Data & AI policies united: Picked to pieces

A case study on the components of data & AI policies in the insurance industry

Author: Tom Bruggink

Date: 27/06/2022

Study and specialization: Master Business Administration, Digital Business & Analytics

Supervisors: Robin Effing & Ton Spil

Version number: 1

UNIVERSITY OF TWENTE.

Preface

I am Tom Bruggink, a student currently studying the Master of Business Administration at the University of Twente. I have always shown a great interest in data, data analysis, and organizing. When I was approached with this assignment, I knew it was an instant fit. I have enjoyed working on this thesis and I hope it is of great value to both Alpina and other companies, but also to the academic literature.

During my thesis, I have gotten great help and advice from my supervisors, Robin Effing & Ton Spil at the University of Twente. They have provided me with great insights, and advice and were always positive and excited about my thesis and the topic. For this, I wanted to thank them.

I would also like to thank Jeroen Luigjes since he was my supervisor from Alpina. Jeroen was a great coach that was very enthusiastic, offered great advice and insights, and always showed interest and passion for the topic.

Furthermore, I want to thank the employees at Alpina, and other insurance organizations who helped me with collecting the policy documents, giving interesting insights, and validation. They were also very welcoming and helpful.

Last, but not least, I would like to thank my family and friends. They have always had my back and supported me throughout my entire life.

Abstract

Data & Artificial Intelligence (AI) are both technologies that influence a great many things in this world and organizations are no exception. Organizations need to adapt to these changes in order to stay relevant and comply to laws and regulations. One of these adaptions is in regards to their policies, specifically their digital policies such as the data & AI policies. However, organizations have trouble defining what should be in those policies. In the academic world, studies regarding data and AI policies have been performed. However, these two topics are often separated. The focus is on public policies and not on policies in the private sector, and there is often a lack of a practical element in these studies. This study analyzes the literature on data & AI Policies, it analyzes the policies of an organization, and interviews experts in the data & AI policy field. This is done in the context of the insurance industry.

The research question in this report is: *"To what extent should an insurance organization include data & AI components within their policies?".* This study gave a list of data & AI policy components that should be included. The components are sorted into three dimensions, data policy components, AI policy components, and supplementary policy components. The data policy component dimension has Data security, Data lifecycle, and Roles & Responsibilities. The AI policy components dimension has Explicability and AI models. The Supplementary policy components dimension has General guidelines, Liability, Design & Strategy, Employee, Ethics, and Privacy. With the help of the MoSCoW and a maturity model, the extent to which these components should be included has been determined.

The organization that has been analyzed within this study has received recommendations based on their coverage. These recommendations include that, based on their priorities and maturity level, they should focus on data lifecycle, roles & responsibilities, design & strategy, and ethics first. These components and their respective sub-components should be defined and implemented. After this, the employee component should be focused on. The employee component is about translating and communicating the policy and its contents to employees.

The practical implication of this study is that organizations can use this list as an index of what they should include in their data & AI policies. The extent to which this should be included is influenced by an organization's priorities and maturity level. The academic implication is that the results from this study are based on a combination of both literature and practice, thereby making a connection between the two worlds. These results can be used for future studies that want to dive deeper into topics that could not get explored within this study. One of these topics is a more in-depth analysis of AI components since this study focused on the combination. However, in the end, the study was more favored towards data instead of AI. Future research could also focus on other variables that influence the list of components and their prioritization. The interviews also showed the gap between a policy's content and its compliance. Principles within policies are often perceived as detached from reality and they should be translated to practice. Future research could be conducted regarding this gap and how this gap can be minimized.

Table of Contents

1 Introduction

Digitalization is influencing organizations and they are increasing their technological capabilities (Blackburn et al., 2022). This also influences policies, which need to be changed to stay relevant. This is especially true for the more digital policies such as data policies. Academics such as Alhassan et al. (2019), Abraham et al. (2019), and Cheng et al. (2017) stress the importance of data governance and policies. Alhassan et al. (2019) state that data governance, and in extension data policies, require more attention from both academics and practitioners. It is important to define (van Mil & Henman, 2016), specifically data and data policies (Richardson & Hoffman-Kim, 2010).

Within the practice, certain events also pushed the importance of data policies. A big event that influenced data policies for European organizations is the GDPR. This legislation required organizations to comply with certain guidelines and regulations, but also follow good practices. Organizations needed to adapt their existing practices and implement certain measurements and guidelines in their policies to comply with these new regulations. However, these changes can often be overwhelming and confusing for organizations and their employees (Hrynaszkiewicz et al., 2020). They do not know what to do. A framework or overview could assist organizations in this process and can allow for a smoother transition. The framework or overview could also address other data-related challenges organizations are facing, such as inaccurate and incomplete data (Kim & Cho, 2018).

There has been previous research regarding specific topics within the data policy field, such as data security (Veiga & Eloff, 2007), data storage (van Zeeland & Ringersma, 2017), or even a more general approach to data governance (Abraham et al., 2019). These studies focus on specific parts of data, but almost none focus on the entirety of data policy. There is also a lack of focus within the academic world on data policies within the private sector, specifically focused on companies. Typing “Data policy” into SCOPUS, which is a database with academic articles, will result in articles mostly focusing on either public policies or policies aimed at universities.

This study will try to touch upon all relevant components of private data policies by using the knowledge that has already been gathered by these studies. The added value of this study is its overview of the components within private data policies, but it also tries to fill in gaps that have been mentioned in the limitation of some of the data policy-related studies. One of the limitations that stood out in the data governance article of Abraham et al. (2019) is a missing practical aspect. This practical aspect will be achieved by comparing the findings from this study to that of the policies within an organization. This will help connect the literature to real-life cases, which will be Alpina in this research.

1.1 Company background

Alpina group (further referenced as Alpina) is a financial advisor that is mainly active in the insurance industry. They cover all kinds of insurances, ranging from home insurance to health insurance. They collect and process a lot of data and must take preliminary steps in their policy to secure their data. Alpina is the result of the merger of Heilbron and Voogd & Voogd (Further referenced as V&V). Heilbron acts as a financial service provider and risk manager in several areas, such as insurance, mortgages, absenteeism, pensions, income, capital, and brokerage (Heilbron, 2020). V&V is a link between advisers and insurers and acts as a proxy part for the insurance companies, but also has an IT-based platform where they provide services such as administrative tasks (Voogd & Voogd, 2022).

Together they now form Alpina, which combines the expertise, resources, and possibilities of both companies (Alpina Group, 2021). Alpina is still active in the insurance industry, but also the mortgage industry, pension landscape, and brokers market (Alpina Group, 2021). Alpina chose to merge the existing data platforms into a new data platform from scratch. The problem here is that the policies

that are being used to base the platform on, namely the data policy, are incomplete. Alpina wants to revamp its data policy and is in need of an overview of the components and requirements that data policy should have. This falls in line with what the literature lacks, an overview of what a data policy should have. This is because this information is, within the literature, mostly spread across different articles. Another reason is a general lack of focus on private data policies.

1.2 Goal and scope

The goal of this study is to form an overview of the different components an organization should include within their private data & AI policies. This is carried out using a case study method and the list is based on both the practice and literature. Both will be compared for similarities and differences, which will result in a final list of components that should be present within data & AI policies. This list can be used as an index of what should be included in data & AI policies for insurance companies. However, this research does not claim that by including all these components, a data & AI policy is complete and 100% compliant. The list merely serves as a guide of what is important to include in data & AI policies.

The scope of this research is limited to data & AI policies, specifically private data & AI policies. A list of what topics a data & AI policy should have will be made in the theoretical framework and based on the already existing literature. The research is focused on the insurance industry since policies from other industries could not be analyzed within the given time. Another limitation to the scope is the focus on Data and AI. Data on its own is too broad and too many topics are connected to data, such as social media and cloud computing. AI has been chosen because Alpina is experimenting with AI technologies and would like to know more on this topic regarding policies and guidelines. From an academic perspective, this also offers value since AI is also relatively new to the academic world. The last scope limitation is the focus on the content of a policy, rather than on the development process of a policy.

1.3 Research questions

To contribute to the literature on data & AI policies, a research question was drafted. The answer to this research question can help Alpina and other similar organizations in forming their data & AI policies. The result can serve as an index of what should be included in a data & AI policy. The research question is:

"To what extent should an insurance organization include data & AI components within their policies?"

The underlying sub-questions, that will help answer the research question, are:

- What are the components private data & AI policies should have according to the literature?
- What policies within Alpina are relevant regarding data & AI, and what components do these policies consist of?
- To what extent do the Alpina policies cover the list of components?
- How can Alpina improve its data & AI policies?

1.4 Definitions

The research question brings forward several terms that have to be defined. One of the most important definitions is that of a private data-policy. A policy has been defined by several academics over several years, and these definitions all differ from each other (Hugh Heclo, 1972; Meehan, 1985; Swain, 2011; Yang, 2014). However, there is a common theme, which is that most of these academics talk about rules and regulations. So, for this research, a policy is focused on rules and regulations, focused on a

specific area. While this is still a very broad definition, the theoretical framework dives deeper into what a (data) policy is and what it should contain.

Two sorts of policies exist, these are public policies and private policies (Yang, 2014). Public policies are usually made and used by governments, whereas private policies are made by companies (Yang, 2014). The main difference is who uses it and the target audience of the policy. Private policies will be the focus of this research.

There are a lot of different policies. However, this research will focus on data & AI policies specifically. A data policy is a policy that focuses on rules and regulations revolving around data usage, security, storage, and everything else that involves data (Rainey & Tolk, 2015). The purpose of a data policy is to serve the objectives of the organization that is collecting the data (Parsons, 2013). However, not all these topics will be analyzed in this research. As previously noted, the focus will be on data & AI. So, within this research, a private data policy is a document, made and used by organizations, which focuses on rules and regulations regarding data and AI.

AI is very broad, so it is important to define this. AI within this research is using the definition from Grewal (2014): "Artificial Intelligence is the mechanical simulation system of collecting knowledge and information and processing intelligence of universe: (collating and interpreting) and disseminating it to the eligible in the form of actionable intelligence.". So within the context of this study, AI is the collection of information and knowledge. This information is interpreted and disseminated in the form of actionable intelligence. The collection will be mainly from the databases of Alpina or external databases. Interpretation will be done by employees who will build algorithms to combine certain data and look for possible connections, e.g. for fraudulent activities. The algorithms will then output actionable intelligence.

There are synonyms for these definitions. However, some words closely resemble definitions but are not quite the same. Data policy is, for example, closely related to data management and data governance. All these three words have different meanings, and it is important to know the differences between them. Abraham et al. (2019) state that data governance is about decisions, what they should be, and who should make them. Data management is about making the decisions that data governance puts out. Data policies are then the tools that are used to communicate the decisions of data governance to the people who must conduct the data management activities (Abraham et al., 2019, pp). The focus of this research is still on data policies, but data governance and data management can be used to limit the resulting articles in the literature review.

Components are also important to define. Within this research, components are seen as elements or topics of a policy. The components can be about the security of data for example. Within this research, components are mostly tied to Data & AI policies. The importance of these components can differ for each organization, but this will be expanded upon later in the research. To conclude, within this research, components are topics or elements an organization should include in their data & AI policy.

1.5 Relevance

This research will contribute to the literature on private data & AI policies. Several academics have pointed out the need for such research. Hrynaszkiewicz et al. (2020) argue that people and organizations are getting more confused with the addition of more policy requirements, and Girard (2020) says organizations of all sizes, would benefit from a coherent suite of data governance standards. While this article is mainly about data governance, it also talks about a lack of policy-oriented standards covering data governance.

Abraham et al. (2019), who also conducted a similar study on data governance, mentioned that a practical component in these studies is missing. This practical component can be covered by comparing the overview to already existing policies of Alpina, but also validating results from both analyses by interviewing experts. The results of these comparisons can lead to a final list, that consists of both components from the literature and practice, which are validated by experts.

From a practical standpoint, it will also be valuable to Alpina. The overview will be compared to the already existing coverage of policies within Alpina. If they have any gaps in this coverage, this research will show how to fill in the gaps. This also helps give a proper direction to employees who work on the data platform, since this overview can support their decision-making. Furthermore, it can give companies within the insurance industry, or similar industries, an overview of components that should be included in a data & AI policy.

1.6 Structure of this report

This report will first dive into the already existing literature on private data & AI policies by doing a literature review in chapter 2. This literature review will provide several articles on which the overview can be based. After this, an overview of components that should be included in private data & AI policies are listed. This is based on the articles from the literature review. Chapter 3 will explain the methodology that will be used in this research. This chapter will also discuss the research design, selection criteria, measurement tools, data collection, data analysis, reliability and validity.

Chapter 4 will start with the analysis of the policies of Alpina, followed by a summary of the interviews and their key findings. The interviews will be used as a validation method to ensure the findings from both the literature and practice, are in line with reality. The chapter will then compare the overview from the literature, the overview of the policies, and the input from the interviews. The results of this comparison will then be used to form a final list of components. The chapter will end with an analysis of Alpina's coverage compared to the final list of components.

Chapter 5 will be the concluding chapter in this research, in which the results are used to answer the research question. This chapter will also give recommendations to Alpina regarding their data & AI policies and how to improve these, the appendices will feature in-depth overviews. Chapter 6 will discuss the limitations and possible future research.

2 Theoretical framework

The theoretical framework is developed by conducting a literature review of previous academic literature revolving around private data & AI policies. Within the literature review, articles are selected and then relevant passages within these articles are coded. These codes are explained in-depth in appendix 1. This will result in a final list of components that should be present in data & AI policies, based on the literature.

The literature review is based on the approach that is described by Wolfswinkel et al. (2013). They describe five steps that need to be performed to conduct a rigorous literature review. Within their method, they also include the analysis of the documents (Wolfswinkel et al., 2013). The reason for selecting this method is that it offers a rigorous systematical approach that can offer a complete overview of a specific area within the literature.

The first step is to define the search criteria, which starts by defining the inclusion and exclusion criteria (Wolfswinkel et al., 2013). The only criterium that was used at first was regarding the publication year, in which the literature review only included articles that were published in the last 10 years (between 2012-2022). Following this, research fields were chosen. For this review, the focus was mainly on data & AI policy research and data governance. However, the selected database did not include these kinds of research fields, so unrelated research fields were excluded. These are explained in the in-depth literature review, which can be read in appendix 1. The database that will be used for the literature review is SCOPUS, this was chosen because it is easily accessible to the researcher, since the University of Twente has a license to most of the articles within this database. The database also has a lot of articles and should cover most of the literature. The last action within the first step is to define the search terms which will be used to search for the articles. These search terms are based on previous knowledge of the topic and information from Alpina. Table 1 shows the selected terms, the search also included the plural forms. An in-depth explanation of these terms can be found in appendix 1.

Table 1: Search keywords

Keywords:	**ID:**
Data Platform Policy	DPP
Data Policy Governance	DPGo
Data Governance	DG
Information Governance	IG
Data Management	DM
Data Policy Classification	DPC
Data Policy Merger	DPM
Data Policy Security	DPSe
Data Policy Storage	DPSt
Information Policy Storage	IPS
Data Policy Artifical Intelligence Ethic (1st)	DPAIE1
Data Policy Artifical Intelligence Ethic (2nd)	DPAIE2
Data Governance Artificial Intelligence Ethic	DGAIE

The second step is to search through the database with the selected criteria and search terms (Wolfswinkel et al., 2013). This step will lead to a sample of texts, which will be used for the third step. Within this step, several alterations have been made to the search terms, since previous search terms did not always result in the relevant articles needed for the overview. Again, an in-depth explanation of several decisions and the used search queries and their search terms can be found in appendix 1. The search queries led to a sample of 227 articles.

227 articles were still too many articles, so the sample was refined based on the relevance of their title. This is already part of step 3, selection. After selecting articles based on the relevance of their title, 57 articles remained, of which 6 were duplicates. When duplicates were excluded, 51 articles remained. From these 51 articles, further refinement has been done based on the relevance of their abstract. If the abstract was missing or unclear, a quick scan of the article was done to determine the relevance. This process led to a selection of 16 articles. A backward citation tracking method was used, and this method added 2 articles, which comes to a total of 18 articles. The in-depth explanation and decision-making can be seen in appendix 1.

These 18 articles are the final articles that are used for the analysis step. This step consists of analyzing the content of the articles. The methods that are used are open coding, axial coding, and selective coding. Open coding is reading the articles and labeling relevant parts of the article with a fitting small description, that usually is between 1 and 5 words long (Wolfswinkel et al., 2013). After the open coding, axial coding is performed, in which the researcher is looking for the relations between the categories and sub-categories. These form the main theme and patterns of the articles and their findings (Wolfswinkel et al., 2013). The last part is selective coding, in which the categories are refined (Wolfswinkel et al., 2013). A theory or concept can be thought of. Within this research, this would be a list of components that should be in data & AI policies and to what extent they should be included. An in-depth process of the coding can also be seen in appendix 1.

The last step is presenting the results, which is done in this chapter. The three categories that have been identified within the coding process are data, AI, and general policy components. Table 2 shows the components that fall under these categories. Appendix 1 gives an in-depth explanation of these components and what codes have been extracted from what articles.

Table 2: The dimensions and their components, based on the literature review

Dimension:	***Component:***
Data policy components	Data security Data lifecycle Roles & responsibilities Data quality Metadata
AI-Policy components	Explicability Beneficence Non-maleficence Autonomy Justice Fairness Privacy (AI) General (AI) Technical guidelines
General policy components	General policy guidelines Liability National organizations Employee Ethics Education and training Reinforcement General policy information Design and strategy

2.1 Data policy components

The literature review resulted in a list of several components which are important in different areas. The first one is the components of a data policy. The components are data security, data lifecycle, roles & responsibilities, data quality, and metadata.

Data security is the first component, which might not be surprising since data security has always been a big concern (Siddiqa et al., 2016) and the difficulty of securing data has been growing (Siddiqa et al., 2016). Several components of security are privacy, integrity, confidentiality, and availability (Siddiqa et al., 2016). Privacy is regarding the safety of the user data and preventing a data leakage (Siddiqa et al., 2016), while confidentiality is about protecting this data using techniques such as encryption (Siddiqa et al., 2016). Integrity is about the usefulness and trustworthiness of the data (Siddiqa et al., 2016). The last component, availability, is about plans or procedures ensuring the data is always available (Siddiqa et al., 2016). Siddiqa et al. (2016) gave a lot of insights into data security and relevant aspects regarding it, others (Abraham et al., 2019; Al-Badi et al., 2018; Parsons, 2013) only mentioned it as a component. Other relevant aspects regarding data security are data classification levels and access requirements (Al-Badi et al., 2018; Parsons, 2013), in which data can be labeled based on the security level. Each level should be connected to specific roles and responsibilities, and this adds an extra security level to the data.

Data lifecycle, according to Abraham et al. (2019), is "the approach of defining, collecting, creating, using, maintaining, archiving and deleting data". This component includes a definition of the process/lifecycle of data within the organization, which includes a clear overview of the several steps it takes within the organization. Other components that need to be defined within the data lifecycle are the production of data (Alhassan et al., 2016), the collection of data (Parsons, 2013), type of data (Hrynaszkiewicz et al., 2020), and retention periods of data (Abraham et al., 2019; Al-Badi et al., 2018; Alhassan et al., 2016; van Zeeland & Ringersma, 2017). It was also decided to put data storage under this component since it overlaps with the definition of a data lifecycle. Data storage is about how data is stored and managed, but also about the decisions regarding infrastructure and different applications that are used to store data (Abraham et al., 2019). A data policy that discusses data storage could even mention several preferred techniques to store data in an efficient way, such as clustering, replication, or indexing (Siddiqa et al., 2016).

Another important component of data policies is the roles & responsibility component. According to Parsons, (2013), data policies should have a clear description of the roles & responsibilities. This is mainly focused on defining the roles & responsibilities and then assigning stakeholders (such as employees) to these roles & responsibilities (van Zeeland and Ringersma, 2017). By defining these, stakeholders know what they can and what they are responsible for. When the roles & responsibilities are defined and implemented, they should also be monitored (Alhassan et al., 2016). Data ownership also falls under the roles & responsibilities component and should be defined before roles & responsibilities are defined and assigned (Vilminko-Heikkinen & Pekkola, 2019). Data ownership is about allocating accountability for data (Abraham et al., 2019), this accountability can also be placed upon specific roles such as a data steward. Finally, the roles & responsibilities need to be communicated properly (Abraham et al., 2019). This way, employees are kept up to date on their current roles and what responsibilities. This component can be seen as an extension of data security since it amplifies the data security, but it is considered a separate component, since it is mentioned separately by a lot of authors, and it also differentiates itself compared to data security.

Within the data policy, data quality is also a component. An organization should define the standards for data quality within its policy (Alhassan et al., 2016). Data quality is about the extent to which data can satisfy a user and their requirements (Abraham et al., 2019). Data quality in the context of data

policy needs to include data standards, but also the measurement and monitoring of data quality. Standards should be defined first, then metrics for measuring data quality, and then a process to monitor the data quality (Abraham et al., 2019). Measurement criteria can be timeliness, meaningfulness, and trustfulness. Another addition can also be the management of data quality and potential issues (Abraham et al., 2019).

The last component within the data policy component dimension is metadata. Metadata is, in short, data about data (Abraham et al., 2019). Metadata is used to describe things, mostly other data (Riley, 2017). Examples of metadata are titles or names. Within a policy, an organization needs to specify how metadata is defined, captured, and used, but the retention periods of metadata also need to be specified. This is similar to the data lifecycle, but this applies to metadata instead of data from customers or other parties (Riley, 2017). However, metadata has a low frequency in the literature (Alhassan et al., 2016), which can also be seen in the number of times it is mentioned in the codes.

2.2 AI-policy components

There is a difference between data and AI policies. Data policy focuses on general data, its security, lifecycle, and the roles and responsibilities connected to it. AI policy focuses on AI, and how data within AI and algorithms are used. Based on the literature these are Explicability, Beneficence, Non-maleficence, Autonomy, Justice, Fairness, Privacy (AI), General (AI), and technical guidelines.

Explicability is a component that is quite frequently showing up in the literature review. Explicability consists of two components, which are intelligibility and accountability (Floridi et al., 2018). Intelligibility is about the explainability of AI and how it works (Floridi et al., 2018). The explainability is important. However, when the explainability of AI rises, the accuracy lowers (Aggarwal et al., 2021). Therefore, a good balance should be made, and an organization should decide how much accuracy they are prepared to sacrifice for explainability. Transparency is seen as a causal component of explainability since non-transparency in AI systems is caused by the complexity of the system (Hagendorff, 2020). The literature, therefore, refers to transparency as a solution to achieve explainability (Floridi et al., 2018), since a more transparent process is easier to interpret and explain (Hagendorff, 2020). Accountability is about the responsibility of the AI, however, it is more difficult to make AI accountable, since it is a system. This component is therefore overlapping with roles & responsibilities since the roles & responsibilities make someone accountable (Alhassan et al., 2016). There are several options to solve this problem, the first one is to modify existing roles & responsibilities to include the responsibility of AI (Aggarwal et al., 2021). This is the option that is most in line with this research. Other options can be to implement AI-personhood or common enterprise liability (Aggarwal et al., 2021). An additional benefit of accountability and transparency is that they both achieve the trust of stakeholders (Mullins et al., 2021). Speaking of trust, this is also something that must be defined, and metrics, to measure the trustworthiness of AI, should be developed (Floridi et al., 2018). This can be in the form of a code of conduct, in which ethical AI and its merits can be explained and how to measure certain aspects relating to AI ethics (Floridi et al., 2018). Floridi (2019) has given an overview of 7 principles needed for trustworthy AI: “human agency and oversight”, “robustness and safety”, “privacy and data governance”, “transparency”, “diversity, non-discrimination and fairness”, “social and environmental well-being”, and “accounting”. These have already been included in different components within the AI policy components.

Some components are heavily connected to (technical) ethics. These are beneficence, non-maleficence, autonomy, and justice. Beneficence is about promoting the well-being of people and the planet (Floridi et al., 2018). This means that AI should be there to serve the benefit of people or the planet, and its goals should be similar to those of people. Non-maleficence is about protecting and doing no harm (Floridi et al., 2018). This can be achieved by ensuring privacy and security, but also by

limiting AI in certain areas such as self-improvement (Floridi et al., 2018). The limitation of AI is further explored in the autonomy component, in which an organization should find a balance between the decision-making AI can perform without human intervention (Floridi et al., 2018). The last component is justice, in which AI can be used to correct past mistakes (Floridi et al., 2018). These are all connected to ethics, but there is a difference. These components focus specifically on the ethical principles of AI, and how AI should be designed, but also how AI systems can act ethically. While Floridi et al. (2018) are specifically mentioning and explaining all these principles, other articles such as Morley et al. (2020), Hagendorff (2020), and Aggarwal et al. (2021) mention just a few of these principles.

An additional component that is somewhat connected to the other components is fairness. Fairness is about treating everyone fair (Mullins et al., 2021). This statement might not clarify a whole lot, but it can be concluded that in a fair AI system, all parties should receive equal treatment. The literature mentions several examples of unfair AI practices that are tied to discrimination (Aggarwal et al., 2021; Floridi, 2019; Mullins et al., 2021), these are called algorithm discrimination. This is where the AI can discriminate because of a bias of the developer or a bias in the training data (Aggarwal et al., 2021). To prevent such discrimination and ensure 'fairness', several approaches can be taken in advance. Examples of these approaches are following industry standards regarding AI, defining principles of 'fairness' so that developers know what rules to follow, or recruiting developers with a diverse background (Aggarwal et al., 2021). The last approach increases the odds a developer might recognize a certain discriminating factor in an algorithm sooner (Aggarwal et al., 2021).

Privacy, which has been previously mentioned as a technique to ensure non-maleficence, should also be part of AI-policy components according to the literature. This is because privacy is often overlooked when working with AI (Hagendorff, 2020) since AI systems have difficulty recognizing when privacy is being violated (Hagendorff, 2020). A potential solution for this might be to reduce the 'sight' of AI systems through cryptography of data (Hagendorff, 2020), this way AI cannot access the data.

Some general sub-components were put under the "general AI" component because they did not fit under any of the other components. The first sub-component is that AI policies need to change continuously. AI changes continuously and therefore, a policy relating to AI should also change (Floridi et al., 2018). A one-off AI policy is not enough, there should be further investments into action points regarding AI (Floridi et al., 2018). It should also be noted that AI is different in each organization (Aggarwal et al., 2021) and thus it should be adapted to the organization. If an organization changes, so do the AI and so do the policies. There are also other general requirements an AI policy should comply with such as robustness and safety. An AI policy should be strong and well build, both technically and relating to the social environment (Floridi, 2019). This means that AI policies should include all necessary components of AI and offer guidelines regarding them, but they should also help in improving the social environment of an organization regarding AI usage (Floridi, 2019). Improving this social environment can be done by enhancing sustainable usage of certain systems or enhancing positive social change (Floridi, 2019), such as helping teachers improve their teaching skills. AI should also offer guidelines during uncertain times or when mistakes have been made concerning AI (Floridi, 2019). Lastly, an AI policy should be lawful, meaning it should adhere to and apply law and regulations accordingly (Floridi, 2019), which also includes regulations set by national bodies within the industry.

The last component, which is related to some of the general AI sub-components is the need for technical guidelines. As mentioned previously, AI is different for each organization (Aggarwal et al., 2021) and thus it is difficult to make a generic set of guidelines. Therefore, Hagendorff (2020) puts the focus on a more in-depth explanation of guidelines, but also on defining certain 'obvious' terms. Hagendorff (2020) explains that it is difficult for developers to implement 'justice' or 'transparency'. Hagendorff (2020) also explains that AI in itself is very vague, a policy should not just be about AI, since

AI is just a term that is used for a collective set of technologies. A policy and its (ethical) guidelines should go into greater (technical) detail (Hagendorff, 2020). Policymakers should know the process of data and its details to make better guidelines that are easier to implement (Hagendorff, 2020).

2.3 General policy components

The general policy components are components that can be within almost all policies but support the data & AI policies in a way. The components of a general policy in this research are: general policy guidelines, liability, national organizations, employee, ethics, education & training, reinforcement, general policy information, and design & strategy.

The general guidelines are similar to "general AI", in that this component consists of sub-components that did not fit the other components. These sub-components can fit into the introduction of the policy or could be put with other relevant components. The first component is that a policy should have a definition of its subject (Al-Badi et al., 2018). In the case of data policies, this should be the definition of data, but also what type of data is used (Hrynaszkiewicz et al., 2020) and even what data is not covered under this policy (Hrynaszkiewicz et al., 2020). An example could be that only traditional data is used, and no big data (Abraham et al., 2019). On top of the definition, a policy should also discuss the scope of the policy (Al-Badi et al., 2018). The relevant stakeholders should also be mentioned in a policy, so it gives an overview of what stakeholders are connected to the policy and might even be impacted (Al-Badi et al., 2018). In data policies, this could also include data stewards or external parties such as the data owners. Another sub-component of policies is the compliance with the law and rules/guidelines of the policy and mechanisms that help with this compliance (Parsons, 2013). A policy could mention the actions that will be taken to ensure compliance. These actions/mechanisms could also be based on national bodies and their frameworks (Aggarwal et al., 2021). The last sub-component of general guidelines is that they should communicate the objectives of the policy, this way it is clear what a policy is trying to accomplish (Abraham et al., 2019).

A policy should also discuss potential risks and liability, especially regarding AI (Aggarwal et al., 2021). It is important to discuss the potential problems that could occur, in both AI-related policies and data-related policies. There should also be a description of what will be done when a problem arises (Al-Badi et al., 2018), and it should refer to the roles & responsibilities section regarding who is responsible for the problem. The liability determination and risk assessment are also connected to data security. This component is discussed in this section because it also applies to other policies, such as AI.

Several components, such as privacy, ethics, and infractions, differ on a local level (Delacroix & Wagner, 2021; Stahl & Wright, 2018). These components must be partly formed by local organizations or national bodies (Delacroix & Wagner, 2021). It could also be that local organizations work together to form certain standards regarding privacy, ethics, and infractions (Delacroix & Wagner, 2021). The impact of this on the policy components is that there could be a section discussing the related national bodies or agreements between organizations.

Another optional component within a policy is regarding the employee and a description of potential scenarios that impact the employees, such as potential workplace disruption (Aggarwal et al., 2021). Policies should help in guiding the transition when work is displaced or impacted by new technologies (Aggarwal et al., 2021). This is important when policies implement relatively big or new actions or guidelines, such as AI, which has a big impact on employees (Hagendorff, 2020). This component also includes communication with employees regarding policies or topics such as roles & responsibilities.

Ethics is a component, which not only applies to data policies but also to AI policies and other policies. Parsons, (2013) and Abraham et al. (2019) mention that there are ethical concerns regarding data

usage, but also AI usage. They also mention that ethical standards should be recorded in the policies since it is often unclear what ethical practices are. However, ethics should not only be communicated to employees in the form of guidelines since ethical guidelines alone do not have an impact on human decision-making or their behavior (Hagendorff, 2020).

Several techniques must be implemented to ensure ethical behavior, decision making, and, in the end, an ethical organization. One of these techniques, which has proven to improve the culture and be effective (Abraham et al., 2019), is education and training revolving around ethics, security, and the impact of AI. It is also important to educate employees, especially engineers who work on AI and data applications, on privacy and ethical related practices (Siau & Wang, 2020). Training is therefore seen as an important component of a policy and could solve several other problems, such as changing the culture and behaviors of employees.

Another technique regarding ethics, which is also important, is reinforcement. According to Hagendorff (2020), ethics, in general, lack a reinforcement mechanism to reinforce ethical claims. This means that certain consequences should be connected to infractions (Delacroix & Wagner, 2021) and a procedure for this scenario should be implemented (Floridi et al., 2018). This reinforcement could be based on organizational standards, but also standards of national bodies within the industry. Auditing is also important since it helps with ensuring that appropriate mechanisms are in place and are effective (Floridi et al., 2018). This is especially important regarding AI since mechanisms for AI are relatively new compared to others within the data world (Floridi et al., 2018). Audits should also look at the application of ethical principles and if they are applied correctly (Morley et al., 2019).

General information is also a component within the general policy components. This component consists of relevant tips and useful information that could improve a policy. These cannot usually be translated into guidelines within the policy but could offer an improved policy when this information is accounted for. One of the first sub-components is the context-dependency of a policy and that the policy differs for each industry and might even differ between organizations (Abraham et al., 2019; Parsons, 2013). In contrast to what has been mentioned earlier, guidelines do still need to maintain some form of generalizability to be helpful for an organization (van Zeeland & Ringersma, 2017). The challenge is to find a balance between being too general and being too specific. An option would be to use best practices as a method (van Zeeland & Ringersma, 2017) of showing the applicability of the guidelines, while still maintaining a general image. The policy could also, as an addition to the general guidelines, have a section in which the applicability and implementation of the guidelines are discussed. Using best practices as a method can also increase the credibility of the guidelines (van Zeeland & Ringersma, 2017), which adds to the 'trust' factor that has been discussed in the AI-policy components section. Another sub-component is that a policy should not limit processes and employees too much with restrictions (Abraham et al., 2019). A policy and its guidelines should empower its employees in making decisions and not limit them too much (Hagendorff, 2020). A final sub-component is regarding the integration of a policy into the behavior and culture of an organization (van Zeeland & Ringersma, 2017). The best policies are those that are integrated into the culture and behavior of the organization (van Zeeland & Ringersma, 2017); however, this is often perceived as difficult.

The difference in policies also comes from the fact that policies are in line with the strategies of an organization (Abraham et al., 2019), which policymakers can use as an instrument to form their guidelines (Klievink et al., 2017). This in turn guides the process of adapting the data architecture (Abraham et al., 2019). However, it is best practice to implement important principles from the strategy already in the design, instead of adapting the architecture at a later stage (Morley et al., 2020).

2.4 Analysis of results

Table 2 shows in total 23 components. These are quite a lot of components to analyze and not every component might be of equal importance. Therefore, redistribution and prioritization will be done. The MoSCoW method is used for this prioritization. The MoSCoW method is a method of prioritization, in which several requirements are prioritized (Hudaib et al., 2018). It consists of four levels, Must-have, Should-have, Could-have, and Won't-have (Hudaib et al., 2018). Must-have components are key to the policy and without them, a policy would fail. Should-have components are still a high priority, but not as critical as the Must-haves. Could-have components are a nice addition to the requirements, but not necessary. Won't-have components will not be used right now but could be added in the future (Hudaib et al., 2018).

2.4.1 Redistribution and prioritization components

The first three data policy components (data security, lifecycle, and roles & responsibilities) all have a high frequency of being mentioned in the several articles, which can be seen in table 3. The articles within the literature review also identify these components as a high priority. This is why these three components are marked as Must-haves. Without these components, a data policy would not work. The frequency of data quality and metadata are notably lower and Alhassan et al., (2016) also noticed the difference in frequency (regarding metadata) within the literature. The components are also not very strong on their own and are therefore added to the data lifecycle since they both fit better into this component compared to the other components (data security and roles & responsibilities).

The AI-policy components have several overlapping components which can be grouped. Beneficence, non-maleficence, autonomy, justice, and fairness can all be grouped under the 'AI Principles'. This is because these components are often mentioned together and are all considered vital and thus marked as Must-have. Explicability will remain a component on its own since this component is mentioned considerably more than the 'AI-principle' components. Explicability is also marked as a Must-have since, without this, an AI policy would work. Privacy will also remain a component on its own since it differs from the 'AI-principle' components in that it is a more significant aspect when compared to the principles. Privacy is also mentioned separately because of the GDPR and its requirements surrounding privacy. Privacy is not as important within the AI policy as the first two components according to the articles, but it is still an important component. Therefore, it is marked as Should-have. The technical guidelines component is only mentioned in one article, but the importance of this component is relatively high. Hagendorff (2020) mentions the importance is high because it is often forgotten within literature and practice. This might also be the reason for its low frequency within the literature review. However, this component is not as crucial to an AI policy since a policy would still function without it. The policy would just not function effectively, therefore the component is marked as Should-have. General AI is a mix of several not significant sub-components that were mostly mentioned once and do not influence the policy in a significant way but might be beneficial to add in the future. Therefore, the component is marked as Won't-have.

Some of the general policy components overlap. This is the case for the employee, education & training, and reinforcement components. They are connected and are grouped under the employee component since they are either focused on communication and education of employees or making sure they follow the policy guidelines through reinforcement. These three components are all considered to be important, since they have a reasonable frequency, and they are also mentioned as relatively important by the authors that mention these components. However, the components have a supportive nature in which they support the other components, so they are marked as Should-have. Without these components, a policy can be functional, but it is strongly recommended they are included to improve the policy. The general policy guidelines have several sub-components, but the

priority of these components differs. Their frequency is also on the low side, as can be seen in table 3. However, they do have recurring sub-components in contrast to the General AI component. Therefore, this component is considered a Could-have since it could provide benefits to a policy. Ethics is considered to be an important component by the authors, and it also has a relatively high frequency. The National organizations component is removed since it is just an influence on the ethics regulations and guidelines. It could be argued that it should be added to ethics, but from the researcher's perspective, this would be out of place and not fit the current ethics component. The ethics component is marked as a Must-have, since the frequency is high, and the authors also indicate the importance of including this topic as high. The liability and design & strategy components have a low frequency and when they are mentioned, the articles do not specifically mention them as important components. Therefore, they are both marked as a Won't-have. The last component, General policy information is marked as Won't-have. This is because this component is not focused on guidelines, but just general advice regarding a policy. This is just the prioritization based on the literature, but the interviews and policy analysis might change this prioritization. Table 3 gives an overview of the reference frequency of each component and its source. Table 4 gives an overview of the prioritization level of the final components based on the literature review.

Table 3: Source of reference of each component

	Article: / Components:	Data security	Data lifecycle	Roles & Resp.	AI-principles	Explicability	Privacy (AI)	Tech guidelines	General AI	Employee	General guidelines	Ethics	Liability	Design & Strategy	General info.
1	Parsons (2013)	x	x	x						x	x				x
2	Alhassan et al. (2016)		x	x											
3	Siddiqa et al. (2016)	x	x												
4	Klievink et al. (2017)													x	
5	van Zeeland and Ringersma (2017)		x	x											x
6	Al-Badi et al. (2018)	x	x	x							x		x		
7	Stahl and Wright (2018)				x	x	x			x		x			
8	Floridi et al. (2018)				x	x			x	x					
9	Abraham et al. (2019)	x	x	x		x				x		x		x	x
10	Vilminko-Heikkinen and Pekkola			x											
11	Floridi (2019)				x	x	x		x			x			
12	Hrynaszkiewicz et al. (2020)										x				
13	Siau and Wang (2020)				x	x	x			x		x			
14	Morley et al. (2020)				x	x						x		x	
15	Hagendorff (2020)				x	x	x	x		x					
16	Delacroix and Wagner (2021)					x				x		x			
17	Aggarwal et al. (2021)		x	x	x	x	x		x	x			x		
18	Mullins et al. (2021)				x	x									

Table 4: Components and their prioritization according to the MoSCoW model

Data components:		**AI-Policy components:**		**General components:**	
Data security	**M**	AI-Principles	**M**	Employee	**S**
Data lifecycle	**M**	Explicability	**M**	General guidelines	**C**
Roles & responsibilities	**M**	Privacy	**S**	Ethics	**M**
		Technical guidelines	**S**	Liability	**W**
		General AI components	**W**	Design & strategy	**W**
				General policy information	**W**

2.4.2 Similarities and differences between components

Some of these components are still connected, albeit in small areas. The roles & responsibilities component is connected to explicability since both try to create accountability. The difference between them is that the roles & responsibilities component is mainly connected to data and making insightful what roles everybody has and what responsibilities are connected to those roles. Explicability is more about helping people understand AI, making the process/algorithm transparent, and making sure the right party is accountable for certain actions that the AI might cause.

