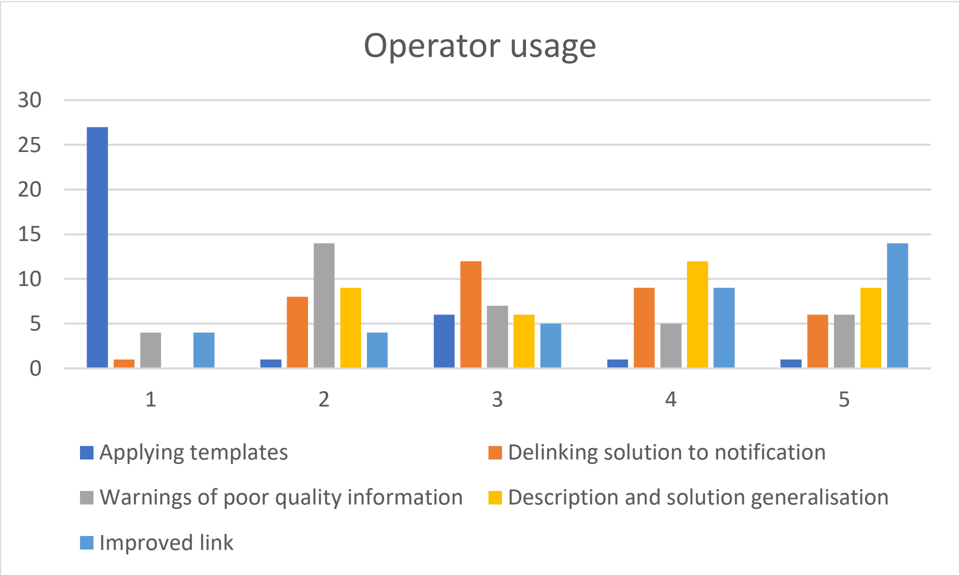

Figure 15 Operator usage non-AIOps-specific concepts

With regards to operator usage, it can be seen that *Applying templates* is favoured, and by a significant margin. The other concepts are much more equally distributed, where *Warnings of poor quality information* distinguishes itself from the others, which is also reflected in Figure 15, showing that the concept is chosen as second the most. Table 13 also shows that *Improved link* has the worst average, which is in line with Figure 15, showing that Improved link is also chosen as the worst most often.

Table 13 Operator usage of non-AIOps-specific concepts values

Concept	**Total average**	**MAD**
Applying templates	1.56	0.83
Delinking solution to notification	3.31	0.91
Warnings of poor quality information	2.86	1.08
Description and solution generalisation	3.58	0.97
Improved link	3.69	1.17

7.2.4 Overall

Table 14 Overall view of non-AIOps-specific concepts values

Concept	**Total average**	**MAD**
Applying templates	1.44	0.69
Delinking solution to notification	3.56	0.94
Warnings of poor quality information	2.89	1.01
Description and solution generalisation	3.22	1.14
Improved link	3.89	1.04

Overall, *Applying templates* is consistently ranked as the best concept, with *Warnings of poor quality information* being the second most favoured. *Improved link* is considered to be the worst concept. The generalisation concept is the most disputed concept. This is not surprising, as although this concept has some strengths, it is more dependent on how well it can be executed than the other concepts.

8 Conclusion

Based on the analysis shown in chapter 7, multiple conclusions can be drawn. This section of the report presents these conclusions. It aims to highlight the strengths and weaknesses of the proposed concepts, based on the evaluation. First, conclusions about the AIOps-specific concepts are drawn. Then, the non-AIOps-specific concepts are examined.

8.1 AIOps-specific conclusions

With regard to the *Azure DevOps boards*, it becomes clear that it is considered easy to implement and reasonably user-friendly, but not effective, versatile or flexible. This means that especially concerning the solution recommender, it is considered to be fitting for the current situation, where the required data is relatively straightforward: text strings or integers. However, considering that the concepts might be used to test the non-AIOps-specific concepts, the *Azure DevOps boards* are considered unfit. Apart from this, if the agile concept of the solution recommender means a change in the required data, the *Azure DevOps boards* might prove unable to adapt. This is likely one of the main reasons that this concept is rated as fourth overall.

Pure data mining is the most divisive concept. Not only has this concept the highest MAD at every single criterion, the histograms also make clear that there is high polarisation. The number of ones and fives given is significantly higher than twos, threes or fours. As a result of this, *Pure data mining* mostly scores middle-of-the-road, with the exception of effectiveness and efficiency, mostly due to a high perceived effectiveness. It is also interesting that this concept is perceived as the best overall. Once again, it does not do so consistently, as shown in a high MAD. This is likely due to varying levels of complexity that his concept can perform at, based on the quality of the entered data and what concepts are implemented (see section 6.1.1). It must be concluded that this concept has potential, but that the degree to which this potential can be used is highly dependent on the level of complexity that the system is built with.

Automatically Send Forms is the concept that gets the most second-place ratings, as can be seen in the histograms of chapter 7. This means that this concept does reasonably well. It is considered the easiest to implement with decent effectiveness and efficiency and operator usage. It also gets the second place in the overall rating with a relatively stable MAD. The most worrying part of this concept is that the perceived annoyance at the side of the operator is considered high. Other than that, this concept is considered decent: it is an easy to implement, easy to use and versatile concept.

The *Splunk dashboard* has one clear strength: it is considered to be easy to use at the side of the operator. However, the concept is also considered inefficient, inflexible and difficult to implement. Likely as a result of this, this concept is rated the worst overall. The conclusion for this concept is that although it certainly has its benefits, the restrictiveness presented by Splunk resulting in the difficult implantation and inflexibility means that this concept is unsuitable for the task at hand.

The *Web-based application* is considered to be over-engineering. Although the effectiveness, flexibility, and versatility are rated highly, the difficulty of implementation and the difficulty for operators to get used to another application drop this concept in the ranking. Especially in terms of money, this concept scores poorly. For what could be a relatively straightforward task, this concept is too complicated.

8.2 Non-AIOps-specific conclusion

Applying templates is the best scoring concept by a significant margin. It is rated the highest in every single criterion, and as a result rated the best overall as well. Both the histograms and the MADs support this consistent rating. One can conclude that this concept is easy to implement, effective, and user-friendly.

Delinking the customer notification from the solution field is a middle-of-the-road concept. Its distribution is centred around 3 in the evaluation of the criteria. This concept is rated highly for its low requirement of differentiation between departments, but the expected effectiveness is also relatively low. Other than that, the ratings are mostly neutral. Overall, one can conclude that this concept is relatively low-risk, low-reward. It is not extremely difficult to implement or use, but the expected results are relatively low as well.

Warnings of poor quality information is the concept that gets the most second-place ratings. It is considered rather easy to implement and effective, but gets less consistent ratings based on the required differentiation between departments and the operator usage. This is not necessarily surprising. The reaction to a low-quality data message will highly differ among the recipients. Some will find it annoying, others will ignore it, and others will adhere to the message. Besides this, what can be considered high-quality data for one department, might not necessarily be considered as such for another department. The key conclusion here is that this is a concept with potential, but that experimentation must take place to address whether it is effective, and how it is received.

The description and solution generalisation is the concept with the 'flattest' distribution, meaning that the participants are divided on this concept. Once again, this is not surprising. This is a high-risk, high-reward concept. The key here is the required differentiation. This is where effort must be taken during implementation: if the proposed generalisations are done well, the reward is high. If not, there is no reward, but rather damage through annoyance on the side of the operator. The high potential of this concept is worth exploring, but this must be done cautiously.

The concept of an *improved link* between tickets in Incident Management with items in Knowledge Management and Problem Management, is consistently ranked the worst concept. It is considered annoying, difficult to address the required differentiation, and difficult to implement. The latter is mostly due to expected problems with Service Manager. This combined, with an average effectiveness makes this concept high-risk, average-reward. It must be concluded that this concept is unsuitable for the task at hand.

9 Future research

In this section of the report, various possibilities for future research are shortly explored. These are subjects that came to light at any point in the research, such as being a by-product of the interviews to determine the scope, the literature research to determine feasible solutions, or through meetings that were part of day-to-day work. The subjects are discussed, and the opportunities and expected challenges are highlighted, based on literature (see section 2.4.4.).

9.1 Self-healing

When the solution recommender has been implemented, a natural continuation of the automation of Rabobank's process is to start implementing self-healing (van Landegem et al., 1994). As Mahdavi-Hezahevi et al. (2017) state, the main objective of adding self-healing features is to "maximize the availability, survivability, maintainability, and reliability of the system.". An efficient way to do this is to utilise the information that has been gathered using the solution recommender. The logic used here is that when a solution can be predicted, the actual implementation of said solution can immediately be executed if there is enough confidence in the given solution.

Especially in combination with the anomaly detector, self-healing becomes an interesting topic of study. If the two fields are combined, a system can self-diagnose and self-heal all without the actual incident happening, meaning that there is no disruption of service. This effect becomes even stronger given that the anomaly detector already proposes a root cause identification, something which not every self-healing mechanism possesses. To illustrate, Dai et al. (2011), have proposed a self-healing system that is consequence-oriented, meaning that the root cause is never identified, but rather the consequences based on the symptoms are used to solve the (potential) incident.

Apart from this, the linking with other knowledge items in the Service Management system provides an interesting perspective on self-configuration (Petrenko, 2020). In this case, when significant anomalies are detected which are expected to be the result of a change, the self-healing system can automatically configure itself to be in the 'correct' configuration, thus solving the problem. Caution must be applied here, as a change will likely result in some alteration of the normal behaviour of a system, meaning that the sensitivity of the anomaly detector must be carefully determined.

The implementation of such a system comes with challenges. One of these challenges is the responsibility, especially considering the infrastructure risk. Braschler et al. (2019) describe this as caution with fully automating workflows, decision processes, and business plans. In such cases, once the technology has been deployed, it is extremely difficult to stop or reduce it, for technical, social, and economic reasons. A self-healing algorithm might sound good, but who does the responsibility lie with if the system causes serious issues due to predicting a wrong solution? Here, explainable AI starts playing a major role, according to Samek and Müller (2019) one of the challenges hindering the prevalence of AI in some applications. It is key, to program the AI in such manner that the verifiability and transparency of the AI are of a sufficient level.

The biggest challenge with implementing a self-healing system is to have the system be accurate enough. There must be a careful balance between ensuring that a solution is achieved and causing the least amount of disruption possible. To illustrate this, completely restarting a system might solve the problem, but does cause a loss of functionality that would rather be avoided. However, any moment that the self-healing system does not produce the solution is another second that the problem persists. One way to combat this problem is to utilise the knowledge of operators by having them authorise the query calls by the self-healing system. This reduces the state of automation, but in doing so collects valuable information from operators.