Another connection between components is that of AI principles and ethics. The AI principles are closely tied to desired ethical behavior of algorithms. This desired behavior is also key within the ethics component. The difference here is that AI principles are about what principles an ethical algorithm or AI should have, whereas ethics is about what behavior is expected from employees. Ethics is also about creating a desired culture and behavior. AI principles are mainly focused on algorithm developers, while general ethics apply to every employee.

Privacy is also a component that has some connections with other components. Privacy and data security are close to each other since privacy is a sub-component of data security. However, the privacy component is focused on privacy within AI, whereas privacy the sub-component of data security, is focused on the privacy of all data used within the company. The context is what differentiates these two and the context will thus determine if a policy document discusses data security or privacy within AI. Another difference is that privacy within AI is also about training the AI to recognize privacy. Privacy also has a connection to the AI-principles and ethics. The difference between these is that privacy is heavily connected to the law and governmental regulations such as the GDPR, while AI principles and ethics are more connected to regulations from national organizations they affiliate with. This is not always the case since some ethical standards are mandatory by law. These connections make privacy a component that might belong to the general policy components since it belongs to both AI and data policies. For now, the literature defines a difference and privacy will both be a component in AI policies and a sub-component of data security. However, the interviewees or policy analysis might change this perspective.

These are the major connections and similarities between the components. Appendix 1 has an in-depth overview of the components and their sub-components. These will be compared to the results from the policy analysis.

3 Research methodology

This chapter will discuss the methodology that is used to solve the different sub-question and ultimately the research question of this report. This is an inductive study, where information from the literature, experts, and policies, is used together to form an overview of components that should be included in a data & AI policy. This chapter starts with the research design and how the information, that is necessary to answer each question, will be collected. This also includes details regarding the interview and implementation of the analysis method. After the research design, the selection process will be described. Following this will be a description of the measurement instruments, data collection, and data analysis. It ends with a description of the reliability and validity of this research and how it will be safeguarded.

3.1 Research design

A case study is chosen as the preferred approach. The case study allows for an in-depth approach, that focuses on a 'case' (Yin, 2018). This is important since a detailed approach is needed to fully understand to what extent certain components should be included within data & AI policies. The case study approach also allows for a real-world perspective (Yin, 2018), which is necessary to give relevant recommendations to Alpina. This also covers the limitation that previous studies had, which is the lack of a practical component and therefore a not-so-strong connection with reality. Yin (2018) also proposes three questions that, when answered, state what research design is best suited. These three questions are: "What is the form of the research questions?", "Does it require control over behavioral events?", and "Does it focus on contemporary events?". When these three questions are answered, a case study is recommended.

The research design that is used for the case study, is a single-case design with the rationale of a critical case (Yin, 2018). This is where the case is used to test whether something is true or not (Yin, 2018). The case study is single since it focuses on a single case, which is Alpina. However, the case method features several interviews which will be used as a validation method for the results from both the literature review and policy analysis.

3.1.1 Case study protocol

Yin (2018) recommends using a case study protocol to keep the focus on the topic of the case study. It also helps in preparing for potential problems and increases the reliability of the report. The protocol has four sections: an overview of the case study, data collection procedures, protocol questions, and a tentative outline for the case study report.

The overview of the case study is about the background information regarding the case study. The goal of the case study is to gain a better understanding of what components should be in data & AI policies and to what extent. The research questions, which can be seen in table 5, help in achieving this goal. The overview should also contain the key readings. Key readings regarding the literature review can be seen in appendix 1. Table 34 in appendix 4 contains the key readings regarding the policy analysis.

Table 5: Research questions, where and how they are answered

Q:	Sub-question:	Where:	How:
Q1	What are the components private data & AI policies should have according to the literature?	Chapter 2	Literature review (desk research)
Q2	What policies within Alpina are relevant regarding data & AI, and what components do these policies consist of?	Chapter 4	Interviews and policy analysis (desk research)
Q3	To what extent do the Alpina policies cover the list of components?	Chapter 4	Interviews and desk research
Q4	How can Alpina improve its data & AI policies?	Chapter 5	Recommendations

The next section is about the data collection procedures. Regarding the literature review, the articles are the data that is collected. The policy analysis uses the policy documents, which have been obtained through an Alpina employee. The procedure for both collection methods has been explained in appendices 1 and 4. The collection of data from the interviews will be done by recording the interviews and transcribing them. The interviews will be done using Teams. Further details regarding the interviews and used methods will be described later in this chapter. Appendix 3 will contain an overview of the used framework during the interviews.

The third section is regarding protocol questions. These are questions that are directed to the researcher and should help keep the focus of the research clear while data is collected and analyzed (Yin, 2018). The protocol should contain a list of these questions, together with likely sources of answers to these questions. The literature review provided a list of components, which can be used to draft the protocol questions. However, since a list of components is the desired outcome of the policy analysis, the questions will mostly be the same with the only differences being the components. The protocol questions and their likely source can be found in appendix 2.

The last section is a tentative outline for the case study report. Within this step, it is important to form the report outline based on the target audience of this case study (Yin, 2018). The target audience of this case study is the University of Twente and Alpina. The outline of the report will be based on the structure provided by documentation from the University of Twente. This structure is described in chapter 1.6. Another aspect of this section is the documentation regarding this research. This will be added within appendix 4, similarly to what has been done for the literature review.

3.1.2 Evidence

A case study can have six sources of evidence: documentation, archival records, interviews, direct observations, participant observations, and physical artifacts (Yin, 2018). Within this study, documentation and interviews will be used as sources.

3.1.2.1 Documentation

Documentation, otherwise known as desk research, within this research is used within the literature review. The documentation uses secondary published data (Zhou & Nunes, 2016). Secondary data means that it already exists and has been collected by someone else (Zhou & Nunes, 2016). In this research, these would be the academic articles, but also other relevant documents from Alpina.

Using documentation has several advantages and disadvantages. An advantage is that documentation is stable and unobtrusive. These components add to the reliability of the research since repetition of the documentation can be done relatively easily. A possible disadvantage of documentation is that documents can be hard to find (Yin, 2018). This was the case for some articles within the literature review since the researcher did not have access to all articles. This is a limitation of this research and should thus be considered when interpreting the results. Another disadvantage of documentation is the potential bias in the selection process. To prevent this, thorough documentation of the decisions and steps will be included in appendix 4.

3.1.2.2 Interviews

This research will, besides desk research, also use interviews, since Yin (2018) specifies that using multiple sources of evidence will increase the validity of the case study. Yin (2018) defines three case study interviews: prolonged interviews, shorter interviews, and survey interviews. Babbie (2012) mentions a similar explanation of interviews but looks specifically at the structure of the interview. Babbie (2012) mentions that there are also three kinds of interviews: unstructured, semi-structured, and structured.

For this research, the shorter interviews, otherwise known as the semi-structured approach, will be used. The reason for this is that the interviews need to have a pre-determined list of topics and questions that are related to the components of a data & AI policy. However, there should also be the flexibility to change topics, for example when a topic is not on the list. Another reason for choosing this method is that the interviews will be used as a validation mechanism for the results, which fits the description of a shorter interview. An added benefit of the shorter interview method is that these kinds of interviews provide specific insights and can be directed to a selection of desired topics (Yin, 2018). Disadvantages such as response bias or bias due to poorly articulated questions (Yin, 2018) will be solved by letting the questions be reviewed and creating a comfortable environment for the interviewees.

Within this research, there will be two internal interviews to act as validators. These will be employees who work within Alpina and have knowledge of data & AI policies. Both interviews will be recorded (with the consent of the interviewee) and transcribed. There will also be external interviews with experts in the field of insurance and data & AI policies. The format of the interview and the questions are the same for each interview and can be seen in appendix 3.

3.1.3 Analysis

Yin (2018) also discusses different ways to begin analyzing the data from the case study. However, the literature review already made use of the approach from Wolfswinkel et al. (2013). Because the literature review and policy analysis need to be compared, the same method will be used to ensure a similar format of the results. This similar format is beneficial in analyzing and comparing the results with each other. Other advantages of the method are that it is structural and provides deep insights into how the overview is formed. This will add to the reliability of the research. To add to the method of Wolfswinkel et al (2013), some aspects of Yin (2018) will also be used. The first is basing the concepts (dimensions) on data and changing these based on the data (Yin, 2018). The second is using a theoretical basis for the analysis, which will be the results from the literature review. There are also analytical techniques that Yin (2018) discusses. The one that applies the most is that of pattern matching, where the findings from one case study are compared to the results from a predicted one. Again, these are the results from the literature review.

3.1.4 Ethics

Regarding ethics, Yin (2018) states that researchers must report their findings as honestly, integer, and objectively as possible, to preserve a high ethical standard. Within this study, the interviewees will be anonymized and only their pronouns and whether they are internal or external interviewees will be used. The interviewees are also asked to give consent in advance for recording, transcribing, and using the results from the interviews. They always have the right to withdraw this consent. To ensure the interviewees agree with what is being used for this research, they are asked to greenlight the transcript, which will be sent to the interviewee. If the interviewees confirm that this is what has been said, the transcript can be used for the research.

3.2 Selection

The selection of relevant articles has been done in the literature review, which also explains the decisions behind the selection. Internal interviewees will be selected based on communication with an Alpina employee. The external interviewees will be people who are experienced within the insurance industry and with data & AI policies. Selection for these will be done based on references from the interviewees or connections through the University of Twente. The selection of what policies from Alpina will be used is based on what has been sent to the researcher through internal e-mail.

3.3 Measurement

As previously mentioned, to measure whether the results from this research are close to reality, several interviews will be held. This will serve as a validation tool to validate whether the results from the literature review and policy analysis are accurate. Based on the interviews, the results will be deemed representable or not. Representable means that the result can be applied to other organizations within the insurance industry. It could even be assumed the results apply to similar organizations and industries, such as banks or consultancy bureaus. Not representable will mean that the results are not in line with reality. If this is the case, irregularities will be identified throughout the interview and revised.

3.4 Data analysis

This research features two analyses. One has already been done, which is the literature review. This process has been explained in-depth in chapter 2 and appendix 1. The other analysis will be done in chapter 4, in which the policies will be analyzed. The plan for this analysis has already been described in chapter 3.1. Both analyses will lead to a final overview of the components that should be in data & AI policies and to what extent these should be included.

3.5 Reliability, validity, and other limitations

The reliability of research is whether the result will be similar if the same technique is applied again (Babbie, 2012). To ensure reliability, a technique has to be able to be conducted by others and decisions should not involve subjectivity, if they do, these decisions should be described in detail (Babbie, 2012). Reliability within this study is increased by documenting every step made within the literature review. If someone were to conduct the same steps, they (with exceptions aside) should come to a similar result. The reliability of the policy analysis will be increased similarly. The different steps that will be taken in this analysis will be described in-depth in appendix 4, so the train of thought is visible to others. The case study protocol also helps in increasing the reliability, since it gives a clear overview of the different steps that have been taken within this research (Yin, 2018).

Reliability ensures similar results when the technique is applied again, but it does not ensure the result is accurate. To ensure accuracy, validity is used. Validity is the extent to which a measure adequately reflects reality (Babbie, 2012). To ensure that the results from this study reflect reality, internal and external interviews will be held. These interviews will validate the results from both the literature and practice. It should be noted that case studies are often critiqued on their generalizability and this is a valid concern. However, the goal of case studies, in general, is not simply to be generalized (Yin, 2018). The goal of this case study is to provide a deep insight into components of a data & AI policy based on the literature and a case. The findings from this case study offer knowledge and learning points for other researchers or organizations. There will be no claim that these findings apply to other organizations, but they might find some overlap that they can use in their organization.

Besides validity and reliability, there are also other limitations and concerns within this research regarding the methodology. One of these is that the case study method is seen as not rigorous enough (Yin, 2018). This originates from researchers who performed a case study but did not follow systematic procedures or did not document their decisions enough (Yin, 2018). To limit the influence of this limitation, intensive documentation for the analyses has been added to the appendices.

4 Results

This chapter will feature the results from the policy analysis, the results from the interviews, and the comparison between the literature, policy analysis, and interview results. This chapter also gives an overview of the final components and their respective prioritization. The chapter ends with an analysis of Alpina's coverage of these components.

4.1 Policy analysis

The documents that were provided by the internal employee for the policy analysis can be divided into two categories. These two categories are main policies and supporting documents. Main policies are documents that are entire policies while supporting documents are documents that are not policies but act as a supporting mechanism for the policies. The main policies are the ones being analyzed and these can be seen in table 34 within appendix 4.

As has been mentioned in chapter 3, the policy analysis has used a similar approach to the literature review. Each policy has been read and relevant sections were coded. These codes are then put into an overview and grouped based on their similarities. Table 36 within appendix 4 has an in-depth overview of the codes, where they originate from, and under what component they fall. Table 37 shows the sub-components, their components, and their dimension. Table 6 shows the results of the policy analysis.

Table 6: Policy components based on policy analysis

Data Policy components	**Data security** **Data lifecycle** **Roles & Responsibilities**
AI-Policy components	**Explicability** **AI-Principles** **AI-models**
General policy components	**General guidelines** **Risk assessment** **Design and strategy** **Others** **Employee** **Ethics** **Privacy**

4.1.1 Data policy components

Data security within the policy documents was discussed in different contexts and policy documents, outside of the expected security policies. These documents were the privacy and ethical policies. When security was mentioned, it was about specific measures to ensure security. These could be physical and digital measures. Security had, similar to the literature, several sub-components and methods relating to these sub-components. The sub-components are privacy, integrity, confidentiality, and availability. The definitions of these sub-components are the same as the literature. Privacy is within this component, more about measures and rules regarding privacy. The policy also discussed classification and methods to classify.

Data lifecycle was also discussed in all kinds of different policies, such as the information security policy, privacy policy, retention policy, and even the ethical framework. However, the term data lifecycle was not used, this was mostly called the data process. While the terms are different, they both have the same goal: To describe the process of data through the entire lifecycle within the organization. This process begins from the first time it is collected till the deletion of the data. Sub-

components were mostly focused on data retention, deletion, and archiving, but the policies also mention the data process, its description, importance of quality, and data storage.

Data roles & responsibilities were also discussed in more policies than anticipated but in a different context. The policies they appeared in ranged from the ethical framework, to access control, to information security. Roles & responsibilities were mostly discussed in the first chapter of the policy and are regarding the roles & responsibilities within a policy. They state the stakeholders that are affected by the policy, what their role is, and their connected responsibility related to this policy. This way, they connect employees with specific responsibilities within a policy. However, a specific mechanism to enforce the responsibilities was not mentioned. This component overlaps with the sub-component stakeholders' roles & responsibilities.

4.1.2 AI policy components

Most AI components were found within the AI policy document. The AI document touched on explicability, which consists of explainability, accountability, and transparency. This is very similar to the literature definition.

This is also the case for the AI principles, which consist of beneficence, non-maleficence, autonomy, justice, and fairness. Their definition within the policy documents lines up with the literature-based definition of these sub-components.

The AI models component is, compared to the literature, a new component. A sub-component of this component is AI-decision making, which are measures and information regarding decision making in AI models. A second sub-component is the data quality of the input of AI models. The importance is placed upon high quality, to ensure reliable results from AI models. The last sub-component within AI models is the monitoring of these models and the frequency of monitoring them.

4.1.3 General policy components

The general policy components are, as anticipated, found in most policies. However, this is not always the case for each component. Some components are very general and support data & AI policies but are not found in most policies.

The first component, general guidelines, consists of several sub-components. The goal & scope is a component that can be found in most policies and is about defining the goal of the policy and its scope. Another component with a high frequency is the definition sub-component, which discusses the definitions of several terms used within the policy. The stakeholders and their roles & responsibilities are also mentioned in most policies, usually at the start of a policy. They describe the stakeholders related to the policy, their roles, and what responsibilities they have in relation to this policy. Within the general guidelines, compliance & audits are also a sub-component. This sub-component mentions several mechanisms to ensure compliance with the policy. The maintenance of a policy was also mentioned frequently, which discussed the evaluation and upkeeping of the document and when it will be done. The last two sub-components are documentation and procedures. Documentation within the policy is about making sure several decisions, incidents, and plans are documented for future use. The procedures component is about several procedures in case something went wrong, or something changes, usually focused on emergency cases or scenarios.

Risk assessment is also a component that was in several policies. This component is mainly about assessing the risks of several tools and threats concerning that policy. It also includes the time interval between each assessment and thus when a risk assessment should be done.

Design & strategy is about the strategy, architecture, and design of systems within Alpina. This component can mostly be found in policies that focus on the development of applications, systems, or networks. It usually emphasizes that a policy should be in line with strategy, but also that systems and applications should be designed to encourage policy-correct behavior.

The employee component consists of three sub-components, that were mentioned in several policies. Communication, education & training, and reinforcement. Communication is about the organization or employees themselves communicating with employees, but also how and what employees should communicate with customers. Education & training is about educating & training employees and stakeholders regarding laws and regulations but also increasing their awareness. The education & training can also be in the area of educating employees on new technologies or systems and how these should be used. Reinforcement is not mentioned in a high frequency, but it is about measurements used to enforce behavior.

Ethics is also mentioned within the policies. As expected, most mentions originated from the ethical framework. The component is mainly about ethical procedures and methods to ensure ethical behavior.

Privacy is, as has been previously mentioned, something that can be placed under different components. It is part of the data security component, which is mainly about measures to ensure data privacy. However, privacy can also be about communicating to the customers what their rights are regarding privacy. This component also mentions several laws and regulations in which this communication is mandatory. The policy documents mention privacy communication in both the AI context and the data usage context. Because it is present in both contexts, it is placed under general policy components.

4.2 Interview results

To validate the findings from both the literature and practice, four interviews were held in April and May 2022. These interviews should assist in confirming the results from the literature and practice. The interview framework can be seen in appendix 3. In short, the interviews are used to gain an understanding of the perspective of the interviewees, but also to gain insights into what they think a data & AI policy should include. Their perspective on the prioritization of the different components will also be asked. This input will be written down, then the results from the analyses will be shown and their input is asked again. To effectively show the results, a preliminary overview of the components and their differences was made. A general overview can be seen in table 7. The in-depth overview can be seen in appendix 4.

Table 7: General overview of components from literature and practice

	Literature	Practice
Data policy components	Data security	Data security
	Data lifecycle	Data lifecycle
	Roles and responsibilities	Roles and responsibilities
AI-policy components	AI-Principles	AI-principles
	Explicability	Explicability
	Privacy	-
	Technical guidelines	-
	General AI	-
	-	AI-models
General policy components	General guidelines	General guidelines
	Liability	Risk assessment
	Design & Strategy	Design & Strategy
	General Policy information	Others
	Employee	Employees
	Ethics	Ethics
	-	Privacy

Mostly the same
Very similar, but difference in sub-components
Big difference in sub-components
Missing in other analysis

Within this chapter, the most important insights from the interviews will be discussed. This will be done by looking at each component and what the interviewees mentioned regarding this. A more in-depth look at what each interviewee mentioned can be seen in appendices 6, 8, 10, and 13.

4.2.1 Data policy components

Within the data policy components dimension, data security, data lifecycle, and roles & responsibilities are the main components identified from both the literature and practice. The interviewees verified most of these components, while some interviewees stated possible changes to the overview of components. Data security and its sub-components are verified by three interviewees and no changes in regards to the overview were mentioned.

Data lifecycle does differ based on the perspective of the interviewees. Interviewee 2 mentions data flows and defines these as "How does data come in, how does it get processed and how does it go out?" (Appendix 7). This is similar to the definition from the literature, which states that a data lifecycle is "the approach of defining, collecting, creating, using, maintaining, archiving and deleting data" (Abraham et al., 2019). Interviewee 2 states that these data flows need to be insightful, in order to secure them, but also to be compliant with laws and regulations. Identifying these is the first step in this compliance process. This shows, that the different data flows and identifying them, are important, since it influences other components. Interviewee 1 mentions that "when maintaining the platform, reproducibility is important, that we can see what codes are used, and can reproduce situations that resulted in specific results" (Appendix 5). So in addition, a data lifecycle should also include documentation, which will make sure situations are reproducible. The documentation of data is also important since this uncovers the lineage of that data (Appendix 5). This will assist in deciding what data can be used for specific purposes, without breaking the law (Appendix 5).

The roles & responsibilities component should also be changed according to interviewees 2 and 3. Interviewee 2 (Appendix 7) stated that data ownership should not be the main focus, but process ownership. She stated that: "It is required by the GDPR that several assessments are conducted, but to do this, an owner of the process is required." (Appendix 7). This is why interviewee 2 (Appendix 7) thinks an organization should focus on process ownership since, without it, compliance with the GDPR will become more difficult. Interviewee 3 stated: "You only have the roles and responsibilities, but it goes further than that. How will those roles and responsibilities work together to ensure the process will be safeguarded." (Appendix 9). According to interviewee 3, the audit and compliance of the roles and responsibilities should also be included in the roles & responsibilities component.

A final addition that influenced all the components in the data policy components dimension is a model that was mentioned by interviewee 3 (Appendix 9). This is the DAMA-DMBOK framework, which gives an overview of " Aspects that should be present (within a data policy) to make it a success" (Appendix 9). This framework is used by organizations, where they directly reference that a policy should be compliant with the DAMA-DMBOK framework (Appendix 9). This way, they are sure that they have covered most of the data policy components. Interviewee 3 also stated that the DAMA-DMBOK framework could be used to check whether the results from the analyses are in line with this framework. Appendix 11 shows the sub-components which have been identified from the framework. For this dimension, data security and roles & responsibilities had similar components to that of the DAMA-DMBOK framework. For the data lifecycle, some components were similar, but some components were also added. A special case is data quality, which consists of several sub-criterions according to interviewee 3 (Appendix 9) and the DAMA-DMBOK framework (International, DAMA, 2017). A more in-depth look at data quality and the other recommended changes from the DAMA-DMBOK framework can be seen in appendix 11.

4.2.2 AI policy components

The components with the AI policy components dimension based on the literature and practice are explicability, AI principles, general AI, technical guidelines, privacy, and AI models. Since interviewees 2, 3, and 4 all stated that they were not experts on AI, they gave no input regarding this dimension and the components within it. The only comment interviewee 4 gave is that they have: "A policy document on how to deal with AI-like applications, formulated from the policy. This contains several principles that we apply." (Appendix 12). This adds to the verification that AI principles are used in the organization of interviewee 4 and that they should be present.

Interviewee 1 gave several insights regarding AI. His first comment was that "Despite the technique, you need to be able to explain the algorithm and how it came to a result." (Appendix 5). He wants to prevent a black box scenario and earn the trust of the customer by granting transparency in their decisions. He thinks that "Without the explainability, certain things, such as discrimination, can be difficult to spot" (Appendix 5). Therefore, he also suggested including ethics within an AI policy since algorithms can easily disregard ethical standards. An example interviewee 1 gave that showed the importance of including ethics is when his team tried to map car damages. "Car damages are way higher for houses with number additions. The reasoning for this is that these are usually flats, where cars are parked closer together and the change for car damage is higher." (Appendix 5). However, the implications of changing the insurance fee are more ethical, since "In general, people that live in flats, usually have a lower wage." (Appendix 5). This would mean that if the fee is increased, people with lower wages are indirectly discriminated against. Interviewee 1 then argues that if this is done, certain groups of the population can be made uninsurable (Appendix 5).

AI technology should also serve a purpose. Interviewee 1 (Appendix 5) noticed that AI is not always used appropriately and this can drive the costs up. He stated that: "In your AI-policy, you want to achieve the goal, with the most fitting and simple technology." (Appendix 5). This is why he thinks that AI technology should be aligned with the goal. He also noted (Appendix 5) that this goal can change, which would mean that the technology usage could also change.

Algorithms also need to be changed continuously, based on the context it is based upon. If the context changes, the algorithm should be adapted to fit this new context. Interviewee 1 stated that: "If an algorithm is used to detect fraud, criminals might catch wind that an organization is using a specific algorithm. These criminals might then change their behavior to dodge the algorithm, therefor it is key the algorithm keeps being trained." (Appendix 5). Therefore, continuous change is added to the AI-models component.

A last addition interviewee 1 mentioned (Appendix 5) is the quantity of data. He stated that: "In order to achieve great algorithms, a lot of data is often needed." (Appendix 5). This high quantity of data does not necessarily pose a higher risk of leakage, but the impact of a potential leak is higher. The solution interviewee 1 proposes is to reduce the data quantity as much as possible (Appendix 5). The data minimalization falls under data leakage, which is part of the privacy component.

4.2.3 General policy components

General policy components consist of, based on the literature and practice, general guidelines, liability, design & strategy, employee, ethics, general policy information, and privacy. Again, some components have been validated by the employees, while interviewees suggest changing some others.

General guidelines have been validated by interviewees 1, 2, and 3. Interviewee 2 stated (Appendix 7) that defining responsibilities and the scope of a policy is important. Employees are asking: "Are we, as a team, responsible for all entities within the organization? Or are we only responsible for entities that

use our ICT?" (Appendix 7). This is an example of why responsibilities and the scope need to be specified within every policy, not just the data & AI policies. Interviewee 2 also stated that: "It should also be specified who is responsible for a policy." (Appendix 7). She added to this that this person should also maintain the policy. Interviewee 3 noted (Appendix 9) that a policy should include the activities and interests of an organization in relation to the policy. This gives the needed context a policy should have, since "Everything is context, the local water company has other components they desire for their policy than the primary school does" (Appendix 9).

The liability component also received validation. Interviewee 2 mentioned: "Risk assessment is important, it is key to correctly identify the risks since they change continuously." (Appendix 7). Without correctly identifying these risks, an organization makes themselves vulnerable. Interviewee 2 also states: "If you look at that (liability), you know your security." (Appendix 7). She uses the risks from the risk assessment as a basis for their security. The DAMA-DMBOK framework, which was mentioned by interviewee 3 (Appendix 9), also highlighted identifying risks as a sub-component.

The DAMA-DMBOK framework also highlighted several sub-components from the design & strategy component. Sub-components that are validated are design and architecture. A new sub-component is 'identifying data needs and requirements', which is also mentioned by interviewee 3 (Appendix 9). Interviewee 3 states: "In your organization, you need to look at what kind of user groups you have, what they use, and what important is for them." (Appendix 9). Identifying these needs and requirements is important in the design of systems and policies.

Culture and behavior, also known as the employee component, are mentioned by every interviewee. Interviewee 1 emphasized the importance of culture and behavior, stating that: "Employees need to understand and be aware of the risks of sharing data files." (Appendix 5). This is just one of the many desired behaviors that interviewee 1 expects employees should follow. He gives a few other examples, but they all follow the trend that employees should be aware of the content of a policy and related laws and regulations. Interviewee 2 states that "A policy's success or failure depends on an employee's compliance" (Appendix 7). A policy is merely just another document if the employees do not comply with it, this is why she thinks it is important employees are motivated to comply with the different policies (Appendix 7). She also gave an example where the IT department was not aware they made a mistake since they were not aware of the policy that tries to prevent such a mistake. This example showed her that employee awareness is also important and should not be underestimated. Interviewee 4 (Appendix 12) uses a data science framework, which is: "A kind of roadmap to make sure you comply to all kinds of business formalities by using the framework. An application of the policy in practice" (Appendix 12). They communicate this via a three-step approach, "Show it, do it together, do it yourself" (Appendix 12). The first step is showing employees how the data science framework works and should be used (Appendix 12). After this, the framework is used together with the employee. The third step is letting the employee use the framework themselves (Appendix 12). This approach makes it, so employees are in a way 'trained' to follow certain steps and desired behavior.

The ethics of a policy is mentioned less and only mentioned by interviewees 1 and 4. Interviewee 1 stated that "The awareness (of ethical standards) is very important". He mentions that employees should be aware of the ethical standards of their organization, or they cannot practice those standards. Interviewee 4 (Appendix 12) noted, in regards to their ethical framework: "We use the (policy) documents from the Verbond (van Verzekeraars (VVV)), we do not have a separate document.". This means that their ethical framework is, on the highest level, the same as that of the VvV. However, the policy is quite generic and, as the interviewee states: "In practice, you cannot say I comply, or I do not comply (with the policy from VvV)" (Appendix 12). This is why they have different levels of policy. However, this falls under the general policy information and will be explained there.

The general policy information component has a small overlap with some of the comments from the interviewees, but most sub-components are not mentioned by the interviewees. Again, this is due to the nature of this component, which is that it is mostly background information that is used when making or adapting this component. Interviewee 4 (Appendix 12) mentions that they have three different levels of policy. These different levels can be used to translate guidelines from policies into practical rules employees can use in their everyday activities (Appendix 12). Interviewee 4 explains the levels as follows: "The highest level is the policy from the VvV, which gives generic guidelines on how to handle AI." (Appendix 12). The second level is from a privacy perspective, here the interests of a customer are evaluated. The third level is the more practical level, which is regarding the actions of a policy, "What will we do with this" (Appendix 12). These three levels are used to encompass all mandatory elements, but also what the organization strives to do in its strategic goals. Another comment that was made by interviewees 1, 2, and 3 is that the priority of components changes based on the context. An example is given by interviewee 1 where he states: "The focus, as we grow more mature, goes more to data quality." (Appendix 5). Interviewee 3 also confirms that the priorities differ for each organization by saying: "What drives an organization, what is the essence, this determines the value and priority." (Appendix 9).

Interviewees also suggested that privacy included more components than the practice and literature suggested. Interviewee 2 (Appendix 7) mentioned: "A data leak prevention (DLP) can be a solution" in regards to the problem of sending the wrong data and violating privacy. Interviewee 1 (Appendix 5) also noted that a data leak procedure is important. Another addition, which was also mentioned in the AI policy components dimension, is data reduction and minimalization. Interviewee 1 mentioned this in AI policy components, but not in the general policy components context. However, the DAMA-DMBOK Framework that interviewee 3 (Appendix 9) mentioned did contain data reduction in this context. A more in-depth look at this can be seen in appendix 11.

An outlier, compared to the literature and policy analyses, is the retention of documents, which is only mentioned by interviewee 1 (Appendix 5) and has not been found in the literature or practice. The retention of documents is different from the retention of data, in that document retention applies to documents (which includes data) and data retention only applies to data (Interviewee 1, Personal communication, 2022). This is an outlier since document retention is not mentioned in either analysis. Another outlier is the general security component that was mentioned by interviewee 1. This component consists of several security-related sub-components, such as requirements for applications, or more general security measures that do not specifically apply to data. These are outliers since they are only mentioned by interviewee 1. This might be because both analyses focused on data security and not a more general security perspective.

4.3 Comparison and final adjustments

Now that both analyses and the interviews have been done, the results from these can be compared. This will be done by looking at each component and their respective input from the literature, practice, and interviews. An overview of these inputs can be seen in table 8.

When an interviewee did not have any notes about a component, it is assumed they agree with the component and its place within the overview. An in-depth overview of the specific differences in each component can be found in appendix 14. "Validated by interviewee" ("X") means that an interviewee mentioned the component, but also that their definition of this component aligns with the definition in this study. If the definition alignment is not the same, it is partly validated ("O").

Table 8 shows that some components have been mentioned more than others. The high frequency of mentions can be connected to the importance of a component, but it can also be tied back to the background of the interviewees. Each component will explain the frequency of mentions.

Table 8: Overview components from literature and practice and whether they are validated by interviews

	Literature	Practice	1	2	3	4
Data policy components	Data security	Data security	X	X	X	
	Data lifecycle	Data lifecycle	O	X	O	X
	Roles & Responsibilities	Roles & Responsibilities	X	O	O	
AI-policy components	Explicability	Explicability	X			
	AI-Principles	AI-principles	X			X
	General AI	-	O			
	Technical guidelines	-				
	Privacy	-	O			
	-	AI-models	O			
General policy components	General guidelines	General guidelines	X	X	O	
	Liability	Risk assessment		X	X	
	Design & Strategy	Design & Strategy		X	O	
	Employee	Employee	X	X	X	O
	Ethics	Ethics	X			O
	General Policy information	Others	O	O	O	O
	-	Privacy	O	O	O	

Very similar	
Differences between analyses or additional sub-components	
Missing in other analysis	
Validated by interviewee	X
Partly validated by the interviewee	O

4.3.1 Data policy components

As can be seen from table 8, the data policy components dimension consists of data security, data lifecycle, and roles & responsibilities. The table shows that the data policy components have been mentioned by most interviewees. The high frequency of these components is probably because all four interviewees are considered to be data policy experts and would most likely mention these three components. Table 8 shows that data security has been mentioned five times and has also stayed consistent with its definition and sub-components. Therefore, the sub-components of data security stay the same and will be included in the final list. All three interviewees also agreed on the definition of data security.

The data lifecycle component is always mentioned, however, table 8 shows that this definition is not always the same. The difference in definition is mostly based on what sub-components this component should include. The literature has several sub-components that fall under data lifecycle, whereas the practice lacks some of these components. Interviewees 1 (Appendix 5) and 3 (Appendix 9) also give several additions to what the data lifecycle component should discuss, while interviewees 2 (Appendix 7) and 4 (Appendix 12) only validate previous results from the literature and practice. The final sub-components of the data lifecycle can be seen in appendix 17.

The roles & responsibilities component has been mentioned a lot, but the inclusion of sub-components differs. Again, the high frequency is most likely coming from the fact that the interviewees are data

policy experts. The differences are mainly regarding what function the roles & responsibilities should have and how this should be connected to the policy. Interviewee 2 suggests changing the focus from data ownership to process ownership, whereas interviewee 3 suggests adding the audit and compliance of the roles & responsibilities to this component. The final list of sub-components of roles & responsibilities can be seen in appendix 17.

4.3.2 AI policy components

Table 8 shows that the AI-policy components dimension consists of explicability, AI principles, general AI, technical guidelines, privacy, and AI models. As can be seen from table 8, most of these components are not validated, since the interviewees are experts on data policies and not AI policies. Those that indicated that they are not experts on AI and thus have limited knowledge, did not mention anything regarding the AI components. An exception is interviewee 4, who mentioned one thing about AI principles and how these are used in his organization. Interviewee 1 is the only one who mentioned multiple components and discussed them more in-depth and with examples, which is probably because he was familiar with the topic of AI.

When looking at the components of AI, explicability is the same in both analyses and confirmed by interviewees 1 and 4. The AI principles component is also the same in both analyses and confirmed by interviewee 1. Therefore, they will not change and be the same in the final list of components.

General AI is missing in the practice entirely, but it has been partially validated by interviewee 1 (Appendix 5). However, the lack of mention of this component might be because this component consists of background information that should be remembered in the back of a policy maker's head when making an AI policy. It also contains information that is usually placed in the 'others' category. The technical guidelines component has only been mentioned in the literature and nowhere else. This might again be because it is often forgotten (Hagendorff, 2020).

Privacy is similar to the general AI component, in that it is only mentioned in the literature and partly validated by interviewee 1 (Appendix 5). However, the privacy component is also present within the general policy components dimension, where the practice does mention privacy, and interviewees 2 and 3 also validate privacy there. So, the table might suggest that this component is not supported enough, but this is because the component is divided between two dimensions.

The AI-models component is only present in the practice and has only been validated once by interviewee 1 (Appendix 5). Only a small addition was suggested regarding this component. The lack of mentions in the literature might be because the AI models component is more practical, whereas the literature focuses more on the ethical side of AI. However, this is merely speculation.

4.3.3 General policy components

The general policy components dimension consists of general guidelines, liability, design & strategy, employee, ethics, general policy information, and privacy. Liability and design & strategy are very similar in both analyses and validated by the same interviewees, interviewees 2 and 3. A small addition was suggested by interviewee 3 regarding design & strategy, which can be seen in appendix 9.

The general guidelines are for the most part the same and validated by the interviewees. However, the practice goes more in-depth and has more sub-components, which is the main difference between the practice and literature. Interviewee 3 (Appendix 9) also suggests adding a small description of the organization's key activities to the data & AI policy. Table 8 also shows a relatively high frequency of mentions, where two interviewees validate the findings from the literature and practice. Potential reasoning for this might be that these sub-components are commonly found in the introductions of most policies and therefore have a high frequency.

The employee component is mentioned in both analyses and by all interviewees. Most interviewees also validate and thus agree with the results from the analyses, interviewee 4 being the exception. Interviewee 4 (Appendix 12) had a different way of educating and training employees. The analyses only differ on the direction of the communication. The high frequency of mentions by interviewees might be explained by the fact that all interviewees mentioned the importance of employees. Since interviewees perceive this component to be important, it gets mentioned a lot.

Ethics differs between the literature and practice. The literature focuses on specific standards, whereas the practice focuses on procedure and methods. Ethics was also mentioned by two interviewees, where interviewee 1 validated the sub-components. Interviewee 4 (Appendix 12) bases their ethics on national organizations, which explains where their standards and methods are based upon.

The general policy information component always differs every time it is mentioned. The literature and practice have a lot of different sub-components. This is because this component, just like the general AI component, consists of background information that should be remembered in the back of a policy maker's head when making a data or AI policy. They differ from each other because the sub-components are usually only mentioned once or twice. The general policy information component does however have a high frequency of mentions by interviewees. This can most likely be tied to one sub-component that three interviewees all mentioned, that priorities of components change, and that they are context-dependent. According to interviewee 3: "What drives an organization, what is the essence, this determines the value and priority." (Appendix 9). Most of the other general policy information sub-components are not mentioned, with a few exceptions.

Privacy was mentioned within the AI policy components dimension, but the practice and interviewees place it under the general policy component dimension. The difference between the literature and practice is that privacy in the literature is focused on the AI context, whereas in practice, it is seen in a broader context. The sub-components are very similar. In regards to the interviewees, interviewees 1 (Appendix 5) and 2 (Appendix 7) state that data reduction and a data leak procedure should be a part of privacy.

Finally, general security and retention periods are the last two components. Both have been mentioned by interviewee 1 (Appendix 5) but were not present in the literature and practice, so no comparison can be made.

4.3.4 Redistribution and prioritization components

All the components and their respective sub-components have been identified. However, some do not fit under the respective dimension or even their respective component. Therefore, some need to be redistributed and put under more fitting components or dimensions.

The first change will be regarding ethics. Ethics had a relatively small number of sub-components, however, after the comparison had been done, some sub-components could be added to the ethics component. The AI principles component is one of these components that can be added to the ethics component since the principles are heavily focused on and inspired by ethics. However, data also needs some sort of ethical principles. Therefore, ethics will have an entirely new component called "Ethical principles", which consist of AI principles, but also data principles. These principles are regarding ethical data & AI usage, but also regarding ethical (AI) decision making. Technical guidelines are also added to the ethics component since technical guidelines are focused on the translation of ethical principles to practice. To clarify the meaning behind this component, the technical guidelines component is renamed to "connect principles with practice". General policy information also contains a sub-component that is related to ethics, which is the "integration of policy into culture and behavior".

This is also added to the ethics component. A final addition to the ethics components is that of national organizations and their influence. This was removed in the literature review, but because its influence was mentioned again in interviewee 4, it is added as a sub-component within ethics. The sub-component is called “Translation national organization guidelines”.

The employee component also gains some sub-components. The first one is the “access to laws and regulations”. Access to laws and regulations is key, according to interviewees 1 (Appendix 5) and 2 (Appendix 7). But it does not fit under the general policy information and is something that can be included in a policy by adding a specific guideline revolving around this. Another addition is connecting principles with practice, which can also be seen in the ethics component. However, since the connection between principles and practice is also important for employees, it is added to the employee component.

Since privacy was present in two different dimensions, one version needs to be moved to another dimension. Because the privacy component can be applied in both the data and AI context, it has been added to the general policy components dimension and removed from the AI policy components dimension. This was done based on the input from the literature and interviewees. The sub-components from both the literature and practice were very similar, so they can easily be merged. The input from the interviewees will be added to this component. These are the ‘data leak procedure’ and ‘data minimalization’ sub-components. However, it should be noted that privacy is more present in data than it is in AI. This is also because privacy is a sub-component in the data security component.

General AI contains sub-components that apply to the design of AI and align with the AI-model component. Interviewee 1 (Appendix 5) already suggested the ‘continuous change’ component should be placed under AI models. The other two sub-components are about aligning AI with the organization, these are also placed under the AI-model component as “Aligning AI with the organization”. This leaves the general AI component with no sub-components, and it is therefore removed.

General policy information is, as previously mentioned, a special component since it consists of information that should be kept in mind when making or adjusting policies. The priority, according to appendix 15, is the lowest and marked as a Would not have. Therefore, it has been decided to remove this component from the list. However, it should be noted that this information can still be useful when making or adapting policies. A policy maker could read these before making a policy.

The other final components consist of the sub-components identified throughout this research. The in-depth overview can be seen in appendix 17. To increase the clarity of the components and dimensions, the general policy components dimension has been renamed to “Supplementary policy components”. This indicates that these components are used to support and enhance the data & AI policies. However, the supplementary components can still be important and labeled as a Must-have.

The general guidelines component also includes the “general” in their name, but, a more fitting name has not been found. However, they are general guidelines for what should be included within a policy, these are also important for data & AI policies.

Some components that did not fit under any other components are general security and retention periods. These were only mentioned by interviewee 1 (Appendix 5) and are thus seen as outliers. Because they are outliers, only mentioned by one interviewee, they are disregarded in the final list.

4.3.5 Concluding marks comparison

If the changes are implemented, table 9 is the final list of components. Table 9 also includes the priority indication, which is based on the literature and interviews. Appendix 15 has an in-depth overview of

how this prioritization came to be. It should be noted that the priority of these components is dependent on an organization and its context and that the priority indication in table 9 is focused on Alpina.

Table 9: Final list of components with their priority indication

Dimension	Component	Priority
Data policy components	Data security	M
	Data lifecycle	M
	Roles & responsibilities	M
AI-policy components	Explicability	M
	AI-models	C
Supplementary policy components	General guidelines	S
	Liability	C
	Design & strategy	C
	Employee	M
	Privacy	M
	Ethics	M

Some of these components have connections with each other, which should not come as a surprise, since they are all connected to data or AI. There were some problematic connections, however, the most problematic connections have been resolved in chapter 4.3.4. Other noteworthy connections are between the employee component and ethics, these two are connected in trying to communicate ethical principles to employees. Privacy is also connected to data security since privacy is a sub-component of data security. Furthermore, in Alpina, the liability component is connected to data security. Alpina uses the risk assessment to design its data security (Appendix 7). Other connections are mainly between the supplementary policy components and the other components since the supplementary components have a more supportive nature.

As can be seen from table 9, all the data policy components and explicability are all marked as Must-have and should be the focus of a data & AI policy within Alpina. This is also the case for the employee, privacy, and ethics component. However, this does not mean that the other components are less important. The Must-have priority for Alpina mean that in the current stage of Alpina, these components should be their main focus.

4.4 Alpina coverage

Now that the final list of components is known, Alpina's coverage of these components can be analyzed. Table 10 shows the list of components again, with colors indicating to what extent Alpina covers these components. An in-depth overview can be seen and read in appendix 16.

4.4.1 Coverage components

As can be seen from table 10, Alpina has 5 components that have a high priority and are incomplete. These should be the main focus for Alpina.

Table 10: Alpina coverage

Dimension	Component
Data policy components	Data security
	Data lifecycle
	Roles & Responsibilities
AI policy components	Explicability
	AI-models
Supplementary policy components	General guidelines
	Liability
	Design & Strategy
	Employee
	Ethics
	Privacy

	Covered
	Incomplete, low priority
	Incomplete, high priority

Data lifecycle and roles & responsibilities both have sub-components that Alpina does not cover. Noteworthy sub-components within the data lifecycle that Alpina does not cover are data usage, data quality, and data lineage. Data usage is about describing how data is used within Alpina. This is important since confusion about how data is and should be used can lead to unethical practices (Parsons, 2013; Abraham et al., 2019). Data quality is also a big sub-component that consists of multiple other sub-components according to interviewee 3 (Appendix 9) and the DAMA-DMBOK framework (International, DAMA, 2017). Alpina does not currently cover this component in the right context, they only touched upon it in the context of AI models. Data lineage has been mentioned by interviewee 1 and is connected to the inclusion of documentation. Alpina currently lacks the required data documentation and cannot see what the lineage of their data is. Without knowing the origin of data, the classifications of data can be incorrect. This in turn can lead to incorrect usage of data, which can lead to even more problems. Therefore, the data documentation and lineage must be in order.

The roles & responsibilities component also has some sub-components that Alpina does not cover. The most noteworthy are process and data ownership. According to Abraham et al. (2019), data ownership is about allocating accountability for data, which can be placed upon a specific role. Process ownership is about allocating accountability for a process. This is important since interviewee 2 stated that according to the law an accountable employee is needed for mandatory assessment. However, Alpina is lacking employees that are appointed as accountable for processes and data.

Within the design & strategy component, Alpina lacks one sub-component in their coverage, which has been identified by interviewee 3 and the DAMA-DMBOK framework (International, DAMA, 2017). This sub-component is 'identify data needs and requirements'. They could have done this and not have it be included within their policies, which would explain why it was not found in the policy analysis. In that case, they could simply add this to their existing policies to make sure the written data needs and requirements are aligned with the data needs and requirements of the employees. If Alpina did not already identify these, they should do so. This is important since these needs and requirements are

used as a basis on which the policy is designed. It influences how the components are formed within Alpina.

The employee component is covered mostly by Alpina, where only the connection between the principles and practice is lacking. This is an important connection since Hagendorff (2020) mentions that guidelines alone are not enough, guidelines need to be more in-depth and translated to practice. At the moment, Alpina has some techniques in place, like education & training activities. These try to close this gap between the policies and practice. However, this can be improved upon, since interviewee 2 stated that this gap is still a problem within Alpina.

Ethics is the last of the 5 components that Alpina does not cover completely but also has a high priority. Alpina covers most sub-components, but they are not completely covered, there is still something lacking from these sub-components. The ethical principles, which consist of AI principles and data principles, are not completely covered, since Alpina only covers the AI principles completely. They do not cover data principles. This is important to cover, since the data principles are influencing the data lifecycle (data usage in particular), but also data security. Alpina also lacks integration of its policy into culture and behavior. This integration can be done through education and training (Abraham et al, 2019; Siau & Wang, 2020), which Alpina already has in place. However, as previously mentioned, they can still improve in this area. Future research regarding the implementation of these methods could be performed to show how big the problem is and what the best way to solve it would be.

These were the 5 high-priority components that Alpina still lacks. The lower priority components, AI models and general guidelines, only have 1 component that is lacking. For AI models this is aligning AI with the organization, and for general guidelines, this is a description of key activities of the organization. An in-depth overview of the components, their sub-components, and what Alpina covers can be seen in appendix 16.

4.4.2 Maturity levels

While the components in table 9 have a priority indication, the interviews stated that these priorities are dependent on the context and phase of an organization. This means that the priority of components can shift when the context of an organization's phase is changing. Interviewees 1 and 2 gave some examples of their shifting focus. Interviewee 1 stated: "So in terms of attention, as we become more mature, the focus shifts more towards data quality." (Appendix 5). Interviewee 2 mentioned that a similar shift is happening for the design & strategy component. Their reasoning for this shift is that in the beginning, measures regarding data security, but also design & strategy, are described and put into place. If this is done correctly, the most important activity is to (re-)evaluate this and adapt when necessary.

This shows that Alpina bases its priority on its current phase. For this research, the concept of knowledge management maturity (Jiankang et al., 2011) is used. This maturity model has different levels, where a low level is equal to new organizations, where processes are in their starting phase (Jiankang et al., 2011). The higher the levels go, the more advanced the organizations becomes with more advanced and improved processes (Jiankang et al., 2011). For this research, three levels are used, levels 1, 2, and 3.

Level 1 is seen as a beginner organization, this organization has chaotic processes and only basic processes are defined (Jiankang et al., 2011). A level 1 organization is very chaotic and should focus on building and defining processes. This organization is thus focused on the components that are critical and need to be defined and built, such as design & strategy and roles & responsibilities. Since there is a lot of chaos in a level 1 organization, components such as liability, data security, ethics, and privacy

are also important. This is because a lack of these components can cause even more chaos and sometimes even irreversible damage. The data lifecycle component can also help with this chaos since this component offers a good overview of the data flows. It can also be used in combination with the design & strategy component to optimize the design and architecture of the organization.

Level 2 is a more advanced organization, where processes are defined and some form of standardization is done (Jiankang et al., 2011). A level 2 organization should focus on making their processes more manageable and make sure they can optimize improvement. Since data lifecycle is very focused on processes, optimization, and improvement, an organization should prioritize this component. Improvement and optimization are also key to the employee component, so this should also be an important component. The employee component can also be put at the first level since it is very important to communicate a policy's content to employees. However, a policy's content needs to be defined first, after this, communication is key. Level 2 also allows developing components that are seen as an extension, such as the AI components. Since AI is still in the developing phase and not yet critical to Alpina's business processes, it is put into level 2.

A level 3 organization is the highest level and is focused on continuously improving its processes (Jiankang et al., 2011). Again, data lifecycle and employee are both important components, since they help with the continuous improvement part. This level also gives space for organizations to improve on their other less critical components, that might not even be in the list of components that is a result of this study.

It should be noted that, while the maturity levels of an organization can influence the prioritization of the components, it is often not the only influence. Therefore, components such as AI models might be critical to an organization and are touched upon in a level 1 organization.

Alpina is currently in-between level 1 and level 2 of this maturity model. They are not a beginner organization, but because of the merger, they have to redo and transform certain processes. Recommendations regarding this will be discussed in chapter 5.

5 Conclusions & recommendations

Both analyses have been performed, interviews have been held and the results from these were analyzed and compared. The coverage of Alpina has also been analyzed. This chapter will feature the conclusions and recommendations of this study and will answer the research question.

5.1 Conclusions

A literature review and policy analysis have been conducted to identify several components that should be present in data & AI policies. These findings have been validated and elaborated upon through interviews with experts in the data & AI policy field. This resulted in a list of components that an insurance organization should include in their data & AI policies. This list of components can be seen in table 11. However, not all of these components have the same priority and do not need to be included to the same extent. Therefore table 11 also includes the prioritization of components within Alpina, which is communicated through the different colors. These priorities are specifically for Alpina, and other organizations can differ in their prioritization since their context might differ. Table 11 also includes the maturity level of the components, which is again, directed towards Alpina.

Table 11: Final list of components in data & AI policy with their priority included

Dimension	Component	Maturity
Data policy components:	Data security	1
	Data lifecycle	1
	Roles & Responsibilities	1
AI policy components:	Explicability	2
	AI-models	2
Supplementary policy components	General guidelines	2
	Liability	1
	Design & Strategy	1
	Employee	2
	Ethics	1
	Privacy	1

	Must
	Should
	Could

The research question within this research is: *"To what extent should an insurance organization include data & AI components within their policies?*. The list of components answers what components should be included. The "To what extent" part of this question is answered through the use of the MoSCoW and maturity model. The MoSCoW model indicates what components should be prioritized. However, this priority is context dependent and one of these context factors is the maturity of an organization. Therefore, the maturity model can be used to give more context to what an organization should prioritize. So the extent to which an insurance organization should include data & AI components is based on their context which consists of variables such as prioritization and maturity level.

An insurance organization, or similar organizations, can use the results from this study as an index of what they should include in their data & AI policies. However, they should perform an analysis of their organization first. They should identify several variables such as their prioritization and maturity level. Other variables can also be important, but they are not discussed within the scope of this study.

5.2 Recommendations

Now that the research question has been answered, recommendations based on this answer can be given to Alpina and its policies. Chapter 4.4 analyzed the coverage of Alpina compared to the final components. This chapter concluded that Alpina has five components they do not fully cover, but do

have a high priority. The high priority is based on their maturity level, but also on what interviewees and the literature have prioritized. These components are data lifecycle, roles & responsibilities, design & strategy, employee, and ethics.

For the data lifecycle, Alpina's biggest concerns are data usage, data quality, and data lineage. For these sub-components, defining and describing them is important. Data usage and data quality should be defined and, together with data lineage, the process around this should also be described. Employees can be involved in this description process since they can offer valuable insights into what data is used for and how data quality can be safeguarded for example. The employees offer a perspective on how the data lifecycle functions within Alpina. Employees from the data team could offer the most valuable insights, but employees from other teams could also offer different insights.

In regards to the roles & responsibilities component, Alpina should focus on the sub-components process and data ownership. As previously mentioned, if Alpina does not have these defined, they cannot perform legally mandatory assessments. This would lead to non-compliance to law and regulations and will most likely result in significant damages. Therefore, it is important Alpina defines the process and data ownership, their responsibilities, and appoints these roles to employees.

Within the design & strategy component, one sub-component is lacking, which is defining data needs and requirements. As previously mentioned, Alpina could have done this but did not put it into their policies. For now, it is assumed that Alpina did not do this and is lacking an overview of data needs and requirements. Alpina should identify this by having conversations with key employees in different departments, this would give them an overview of the needs and requirements of each department. This overview can then be used to design or change certain architectures and systems.

For the employee component, Alpina's focus should be on connecting the principles in their policy with the practice, since merely stating the guidelines is often not enough (Hagendorff, 2020). Training and education are both methods that are already used to close the gap between policy and practice, however, Alpina can improve in this area. Topics within policies such as ethics and desired behavior regarding data usage can be communicated to Alpina employees through training and education. Therefore, it is recommended that Alpina expands upon training and education activities.

Within the ethics component, Alpina should focus on translating ethics to practice and integrating this into their culture. This translation is key (Hagendorff, 2020) and without it, the content of a policy could be perceived as abstract and not realistic. The ethical principles should also be integrated into the culture of an organization, where employees' behavior aligns with guidelines. It is recommended that Alpina defines the ethical principles and then, together with employees, tries to link this with practical guidelines that employees can implement in their daily work. Training, education, or a data science framework (interviewee 4, appendix 12), can help with this integration.

From these five components, common themes that Alpina is lacking can be seen. The first is that certain (sub-)components should be defined and described. The other common theme is translating policy to practice. This is in line with their maturity level, which is between levels 1 and 2. Alpina is focused on merging systems and policies of Heilbron and V&V. So they, according to their maturity level, should be focused on defining (sub-)components and contents first. The components that have priority are data lifecycle, roles & responsibilities, design & strategy, and ethics. After the content and (sub-)components of a policy have been defined, Alpina should focus on translating the policies to practice. The main component that should be focused on here is the employee component since this is focused on communicating and translating policy to practice. Alpina should also work on its evaluation metrics and optimize continuous improvement of its data policies, which are part of the general guidelines and employee components.

6 Discussion

This research has resulted in a list of components that should be present in data & AI policies and recommended Alpina on what they should improve regarding their data & AI policies. This list of components can be used by other organizations as an index of what they should include within their data & AI policies. The list is mostly aimed at organizations within the insurance industry, but other organizations in similar industries might also benefit from using the list of components. Similar industries are industries that are active within the financial sector, such as banking or mortgages.

From an academic perspective, this list offers an overview of important components of data & AI policies. This list is based on the literature and practice, which offers a broader perspective. The research also uses interviews to validate and sometimes even expand the findings. These combinations make for a list that is based on different sources, instead of focusing on one area. With the addition of the practical component, the list becomes easier applicable for companies since the practical component gives the list a connection with reality.

6.1 Limitations

While the results of this research might be of value to some organizations, there are also some limitations to this study. A limitation of the interviews is that several interviewees noted that they are experts in data policy, but not in AI policy. This in turn limits the validity of the AI policy results, since not all interviewees can be marked as experts in both fields. The validity of the last interview is also limited in that it was only half an hour long (instead of the usual hour) and only the first part of the interview could be done. This resulted in gaining no feedback and validation of the results from the literature review and policy analysis.

Regarding the literature review, it is using sources from only one database, the SCOPUS database. All articles within the literature review are from this database, which might give a limited perspective. Appendix 1 dives into this issue by analyzing results from other databases, however, this only shows that other databases such as Google Scholar have different articles. It should also be noted that access to articles within SCOPUS was limited and some articles were excluded since the researcher did not have access to them. Articles were also filtered on English and Dutch articles since the researcher can only understand these articles. The limitation here is that other languages, which could have provided additional insights, were excluded.

The literature review is also limited in that specific terms were used to search for data & AI-related articles. However, not all synonyms and terms could be used, since this would make the final selection of articles too big.

The frequency of codes found in articles and policies could have been included to provide a better insight into what codes are mentioned a lot and what codes are barely mentioned. The only time it played a role is when the priority indication for the literature was made. This means that a code that was only mentioned once or twice had the same impact as a code that was mentioned 20 times. This is a limitation of this study, since codes that were only mentioned once, are treated similarly to codes that have been mentioned more than 5 times. This should be accounted for in follow-up studies.

Another limitation connected to the literature review is its general perspective. The literature review used search terms that were not connected to the insurance industry, which lead to articles that were not specifically focused on the insurance industry. This in turn leads to a list of components that is not necessarily focused on the insurance industry, but more a general list of components for data & AI policies. This general perspective of the literature might steer the final list of components more towards a general perspective. However, the policy analysis and interviews might have reduced the

impact of this limitation by focusing solely on policies from an insurance company and interviewing employees who are experts in their field, within the insurance industry.

The bias of the researcher is another limitation of this research. The researcher could have had an unknown bias at the time of the analyses, this could explain why some components from the interviews were not found in the analyses. This bias might be that the researcher is focused on data & AI, which could lead to the researcher missing other relevant components.

The generalizability of the list of components to other organizations is also a limitation within this study. A case-study method is often critiqued for its lack of generalizability. However, some generalizability is achieved by using interviews to validate the findings from both analyses. These interviews increase the external validity of the study. Nonetheless, the generalizability and thus the external validity is still limited because of the chosen study method.

The reliability of this study is also limited. During the making of this study, new policies within Alpina have been made and existing ones are being changed. Because of this, another analysis, with the same steps, might give a slightly different result. The data & AI field is also a very volatile one. This volatility can be the reason that a similar study gives different outcomes. However, this is not because the study changes, but because the source materials change.

6.2 Future research

While this research is focused on the content of a policy, it is also important that employees follow what is written in the policy. Interviewee 2 mentioned that this is not always the case and that there is a gap between what is written down and what is performed. This falls outside the scope of this study, but future researchers could research this gap and they could provide insights into how to close this gap.

Future research could also assist in better understanding the integration of the policy within the culture and behavior of employees and how to improve this culture and behavior. If this research is performed within Alpina, the internal audit team could provide insights regarding this topic. It is also recommended that future research regarding this includes insights from other companies since it was mentioned Alpina does not currently have an effective strategy regarding this integration. An interesting model that was found during this research that future researchers could use is the Protection Motivation Theory (PMT). This framework can help in gaining a better understanding of how to influence employees through fear appeals (Conner & Norman, 2015).

This research is focused on the combination of Data & AI, but most interviewees mentioned that they were not experts on AI, which resulted in a lower input regarding AI compared to data. Future research could focus on AI specifically, where experts in the field of AI are interviewed. The combination of data & AI could also be researched again, but it is then recommended to interview experts on data and AI equally.

While this research is based on policies and interviewees from the insurance industry, possible future research could be based on other similar or completely different industries. The results from those kinds of studies can be compared to the results from this study. If they align, this can serve as validation. If the results differ, that research could dive into why that industry differs from the insurance industry.

Future research could also dive deeper into the combination of the components and the maturity model and prioritization. Within this research, this was only touched upon. Future research could dive deeper into how a maturity model would work in this context, but also the different scenarios.

Appendices

Appendix 1: Literature review in-depth steps and explanation:

Step 1: Define

Part 1: Inclusion and exclusion criteria

Filter on the last ten years, where 2012 is still included since 2022 just started. (2012-2022). This is to only include recent articles and updated information.

Part 2: Field of research

Results should mainly be relevant to the research. So, they should be about the data policy, or something closely related (data governance, data management).

Part 3: Outlets and databases

The database that mainly will be used is SCOPUS since it is easily accessible for the researcher and has several filtering options. Other relevant databases can be the web of science or google scholar.

Part 4: Search terms

The search terms are what will be typed in the search bar of SCOPUS. These are:

- Data. Where information can be a synonym. This is one of the main topics of the research and is necessary to gather articles related to data policy components.
- Policy. Where governance can be a synonym or addition to filtering the search. This is also a main topic of the research and is necessary to gather articles related to data policy components. To search for both the word “policy” and “policies” the symbol * is used.
- Taxonomy (classification). This is to focus the search on a possible classification of data policies.
- Platform, Security, Storage, Ethics & AI. This is related to Alpina and what they have communicated. A contact from within Alpina also communicated some topics which might be of interest, these have been included in the search query.

A small deviation from Wolfswinkel’s article in this literature review is that the search process overlaps somewhat with that of the selection process. The search process will consist of entering search queries and getting the results from these queries to be as small as possible. The selection process will then immediately start, in which queries will be selected on their fit to the research. A search query should lead to ideally somewhere around 20 results, but exceptions can be made if results cannot be filtered down solely through the query. Queries that deliver relevantly and a relatively small number of articles will be selected. This selection is based on a quick analysis of the top 10 articles, if some articles stand out, the query will be selected.

The inclusion criterium for the first search is that an article should be about the data policy, or something closely related to it, such as data governance or data management. Articles will be excluded if they are not mainly about policy and just mentions it briefly. If a search query still has a lot of results left after limitations, older articles with few citations will also be excluded. A tool to limit the results is the different logical operators within the SCOPUS search engine.

- AND: It should contain both words
- OR: It should contain at least one of these words
- Pre/0: One word has to be directly in front of the other word

Another tool to limit the results is focusing on only the title and keywords, instead of title, keywords, and abstracts. The search query can also exclude certain irrelevant disciplines, these would be:

- Agricultural and Biological Sciences
- Arts and Humanities
- Biochemistry, Genetics, and Molecular Biology
- Earth and Planetary Sciences
- Energy
- Engineering. Excluded because it was too technical
- Environmental Science
- Materials Science. Excluded because it was too technical
- Mathematics
- Neuroscience
- Nursing
- Pharmacology, Toxicology, and Pharmaceutics
- Psychology
- Physics and Astronomy

After this first selection process, there will be an overview of the relevant queries and how many results they gave. Then the articles will be analyzed based on their title and potential relevancy to the research individually. Interesting and relevant articles will be selected and put into a list. The goal is to have somewhere around 40 articles on this list so that the next step will not take up too much time. Articles that are directly related to data policy or some sort of overview, guidelines, or framework that discusses policies, AI, or something related. Articles that are too technical will be excluded.

The list will then be used to conduct the final step in the selection process. If any articles were selected twice or more, they will be filtered out. The abstracts of each article will be analyzed, and relevant articles will be selected. If an abstract is missing or vague, a quick scan of the article will be done. Articles that are deemed relevant but cannot be accessed because it does not fall under the available databases of the University of Twente, will be categorized separately. These selected articles will be used to form an overview of the components a data & AI policy should include. The criteria here is that an article should be about data policies or AI, but it should also add value to the report.

Step 2: Search

Each query and its steps will be shortly described, the query will finish with a small table in which the results from each step within that query can be seen. The table also shows which queries, and their respective results, are selected and will move on to the next selection step. The selected step, connected to the query, is marked in yellow.

Data AND Polic*

1. Searched on Scopus, started with Data AND polic* in Title, Abstract, and Keywords (TAK). Got 374,033 results, which is too many. Results can be seen in table 12.
2. Searched on specifically Data polic*, in which data has to be directly in front of polic* to form the words: Data policy or Data policies. This is done by searching for Data Pre/0 Polic*. This will be used in all searches to refine results. This resulted in 973 results, which is still too much.
3. Only searched for data Pre/0 polic* in Titles and Keywords (TK), not in abstracts, since those could briefly mention the words and it could not be the main topics of the article. The result was 266 articles, still a lot.
4. Corporate was added to the search query to refine the results and focus on the private of private data policies. Only 12 articles were found, which were not relevant. A further approach is by using the second query (Data Pre/0 Polic* within TAK) and filtering down on those results by adding another word. This will be done in the next few queries.

*Table 12: Data AND Polic**

	Combination of words:	**Data/Policy**	
Step:	*Query:*	*Meaning behind query:*	*Results*
1	TITLE-ABS-KEY (data AN	Combination of data and policy	374,033
2	TITLE-ABS-KEY (data PR	Data has to be directly in front of policy	973
3	TITLE (data PRE/0 polic	Abstract is excluded	266
4	(TITLE (data PRE/0 pol	Filtered on articles with corporate in it, to focus on private policies	12

Data AND Polic* AND platform

1. Searched for platform AND Data PRE/0 Polic* in TAK, 94 results, which is a lot less, but top articles do not appear to be relevant and there are still too many articles to scan. Results can be seen in table 13.
2. Searched for the same keywords, but now only in TK. Only 5 irrelevant articles.
3. Went back to the first search query but searched for Data Pre/0 platform AND data pre/0 polic* in TAK. Showed 10 articles, of which some were interesting, these are included in the second search. (DPP)

Table 13: Data AND Polic AND Platform*

	Combination of words:	**Data/Policy/Platform**	
Step:	*Query:*	*Meaning behind query:*	*Results*
1	TITLE-ABS-KEY (platforn	Data has to be directly in front of policy and search should include platform	94
2	TITLE (platform AND d	Only searched in title and keywords	5
3	TITLE-ABS-KEY (data PR	Searched for data platform AND data policy, where data has to be directly in front	10

Data AND Polic* AND Governance

1. Searched for Governance AND data Pre/0 polic* in TAK. Relevant articles, but too many results (79). Results can be seen in table 14.
2. Excluded abstract in search, got 8 results. Results had some interesting articles, which are included in the second search. (DPGo)
3. Also tried to search for interesting articles using the word information as a substitute for data. Searched for governance AND information preceding polic*, got 68 results, however, the top 10 articles were not relevant.
4. Limited the search to exclude the abstract, got 23 results, but still no interesting articles. It might be that the combination of polic* and governance is not a good fit.

5. Since governance and polic* might not be a good fit, exclude polic* and search for data pre/0 governance in TAK. Got 1,234 results, which is too many.
6. Excluded abstract from search and only searched within TK. Still too many results with 770 articles, but there are some interesting articles when sorting on “most cited by”.
7. Exclude older articles (before 2012), but this did not remove that many articles, 733 are still left.
8. Exclude the list of irrelevant disciplines (See step 1), which only leaves “Computer science”, “Social sciences”, “Business, Management and Accounting”, “Decision sciences”, “Medicine”, and “Economics, Econometrics, and Finance”. This query has 389 results, which is still quite a lot.
9. To limit the query, “Corporate” was added to limit the articles on the private sector. This limited the results down to 96.
10. Another extra limitation is by increasing the limit from 2012 to 2016, so articles before 2016 are excluded. This results in 83 results.
11. Further limitations to the query are not done, but the top-cited articles are used. Any articles that are cited 10 times or more are selected. To be fair to new articles, those that have come out in the last year (2021 and 2022) will also be used. This results in 25 articles that are cited a lot and 17 new articles, in a total of 42 articles. These are added to the second search. (DG)
12. Information was used as a substitute for data, 21 results came out of the query.
13. When the same method was used for information, only 4 articles were cited 10 times or more and only 4 articles came out in 2021 or 2022. This is a total of 8 articles, which are included in the second search. (IG)

Table 14: Data AND Polic AND Governance*

	Combination of words:	**Data/Policy/Governance**	
Step:	*Query:*	*Meaning behind query:*	*Results*
1	TITLE-ABS-KEY (data PR	Data has to be directly in front of policy and contain governance	79
2	TITLE (governance AND	Exclude abstract in search	8
3	TITLE-ABS-KEY (informa	Replace data with information	68
4	TITLE (governance AND	Exclude abstract in search	23
5	TITLE-ABS-KEY (data PR	Tried searching on only Data Pre/0 governance, without policy	1,234
6	TITLE (data PRE/0 gove	Excluded abstract in search	770
7	TITLE (data PRE/0 gove	Exclude older years (Before 2012)	733
8	(TITLE (data PRE/0 gov	Exclude irrelevant disciplines	389
9	((TITLE (data PRE/0 g	Limit search by adding corporate to the search to focus on "private" governance	96
10	((TITLE (data PRE/0 g	Exclude articles before 2016	83
11	Same query as above, b	Only using articles that have more than 10 citations or are from the last year (2021 or 2022)	25
12	(TITLE (information PR	Same can be done for information Pre/0 governance AND corporate	21
13	Same query as above, b	Take the articles with 10 or more cited by and last year (2021 and 2022)	4

Data AND management AND Corporate

1. Data management is closely related to data polic* so the combination Data Pre/0 management in TAK was searched, which resulted in 45,332 results, which are too many articles. Results can be seen in table 15.
2. The abstract was excluded from the search, the query only looked at TK, which resulted in 23,843 articles, which was still too many.
3. A strong filter was needed to drastically limit the results, all articles older than 2016 were excluded. This resulted in a search of 8,169 articles, which was still a lot.
4. Irrelevant disciplines (from step 1) were excluded from the search. This brings the results down to 3,579.
5. “Corporate” was added to filter the search solely on the private sector and limit the results by a much-needed number, with this filter enabled the results came down to 98 articles.

6. To limit the results, only the articles that got cited by 10 or more sources will be used or that got released within the last year (2021 and 2022). These are 19 articles that got cited 10 or more times and 19 articles that got released in 2021 or 2022, which are 38 articles in total. These were added to the second search. (DM)
7. Information was used as a substitute for data to see if there were interesting articles for information management, with a filter on corporate. This resulted in 1,916 articles, which are way more articles than the query with data management gave.
8. To limit the results again, only the articles that got cited 10 or more times were selected, which were 230 articles. The 230 plus the 196 articles that got released in 2021 and 2022 accumulated to 427 articles. This is still too many results, and the top 10 cited articles were not deemed relevant. The decision was made to not include these results in the second search.

Table 15: Data AND management AND Corporate

	Combination of words:	**Data/Management/Corporate**	
Step:	*Query:*	*Meaning behind query:*	*Results*
1	TITLE-ABS-KEY (data PR	First search for data management	45,332
2	TITLE (data PRE/0 man	Exclude abstract	23,843
3	(TITLE (data PRE/0 ma	Excluding older years (before 2016)	8,169
4	(TITLE (data PRE/0 ma	Excluding irrelevant disciplines	3,579
5	((TITLE (data PRE/0 m	Filter on corporate to focus on private data management	98
6	Same query as above, b	Only using articles that have more than 10 citations or are from the last year (2021 or 2022)	19
7	((TITLE (information P	Same can be done for information Pre/0 management AND corporate	1,916
8	Same query as above, b	Take the articles with 10 or more cited by and last year (2021 and 2022)	230

Data AND Polic* AND taxonomy/classification

1. Searched for taxonomy to find if there were maybe interesting articles that divided data polic* into different categories. Got 10 articles, as a result, were not relevant to the research. Results can be seen in table 16.
2. Also tried it with information as a substitute for data, but to no success, 8 articles that were not of interest.
3. This was also done by changing taxonomy into the synonym “classification”. Information Pre/0 polic* AND classification was used, and this resulted in 30 articles, which could be smaller.
4. The next step is to remove the abstract from the query, and this resulted in two articles. At least one was interesting. These were included in the second search. (DPC)
5. Also searched for the combination Data Pre/0 polic* AND classification and 28 results were found, which could also be limited more.
6. The abstract was also removed here, this resulted in 8 articles, but there were no interesting articles within this search.

Table 16: Data AND Polic AND taxonomy/classification*

	Combination of words:	**Data/Policy/Taxonomy/Classification**	
Step:	*Query:*	*Meaning behind query:*	*Results*
1	TITLE-ABS-KEY (taxonor	Data has to be directly in front of policy and contain taxonomy	10
2	TITLE-ABS-KEY (taxonor	Replace data with information	8
3	TITLE-ABS-KEY (classific	Replace taxonomy with classification	30
4	TITLE (classification AN	Removed abstract	2
5	TITLE-ABS-KEY (classific	Replace information back with data again	28
6	TITLE (classification AN	Removed abstract	8

Data AND polic* AND merger

1. Tried to search on the combination of Merger AND Data Pre/0 polic* in TAK but got only 1 result. While the article only briefly mentions mergers, it is an interesting article nonetheless

that might offer value to the research. This article is added to the second search (DPM). Results can be seen in table 17.

2. Since the previous search only showed one article, the requirement of Pre/0 has been loosened a bit. This search will put the 0 to a 5, meaning data has to be 5 places or less in front of polic*. This resulted in 22 articles, but, apart from the previous article, no articles were interesting in this query.
3. The first query was used again, but data was replaced with information. This resulted in 5 articles, but there were not relevant.
4. The requirement of Pre/0 has loosened again, which resulted in 40 articles, but these were also not relevant.

Table 17: Data AND Polic AND merger*

	Combination of words:	**Data/Policy/Merger**	
Step:	*Query:*	*Meaning behind query:*	*Results*
1	TITLE-ABS-KEY (merger	Data in front of policy and it should include merger	1
2	TITLE-ABS-KEY (merger	Tried to loosen the search a bit by allowing data to be in front of policy within 5 spaces	22
3	TITLE-ABS-KEY (merger	Replaced data with information	5
4	TITLE-ABS-KEY (merger	Tried to also loosen the search a bit again, same as with data.	40

Data AND polic* AND security

1. Searching for Security AND Data pre/0 polic*. Got 105 results, which is quite a lot. The top 10 had some interesting articles. Results can be seen in table 18.
2. Excluded the abstract from the search and got 10 results. It showed some interesting articles, but excluded others that were noticed in the first search of security AND data PRE/0 polic*
3. Go back to the first search in this query and exclude older articles (before 2012). This leaves only 80 articles, which is still too many.
4. Irrelevant disciplines (step 1) are also excluded. Now there are only 42 articles left. This can still be smaller.
5. Only articles that are cited 10 or more times are included in the second search, which is 14 articles. The articles that are released in 2021 or 2022 are also included, these are 6 articles. 20 articles in total are added to the second search. (DPSe)
6. Use the same query but replace data with information. This will result in 44 articles, but the top 10 are not interesting, so none of them are included in the second search.

Table 18: Data AND Polic AND security*

	Combination of words:	**Data/Policy/Security**	
Step:	*Query:*	*Meaning behind query:*	*Results*
1	TITLE-ABS-KEY (security	Data in front of policy and it should also include security	105
2	TITLE (security AND da	Exclude abstract	10
3	TITLE-ABS-KEY (security	Include abstract again, but exclude older articles (before 2012)	80
4	TITLE-ABS-KEY (security	Limit years to last 10 years (2012-2022)	42
5	Same query as above, b	Only using articles that were cited 10 times or more, or are released in 2021 or 2022	14
6	TITLE-ABS-KEY (security	Try the same query, but replace data with information	44

Data AND polic* AND storage

1. Searching on Storage AND data Pre/0 Polic* will result in 40 articles. Still relatively big, but interesting articles at the top. Results can be seen in table 19.

2. Excluding the abstract will result in only 10 articles, does have some interesting articles, but not all from the previous search.
3. Widen the search a bit by going back to the first search query and excluding older articles (before 2012). This results in 36 articles, which can still be reduced some more.
4. Filter irrelevant disciplines out (step 1). This results in 16 articles. These are included in the second search. (DPSt)
5. The same query, but with data replaced by information, gave 5 articles. There were some interesting articles, so this is also included in the second search. (IPS)

Table 19: Data AND Polic AND Storage*

	Combination of words:	**Data/Policy/Storage**	
Step:	*Query:*	*Meaning behind query:*	*Results*
1	TITLE-ABS-KEY (storage	Do a simple search for combination storage with data Pre/0 policy	40
2	TITLE (storage AND da	Exclude abstract	10
3	TITLE-ABS-KEY (storage	Include abstract again, exclude older articles	36
4	TITLE-ABS-KEY (storage	Go back to first query and filter irrelevant disciplines	16
5	TITLE-ABS-KEY (storage	Also search for combination of Information, Policy and storage	5

Data AND Polic* AND Artificial AND Intelligence AND Ethic*

1. Searching on Data Pre/0 Polic* AND Artificial Pre/0 intelligence gives 16 articles.
2. Older articles (before 2012) are excluded, but there are no relevant articles within the 14 results from this query. Results can be seen in table 20.
3. Broaden the search by loosening the Pre/0 from data polic* to a Pre/5, which gives 266 results, but still no relevant articles in the top 10 cited by articles.
4. Do not want to filter on disciplines, since Artificial Intelligence could have relevant articles in other disciplines. Adding a new keyword to focus the search might help. Adding Ethic* so that the documents are focused on the Ethics of Artificial Intelligence and data policies. This results in 13 articles, some of which are relevant.
5. Loosen the search a bit more, so data pre/10 polic* is used instead of pre/5. This resulted in 21 articles, which are included in the second search. (DPAIE1)
6. To be complete in the search, there was also a query that looked at the combination of Artificial Intelligence AND Data AND polic* AND ethic* in only the TK and before 2012. This resulted in 27 articles, which were also used in the second search. (DPAIE2)
7. Also, replace “policy” with governance, maybe there are a lot of missing articles. Turns out there are some overlapping, but also some interesting missing articles compared to the previous searches. 29 results came from this query, these will also be used.

Table 20: Data AND Polic AND Artificial AND Intelligence AND Ethic**

	Combination of words:	**Data/Policy/Artificial/Intelligence/Ethic**	
Step:	*Query:*	*Meaning behind query:*	*Results*
1	TITLE-ABS-KEY (artificia	Search for combination of AI and Data pre/0 policy	16
2	TITLE-ABS-KEY (artificia	Exclude older articles	14
3	TITLE-ABS-KEY (artificia	Loosen search, change pre/0 from data policy and change it pre/5	266
4	TITLE-ABS-KEY (ethic*	Include Ethic*	13
5	TITLE-ABS-KEY (ethic*	Change Pre/5 into Pre/10	21
6	TITLE (artificial PRE/0 i	Remove Pre/10 and focus on combination: Data Policy Artificial Intelligence Ethic in TK	27

This step resulted in 227 articles. Table 21 shows all the yellow marked queries and how many results the query gave.

Table 21: Keywords and number of articles from the query

Keywords:	ID:	Number of articles from query:
Data Platform Policy	DPP	10
Data Policy Governance	DPGo	8
Data Governance	DG	42
Information Governance	IG	8
Data Management	DM	38
Data Policy Classification	DPC	2
Data Policy Merger	DPM	1
Data Policy Security	DPSe	20
Data Policy Storage	DPSt	16
Information Policy Storage	IPS	5
Data Policy Artifical Intelligence Ethic (1st)	DPAIE1	21
Data Policy Artifical Intelligence Ethic (2nd)	DPAIE2	27
Data Governance Artificial Intelligence Ethic	DGAIE	29
Total:		227

Other databases

To increase the validity of the literature review, the Web of Science database and Google scholar were used as verification that the articles that are found on Scopus, are somewhat even with those of other databases. This is done by using some of the interesting queries to see if they give the same number of articles, but also the same titles.

Web of Science

For Web of Science, the queries that are used for SCOPUS can be somewhat reused. The queries have to be transformed to match the search tools of the Web of Science. To search within TAK, TS (Topic) is used. To search for words that have to be in front of another word, NEAR/0 is used.

If we apply this knowledge to the query of Data Platform Policy (TITLE-ABS-KEY (data PRE/0 platform AND data PRE/0 policy)), this will make the following query: "TS= (data NEAR/0 platform AND data NEAR/0 policy)". This query gives 5 articles as a result, in comparison to the 10 articles Scopus gave. Four of the articles in Web of Science also showed up in the SCOPUS.

Two other comparisons were made on queries that gave a smaller number of results; this is because those are easier to compare. The query of Data Policy Artificial Intelligence Ethic* was used in Web of Science ((TS= (ethic* AND artificial NEAR/0 intelligence AND data NEAR/10 polic*))), which gave 17 results, compared to that of 21 in Scopus. Of the 17 results, 10 were the same as in Scopus.