These operators however, need to be properly informed of what is being done and why. Not only from an ethical point of view, but also from a utility-orientated perspective. If the operators feel like they are working towards the extermination of their own job, even though this is not the case (especially since solving incidents is not an operator's entire job), they will unlikely be willing to properly cooperate. Informing the operators and giving them a voice is of major importance in this case (Umiker, 1997).

9.2 Impact determination

Where an opportunity lies for an AI project is to utilise AI to assess whether an anomaly in a system correlates with anomalies in other systems. Especially for major incidents, where it is crucial to precisely and swiftly determine the impact of an incident. The difficulty here lies in the fact that a lot of systems belong to multiple chains and/or processes. To give a (simplified) example, the Rabobank app is both part of payment processes, and of customer support. This means that whenever it is determined that there is a (potential) loss of functionality in one of the chains, that it is difficult to assess how far this loss 'ripples' to other processes or chains. Currently, this work is being done by humans, who analyse the signals produced by the executed measurements (e.g. Twitter messages, number of transactions, crashes etc.).

The goal of this is first to determine the customer impact, the degree to which the customer is hindered by an incident or group of incidents, and to act accordingly. Second, this is a crucial process from a retrospective viewpoint. Determining correlation between incidents means that eventually, one can prevent disturbances, or at the least decrease the ripple effect and thus reduce the customer impact.

The potential of AI comes into the picture in two parts of this process. The first part is in structuring the great quantity of unstructured data that is delivered as input to this process. Using AI, it would be possible to automatically unify and standardise this data and get key values, meaning that the analysts can allocate their time to the actual value-adding tasks. Apart from this, AI can make relationships between different events better than humans can. These relationships can then be used to determine customer impact quicklier and more accurately, and they provide insight into how different applications or chains are interconnected, and what this will mean for the future with regard to (Major) Incident Management.

Challenges here concern explainable AI. Decisions and analyses, especially with major incidents need to be accounted for. When, in hindsight, the wrong decision is made based on a relationship proposed by the AI, then it must be known why that relationship was made.

9.3 Know Your Customer

Banks are required to perform Know Your Customer (KYC) checks upon clients or customers, to combat criminal activities such as money laundering, but also drug dealing and terrorism (De Nederlandsche Bank, 2022). Rabobank already uses AI to detect suspicious streams of money (Rabobank, 2021b). However, there also lie opportunities to analyse the paperwork that is delivered to Rabobank. Currently, this process is semi-automated. Programs are used that summarise (parts) of papers, yet the responsible employee still largely starts at a random paper, and judges whether a transaction is suspicious or not. Doughty (2005) stresses the importance of automation in the field of KYC in order to stay compliant with legislation.

A way to improve this process is by using a similar system as ASReview (van de Schoot et al., 2021). This model, recommends a scientific paper to the user based on whether they ranked similar (determined using NLP) papers relevant or not. The key here is that the human interaction of this active learning cycle is also the decision maker. In the case of KYC, this means that the AI does not make the decision, which would be ethically questionable, but the decision whether the presented paper contains suspicious activity or not is taken by a human. Instead of deciding that a person or activity is suspicious, the AI presents evidence that the person might be suspicious.

There are two major difficulties at play in this design. The first difficulty is once again, explainable AI. Even though the presented system takes away the decision from the AI, the reasoning behind why the presented paper is shown must still be explainable. Apart from this, any biases with regards to race, gender etc. must be carefully navigated to prevent situations in which these are presented as leading causes to show certain papers above others (Coeckelbergh, 2020).

Other than that, there is the issue of when to stop presenting new papers. One can keep presenting papers until every paper has been read, but that is inefficient and likely quite ineffective. A confidence score in the form of a probability that the document contains suspicious activity is doable, but a low cut-off point means that documents must be read that likely do not add value, while a relatively high cut-off point means that (potential) criminal activity will remain unseen.

9.4 Security

The currently used implementation of anomaly detection is mostly based on monitoring data and the performance of systems or applications rather than the security aspects of said systems and applications. Although Rabobank is compliant with regards to security, meaning that the legislated standards, as well as the internal standards are conformed to, there is always room for improvement.

Extensive research has been conducted to apply AI in the security field of information systems, and Batina et al. (2022) provide an overview of the state-of-the-art with regard to security and AI. They discuss the use of AI in encryption, authentication and privacy, but for now the focus is on anomaly detection for intrusion detection, due to the familiarity with anomaly detection.

Rimmer et al. (2022) state the benefit of not assuming complete knowledge of prior malicious patterns and instead learning normal behaviour of good-natured traffic, especially due to the continuously evolving nature of attacks. However, they also present challenges that belong with using anomaly detection for intrusion detection. Challenges that can be expected if intrusion detection were to be implemented at Rabobank are: high false positive rates, getting representative and clean data, and attacking techniques that anomaly detection is susceptible to such as mimicry attacks (attacks that mimic normal traffic) and poisoning attacks (slowly letting the detection model learn an attack as normal behaviour) (Biggio et al., 2014; Fogla et al., 2006).

Besides the technical challenges, explainable AI and the confidence in the AI are once again key challenges that need to be coped with, especially concerning the sensitive nature of security.

10 Recommendation

In this section a recommendation is given with regards to the problem of poor data quality, keeping in mind the goal and the requirements of building a solution recommender (Appendix A). This recommendation will use the conclusions reached in chapter 8 to provide a way to improve the data quality, the challenges that the implementation will face, and a recommendation on the future research that was examined in chapter 9.

10.1 AIOps-specific recommendation

First a recommendation in order to answer the question: "*How can a data channel be designed in order to collect data for a solution recommender, keeping in mind the ease of implementation and user-friendliness?*" is made. Based on the conclusions drawn in chapter 8, it is advised to use the Microsoft Forms to receive the required user ratings, and to test Service Manager improvements. This channel gives the flexibility to adjust data fields and to easily collect the required information. Giving operators access to the Forms can be done in two manners. First, as has been described, a Microsoft Teams message can be sent to notify operators that a form must be filled in. A bot can be used to make this process both automated and more humanlike, reducing the perceived annoyance. However, the questionnaire made clear that the perceived annoyance is considered to be high.

Therefore, another option is to combine the Forms solutions with the Azure DevOps boards solution. In doing so, the intuitiveness of the boards, which are used in the day-to-day work of operators, is utilised, without having the restrictiveness of the cards hindering the process. How this would work is that a task is created on the board with the name "*Fill in form*", which has the link to the Form in the task description or through a one-click action. This circumvents the major flaws of the Azure DevOps boards concept, while utilising its strength: operator usage. This solution is recommended as the best one, however, given the potential problems with authorisation and the upcoming merging of Azure DevOps boards, the Teams messages is a viable alternative.

It is recommended to first implement this solution in a team that is experienced with the anomaly detector, such as Instant Payments, to assess start-up problems and to evaluate how well the solution is received and how effective it is. As has been established, key to the implementation is educating the operators on the goal of the changed process, and how they benefit from this effort.

The implemented solutions, especially if the non-AIOps-specific solutions are implemented in the Forms, also allow experimenting with NLP for an improved search engine. As the questionnaire shows, *Pure data mining* is a divisive concept, that can be considered high-risk, high-reward. It is recommended to experiment using this type of analysis, however the main constraint here is the data quality. If the data is deemed to be of sufficient quality, a model that combines the quality quantifications of Zhou et al. (2017) and the action extraction using semantic processing of Ali Zaidi et al. (2021) is recommended. This means that a recommendation is given without user ratings, but using similarity measures. If via some type of minimum viable product, it has been proven that the quality of the data is sufficient, the implementation can be expanded companywide. Once again, educating operators on why these changes are being made, and how they will benefit from them, is one of the key components of this process.

10.2 Non-AIOps-specific recommendation

With regard to the non-AIOps-specific concepts, it is recommended to implement the templates, as well as the poor quality warnings. These concepts were rated positively in the questionnaire, and are deemed easy to implement and effective. Besides this, the solution and description generalisation is contested. This concept has potential, but it is key to spend time and effort into ensuring that the differentiation between departments is done accordingly. Once again, the recommendation is to implement these changes into the proposed Form first, to test its effectiveness and reception. Since, this works to a less degree for the generalisation, it is recommended to bring together a group of people to discuss to what degree the generalisation is achievable, and (assuming that this meeting has a positive outcome) to then start experimenting.

10.3 Future research recommendation

Apart from this, Rabobank has some opportunities concerning AI that can be exploited at some point in the future, once more research has been done. The impact determination is advised to be researched first, as this presents the greatest opportunity. The reason behind this is that the impact of Major Incidents, the most impactful incidents, can potentially be reduced, which brings Rabobank the greatest value. Considering that several opportunities can be considered impact determination, a recommendation is given as to what specific subjects can be examined further.

10.3.1 Time-based impact determination

To determine the expected impact of an incident, correlation between applications must be established. One way to do this is to find a correlation based on time series of tickets of incidents.

It has previously been discussed how tickets can be related to each other (see section 7.1.1.). Up until now, tickets and incidents have mostly been related to each other based on inherent qualities (affected Configuration Item (CI), caused by CI) and their textual information (description, solution). However, incidents can also be related to each other based on the time that they occurred (Marcu et al., 2009). Salah et al. (2015) describe implementing time (among others) in the correlation of incident tickets. They implement creation time, resolution time and closing time (called validation time) to relate incidents to each other, both locally (within one application) and globally.

The global aspect of this method can be applied within the context of this study. Specifically, to realise impact determination, the time aspect can be used to see if there is any correlation between applications. That is, to research if the health of one application implies consequences for other applications. The assumption is that if applications often have tickets with creation, resolution or closing times close to each other, that there is some sort of correlation. This can imply a certain causation, which should be researched further.

The greatest issue here is the current inaccuracy of the recorded times by operators. The resolution time and closing time are often very close to each other as operators, for example, enter the information at the end of the workday, instead of when the task is actually completed. Therefore, the creation time is the most trustworthy time in this process, especially considering that tickets are also created automatically by monitoring software, meaning that the times are much more accurate. Apart from this, there is the problem of correlation and causation. It is possible that a correlation between applications has no meaningful causation. Therefore, if there is correlation between applications, there should be a critical assessment on whether this implies causation.

10.3.2 Automatic data structuring

Whenever a Major Incident occurs, the collected data is structured manually. This means that to make the data uniform, present them in one channel, and to filter out the unnecessary data, humans perform labour. AI can do this both faster and more accurately (Gupta & Mangla, 2020). As a result of this, crucial time is lost, especially considering that this relates to Major Incidents, where fast resolution is of greater importance.