The last query that was used in Web of Science is that of Data Policy Governance ((TI= (Governance AND Data NEAR/0 Polic*)) OR AK= (Governance AND Data NEAR/0 Polic*) OR KP = (Governance AND Data NEAR/0 Polic*)). This query gave 6 results, compared to the 8 results Scopus gave. All the articles on Web of Science are also in the results from SCOPUS.

Google Scholar

Google Scholar is different compared to Scopus and Web of Science. Google has an advanced search option, but it is not possible to search for keywords, only in the entirety of the article or the title. This makes the search different and will probably not give the same results that Scopus or Web of Science gave.

If we enter a similar query, in which we search in the entire article, we get this query: ("Data policy OR Data policies" AND "Data platform"). This results in 1,130 results, which are way more results than those of Scopus and Web of Science. This is because Google Scholar is a web search engine and not just a database, thus Scholar has more documents, not only articles. Scopus and Web of Science also have some results that are not articles, but specific book chapters for example. However, this is not close to what Google Scholar has. There is an option to only search for review articles, this limits the results to 39 articles. But none of the articles in Scopus show up in these 39 results.

The query for Data Policy Artificial Intelligence Ethic* in Google Scholar ("Data Policy" "Ethics" "Artificial Intelligence") also shows a lot of results, 1680. If the 'review articles' option is selected, only 98 results remain. None of the articles in Scopus shows up in these 98 results. (Check pages 1,2 and 8)

The last query, for Data Policy Governance, is: (allintitle: Governance "Data Policy"). As previously mentioned, Google Scholar does not have an option to search within Keywords, so this search only focuses on the title. This might result in fewer results since keywords are not included in this search. The query on Google Scholar gave 23 results, which is relatively close to the results from Scopus and Web of Science. However, only two articles from Scopus were in these 23 results. This makes sense since these are the only articles from Scopus which have Data policy AND Governance in the title. If the option "review articles" is selected, only one result is shown, which does not appear in the Scopus results.

Table 22: Results queries other databases

Keywords:	**Data Platform Policy**		
	SCOPUS	*Web of Science*	*Google Scholar*
Query:	TITLE-ABS-KEY (data PRE/0 platform AND data PRE/0 policy)	TS=(data NEAR/0 platform AND data NEAR/0 policy)	"Data policy OR Data policies" AND "Data platform"
Number of results:	10	5	1130 (39)
% in line with SCOPUS articles:	100%	80%	N.A.

Keywords:	**Data Policy Artificial Intelligence Ethic**		
	SCOPUS	*Web of Science*	*Google Scholar*
Query:	TITLE-ABS-KEY (ethic* AND artificial PRE/0 intelligence AND data PRE/10 polic*) AND PUBYEAR > 2011	(TS= (ethic* AND artificial NEAR/0 intelligence AND data NEAR/10 polic*))	"Data Policy" "Ethics" "Artificial Intelligence"
Number of results:	21	17	1680 (98)
% in line with SCOPUS articles:	100%	59%	N.A.

Keywords:	**Governance Data Policy**		
	SCOPUS	*Web of Science*	*Google Scholar*
Query:	TITLE (governance AND data PRE/0 policy) OR KEY (governance AND data PRE/0 policy)	(TI=(Governance AND Data NEAR/0 Polic*)) OR AK=(Governance AND Data NEAR/0 Polic*) OR KP =(Governance AND Data NEAR/0 Polic*)	allintitle: Governance "Data Policy"
Number of results:	8	6	23 (1)
% in line with SCOPUS articles:	100%	100%	9%

An overview of all the results can be seen in table 22. Regarding the validation, Web of Science validates the results from SCOPUS relatively well, but Google Scholar does not. This is probably connected to the difference in search queries and available documents between SCOPUS/Web of Science and Google Scholar. Because most articles between Scopus and Web of Science in this small exercise are similar, it is assumed that if a similar search in Web of Science was done, it would deliver similar results. This is not the case for Google Scholar, since it is so different compared to Web of Science and Scopus.

Step 3: Selection

This step is about selecting articles from the second step. This is based on the title of the articles.

Articles were chosen based on the following criteria:

- Articles are about the data policy, governance, management, or anything else related to the research.
- Articles can provide valuable information to the research. Not just an interesting article, but it is relevant to the research.

The search is done by entering the relevant queries again and going through the article titles. Relevant article titles were selected and put into a list, that will be used for the final search. Table 23 shows the number of relevant articles for each search query.

Table 23: Number of relevant articles for each query

Keywords:	ID:	Number of articles from query:	Articles that are relevant based on title:
Data Platform Policy	DPP	10	4
Data Policy Governance	DPGo	8	4
Data Governance	DG	42	9
Information Governance	IG	8	2
Data Management	DM	38	2
Data Policy Classification	DPC	2	1
Data Policy Merger	DPM	1	1
Data Policy Security	DPSe	20	7
Data Policy Storage	DPSt	16	5
Information Policy Storage	IPS	5	1
Data Policy Artifical Intelligence Ethic (1st)	DPAIE1	21	5
Data Policy Artifical Intelligence Ethic (2nd)	DPAIE2	27	10
Data Governance Artificial Intelligence Ethic	DGAIE	29	6
Total:		227	57
Number of double articles:			6
Total second search, excluding the doubles:			51

As can be seen from table 23, the number of articles was reduced from 227 to 57 articles. 6 articles were selected twice and when the duplicates were filtered out, 51 articles remained.

One exception was made. In the DPSe query, there was an article from 2013 called “Data policy”. This falls outside of the range that was set (10 or more times cited or from 2021 or 2022), but the title was deemed too relevant and was included in the list. The titles of all the selected articles can be seen in the next search.

The final step in selection is done by reading the abstracts of the articles, or when they are missing, parts of the article. The selected articles will be used in the formation of the overview of components needed for a data & AI policy. The list of the selected articles, followed by whether they will be chosen or not, can be seen in table 24.

Table 24: Decision articles

Nr:	ID:	Title:	Will be used:	Reason:
1	DPP	Usability evaluation of an open data platform	No	Not useful for overview but might give interesting advice regarding platform usability. Mentions several factors for good usability of a platform
2	DPP	Implementation of Digital-Era Governance: The Case of Open Data in U.S. Cities	No	Heavily focused on an open data platform and implementation, politics, and risks regarding this. Does not offer details regarding essential components of data policies
3	DPP	How Indonesia uses big data “Indonesian one data” for the future of policy making	Can't access	Might be interesting, but the article cannot be accessed
4	DPP	Open government data policy and Indian ecosystems	No	Mentions some interesting aspects, but the main focus is on implications of the NDSAP towards governments and politics
5	DPGo	Data Policy Definition and Verification for System of Systems Governance (Book Chapter)	Can't access	Its focus is on systems of systems, but it does mention the definition of data policy and the components it should have. However, it cannot be accessed
6	DPGo	Open government data policy & governance: Applicability of ecosystem approach. Comparative cross-country analysis	No	It is a summary of a panel and doesn't include any relevant insights for the research
7	DPGo	Digital strategies in action -A Comparative analysis of national data infrastructure development	Yes	It mentions NDI and its strategies, it might give insights into relevant components
8	DPGo /DG	Emerging models of data governance in the age of datafication	No	Don't think the data governance models will mention specifics on important components. Might be used in addition to something else
9	DG /IG	Data governance: A conceptual framework, structured review, and research agenda	Yes	Provides a framework that includes major building blocks of data governance. This could include relevant components needed for a data policy
10	DG	Data governance, data literacy, and the management of data quality	No	It has interesting aspects, but it focuses on the research data management of librarians, and this limits applicability. Might be used in addition
11	DG	Data governance activities: an analysis of the literature	Yes	Has an overview of data governance activities

12	DG	Exploring big data governance frameworks	Yes	Mentions of a framework for big data governance could be used as the basis for what a policy requires
13	DG	Data Governance Framework for Big Data Implementation with a Case of Korea	No	Focuses on data quality and not on data policy and what it should contain
14	DG /DM	Changes in roles, responsibilities, and ownership in organizing master data management	Yes	Discusses how roles, responsibilities, and ownership should fit into data management, governance, and policies
15	DG	Data governance activities: a comparison between scientific and practice-oriented literature	Can't access	Might be similar to the 11th article in this list, but this article also mentions practical components. However, it cannot be accessed
16	DG	Constructing a mutually supportive interface between ethics and regulation	Yes	Argues technology needs an ethical framework and gives a framework. Also discusses AI ethics and the framework
17	IG	The role of information governance in big data analytics-driven innovation	No	Discusses the importance of information governance, while interesting and somewhat relevant, it does not fit into the overview
18	DM	A survey of big data management: Taxonomy and state-of-the-art	Yes	Gives an overview of data management techniques, divided into data storage, pre-processing, processing, and data security.
19	DPC	A conceptual data quality framework for IPCHEM – The European Commission Information Platform for chemical monitoring	No	Is focused on the methodology of an IPCHEM architecture, not relevant
20	DPM	Corporate data quality management in context	No	The abstract seems to be about the motivation for data quality management and other factors, not about guidelines or something regarding this
21	DPSe	Big data, open government, and e-government: Issues, policies, and recommendations	No	Does provide an overview of important categories within Policies? But not much substance. Check the article, conclusions are too general and don't have real guidelines within them.
22	DPSe	Reconciling contradictions of open data regarding transparency, privacy, security, and trust	No	Discusses interesting aspects, but those are based on public value, not necessary components. Also, very focused on open data, which limits applicability. Might use this as a public value of what is important?

23	DPSe /SPSt	Research data management: a conceptual framework	Can't access	Could be interesting, since it provides a conceptual framework of something that resembles a data policy framework somewhat, but it cannot be accessed
24	DPSe	Achieving privacy and security in multi-owner data outsourcing	No	Is about data outsourcing and it might not discuss key components of a policy. Presents a framework but is about access control for protecting outsourced data. Not relevant
25	DPSe	Impact of open data policies on consent to participate in human subject's research: Discrepancies between participant action and reported concerns	No	Is more about the human subjects and how it influences them than it is about data policies
26	DPSe	Control Design of Information Security Related to Privacy in the Smart SIM Business Process	No	Might provide several security standards, but it is way too specific? Limits applicability, might be used as an addition
27	DPSe	Data policy	Yes	An article that talks about the definition of data policy, what it needs, and other relevant topics
28	SPSt	The development of a research data policy at Wageningen university & research: Best practices as a framework	Yes	Shows several criteria from existing frameworks, principles, and best practices
29	SPSt	Curation and policy issues in collaborative research data management communities: Perspectives from key stakeholders	No	It talks about stakeholders solely and it is a panel, so it does not give enough information
30	SPSt	Data governance for SoS	Can't access	It might have some interesting aspects, but can't access the full document
31	SPSt	Developing a research data policy framework for all journals and publishers: An output of the data policy standardization and implementation interest group (IG) of the research data alliance (RDA)	Yes	Provides an overview of what policies should have. It is about research data policy framework, but it might have some interesting components for private data policies
32	IPS	An information policy perspective on learning analytics	No	Abstract didn't give enough information when reading the article: interesting topics around how learning analytics influence laws and regulations, but it does not provide steps or

				guidelines. Might use this in addition to the AI subchapter.
33	DPAIE1	Ethics and Privacy in AI and Big Data: Implementing Responsible Research and Innovation	Yes	The article might provide several interesting components, such as privacy by design and the right to be forgotten
34	DPAIE1	AI ethics in predictive policing: From models of threat to an ethics of care	No	Misunderstood, this article is about AI Ethics regarding the Police, not policies
35	DPAIE1	The data firehose and AI in government: Why data management is a key to value and ethics	No	It discusses data and AI in governments, but does not talk about policy guidelines on AI or Ethical guidelines
36	DPAIE1	The ethical switch: An automated policy and regulation framework for data dissemination in cloud environments	No	It provides information on challenges regarding policies on ethics but does not go in-depth about guidelines or something
37	DPAIE1	Artificial Intelligence in Healthcare from a Policy Perspective	Yes	Provides a list of challenges and opportunities for AI (in healthcare, but can be applied to other industries)
38	DPAIE2 /DGAIE	AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations	Yes	Interesting take on AI, its ethics, and how policies can help
39	DPAIE2	Data, privacy, and the greater good	No	It discusses data, privacy, and ethics. However, it might not be relevant to the research, since it does not provide some sort of guidelines or framework
40	DPAIE2	Artificial Intelligence in Health Care: A Report from the National Academy of Medicine	No	Mentions several interesting aspects and the opportunities of AI in health care, does not go over guidelines or some sort of framework
41	DPAIE2	Artificial intelligence (AI) Ethics: Ethics of AI and ethical AI	Yes	Mentions that the article will explain ethics rules, regulations, and policies regarding AI
42	DPAIE2	AI in healthcare: Ethical and privacy challenges	No	Focuses too much on healthcare
43	DPAIE2	Ensuring trustworthy use of artificial intelligence and big data analytics in health insurance	No	Abstract focuses on health insurance data, which might be interesting, but not sure whether this will be relevant. Use as additional info, not in the overview
44	DPAIE2	Trustworthy Artificial Intelligence in Medical Imaging	No	Not for the overview, might be complementary for the AI part

45	DPAIE2 /DGAIE	The 'Ethification' of ICT Governance. Artificial Intelligence and Data Protection in the European Union	No	It speaks in a very general manner about ethics and does not mention ethics regarding data or data policies
46	DPAIE2	AI Ethics: Algorithmic Determinism or Self-Determination? The GPDR Approach	No	However, it does contain some interesting subjects that could be used as additional information outside the overview
47	DPAIE2	Ethics of Artificial Intelligence: Research Challenges and Potential Solutions	No	Not for the overview, but has several solutions to ethical issues of AI
48	DGAIE	From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods, and Research to Translate Principles into Practices	Yes	Contains interesting findings regarding the ethical implications of AI and provides interesting future scenarios
49	DGAIE	A high-level overview of AI ethics	No	A very general high-level overview is interesting but does not discuss applications for organizations or important guidelines
50	DGAIE	We Haven't Gone Paperless Yet: Why the Printing Press Can Help Us Understand Data and AI	No	Interesting paper, but not relevant, since it goes in-depth about understanding data and AI and how it influences communities and people. Might have been interesting in a broader context that did not focus on Data & AI policies
51	DGAIE	Creating ethics guidelines for artificial intelligence and big data analytics customers: The case of the consumer European insurance market	Yes	Is focused on all the right areas and the abstract presents relevant information to the research

Table 25: Totals from table 24

The number of yes articles:	16
The number of no articles:	30
The number of articles that cannot be accessed:	5
Total:	51

As can be seen in table 25, this selection step leads to 16 articles that have been chosen.

The literature review also suggests doing a backward citation search. In this backward citation search, the references of the chosen articles have been checked for relevance based on their title (wolfs Winkel). If the title was deemed relevant, the abstract was analyzed, and this will decide if the article will be included in the literature review. Table 26 shows the articles that were relevant according to their title. The color shows the decision that has been made regarding their relevance. This search has led to two additional articles, which are the two green-marked articles in table 26. The backward citation method has also been applied to these two articles. Table 27 shows an overview of the article titles from the backward citation search.

Table 26: Backward citation overview

Year	Title	Interesting article in references:				
2017	Digital strategies in action -A Comparative analysis of national data infrastructure development	https://ww	-			
2019	Data governance: A conceptual framework, structured review, and research agenda	https://ww	https://w	https://w	https://w	https://w-
2016	Data governance activities: an analysis of the literature	https://ww	https://w-			
2018	Exploring big data governance frameworks	https://ww	https://w	https://w	https://w	-
2019	Changes in roles, responsibilities and ownership in organizing master data management	https://ww	https://w-			
2021	Constructing a mutually supportive interface between ethics and regulation	https://ww	https://w	https://w	-	
2016	A survey of big data management: Taxonomy and state-of-the-art	https://ww	https://w	https://w	-	
2020	Developing a research data policy framework for all journals and publishers: An output of the data policy standardisation and implementation interest group (IG) of the research data alliance (RDA)	https://ww	-			
2013	Data policy	https://ww	-			
2017	The development of a research data policy at wageningen university & research: Best practices as a framework	https://ww	-			
2018	Ethics and Privacy in AI and Big Data: Implementing Responsible Research and Innovation	https://ww	https://w	https://w	-	
2021	Artificial Intelligence in Healthcare from a Policy Perspective	https://ww	-			
2018	AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations	-				
2020	Artificial intelligence (AI) Ethics: Ethics of AI and ethical AI	https://ww	https://w	https://w	-	
2020	From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices	https://ww	https://w	https://w	https://w	-
2021	Creating ethics guidelines for artificial intelligence and big data analytics customers: The case of the consumer European insurance market	https://ww	https://w-			
2020	The Ethics of AI Ethics: An Evaluation of Guidelines	https://ww	https://w	https://w	https://w	-
2019	Establishing the rules for building trustworthy AI	https://ww	-			

Title seemed interesting, but abstract did not
Interesting but not within scope. Can be used as reference further in the research
Article was one of the already chosen articles
Interesting article, will be added
Can't be accessed
Relatively old source

Table 27: Overview titles of articles backward citation

Year	Title	Possibly interesting articles:
2017	Digital strategies in action -A Comparative analysis of national data infrastructure development	A comparison of national open data policies: Lessons learned
2019	Data governance: A conceptual framework, structured review, and research agenda	Exploring big data governance frameworks
		Exploring big data governance frameworks

		Data governance activities: a comparison between scientific and practice-oriented literature
		Critical Success Factors for Data Governance: A Theory Building Approach
		The need for data governance: A case study
2016	Data governance activities: an analysis of the literature	Information governance in NHS's NPfIT: A case for policy specification
		Organizing Data Governance: Findings from the telecommunications industry and consequences for large service providers
2018	Exploring big data governance frameworks	Corporate governance of big data: Perspectives on value, risk, and cost
		Data Governance Framework for Big Data Implementation with a Case of Korea
		Designing data governance
		Data governance activities: an analysis of the literature
2019	Changes in roles, responsibilities and ownership in organizing master data management	Designing data governance
		Organizing Data Governance: Findings from the telecommunications industry and consequences for large service providers
2021	Constructing a mutually supportive interface between ethics and regulation	Principles and business processes for responsible AI
		The Ethics of AI Ethics: An Evaluation of Guidelines
		Ethics and Privacy in AI and Big Data: Implementing Responsible Research and Innovation
2016	A survey of big data management: Taxonomy and state-of-the-art	Considerations for big data: Architecture and approach
		Designing data governance
		Challenges and opportunities with big data
2020	Developing a research data policy framework for all journals and publishers: An output of the data policy standardization and implementation interest group (IG) of the research data alliance (RDA)	FAIRsharing as a community approach to standards, repositories, and policies
2013	Data policy	Professional ethics and practice in archives and records management in a human rights context
2017	The development of a research data policy at Wageningen university & research: Best practices as a framework	Comment: The FAIR Guiding Principles for scientific data management and stewardship
2018	Ethics and Privacy in AI and Big Data: Implementing Responsible Research and Innovation	The dark secret at the heart of AI

		Towards an "Ethics by Design" Methodology for AI Research Projects
		Regulating big data. The guidelines of the Council of Europe in the context of the European data protection framework
2021	Artificial Intelligence in Healthcare from a Policy Perspective	Integrating artificial intelligence into health care through data access: Can the GDPR act as a beacon for policymakers?
2018	AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations	-
2020	Artificial intelligence (AI) Ethics: Ethics of AI and ethical AI	Building trust in artificial intelligence, machine learning, and robotics
		Ethical and Moral Issues with AI - A Case Study on Healthcare Robots
		Artificial intelligence, machine learning, automation, robotics, future of work and future of humanity: A review and research agenda
2020	From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods, and Research to Translate Principles into Practices	Principles and business processes for responsible AI
		Governing artificial intelligence: ethical, legal, and technical opportunities and challenges
		Establishing the rules for building trustworthy AI
		AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations
2021	Creating ethics guidelines for artificial intelligence and big data analytics customers: The case of the consumer European insurance market	Artificial intelligence governance principles: towards ethical and trustworthy artificial intelligence in the European insurance sector
		AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations
2020	The Ethics of AI Ethics: An Evaluation of Guidelines	Incorporating Ethics into Artificial Intelligence
		AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations
		Why teaching ethics to AI practitioners is important
		Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions
2019	Establishing the rules for building trustworthy AI	AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations

Table 28 displays the final list of articles and their reason for selection. These articles will be coded.

Table 28: Final articles that have been chosen for the literature review

Year	Article nr.	Author	Title	Reason for its selection
2013	1	Parsons M.A.	Data policy	Talks about the definition of data policy and the objectives of a data policy. Can give several important components for the overview
2016	2	Alhassan I. & Sammon D. & Daly M.	Data governance activities: an analysis of the literature	Has an overview of data governance activities, might also include relevant policy-related components
2016	3	Siddiqa, A., Hashem, I. A. T., Yaqoob, I., Marjani, M., Shamshirband, S., Gani, A., & Nasaruddin, F.	A survey of big data management: Taxonomy and state-of-the-art	Gives an overview of data management techniques, divided into data storage, pre-processing, processing, and data security. These could serve as components as part of a data policy
2017	4	Klievink, B., Neuroni, A., Fraefel, M., & Zuiderwijk, A.	Digital strategies in action -A Comparative analysis of national data infrastructure development	Mentions components of data policy and strategy. These could be used for the overview
2017	5	van Zeeland, H., & Ringersma, J.	The development of a research data policy at Wageningen university & research: best practices as a framework	Shows several criteria from existing frameworks, principles, and best practices, which can be used for the overview
2018	6	Al-Badi, A., Tarhini, A., & Khan, A. I.	Exploring big data governance frameworks	Mentions of a framework for big data governance could be used to gain relevant components for a data policy
2018	7	Stahl, B. C., & Wright, D.	Ethics and Privacy in AI and Big Data: Implementing Responsible Research and Innovation	The article might provide several interesting components regarding what an AI policy needs. This could be used to back up the AI portion of the data policy overview
2018	8	Floridi et al.	AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks,	Provides several ethical principles for AI, which could form the base for

			Principles, and Recommendations	the AI section of the overview
2019	9	Abraham, R., Schneider, J., & vom Brocke, J.	Data governance: A conceptual framework, structured review, and research agenda	Provides a framework that includes major building blocks of data governance. This could include relevant components needed for a data policy
2019	10	Vilminko-Heikkinen, R., & Pekkola, S.	Changes in roles, responsibilities, and ownership in organizing master data management	This could give background info regarding roles, responsibilities, and ownership and how they should fit into data management, governance, and policies
2019	11	Floridi	Establishing the rules for building trustworthy AI	Provides several components needed for trustworthy AI
2020	12	Hrynaszkiewicz, I., Simons, N., Hussain, A., Grant, R., & Goudie, S.	Developing a research data policy framework for all journals and publishers: An output of the data policy standardization and implementation interest group (IG) of the research data alliance (RDA)	Provides an overview of what policies should have. It is about research data policy framework, but it might have some interesting components for private data policies
2020	13	Siau, K., & Wang, W.	Artificial intelligence (AI) Ethics: Ethics of AI and ethical AI	Can give insights into risks and solutions regarding ethics and AI, these could contain components
2020	14	Morley et al.	From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods, and Research to Translate Principles into Practices	Contains interesting findings regarding the ethical implications of AI and these could help form the AI components
2020	15	Hagendorff	The Ethics of AI Ethics: An Evaluation of Guidelines	Can provide several guidelines for Ethics within AI
2021	16	Delacroix, S., & Wagner, B.	Constructing a mutually supportive interface between ethics and regulation	Gives AI and ethical framework could use several components from this regarding AI and ethical components in the data policy

2021	17	Aggarwal, M., Gingras, C., & Deber, R.	Artificial Intelligence in Healthcare from a Policy Perspective	Provides a list of challenges and opportunities for AI (in healthcare but might also apply to other industries). These challenges and opportunities could have relevant policy components
2021	18	Mullins et al.	Creating ethics guidelines for artificial intelligence and big data analytics customers: The case of the consumer European insurance market	Is focused on all the right areas and the abstract presents relevant information to the research. The article could provide interesting components for the overview, which are specific to the insurance industry

Step 4: Analyze

The action within this step is open coding, which adds codes to specific parts of text within the chosen articles. Table 29 has an overview of all the codes that have been used in each article. It also already has the codes or the second action within this step, Axial coding. Axial codes are overarching categories and relations between the different codes.

Table 29: Overview of open and axial codes

Article	Open coding	Axial coding
1	**Data policy is context dependent**	**General policy information**
	Data access requirements	**Data security**
	Data preservation and stewardship requirements	**Roles & Responsibilities**
	Standards and compliance mechanisms	**General policy guidelines**
	Data security	**Data security**
	Privacy and ethical concerns	**Ethics**
	Specific collection protocols and defined data flows	**Data lifecycle**
2	**Define domains**	**Roles & Responsibilities**
	Data lifecycle	**Data lifecycle**
	Production, retention, and retirement of data assets	**Data lifecycle**
	Define roles and responsibilities	**Roles & Responsibilities**
	Define, implement, and monitor roles and responsibilities	**Roles & Responsibilities**
	Roles and responsibilities actions	**Roles & Responsibilities**
3	**Data storage is fundamental**	**Data lifecycle**
	Security is big concern	**Data security**
	Security; privacy	**Data security**
	Security; integrity	**Data security**
	Security; confidentiality	**Data security**
	Security; availability	**Data security**

	Security measure; Data checking	Data security
	Type of data stored	Data lifecycle
	Data storage technique; clustering	Data lifecycle
	Data storage technique; replication	Data lifecycle
	Data storage; technique; indexing	Data lifecycle
	Security; integrity major challenge	Data security
	Different cleaning methods -> high quality	Data quality
	Increase in difficulty of securing data	Data security
4	Using strategy documents as policy instrument	General policy information
5	Data retention	Data lifecycle
	Data ownership	Roles & Responsibilities
	Responsibilities stakeholders	Roles & Responsibilities
	Use best-practices, increase credibility guidelines	General policy information
	Integrate policy into behavior and culture	General policy information
	Generalize guidelines	General policy information
6	Stewardship	Roles & Responsibilities
	Data definition and usage standards	General policy guidelines
	Metadata management	Metadata
	Data lifecycle management	Data lifecycle
	Risks	Liability
	Relevant stakeholders	General policy guidelines
	Scope of (big) data	General policy guidelines
	Storage of data	Data lifecycle
	Communication and management of data	Data lifecycle
	Data capture, management, and consumption	Data lifecycle
	Privacy	Data security
	Security	Data security
	Risks	Liability
	Retention	Data lifecycle
	Regulatory compliance	General policy guidelines
	Data classification requirements	Data security
7 (AI)	Privacy	Privacy (AI)
	Data protection	Data security
	Data protection	Data security
	Privacy	Privacy (AI)
	Education of ethics and security	Education and training
	Accountability	Explicability
	Using professionalism	National organizations
8 (AI)	Beneficence	Beneficence
	Non-maleficence	Non-maleficence
	Autonomy	Autonomy
	Justice	Justice
	Applicability (Intelligibility (Explainability) and accountability (transparency))	Explicability
	AI; Beneficence	Beneficence
	AI; Non-maleficence	Non-maleficence
	AI; Autonomy	Autonomy
	AI; Justice	Justice
	AI: Applicability, consists of Intelligibility (Explainability) and accountability (transparency)	Explicability

	Auditing mechanism AI	**Reinforcement**
	Implement procedure when harm is done AI	**Reinforcement**
	Develop metrics trustworthiness AI	**Explicability**
	Code of conduct Data and AI	**Explicability**
	Education of impact AI	**Education and training**
	Continuous change AI policy	**General AI**
9	**Policies should be in line with strategy**	**General policy information**
	Data quality	**Data quality**
	Data security	**Data security**
	Data architecture	**Design and strategy**
	Data lifecycle	**Data lifecycle**
	Metadata	**Metadata**
	Data storage and infrastructure	**Data lifecycle**
	Data creation and acquisition	**Data lifecycle**
	Data storage	**Data lifecycle**
	Data security	**Data security**
	Data quality	**Data quality**
	Permissible use of data	**Data lifecycle**
	Communicate objectives	**General policy guidelines**
	Communicate accountabilities	**Explicability**
	Communicate roles and responsibilities	**Roles & Responsibilities**
	Communicate retention periods	**Data lifecycle**
	Policy training stakeholders	**Education and training**
	Data quality	**Data quality**
	Data security	**Data security**
	Data architecture	**Design and strategy**
	Data lifecycle	**Data lifecycle**
	Metadata	**Metadata**
	Data storage and infrastructure	**Data lifecycle**
	Ethical concerns	**Ethics**
	Improvement of culture from training	**Education and training**
	Define and audit data security procedures	**Data security**
	Accountability data	**Explicability**
	Limit restrictions policy	**General policy information**
	Industry context is important	**General policy information**
10	**Define responsibilities**	**Roles & Responsibilities**
	Define ownership data	**Roles & Responsibilities**
	Define responsibilities	**Roles & Responsibilities**
	Define data owner	**Roles & Responsibilities**
	Define roles and responsibilities	**Roles & Responsibilities**
	Understand definition data owner	**Roles & Responsibilities**
11 (AI)	**AI; lawful**	**General AI**
	AI; Ethical	**Ethics**
	AI; robust	**General AI**
	Human agency and oversight	**Autonomy**
	Robustness and safety	**General AI**
	Privacy and data governance	**Privacy (AI)**
	Transparency	**Explicability**
	Diversity, non-discrimination, and fairness	**Fairness**
	Societal and environmental well-being	**General AI**

	Accountability	**Explicability**
12	**Define type of data**	**General policy guidelines**
	Define not covered data	**General policy guidelines**
	Define data format	**Metadata**
13 (AI)	**Achieve trust through transparency**	**Explicability**
	Security	**Data security**
	Privacy	**Privacy (AI)**
	Accountability AI	**Explicability**
	Define ethical standards	**Ethics**
	Educate engineers regarding privacy	**Education and training**
	Define ethics	**Ethics**
14 (AI)	**Beneficence**	**Beneficence**
	Non-maleficence	**Non-maleficence**
	Autonomy	**Autonomy**
	Justice	**Justice**
	Explicability (Applicability)	**Explicability**
	Explicability -> more transparent and accountable	**Explicability**
	Implement principles already in design	**Design and strategy**
	Need for AI ethics	**Ethics**
15 (AI)	**Accountability**	**Explicability**
	Privacy	**Privacy (AI)**
	Fairness	**Fairness**
	Importance of diversity in AI guidelines	**Fairness**
	Impact of AI on employees is big	**Employee**
	More detailed technical explanation to guidelines	**Technical guidelines**
	Reinforcement mechanisms	**Reinforcement**
	AI should focus on technological details	**Technical guidelines**
	AI can assist in removing bad routines	**Education and training**
	Reinforcement mechanisms	**Reinforcement**
	Impact of guidelines on human decision-making is low	**Education and training**
	Impact of guidelines on behavior is low	**Education and training**
	Non-empowerment of employees in raising ethical concerns	**Employee**
16 (AI)	**Transparency**	**Explicability**
	Accountability	**Explicability**
	Connect consequences to infractions	**Reinforcement**
	Assistance of national professional bodies	**National organizations**
	Assistance of national professional bodies	**National organizations**
	Connect consequences to infractions	**Reinforcement**
	Unclear ethical practices	**Ethics**
	Working together with local organizations	**National organizations**
17 (AI)	**Address consequences of AI innovation**	**Reinforcement**
	Establish self-regulatory and oversight mechanisms	**Reinforcement**
	Unclear responsibilities	**Roles & Responsibilities**
	Accuracy	**Explicability**
	Fairness	**Fairness**
	Transparency	**Explicability**
	Data privacy and consent	**Privacy (AI)**
	Accountability	**Explicability**

	Workplace disruption	**Employee**
	Lack of transparency	**Explicability**
	Describe collection and usage of data	**Data lifecycle**
	Liability determination	**Liability**
	Accountability determination	**Explicability**
	AI is different in each organization	**General AI**
	Algorithm discrimination	**Fairness**
	More explainability -> less accuracy	**Explicability**
18 (AI)	**Transparency**	**Explicability**
	Transparency in some degree is needed for explainability and trust	**Explicability**
	Fairness	**Fairness**

The Axial codes can then also be put together to form dimensions, called aggregate dimensions. Table 30 shows these codes. This final step is the final refinement of the codes, also called selective coding.

Table 30: Overview of open codes, axial codes, and aggregate dimensions

Open codes	Axial codes	Aggregate dimension
Data security	Data security	Data policy components
Security is big concern		
Security; privacy		
Security; integrity		
Security; confidentiality		
Security; availability		
Security measure; Data checking		
Security; integrity major challenge		
Increase in difficulty of securing data		
Security		
Data security		
Data security		
Data security		
Security		
Data protection		
Data protection		
Privacy		
Define and audit data security procedures		
Data classification requirements		
Data access requirements		
Data lifecycle	Data lifecycle	
Permissible use of data		
Production, retention, and retirement of data assets		

Data storage is fundamental		
Type of data stored		
Data storage technique; clustering		
Data storage technique; replication		
Data storage; technique; indexing		
Data retention		
Data lifecycle management		
Storage of data		
Retention		
Data storage and infrastructure		
Data creation and acquisition		
Communicate retention periods		
Data storage		
Data lifecycle		
Data storage and infrastructure		
Describe collection and usage of data		
Data capture, management, and consumption		
Communication and management of data		
Data lifecycle		
Specific collection protocols and defined data flows		
Define roles and responsibilities	Roles & Responsibilities	
Define, implement, and monitor roles and responsibilities		
Roles and responsibilities actions		
Responsibilities stakeholders		
Stewardship		
Unclear responsibilities		
Data ownership		
Data preservation and stewardship requirements		
Define responsibilities		
Define ownership data		
Define responsibilities		
Define data owner		
Communicate roles and responsibilities		
Define roles and responsibilities		
Define domains		
Understand definition data owner		
Data quality		

Data quality	Data quality	
Data quality		
Different cleaning methods -> high quality		
Metadata management	Metadata	
Metadata		
Metadata		
Define data format		
Accountability	Explicability	AI-policy components
Explicability (Intelligibility (Explain ability) and accountability (transparency))		
AI: Explicability, consists of Intelligibility (Explain ability) and accountability (transparency)		
Accountability data		
Transparency		
Accountability		
Achieve trust through transparency		
Accountability AI		
Explicability		
Explicability -> more transparent and accountable		
Accountability		
Transparency		
Accountability		
Transparency		
Accountability		
Lack of transparency		
Accountability determination		
More explain ability -> less accuracy		
Transparency		
Communicate accountabilities		
Develop metrics trustworthiness AI		
Code of conduct Data and AI		
Transparency in some degree is needed for explainability and trust		
Accuracy (AI)		
Beneficence	Beneficence	
AI; Beneficence		
Beneficence		
Non-maleficence		
AI; Non-maleficence		

Non-maleficence	Non-maleficence	
Autonomy	Autonomy	
AI; Autonomy		
Autonomy		
Human agency and oversight		
Justice	Justice	
AI; Justice		
Justice		
Diversity, non-discrimination, and fairness	Fairness	
Fairness		
Importance of diversity in AI guidelines		
Fairness		
Algorithm discrimination		
Fairness		
Data privacy and consent	Privacy (AI)	
Privacy		
Privacy		
Privacy		
Privacy		
Privacy and data governance		
Continuous change AI policy	General AI	
AI; lawful		
AI; robust		
Robustness and safety		
Societal and environmental well-being		
AI is different in each organization		
More detailed technical explanation to guidelines	Technical guidelines	
AI should focus on technological details		
Scope of (big) data	General policy guidelines	General policy components
Define type of data		
Data definition and usage standards		
Define not covered data		
Relevant stakeholders		
Standards and compliance mechanisms		
Regulatory compliance		
Communicate objectives		

Liability determination	Liability
Risks	
Risks	
Using professionalism	National organizations
Assistance of national professional bodies	
Assistance of national professional bodies	
Working together with local organizations	
Workplace disruption	Employee
Impact of AI on employees is big	
Non-empowerment of employees in raising ethical concerns	
Privacy and ethical concerns	Ethics
Ethical concerns	
Need for AI ethics	
Define ethical standards	
Define ethics	
AI; Ethical	
Unclear ethical practices	
Education of ethics and security	Education and training
Education of impact AI	
Policy training stakeholders	
Improvement of culture from training	
Impact of guidelines on human decision-making is low	
Impact of guidelines on behavior is low	
AI can assist in removing bad routines	
Educate engineers regarding privacy	
Reinforcement mechanisms	Reinforcement
Reinforcement mechanisms	
Connect consequences to infractions	
Connect consequences to infractions	
Address consequences of AI innovation	
Implement procedure when harm is done AI	
Auditing mechanism AI	
Establish self-regulatory and oversight mechanisms	
Data policy is context dependent	
Industry context is important	
Generalize guidelines	

Use best-practices, increase credibility guidelines	General policy information	
Limit restrictions policy		
Integrate policy into behavior and culture		
Implement principles already in design	Design and strategy	
Using strategy documents as policy instrument		
Policies should be in line with strategy		
Data architecture		
Data architecture		

Step 5: Present

Table 31 shows the three main categories (aggregate dimensions) and their components.

Table 31: Overview of three main categories and their components

Data Policy components	**Data security** **Data lifecycle** **Roles & responsibilities** **Data quality** **Metadata**
AI-Policy components	**Explicability** **Beneficence** **Non-Maleficence** **Autonomy** **Justice** **Fairness** **Privacy (AI)** **General AI** **Technical guidelines**
General policy components	**General policy guidelines** **Liability** **National organizations** **Employee** **Ethics** **Education and training** **Reinforcement** **General policy information** **Design and strategy**

However, after the transformation process, several components have been combined and form new components together. Table 32 shows the dimensions, their components, and sub-components.

Table 32: In-depth overview of dimensions, components, and their sub-components

Dimensions	*Component:*	*Sub-component (code):*
Data policy components:	Data security	Data security Data privacy Data integrity Data confidentiality Data availability Data classification Data access
	Data lifecycle	Data lifecycle Data storage Data production/creation Data collection/acquisition Data retention Data usage (description) Data quality

		Metadata
	Roles & Responsibilities	Data Roles & Responsibilities Data ownership
AI policy components:	Explicability	Explainability Explicability Accountability Transparency
	AI-Principles	Beneficence Non-maleficence Autonomy Justice Fairness
	General AI	AI depends on the organization Continuous change
	Technical guidelines	Technical details Connecting principles with practice
	Privacy (AI)	Privacy Communication of privacy within AI usage Consent of usage data
General policy components	General guidelines	Goal & Scope Definitions Stakeholders Compliance/audit
	Liability	Risk Liability
	Design & Strategy	Strategy Design Architecture
	Employee	Communication to employees Education & Training Reinforcement of policy
	Ethics	Ethical standards National organizations
	General Policy information	Context dependent Limit restrictions policy Integration policy into culture and behavior Best practices

Appendix 2: Protocol questions:

Table 33 shows the protocol questions, with the codes and their likely source. The codes are based on table 32.