AI can be applied in this structing process in several ways. First, AI can be used to extract relevant information from external indicators, such as Twitter messages or user reports. Second, AI can help transform data into the relevant format, by detecting what type the data is (date, location, numerical, etc.). This decreases the time needed by analysts to transform the data and to structure it into a format that can be easily analysed.

Thirdly, AI can be applied to not only transform the collected data, but also to analyse it. Right now, anomaly detection is only used on monitoring logs of application, meaning that only internal indicators are taken into account. However, anomaly detection can also be applied to external indicators. Especially considering that perceived customer impact often follows a somewhat normal distribution, where the number of complaints increases over time, reaches a peak, and then decreases (Rabobank, personal communication, 2022). This means that instead alerts being fired at certain (often defined by humans) thresholds, the pattern of complaints can be analysed and alerts can be fired at earlier stages, based on the normal behaviour of the complaints. Due to such earlier alerts, the resolution can be achieved at an earlier stage. Guille and Favre (2015) present a model that extracts information from Twitter to detect events based on an anomaly approach, combining both time, mentions (the linking of other user accounts), and text. This model can be used to collect information from Twitter, but can also be extended to other social media that are used as input channels.

10.3.3 Anomaly detection for chains

Rabobank has expressed interest in possibilities to more effectively monitor chains of application. As has been described, the difficulty lies with applications or systems often being part of multiple chains/processes. AI can be used to improve the monitoring for chains. The first of two ways in which this can be done is from a top-down perspective, where data is collected globally, purely for chain monitoring purposes. Second, a bottom-up perspective can be applied, where the already existing anomaly detectors are used to create a chain overview.

Top-down perspective

Using the top-down perspective, it can be ordained that every application within a chain proposes several major indicators that are measured. These can both be internal (application logs, etc.) indicators and external (user reports, etc.). The indicators from the various indicators are then combined to analyse whether the behaviour of the chain as a whole is normal, using anomaly detection. If the relation between several applications is unusual, then this might be due to an incident.

The main downside of this perspective is that the root cause identification, one of the main advantages of anomaly detection, is not truly applicable to such a system due to the limited number of indicators. The goal of such a system is to provide an overall health overview for higher management, with minimal effort for (and advantages) for singular teams

Bottom-up perspective

Using the bottom-up perspective, the existing anomaly detectors are combined into a system that presents a health score of the entire chain. This is done by combining the individual health scores of the application into one singular health score. The key challenge here is that the contributor scores on which these health scores are based are not normalised. To illustrate, a contributor score of 7 in one application might be a severe anomaly, while it is considered a small anomaly in a different application. Similarly, more volatile applications will have health scores of which deviations must be higher to be significant. However the health scores that are shown in the dashboard have colour coding to them already, where a healthy score is given a green outlay and a bad score a red outlay (and 2 stages in-between are different shades of orange). The colour can be visually combined to show an overall indication of the health of a chain.

The idea of ‘gluing’ together the different existing anomaly detectors provides the advantages of having one overview and depth with regard to root cause identification of the singular anomaly detectors. However in terms of comparing applications within a chain, this concepts provides little. It must be taken into account that under the assumption that the required anomaly detectors are already in place, this idea requires a low amount of effort.

10.4 Discussion

There are several things that must be taken into account with this recommendation, both from a scientific perspective as well as from the business-orientated perspective. This section of the report aims to highlight these problems and to establish the underlying reasons.

First, the conducted evaluation of the solution concepts is not objective. A questionnaire does not provide the ‘hard’ quantitative data that is able to clearly measure strengths and weaknesses like an experiment could. However, due to the time and effort that proper experiments would take (implementing, explaining, and getting reliable data), the choice has been made to use a questionnaire. This means that in practice, a non-recommended concept might work better.

Second, the questionnaire has been filled in by nine participants. Care has been taken to reduce the bias of the individual participants by using the input of several groups of people, who have a different stake in the project (or none at all; see section 2.4.3). However, the low response rate means that the significance of the questionnaire diminishes. Preferably, the response rate would have been higher, which ensures that personal biases are less meaningful, but also that the different groups can be compared with some statistical significance.

Third, the participants of the questionnaire gained most of their information from the presentation which has been given to them by the researcher. This means that any bias of the researcher might have influenced the responses of the participants. Care has been taken to present as neutrally as possible: keeping descriptions objective, and not over- or under-representing any concepts. This does not mean however, that there was no bias at all.

It must be noted that the scientific significance of this study is not great. This study must mostly been seen as a case study of a specific and detailed problem within one company. If any other company is to implement a solution recommender or aims to improve the overall data quality, then this study might prove useful as inspiration. The likely differences in software used, current systems, and work styles, might prove ground to disregard the outcome of this study. The methodology, proposed ideas, and described literature however can be useful information.

Bibliography

Aggarwal, C. C. (2016). *Recommender Systems: The Textbook* (1st ed. 2016 ed.). Springer.

Ali Zaidi, S. S., Fraz, M. M., Shahzad, M., & Khan, S. (2021). A multiapproach generalized framework for automated solution suggestion of support tickets. *International Journal of Intelligent Systems*, *37*(6), 3654–3681. https://doi.org/10.1002/int.22701

Alpaydin, E. (2014). *Introduction to Machine Learning (Adaptive Computation and Machine Learning series)* (3rd ed.). The MIT Press.

Amintabar, V., Heydarnoori, A., & Ghafari, M. (2015). ExceptionTracer: A Solution Recommender for Exceptions in an Integrated Development Environment. *IEEE 23rd International Conference on Program Comprehension*.

Babu, S., Pitchai, R., & Anjanayya, S. (2021). Web log analysis using Spark with solution recommendation. *Materials Today: Proceedings*. https://doi.org/10.1016/j.matpr.2021.03.313

Barton, D. (2018). *One Bank's Agile Team Experiment*. Harvard Business Review. https://hbr.org/2018/03/one-banks-agile-team-experiment

Batina, L., Bäck, T., Buhan, I., & Picek, S. (2022). *Security and Artificial Intelligence: A Crossdisciplinary Approach* (1st ed. 2022 ed.). Springer.

Batini, C., & Scannapieco, M. (2006). *Data Quality: Concepts, Methodologies and Techniques (Data-Centric Systems and Applications)* (1st ed.). Springer.

Bauer, E. (2016). *Lean Computing for the Cloud*. Wiley.

Biggio, B., Rieck, K., Ariu, D., Corona, I., Roli, F., Wressnegger, C., & Giacinto, G. (2014). Poisoning behavioral malware clustering. *Proceedings of the 2014 Workshop on Artificial Intelligent and Security Workshop*.

Booch, G., Rumbaugh, J., & Jacobson, I. (2005). *The Unified Modeling Language User Guide* (2nd ed.). Addison-Wesley Professional.

Braschler, M., Stadelmann, T., & Stockinger, K. (2019). *Applied Data Science: Lessons Learned for the Data-Driven Business* (1st ed. 2019 ed.). Springer.

Cho, S., Ensari, I., Weng, C., Kahn, M. G., & Natarajan, K. (2021). Factors Affecting the Quality of Person-Generated Wearable Device Data and Associated Challenges: Rapid Systematic Review. *JMIR mHealth and uHealth*, *9*(3), e20738. https://doi.org/10.2196/20738

Coeckelbergh, M. (2020). *AI Ethics (The MIT Press Essential Knowledge series)*. The MIT Press.

Cooper, D. R., Schindler, P. S., & Sharma, J. K. (2018). *Business Research Methods*. McGraw-Hill Education.

Corporate Finance Institute. (2021, September 14). *Top Banks in the Netherlands*. https://corporatefinanceinstitute.com/resources/careers/companies/top-banks-in-the-netherlands/?msclkid=4ed05cfbb58c11ec80ffd6524e3153ee

da Silva, R. F. G., Roy, C. K., Rahman, M. M., Schneider, K. A., Paixão, K., Dantas, C. E. D. C., & Maia, M. D. A. (2020). CROKAGE: effective solution recommendation for programming tasks by leveraging crowd knowledge. *Empirical Software Engineering*, *25*(6), 4707–4758. https://doi.org/10.1007/s10664-020-09863-2

Dai, Y., Xiang, Y., Li, Y., Xing, L., & Zhang, G. (2011). Consequence Oriented Self-Healing and Autonomous Diagnosis for Highly Reliable Systems and Software. *IEEE Transactions on Reliability*, *60*(2), 369–380. https://doi.org/10.1109/tr.2011.2136490

Darr, E. D., & Kurtzberg, T. R. (2000). An Investigation of Partner Similarity Dimensions on Knowledge Transfer. *Organizational Behavior and Human Decision Processes*, *82*(1), 28–44. https://doi.org/10.1006/obhd.2000.2885

Dayarathna, M., Akmeemana, P., Perera, S., & Jayasinghe, M. (2017). Demo: Solution Recommender for System Failure Recovery via Log Event Pattern Matching on a Knowledge Graph. *Proceedings of DEBS '17*.

De Nederlandsche Bank. (2022). *Witwassen en crimineel geld bestrijden*. www.dnb.nl. https://www.dnb.nl/betrouwbare-financiele-sector/witwassen-en-crimineel-geld-bestrijden/?faq=V2F0IghvdWR0IghldCBrZW4tdXcta2xhbnQgcHJpbmNpcGUgaW4_IA

Doughty, C. (2005). Know your customer. *Business Information Review*, *22*(4), 248–252. https://doi.org/10.1177/0266382105060603

Dove, R. (1996). Agile knowledge transfer: Reusable, reconfigurable, scalable. *Production*.

Durham, D., & Michel, C. (2021). *Lean Software Systems Engineering for Developers: Managing Requirements, Complexity, Teams, and Change Like a Champ* (1st ed.). Apress.

Fogla, P., Sharif, M. I., Perdisci, R., Kolesnikov, O. M., & Lee, W. (2006). Polymorphic blending attacks. *USENIX Security Symposium*.

Girault, C., & Valk, R. (2003). *Petri Nets for Systems Engineering: A Guide to Modeling, Verification, and Applications* (Softcover reprint of hardcover 1st ed. 2003 ed.). Springer.

Google (2022). *DETECTLANGUAGE - Google Docs Editors Help*. https://support.google.com/docs/answer/3093278?hl=en

GDPR.eu. (2022). *General Data Protection Regulation (GDPR) Compliance Guidelines*. https://gdpr.eu/

Gugiu, P. C., & Rodríguez-Campos, L. (2007). Semi-structured interview protocol for constructing logic models. *Evaluation and Program Planning*, *30*(4), 339–350. https://doi.org/10.1016/j.evalprogplan.2007.08.004

Guille, A., & Favre, C. (2015). Event detection, tracking, and visualization in Twitter: a mention-anomaly-based approach. *Social Network Analysis and Mining*, *5*(1). https://doi.org/10.1007/s13278-015-0258-0

Gupta, N., & Mangla, R. (2020). *Artificial Intelligence Basics: A Self-Teaching Introduction*. Mercury Learning and Information.

Guven, S., Murthy, K., Shwartz, L., & Paradkar, A. (2016). Towards establishing causality between change and incident. *NOMS 2016 – 2016 IEEE/IFIP Network Operations and Management Symposium*. https://doi.org/10.1109/noms.2016.7502929

Haita-Falah, C. (2017). Sunk-cost fallacy and cognitive ability in individual decision-making. *Journal of Economic Psychology*, *58*, 44–59. https://doi.org/10.1016/j.joep.2016.12.001

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining*. Morgan Kaufmann.

Harinarayanan, V. P. (2021). *Building the Modern Workplace with SharePoint Online: Solutions with SPFx, Power Automate, Power Apps, Teams, and PVA* (1st ed.). Apress.

Hassenstein, M. J., & Vanella, P. (2022). Data Quality—Concepts and Problems. *Encyclopedia*, *2*(1), 498–510. https://doi.org/10.3390/encyclopedia2010032

Heerkens, H., & Winden, V. A. (2021). *Solving Managerial Problems Systematically (Routledge-Noordhoff International Editions)* (1st ed.). Routledge.

Heras, S., García-Pardo, J. N., Ramos-Garijo, R., Palomares, A., Botti, V., Rebollo, M., & Julián, V. (2009). Multi-domain case-based module for customer support. *Expert Systems with Applications*, *36*(3), 6866–6873. https://doi.org/10.1016/j.eswa.2008.08.003

Jannach, D. (2010). *Recommender Systems: An Introduction* (1st ed.). Cambridge University Press.

Luo, C., Lou, J., Lin, Q., Fu, Q., Ding, R., Zhang, D., & Wang, Z. (2014). Correlating Events with Time Series for Incident Diagnosis. *20th ACM SigKDD Conference*.

Machado, C., & Davim, P. J. (2021). *Knowledge Management and Learning Organizations* (1st ed. 2021 ed.). Springer.

Mahdavi-Hezavehi, S., Durelli, V. H., Weyns, D., & Avgeriou, P. (2017). A systematic literature review on methods that handle multiple quality attributes in maharchitecture-based self-adaptive systems. *Information and Software Technology*, *90*, 1–26. https://doi.org/10.1016/j.infsof.2017.03.013

Mallick, A., Hsieh, K., Arzani, B., & Joshi, G. (2022). Matchmaker: Data Drift Mitigation in Machine Learning for Large-Scale Systems. *Proceedings of Machine Learning and Systems 4 (MLSys 2022)*.

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval* (1st ed.). Cambridge University Press.

Marcu, P., Grabarnik, G., Luan, L., Rosu, D., Shwartz, L., & Ward, C. (2009). Towards an optimized model of incident ticket correlation. *2009 IFIP/IEEE International Symposium on Integrated Network Management*.

Màrquez, L., Carreras, X., Litkowski, K. C., & Stevenson, S. (2008). Semantic Role Labeling: An Introduction to the Special Issue. *Computational Linguistics*, *34*(2), 145–159. https://doi.org/10.1162/coli.2008.34.2.145

Mehrotra, K. G., Mohan, C. K., & Huang, H. (2018). *Anomaly Detection Principles and Algorithms (Terrorism, Security, and Computation)* (1st ed. 2017 ed.). Springer.

Melkas, H., & Harmaakorpi, V. (2008). Data, information and knowledge in regional innovation networks. *European Journal of Innovation Management*, *11*(1), 103–124. https://doi.org/10.1108/14601060810845240

Micro Focus. (2022). *IT Service Manager Software*. Microfocus.com. https://www.microfocus.com/en-us/products/servicemanager/overview

Microsoft. (2022a). *Compare Microsoft 365 E3, E5 & F3 | Microsoft 365 Enterprise*. https://www.microsoft.com/en-gb/54abobank54-365/compare-microsoft-365-enterprise-plans

Microsoft. (2022b). *What is DevOps? DevOps Explained*. Microsoft Azure. https://azure.microsoft.com/en-gb/overview/what-is-devops/#culture

Microsoft. (2022c, February 9). *Add custom work item type to inherited process – Azure DevOps Services*. Microsoft Docs. https://docs.microsoft.com/en-us/azure/devops/organizations/settings/work/add-custom-wit?view=azure-devops

Object Management Group. (2006). *Business Process Modeling Notation Specification*.

Object Management Group. (2011). *Business Process Model and Notation (BPMN)*.

Olson, J. E. (2003). *Data Quality: The Accuracy Dimension (The Morgan Kaufmann Series in Data Management Systems)* (1st ed.). Morgan Kaufmann.

Petrenko, S. (2020). *Developing a Cybersecurity Immune System for Industry 4.0 (River Publishers Series in Security and Digital Forensics)*. River Publishers.

Petry, F. E., & Zhao, L. (2009). Data mining by attribute generalization with fuzzy hierarchies in fuzzy databases. *Fuzzy Sets and Systems*, *160*(15), 2206–2223. https://doi.org/10.1016/j.fss.2009.02.014

Project Management Institute. (2017). *Agile Practice Guide* (1st ed.). Project Management Institute.

Rabobank. (2018). *Code of Conduct*. https://www.rabobank.com/en/images/code-of-conduct-rabobank-en.pdf

Rabobank. (2021a). *A new way of working for a rapidly changing world*. Rabobank.Com. https://rabobank.jobs/en/grow-magazine/simplify-scale-a-new-way-of-working-for-a-rapidly-changing-world/?msclkid=8598b550b58d11ec9f5bcfc202525e54

Rabobank. (2021b, October 7). *KYC Data & Analytics*. 54abobank.nl. https://rabobank.jobs/nl/vakgebied/know-your-customer/data-analytics-kyc/

Rabobank. (2022a). *Annual Report 2021*.

Rabobank. (2022b). *Cooperative*. Rabobank.Com. https://www.rabobank.com/en/about-rabobank/cooperative/index.html?msclkid=1d0c6ebda91911ecb6e8a9caa4d27118

Redman, T. C., & Godfrey, B. A. (1996). *Data Quality For The Information Age (Artech House Computer Science Library)* (First Edition). Artech House Publishers.

Resnick, P., & Varian, H. R. (1997). Recommender systems. *Communications of the ACM*, *40*(3), 56–58. https://doi.org/10.1145/245108.245121

Ricci, F., Rokach, L., & Shapira, B. (2022). *Recommender Systems Handbook* (3rd ed. 2022 ed.). Springer.

Rimmer, V., Nadeem, A., Verwer, S., Preuveneers, D., & Joosen, W. (2022). Open-world network intrusion detection. *Security and Artificial Intelligence*.

Salah, S., Maciá-Fernández, G., Díaz-Verdejo, J. E., & Sánchez-Casado, L. (2015). A Model for Incident Tickets Correlation in Network Management. *Journal of Network and Systems Management*, *24*(1), 57–91. https://doi.org/10.1007/s10922-014-9340-6

Samek, W., & Müller, K. (2019). Towards explainable Artificial Intelligence. *Explainable AI*.

Sansone, C., & Harackiewicz, J. M. (2000). *Intrinsic and Extrinsic Motivation: The Search for Optimal Motivation and Performance (Educational Psychology)* (1st ed.). Academic Press.

Santos Júnior, P. S., Barcellos, M. P., Falbo, R. D. A., & Almeida, J. P. A. (2021). From a Scrum Reference Ontology to the Integration of Applications for Data-Driven Software Development. *Information and Software Technology*, *136*, 106570. https://doi.org/10.1016/j.infsof.2021.106570

Scott, D. (2015). Best Practices for Continuous Application Availability. *Gartner IT Summit*.

Slack, N., Brandon-Jones, A., & Johnston, R. (2016). *Operations Management* (8th ed.). Pearson Canada.

Splunk Inc. (2022). *Splunk | The Data Platform for the Hybrid World*. Splunk. https://www.splunk.com/

Tang, L., Shwartz, L., & Grabarnik, G. (2013). Recommending Resolutions for Problems Identified by Monitoring. *2013 IFIP/IEEE International Symposium on Integrated Network Management (IM 2013)*.

Umiker, W. (1997). Removing Barriers to Change. *Laboratory Medicine*, *28*(1), 14–16. https://doi.org/10.1093/labmed/28.1.14

Van de Schoot, R., de Bruin, J., Schram, R., Zahedi, P., de Boer, J., Weijdema, F., Kramer, B., Huijts, M., Hoogerwerf, M., Ferdinands, G., Harkema, A., Willemsen, J., Ma, Y., Fang, Q., Hindriks, S., Tummers, L., & Oberski, D. L. (2021). An open source machine learning framework for efficient and transparent systematic reviews. *Nature Machine Intelligence*, *3*(2), 125–133. https://doi.org/10.1038/s42256-020-00287-7

van Landegem, T., Vankwikelberge, P., & Vanderstraeten, H. (1994). A self-healing ATM network based on multilink principles. *IEEE Journal on Selected Areas in Communications*, *12*(1), 139–148. https://doi.org/10.1109/49.265713

Wang, R. Y., & Strong, D. M. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems*, *12*(4), 5–33. https://doi.org/10.1080/07421222.1996.11518099

Weske, M. (2019). *Business Process Management: Concepts, Languages, Architectures* (3rd ed. 2019 ed.). Springer.

Wilson, C. (2013). *Interview Techniques for UX Practitioners: A User-Centered Design Method* (1st ed.). Morgan Kaufmann.

Wu, Y., Zhao, S., & Li, W. (2020). Phrase2Vec: Phrase embedding based on parsing. *Information Sciences*, *517*, 100–127. https://doi.org/10.1016/j.ins.2019.12.031

Xu, J., Mu, J., & Chen, G. (2020). A multi-view similarity measure framework for trouble ticket mining. *Data & Knowledge Engineering*, *127*, 101800. https://doi.org/10.1016/j.datak.2020.101800

Yager, R. R., & Petry, F. E. (2006). A Multicriteria Approach to Data Summarization Using Concept Ontologies. *IEEE Transactions on Fuzzy Systems*, *14*(6), 767–780. https://doi.org/10.1109/tfuzz.2006.879954

Zeng, C., Zhou, W., Li, T., Shwartz, L., & Grabarnik, G. Y. (2017). Knowledge Guided Hierarchical Multi-Label Classification Over Ticket Data. *IEEE Transactions on Network and Service Management*, *14*(2), 246–260. https://doi.org/10.1109/tnsm.2017.2668363

Zhou, W., Tang, L., Zeng, C., Li, T., Shwartz, L., & Ya. Grabarnik, G. (2016). Resolution Recommendation for Event Tickets in Service Management. *IEEE Transactions on Network and Service Management*, *13*(4), 954–967. https://doi.org/10.1109/tnsm.2016.2587807

Zhou, W., Xue, W., Baral, R., Wang, Q., Zeng, C., Li, T., Xu, J., Liu, Z., Shwartz, L., & Ya. Grabarnik, G. (2017). STAR. *KDD 2017 Applied Data Science Paper*. https://doi.org/10.1145/3097983.3098190

Appendix A Intended recommender system

The goal is to build a recommender system that recommends solutions to problems to operators, using the output of the anomaly detector. This section of the report shortly explains the proposed (by the AIOps team at the time of writing) approach to a content-based recommender system and its implication for the required data quality (Aggarwal, 2016; Ricci et al., 2022). It is important to note that the proposed recommender system is to be implemented within the context of one application, as the anomaly detector used as input, is also deployed within the context of one application due to data differences.