Table 33: Protocol questions and their likely source

Question:	Codes	Likely source:
What information regarding data security is important within data policies? What priority would this component have?	Data security Data privacy Data integrity Data confidentiality Data availability Data classification Data access	Alpina incidentenmanagement Heilbron Beveiliging en applicaties Heilbron Continuïteitsbeheer Heilbron Dreiging en incidentenbeheer Heilbron Informatiebeveiligingsbeleid V&V Beveiligingsbeleid V&V Continuïteitsbeheer V&V Dreiging en incidentenbeheer V&V Informatiebeleid V&V Technische beveiliging V&V Toegangsbeheer
What information regarding data lifecycle is important within data policies? What priority would this component have?	Data lifecycle Data storage Data production/creation Data collection/acquisition Data retention Data usage (description) Data quality Metadata	Heilbron Bewaartermijn Heilbron Systeembeheer V&V Applicatiebeheer V&V Netwerk en communicatie V&V Systeembeheer V&V Systeemontwikkeling
What information regarding Roles & Responsibilities is important within data policies? What priority would this component have?	Data Roles & Responsibilities Data ownership	Alpina Rolbeschrijvingen, V&V Toegangsbeheer
What information regarding AI-Principles is important within AI policies? What priority would this component have?	Beneficence Non-maleficence Autonomy Justice Fairness	V&V AI-beleid
What information regarding explicability is important within AI policies? What priority would this component have?	Explainability Explicability Accountability Transparency	V&V AI-beleid
What information regarding the Privacy of AI is important within AI policies? What priority would this component have?	Consent of usage data Communication of privacy within AI usage Privacy	Heilbron Privacy-beleid, V&V Privacy-beleid V&V AI-beleid

What information regarding technical guidelines is important within AI policies? What priority would this component have?	Technical details Connecting principles with practice	V&V AI-bellied
What information regarding General AI is important within AI policies? What priority would this component have?	AI depends on the organization Continuous change	V&V AI-bellied
What information regarding employees is important within policies? What priority would this component have?	Communication to employees Education & Training Reinforcement of policy	All policies
What information regarding general guidelines is important within policies? What priority would this component have?	Goal & Scope Definitions Stakeholders Compliance/audit	All policies
What information regarding ethics is important within policies? What priority would this component have?	Ethical standards National organizations	Alpina Ethisch kader
What information regarding liability is important within policies? What priority would this component have?	Risk Liability	All policies
What information regarding Design & Strategy is important within policies? What priority would this component have?	Strategy Design Architecture	All policies
What information regarding General policy information is important within policies? What priority would this component have?	Context dependent Limit restrictions policy Integration policy into culture and behavior Best practices	All policies
What information is presented in the policies that are not in the literature review results? What would be the priority of these components?	-	All policies

Appendix 3: Interview framework

The interview framework is used during the interview and contains a general set of guidelines that will be used during the interview. The framework is as follows:

- General introduction
    - Discus the goal of the interview and research. It also features a small overview of what can be expected.
- The main part of the interview
    - Q1:
        - What components should Data policies have?
        - What priority, based on the MoSCoW model, would you assign to each component. (MoSCoW model will be briefly explained if the interviewee is not familiar with the model)
    - Q2:
        - What components should AI policies have?
        - What priority, based on the MoSCoW model, would you assign to each component. (MoSCoW model will be briefly explained if the interviewee is not familiar with the model)
    - Q3:
        - What components should General policies have?
        - What priority, based on the MoSCoW model, would you assign to each component. (MoSCoW model will be briefly explained if the interviewee is not familiar with the model)
    - The list of components and priorities of each component will be shown to the interviewee. Their previously mentioned components and priorities will be compared to the list based on the literature and practice.
    - Ask input regarding the literature and practice-based list of components and priorities.
- Closing statements
    - Ask if there are questions.
    - Thank the interviewee for the interview and input

Appendix 4: Policy analysis in-depth steps and explanation:

The analysis was done based on a selection of policy documents. This selection was based on documents sent by the internal employee. They had sent policies regarding data & AI via the mail. The overview of all the documents can be seen in table 34. This is divided into two categories, main policies, and supporting policies. Main policies are documents that contain the main policies while supporting documents are documents that support the main policies. These can be overviews, extra explanations, or appendices.

Table 34: Overview of available documents from policy analysis

Main policies	**Supporting documents**
Alpina Ethisch-kader 2020	Alpina Ethisch kader toelichting
Alpina Incidentenmanagement 2021	Alpina uitbestedingsbeleid flowchart 2021
Alpina Rolbeschrijvingen	Alpina uitbestedingsbeleid 2021 bijlagen
Alpina uitbestedingsbeleid 2021	Alpina uitbestedingsbeleid flowchart 2021
Heilbron beveiliging en applicaties	V&V AI-beleid advies 2021
Heilbron bewaartermijn 2020	V&V Bewaartermijn overzicht
Heilbron Continuïteitsbeheer 2017 (Extern)	V&V Privacy verklaring 2021
Heilbron Dreiging en incidentenbeheer	Zzz overig AI-beleid verzekeringssector
Heilbron informatiebeveiligingsbeleid 2020	Zzz overig checklist KOAT
Heilbron Privacy-beleid	Zzz overig introductie checklist OT
Heilbron Systeembeheer (Extern)	
V&V AI-beleid 2021	
V&V Applicatiebeheer	
V&V Beveiligingsbeleid 2021	
V&V Bewaartermijn	
V&V Continuïteitsbeheer	
V&V Data classificatie	
V&V Dreiging en incidentenbeheer	
V&V Informatiebeleid 2019	
V&V Netwerk en communicatie	
V&V Privacy-beleid 2021	
V&V Systeembeheer	
V&V Systeemontwikkeling	
V&V Technische beveiliging	
V&V Toegangsbeheer	

Each document within the main policy section was read and coded for relevant sections. The codes can be seen in table 35.

Table 35: Codes within each policy

Policy document:	**Codes:**
Alpina Ethisch-kader 2020	AI-Principles
Alpina Ethisch-kader 2020	Autonomy
Alpina Ethisch-kader 2020	Autonomy
Alpina Ethisch-kader 2020	Data security; privacy

Alpina Ethisch-kader 2020	Data security
Alpina Ethisch-kader 2020	Availability
Alpina Ethisch-kader 2020	System should be designed to optimize control
Alpina Ethisch-kader 2020	Ethic reliability
Alpina Ethisch-kader 2020	Goal and achievability
Alpina Ethisch-kader 2020	Archive
Alpina Ethisch-kader 2020	Data quality
Alpina Ethisch-kader 2020	Integrity
Alpina Ethisch-kader 2020	Data quality
Alpina Ethisch-kader 2020	Communication customer
Alpina Ethisch-kader 2020	Privacy
Alpina Ethisch-kader 2020	Privacy
Alpina Ethisch-kader 2020	Confidentiality
Alpina Ethisch-kader 2020	Autonomy
Alpina Ethisch-kader 2020	Training
Alpina Ethisch-kader 2020	Transparency
Alpina Ethisch-kader 2020	Communication customer
Alpina Ethisch-kader 2020	Autonomy
Alpina Ethisch-kader 2020	Justice
Alpina Ethisch-kader 2020	Fairness
Alpina Ethisch-kader 2020	Fairness
Alpina Ethisch-kader 2020	Fairness
Alpina Ethisch-kader 2020	Reinforcement
Alpina Ethisch-kader 2020	Beneficence
Alpina Ethisch-kader 2020	Responsibilities
Alpina Ethisch-kader 2020	System should be designed to optimize control
Alpina Ethisch-kader 2020	Non-maleficence
Alpina Ethisch-kader 2020	Ethic
Alpina Ethisch-kader 2020	Employee
Alpina Ethisch-kader 2020	Documentation
Alpina Ethisch-kader 2020	Definitions
Alpina Ethisch-kader 2020	Risk assessment
Alpina Ethisch-kader 2020	Documentation
Alpina Ethisch-kader 2020	Responsibilities
Alpina Ethisch-kader 2020	Data security; privacy
Alpina Ethisch-kader 2020	Roles & Responsibilities
Alpina Ethisch-kader 2020	Data classification
Alpina Ethisch-kader 2020	Roles & Responsibilities
Alpina Ethisch-kader 2020	Reinforcement
Alpina Ethisch-kader 2020	Autonomy
Alpina Ethisch-kader 2020	Communication customer
Alpina Ethisch-kader 2020	Communication customer
Alpina Ethisch-kader 2020	Training
Alpina Ethisch-kader 2020	Education
Alpina Ethisch-kader 2020	Data retention

Alpina Ethisch-kader 2020	Conformance
Alpina Ethisch-kader 2020	DPIA
Alpina Incidentenmanagement 2021	Definitions
Alpina Incidentenmanagement 2021	Roles & Responsibilities
Alpina Incidentenmanagement 2021	Documentation
Alpina Incidentenmanagement 2021	Communication customer
Alpina Incidentenmanagement 2021	Identifying roles and responsibilities regarding incidents
Alpina Incidentenmanagement 2021	Documentation
Alpina Rolbeschrijvingen	Goal & Scope
Alpina Rolbeschrijvingen	Maintenance policy
Alpina Rolbeschrijvingen	Communication employees
Alpina Rolbeschrijvingen	Goal & Scope
Alpina Rolbeschrijvingen	Roles & Responsibilities
Alpina Rolbeschrijvingen	Audit
Alpina Rolbeschrijvingen	Awareness culture
Alpina Rolbeschrijvingen	Policy in line with strategy
Alpina Rolbeschrijvingen	Audits
Alpina Rolbeschrijvingen	Definitions
Alpina Rolbeschrijvingen	Responsibilities
Alpina Rolbeschrijvingen	Definitions
Alpina Rolbeschrijvingen	Tasks
Alpina Rolbeschrijvingen	Employees own responsibilities
Alpina uitbestedingsbeleid 2021	Goal & Scope
Alpina uitbestedingsbeleid 2021	Goal & Scope
Alpina uitbestedingsbeleid 2021	Stakeholders
Alpina uitbestedingsbeleid 2021	Responsibilities
Alpina uitbestedingsbeleid 2021	Tasks
Alpina uitbestedingsbeleid 2021	Goal & Scope
Alpina uitbestedingsbeleid 2021	Goal & Scope
Alpina uitbestedingsbeleid 2021	Goal & Scope
Alpina uitbestedingsbeleid 2021	Goal & Scope
Alpina uitbestedingsbeleid 2021	Duration
Alpina uitbestedingsbeleid 2021	Compliance
Alpina uitbestedingsbeleid 2021	Related laws and regulations
Alpina uitbestedingsbeleid 2021	Definitions
Alpina uitbestedingsbeleid 2021	Definitions
Alpina uitbestedingsbeleid 2021	Maintenance policy
Alpina uitbestedingsbeleid 2021	Exceptions to previous guidelines
Heilbron beveiliging en applicaties	Roles & Responsibilities
Heilbron beveiliging en applicaties	Data security communication; privacy
Heilbron beveiliging en applicaties	Data protection measure
Heilbron beveiliging en applicaties	Data quality; input
Heilbron bewaartermijn 2020	Data retention
Heilbron bewaartermijn 2020	Data retention

Heilbron bewaartermijn 2020	Maintenance policy
Heilbron Continuïteitsbeheer 2017 (Extern)	Definitions
Heilbron Continuïteitsbeheer 2017 (Extern)	Definitions
Heilbron Continuïteitsbeheer 2017 (Extern)	Goal & Scope
Heilbron Continuïteitsbeheer 2017 (Extern)	Responsibilities
Heilbron Continuïteitsbeheer 2017 (Extern)	Maintenance policy
Heilbron Continuïteitsbeheer 2017 (Extern)	Communication roles and responsibilities
Heilbron Continuïteitsbeheer 2017 (Extern)	Education and training
Heilbron Continuïteitsbeheer 2017 (Extern)	Responsibility communication
Heilbron Continuïteitsbeheer 2017 (Extern)	Data availability
Heilbron Continuïteitsbeheer 2017 (Extern)	Data availability
Heilbron Continuïteitsbeheer 2017 (Extern)	Data security
Heilbron Continuïteitsbeheer 2017 (Extern)	Data access requirement
Heilbron Continuïteitsbeheer 2017 (Extern)	Stakeholders
Heilbron Continuïteitsbeheer 2017 (Extern)	Risk assessment
Heilbron Continuïteitsbeheer 2017 (Extern)	Risk explanation
Heilbron Dreiging en incidentenbeheer	Data security measures
Heilbron Dreiging en incidentenbeheer	Definitions
Heilbron Dreiging en incidentenbeheer	Maintenance policy
Heilbron Dreiging en incidentenbeheer	Procedures when incidents occur
Heilbron Dreiging en incidentenbeheer	Documentation
Heilbron Dreiging en incidentenbeheer	Communication
Heilbron informatiebeveiligingsbeleid 2020	Goal & Scope
Heilbron informatiebeveiligingsbeleid 2020	Data availability
Heilbron informatiebeveiligingsbeleid 2020	Data integrity
Heilbron informatiebeveiligingsbeleid 2020	Data integrity
Heilbron informatiebeveiligingsbeleid 2020	Data classification protection
Heilbron informatiebeveiligingsbeleid 2020	Goal & Scope
Heilbron informatiebeveiligingsbeleid 2020	Data availability
Heilbron informatiebeveiligingsbeleid 2020	Goal & Scope
Heilbron informatiebeveiligingsbeleid 2020	Awareness responsibilities
Heilbron informatiebeveiligingsbeleid 2020	Compliance to law and regulations
Heilbron informatiebeveiligingsbeleid 2020	Data availability
Heilbron informatiebeveiligingsbeleid 2020	Communication employees update policy
Heilbron informatiebeveiligingsbeleid 2020	Data security remote access
Heilbron informatiebeveiligingsbeleid 2020	Yearly risk assessment
Heilbron informatiebeveiligingsbeleid 2020	Security measures suppliers
Heilbron informatiebeveiligingsbeleid 2020	Documentation incidents
Heilbron informatiebeveiligingsbeleid 2020	Communication incidents
Heilbron informatiebeveiligingsbeleid 2020	Access to laws and regulations
Heilbron informatiebeveiligingsbeleid 2020	Documentation
Heilbron informatiebeveiligingsbeleid 2020	Communication update policy
Heilbron informatiebeveiligingsbeleid 2020	Back-up data availability
Heilbron informatiebeveiligingsbeleid 2020	Responsibilities
Heilbron informatiebeveiligingsbeleid 2020	Roles & Responsibilities

Heilbron informatiebeveiligingsbeleid 2020	Data classification security risk
Heilbron informatiebeveiligingsbeleid 2020	Risk classification
Heilbron informatiebeveiligingsbeleid 2020	Compliance measures
Heilbron informatiebeveiligingsbeleid 2020	Data security; privacy
Heilbron informatiebeveiligingsbeleid 2020	Data security measure
Heilbron informatiebeveiligingsbeleid 2020	Data access requirement
Heilbron informatiebeveiligingsbeleid 2020	Data privacy procedure
Heilbron informatiebeveiligingsbeleid 2020	Data security measure
Heilbron informatiebeveiligingsbeleid 2020	Education and training increases awareness
Heilbron informatiebeveiligingsbeleid 2020	Security measure mail
Heilbron informatiebeveiligingsbeleid 2020	Roles and responsibilities access limited
Heilbron informatiebeveiligingsbeleid 2020	Roles and responsibilities RBAC
Heilbron Privacy-beleid	Goal & Scope
Heilbron Privacy-beleid	Description usage data; privacy
Heilbron Privacy-beleid	Description data processing goal; privacy
Heilbron Privacy-beleid	Description data processing goal; privacy
Heilbron Privacy-beleid	Description access, personal data; privacy
Heilbron Privacy-beleid	Data security privacy
Heilbron Privacy-beleid	Privacy; access to personal data
Heilbron Privacy-beleid	Privacy; special rights
Heilbron Privacy-beleid	Privacy; special rights
Heilbron Privacy-beleid	Privacy; special rights
Heilbron Privacy-beleid	Privacy; special rights
Heilbron Privacy-beleid	Data retention
Heilbron Privacy-beleid	Privacy; process special rights
Heilbron Privacy-beleid	Data availability
Heilbron Privacy-beleid	Data integrity
Heilbron Privacy-beleid	Data trust
Heilbron Privacy-beleid	Data privacy measures
Heilbron Privacy-beleid	Data privacy measures
Heilbron Privacy-beleid	Classification information
Heilbron Privacy-beleid	Data availability
Heilbron Privacy-beleid	Data availability
Heilbron Privacy-beleid	Communication updates and patches
Heilbron Privacy-beleid	Data protection measure
Heilbron Privacy-beleid	Data storage
Heilbron Privacy-beleid	Responsibilities assigned
Heilbron Privacy-beleid	Procedure security incidents
Heilbron Privacy-beleid	Procedure incident process
Heilbron Privacy-beleid	Documentation incidents
Heilbron Privacy-beleid	Documentation actions after incident
Heilbron Privacy-beleid	Procedure data leakage
Heilbron Systeembeheer (Extern)	Data security; availability
Heilbron Systeembeheer (Extern)	Data security; integrity
Heilbron Systeembeheer (Extern)	Trust

Heilbron Systeembeheer (Extern)	Classification information
Heilbron Systeembeheer (Extern)	Data security back-up
Heilbron Systeembeheer (Extern)	Data security, restore test
Heilbron Systeembeheer (Extern)	Description updates and patches
Heilbron Systeembeheer (Extern)	Protection and detection malware
Heilbron Systeembeheer (Extern)	Data storage responsibility assigning
Heilbron Systeembeheer (Extern)	Security incident procedure
Heilbron Systeembeheer (Extern)	Incident process framework
Heilbron Systeembeheer (Extern)	Observation incident
Heilbron Systeembeheer (Extern)	Observation actions incident
Heilbron Systeembeheer (Extern)	Procedure data leakage
V&V AI-beleid 2021	Goal & Scope
V&V AI-beleid 2021	Goal & Scope
V&V AI-beleid 2021	Maintenance policy
V&V AI-beleid 2021	Responsibilities
V&V AI-beleid 2021	Stakeholder
V&V AI-beleid 2021	Data quality ensurance
V&V AI-beleid 2021	Transparency AI
V&V AI-beleid 2021	Customer prioritization
V&V AI-beleid 2021	Non-maleficence
V&V AI-beleid 2021	Beneficence
V&V AI-beleid 2021	Explicability
V&V AI-beleid 2021	Non-maleficence
V&V AI-beleid 2021	Beneficence
V&V AI-beleid 2021	Ethic method to ensure ethical practices
V&V AI-beleid 2021	Responsibility AI
V&V AI-beleid 2021	Monitoring AI models
V&V AI-beleid 2021	Open communication departments regarding AI
V&V AI-beleid 2021	Education AI
V&V AI-beleid 2021	Data quality input AI model
V&V AI-beleid 2021	Data quality input
V&V AI-beleid 2021	Data quality
V&V AI-beleid 2021	Data quality input
V&V AI-beleid 2021	Risk assessment
V&V AI-beleid 2021	Data quality
V&V AI-beleid 2021	AI technology choice based on needs department
V&V AI-beleid 2021	Explainability
V&V AI-beleid 2021	Involvement employee decision making AI model
V&V AI-beleid 2021	Test AI model
V&V AI-beleid 2021	Yearly tests AI model
V&V AI-beleid 2021	Explainability depends on usage model
V&V AI-beleid 2021	High explainability support employees in decision making
V&V AI-beleid 2021	Explainability measures
V&V AI-beleid 2021	Transparency measures

V&V AI-beleid 2021	Fairness
V&V AI-beleid 2021	Justice
V&V AI-beleid 2021	Fairness
V&V AI-beleid 2021	Justice
V&V AI-beleid 2021	Diversity inclusion AI decision making
V&V AI-beleid 2021	Monitoring AI
V&V AI-beleid 2021	Data quality ensurance
V&V AI-beleid 2021	Responsibilities outsourcing systems
V&V AI-beleid 2021	AI validation
V&V AI-beleid 2021	AI outsourcing requirements
V&V AI-beleid 2021	Explainability
V&V AI-beleid 2021	Framework as check mechanism for AI model
V&V AI-beleid 2021	Maintenance policy
V&V AI-beleid 2021	Risk assessment
V&V Applicatiebeheer	Trust
V&V Applicatiebeheer	Integrity
V&V Applicatiebeheer	Goal & Scope
V&V Applicatiebeheer	Definition roles and responsibilities
V&V Applicatiebeheer	Application owner
V&V Applicatiebeheer	Data protection
V&V Applicatiebeheer	Responsibilities
V&V Applicatiebeheer	Data integrity
V&V Applicatiebeheer	Data integrity measures
V&V Applicatiebeheer	Data access requirements
V&V Applicatiebeheer	Risk assessment
V&V Applicatiebeheer	Data validity; integrity; quality
V&V Applicatiebeheer	Data access requirements; roles
V&V Applicatiebeheer	Data confidentiality; encryption
V&V Applicatiebeheer	Users authentication
V&V Applicatiebeheer	Data security measures
V&V Applicatiebeheer	Data availability
V&V Applicatiebeheer	Audit conformance to requirements yearly
V&V Beveiligingsbeleid 2021	Definitions
V&V Beveiligingsbeleid 2021	Goal & Scope
V&V Beveiligingsbeleid 2021	Definition roles and responsibilities
V&V Beveiligingsbeleid 2021	Goal & Scope
V&V Beveiligingsbeleid 2021	Policy in line with strategy
V&V Beveiligingsbeleid 2021	Communication employees
V&V Beveiligingsbeleid 2021	Data security clarity
V&V Beveiligingsbeleid 2021	Awareness data security risks
V&V Beveiligingsbeleid 2021	Awareness data security risks
V&V Beveiligingsbeleid 2021	Sharing information limited to known parties
V&V Beveiligingsbeleid 2021	Goal & Scope
V&V Beveiligingsbeleid 2021	Goal & Scope
V&V Beveiligingsbeleid 2021	Maintenance policy

V&V Beveiligingsbeleid 2021	Implementation of data security in culture
V&V Beveiligingsbeleid 2021	Security applications mobile
V&V Beveiligingsbeleid 2021	Best practices
V&V Beveiligingsbeleid 2021	Security checks applications
V&V Beveiligingsbeleid 2021	Data integrity
V&V Beveiligingsbeleid 2021	Roles and responsibilities; RBAC
V&V Beveiligingsbeleid 2021	Risk assessment applications
V&V Beveiligingsbeleid 2021	Data availability; back-ups design
V&V Beveiligingsbeleid 2021	Data availability; back-ups
V&V Beveiligingsbeleid 2021	Communication encryption; data confidentiality
V&V Beveiligingsbeleid 2021	Data security physical; data privacy
V&V Beveiligingsbeleid 2021	Data confidentiality; encryption
V&V Beveiligingsbeleid 2021	Procedure security incidents
V&V Beveiligingsbeleid 2021	Responsibility security profiles
V&V Beveiligingsbeleid 2021	Process data availability; continuity
V&V Beveiligingsbeleid 2021	Data privacy integration in culture
V&V Beveiligingsbeleid 2021	Roles and responsibilities; data ownership
V&V Beveiligingsbeleid 2021	Data security remote access
V&V Beveiligingsbeleid 2021	Education awareness
V&V Beveiligingsbeleid 2021	Education and training employees
V&V Beveiligingsbeleid 2021	Hardware measures security
V&V Beveiligingsbeleid 2021	Security requirements design
V&V Beveiligingsbeleid 2021	Security protection applications
V&V Beveiligingsbeleid 2021	Risk assessment
V&V Beveiligingsbeleid 2021	Role based access
V&V Beveiligingsbeleid 2021	Role based access
V&V Beveiligingsbeleid 2021	Security; protection servers
V&V Beveiligingsbeleid 2021	Data availability; back-ups
V&V Beveiligingsbeleid 2021	Data confidentiality; encryption
V&V Beveiligingsbeleid 2021	Data security; architecture
V&V Beveiligingsbeleid 2021	Data security measure
V&V Beveiligingsbeleid 2021	Data access requirements
V&V Beveiligingsbeleid 2021	Data security; alarm intruder
V&V Beveiligingsbeleid 2021	Data privacy; data leakage protection
V&V Beveiligingsbeleid 2021	Data confidentiality; encryption
V&V Beveiligingsbeleid 2021	Process security incident
V&V Beveiligingsbeleid 2021	Responsibility security
V&V Beveiligingsbeleid 2021	Data security; physical protection
V&V Beveiligingsbeleid 2021	Data security; protection power outage
V&V Beveiligingsbeleid 2021	Data security; protection natural causes
V&V Beveiligingsbeleid 2021	Data security; continuity within culture
V&V Beveiligingsbeleid 2021	Data security; availability; back-up
V&V Beveiligingsbeleid 2021	Procedure data security incident
V&V Beveiligingsbeleid 2021	Documentation plans
V&V Bewaartermijn	Definition

V&V Bewaartermijn	Data lifecycle process
V&V Bewaartermijn	Goal & Scope
V&V Bewaartermijn	Definitions
V&V Bewaartermijn	Data retention
V&V Bewaartermijn	Data permission
V&V Bewaartermijn	Data lifecycle permission
V&V Bewaartermijn	Data deletion rules
V&V Bewaartermijn	Special emergency process
V&V Bewaartermijn	Categorization
V&V Bewaartermijn	Definitions
V&V Bewaartermijn	Process deletion
V&V Bewaartermijn	Archiving
V&V Bewaartermijn	Special emergency process
V&V Continuïteitsbeheer	Goal & Scope
V&V Continuïteitsbeheer	Goal & Scope
V&V Continuïteitsbeheer	Definition roles and responsibilities
V&V Continuïteitsbeheer	Risk assessment
V&V Continuïteitsbeheer	Data security; robust hardware
V&V Continuïteitsbeheer	Data security; 24/7 employee availability
V&V Continuïteitsbeheer	Maintenance policy
V&V Data classificatie	Data classification guidelines
V&V Dreiging en incidentenbeheer	Goal & Scope
V&V Dreiging en incidentenbeheer	Goal & Scope
V&V Dreiging en incidentenbeheer	Responsibilities of teams
V&V Dreiging en incidentenbeheer	Risk assessment vulnerabilities
V&V Dreiging en incidentenbeheer	Risk assessment vulnerabilities
V&V Dreiging en incidentenbeheer	Retention period logs
V&V Dreiging en incidentenbeheer	Risk assessment threats
V&V Dreiging en incidentenbeheer	Communication risks
V&V Dreiging en incidentenbeheer	Roles and responsibilities incidents
V&V Dreiging en incidentenbeheer	Process regarding incidents
V&V Dreiging en incidentenbeheer	Maintenance policy
V&V Informatiebeleid 2019	Goal & Scope
V&V Informatiebeleid 2019	Responsibilities stakeholders
V&V Informatiebeleid 2019	Goal & Scope
V&V Informatiebeleid 2019	Goal & Scope
V&V Informatiebeleid 2019	Goal & Scope
V&V Informatiebeleid 2019	Maintenance policy
V&V Informatiebeleid 2019	Conformance responsibility
V&V Informatiebeleid 2019	Data architecture
V&V Informatiebeleid 2019	Data right definition
V&V Informatiebeleid 2019	Data security
V&V Informatiebeleid 2019	Data processing
V&V Informatiebeleid 2019	Data lifecycle
V&V Informatiebeleid 2019	Data classification

V&V Informatiebeleid 2019	Data classification measures
V&V Informatiebeleid 2019	Information processing responsibility
V&V Informatiebeleid 2019	Accountability responsible person
V&V Informatiebeleid 2019	Application responsibility
V&V Informatiebeleid 2019	Communication classification employees
V&V Informatiebeleid 2019	Risk assessment data flow
V&V Informatiebeleid 2019	Data classification sensitive info privacy
V&V Informatiebeleid 2019	Data access decision maker
V&V Informatiebeleid 2019	Classification information determined when created
V&V Informatiebeleid 2019	Privacy: personal data is not allowed to be asked
V&V Informatiebeleid 2019	Privacy: process when too much information is given
V&V Informatiebeleid 2019	Privacy: special requirements access to information
V&V Informatiebeleid 2019	Requirements copying classified info
V&V Informatiebeleid 2019	Requirements printing classified info
V&V Informatiebeleid 2019	Data deletion requirements
V&V Informatiebeleid 2019	Archiving
V&V Informatiebeleid 2019	Pseudonymizing info
V&V Informatiebeleid 2019	Data deletion alternative
V&V Informatiebeleid 2019	Data retention
V&V Informatiebeleid 2019	Data security; physical protection
V&V Informatiebeleid 2019	Data security
V&V Informatiebeleid 2019	Security applications
V&V Informatiebeleid 2019	Roles and responsibilities access
V&V Informatiebeleid 2019	Roles and responsibilities access
V&V Informatiebeleid 2019	Privacy; Requirements sending data
V&V Informatiebeleid 2019	Process data requests employees
V&V Informatiebeleid 2019	Data protection privacy
V&V Informatiebeleid 2019	Data security of BI environment
V&V Informatiebeleid 2019	Data security; integrity measures
V&V Informatiebeleid 2019	Data security; integrity measures
V&V Informatiebeleid 2019	Data security; availability BI environment
V&V Informatiebeleid 2019	Availability classification
V&V Informatiebeleid 2019	Integrity classification
V&V Informatiebeleid 2019	Data classification
V&V Informatiebeleid 2019	Risk assessment
V&V Informatiebeleid 2019	Monitor and report
V&V Informatiebeleid 2019	Stakeholders' responsibilities and roles
V&V Netwerk en communicatie	Goal & Scope
V&V Netwerk en communicatie	Goal & Scope
V&V Netwerk en communicatie	Roles & Responsibilities
V&V Netwerk en communicatie	Security physical and digital networks
V&V Netwerk en communicatie	Architecture
V&V Privacy-beleid 2021	Goal & Scope
V&V Privacy-beleid 2021	Goal & Scope
V&V Privacy-beleid 2021	Stakeholders' responsibilities and roles

V&V Privacy-beleid 2021	Goal & Scope
V&V Privacy-beleid 2021	Requirements processing data
V&V Privacy-beleid 2021	Data privacy
V&V Privacy-beleid 2021	Data integrity
V&V Privacy-beleid 2021	Data availability
V&V Privacy-beleid 2021	Data security personal data
V&V Privacy-beleid 2021	Ethics; justice; non-maleficence
V&V Privacy-beleid 2021	Transparency
V&V Privacy-beleid 2021	Goal & Scope
V&V Privacy-beleid 2021	Goal & Scope
V&V Privacy-beleid 2021	Goal & Scope
V&V Privacy-beleid 2021	Maintenance policy
V&V Privacy-beleid 2021	Reinforcement policy
V&V Privacy-beleid 2021	Definitions
V&V Privacy-beleid 2021	Compliance laws and regulations privacy
V&V Privacy-beleid 2021	Requirements for compliance
V&V Privacy-beleid 2021	Roles and responsibilities privacy
V&V Privacy-beleid 2021	Documentation
V&V Privacy-beleid 2021	Privacy: counter that answers problems
V&V Privacy-beleid 2021	Privacy checklist improves awareness
V&V Privacy-beleid 2021	Yearly review privacy requirements
V&V Privacy-beleid 2021	Privacy awareness
V&V Privacy-beleid 2021	Data leakage prevention team
V&V Privacy-beleid 2021	Privacy by design
V&V Privacy-beleid 2021	Compliance privacy policy
V&V Privacy-beleid 2021	Privacy contracts
V&V Privacy-beleid 2021	Compliance privacy
V&V Privacy-beleid 2021	Privacy requirements suppliers
V&V Privacy-beleid 2021	Transparency
V&V Privacy-beleid 2021	Transparency
V&V Privacy-beleid 2021	Privacy
V&V Privacy-beleid 2021	Data access decision making management
V&V Privacy-beleid 2021	Privacy requirements
V&V Privacy-beleid 2021	Privacy requirements
V&V Privacy-beleid 2021	Privacy requirements
V&V Privacy-beleid 2021	Data security; privacy
V&V Privacy-beleid 2021	Data integrity
V&V Privacy-beleid 2021	Data retention personal data
V&V Privacy-beleid 2021	Privacy guidelines
V&V Privacy-beleid 2021	Data retention
V&V Privacy-beleid 2021	Privacy; guidelines recording calls
V&V Privacy-beleid 2021	Autonomy; responsibilities
V&V Privacy-beleid 2021	Data security; privacy rights
V&V Privacy-beleid 2021	Transparency data
V&V Privacy-beleid 2021	Risk assessment

V&V Privacy-beleid 2021	Maintenance policy
V&V Systeembeheer	Goal & Scope
V&V Systeembeheer	Goal & Scope
V&V Systeembeheer	Roles & Responsibilities
V&V Systeembeheer	Data security requirements
V&V Systeembeheer	Requirements network storage system
V&V Systeembeheer	Documentation
V&V Systeembeheer	Conformance testing/auditing
V&V Systeembeheer	Data availability; back-ups
V&V Systeembeheer	Manage change -> make procedure
V&V Systeembeheer	Document change
V&V Systeemontwikkeling	Goal & Scope
V&V Systeemontwikkeling	Goal & Scope
V&V Systeemontwikkeling	Roles & Responsibilities
V&V Systeemontwikkeling	Process description
V&V Systeemontwikkeling	System architecture requirements
V&V Systeemontwikkeling	Documentation design systems
V&V Systeemontwikkeling	Data security new system acquirement
V&V Systeemontwikkeling	Procedures new systems
V&V Systeemontwikkeling	Documentation
V&V Technische beveiliging	Roles and responsibilities regarding upkeeping
V&V Technische beveiliging	Security guidelines architecture
V&V Technische beveiliging	Data architecture
V&V Technische beveiliging	Data security architecture guidelines
V&V Technische beveiliging	Communication employees guidelines
V&V Technische beveiliging	Data protection measure
V&V Technische beveiliging	Connection roles and responsibilities to ID
V&V Technische beveiliging	Transparency data access rights
V&V Technische beveiliging	Data privacy
V&V Technische beveiliging	Confidentiality
V&V Toegangsbeheer	Roles & Responsibilities
V&V Toegangsbeheer	Data access requirements
V&V Toegangsbeheer	Categorization
V&V Toegangsbeheer	Awareness responsibilities
V&V Toegangsbeheer	Function separation; roles and responsibilities
V&V Toegangsbeheer	Appointing roles and responsibilities
V&V Toegangsbeheer	Roles
V&V Toegangsbeheer	Link categorization to roles
V&V Toegangsbeheer	Special roles and responsibilities
V&V Toegangsbeheer	Data security; access
V&V Toegangsbeheer	Data security; access

After each document was coded, the codes were grouped based on similarities and formed sub-components. These can be seen in table 36.

Table 36: Overview codes policy analysis as part of their sub-components

Codes	**Sub-component:**
Data security	Data security
Security checks applications	
Data security; 24/7 employee availability	
Data security	
Data security measures	
Data security remote access	
Security measures suppliers	
Data security measure	
Data security measure	
Security measure mail	
Procedure security incidents	
Data security measures	
Data security clarity	
Awareness data security risks	
Awareness data security risks	
Security applications mobile	
Data security remote access	
Hardware measures security	
Security requirements design	
Security protection applications	
Security; protection servers	
Data security; architecture	
Data security measure	
Data security; alarm intruder	
Data security; physical protection	
Data security; protection power outage	
Data security; protection natural causes	
Data security; robust hardware	
Data security	
Data security; physical protection	
Data security	
Security applications	
Data security of BI environment	
Security physical and digital networks	
Data security personal data	
Data security requirements	
Data security new system acquirement	
Security guidelines architecture	
Data security architecture guidelines	
Data protection measure	
Users authentication	
Data protection measure	

Data protection	
Data protection measure	
Data security, restore test	
Protection and detection malware	
Data security; privacy	Privacy
Data security; privacy	
Data security communication; privacy	
Data security; privacy	
Data privacy procedure	
Data security privacy	
Data privacy measures	
Data privacy measures	
Data security physical; data privacy	
Data privacy integration in culture	
Data privacy; data leakage protection	
Data classification sensitive info privacy	
Privacy: personal data is not allowed to be asked	
Privacy: process when too much information is given	
Privacy: special requirements access to information	
Privacy; Requirements sending data	
Data protection privacy	
Data privacy	
Data security; privacy	
Data security; privacy rights	
Data privacy	
Requirements copying classified info	
Requirements printing classified info	
Sharing information limited to known parties	
Data leakage prevention team	
Integrity	Integrity
Data integrity	
Data integrity	
Data integrity	
Integrity	
Data integrity	
Data integrity measures	
Data validity; integrity; quality	
Data integrity	
Data security; integrity measures	
Data security; integrity measures	
Integrity classification	
Data integrity	
Data integrity	
Data trust	
Trust	

Data security; integrity	
Trust	
Data confidentiality; encryption	Confidentiality
Communication encryption; data confidentiality	
Data confidentiality; encryption	
Data confidentiality; encryption	
Data confidentiality; encryption	
Confidentiality	
Availability	Availability
Data availability	
Data availability	
Data availability	
Data availability	
Data availability	
Back-up data availability	
Data availability	
Data availability	
Data availability	
Data availability	
Data availability; back-ups design	
Data availability; back-ups	
Process data availability; continuity	
Data availability; back-ups	
Data security; availability; back-up	
Data security; continuity within culture	
Data security; availability BI environment	
Availability classification	
Data availability	
Data availability; back-ups	
Data security; availability	
Data security back-up	
Data classification	Data classification
Data classification protection	
Classification information	
Data classification guidelines	
Data classification	
Data classification measures	
Classification information determined when created	
Data classification	
Categorization	
Categorization	
Data classification security risk	
Classification information	
Data security; access	Data access
Data security; access	

Data access requirement	
Data access requirement	
Data access requirements	
Data access requirements	
Data access decision maker	
Data access decision making management	
Transparency data access rights	
Data access requirements	
Data lifecycle	Data process
Data processing	
Process description	
Requirements processing data	
Process data requests employees	
Data lifecycle permission	
Data quality; input	
Data storage	Data storage
Requirements network storage system	
Data retention	Data retention, deletion, and archiving
Data retention	
Data retention	
Data retention	
Data retention	
Retention period logs	
Data retention	
Data retention personal data	
Data retention	
Data deletion rules	
Process deletion	
Data deletion requirements	
Data deletion alternative	
Pseudonymizing info	
Archiving	
Archive	
Archiving	
Roles & Responsibilities	Roles & Responsibilities
Roles & Responsibilities	
Roles & Responsibilities	
Identifying roles and responsibilities regarding incidents	
Roles & Responsibilities	
Employees own responsibilities	
Roles & Responsibilities	
Roles & Responsibilities	
Roles and responsibilities access limited	
Roles and responsibilities RBAC	
Responsibilities	

Roles and responsibilities; RBAC	
Roles and responsibilities; data ownership	
Responsibility security profiles	
Role based access	
Role based access	
Roles and responsibilities incidents	
Roles & Responsibilities	
Roles and responsibilities privacy	
Roles & Responsibilities	
Roles & Responsibilities	
Roles and responsibilities regarding upkeeping	
Roles & Responsibilities	
Function separation; roles and responsibilities	
Appointing roles and responsibilities	
Roles	
Special roles and responsibilities	
Responsibilities	
Responsibilities	
Responsibilities	
Awareness responsibilities	
Responsibilities assigned	
Responsibility AI	
Responsibilities outsourcing systems	
Responsibilities	
Responsibility security	
Responsibilities of teams	
Conformance responsibility	
Information processing responsibility	
Application responsibility	
Awareness responsibilities	
Application owner	
Data storage responsibility assigning	
Connection roles and responsibilities to ID	Connecting roles and responsibilities digitally
Roles and responsibilities access	
Roles and responsibilities access	
Link categorization to roles	
Data access requirements; roles	
Roles and responsibilities regarding upkeeping	
Explicability	Explicability
Explainability	Explainability
Explainability depends on usage model	
Explainability	
Explainability measures	
High explainability support employees in decision making	
Accountability responsible person	Accountability

Transparency AI	Transparency
Transparency	
Transparency measures	
Transparency	
Transparency	
Transparency	
Transparency data	
AI-Principles	AI-Principles
Beneficence	Beneficence
Beneficence	
Beneficence	
Non-maleficence	Non-Maleficence
Non-maleficence	
Non-maleficence	
Autonomy	Autonomy
Autonomy	
Autonomy	
Autonomy	
Autonomy	
Autonomy; responsibilities	
Justice	Justice
Justice	
Justice	
Justice	
Fairness	Fairness
Fairness	
Fairness	
Fairness	
Fairness	
Involvement employee decision making AI model	AI-Decision making
AI technology choice based on needs department	
Diversity inclusion AI decision making	
AI validation	
AI outsourcing requirements	
Data quality	Data quality input
Data quality	
Data quality ensurance	
Data quality input AI model	
Data quality input	
Data quality	
Data quality input	
Data quality	
Data quality ensurance	
Monitoring AI models	Monitoring AI-models
Monitor and report	

Monitoring AI	
Test AI model	
Yearly tests AI model	
Framework as check mechanism for AI model	
Goal & Scope	Goal & Scope
Goal & Scope	
Goal & Scope	
Goal & Scope	
Goal & Scope	
Goal & Scope	
Goal & Scope	
Goal & Scope	
Goal & Scope	
Goal & Scope	
Goal & Scope	
Goal & Scope	
Goal & Scope	
Goal & Scope	
Goal & Scope	
Goal & Scope	
Goal & Scope	
Goal & Scope	
Goal & Scope	
Goal & Scope	
Goal & Scope	
Goal & Scope	
Goal & Scope	
Goal & Scope	
Goal & Scope	
Goal & Scope	
Goal & Scope	
Goal & Scope	
Goal & Scope	
Goal & Scope	
Goal & Scope	
Goal & Scope	
Goal & Scope	
Goal & Scope	
Goal & Scope	
Goal & Scope	
Goal & Scope	
Goal & Scope	
Goal & Scope	
Goal & Scope	
Goal & Scope	

Goal and achievability	
Definitions	Definitions
Definitions	
Definitions responsibilities	
Definition tasks	
Definitions	
Definitions	
Definitions	
Definitions	
Definition roles and responsibilities	
Definitions	
Definition roles and responsibilities	
Definition data lifecycle process	
Definitions	
Definitions	
Definition roles and responsibilities	
Data right definition	
Definitions	
Responsibilities stakeholders	Stakeholders
Responsibilities stakeholders	
Responsibilities stakeholders	
Stakeholders' responsibilities and roles	
Stakeholders' responsibilities and roles	
Stakeholders and their tasks	
Stakeholders	
Compliance to law and regulations	Audit and compliance
Compliance measures	
Compliance	
Conformance through DPIA	
Compliance laws and regulations privacy	
Requirements for compliance	
Conformance testing/auditing	
Audit	
Audits	
Audit conformance to requirements yearly	
Maintenance policy	Maintenance policy
Maintenance policy	
Maintenance policy	
Maintenance policy	
Duration	
Maintenance policy	
Maintenance policy	
Maintenance policy	
Maintenance policy	
Yearly review privacy requirements	

Maintenance policy	
Maintenance policy	
Maintenance policy	
Maintenance policy	
Maintenance policy	
Documentation	Documentation
Documentation	
Documentation	
Documentation	
Documentation	
Documentation incidents	
Documentation	
Documentation incidents	
Documentation plans	
Documentation	
Documentation	
Documentation design systems	
Documentation	
Documentation actions after incident	
Document change	
Description updates and patches	
Procedures when incidents occur	Procedures
Procedure incident process	
Procedure data leakage	
Procedure security incidents	
Process security incident	
Procedure data security incident	
Special emergency process	
Special emergency process	
Process regarding incidents	
Procedures new systems	
Security incident procedure	
Incident process framework	
Observation incident	
Observation actions incident	
Procedure data leakage	
Risk assessment	Risk assessment policy
Risk assessment	
Risk assessment	
Risk assessment	
Risk assessment	
Risk assessment	
Risk explanation	
Yearly risk assessment	
Data classification security risk	

Risk classification	
Risk assessment	
Risk assessment	
Risk assessment applications	
Risk assessment vulnerabilities	
Risk assessment vulnerabilities	
Risk assessment threats	
Risk assessment data flow	
Risk assessment	
Data architecture	Data architecture
Architecture	
Data architecture	
System architecture requirements	
Policy in line with strategy	Strategic alliance
Policy in line with strategy	
System should be designed to optimize control and ethical behavior	Design
System should be designed to optimize control and ethical behavior	
Privacy by design	
Exceptions to previous guidelines	Others
Best practices	
Manage change -> make procedure	
Related laws and regulations	
Access to laws and regulations	
Employee	Communication
Open communication departments regarding AI	
Communication incidents	
Communication risks	
Communication roles and responsibilities	
Responsibility communication	
Communication classification employees	
Communication employees	
Communication employees	
Communication employees guidelines	
Communication updates and patches	
Communication employees update policy	
Communication update policy	
Communication	
Communication customer	
Communication customer	
Communication customer	
Communication customer	
Communication customer	
Education awareness	Education & Training

Education and training employees	
Training	
Training	
Education	
Education and training	
Education AI	
Education and training increases awareness	
Awareness culture	
Implementation of data security in culture	
Reinforcement	Reinforcement
Reinforcement	
Reinforcement policy	
Reinforcement	
Ethics	Ethics
Ethic	
Confidentiality	
Ethic reliability	
Ethic method to ensure ethical practices	
Description data processing goal; privacy	Communication privacy
Description data processing goal; privacy	
Description usage data; privacy	
Privacy; guidelines recording calls	
Description accesses personal data; privacy	
Privacy; access to personal data	
Data permission	
Privacy awareness	
Privacy checklist improves awareness	
Privacy: counter that answers problems	
Privacy	Privacy
Privacy	
Privacy; special rights	
Privacy; special rights	
Privacy; special rights	
Privacy; special rights	
Privacy; process special rights	
Privacy contracts	
Compliance privacy policy	
Compliance privacy	
Privacy requirements suppliers	
Privacy	
Privacy requirements	
Privacy requirements	
Privacy requirements	
Privacy guidelines	

These sub-components in turn form the components. The sub-components, their respective components, and their aggregate dimension can be seen in table 37.