The premise of a comparable system is as follows. One can define three objects for a movie recommender: the users, the movies, and the features of the movies (e.g. different genres, duration, age, director etc.) (Jannach, 2010). If users give a rating (say on a 1-5 scale) to various movies, a rating that the user will give to another movie can be predicted, based on how strong certain features exist in other movies. If the user decides to rate the movie, the model will recalibrate, and adjust the recommended movies accordingly.

In this case, one can define the actions, undertaken in order to solve the problem, as the 'movies', as this is the desired output of the recommender system. Apart from this one can group certain contributors to the anomaly together in different 'user profiles' where the movie features are features that actions possess to a varying degree, to be assessed beforehand by an operator (performance related, false positive, Linux-based, etc.). When an anomaly occurs, the contributors are known due to the anomaly detector, and the anomaly can be attributed to one of the predefined groups (for example, by means of k-nearest-neighbour clustering). Based on this group, a recommendation for a certain action is given. For every action the operator has undertaken in order to solve the problem, they rate that action on how well it worked to resolve the problem. This will then improve the model's knowledge and thus its outcome.

Table 6 gives a simple identification of how various profiles have ratings for certain actions and how the action has different values for predefined features. Based on these features a calculation for *X* can be made. Here *X* is a new anomaly of the type '*Profile 3*', of which it is desirable to know how the action '*reset database*' would affect this anomaly. The question marks represent other, for now irrelevant unknown ratings. It must be noted that the given actions and features are to give an idea of what these concepts represent, and do not necessarily represent actions and features that are to be included in the recommender system, nor do the given values.

Table 15 Example of ratings of certain actions and profiles

Actions/Profiles	**Profile 1**	**Profile 2**	**Profile 3**	**Profile 4**	**Features**	**Linux-based**	**False positive**	**Database related**
Do nothing	5	1	1	2		0.1	1	0
Clear memory drive	2	4	4	?		0.9	0	0.2
Reset database	?	3	X	4		0.1	0.2	1
Restart server	2	1	5	3		0.1	0.2	0.4

Based on this, the solution concepts must at least contain a way for operators to give ratings to different actions. This is currently impossible in the used Service Manager software.

Appendix B Interview protocol

Table 16 Interview protocol

Title interview:		
With:		
Date:		
Main goal:		
Main topic	**Questions to ask**	**Time Frame (duration)**
Introduction	Who am I? + context IEM	2
	What is my assignment?	
	What is my goal for this interview?	
	How long? (They probably already know due to planning)	
	What does the interview look like?	
Easing into	What is your current job?	
	In what domain/tribe?	
	What is your background? Study? Previous work?	
Identify key informants	Who do you think I should also talk to about this project? (General)	3
	Who can I contact to get information about… ? (specific part)	
	Who should I keep in mind with regard to this project?	
Basic background and context	What is your task in this process?	10
	What is the history of the process?	
	What is the capacity of the process?	
	What is the goal of the process?	
Service outcome	What is the outcome of the process?	5
	What domain/tribe benefits from this process? And in what way?	
	How is the customer influenced by the process?	
Activities and outputs	Who does what, and when?	10
	Based on who I am talking to: Different micro/macro levels	
Inputs	What resources are used to support the process?	2
	Is there a gap between resources needed and available?	
Importance	How critical is this process?	2
	Where are risks/bottlenecks in this process?	
	Assuming this process stops, what happens?	
Closing	Do you have any questions for me?	3 – Note that at least 15 minutes extra must be calculated to ask further questions
	Can I contact you for further questions?	
	Is there anything else you need to know?	
	How did you experience this interview?	

Approaches

Cooper et al. (2014) differentiate three different approaches to an individual depth interview. The first of these is an unstructured interview. In this approach, no specific questions or order of topics are noted or followed. In general, these interviews start with a participant narrative. In preparation, a list of topics is created.

The second approach is a semi-structured interview. These types of interviews generally start with a few specific questions and then follow the individual's thoughts by probes of the interviewer. A semi-structured interview generally has a list of questions that require asking, but no specifics as to what order these questions should be asked in.

The last approach is the structured interview. This approach is mostly applied in quantitative research, where the interactivity between interviewer and interviewee is of lesser importance. A list of questions is presented as well as the order in which these questions will be asked. The structured review is better equipped to compare the answers from different interviewees, as is desired for a quantitative approach.

In the research of this study, mostly qualitative interviews are held. Only the evaluative phase will have research that can be described as quantitative. As a result of this, the structured approach is deemed unfit. Using this approach loses the interactivity that is necessary to get the explanations and reasoning behind the interviewees' answers. This does lose the ability to easily compare answers from different interviewees, but this is of lesser importance.

This leaves the unstructured and the semi-structured approach. Keeping the learning goal in mind, it is expected that the semi-structured approach will yield the greatest results. This is due to it having a clear guideline of questions, in which the research goal of the interview can be expressed. The semi-structured approach balances the interactivity of the unstructured approach and the basic structure of the structured approach.

Protocol

An interesting approach based on the semi-structured approach is the semi-structured interview protocol for logic models (SSIP) as proposed by Gugiu and Rodriguez-Campos (2007). They developed a protocol with specific research questions based on other interview models. The application of the protocol on logic models is one that suits my assignment quite well, considering that the process needs to be defined in a logical model, where the primary input is data collected from interviews. It is important to note that the main goal of the SSIP is to evaluate processes, not define them. Apart from this, to minimise the time an interview takes, it is necessary to be critical of what questions need to be included in the interview protocol (to be found in Table 7). The bulk of the protocol is based on the SSIP, however, it must be noted that the main goal of the SSIP is to evaluate a model. Questions regarding an evaluation are mainly ignored, to ensure the brevity and versatility of the protocol. It must also be taken into consideration that the protocol is designed for programs, such as a youth mentoring program, and not necessarily for services. The protocol is adjusted accordingly.

Apart from this Cooper et al. (2014), but more importantly Wilson (2013) describes a method to use the semi-structured interview. First, Wilson (2013) distinguishes a general schedule for such an interview.

- Introduction of oneself, purpose, and topic of the interview
- A list of topics and questions to ask about the topics
- Suggested probes and prompts

- Closing comments

The general structure of the framework is based on this paper from Wilson (2013). This includes the structure of the research goal, topics, questions, and the time it is approximated to take. An important takeaway from Wilson (2013) is the importance of a short interview briefing with the interviewee, to describe what is the goal, how long it is roughly going to take, and what topics will be discussed. Another takeaway is to ease the interviewee into the interview. A way to do this is to ask about their background, and what they are currently working on. This also shows the context of this specific person to the project. It is important to note that due to the versatility of the framework, there is little sense in absolutely pinning down the time frame of certain parts of the interview. Instead, a rough estimate is given, which can be adjusted if necessary.

Appendix C Measurements protocol

The following appendix aims to give transparency regarding the measurements taken in section 1.5 by providing the performed actions that led to the presented values

All measurements are taken from one month of incident data, as registered in Service Manager. This data has been extracted to Excel in a different process and using this extraction, the data has been analysed. For the following measurements, the 'Solution' field has been analysed. Besides this field, one could analyse the 'Description' field, however considering that this field is often (partially) filled with machine-generated data, the choice has been made to only measure the 'Solution' field.

The extraction to an Excel file means that certain entries cannot be viewed due to the General Data Protection Regulation (GDPR) of the European Union (GDPR.eu, 2022). These 577 entries have been removed from the 46,207 total entries. In order to not manually review 45,630 entries, the data has been sorted on the number of times that an entry occurs. This provides 18,427 unique entries, which is still considered too large to manually analyse. Therefore a cut-off point is determined. Reducing the data to only the entries that occur five times or more, gives 802 unique entries (representing 25,637 entries total). This is considered to be satisfactory in both the number of different entries, as well as the significance towards the calculated number. Reduced data refers to this set of data entries with at least 5 occurrences. It must be noted that for the calculations the total number of entries (45,630) is used, since unique values are likelier to contain valuable data. Using the reduced data would therefore skew the data in favour of the researcher.

Completeness

Batini and Scannapieco (2006) define data completeness as "the extent to which data is of sufficient depth, breadth, and scope for the task at hand". As such, in this study, data completeness is measured as the number of fields that are considered non-empty.

From the reduced data, it has been manually (that is, by the author) decided whether the entries have some form of solution filled in. The main criteria here is that if there is any data that is considered to say something about the solution, however valuable that entry is, the field is considered filled in. (e.g. "reset", "user error"). False positive or tests are not taken into account for this calculation as these are considered to present information ('do nothing' is still a solution). Table 8 provides the entries that were considered empty, as well as their count. The entries a fully copied unless specifically mentioned otherwise. Entries made bold are also considered non-English for the measurement of interpretability.

Table 17 Data entries

String	Count
Empty[2]	7,367
opgelost	419
Closing the incident	382
Completed manually	341
solved	123
.	90
Issue has been fixed now.	83
done	82
Issue has been fixed.	78
offline	73

[2] This entry is not literally "Empty" but rather not filled in with a value at all

x	54
required	47
fixed	47
Solved.	44
Issue Resolved	43
closed	43
complete	37
resolved	36
Uitgevoerd	35
All good now.	34
Done	34
Opgelost.	34
User confirmed issue is resolved	32
/	32
Server is up	18
Solved	16
all done	16
all ok	16
completed	16
on	16
-	15
by hand	13
The issue is resolved. Hence, closing the incident.	13
DONE -Albert[3]	13
manual	13
Complete OK	12
opgelost	12
Resoled	10
Unknown	9
n/a	9
close	9
Resolved	9
Issue resolved	8
ok	8
Completed	8
na	7
All fine now	7
Fixed	7
resolved.	7
Closed without customer, financial, or user impact	6
already done	6
Temporary issue, resolved now	5
The issue has been resolved	5
fixed.	5
Total	9,934

Therefore the completeness as the percentage of fields filled in is $100\% - \frac{9{,}934}{45{,}630} = 78.2\%$

[3] Name has been changed in order to anonymise

Interpretability

Batini and Scannapieco (2006) define data completeness as "the extent to which data is of sufficient depth, breadth, and scope for the task at hand". As such, in this study, data interpretability is measured as the percentage of fields that are in English.

Over all the entries the Google Sheets function DETECTLANGUAGE is used on the entries (Google, 2022). Here all entries are used (save the censored due to GDPR ones), as there is no manual process necessary. The results are presented in Table 18. It must be noted that initially there was a larger array of languages (eg. Afrikaans, Quechua, Danish). The smaller languages have been manually investigated and have been redistributed if deemed necessary (216 entries). Table 18 is the result of this redistribution.

Table 18 Counts of major languages

Language	**Count**
English	30491
Dutch	7227
Portuguese	302
Spanish	4
Total	38024
Empty	7367
Undefined	239

In Table 18, 'Empty' refers to the cell being completely empty. Undefined means that no language can be detected, either as a result of links being entered, or no text (eg. "?", "/"). These values are taken out of consideration for the calculation.

Therefore, the interpretability as the percentage of fields filled in English is $\frac{30491}{38024} = 83.3\%$

Relevancy

Batini and Scannapieco (2006) define data completeness as "the extent to which data is applicable and useful for the task at hand". Considering that one of the main tasks at hand is the construction of a solution recommender, data can be considered more useful if it is standardised to a certain degree. This is difficult to precisely measure in a meaningful way, especially taking into account the technical knowledge that is required to analyse fields. Therefore, the following measurement is not necessarily meant as a precise figure, but rather as an example of how the data can be of low quality due to poor standardisation. As such, in this study, data relevancy is measured as the number of different entries used for the concept 'false positive'.

The concept that was chosen to show how many different entries are used is 'false positive': an incident that was created because a certain threshold was exceeded, however there is no actual incident occurring. One of the reasons that this concept was chosen is that it is easy to recognise without having extensive knowledge of the applications. The second reason is that it is a concept that can relatively easily be generalised into one term without losing too much details. Similarly, the solution will mostly be uniform as well: doing nothing.

To measure this, terms are manually selected from the reduced dataset (at least 5 occurrences). The reduced dataset is used to keep the manual labour at a manageable level. The selected terms are terms that are considered to be some form of 'false positive', excluding types of tests. Entries that are measured for data completeness are not measured for relevancy. Table 19 shows the selected terms, as well as a manually (that is, by the researcher) assigned group. The terms are fully copied, unless mentioned otherwise

Table 19 Various entries that are considered as 'false positive'

Term	Group
ABC[4] is working fine.	Working fine
Application is working fine.	Working fine
The application is up and working fine	Working fine
Working fine.	Working fine
false positive alarms.	False positive
software upgraded. false positive alarms.	False positive
False/positives because of update	False positive
false positive	False positive
False positive alarm. Closing.	False positive
False positives, no impact	False positive
Due to false-positive reported during maintenance	False positive
False alarm	False positive
False alert	False positive
Software upgraded, false positive alarms.	False positive
service disruption, all service were up and running	Up and running
all services are up and running	Up and running
Up n running.	Up and running
Network hiccup	Hiccup
Hiccup	Hiccup
Checked, saw nothing, maybe a hick-up or something?	Hiccup
Hick-ups.	Hiccup
Network hiccup	Hiccup
Due to network hiccup	Hiccup
XYZ[5] hickup	Hiccup
Due to a network hiccup	Hiccup
Its only a warning – no ticket needed	Warning
Warning	Warning
Expected to be down	Expected
These are expected to e down at this time	Expected
Task is expected to be down at this time	Expected
No issues found	No Issue
No issue	No Issue
No issue	No issue
No actions needed, reboots	No action needed
No action needed	No action needed
Compliant	/
All results normal.	/

All of the 37 terms in table 19 are some variation of 'false positive' and can be divided into 10 groups which are considered to be difficult to relate to each other using some form of NLP. These groups are not an objective measurement, but aim to show that for a relatively generic concept, a variety of terms

[4] Not the actual name of the application

is used. Although 10 terms may not seem as too much trouble, this is only one concept, totalling only 1.9% of the total amount of tickets.

Appendix D Summary of the theoretical framework

Table 20 Summarisation of key takeaways of the theoretical framework

Subject	Key takeaway	Main sources
Knowledge transfer	If an improved incident registration process is implemented, it is paramount to stress the strategic similarity between AIOps and the user's domain.	Dove (1996) Darr & Kurtzberg (2000)
Data quality dimensions	Data quality can be divided into different categories, which can further be divided into dimensions to assess different parts of data quality.	Wang and Strong (1996) Batini and Scannapieco (2006)
Modelling language	The BPMN provides a modelling language that is flexible rather than restrictive, providing great usability to the modeller.	Weske (2019) Object Management Group (2011)
Data generalisation	Data generalisation needs to balance the usefulness of the generalisation and the loss of specificity. Concept hierarchies can help in this.	Han et al. (2012) Yager & Petry (2006) Petry & Zhao (2009)
Lean management	Following the principles of lean management, automation must be applied where appropriate	Slack et al. (2016) Durham & Michel (2021) Bauer (2016)

Appendix E Questionnaire

All of the below questions are questions in which the given options are ranked based on which of the solution conforms to the given statement the strongest. Rankings are chosen as opposed to scores or individual assessments to force the participants to compare the various concepts, instead of choosing one 'good' and rating the other simply as bad, without any distinctions (Cooper et al., 2014). After every question, people are given the option to fill in a "No opinion" option. For the sake of brevity, these are not included in the overview of the questions. Besides, once again for the sake of brevity, the options which are to be ranked are only given for the first question. The criteria that are given for the questions were not presented to the participants, in order to not influence the answers of the individual question.

AIOps-specific concepts

Ease of implementation

1. I believe this concept can be implemented in a short time frame
    a. Azure DevOps boards
    b. Pure data mining
    c. Automatically Send Forms
    d. Splunk dashboard
    e. Web-based application
2. I believe this concept requires little labour to implement.
3. I believe this concept requires little expert knowledge to implement.
4. I believe this concept costs little money to implement.

Effectiveness and efficiency

5. I believe this concept is effective in gathering the data required for a solution recommender.
6. I believe this concept is versatile (i.e. can handle different types of data).
7. I believe this concept is flexible (i.e. can easily be changed according to new requirements).
8. I believe this concept is NOT over-engineered (i.e. too complicated or too capable for the task at hand).

Operator usage

9. I believe this concept can easily be added to the work routine of an operator.
10. I believe this concept will be properly used by operators (i.e. taking the time to fill in the solutions, ratings etc.)
11. I believe this concept will have a low perceived annoyance at the side of the operator.
12. I believe operators have the required knowledge to make use of this concept.

13. Overall, I believe this to be the superior concept.

Non-AIOps-specific concepts

Ease of implementation

1. I believe this concept can be implemented in a short time frame
    a. Applying templates
    b. Delinking solution to customer notification
    c. Warnings of poor quality information
    d. Description and solution generalisation
    e. Improved link with Problem Management and Knowledge Management
2. I believe this concept requires little labour to implement.
3. I believe this concept requires little expert knowledge to implement.
4. I believe this concept costs little money to implement.

Effectiveness and efficiency

5. I believe this concept is effective in improving the data quality.
6. I believe this concept will require little differentiation between departments.
7. I believe this concept can handle the required differentiation between departments.
8. I believe this concept can be properly implemented in Service Manager.

Operator usage

9. I believe operators will adhere to this concept (i.e. use the concept voluntarily and properly).
10. I believe the perceived annoyance of this concept will be low.

11. Overall, I believe this to be the superior concept.

Appendix F Responses

This appendix covers all the responses as given by the participants. In order to ensure the privacy that was promised to the participants, the participants have been anonymised. Considering that the number of participants from a single group is low, the groups have been omitted for the same reason.

I believe this concept requires little labour to implement.					
Rating → /Participant	**1**	**2**	**3**	**4**	**5**
A	Pure data mining	Azure DevOps boards	Automatically Send Forms	Splunk dashboard	Web-based application
B	Pure data mining	Automatically Send Forms	Azure DevOps boards	Web-based application	Splunk dashboard
C	Azure DevOps boards	Splunk dashboard	Automatically Send Forms	Web-based application	Pure data mining
D	Automatically Send Forms	Pure data mining	Azure DevOps boards	Splunk dashboard	Web-based application
E	Azure DevOps boards	Automatically Send Forms	Splunk dashboard	Pure data mining	Web-based application
F	Pure data mining	Web-based application	Azure DevOps boards	Splunk dashboard	Automatically Send Forms
G	Automatically Send Forms	Azure DevOps boards	Splunk dashboard	Web-based application	Pure data mining
H	Splunk dashboard	Azure DevOps boards	Automatically Send Forms	Web-based application	Pure data mining
I	Pure data mining	Automatically Send Forms	Azure DevOps boards	Web-based application	Splunk dashboard

I believe this concept requires little labour to implement.					
Rating → /Participant	**1**	**2**	**3**	**4**	**5**
A	Pure data mining	Azure DevOps boards	Automatically Send Forms	Splunk dashboard	Web-based application
B	Pure data mining	Automatically Send Forms	Azure DevOps boards	Web-based application	Splunk dashboard
C	Azure DevOps boards	Splunk dashboard	Automatically Send Forms	Web-based application	Pure data mining
D	Automatically Send Forms	Pure data mining	Azure DevOps boards	Splunk dashboard	Web-based application
E	Azure DevOps boards	Automatically Send Forms	Splunk dashboard	Pure data mining	Web-based application
F	Pure data mining	Web-based application	Azure DevOps boards	Splunk dashboard	Automatically Send Forms
G	Automatically Send Forms	Azure DevOps boards	Splunk dashboard	Web-based application	Pure data mining
H	Splunk dashboard	Azure DevOps boards	Automatically Send Forms	Web-based application	Pure data mining
I	Pure data mining	Automatically Send Forms	Azure DevOps boards	Web-based application	Splunk dashboard