Table 37: Overview of sub-components, their components, and the dimension they belong to according to the policy analysis

Data security	Data security	Data policy components:
Privacy		
Integrity		
Confidentiality		
Availability		
Data classification		
Data access		
Data process	Data lifecycle	
Data storage		
Data retention, deletion, and archiving		
Roles & Responsibilities	Roles & Responsibilities	
Connecting roles and responsibilities digitally		
Explicability	Explicability	AI policy components:
Explainability		
Accountability		
Transparency		
AI-principles	AI-principles	
Beneficence		
Non-maleficence		
Autonomy		
Justice		
Fairness		
AI-decision making	AI-models	
Data quality input		
Monitoring AI-models		
Goal & Scope	General guidelines	General policy components
Definitions		
Stakeholders		
Audit & Compliance		
Maintenance policy		
Documentation		
Procedures		
Risk assessment policy	Risk assessment	
Design	Design & Strategy	
Architecture		
Strategy		
Exceptions	Others	
Managing change		
Access to laws and regulations		
Best practices		
Communication	Employee	

Education & Training		
Reinforcement	Employee	General policy components
Ethics	Ethics	
Privacy	Privacy	
Communication privacy		

If the sub-components are removed, the components with their aggregate dimensions remain, as can be seen in table 38.

Table 38: Overview dimensions and their components based on the policy analysis

Data policy components	Data security
	Data lifecycle
	Roles & Responsibilities
AI-policy components	AI-principles
	Explicability
	AI-models
General policy components	General guidelines
	Risk assessment
	Design & Strategy
	Others
	Employee
	Ethics
	Privacy

Appendix 5: Interview 1 transcript

I = Interviewer

R = Respondent

I: Het doel van dit interview is om een beeld te krijgen van wat 'experts' zoals jij cruciale elementen vinden in data & AI beleid. Het plan is om eerst jou te vragen wat jij belangrijke componenten vindt met betrekking tot databeleid, AI-beleid en algemeen beleid en dat deze dan worden vergeleken met de componenten die voortgekomen zijn uit het literatuuronderzoek en de beleidsanalyse. Dus als we beginnen met databeleid, wat vind jij cruciale elementen binnen databeleid?

R: Uhm, nou kijk, ik kijk daarnaar mede vanuit de situatie waar we inzitten, dus waar we nu staan. Dus ik denk dat door de tijd heen, wat ik de belangrijkste onderwerpen vind, dat gaat verschuiven. Dus je kijkt daarmee vooral waar staan we nu, wat gaan we doen. We zitten in de fase waar we een nieuw platform opbouwen, nieuwe reporten en informatieproducten opbouwen. Dus we zitten vooral in de bouwfase en het verzamelen van de data. De eerste keer dat we zo'n data-Lake aanraken en verwerken. Dus in deze fase is voor mij het belangrijkste veiligheid. Security is voor mij het belangrijkste, welke mensen kunnen op het platform terecht, is het platform helemaal mooi ingericht etc. Dus ik zou security denk ik met stip op 1 zetten.

I: Oké, dus de prioriteit staat erg hoog van security, zou dit veranderen naarmate het platform meer volwassen wordt?

R: Ja, want dan zou ik inderdaad dat we het platform zo hebben ingericht en dat het stabieler is en ervaring hebben opgedaan. Mensen hebben hun skills verder ontwikkeld en dat we de inrichting steeds scherper hebben staan, dat we meer overgaan op meta datering. Dus in eerste instantie security, met name ook uitgaand dataverkeer, wie heeft toegang tot welke data en hoe verzenden we dat. Dus dat gaat door op de security, hoe versturen we de data en wie kan er dan ook echt bij. De access policy, dus ik vind dat belangrijk. Vervolgens is het ook belangrijk, in de fase waar we nu staan, is data quality. Het verbeteren van data, wie is hiervoor verantwoordelijk? Wat je merkt in onze organisatie is dat we dat moeilijk van de grond krijgen, omdat er niemand echt verantwoordelijk is voor de data quality. Er is ook weinig last van, want als ze er last van krijgen dan gebeurt er wel wat. Maar over het algemeen is er veel gekeken naar het oplossen van data quality problemen in een report. Dan gingen we een referentie lijstje aanleggen en met dat lijstje gingen we data quality problemen oplossen. Terwijl je eigenlijk zou moeten zeggen, dat het in de bron wordt verbeterd.

I: Oké, zou je dan ook data access zien als een methode om data security te waarborgen?

R: Ja, correct.

I: Zijn er naast data security en data quality, ervan uitgaand dat data access onderdeel is van data security, verder nog cruciale elementen binnen een databeleid?

R: Ik denk dat een derde onderwerp is, als je kijkt naar databeleid, is denk ik ook de manier waarop we het opslaan. Dus echt de opslag, waar slaan we de data op? Dus nu in principe in 1 grote data Lake, waaruit we verschillende datamarkten maken. Vanuit die datamarts geven we iemand daar toegang toe. Vanuit toegangsbeheer. Maar wat ik nu vooral zie is dat er een heleboel database zijn waar dit nog niet zo geregeld is. Dus ik wil veel meer data vanuit het datacentra alles is opgeslagen en vanuit daaruit gedistribueerd wordt.

I: Oké, dus hier gaat het echt om data in een centrale plek op te slaan, zodat het toegangsbeheer makkelijker toegepast kan worden?

R: Ja, dat klopt. Bij het onderhouden van het platform is het belangrijk dat het reproduceerbaar hebben, dat we ook zien wat voor codes zijn gebruikt, dat we dat vastleggen in zogenaamde branches en dat we die structuren vasthouden en reproduceerbaar hebben hoe we tot bepaalde uitkomsten zijn gekomen.

I: De documentatie?

R: De documentatie inderdaad, dus hoe wordt zo'n product gebouwd, ook voor het beheers beleid naar de toekomst toe. Dat ook een medewerkers wissel kan plaatsvinden zonder al te grote problemen en kennis verlies. Dat vind ik ook nog wel een belangrijke daarin. Als we iets verder zijn, wil ik ook meta datering op orde hebben. Dus dat we data classificatie hebben en hoe we daar mee omgaan per klasse. Dus, we moeten bijvoorbeeld weten welke bron data vandaan komen, betreft de data bijzondere persoonsgegevens, etc. Dat zou ik graag in die meta datering vast willen houden.

I: Oké, dat is wel interessant om te horen. Dat leid ik dan ook een beetje terug naar wat die andere *zelfde naam*, die toen bij ons vergaderingsoverleg was in Leusden, vertelde over hoe ze data invoeren in het systeem.

R: Ja, inderdaad. Dat is waarom ik data linea belangrijk vind. De herkomst van de data, wat is de oorsprong van de data, hoe heb ik het verkregen. Dat bepaalt ook hoe ik erom mag en moet gaan. En de data classificatie die daarbij hoort. Dat is weer een stapje verder dan dat logisch toegangsbeheer waar je in eerste instantie aan denkt. Wat ik wel erg belangrijk vind is de cultuur die erom heen ligt, wat wel moeilijker te vatten is in een hard afstreep punt. Dus de hele cultuur om: "Zijn mensen bewust van de data waarmee ze werken?" en de vertrouwelijkheid van die data.

I: Oké, en zou je dit cultuurcomponent meer plaatsen bij databeleid, of in het algemene beleid?

R: Nou kijk ik zou dat toch graag in het beleid willen hebben. Wat betekent dat nou, wanneer zijn we tevreden, welk gedrag laten mensen zien als zij de juiste cultuur uitademen. Wat zeggen ze dan, wat doen ze dan, wat voor gedrag laten ze zien? Dat soort zaken. Dat vind ik wel echt een belangrijk, dat ze de juiste data invoeren, dat ontwikkelaars zodanig een applicatie ontwikkelen dat we er op datagebied ook echt wat aan hebben. Dus je kan als ontwikkelaar af met een open tekstveld, maar qua data is dat echt een draak. Je wilt liever een drop down met vooraf gedefinieerde velden hebben. Je zou ook willen dat het management een data gedreven beslissing neemt, dat data wordt gebruikt in een manager performance cyclus, dat we daarin met elkaar bespreken: "Zo presteer jij, dit is de norm en zo zou je het nog beter kunnen doen". Dat zit hem veel meer in de cultuur, en dat mensen snappen en bewust zijn met de risico's van databestanden delen. En specifieke wetgeving snappen. Zo vind ik dat data engineers moeten weten dat er een specifieke bewaartermijn zit op data en dat er een AVG bestaat met bepaalde eisen. Maar ook als mensen data delen, dan vind ik dat ze bewust moeten zijn van de risico's. Dat als ze iets versturen, dat netjes beveiligen met zo'n beveiligd mailprogramma, bijvoorbeeld of via een SSTP-data bestanden delen. Dat zijn ook wel items die onder die cultuur vallen.

I: En hoe zie je dit voor je? Is dit in het beleid opgenomen als een lijst met gewenst gedragingen?

R: Nou ik denk dat dit een wisselwerking is, vanuit je beleid probeer je invulling te geven aan de gewenste cultuur. Maar dat je ook vanuit die gewenste cultuur dat je elementen terug ziet je in je beleid terug moet zien. Een voorbeeld is dat ik, qua cultuur, het heel belangrijk vind dat mensen, in zeg maar wat, de specifieke wetgeving moeten kennen qua bewaren van data. Dat betekent dat ik in het beleid iets moet opnemen over hoe mensen dat zich tot hen moeten nemen. Dat je bijvoorbeeld, vertaald naar beleid, dat we vanuit het datateam een casus op het intranet moeten publiceren minimaal één keer per maand. Of dat als we zaken tegenkomen zoals dat de cultuur niet actief wordt

nageleefd, dat we dat delen. Als er een case is waar iets is misgegaan, dat we dat ook delen. Dat hebben we een paar gedaan in het afdelingsoverleg, waarbij we de casus hebben genomen waarbij data vanuit externe kantoren met interne kantoren werd vermengd. Dat is echt niet de bedoeling. Dat kan je wel in beleid opschrijven, maar dat gaat pas echt leven als je laat weten als het fout gaat. Die casus moet je dan ook bespreekbaar maken, zonder iemand volledig voor schut te zetten.

I: Oké, dus, als samenvatting, heb ik in het databeleid als cruciale componenten: Data security, met als subonderdeel data access, data quality/verantwoordelijkheid, de manier van opslag, de reproduceerbaarheid, transparantie en documentatie. Dan ook nog meta datering en classificatie en data lineage/herkomst van de oorsprong.

R: Ja, correct. Daar zou ik meer technologische factoren ook bij voegen. Bijvoorbeeld een aantal hele basale afspraken, dat data alleen wordt opgeslagen in de Europese unie. Dat zijn echt dingen die daar standaard in terug horen te komen.

I: Dus vooral wetgeving en regelingen?

R: Ja, correct. Nou hoop ik niet dat ik iets vergeten ben, waarvan jij denkt: “Waarom noem je dat niet.”

I: Voor zover ik het zie hebben we een best goed lijstje.

R: Dat ga jij nu laten zien zeker?

I: Voordat we het resultaat van de analyses bekijken, wil ik eerst even polsen wat jij belangrijk vindt in een AI-policy.

R: Ah, is goed. Ik heb een paar dingen die top of mind zijn. Dat is wel lastig, omdat we nog niet zoveel AI hebben. Wat ik belangrijk vind op het moment is dat je ongeacht de techniek, want de AI-technieken kunnen erg complex zijn. Dat je ongeacht de techniek kunt uitleggen hoe het algoritme werkt, en hoe het tot een resultaat is gekomen.

I: Dus de uitlegbaarheid?

R: Absoluut, je moet niet tot zo’n black box situatie komen. Dat is niet goed voor het vertrouwen in het gebruik van de tool. Ook leidt het er tot toe dat je als AI ontwikkelaar/data scientists, dat je zelf ook niet goed in staat bent om te zien waar je sommige afslagen hebt genomen die helemaal niet goed zijn. Ik vind de explainability heel erg belangrijk, want zonder dit ga je ook iets anders niet zien in de black box. Want met AI kan je heel makkelijk bijvoorbeeld discrimineren. Op geloof, ras, sekse etc. Want het gebeurt heel snel, dat hebben we ook wel gezien in een algoritme dat we gebruikt hebben. Voor inkomens verzekeringen, dat we zien dat vrouwen een hogere kans hebben om uit te vallen. Is het dan verantwoord om, bedrijven waar veel vrouwelijke werknemers zitten, een hogere premie te berekenen? Ik vind dat ethische gedeelte, dus de ethiek, heel erg belangrijk. Vinden we dat wel oké met elkaar? En dit staat los van alleen discriminatie. Zo weet ik een ander voorbeeld case van jaren geleden, dat de schades bij een auto, veel hoger zijn bij huizen met een huisnummer toevoeging. Dat komt omdat dat meestal flats zijn, dus dan staan auto’s dichter bij elkaar geparkeerd, dus heb je een grotere kans op krasjes etc. Nou weten we ook met zen allen, algemeen, dat als je een appartement hebt over het algemeen minder inkomen hebt. Dus, dan moet je jezelf echt afvragen… want we zijn het verzekeren ooit begonnen als een collectiviteitsgedachte en dat je de risico’s met elkaar deelt. En als je op die manier het gaat uitrekenen, dus dan krijg je een bepaalde bevolkingsgroep, als je het helemaal doortrekt, zo je een bepaalde bevolkingsgroep onverzekerbaar kunnen maken.

I: Dan moet je inderdaad afvragen waar de grens ligt.

R: Inderdaad, de ethiek vind ik erg belangrijk. Ik denk ook aan technologie. In verband met kosten, daar zou ik zeggen, in je AI-beleid, wil je dat je voor het doel dat je wilt bereiken, de meest eenvoudige technologie wilt inzetten. Want daar ook weer, kan je helemaal losgaan met alle technologie en exotische verbanden. Maar daar moet je ook kijken, wat is nou de eenvoudigste en ook weer beste uit te leggen technologie. Maar ook voor de kosten die dat met zich meebrengt. Voor de rekeningkracht van Azure, krijg je voor iedere berekening die je doet moet je geld betalen. Dus de eenvoudige technologie raakt meerdere componenten. Ook die kostencomponent vind ik daarin erg belangrijk. Binnen AI vind ik security ook weer erg belangrijk. Welke data stop ik erin en wat zijn de outputs en waar komen die terecht. Dus ik zou het qua beleid, dat het hele AI gebeuren, precies in dezelfde Azure omgeving gebeurt als de rest van het data platform. Wat je heel snel ziet met die AI-jongens is dat ze graag in dockers en containers en weet ik veel wat ze allemaal willen prutsen. Dat kan allemaal, maar vaak kan je hetzelfde bereiken binnen die omgeving.

I: Oké, maar wat is dan het verschil tussen data security binnen de data policy en de AI-policy? Is dat van de AI hetzelfde?

R: Nee, dat denk ik niet. Nou, op het gebied van AI bedoel je dan? Hm, goeie vraag, goeie vraag. Nou als je het binnen hetzelfde platform doet heb je dezelfde security eisen. Maar voor AI heb je vaak andere tools nodig die ook vaker open source tooling met zich meebrengen. Dat zie je ook met die open source tooling, der was laatst een tooling wat een prachtige tool, maar daar zat dus een erg grote fout in, een kwetsbaarheid. Daar loop je dan wel tegen aan, dat heb je niet vaak met standaard tools. Dus in die zin gebruikt AI minder standaard componenten, maar ze nemen wel kwetsbaarheden met zich mee. Dus dat je daar extra alert op moet zijn.

I: Zou dat dan niet meer leunen tegen regels met betrekking tot requirements van applicaties? Bepaalde applicaties die gebruikt worden voor AI zouden moeten voldoen aan bepaalde regels.

R: Ja, maar ook de hoeveelheden data. In je platform heb je data Lake zitten, daar zit al je data in en die moet super goed beveiligd zijn. Daarboven heb je data marts, die moeten ook weer beveiligd zijn en daar mogen alleen bepaalde personen bij. Wat je ziet is dat de AI-jongens vaak veel meer data gebruiken dan dat in zo'n datamarkt staat. Dus de verzamelde data zijn velen malen groter. Want juist om tot goede algoritmes te komen heb je vaak veel data nodig.

I: En wat is dan het probleem van veel data?

R: Nou de impact is veel groter, voor de risico's kijk je altijd naar kans x impact. Nou de kans is in principe gelijk, omdat je in hetzelfde platform zit. Maar de impact is veel groter, omdat je veel meer data gebruikt. Plus, in combinatie met dat je aan het experimenten bent. Dus je kan, tijdens het ontwikkelen van zo'n algoritme, een bevinding doen dat bedrijven met veel vrouwen als werknemers, dat die een grotere verhoging in de premie zouden moeten hebben dan waar veel mannen werken. Dat soort analyses zitten daarachter, om als output te creëren. Later beslis je pas, nou dat vind ik niet ethisch om toe te passen, dat verwerp ik. Dat heb je niet in je gewone data marts zitten, dat is veel meer het doortrekken van sub informatie uit je data Lake. En in je AI-uitkomsten, zitten… ja hoe moet je het zeggen. Daar zitten een soort meningen in.

I: Hoe bedoel je?

R: Nou vanuit algoritmes krijg je veel meer voorstellingen. AI stelt bijvoorbeeld voor dat vrouwen meer premie moeten betalen. En van, nou ik weet helemaal niet of ik dit wil.

I: Zou dit dan ook makkelijker te interpreteren zijn als het meer transparant is? Zodat je weet waar de AI de beslissing op baseert?

R: Ja, inderdaad. Het black box scenario weer.

I: Zijn er naast explainability, ethiek, eenvoudige technologie en data security nog componenten binnen AI-beleid?

R: Uhm, ff kijken. Nou, ik denk data minimalisatie. Dat hebben we ook niet benoemd bij die andere. Dat je altijd probeert om de impact te verkleinen door zo min mogelijk data te gebruiken. Dus de meest simpele techniek, maar ook zo min mogelijk data. Een simpel voorbeeld: Als de afdeling een lijstje nodig heeft met welke dossiers ze moeten verwerken, dan hebben ze een polis nummer nodig en de oudste datum. Misschien ook nog de laatste medewerker die hem heeft aangeraakt en het type mutatie, maar daar heb je geen NAW-gegevens van de klant voor nodig.

I: Dus het voorkomen van onnodige data delen, data minimalisatie.

R: Ja, dat noemen we data minimalisatie inderdaad.

I: Naast dit laatste component, zijn er verder nog AI-componenten die u kunt bedenken?

R: Hm, nee.

I: Oké, dan wou ik graag verder gaan over algemene beleid componenten.

R: Hoe bedoel je dit precies? Algemeen?

I: Nou dit zijn vooral componenten die niet specifiek toe behoren tot databeleid component, maar wel belangrijk zijn als een soort ondersteuning van data & AI beleid. Een voorbeeld is dus de cultuur die je opnoemde, dit is niet specifiek alleen voor data, maar geld voor de hele organisatie. Zijn er verder nog componenten die dan toebehoren aan dit 'algemene' beleid componenten?

R: Ja precies. Dan kun je denken aan toegangsbeleid. Dat gaat verder dan de data kant. Dus ik denk dat logisch toegangsbeleid belangrijk is. Bewaartermijn. Ontwikkelbeleid/applicatie ontwikkelbeleid.

I: Wat is dan belangrijk met betrekking tot ontwikkel en applicatie beleid? De standaarden die gehanteerd worden bij het gebruiken van applicaties?

R: ja, inderdaad de standaarden. Maar ook wat je daar gebruikt qua data flow, welke databases daar achter hangen etc. Maar ook welke keuzes gemaakt worden.

I: Dus de verschillende keuzes die gemaakt worden met betrekking tot applicaties?

R: ja, precies. Wat je bij ons ziet is dat ontwikkelaars het leuk vinden om nieuwe technieken te gebruiken. Dus die hebben een document managementsysteem gemaakt. Waar dan een mongo database achter hangt. Dat is een niet relationele database en is veel moeilijker om uit te lezen vervolgens en die data te gebruiken. Dat heb je helemaal niet nodig voor de snelheid van het systeem en dat is echt een slechte ontwikkelkeuze geweest.

I: Oké, zijn er verder nog meer van deze algemene componenten?

R: Ik denk algemeen securitybeleid. Dus wie toegang heeft tot ons gehele netwerk, werken we met Multi-factor authenticatie etc. Dat is echt iets wat in het algemene beleid staat wat doorwerkt tot de veiligheid van de gegevens. Denk aan een data lekprocedure, die hebben we laats ververst. Logisch toegangsbeleid. Informatie beveiligen en rolbeschrijvingen. Privacy beleid, AVG uiteraard.

I: Oké, dan heb ik nu opgeschreven voor databeleid:

Data security, met access als sub-component

Data quality/verantwoordelijkheid

De manier van opslag, waar wordt het opgeslagen

Reproduceerbaarheid, transparantie en documentatie

Meta datering en classificatie

Data lineage/ herkomst en oorsprong van data

Technische factoren met betrekking tot opslag en wetgeving

Data minimalisatie, voorkom data lek

R: Ja, dat klopt wel.

I: Oké, voor AI-policy heb ik:

Explainability, transparantie. Voorkom black box scenario, vertrouwen creëren

Voorkom discriminatie, ethiek

Eenvoudige techniek - > het doel moet passen bij de technologie om onnodige kosten te voorkomen

Security, vooral gericht op requirements van applicaties

Hoeveelheid data, minimaliseer dit. Verminder impact

R: Hm, dat klopt ook inderdaad.

I: Dan voor general policy componenten heb ik:

Cultuur, gewest gedrag, richtlijnen met betrekking tot gedrag, maar ook bewust wetgeving. Procedures voor handelingen

Toegangsbeleid/ rolbeschrijvingen

Bewaartermijn

Ontwikkelbeleid applicaties, standaarden en architectuur

Algemeen security

Privacy/Datalekprocedure

Informatiebeveiliging

R: Ja, ik denk dat we een mooi lijstje hebben.

I: Oké, dat is goed om te horen. Nu wou ik graag het lijstje afgaan en de componenten indelen op prioriteit niveau gebaseerd op het MoSCoW model, bent u hiermee bekend?

R: Ja, dat ken ik wel, Must, Should, Could en Would toch?

I: Bijna, de W staat voor Won't. Dus iets wat nu niet echt nodig is, maar misschien later wel.

R: Ah oké.

I: Dan benoem ik nu de componenten op, dan moet jij zeggen welke prioritiet jij vindt dat deze componenten hebben. Data security

R: Must

I: Data quality

R: Ja, dat is Should, dat word dus belangrijker. De access en security gaan iets omlaag. Het blijven altijd belangrijke dingen. Dus qua aandacht gaat de focus, als we meer volwassen worden, meer naar data quality.

I: Dataopslag, manier van opslaan

R: Should

I: Reproduceerbaarheid, transparantie en documentatie

R: Must

I: Meta datering/ classificatie

R: Could

I: Data lineage/ herkomst

R: Could

I: Technische factoren/wet- en regelgeving

R: Must

I: Data minimalization

R: Could

I: Explainability

R: Must

I: Voorkom discriminatie, ethiek

R: Must

I: Eenvoudige techniek dat bij het doel past

R: Should

I: Security binnen AI beleid

R: Must

I: Data minimalisatie binnen AI, verklein de impact

R: Could. Hm, nu we zo praten zie ik punten wel overlappen inderdaad.

I: Cultuur awareness

R: Must

I: Toegangsbeleid/rolbeschrijvingen

R: Must

I: Bewaartermijnen

R: Should

I: Ontwikkelbeleid applicaties

R: Should

I: Security algemeen

R: Must

I: Datalekprocedure gekoppeld met privacy

R: Must

I: Informatie algemeen

R: Wat was dat ook alweer?

I: Informatiebeveiliging beleid

R: Owja, ook Must

I: Oké, dan kunnen we nu deze lijst die we net opgesteld hebben, vergelijken met de resultaten van de analyses. Zoals je kunt zien, is hier een overzicht van alle componenten die uit de literatuur en praktijk zijn gehaald. Groen betekent dat er een grote overeenkomst tussen beide componenten is, blauw dat de overeenkomst er nog is, maar wel wat verschil tussen beide componenten. Geel betekent dat er een groot verschil in componenten is en oranje betekent dat deze component niet aanwezig is in de andere analyse. Ik zal zo laten zien waaruit elke component bestaat, maar heb je opmerkingen over dit huidige schema?

R: Nee, nog niet

I: Oké, top. Dan kunnen we alvast gaan kijken naar de lijst met sub-component waaruit elk component bestaat. Als we jouw lijst vergelijken met deze lijst, zijn er best veel overeenkomsten, maar als je zo kijkt naar de component en sub-componenten waaruit een data policy zou moeten bestaan, heb je hier dan nog aanvullingen die erbij moeten staan? Of componenten die vindt die er niet bij horen, maar wel in de lijst staan?

R: Ik denk dat dit uh…. Ik zit nog wel te kijken naar data confidentiality…. Waar zou de reproduceerbaarheid en documentatie zitten?

I: Als we verder naar beneden scrollen kan het teruggevonden worden bij algemene beleid componenten.

R: Owja, betekent dat wij dat niet hebben? Omdat hij in het rood staat?

I: Nee, dit betekent dat de literatuur hier niet over praat.

R: Ah, oké.

I: Nou, binnen de componenten van data lifecycle komt ook data storage voor, dit noemde jij ook al eerder. Dus dit komt ook overeen. Ook komt data quality en metadata voor, wat jij ook aangaf als

belangrijk. Echter waren deze niet altijd te vinden in de beleidsstukken, dus dit komt ook overeen met wat je eerder aangaf.

R: Ja, inderdaad.

I: Zijn er verder nog componenten die opvallen binnen data life cycle?

R: Nee, niet zo snel

I: Binnen roles and responsibilities is ook te zien dat de literatuur data ownership als belangrijk aanduidt, echter dit hebben de beleidsstukken niet besproken.

R: Ja, klopt. Dit is ook iets wat ik nu in mijn beleid beschrijf. Waar ik een makkelijke stel regel heb. Je hebt een eigenaar van de applicatie. Deze is verantwoordelijk voor het inrichten van de applicatie, de juiste technologie wordt gekozen enz. We hebben dan ook mensen die data invoeren, en deze zijn dan ook verantwoordelijk voor de kwaliteit van de ingevoerde data.

I: Oké, dus dit gaat dan weer om medewerkers zoals *Naam van een medewerker*, bij het afdelingsoverleg in Leusden?

R: Ja, inderdaad. Dat team is dan ook verantwoordelijk voor hen invoer. Als er dan fouten worden gemaakt, is dat team verantwoordelijk en moet die fout oplossen.

I: Oké. Als we verder kijken naar de componenten van AI-policy, zien we ook dat er wat componenten zijn die jij ook eerder opnoemde. Met name, de explainability en ethiek (AI-principles).

R: Ja, inderdaad

I: Als we kijken naar componenten zoals general AI, zie je dat veel van deze sub-componenten niet voorkomen in beleidsstukken.

R: Nou, ik denk dat “Continious change” gebruikt kan worden in AI-models. In principe refresh je je model omdat er zoveel verandering is. Een voorbeeld is een algoritme dat fraude kan detecteren, maar als fraudeurs dit door hebben gaan ze andere methodes gebruiken. Dan zou je eigenlijk je model continue moeten retrainen.

I: Oké, dus de Continious change zou eigenlijk ook onder AI-models passen?

R: Ja, die hoort daar inderdaad wel bij.

I: Verder kwam data quality ook voor, alleen dit kwam voor in de context van AI-models, in plaats van waar jij het noemde, databeleid.

R: Oké, dat klink ook wel logisch

I: Zijn er verder nog opmerkingen met betrekking tot de AI-policy componenten? Aangezien deze best veel overeenkomen met wat je eerder zelf hebt samengesteld.

R: Nee, ik zie inderdaad veel overeenkomsten en niks wat mij opvalt.

I: Oké, dan gaan we door naar de laatste dimensie: de general policy componenten. Zoals je kan zien, zijn er hier ook veel groene vlakken, dus veel overeenkomsten tussen de analyses. Als we jouw lijst ernaast houden zien we ook weer wat overeenkomsten, zoals documentatie, maar ook design met betrekking tot architectuur en risico’s identificeren. Toegangsbeleid/rolbeschrijvingen komt ook terug, maar zit hem vooral in de stakeholders roles and responsibilities. Klopt mijn analyse zover?

R: Ja, die klopt. Wel raar dat documentatie niet in de literatuur terugkomt.

I: Ja, ik vond het ook raar. Maar ik denk dat documentatie in beleidsdocumenten apart benoemd wordt, terwijl het in de literatuur meer onderdeel is van. Zo kan het onderdeel zijn van data usage (descriptie).

R: Ja oké, dan kan het inderdaad ook verwerkt zijn in de production/creation of collection/acquisition. Wat ik in mijn beleid vooral in de lifecycle zet.

I: Ja inderdaad. Als we dan weer verder kijken naar algemene policy componenten, denk ik dat employees wel een shout-out verdient. Aangezien het een erg belangrijk onderdeel is en komt ook wat overeen met jouw cultuurcomponent. Zie jij binnen dit component nog iets wat je opvalt of niet duidelijk is?

R: Nee, inderdaad. Ik zie wel wat overeenkomsten tussen de gedragingen die gewenst zijn en dit employees component. Dit zie ik echter ook terug in Ethiek.

I: Dat is net waar ik het over wou hebben. Ethiek vanuit de literatuur gaat meer over definiëren van ethische standaarden, terwijl dat in beleidsstukken meer ingaat op procedures and methodes.

R: Inderdaad

I: Misschien dat dit component samen met employees dan ook gelijkwaardig is met je cultuurcomponent.

R: Ja ik denk het wel, die awareness is gewoon erg belangrijk.

I: Als laatst wou ik je nog de lijst van prioriteiten zien. Zie jij hier nog apartheden in? Zijn er componenten die jij als lagere prioriteiten zou stellen?

R: Nee, ik zie niet zo zeer iets apart. Wel denk ik dat het belangrijk is dat de prioriteiten verschillen per fase. Zoals nu is security erg belangrijk en dus een Must have, maar in de toekomst kan dit verschuiven naar een Should have.

I: Oké, top. Dan wil ik je bedanken voor je tijd en voor dit interview.

Appendix 6: transformation codes interview 1

Table 39 shows what interviewee 1 has mentioned regarding the components. The red marked text refers to the comments made after showing the results from both analyses.

Table 39: Overview mentioned topics interviewee 1

Dimension	Components	Priority
Data policy components	Data security, outgoing data, data access	M -> S
	Data quality and responsibility	S -> M
	Storage, method of storage	S
	Reproducibility, documentation	M
	Meta dating, classification	C
	Data lineage	C
	Technical factors regarding laws and regulations	M
	Data minimalization, prevent data leakage	C
	Data ownership	M
AI-policy components	Explainability, no black box, trust	M
	Ethics; prevent discrimination	M
	Technology should fit goals, prevent high costs	S
	Security (requirements applications)	M
	Data quantity, reduce the impact of risk, minimalization data	C
	Continuous change should be added to AI-models	-
General policy components	Culture, desired behavior, and increase awareness	M
	Retention periods	S
	Development policy regarding applications. Standards and architecture	S
	General security, including access policy and role descriptions	M
	Data leakage procedure and privacy	M
	Information security general	M
	Culture and employee & ethic component might be connected	-
	Priorities of components change during lifetime organization	-

Context of table 39:

As can be seen from table 39, the interviewee has mentioned security as an important component, which focuses mainly on data that is sent out. He mentioned that access is also a component, which is closely related to data security and is important in a policy. However, their priorities are not always the same. He mentioned that due to the transition of the platform, this priority is changing from a Must, to a Should. He explains that while the platform is created, security measures are implemented, and this should take priority. However, after the platform is created, these measures are already implemented. Therefore, security should not be a major problem since the measures should prevent any major problems (Appendix 5). While it is not a major problem anymore, it should still be something that gets enough attention, therefor it gains the Should priority (Appendix 5). For data quality and responsibility, the prioritization will increase after the platform is finished. Data quality should be high, and this responsibility should be assigned. However, while this does have a high priority now, it gains the highest prioritization after the platform is finished (Appendix 5). This is because this will be the main task of the data team. He also mentioned that data storage is important. The policy should determine where data is stored and how this is done, this would, according to the interviewee, make access policies easier to use (Appendix 5).

Another component that has been mentioned is reproducibility and documentation. The interviewee mentioned that he wants results to be reproducible. For results to be reproducible, documentation requirements should be made, which increases transparency (Appendix 5). He marked this as a Must since important knowledge could get lost if there would be a change of functions. The interviewee also mentioned that metadata and classification are important for documentation. He reasons that this

could help give a better overview of the data. This is also connected to the data lineage since the data lineage often is used in determining the classification of data and how it can be used (Appendix 5).

Technical factors regarding laws and regulations are also mentioned as a Must. This is mostly the translation of specific laws and regulations within a data policy. The interviewee mentioned the place of storage within the EU as an example of a law that needs to be translated (Appendix 5). The last component that has been mentioned is data minimalization, to prevent a data leak. The interviewee mentioned that a policy should motivate the reduction of data so that only the necessary data is used. He mentioned an example of data requests done by employees, in which employees often gained way more data than they requested (Appendix 5). Data minimalization could reduce the sent data and thus, reduce the impact if a data leak were to happen.

Regarding the AI components, the interviewee mentioned Explainability as a top priority. He mentioned that regardless of the technique, an algorithm should be explainable and so should the results (Appendix 5). He wanted to dodge the use of a black box scenario and keep the trust of the customer by being transparent. Therefore, he also mentioned ethics and the prevention of discrimination as a priority. He gave an example in which an algorithm was used to calculate the chance of sickness absenteeism (Appendix 5). The result was that female employees had a higher chance, which causes an ethical dilemma about the usability of this result. Another component within AI-policy components is that technology should not just be used, but it should fit with the desired goal and prevent unnecessary costs. The interviewee gave an example of a technology that is being used, where every extra calculation costs more money, so a simple calculation that results in an answer to the question is desired (Appendix 5). He also noted that when AI engineers are working with their tools, they often do so in separate environments, outside of the platform. This is endangering the data they work with, so data security within AI is also a problem, specifically the requirements on what tools and applications are being used. The interviewee mentioned data quantity as the last component within AI policies, specifically reducing the quantity (Appendix 5). He explained that a big quantity of data is being used within AI algorithms. This big quantity is increasing the impact of potential data leakage and is thus not desired. An AI policy should address potential methods to reduce data quantity (Appendix 5).