I believe this concept requires little expert knowledge to implement.					
Rating → /Participant	**1**	**2**	**3**	**4**	**5**
A	Web-based application	Automatically Send Forms	Splunk dashboard	Pure data mining	Azure DevOps boards
B	Automatically Send Forms	Azure DevOps boards	Pure data mining	Web-based application	Splunk dashboard
C	Azure DevOps boards	Web-based application	Automatically Send Forms	Splunk dashboard	Pure data mining
D	Splunk dashboard	Azure DevOps boards	Automatically Send Forms	Pure data mining	Web-based application
E	Azure DevOps boards	Automatically Send Forms	Pure data mining	Splunk dashboard	Web-based application
F	Pure data mining	Web-based application	Azure DevOps boards	Splunk dashboard	Automatically Send Forms
G	Automatically Send Forms	Splunk dashboard	Azure DevOps boards	Web-based application	Pure data mining
H	Pure data mining	Web-based application	Splunk dashboard	Azure DevOps boards	Automatically Send Forms
I	Pure data mining	Automatically Send Forms	Azure DevOps boards	Web-based application	Splunk dashboard

I believe this concept costs little money to implement.					
Rating → /Participant	**1**	**2**	**3**	**4**	**5**
A	Pure data mining	Automatically Send Forms	Azure DevOps boards	Splunk dashboard	Web-based application
B	Pure data mining	Automatically Send Forms	Azure DevOps boards	Splunk dashboard	Web-based application
C	Azure DevOps boards	Automatically Send Forms	Pure data mining	Splunk dashboard	Web-based application
D	Splunk dashboard	Azure DevOps boards	Automatically Send Forms	Pure data mining	Web-based application
E	Azure DevOps boards	Automatically Send Forms	Splunk dashboard	Pure data mining	Web-based application
F	Pure data mining	Web-based application	Azure DevOps boards	Splunk dashboard	Automatically Send Forms
G	Automatically Send Forms	Azure DevOps boards	Splunk dashboard	Pure data mining	Web-based application
H	Splunk dashboard	Azure DevOps boards	Automatically Send Forms	Web-based application	Pure data mining
I	Automatically Send Forms	Pure data mining	Azure DevOps boards	Splunk dashboard	Web-based application

I believe this concept is effective in gathering the data required for a solution recommender.					
Rating → /Participant	**1**	**2**	**3**	**4**	**5**
A	Pure data mining	Automatically Send Forms	Azure DevOps boards	Splunk dashboard	Web-based application
B	Azure DevOps boards	Automatically Send Forms	Splunk dashboard	Pure data mining	Web-based application
C	Pure data mining	Automatically Send Forms	Web-based application	Azure DevOps boards	Splunk dashboard
D	Web-based application	Automatically Send Forms	Azure DevOps boards	Splunk dashboard	Pure data mining
E	Web-based application	Splunk dashboard	Pure data mining	Azure DevOps boards	Automatically Send Forms
F	Pure data mining	Web-based application	Automatically Send Forms	Splunk dashboard	Azure DevOps boards
G	Pure data mining	Web-based application	Splunk dashboard	Azure DevOps boards	Automatically Send Forms
H	Pure data mining	Automatically Send Forms	Splunk dashboard	Azure DevOps boards	Web-based application
I	Web-based application	Pure data mining	Splunk dashboard	Automatically Send Forms	Azure DevOps boards

I believe this concept is versatile (i.e. can handle different types of data).					
Rating → /Participant	**1**	**2**	**3**	**4**	**5**
A	Pure data mining	Automatically Send Forms	Azure DevOps boards	Splunk dashboard	Web-based application
B	Automatically Send Forms	Azure DevOps boards	Splunk dashboard	Pure data mining	Web-based application
C	Pure data mining	Automatically Send Forms	Web-based application	Azure DevOps boards	Splunk dashboard
D	Web-based application	Splunk dashboard	Automatically Send Forms	Azure DevOps boards	Pure data mining
E	Web-based application	Pure data mining	Automatically Send Forms	Splunk dashboard	Azure DevOps boards
F	Pure data mining	Web-based application	Azure DevOps boards	Splunk dashboard	Automatically Send Forms
G	Pure data mining	Web-based application	Automatically Send Forms	Splunk dashboard	Azure DevOps boards
H	Azure DevOps boards	Splunk dashboard	Pure data mining	Automatically Send Forms	Web-based application
I	Web-based application	Pure data mining	Splunk dashboard	Automatically Send Forms	Azure DevOps boards

I believe this concept is flexible (i.e. can easily be changed according to new requirements).					
Rating → /Participant	**1**	**2**	**3**	**4**	**5**
A	Pure data mining	Automatically Send Forms	Azure DevOps boards	Splunk dashboard	Web-based application
B	Automatically Send Forms	Azure DevOps boards	Splunk dashboard	Pure data mining	Web-based application
C	Pure data mining	Automatically Send Forms	Web-based application	Azure DevOps boards	Splunk dashboard
D	Web-based application	Automatically Send Forms	Splunk dashboard	Azure DevOps boards	Pure data mining
E	Web-based application	Pure data mining	Splunk dashboard	Automatically Send Forms	Azure DevOps boards
F	Automatically Send Forms	Splunk dashboard	Web-based application	Azure DevOps boards	Pure data mining
G	Automatically Send Forms	Web-based application	Azure DevOps boards	Splunk dashboard	Pure data mining
H	Azure DevOps boards	Splunk dashboard	Web-based application	Automatically Send Forms	Pure data mining
I	Web-based application	Automatically Send Forms	Pure data mining	Azure DevOps boards	Splunk dashboard

I believe this concept is NOT over-engineered (i.e. too complicated or too capable for the task at hand).					
Rating → /Participant	**1**	**2**	**3**	**4**	**5**
A	Pure data mining	Azure DevOps boards	Automatically Send Forms	Splunk dashboard	Web-based application
B	Pure data mining	Automatically Send Forms	Azure DevOps boards	Splunk dashboard	Web-based application
C	Web-based application	Splunk dashboard	Azure DevOps boards	Automatically Send Forms	Pure data mining
D	Web-based application	Automatically Send Forms	Pure data mining	Splunk dashboard	Azure DevOps boards
E	Automatically Send Forms	Azure DevOps boards	Pure data mining	Splunk dashboard	Web-based application
F	Pure data mining	Splunk dashboard	Azure DevOps boards	Web-based application	Automatically Send Forms
G	Automatically Send Forms	Splunk dashboard	Azure DevOps boards	Pure data mining	Web-based application
H	Splunk dashboard	Azure DevOps boards	Web-based application	Pure data mining	Automatically Send Forms
I	Pure data mining	Web-based application	Splunk dashboard	Azure DevOps boards	Automatically Send Forms

I believe this concept can easily be added to the work routine of an operator.					
Rating → /Participant	**1**	**2**	**3**	**4**	**5**
A	Automatically Send Forms	Splunk dashboard	Azure DevOps boards	Web-based application	Pure data mining
B	Pure data mining	Automatically Send Forms	Azure DevOps boards	Splunk dashboard	Web-based application
C	Splunk dashboard	Azure DevOps boards	Automatically Send Forms	Web-based application	Pure data mining
D	Splunk dashboard	Azure DevOps boards	Automatically Send Forms	Web-based application	Pure data mining
E	Azure DevOps boards	Pure data mining	Splunk dashboard	Automatically Send Forms	Web-based application
F	Splunk dashboard	Azure DevOps boards	Automatically Send Forms	Pure data mining	Web-based application
G	Pure data mining	Splunk dashboard	Azure DevOps boards	Web-based application	Automatically Send Forms
H	Automatically Send Forms	Azure DevOps boards	Splunk dashboard	Web-based application	Pure data mining
I	Splunk dashboard	Automatically Send Forms	Web-based application	Pure data mining	Azure DevOps boards

I believe this concept can easily be added to the work routine of an operator.					
Rating → /Participant	**1**	**2**	**3**	**4**	**5**
A	Automatically Send Forms	Splunk dashboard	Azure DevOps boards	Web-based application	Pure data mining
B	Pure data mining	Automatically Send Forms	Azure DevOps boards	Splunk dashboard	Web-based application
C	Splunk dashboard	Azure DevOps boards	Automatically Send Forms	Web-based application	Pure data mining
D	Splunk dashboard	Azure DevOps boards	Automatically Send Forms	Web-based application	Pure data mining
E	Azure DevOps boards	Pure data mining	Splunk dashboard	Automatically Send Forms	Web-based application
F	Splunk dashboard	Azure DevOps boards	Automatically Send Forms	Pure data mining	Web-based application
G	Pure data mining	Splunk dashboard	Azure DevOps boards	Web-based application	Automatically Send Forms
H	Automatically Send Forms	Azure DevOps boards	Splunk dashboard	Web-based application	Pure data mining
I	Splunk dashboard	Automatically Send Forms	Web-based application	Pure data mining	Azure DevOps boards

I believe this concept will be properly used by operators (i.e. taking the time to fill in the solutions, ratings etc.)					
Rating → /Participant	**1**	**2**	**3**	**4**	**5**
A	Automatically Send Forms	Azure DevOps boards	Web-based application	Pure data mining	Automatically Send Forms
B	Automatically Send Forms	Pure data mining	Splunk dashboard	Web-based application	Automatically Send Forms
C	Automatically Send Forms	Azure DevOps boards	Web-based application	Pure data mining	Automatically Send Forms
D	Azure DevOps boards	Web-based application	Automatically Send Forms	Pure data mining	Azure DevOps boards
E	Pure data mining	Splunk dashboard	Azure DevOps boards	Automatically Send Forms	Pure data mining
F	Pure data mining	Web-based application	Azure DevOps boards	Automatically Send Forms	Pure data mining
G	Azure DevOps boards	Web-based application	Splunk dashboard	Automatically Send Forms	Azure DevOps boards
H	Splunk dashboard	Azure DevOps boards	Web-based application	Pure data mining	Splunk dashboard
I	Automatically Send Forms	Pure data mining	Azure DevOps boards	Splunk dashboard	Automatically Send Forms

I believe this concept will have a low perceived annoyance at the side of the operator.					
Rating → /Participant	**1**	**2**	**3**	**4**	**5**
A	Pure data mining	Splunk dashboard	Automatically Send Forms	Web-based application	Azure DevOps boards
B	Pure data mining	Azure DevOps boards	Splunk dashboard	Automatically Send Forms	Web-based application
C	Pure data mining	Web-based application	Automatically Send Forms	Azure DevOps boards	Splunk dashboard
D	Splunk dashboard	Automatically Send Forms	Web-based application	Azure DevOps boards	Pure data mining
E	Splunk dashboard	Azure DevOps boards	Pure data mining	Automatically Send Forms	Web-based application
F	Pure data mining	Web-based application	Splunk dashboard	Azure DevOps boards	Automatically Send Forms
G	Pure data mining	Splunk dashboard	Azure DevOps boards	Web-based application	Automatically Send Forms
H	Splunk dashboard	Azure DevOps boards	Automatically Send Forms	Web-based application	Pure data mining
I	Pure data mining	Azure DevOps boards	Splunk dashboard	Automatically Send Forms	Web-based application

I believe operators have the required knowledge to make use of this concept.					
Rating → /Participant	**1**	**2**	**3**	**4**	**5**
A	Automatically Send Forms	Azure DevOps boards	Splunk dashboard	Web-based application	Pure data mining
B	Pure data mining	Automatically Send Forms	Azure DevOps boards	Splunk dashboard	Web-based application
C	Automatically Send Forms	Azure DevOps boards	Splunk dashboard	Web-based application	Pure data mining
D	Splunk dashboard	Azure DevOps boards	Automatically Send Forms	Web-based application	Pure data mining
E	Pure data mining	Automatically Send Forms	Azure DevOps boards	Splunk dashboard	Web-based application
F	Splunk dashboard	Azure DevOps boards	Web-based application	Automatically Send Forms	Pure data mining
G	Pure data mining	Azure DevOps boards	Splunk dashboard	Automatically Send Forms	Web-based application
H	Azure DevOps boards	Splunk dashboard	Automatically Send Forms	Web-based application	Pure data mining
I	Web-based application	Automatically Send Forms	Splunk dashboard	Azure DevOps boards	Pure data mining

Overall, I believe this to be the superior concept.					
Rating → /Participant	**1**	**2**	**3**	**4**	**5**
A	Pure data mining	Automatically Send Forms	Splunk dashboard	Azure DevOps boards	Web-based application
B	Azure DevOps boards	Automatically Send Forms	Pure data mining	Splunk dashboard	Web-based application
C	Pure data mining	Automatically Send Forms	Web-based application	Azure DevOps boards	Splunk dashboard
D	Web-based application	Splunk dashboard	Automatically Send Forms	Azure DevOps boards	Pure data mining
E	Splunk dashboard	Azure DevOps boards	Web-based application	Pure data mining	Automatically Send Forms
F	Pure data mining	Web-based application	Azure DevOps boards	Splunk dashboard	Automatically Send Forms
G	Pure data mining	Automatically Send Forms	Web-based application	Azure DevOps boards	Splunk dashboard
H	Pure data mining	Automatically Send Forms	Azure DevOps boards	Web-based application	Splunk dashboard
I	Web-based application	Splunk dashboard	Automatically Send Forms	Pure data mining	Azure DevOps boards

Non AIOps-specific responses

I believe this concept can be implemented in a short time frame.					
Rating → /Participant	**1**	**2**	**3**	**4**	**5**
A	Applying templates	Delinking	Warnings	Generalisation	Improved link
B	Delinking	Applying templates	Warnings	Improved link	Generalisation
C	Generalisation	Warnings	Delinking	Applying templates	Improved link
D	Applying templates	Warnings	Delinking	Generalisation	Improved link
E	Generalisation	Applying templates	Warnings	Delinking	Improved link
F	Applying templates	Warnings	Improved link	Generalisation	Delinking
G	Applying templates	Warnings	Improved link	Delinking	Generalisation
H	Applying templates	Delinking	Warnings	Generalisation	Improved link
I	Applying templates	Warnings	Delinking	Generalisation	Improved link

I believe this concept requires little labour to implement.					
Rating → /Participant	**1**	**2**	**3**	**4**	**5**
A	Applying templates	Warnings	Generalisation	Delinking	Improved link
B	Delinking	Applying templates	Warnings	Improved link	Generalisation
C	Warnings	Generalisation	Applying templates	Delinking	Improved link
D	Applying templates	Generalisation	Warnings	Delinking	Improved link
E	Generalisation	Applying templates	Delinking	Warnings	Improved link
F	Improved link	Generalisation	Warnings	Delinking	Applying templates
G	Applying templates	Warnings	Delinking	Improved link	Generalisation
H	Warnings	Applying templates	Generalisation	Delinking	Improved link
I	Generalisation	Warnings	Delinking	Applying templates	Improved link

I believe this concept requires little expert knowledge to implement.					
Rating → /Participant	**1**	**2**	**3**	**4**	**5**
A	Applying templates	Generalisation	Warnings	Delinking	Improved link
B	Delinking	Warnings	Applying templates	Improved link	Generalisation
C	Warnings	Generalisation	Delinking	Applying templates	Improved link
D	Generalisation	Applying templates	Delinking	Warnings	Improved link
E	Delinking	Applying templates	Generalisation	Warnings	Improved link
F	Improved link	Delinking	Generalisation	Applying templates	Warnings
G	Warnings	Applying templates	Delinking	Improved link	Generalisation
H	Applying templates	Improved link	Delinking	Warnings	Generalisation
I	Applying templates	Warnings	Delinking	Generalisation	Improved link

I believe this concept costs little money to implement.					
Rating → /Participant	**1**	**2**	**3**	**4**	**5**
A	Generalisation	Applying templates	Delinking	Improved link	Warnings
B	Delinking	Applying templates	Warnings	Improved link	Generalisation
C	Warnings	Generalisation	Delinking	Applying templates	Improved link
D	Generalisation	Applying templates	Warnings	Improved link	Delinking
E	Applying templates	Warnings	Generalisation	Delinking	Improved link
F	Applying templates	Warnings	Generalisation	Improved link	Delinking
G	Applying templates	Delinking	Generalisation	Improved link	Warnings
H	Applying templates	Warnings	Delinking	Generalisation	Improved link
I	Applying templates	Warnings	Delinking	Generalisation	Improved link

I believe this concept is effective in improving the data quality.					
Rating → /Participant	**1**	**2**	**3**	**4**	**5**
A	Generalisation	Applying templates	Warnings	Improved link	Delinking
B	Applying templates	Warnings	Delinking	Improved link	Generalisation
C	Warnings	Generalisation	Improved link	Delinking	Applying templates
D	Warnings	Applying templates	Generalisation	Delinking	Improved link
E	Applying templates	Warnings	Generalisation	Delinking	Improved link
F	Applying templates	Improved link	Generalisation	Warnings	Delinking
G	Delinking	Improved link	Applying templates	Generalisation	Warnings
H	Improved link	Warnings	Delinking	Applying templates	Generalisation
I	Applying templates	Warnings	Generalisation	Delinking	Improved link

I believe this concept will require little differentiation between departments.					
Rating → /Participant	**1**	**2**	**3**	**4**	**5**
A	Warnings	Improved link	Applying templates	Delinking	Generalisation
B	Delinking	Warnings	Improved link	Applying templates	Generalisation
C	Delinking	Generalisation	Applying templates	Warnings	Improved link
D	Generalisation	Applying templates	Warnings	Delinking	Improved link
E	Delinking	Generalisation	Applying templates	Warnings	Improved link
F	Applying templates	Improved link	Generalisation	Warnings	Delinking
G	Applying templates	Delinking	Improved link	Warnings	Generalisation
H	Delinking	Generalisation	Applying templates	Improved link	Warnings
I	Warnings	Applying templates	Generalisation	Delinking	Improved link

I believe this concept can handle the required differentiation.					
Rating → /Participant	**1**	**2**	**3**	**4**	**5**
A	Applying templates	Warnings	Generalisation	Delinking	Improved link
B	Applying templates	Generalisation	Warnings	Improved link	Delinking
C	Generalisation	Warnings	Delinking	Applying templates	Improved link
D	Applying templates	Generalisation	Warnings	Delinking	Improved link
E	Generalisation	Applying templates	Delinking	Warnings	Improved link
F	Applying templates	Generalisation	Improved link	Warnings	Delinking
G	Improved link	Delinking	Applying templates	Warnings	Generalisation
H	Generalisation	Delinking	Warnings	Improved link	Applying templates
I	Warnings	Applying templates	Delinking	Generalisation	Improved link

I believe this concept can be properly implemented in Service Manager.					
Rating → /Participant	**1**	**2**	**3**	**4**	**5**
A	Applying templates	Warnings	Improved link	Delinking	Generalisation
B	Delinking	Warnings	Applying templates	Improved link	Generalisation
C	Improved link	Delinking	Warnings	Generalisation	Applying templates
D	Applying templates	Warnings	Delinking	Generalisation	Improved link
E	Applying templates	Generalisation	Delinking	Warnings	Improved link
F	Applying templates	Warnings	Generalisation	Improved link	Delinking
G	Applying templates	Delinking	Improved link	Generalisation	Warnings
H	Applying templates	Improved link	Delinking	Warnings	Generalisation
I	Applying templates	Warnings	Generalisation	Delinking	Improved link

I believe operators will adhere to this concept (i.e. use the concept voluntarily and properly).					
Rating → /Participant	**1**	**2**	**3**	**4**	**5**
A	Applying templates	Warnings	Delinking	Improved link	Generalisation
B	Applying templates	Improved link	Delinking	Generalisation	Warnings
C	Applying templates	Improved link	Delinking	Warnings	Generalisation
D	Warnings	Applying templates	Generalisation	Delinking	Improved link
E	Warnings	Generalisation	Applying templates	Delinking	Improved link
F	Applying templates	Generalisation	Warnings	Improved link	Delinking
G	Improved link	Delinking	Generalisation	Applying templates	Warnings
H	Applying templates	Warnings	Delinking	Improved link	Generalisation
I	Applying templates	Warnings	Delinking	Generalisation	Improved link

I believe the perceived annoyance of this concept will be low.					
Rating → /Participant	**1**	**2**	**3**	**4**	**5**
A	Warnings	Improved link	Applying templates	Generalisation	Delinking
B	Applying templates	Warnings	Improved link	Delinking	Generalisation
C	Applying templates	Delinking	Warnings	Generalisation	Improved link
D	Applying templates	Generalisation	Warnings	Delinking	Improved link
E	Applying templates	Generalisation	Delinking	Improved link	Warnings
F	Applying templates	Warnings	Generalisation	Improved link	Delinking
G	Improved link	Delinking	Applying templates	Generalisation	Warnings
H	Applying templates	Warnings	Delinking	Generalisation	Improved link
I	Applying templates	Delinking	Warnings	Generalisation	Improved link

Overall, I believe this to be the best concept					
Rating → /Participant	**1**	**2**	**3**	**4**	**5**
A	Applying templates	Warnings	Generalisation	Delinking	Improved link
B	Applying templates	Warnings	Delinking	Improved link	Generalisation
C	Warnings	Generalisation	Applying templates	Improved link	Delinking
D	Applying templates	Generalisation	Warnings	Delinking	Improved link
E	Applying templates	Generalisation	Warnings	Delinking	Improved link
F	Applying templates	Generalisation	Improved link	Warnings	Delinking
G	Improved link	Delinking	Applying templates	Generalisation	Warnings
H	Applying templates	Delinking	Improved link	Warnings	Generalisation
I	Applying templates	Warnings	Delinking	Generalisation	Improved link