Regarding the general policy components, culture is perceived as very important for the interviewee (Appendix 5). He mentions that employees should show specific behavior and that they should be aware of the data they are working with. He envisioned that a policy should describe certain desires and what certain stakeholders should be aware of. One of these is that data engineers should be aware of retention periods, the GDPR, and its requirements. Another is that employees should be aware of the risks that come with sharing documents filled with sensitive data, but they should also know what to do in scenarios like these (Appendix 5). The interviewee also mentioned that within this culture, failures need to be shared to learn, but also to show that something has gone wrong.

Furthermore, retention periods and a development application policy are also identified as components of data & AI policies (Appendix 5). Retention periods are different from data retention periods, in that retention periods also apply to the retention of certain documents. Regarding application development, it is about the standards and requirements an application should have to be allowed within the architecture of the organization.

Another general policy component is general security. This is about certain measures to ensure the security of systems, such as multi-factor authentication, but also a logical access policy (Appendix 5). He also mentioned procedures for data leaks and certain privacy-related laws and regulation policies. This is mostly about how the GDPR is translated within a privacy policy. The last component the

interviewee mentioned is a general information security policy (Appendix 5). This information is not just data in storage but could also be sensitive information about the organization or its employees.

The interviewee gave his validation for the components within the data policy dimension and saw a lot of similarities with his answers (Appendix 5). One thing he noticed is the lack of data ownership within the practice and how he sees the importance of this. He is also trying to implement this within future policies.

Regarding the AI policy components, explainability and AI principles were very similar to what the interviewee answered, so nothing was noted about those. He did have something to say about the sub-component continuous change. He thought this was related to the AI models since due to this change, AI models need to adapt to this change (Appendix 5). He gave an example of a scenario where fraudsters know what methods a company is using to track them and how they would likely change their methods. Companies should thus retrain their models to adapt to this change.

He also made a connection between his culture component and the employee and ethic component. The last note he made regarding the list is about how priorities change during the organization's lifetime, such as his priorities for data security and data quality (Appendix 5).

To better analyze the input from the interviews with the analyses, table 39 is reorganized. The result can be seen in table 40. The explanation behind these transformations can be seen in table 41.

Table 40: Reorganized components list interviewee 1

Dimension	Components	Sub-component	Priority
Data policy components	Data security	Data security	M -> S
		Data access	M -> S
		Privacy; data minimalization	C
		Classification	C
	Data lifecycle	Data quality	S -> M
		Data storage	S
		Data lineage	C
		Documentation	M
		Meta dating	C
	Roles & Responsibilities	Responsibilities	S -> M
		Data ownership	M
AI-policy components	Explicability	Explainability	M
	AI-Principles	Non-maleficence	M
	General AI	Align technology with goals organization	S
	AI-models	Continuous change	S
	Privacy	Data minimalization	C
General policy components	General guidelines	Stakeholder roles and responsibilities	M
		Procedure	M
	General policy information	Priorities components change	-
	Employee	Education & Training	M
		Communication	M
	Ethics	Ethics standards	M
	Privacy	Data leakage procedure	M
	General security	Development policy regarding applications	S
		Access policies	M
		General information security	M
		Security requirements application	M
	Retention periods	Retention periods	S

Table 41: Explanation of transformation process interviewee 1

Dimension	Components	Falls under:	Priority	Explanation

Data policy components	Data security, outgoing data, data access	Data security	M -> S	Mentions data security and data access, both related to the data security component
	Data quality and responsibility	Data lifecycle and Roles & Responsibilities	S -> M	Data quality is part of the data lifecycle. Data responsibilities are part of Roles & Responsibilities
	Storage, method of storage	Data lifecycle	S	Data storage is a sub-component of the data lifecycle
	Reproducibility, documentation	Data lifecycle	M	Reproducibility and documentation are not yet part of any component. However, they might fit under the data lifecycle because the data lifecycle is about the entire lifecycle, in which documentation might also play a part.
	Meta dating, classification	Data lifecycle and data security	C	Meta dating falls under the data lifecycle. Classification falls under data security
	Data lineage	Data lifecycle	C	Data lineage is about the history of the data, which falls under the data lifecycle
	Technical factors regarding laws and regulations	x	M	Can be removed, since law and regulation are translated into policy guidelines such as privacy, security, and storage requirements. These are not part of a specific component.
	Data minimalization, prevent data leakage	Data security	C	This falls under data security, specifically a method of privacy. The method can be used to reduce the impact of a data leak.
	Data ownership	Roles & Responsibilities	M	Data ownership is part of the Roles & Responsibilities component
AI-policy components	Explainability, no black box, trust	Explicability	M	Explainability is part of the explicability component
	Ethics; prevent discrimination	AI-Principles	M	Preventing discrimination falls under the non-maleficence sub-component, which is part of the AI-principles
	Technology should fit goals, prevent high costs	General AI (Fit organization)	S	Falls under general AI, since it is about how AI technology should align with an organization's goal
	Security (requirements applications)	General security	M	While the interviewee did explain this scenario within the

				AI context, they also noted it under the general components. This shows that it also affects more than AI, which is why it is placed under the general policy component. Furthermore, this is about the security of applications, which falls under general security.
	Data quantity, reduce the impact of risk, minimalization data	Data security	C	Again, data minimalization is a method to reduce the impact of a data leak, which falls under privacy. Thus, this falls under data security within AI.
	Continuous change should be added to AI-models	AI-models	-	It is an addition to the AI-models component
General policy components	Culture desired behavior and increase awareness	Employee	M	The culture, its desired behavior, and awareness are closely related to the training & education sub-component, which falls under the employee component.
	Retention periods	Data lifecycle	S	Retention periods are part of the data lifecycle
	Development policy regarding applications. Standards and architecture	General security	S	The interviewee talked about making sure applications are secure, by setting requirements for applications. This falls under general security, which is about securing the organization.
	General security, including access policy and role descriptions	General security and general guidelines	M	General security is about securing the organization. Access policy is about only giving access to authorized people, which helps secure the organization. This is why it is part of general security. Role descriptions are part of general guidelines since general guidelines have a sub-component which is the Roles & Responsibilities of stakeholders. This also includes the descriptions of these roles.
	Data leakage procedure and privacy	General guidelines and privacy	M	General guidelines include the sub-component procedures, which fit the data leakage procedure. Data leakage is also

				connected to the privacy component.
	Information security general	General security	M	Information security is about protecting organizational information, which is a part of the general security of an organization. It is different from data security, which solely focuses on data that is stored. Information security focuses on the protection of the organization's information, regarding systems, but also employee information.
	Culture and employee & ethic component might be connected	Employee and ethics	-	This was already discussed in the "culture" component and is a confirmation of the earlier assumption of the connection that culture is connected to employees. However, this can also be added to ethics, as was noted during the interview.
	Priorities of components change during lifetime organization	General policy information	-	An organization's priorities change during the lifetime of an organization, which is connected to some of the sub-components of general policy information, such as context-dependency and managing change. Because of these connections, and the changing of priorities is general policy information, it is placed under this component as a new sub-component.

Appendix 7: Interview 2 transcript

I = Interviewer

R = Respondent

I: Het doel van het interview is om meer inzicht te krijgen in uw perspectief omtrent cruciale elementen in data & AI beleid.

R: Dus het gaat eigenlijk over wat ik belangrijke onderwerpen vind die in het databeleid staan?

I: Ja, dat is dan opgedeeld in drie delen. Databeleidscomponenten, AI-beleidscomponenten en General beleidscomponenten. Waar General vooral ondersteunend is voor Data en AI-beleid.

R: Ja. Dingen die ik zelf belangrijk vind zijn natuurlijk vanuit mijn invalshoek als *Functie* en ik ben *andere functie*, dus ik vind privacy ook belangrijk. Wat ik dan vooral aan data terug wil zien is hoe we data stromen beheren, dus hoe komt de data binnen, hoe verwerken we die vervolgens en hoe gaat de data er weer uit. Hoe zorgen we ervoor dat op een veilige manier gaat. Maar ook dat we aan wettelijke vereiste voldoen, zoals de AVG, maar ook eisen die vanuit meer sectorspecifieke programma's terugkomen. Zoals werkprogramma risicobeheersing volmacht, dat is een programma voor de volmachten, onder andere het mandaat. Voogd is ook een volmacht. Die stellen ook allemaal eisen over back-ups, veilige opslag van data. Dus dat zijn al dingen die ik belangrijk vind dat ze terugkomen. Dat we goed vastleggen hoe we die zaken beveiligen, vooral ook die data transitie. Dat we vaak wel weten hoe we dingen moeten beveiligen, maar die transitie vergeten we vaak. Ook, daarmee samenhangt, wat wij ook nog echt niet doen, is data classificatie. Op basis van die classificatie kunnen we natuurlijk bepalen wat voor maatregelen we moeten nemen om die data veilig te houden. Want hebben we het over bedrijfsgegevens, of NAW-gegevens van klanten of hebben we het over bijzondere gegevens zoals medische gegevens, strafrechtelijke gegevens. Dit zijn ook allemaal gegevens die wij verwerken, want wij hebben een Arbodienst. Dan moeten we natuurlijk wel zwaar classificeren en zwaar beveiligen. Maar classificatie beleid mist dus en eigenlijk mist ook een automatische toepassing waarmee documenten zouden worden geclassificeerd. Dat is dan ook iets wat ik op dit moment mis en ook heel erg belangrijk vind. Meer in algemene zin, is de Scope, waar zijn we van. We zijn erg groot. Zijn we daarmee als staff datateam verantwoordelijk voor alle entiteiten, of is dat onze scope? Is het databeleid van toepassing op alle entiteiten of is het alleen van toepassing op de entiteiten die alleen op onze ICT draaien. Want we hebben ook entiteiten zoals Assicuro die hun eigen ICT doen, vinden wij ook dat die onder de scope vallen? Het is ook in hoeverre je verantwoordelijk je bent. Verder is beleidsevaluatie ook belangrijk, want je hebt al kunnen zien dat sommige beleidsstukken erg oud zijn, of ze worden jaarlijkse ondertekend door een directeur: "Het is goed". Maar het is ook belangrijk dat die up-to-date worden gehouden, want er zijn heel veel ontwikkelingen op datagebied, securitygebied. Maar de risico's veranderen ook, want we hebben nu bijvoorbeeld Corona gehad, een oorlog tussen Rusland en Oekraïne die risico's met zich meebrengt en waar we ons op moeten aanpassen. Dus je moet jaarlijks gewoon kijken of we nog dezelfde risico's lopen en moeten we niet ons beleid daarmee op rangschikken. Dat is denk ik wel voor nu even wat in me opkomt van wat ik belangrijk vind.

I: Oké, een kleine samenvatting dan. Ik heb nu databeleid met inderdaad het databeheer en in de literatuur noemen ze dit de data lifecycle, waar eigenlijk de hele lifecycle van de data gedocumenteerd is en duidelijk is wat hiermee gebeurt. Dus ook opslag enz. Dan heb ik hier de veiligheid/security, dat valt denk ik buiten de lifecycle, maar is inderdaad ook belangrijk. Vanuit de literatuur en volgens u. Dan de AVG/wet- en regelgeving. Dus hoe dat vertaalt in beleid en de toepassing in Alpina. De data transitie, is dit met betrekking tot de fusie? Bedoelt u dat?

R: Ja, maar ook richting klanten. Hoe brengen klanten hun data naar ons. Bijvoorbeeld, wij mogen bijna in geen enkel geval een BSN verwerken, maar er zijn ook klanten die een foto van hun hele paspoort naar ons sturen, zonder iets afgeschermd. Wij hebben bijvoorbeeld Zivver voor beveiligd mailen. Maar in hoeverre wordt naar nou werkelijk gebruik van gemaakt. Wij sturen heel vaak gevoelige informatie via onbeveiligd mail naar onze klanten toe en dat kan gewoon niet. Ik zou het liefst een DLP-oplossing, of te wel een data leak prevention oplossing die ons al … de meest voorkomende data-lek oorzaken bij ons zijn dat ze naar de verkeerde persoon worden verstuurd. Alleen met een DLP-oplossing kan je al zoveel problemen voorkomen. Dat is dan ook wat ik bedoel met de data transitie van hoe ontvangen wij dingen goed beveiligd en hoe hebben wij dingen zelf beveiligd? Nou dat hebben we al goed op orde, dat wij het zelf goed beveiligd hebben. Maar dat beveiligd ontvangen en vooral het beveiligd uitsturen naar derde partijen en klanten is wel een verbeterpunt voor ons.

I: Oké, sturen en beveiligen van info. Dat zal dan onderdeel zijn van security? Of is dat een apart component?

R: Nee, onderdeel.

I: Dan classificatie van data, zou dit ook onderdeel zijn van security?

R: Classificatieschema vind ik een apart component.

I: Dan toepassing van de classificatie had ik ook, maar dat hoort er denk ik bij dan. Dan heb ik ook de data-lek preventie toepassing opgeschreven, hoort dat erbij?

R: Ja

I: Dan heb ik in het algemeen componenten verantwoording/scope en toepassing.

R: Yep

I: En beleidsevaluatie en risk assessment had ik ook neergezet. Zijn er verder nog iets qua componenten?

R: Hm, niet wat ik zo snel kan bedenken.

I: Misschien iets qua AI-beleid? Bent u bekend op dit gebied qua beleid?

R: Nee, niet echt

I: Hebt u vaag idee over wat belangrijk zou zijn in dit gebied? Wat u belangrijk zou vinden?

R: We hebben daar natuurlijk ook wel te maken met wet- en regelgeving. Sivi checklist, Onbemenste toepassing. Dat de eisen die daar worden gesteld, wij dat ook meenemen in het AI-beleid. Wat er in principe instaat weet ik eerlijk gezegd niet, daar heeft *employee 1*meer verstand van. Maar in de volmacht tak wordt daar steeds meer aandacht aan besteed en die Sivi check is daarom ook ontwikkelt.

I: Oké, dank u wel. Dan heb ik ook nog een prioriteit lijst die de prioriteit van verschillende componenten aangeeft. Dit heb ik gedaan door gebruik te maken van het MoSCoW model, bent u hiermee bekend?

R: Uhm, nee.

I: MoSCoW staat kort voor Must have, Should have, Could have, en Won’t have. Wat dus eigenlijk de prioriteit aangeeft van dit heeft een erg grote prioriteit, dus het is een Must have. Dit is belangrijk, maar niet superbelangrijk, dus een Should have. Als we dan uw lijstje afgaan, bijvoorbeeld omtrent het

databeheer/ data lifecycle. Op welke schaal van Must have, Should have, Could have of Won't have zou u dit component prioriteren?

R: Must

I: Veiligheid/security?

R: Ook Must

I: AVG Eisen/wetgeving/regelgeving?

R: Must

I: Datatransitie richting klanten en data-lek prevention

R: Wat mij betreft ook een Must

I: Dan nog de classificatie?

R: De volgende was Could-be?

I: Should

R: Ow dan Should. Uh, ja Should

I: Dan de algemene componenten. Verantwoording/scope/toepassing?

R: Uh... Should

I: Beleidsevalutie/risk assessment?

R: Must

I: En dan van de AI-policy de wetgeving?

R: Ook een Must, wetgeving is altijd een Must

I: Ja, dat zag ik bij *employee 1* ook terugkomen. Dan zal ik even mijn scherm delen. Zie u mijn scherm?

R: Ja

I: Hier ziet u de verschillende componenten die uit de literatuur en praktijk analyses zijn gekomen. Zo zie ik al wat overeenkomst met uw antwoorden, dus dat is goed. Elk component bestaat uit sub-componenten die zal ik u straks laten zien. Maar eerst het kleine overzicht van de componenten. Ziet u tot nu toe iets wat u opgevallen is? Iets wat u denkt dat helemaal niet belangrijk is? Of iets dat er niet tussen staat?

R: Hm, nee nu nog niet

I: Oké, we lopen zo meteen door de verschillende componenten heen en misschien dat dit dan veranderd. De betekenis achter de kleuren is dat groen betekent dat de componenten bijna identiek zijn, blauw is dat ze erg op elkaar lijken, maar er kleine verschillen zijn, geel is dat er een groot verschil is in sub-componenten en oranje betekent dat dit component niet bij de andere analyse naar voren kwam. Zo kunt u zien dat Technical guidelines niet voorkomt uit de analyse van de beleidsstukken. Data policy bestaat uit drie sub componenten volgens de analyses, data security, lifecycle en roles & responsibilities. Wilt u dat ik de securitycomponenten uitleg of begrijpt u ze?

R: Ik begrijp ze. Voor mij is het wel duidelijk en logisch dat dit de sub-componenten zijn. Dus ja

I: Oké, dat is dan goed om te horen. Omtrent data policy is er wel een verschil tussen sub-componenten. Zo is er te zien dat er niet gesproken wordt in de beleidsstukken over metadata, terwijl dit wel zo is in de literatuur. Dit is ook het geval omtrent data production/creation en data collection/acquisition.

R: Wat bedoel je precies met de production/creation en collection/acquisition?

I: Data production/creation gaat vooral om data die gemaakt wordt binnen het bedrijf en wat daarmee gedaan wordt. Data acquisition en collection is dus echt het verzamelen van data en regels omtrent dat.

R: Dat verzamelen van data vind je wel terug in het privacy beleid, omdat wij een grond moeten hebben om dit te verzamelen. In de verwerkersovereenkomst komt dit ook weer terug. Dat is wel dat we hebben aan de verwerking van data. Ik geloof wel dat we niks hebben omtrent data creatie enz. Dat klopt. Dan moet ik wel zeggen dat data retentie iets is dat we wel opgeschreven hebben, maar niet naleven ofzo. Ik weet dat we aan de Heilbron kant een keer een test hebben gehad met als resultaat dat het moeilijk uitvoer is. Dus daar moeten we nog eens goed naar kijken. We hebben ook veel data die we al veel langer moeten opschonen of dat bijvoorbeeld… in 2018 trad de AVG in werking. Daarvoor hadden we veel kopie paspoorten verzameld en toen kwam de AVG in werking en mochten we die BSN niet meer verwerken, maar dat betekent wel dat toen die tijd wij geen opschoningsactie hebben gedaan. Dus voor 2018 hebben we nog heel veel BSN in ons systeem staan. Dus dat bewaarbeleid is er wel, maar er moet nog wel wat gebeuren omtrent de uitvoering.

I: Oké, dat is goed dat u dat zegt. Ik denk inderdaad dat er wel een verschil zit tussen wat opgeschreven wordt en hoe het nou werkelijk is.

R: Nee en dat is dus vaak het probleem in onze organisatie, dat er wel beleidsstukken zijn, maar niet bekend bij iedereen en dus ook niet worden nageleefd. En dat is nu ook met de fusie, want nu gaan we al die beleidsstukken herschrijven en dan komen we er toch wel achter dat er dingen in staan die we helemaal niet doen. Dus dat is altijd wel een moeilijk ding van beleid. Het is makkelijk om iets op te schrijven, maar om het te doen is een tweede.

I: Dat is inderdaad waar, misschien wel interessant voor straks, want het komt weer voor in een ander component. Om de data policy lijst af te maken, roles and responsibilities. Ik merkte een verschil tussen wat de literatuur bedoelde en wat de praktijk had. De beleidsstukken gingen vooral in op de stakeholders in het begin van het beleid, wat hun verantwoordelijkheden waren etc. Terwijl de literatuur ook doelde op het verantwoordelijk maken omtrent data ownership.

R: ja, de literatuur zegt meer wie is eigenaar van de data. Dat is sowieso iets dat eigenschap, iets wat erg mist bij data.

I: Is er een ontwikkeling gaande op dit gebied?

R: Uhm, nou wij zijn meer opzoek naar eigenaarschap van processen. Vanuit de AVG is het verplicht om DPIA's uit te voeren, privacy impact assessments. Maar ook vanuit security perspectief om business impact assessments uit te voeren, BIA's. Maar om dat te kunnen doen heb je de eigenaar van dat proces nodig en die missen we gewoon. Het is dan ook erg moeilijk, vooral omdat we net gefuseerd zijn. Dus qua data eigenaarschap weet ik niet zo goed. Maar vanuit mijn perspectief zijn we inderdaad bezig met proces eigenschap en dit is heel belangrijk.

I: Ik kan mij herinneren dat *employee 1* het ook over een soort gelijk iets had.

R: Ja, maar eigenschap is ook erg belangrijk. *employee 1* en ik hebben eigenlijk een tweedelijns functie, dus eigenlijk verantwoordelijk van de eerste lijn om dit goed uit te voeren. Dus als je iets constateert dat niet goed is, dan moet dat in de eerste lijn belegd worden om dit te herstellen. Maar wie is er verantwoordelijk in de eerste lijn? Dat weten we niet en dat is dan ook iets waar we vaak tegen aan lopen.

I: Oké, interessant. Is dat al vaker voorkomen? Dat er een probleem was en geen eigenaar?

R: Ja, dat kom ik super vaak tegen en is super frustrerend. Het is dan ook bij Heilbron een tijdje geweest dat we bij het compliance team de zaak zelf gingen oppakken, terwijl we eigenlijk een adviserende rol hebben richting de business. Een tweedelijns, terwijl de eerste lijn uitvoerend is. Nu proberen we die schakel te maken om meer bij de eerste lijn te leggen. Omdat er geen eigenaren waren hebben we het zelf maar opgepakt, maar dat is niet hoe de rolverdeling hoort. Dat is dus zo gegroeid en dat gaat nu tijd kosten om te op te lossen, om die bewustwording bij de eerste lijn te creëren dat zij verantwoordelijk zijn en niet meer bij ons neer kunnen gooien.

I: Oké. Dan de AI-policy componenten. Dit is vooral gedreven door ethisch gebruik. Explicability is vooral dat beslissingen uitlegbaar zijn, transparant.

R: Ja, dat er niet gediscrimineerd wordt.

I: Ja, soort van. Discriminatie is dan meer van de AI-principles, specifiek Non-maleficence. De eerste twee componenten is vooral ervoor zorgen dat AI-policy ethisch is. De onderste zijn meer gericht op praktijk en het vertalen van ethiek naar praktijk. Echter kwam ik dit niet echt tegen in het AI-beleid. Privacy kwam ook weer naar voren, vooral omtrent communicatie. Aangezien veel mensen een schrik reactie kunnen hebben op AI, dus het is belangrijk om dit goed te communiceren.

R: AVG verwijst hier ook naar hé, er mogen geen geautomatiseerde beslissingen worden genomen op basis van bepaalde kenmerken van mensen. Dat is specifiek genoemd.

I: Ja, er zit inderdaad ook een overlap tussen het sub-component van dat security. Wat ik wel zag in beleid stukken is dat AI-models wel benoemd werd, maar niet in de literatuur. Verder zag ik general AI niet in beleidsstukken voorkomen, maar dit kan ook komen door de lage prioriteit en dat het niet vaak benoemd werd.

R: Ja

I: Verder zijn er nog de general policy componenten. Zo is te zien dat er general guidelines zijn die overeenkomen met componenten die u opnoemde.

R: Een aanvulling hierop vind ik dat er ook benoemd moet worden wie er verantwoordelijk is voor het beleid. Ik weet niet of dat in ons beleid staat, maar…

I: Hm, ik zal het erbij zetten. Ik was wel onder de impressie dat diegene die het beleid opstelt, ook verantwoordelijk is voor updates. Is dit het geval?

R: Ja, dat is wel vaak hoe het gaat ja

I: Hebt u hier nog verdere toevoegingen bij?

R: Nee, nou ja het onderhoud en wie verantwoordelijk is voor het onderhoud. Met risk assessment is het wel belangrijk dus om dit risico’s goed te identificeren, want die veranderen continue. Dus als wij niet periodiek kijken welke risico’s er zijn, maken we ons kwetsbaar. Nu zijn natuurlijk allemaal mensen gaan thuis werken, dus als je niet kijkt hoe je daar jezelf op moet aanpassen, nou data gaat alle kanten

op. Het wordt mee in huis genomen en daar moet je jezelf we op aanpassen. Dus ik vind een risk based aanpak belangrijk.

I: Oké, dan zet ik er een M'etje achter van Must. Nou de volgende component, liability werd niet echt gezien als een prioritiet. Maar u geeft dus echt aan dat er prioriteit achter zit.

R: Ja, want als je daarnaar kijkt, weet je ook je security. Dat is aan elkaar gekoppeld.

I: Dat is eigenlijk een connectie die ik nog niet had gemaakt.

R: Het is natuurlijk wel zo dat in de security iedereen zen eigen aanpak heeft, maar de meest gebruikt aanpak is gewoon op basis van risico's te werken en zo je security in te richten. Dat is dan ook hoe we het graag bij Alpina willen doen, dus vandaar dat er wel een verschil in kan zitten.

I: Oké, goed om te horen, het is genoteerd. Design & Strategy is dus vooral omtrent inrichting en architectuur. De literatuur geeft dit een lage prioriteit, maar wat voor prioriteit zou u dit component geven?

R: Nou ja, in de situatie waar we nu in zitten, de fusie met het platform, is architectuur heel erg belangrijk. Want wat wij nu bijvoorbeeld missen is dat we gaan migreren naar het Voogd platform, maar we missen inzicht in wat er allemaal eigenlijk bij Heilbron zit, dus wat er moet er allemaal mee, wat missen we, wat moeten we uitbesteden, wat moet er eigenlijk opgepakt worden bij Heilbron en bij voogd opzetten. Omdat we geen zicht hebben in dit architectuur, kost dat heel veel tijd en misschien mis je dingen. Dus nu hebben we ook besloten om een architect te gaan aannemen, nou eigenlijk iemand die in dienst is architect maken. Omdat inzichtelijk te krijgen, want zo krijg je ook je datastromen inzichtelijk en waar je beveiliging moet toepassen en wat de toegangspunten zijn. En samen eigenlijk een architectuur en netwerktekening en zo krijg je alles inzichtelijk, wat zijn de toegangspunten, waar moeten we ons beveiligingen en dan heb je een compleet plaatje. Dat missen we nu wel, en dat is in de situatie waar wij nu zitten wel jammer en dat kost ons veel tijd.

I: oké, dus het ligt aan de fusie. Misschien dat na de fusie de prioriteit lager ligt?

R: Ja, als je het al gedaan hebt is het goed om bij te houden. Want zoals ik al zei, voor mijn vakgebied is het belangrijk om die datastromen te weten en dus de toegangspunten te weten, hoe kunnen ze binnenkomen, waar gaat het allemaal naar toe waar moeten we opletten. Het inzicht is heel erg belangrijk, want als je het toch gaat opstellen dan is het toch goed om bij te houden.

I: En om het duidelijk te hebben, welke prioriteit zou u aan dit component geven tijdens de fusie?

R: Tijdens echt een must

I: na de fusie?

R: een should, want dan heb je het aardig inzichtelijk en ik ga er dan vanuit dat er dan niet veel meer veranderd want tijdens de fusie heb je al beslissingen gemaakt waarmee je gaat werken qua applicaties en welke je gaat uitbesteden. Dan ga je echt kijken waarmee wil je werken. Dus ik verwacht na de fusie dat dat niet zo snel meer veranderd.

I: Hm, volgens u uitleg komt het meer overeen met de Could, waar het niet cruciaal is, maar wel onderhouden moet worden.

R: Hm, ja een Could dan. Nee, dat klopt daar heb je gelijk in.

I: Ik wil uw woorden niet verdraaien

R: Nee, maar je hebt er gelijk in

I: De general policy information zijn vooral losse componenten die ik niet heb gezien in jullie beleid, maar dit zijn dan ook vooral componenten die je gebruikt tijdens het bouwen van het beleid. Zo weet je bijvoorbeeld dat je regels in het beleid niet voor elk scenario gelden.

R: Dus meer dingen om mee te nemen in je achterhoofd.

I: Ja, de enige overeenkomst waren de best practices. Dat vond ik wel een leuke, een maatregel waar je rekening houdt met context. Wat misschien wel interessant is, is Access to laws and regulations. Zou u dit bestempelen als belangrijk?

R: Nou als medewerkers weten wat wet- en regelgeving is, dan geef je meer intrinsieke motivatie om het beleid te volgen en snappen ze het beter waar het beleid op gebaseerd is en wat de gedachte is. Als je mensen zomaar iets oplicht omdat het in het beleid staat dan, ja… "boeie, weet je wel, ik doe gewoon men werken en dat kost me extra tijd en heb ik geen zin in". Maar als mensen weten wat de wet- en regelgeving is, dat we boetes krijgen enz., dan geloof ik er wel in dat mensen meer intrinsieke motivatie krijgen om zich eraan te houden.

I: Dus dat ze een soort pop-up krijgen: "Waarschuwing als dit verkeerd gaat, betaal je 5.000 euro".

R: Haha, ik snap wat je bedoelt

I: Misschien dat access to laws and regulations dan ook wel onder de volgende component kan worden gezet Employees. Dit kan dan onder communicatie komen te staan?

R: Ja, denk ik ook

I: Nou ja, de analyses vinden employees erg belangrijk. Het is vooral over de toepassing, hoe het beleid invloed heeft op de medewerkers. Hoe het gecommuniceerd wordt, dit is te doen door educatie & training. Reinforcement is dan het koppelen van acties en gevolgen. Dat iemand ontslagen wordt, of dat er bijvoorbeeld in het beleid staat dat als iemand de regels over treedt, diegene dan meteen wordt ontslagen of zoiets.

R: Ja, je zou dan een algemeen sanctiebeleid kunnen hebben omtrent overtreding van de regels van Alpina.

I: Is dat er? Een algemeen sanctiebeleid?

R: Nee, dat is er volgens mij niet. Het is bij mij niet bekend, dus ik denk het niet. Maar ja ik denk inderdaad dat medewerkers inderdaad erg belangrijk zijn, want je kan van alles opschrijven maar als medewerkers zich er niet bewust van zijn of het wordt niet goed gecommuniceerd of ze voeren het niet uit, dan is het beleid gewoon… dan heb je het document gewoon voor niks. Zij zijn dus echt het belangrijkste, want zij moeten het beleid uitvoeren.

I: En voor de duidelijkheid, wat zou uw prioriteit omtrent employees dan zijn?

R: Ja een Must, als zij gewoon niet weten wat ze moeten doen, wat de regels zijn, dan wordt dus het beleid niet nageleefd en dan had je eigenlijk, in mijn ogen, net zo goed geen beleid kunnen hebben.

I: Dat is een goede, die uitspraak ga ik misschien wel quoten

R: Ja, want het is zeg maar… we hadden laatst iets met een beleidsstuk, er was iets gebeurt bij de IT-afdeling van Voogd wat in het beleid stond dat niet mocht. Toen zijn we met dat beleidsstuk naar de IT-afdeling gegaan en toen hebben we gezegd: "Maar het staat hier" en toen zeiden ze: "Maar wij zijn

hier niet meer bekend". Dus, snap je. Dat beleid is leuk dat we dat op papier hebben en als we een audit krijgen, dan krijgen we een vinkje want dan kunnen we laten zien: "Kijk we hebben dat beleidsstuk wat we moeten hebben volgens jullie". Maar als het dan niet gebruikt wordt in de organisatie ja, weet je. Het staat of valt met de werknemers die het moeten naleven.

I: Ja, zoals u het nu uitlegt klinkt het logisch. Het laatste component is eigenlijk ethiek. De literatuur definieerde dit is ethische standaarden, in de beleidsstukken werd dit besproken door te praten over procedures en methodes om dit toe te passen. Ik weet niet hoever deze methode gebruikt worden, want u gaf aan dat er een verschil is tussen wat er op papier staat en daadwerkelijk gedaan wordt?

R: Ja, daar hebben we nog wel een slag te slaan.

I: Oké, ik heb ook een interview met iemand van Ditzo, misschien dat hij ook een soortgelijk iets aangeeft.

R: Ik weet niet hoever je daar nog tijd voor hebt, maar wat daar misschien wel interessant is voor jou is dat we een internal audit team hebben. Zij toetsen beleid praktijk, dus in hoever wat in het beleid staat, wordt dat nageleefd door middel van controles kijken zij of we het beleid goed hebben geïmplementeerd. Ik weet niet of je dit nog kan toevoegen aan je onderzoek? Ik weet niet hoe realistisch dat nog is?

I: Nou het is goed dat u dat zegt, ik denk echter niet dat ik deze toetsing kan toevoegen aan mijn onderzoek. Maar het is misschien wel interessant voor een follow up onderzoek.

R: Maar je zou dus aanbevelingen kunnen opnemen. Internal audit heeft daar dus heel veel kennis over, tussen het bestaan van het beleid en de werking daarvan in de organisatie. Dus dat geeft wel goede inzichten, dus het is wel een goede aanbevelingen voor je vervolgonderzoek.

I: Oké, ik zou het opschrijven. Dan als laatst heb ik nog de prioriteiten lijst. U zei net dat employees een Must moet zijn, dus dat schrijf ik op. Is er echter nog iets wat volgen u veranderd moet worden in deze lijst?

R: Privacy is wat mij betreft een must, omdat het wet- en regelgeving is. Omdat je hier een boete van krijgen, maar ook reputatieschade. Verder ben ik het wel eens met deze lijst. Employee en privacy. *Liability more hier ook bij, maar is gemist. Verander liability naar must*

I: Oké, top. Hebt u verder nog vragen?

R: Nee, het was duidelijk.

I: Oké top, dat betekent dat ik iets goed heb gedaan.

R: Ja, nee het is goed dat dit onderzoek gedaan wordt en ben heel benieuwd naar je verslag.

I: Jup, eind mei hoop ik het af te hebben. Verder wil ik u graag bedanken voor uw tijd en input.

R: Jij ook bedankt en succes nog met de laatste lootjes.

I: Dank u, dan wens ik u nog een fijne koningsdag morgen.

R: Ik jou ook.

Appendix 8: transformation codes interview 2

Table 42 shows what interviewee 2 has mentioned regarding the components. The red marked text refers to the comments made after showing the results from both analyses.

Table 42: Overview mentioned topics interviewee 2

Dimension	Components	Priority
Data policy components	Data management, data lifecycle	M
	Data security, outgoing data. Data transition towards customers.	M
	GDPR requirements, laws, and regulations	M
	Classification data, connect measurements, apply classification	S
	Focus on process ownership instead of data ownership	M
AI-policy components	Laws and regulations. How to comply with these	M
General policy components	Scope, responsibilities, and application policy	S
	Policy evaluation	M
	Risk assessment	M
	Include who is responsible for the evaluation of policy	-
	Design & Strategy, specifically the architecture	M -> C
	Access to laws and regulations under the employee component	-
	Employee	M
	Increase awareness to increase compliance	-
	Priorities components change	-

Context of table 42:

The first component she mentions is data management, or as the literature defines it, data lifecycle (Appendix 7). She specifically talks about data flows and how they are managed within the organization. They start from when they enter the company and end when they leave the company. The problem she has with the data flows is the lack of visibility, she thinks an overview of the different data flows within Alpina is key. Another component that is closely related to the data flows is the data transition, which is about the incoming and outcoming flow of data from and to customers (Appendix 7). She gives an example where customers sent entire photos of their passwords, but they do not cover up excess sensitive information. This in turn leads to questions on what employees should do with this. Another example is the usage of a secure mailing tool, for sensitive data. Both these examples show the importance of securing the data transition. Another component of data policy should be the classification of data (Appendix 7). On basis of the classification, the measurements needed to secure data can also be identified. This way, sensitive data that needs more drastic security measures, can be classified as such, and measurements to ensure stronger security can be taken. These components are part of why data security is important. She is mostly focused on how the data flows, when they leave the company, are secured. She states that the most frequent data leaks within Alpina are from sending the data to the wrong recipient and that this could be prevented using a Data leak prevention solution (Appendix 7). This should also include a data leakage procedure, for when it goes wrong.

Regarding AI, the interviewee mentions that she is not up to date on what an AI policy should contain and that interviewee 1 is focusing on that at the moment (Appendix 7). She did mention that it is key to comply with laws and regulations, but also a checklist from SIVI, which applies to all financial institutions. This checklist is on AI and automated decision making and this should be included within an AI policy (Appendix 7).

The interviewee mentioned that, regarding general policy components, defining responsibilities, scope and the application of a policy are key (Appendix 7). Since Alpina has grown quite big, it is increasingly more important to specify the departments a policy applies to. Besides this, it is also important to

specify the responsibilities within a policy. This ties together with the scope, “are we (Alpina) responsible for that organization?”. Interviewee 2 (Appendix 7) also mentions that policy evaluation is important, but this was not always the case for Alpina. Policies were not even evaluated, or directors just signed without evaluating the policy in enough depth (Appendix 7). The evaluation is especially important in data and security policies since there are a lot of developments within these areas. Another component that is of importance is the risk assessments (Appendix 7). She stated different risks, such as Corona and the war between Russia and Ukraine, that have an influence on Alpina and thus pose threats. She mentioned that a yearly review is needed as a minimum to assess if a policy is still up-to-date and if it covers the risks in a sufficient manner (Appendix 7).

The interviewee had no comment on the data security component and found the sub-components logical (Appendix 7). Regarding the data lifecycle, she mentioned that data collection and acquisition could be found in the privacy policy since the GDPR states that this is necessary. She also mentions that there is currently nothing regarding the creation of data (Appendix 7). Data ownership was also mentioned, but she does not think the focus should be on data ownership, but on the ownership of processes. From a data security perspective, certain assessments are mandatory from a law perspective. However, to perform some of these assessments, process owners are needed (Appendix 7). Therefore, this is regarded as very important, since without these process owners, assessments cannot be made and thus the company could break the law. Another problem that occurs, when there are no process owners, is that whenever there is a problem within that process, there is nobody responsible to fix that problem. This often meant that the compliance team had to fix it, which was not part of their job description (Appendix 7). Therefore, Alpina tries to fix this, by making sure a ‘first-line’ is made. The ‘first-line’ is an employee who is responsible for that process. The compliance team then acts as a ‘second-line’, which is an advising role towards the ‘first-line’. There were no further comments regarding the data policy components (Appendix 7). There were no notes about the AI-policy components. She only explained that the GDPR also has some explanations regarding automated decision-making on certain characteristics (Appendix 7).

Regarding the general policy components, the interviewee did have some comments. She thought that the general guidelines should also include who is responsible for the policy and updating this. She also noted that within Alpina, a Risk-based approach is used to identify risks and based on those risks, make, or adapt a security policy. Therefore, she thinks liability is a Must-have component (Appendix 7). However, she did note that this could be different in other organizations. She also wanted to change the priority of the design & strategy component. She reasons that during the fusion of the platform, architecture is key and thus a Must. This would change to a Could after the fusion of the platform (Appendix 7). Regarding the general policy information, the sub-component “Access to laws and regulations” was more connected to the employee component (Appendix 7). The argument for this is that when employees have access to laws and regulations, their intrinsic motivation to follow the policy guidelines increases. They now know what the policy is based upon and thus have more motivation to follow the policy guidelines (Appendix 7). The interviewee also marked the employee component as a Must-have. She stated that if employees do not know what to do, and what the rules are, the policy will not be complied with, and, from her perspective, there is no use in having a policy. She gave an example of a case in which the IT departments made a mistake and the compliance team confronted them. The IT department answered that they were not aware of the policy, thus they could not know they made a mistake. The policy is completely dependent on whether the employee complies with the policy, without this, a policy is useless (Appendix 7).

The interviewee also noted that there is often a difference between what is written down and what is happening. She gave the retention periods as an example of something that is written down but not

practiced within Alpina and that this is a point of improvement for them (Appendix 7). She thinks that awareness is the main problem, employees are not aware of the policies, which was also proven with the IT department case. She also stated that there is a difference regarding ethics, the difference is between what is written down and what is performed (Appendix 7). Finally, she indicated that the priorities of these components change for an organization based on what phase they are in now (Appendix 7). For her, this would be the change of priority regarding design & strategy.

To better analyze the input from the interviews with the analyses, table 42 is reorganized. The result can be seen in table 43. Table 44 contains the explanation of this transformation.

Table 43: Reorganized components list interviewee 2

Dimension	Components	Sub-component	Priority
Data policy components	Data security	Data security	M
		Classification	S
	Data lifecycle	Data management	M
	Roles & Responsibilities	Process ownership	M
General policy components	General guidelines	Scope & Goal	S
		Responsibilities stakeholders	S
		Policy maintenance	-
		Audit & Compliance	-
	Risk assessment	Risk assessment	M
	Design & Strategy	Architecture	M -> C
	General policy information	Priorities components change	-
	Employee	Communication	M
		Education & Training	M

Table 44: Explanation of transformation process interviewee 2

Dimension	Components	Falls under:	Priority	Explanation
Data policy components	Data management, data lifecycle	Data lifecycle	M	Data management is part of the data lifecycle component. The data lifecycle is about managing and documenting the data through its lifecycle.
	Data security, outgoing data. Data transition towards customers.	Data security	M	Data security regarding the transition is part of data security.
	GDPR requirements, laws, and regulations	x	M	Can be removed, since law and regulation are translated into policy guidelines such as privacy, security, and storage requirements. These are not part of a specific component.
	Classification data, connect measurements, apply classification	Data security	S	Data classification is a sub-component of data security.
	Focus on process ownership instead of data ownership	Roles & Responsibilities	M	Is part of the Roles & Responsibilities but changes the data ownership to process ownership.

AI-policy components	Laws and regulations. How to comply with these	x	M	Can be removed, since law and regulation are translated into policy guidelines such as privacy, security, and storage requirements. These are not part of a specific component.
General policy components	Scope, responsibilities, and application policy	General guidelines	S	Scope and responsibilities are both sub-components of the general guideline component
	Policy evaluation	General guidelines	M	Policy evaluation is similar to audit & compliance but looks specifically at the policy and how it is up-to-date. It is also similar to maintenance policy since an evaluation could be part of the maintenance.
	Risk assessment	Risk assessment	M	Risk assessment is a component on its own, a synonym is "liability".
	Include who is responsible for the evaluation of policy	General guidelines	-	Evaluation responsibility falls under general guidelines since it is part of the maintenance, part of the role and responsibilities of stakeholders, and part of the audit & compliance.
	Design & Strategy, specifically the architecture	Design & strategy	M -> C	Design & strategy is a component on its own.
	Access to laws and regulations under the employee component	Employee	-	Access to laws and regulation is a method of communicating with employees but also educating & training them These are both parts of the employee component.
	Employee	Employee	M	The employee is a component on its own.
	Increase awareness to increase compliance	Employee	-	Part of the educating & training sub-component, but also the communicating component. Which fall both under the employee component.

Appendix 9: Interview 3 transcript

I = Interviewer

R1 = Respondent 1

R2 = Respondent 2

I: Aan jullie eerst de vraag: Wat vinden jullie cruciale componenten binnen data & AI?

R1: Allereerst, qua beleid, de meeste cases die wij tegen komen gaan vaak over strategie en hoe pak je dat aan tegenover de bedrijfsstrategie die er zijn. Dus dat is de focus, maar als je het over beleid hebt, is het belangrijk dat je nog verder kijkt dan dat. Als je vaak management interviewt krijgt je vaak een visie uiteraard, waar wil je met het bedrijf naar toe. Als je kijkt met beleid is dat nog fundamentelere voor het bedrijf zelf, nog meer langere termijn. Ook moet het los staan van de strategie van het bedrijf, waar wil het naar toe. Waar je het beste mee kan beginnen qua beleid is gewoon kijken wat in essentie doet het bedrijf. Wat is Heilbron, want hier gaat het over toch?

I: Alpina eigenlijk

R2: Oud Heilbron, Heilbron is intermediair kanaal. Ze hebben een volmacht licentie, en verzuim. Ze zijn gefuseerd met Voogd & Voogd en die zitten in de service providing. Die zijn samen gefuseerd tot Alpina groep.

R1: Ah, oké. Ten opzichte van het beleid is het belangrijk om te beginnen met het verdienmodel van het bedrijf of van de instantie. Wat in essentie gaat het om, om je beleid daarom te bouwen. Wat betekent dat concreet voor data, hoe je daar om mee zou moeten gaan. Daaruit krijg je automatisch een wens ten opzichte van de data. Wat drijft een onderneming, waar doet men het voor. Dat verdienmodel dat vraagt automatisch hoe je met data om zou moeten gaan, maar ook wat voor type klant bedien je. Dat zijn aspecten hoe je je beleid moet opstellen als basispunt.

I: Oké, ik snap waar u vandaan komt. Als je dan kijkt naar belangrijke aspecten of componenten binnen data beleid?

R2: Wat *respondent 1* probeert te zeggen. Is dat voordat je met je beleid begint, je moet weten waar je bedrijf van is, wat doet het bedrijf, wat voor gegevens leggen zij vast en wat voor belangen komen daarbij kijken. Als je dat als essentie neemt op wat voor bedrijfselementen dat raakt, dat is de kapstok waar je het aan op hangt. Dat vertrekpunt zou ik ook doen, eerst scherpstellen. De organisatie, waar zijn we van, waar staan we voor. Dat zou ik altijd als inleiding nemen bij men beleid.

R1: ja, dat klopt. Dat is je context waarvan je gaat starten.

I: Misschien goed om te zeggen, ik heb het onderzoek in drie delen opgedeeld. Data beleidscomponenten, AI beleidscomponenten en Algemene beleidscomponenten. Deze laatste dienen als ondersteuning voor de andere twee. Ik denk dan ook dat deze inleiding in elk beleid naar voren moet komen en dan ook onder de algemene dimensie valt.

R1: ja, in feite wel. Daarom is het ook krachtig dat dit gekoppeld is aan je data en AI beleid. Alles is context, als je het lokale waterbedrijf hebt leidt dat tot heel andere gewenste componenten in het beleid vergeleken met een basisschool.

R2: Dus het geen uitvoerig stuk denk ik, die context, dat is iets wat vaak zien bij andere organisatie. Dat is echt het vetrekpunt, hoe ziet jouw corporate organisatie eruit. Vanuit welke context kijken wij, als we vanuit Alpina kijken, volmacht etc. Dus over welke data heb je het over. Daar bestaat het beleid uit, maar je begint dus met context. Dat je zaken hebben die algemeen gelden verder op het beleid,

dus een soort algemene noemer, dat lijkt me wel handig inderdaad. Dat lijkt me goed dat je dat hebt aangebracht.

I: en als je die context duidelijk hebt, op gebied van data, dan probeer ik toch een beetje terug te gaan naar data beleid. Wat voor componenten, bijvoorbeeld data security, zijn er nog meer belangrijk op dat gebied?

R1: Welke componenten? Goede vraag, maar in het algemeen is het belangrijk om te kijken hoe mensen, zowel beslissers tot en met gebruikers, als producenten als consumenten van data, een bepaalde richtlijnen mee moeten krijgen, met hoe men moet omgaan met die data. Hoe moet men die data behandelen. Die basisprincipes moeten terugkomen.

I: dus, als samenvatting: richtlijnen omtrent data.

R1: ja, dat is een component. Dan heb je het over zaken zoals gegevensbescherming, maar ook de wens die de verschillende business units hebben voor het gebruik van data om effectief hun rol te kunnen vervullen binnen het bedrijf. Als 1 deel van het bedrijf als taak heeft om te zorgen voor de boekhouding, komen daar weer specifieke data wensen uit. Dus er zijn allemaal verschillende componenten binnen het bedrijf die een actor zijn die een activiteiten uitvoeren om een bedrijf succesvol te laten zijn. Die actoren heb je modellen voor. Die moeten voldoende terugkomen in het beleid. Misschien een beetje wollig, ik zit misschien te kijken of ik daar een plaatje van heb.

R2: ik zal ook even wat proberen samen te vatten, omdat het misschien wat onduidelijk nog is. Je begint in principe met het opstellen van je dataprincipes. Hoe ga je om met data. Dus die richtlijnen waar je het over had, volgens mij is dat in de theorie de data principes opstellen. Vervolgens moet je in je organisatie kijken wat voor gebruikersgroepen hebben, wat gebruiken ze en wat voor dingen zijn daarvoor belangrijk. Daar maak je en breakdown van.

R1: ja, ik moet wel zeggen dat mijn specialiteit zit bij data governance en data management. Dat is ook onderdeel van je beleid. Ik moet eerlijk bekennen, ik heb geen of nauwelijks ervaring op AI beleid. Mijn specialiteit zit vooral op data governance en management, dus hoe moeten mensen met data omgaan, wat zou je logische wijs in moeten richten om daar succesvol in te zijn, om dat te management. Ook een rollenmodel erbij, wie moet wat doen om het voldoende succesvol te laten lopen. Dat zijn ook beleidsaspecten.

R2: ja, sowieso een flink chapter data governance.

R1: ja, 9 van de 10 keer zou je daar wel op uit komen. Dat komt omdat data governance een super belangrijk aspect is om fatsoenlijk met je data om te gaan en ook het überhaupt bruikbaar te laten worden. Daarom is het vaak een zwaar onderdeel.

I: Om het even duidelijk te hebben, wat vat u precies onder data management?

R1: Ja, dan kan ik het beste het DAMA modelletje laten zien.

R2: ben je daar mee bekend?

I: Vaag, het zeg mij zo niks.

R1: Ik zal hem even laten zien.

R2: Dat gaat helpen, het is een model dat vaak gebruikt wordt door andere organisaties. Dit zal dan ook het grootste component zijn in het data beleid.

R1: Dit is het framework van DAMA-DMBOK. Het is een verzameling van best-practices. Dus een aantal knappen koppen hebben even gekeken naar de ervaringen van bedrijven en instanties en hoe hebben zij het gedaan, maar wat werkt het beste. Dit is eruit gerold, het DAMA wiel. Deze aspecten moeten allemaal aanwezig zijn om het tot een succes te maken. Deze aspecten moeten allemaal in een bepaalde mate aanwezig zijn om het succesvol te laten zijn. Het is vaak wat je ook ziet aan het beleid van bedrijven dat men er automatisch voor kiest dit te benoemen expliciet in het beleid. Dus conform DAMA-BMBOK, zijn dit stappen die we in ons beleid ingericht zien worden. Dus even heel kort door de bocht gezegd, dus als je geen fatsoenlijke data kwaliteit management hebt, maar je doet de rest wel, kom je een heel eind, maar uiteindelijk ga je tegen slechte data aanlopen en vastlopen. Doe je wel data kwaliteitsmanagement, maar als je data development over, is ook even open deuren intrappen, maar dan zal je vastlopen in non-functionals. Dit zal ervoor zorgen dat je geen fatsoenlijke tools zal kunnen voorzien voor mensen. Doe je geen security management, ook wel logisch, bescherming van je eigen data of die van je klanten, ook essentieel. Zo zie je dat alle onderdelen belangrijk zijn in je beleid en dat je hier wat mee doet.

I: Oké, wel interessant. In mijn literatuuronderzoek en beleidsanalyse is dit model niet naar voren gekomen, maar wel de verschillende onderdelen.

R1: Ik gooi het linkje wel even in de chat. Ik kan me voorstellen dat je dat wilt doornemen. Dit is een super mooi middel om te kijken of je alle onderdelen hebt en het kunt afdichten. De buitenste zijn data management componenten, de data governance is de mens kant. Dus dat pietje en jantje daadwerkelijk dat gaat doen. Dus dat je met elkaar afspreekt die doet dat, die dat. Ook regels, ik noem maar wat, dat je een afspraak met data security management. Voordat je een aanvraag voor analyse doet, moet je altijd een aanvraag doet via een bepaald loket. Jantje of pietje ziet hier en controleert dat.

I: Oké, heeft het nog een verschillende prioriteit? Is het zo data quality een hoge prioriteit is?

R1: Dat vind ik een goede vraag en ook een belangrijke vraag. Dat is puur afhankelijk van de context, waar we het in het begin overhadden. Dat je altijd begint met wat drijft een onderneming, waar gaat het om in essentie. Die bepaalt de waarde en prioriteit. Dus stel dat je bijvoorbeeld een waterbedrijf hebt, dan heb je een hele andere prioriteiten. Bij een basisschool kan ik me voorstellen dat je heel goed de privacy gegevens van je leerlingen beschermt. Andere zaken zullen dan minder zwaar zijn. Zo kan je met de architectuur op een basisschool bezig zijn, maar als er maar twee systemen zijn heeft dit een lagere prioriteit. Bij een waterbedrijf zal dit dan weer anders zijn. Hier zal meta en document management misschien meer prioriteit hebben. Als antwoord op: is het allemaal even zwaar? Nee ik denk het niet, het is heel zelden dat het allemaal gelijk is. De essentie van het bedrijf bepaalt de prioriteit.

R2: Dat is wel interessant als je het koppelt aan Alpina. We hebben veel verschillende dingen in huis. Je hebt een Arbodienst die ene heleboel gevoelige informatie registeren. Dus daar wil je waarschijnlijk veel meer op security achtige componenten letten en rekening mee houden. Maar je hebt ook vanuit de service providing een portefeuille niet veel gevoelige dingen. Dus die context is bij Alpina een uitdaging om in je beleid die context te verwerven. Dat is dan ook het uitdagende en interessanter bij het onderwerp Alpina.

I: Hm, inderdaad. Dat is ook iets wat ik bij andere interviews te horen kreeg. Zoals het veranderen van de prioriteit van de componenten als het platform klaar is.

R1: Ja precies, het is wel zo dat je wel heel scherp naar beleid moet kijken dat het… dat is ook wel belangrijk voor jouw stuk. Beleid moet boven strategie staan als het gaat om langere termijn. Het mag

niet zo zijn dat je over een half jaar, of een jaar, een beleidsaspect totaal niet of geen toegevoegde waarde meer heeft. Dat is een goede toetsing voor als je je beleid schrijft. Beleid blijft voor zeer langere termijn relatief actueel.

I: Misschien kan ik meer verduidelijk krijgen over hoe dit erboven staat, ik begrijp het nog niet echt.

R1: Ik kan wel even een voorbeeld geven. Klant van ons, Vodafone Ziggo. Een Telco zoals dat heet, dat bedrijf heeft gedurende de jaren verschillende strategieën gevolgd. Je ziet dat per jaar dat die koers wijzigt. Een koerswijziging is dat je bijvoorbeeld voldoende snel live gaat, met hogere snelheden. Voorheen was Vodafone Ziggo 1 van de snelste en was dat geen relevant ding in je strategie. Door de druk van glasvezel is Vodafone Ziggo gaan versnellen in hun snelheden van kabelinternet. Dat is iets strategisch. Als je gaat kijken naar beleid heb je … is het veel minder veranderlijk van strategie. Als KPN morgen zegt dat ze morgen iets baanbrekends op de markt gaan brengen en de competitie daarmee gaat beïnvloeden moet je je strategie aan gaan passen. Maar beleid zal dan niet veel veranderen. Beleid zal altijd langer meegaan. Beleid moet zijn dat je je netwerk altijd een bepaalde mate van stabiliteit is, voldoende redundant zoals dat heet. Dat de verbinding op orde blijft. Dat blijft ongeacht dat je je koers moet wijzigingen van strategie.

R2: Dus dat beleid waar *respondent 1* het over heeft, dat blijft onveranderd ondanks dat je strategie veranderd. Dus Alpina zegt we gaan super budget in de markt zitten, en alles digital doen. Dan zou je nog kunnen zeggen dat het beleid hetzelfde of minimaal veranderen moet.

R1: Nog een voorbeeld van Vodafone Ziggo, stel dat ze onder druk komen te staan dat de snelheden omhoog moeten, maar het beleid zegt dat de stabiliteit een minimale waarde heeft. Dus het kan dat het beleid en strategie elkaar steken, dan zou het beleid krachtiger zijn. Het beleid zou leidend moeten zijn hier.

I: Oké, maar in het begin zou u juist dat het beleid gebaseerd moet zijn op strategie van het bedrijf toch?

R1: nee, nee. Het heeft een onlosmakelijk verband met elkaar. Maar strategie is korter en kan snel veranderen. Een beleid moet je neer zetten dat het heel robuust is en niet snel kan veranderen, het gaat nog een langere termijn mee.

R2: Dus je schets een context van een bedrijf waarmee gaan we beginnen, dus niet de strategie hoe positioneert het bedrijf zich op de markt, maar wat is de core van wat zij doen. Dat is de context.

R1: Ja, wat je net ook zag in de DAMA model, die cirkel diagram. Als ons beleid zegt, conform DAMA gaan we het inrichten. Dan zeg je eigenlijk dit is het beleid, dit gaan we volgen, die diagram geven we verder invulling aan. Maar vervolgens kun je wel zeggen, dit accepteren we als beleid, maar in de uitvoering kan je wel zeggen hier willen we qua strategie naar toe. Dan kunnen we prioriteiten bepalen ook ten opzichte van die diagram. Dus wat sluit daar goed bij aan, maar beleid blijft hetzelfde, conform DAMA. Dit is ook zo voor bedrijven die volgens ISO werken. Ik zit nu ook bij een baggermaatschappij, die heeft ontzettend veel ISO normen, die ontkomen daar niet aan, dat is onderdeel van het beleid.

I: Ah oké, dan heb ik dat verkeerd gehoord. Dat model is inderdaad wel een hele goed en de connectie met mijn resultaten. Is het goed dat ik nu de componenten laat zien ik gevonden heb als resultaat van mijn onderzoek?

R1: Ja, maar ik moet er zo wel vandoor. Ik ben echter wel nieuwsgierig, dus ik meld even dat ik iets later ben.

I: Oké. Zoals ik net vertelde, zien we hier de drie dimensies waar ik het over had. Zoals u kunt zien zijn de groene dingen die overeenkomen met elkaar. Dan heb ik ook de interviews en hoe deze de componenten en sub-componenten hebben gevalideerd. Hier ziet u dan ook de componenten en waar elk component uit bestaat. Mijn idee was om te kijken of jullie iets zien wat niet volgens jullie hoort onder dat component. U zou net bijvoorbeeld het rollenmodel, dat komt denk ik veel overeen met de Roles & responsibilities. Het DAMA model komt denk ik vooral overeen met data security en data lifecycle.

R1: Dan moet ik even kijken, wat ik wel zou doen, voor AI geld dit misschien niet zo, maar voor data kan je het DAMA lijstje pakken en die afvinken. Voor AI weet ik dit niet.

R2: Mijn kennis qua AI is ook slecht hoor.

I: Oké, ondanks dat, als u hier iets kwijt over wilt mag dat gerust.

R2: Die drie onderdelen van data beleid, heb je die zelf bedacht?

I: nee, die komen van de literatuur. Data security is iets wat veel naar voren komt, data lifecycle is iets wat in de literatuur naar voren komt, maar in het beleid anders vertaald is.

R1: Ik zou eigenlijk… die je net noemde wel onder data management zetten. Uhm… metadata

R2: Ik denk inderdaad dat een heleboel componenten in DAMA-DMBOK staan. Heb jij dingen die je echt mist vanuit je kennis van DMBOK?

R1: Ja, ik zit even te kijken. Ja ik zie wel onderdelen, je hebt bijvoorbeeld data access, data availability bij security benoemd. Eigenlijk is data availability onderdeel van data kwaliteit. Data kwaliteit kent een heleboel aspecten, een daarvan is data availability. De kwaliteit van je data wordt bepaald door availability.

I: Oké, vanuit de literatuur wordt het naar voren gebracht als: zorg voor back-ups etc.

R1: Bij data kwaliteit is een van de aspecten data availability, dat kent in totaal 7 aspecten, betrouwbaarheid, toegankelijkheid, beschikbaarheid… uhm en nog wat andere componenten.

R2: Ja, inderdaad. Ik denk wel dat het waardevol is om het DMA-DMBOK gebruiken en inlezen. Misschien zie je dan iets zoals Master data management, wat ik zo snel niet terug zie komen.

R1: Ik moet gaan, maar vind het wel echt super interessant. Geef aan als je meer vragen hebt. Ik wens je sowieso veel succes.

I: Dank u voor uw tijd en fijn weekend alvast.

R2: Wat mij opvalt is dat je, als je bijvoorbeeld naar de eerste kolom kijkt van de literatuur, als je kijkt naar de data kwaliteit, vallen daar een paar aspecten onder. Bijvoorbeeld betrouwbaarheid, maar dat is wel in die hiërarchie, dan moet je een paar dingen doen in je beleid. Dat moet je even kijken, maar die moeten wel terugkomen in je beleid.

I: Weet u toevallig de andere aspecten ook?

R2: Ik zal even kijken of ik een linkje kan vinden, want het is iets wat elk bedrijf eigenlijk wel heeft. Ik heb helaas echter niet iets liggen.

I: Ik zal het anders zelf even opzoeken.

R2: Dat is inderdaad wel handig.

I: Dan wou ik nog heel even iets verder gaan qua beleid. AI, hebt u iets wat u belangrijk vindt omtrent dit beleid?

R2: Ik denk dat een paar algemene principes gelden, dus opslag, bewaartermijnen, mag ik deze data opslaan. Dus dat hele privacy gedeelte. Met AI wil je innoveren, dus je gaat voorspellingen doen, maar je moet wel afvragen, mag ik dat en wil ik dat. Zeker als je het gaat gebruiken, dus de data ethiek, dat is een stuk complexer en dat moet je in je beleid terug laten komen. Dat is ook voor data beleid, maar dat is zwaarder in de nieuwe technologieën.

I: ja oké, dat is eigenlijk iets wat terugkomt in je AI-principes.

R2: Wat mij verder opvalt, als je kijkt naar je AI beleid, zie ik een paar aspecten zoals privacy. Als je die in je AI-beleid neer zet zou je die ook in je data beleid neer moeten zetten.

I: Als we iets naar beneden gaan, zien we privacy bij de algemene componenten. Zo is privacy belangrijk in beide AI en data. Daarom valt het onder algemene componenten.

R2: Waar ik verder aan zou denken, buiten Alpina. Ik ben gewend dat DPIA's gedaan worden, ik weet niet of die term je wat zegt?

I: Ja, dat ken ik

R2: Als je die verandering organiseert. Je wil een proces hebben om een DPIA op te stellen, met een beoordeling daarvan en die ook goed te keren, vaak ik een architectuur setting. Dan zie je ook vaak een architectuur board die dat bewaakt conform beleid. Dus je wil een soort assessment van die DPIA's. Wat ik ook wel soms bij bedrijven is dat er een soort data usage board is, of je het recht hebt om die data ook echt te gebruiken vanuit een geldig doel en grondslag. Dus hoe gaan we om met AVG-wetgeving en hoe vertalen we dit naar onze organisatie.

I: Oké, ik denk dat die laatste vooral valt onder privacy.

R2: Je hebt dan ook verschillende beleidstukken en het is ook belangrijk om te verwijzen, wat doe je waar. Dat is dan ook wel interessant.

I: Oké, dan wou ik heel even nog een paar andere componenten doorlopen. Jullie hadden het net over de algemene organisatie en introductie, met de context. Deze staat onder deze componenten (wijst naar general policy information/others). Echter is dit meer iets wat achter in het hoofd gehouden moet worden. Verder zijn er general guidelines zoals u kunt zien, zoals scope bepalen, maar ook definities.

R2: Ja zeker, definities is zeer belangrijk.

I: Verder is er nog de liability component. Een andere respondent gaf aan dat dit erg belangrijk is bij Alpina, omdat hierop de data security ingericht wordt. Merk je ook dat andere bedrijven, of de verschillende bedrijven waar jij toezicht op hebt, dezelfde prioriteit hebben?

R2: Ik denk dat andere bedrijven dat veel hoger hebben staan dan Alpina. Ik denk dat zeker, overal waar persoonsgegevens worden verwerkt, daar wordt gewoon heel scherp opgelet. Daar wordt beleid voor geschreven en uitgevoerd. Ik ben gewend dat daar een veel grotere organisatie om heen is gebouwd bij andere organisaties. Bij Alpina zijn dit twee of drie personen die dat bewaken en controleren, dus die data governance, dat ben ik veel zwaarder gewent.

I: Oké. Ik heb ook gehoord dat de implementatie erg belangrijk is. De vertaling van het beleid is erg belangrijk.

R2: hoe je dat terug kan laten komen in je beleid, is extra aandacht geven aan je governance. Hoe ga ik in mijn organisatie die ook daadwerkelijk beheersen, wat voor uitgangspunten neem ik mee in mijn beleid. Dit wordt driemaandelijks gereviewed. Misschien niet zo concreet, of bij elke verandering waarbij persoonsgegevens betrokken zijn wordt een DPIA uitgevoerd. Structurele inrichting zaken, dus hoe ga ik het beheersen, daar zou ik aandacht aan besteden.

I: Dan is het ook de truc om daar balans in te vinden.

R2: Inderdaad. Maar ook bij jouw rollen en verantwoordelijkheden, jij hebt alleen de rollen en verantwoordelijkheden, maar het gaat ook verder dan dat. Hoe gaan die rollen en verantwoordelijkheden met elkaar zorgen voor het proces van hoe wordt dat geborgd, dat is je governance.

I: Oké, dus hoe wordt nagegaan dat iedereen zich aan zijn verantwoordelijkheid houdt.

R2: ja, maar iedereen heeft waarschijnlijk een rol en verantwoordelijkheid. Dus pietje heeft A en Jantje B, maar ze willen ook nog C of iets met een organisatie laag of iets anders, dat is allemaal inrichting van je organisatie. Maar wel het model van hoe beheersen wij het, dat is redelijk abstract, uitvoerende controleren macht enz. Maar je moet wel het governance model scherp hebben. Dat kan ook zijn dat governance geregeld en de organisatie heeft een governance model ingericht om bepaalde dingen te bewaken. Dat staat dan in je beleid. Dan wil je dat beleid uitvoeren en checken is er een proces opgesteld, doen ze wel wat ze moeten doen. Dat staat dan niet in je beleid, maar dat wil je wel door vertalen vanuit je beleid. Dus een soort governance-uitgangspunt, dat zou ik adviseren om daar strakker op te zitten.

I: Oké, uhm. Dan wou ik nog even snel employee component bespreken. Dat is eigenlijk de invloed van het beleid op de medewerker, hoe je dat communiceert, dus de awareness, educatie en training, hoe vertaal je die dingen. Zorg ervoor dat ze makkelijker risico's herkennen, en dan de reinforcement. Wat doe je als ze zich niet aan de regels houden.

R2: Nou die herken ik allemaal. Die awareness is wel een ding. Volgens mij ben je zelfs verplicht om beleidsstukken toegankelijk te maken voor al je medewerkers.

I: Ja, dat kwam inderdaad ook voor in de beleidsstukken.

R2: De grootste feedback is dus pak dat DAMA-DMBOK erbij en kijk even of je wat mist.

I: Ja dat ga ik zeker doen. Bedankt voor uw tijd en input. Ook alvast een fijn weekend.

R2: Dank je, jij ook.

Appendix 10: transformation codes interview 3

Table 45 shows what interviewee 2 has mentioned regarding the components.

Table 45: Overview mentioned topics interviewee 3

Dimension	Components	Priority
Data policy components	DAMA-DMBOK model	-
	Model for roles	-
	Guidelines revolving around using data (Data principles)	-
	Data quality has 7 underlying components	-
	Safeguard roles and responsibilities	-
AI-policy components	Privacy (Storage, retention periods)	-
	Ethics	-
General policy components	Privacy	-
	Base policy on the core (activities) of the company	-
	Context depended priorities	-
	An intro that discusses companies core	-
	Data governance should be used to safeguard policy application	-
	Translate GDPR laws and data usage	-

Context of table 45:

Two parties, that do not work for Alpina, have been interviewed. These parties are experts in the field of data policy within the insurance world. This party consisted of two interviewees who both worked together and helped several other companies in forming their data policies.

The interviewees (Appendix 9) mentioned the DAMA-DMBOK model as an important model that gives the components a data policy should have, specifically within the data security and lifecycle components. They also mentioned that this model is used by a lot of organizations (Appendix 9). The DAMA-DMBOK model mentions 10 components (International, DAMA, 2017). These 10 components are: Modeling & design, Storage & operations, security, integration & interoperability, document & content management, reference & master data, warehousing & BI, metadata, quality, and architecture (International, DAMA, 2017). The interviewees (Appendix 9) mentioned that these components can be compared to the results from the analyses, which have been done in appendix 11. The results from this analysis will be discussed in chapter 4.3, the comparison between both analyses and the interviews.

Data quality is also mentioned by the interviewees, but they stated that data quality has seven underlying components, one of these is data availability. Other components they mentioned were reliability and accessibility (Appendix 9). They did not know all seven components at the time of the interview, however, authors such as Cai and Zhu (2015) mention the several aspects of data quality. These components are similar to sub-components from data security and data lifecycle. Another important aspect they mentioned is the description of data principles, specifically, the guidelines on the usage of data. Employees have to be aware of these principles and guidelines.

Regarding the roles and responsibilities within data policy, the interviewees validated the usage of a model with roles that specify the roles and their connected responsibilities. However, they also stated that these roles and responsibilities should be safeguarded and audited on whether employees fulfill these roles (Appendix 9).

Both interviewees noted that they do not know AI policy and its specifics. However, they did give some important aspects that might be of importance to an AI policy. They stated that within AI, privacy must

be important, specifically, the permission of whether data can be used and stored (Appendix 9). They also stated that ethics within AI is key, specifically the question of whether certain data can be used from an ethical point of view.

A component that was mentioned repeatedly and was emphasized is that a policy is dependent on the context (Appendix 9). This is also the reasoning for why they did not give a priority indication, the priority is also dependent on the context. It was also emphasized that a policy should be based on the core (activities) of a company, a short description of this should also be included in the introduction section of the policy.

Furthermore, privacy was again mentioned under the general policy components. The interviewees stated that privacy is important in both data & AI policies. They also mentioned that policies should include an explanation of how the GDPR laws and regulations are implemented, specifically regarding data usage (Appendix 9). This applies to both AI and data policies.

The last addition is regarding the employee component. The interviewees mentioned the importance of data management, but also governance (Appendix 9). Data governance is about making sure employees act according to what the policies state, employees have to comply with the policies' content. The interviewees mentioned that part of data governance is also about safeguarding this compliance.

To better analyze the input from the interviews with the analyses, table 45 is reorganized. The result can be seen in table 46. Table 47 explains the reasoning behind this transformation. Appendix 11 has the explanation and interpretation of the DAMA-DMBOK model, which is not included in tables 46 and 47.

Table 46: Reorganized components list interviewee 3

Dimension	Components		Priority
Data policy components	Data security	DAMA-DMBOK model	-
	Data lifecycle	Guidelines revolving around using data (Data principles)	-
		Data quality has 7 underlying components	-
	Roles & responsibilities	Model for roles	-
		Safeguard roles and responsibilities	-
AI-policy components	AI-principles	Ethics	-
	Privacy	Privacy (Storage, retention periods)	-
General policy components	General guidelines	An intro that discusses companies core	-
	Employee	Data governance should be used to safeguard compliance with policy	-
	General policy information	Context depended priorities	-
		Base policy on the core (activities) of the company	-
	Privacy	Translate GDPR laws and data usage	-
		Privacy	-

Table 47: Explanation of transformation process interviewee 3

Dimension	Components	Falls under:	Priority	Explanation
Data policy components	Data security	Data security	-	Self-explanatory
	Model for roles	Roles & responsibilities	-	Self-explanatory
	Guidelines revolving around using data (Data principles)	Data lifecycle	-	Data usage was already a sub-component of the data lifecycle

	Data quality has 7 underlying components	Data lifecycle	-	Data quality was already a sub-component of the data lifecycle
	Safeguard roles and responsibilities	Roles & responsibilities	-	Safeguarding was not part of the roles & responsibilities component, but it does fit within this component.
AI-policy components	Privacy (Storage, retention periods)	Privacy	-	Self-explanatory
	Ethics	AI-Principles	-	They mentioned AI ethics, within this research, this is called AI principles.
General policy components	Privacy	Privacy	-	Self-explanatory
	Base policy on the core (activities) of the company	General policy information	-	They mentioned that this is something that should be accounted for when making a policy, which fits the description of general policy information.
	Context dependent on priorities	General policy information	-	This was already a sub-component of general policy information.
	An intro that discusses the company's core	General guidelines	-	In addition to the general guidelines, the interviewees mentioned that this should be in the intro of a policy. This could be placed within the Goal & Scope, stakeholder section.
	Data governance should be used to safeguard policy application	Employee	-	Policy application and the translation of this fall under the integration of policy, which is part of the employee component.
	Translate GDPR laws and data usage	Privacy	-	Translating GDPR guidelines is part of the privacy component.

Appendix 11: Comparison of DAMA-DMBOK and the results study

Table 48 shows the components of the DAMA-DMBOK, what they mean, and whether they validate the findings from the literature and practice. The definitions are all based on the DAMA-DMBOK book (International, DAMA, 2017). The bold underlined components are, compared to the two analyses, new sub-components.

Table 48: DAMA-DMBOK comparison and explanation for connection to existing components

Components DAMA-DMBOK:	**Definitions (based on** (International, DAMA, 2017)).	**Explanation:**	**Connected to components:**
Data Architecture	Identify data needs organization, then design and maintain these needs. The goals are to identify data storage and processing requirements, and design structures to fulfill data requirements.	This overlaps somewhat with the design and strategy component. Identifying data needs is missing. This might be a step that is not described in the policies.	Design & strategy: - Architecture - Design - **Identify data requirements and needs**
Data Modeling and Design	Discover, analyze and scope data requirements, then present and communicate them through the data model.	This needs to be done in advance of making a policy. Partly discussed in data quality. Partly validated. Data requirements are similar to AI principles. Data requirements	Design & Strategy: - Design - **Identify data requirements and needs**
Data Storage and Operations	Design, implementation, and support of stored data to maximize value. The goals here are managing availability, ensuring integrity, and managing the performance of data transactions.	Availability is described in the previous analyses, the same applies to integrity. In this context, it goes under data quality. Managing performance of data transactions is missing.	Data lifecycle - Data quality (Availability & integrity) **- Managing performance**
Data Security	Define, plan, and execute security policies. Provide authentication, authorization, access, and auditing of data and information assets. The goals here are enabling and preventing access, complying with relevant regulations and policies for privacy,	These components are all described in the data security component. Audits and enforcement are described in the general guidelines and under the employee component.	Data security: - Data security - Data access General guidelines: - Audit & compliance Employee: - Reinforcement

	protection, and confidentiality, and ensuring the privacy and confidentiality of stakeholders. This should also be audited and enforced.		
Data Integration and Interoperability	Managing data movement within and between applications and organizations. The goals are here to provide data securely, identify meaningful events and triggers, and support BI and analytics.	Securing this is about securing the data flows and data lifecycle. Part is covered under the data lifecycle. Identifying meaningful events and triggers could be covered under risk assessment.	Data security: - Data security Data lifecycle: - Data flows Liability: - Identify risks
Document and Content management	Plan, implement and control activities for life cycle management of data and information. The goals are to comply with legal obligations regarding records, ensure effective and efficient storage of documents and content, and ensure integration capabilities between structure and unstructured content.	Legal obligations regarding records are covered under data retention. Effective and efficient storage is covered under data storage. The component "Integration capabilities" is missing.	Data lifecycle: - Data retention - Data storage **- <u>Integrate content between systems</u>**
Reference and Master data	Manage shared data to meet organizational goals, reduce risks of data redundancy, ensure high quality and reduce data integration costs.	Reduce data redundancy has been mentioned by interviewee 1 and falls under privacy. High quality falls under data quality. The other two are missing.	Privacy: - **<u>Data reduction</u>** Data lifecycle: - Data quality **- <u>Integrate content between systems</u>**
Data Warehousing and Business Intelligence	Plan, implement, and control processes to provide data support decisions and support knowledge workers in reporting, queries, and analysis.	Providing support, through control processes, for employees who make data-based decisions has not been mentioned in either analysis.	Data lifecycle: - **<u>Process around data-based decision making</u>**
Metadata	Plan, implement and control activities to enable access to high-quality, integrated metadata. The goals	The definition is covered under the general guidelines. The metadata component covers the others but might not	Data lifecycle: - Metadata - Data quality - Data usage

	here are to define business terms and usage, collect metadata from diverse sources, provide a standard way to access metadata, and ensure metadata quality and security.	discuss diverse sources. Security is covered under the security aspect.	General guidelines: - Definitions Data security: - Data security
Data Quality	Plan, implement and control activities that apply quality management techniques to data to make sure it is fit for usage. The goals here are to develop a governed approach to make data fit, define standards, requirements, and specifications for data quality controls as part of the data lifecycle, define and implement processes to measure, monitor, and report data quality levels, and identify and advocate for opportunities to improve the quality of data.	Data quality is not the same as data quality within the analyses. Data quality within the DAMA-DMBOK is bigger and has several sub-components. These are Quality management techniques to fit usage, define standards requirements, define and implement processes for measures, monitor and report quality levels, and identify & advocate for opportunities to improve the quality of data.	Data lifecycle: - Data quality

An overview of all (sub-)components that have been identified within the DAMA-DMBOK:

- Data security:
    - Data security
    - Data access
- Data lifecycle:
    - Data quality (Availability & integrity)
    - **Managing performance**
    - Data flow
    - Data retention
    - Data storage
    - **Integrate content between systems**
    - **The process around data-based decision making**
    - Metadata
    - Data usage
- General guidelines:
    - Audit & compliance
    - Definitions

- Liability:
    - Identify risks
- Design & Strategy:
    - Architecture
    - Design
    - **Identify data requirements and needs**
- Employee:
    - Reinforcement
- Privacy:
    - Data reduction

The effect of these components, compared to the other analyzes, is as follows:

- Adding the following sub-component:
    - Identify data requirements and needs (Under design & strategy component)
    - Integration between data (Under data lifecycle component)
    - Managing performance (Under data lifecycle component)
    - Process of supporting data-based decisions (Under data lifecycle component)
    - Data reduction (Under privacy)
- The following sub-components are similar to the components from the other analyses, but there are some differences:
    - Data quality (Under data lifecycle)
        - The DAMA-DMBOK model sees data quality as having different sub-components that ensure high data quality. These are within this overview sub-sub-components. Within the other analyses, data quality was not described as a component with several sub-components.
    - Metadata (Under data lifecycle)
        - Ensuring access to metadata, high quality, and security metadata. The DAMA-DMBOK describes metadata in more detail compared to the other analyses.
- And finally, the following sub-components are the same, or very close:
    - Data security (Under data security)
    - Data access (Under data security)
    - Data flows (Similar to data lifecycle)
    - Data storage (Under data lifecycle)
    - Data retention (Under data lifecycle)
    - Define data usage (Under data lifecycle)
    - Define terms (Under general guidelines)
    - Audit & Compliance (Under general guidelines)
    - Identify risks (Under liability)
    - Design (Under design & strategy)
    - Architecture (Under design & strategy)
    - Reinforcement (Under employee)
    - Context dependency (Under general information)
    - Priorities component context dependent (Under general information)

These components, together with other comments interviewees from interview 3 have made, have all been added to the total overview, which can be seen in appendix 14.

Appendix 12: Interview 4 transcript

I = Interviewer

R1 = Respondent 1

R2 = Respondent 2

I: Mijn onderzoek is omtrent data beleid & AI. In opdracht van Alpina group, ze hebben net een merger gehad waar twee grote bedrijven zijn gemerged tot Alpina groep. Ze hadden moeite met het duidelijk vormen van data beleid. Omdat ze ook geïnteresseerd zijn in AI, wouden ze dit ook toevoegen aan het onderzoek. Vanuit de literatuur is er ook geen duidelijk framework wat nou in zo'n beleid precies moet. Vandaaruit was er ook interesse van de UT, waar ik studeer op dit moment, om dit framework te vormen. Ik heb twee analyses gedaan, één analyse gericht op de literatuur en een analyse van de beleidsstukken van Alpina. Nou wou ik graag interview hebben om deze resultaten te valideren. Ik heb twee interviews intern gehad bij Alpina en 1 extern interview. Daar wou ik graag ook jullie validatie bij hebben.

R1: Oké, dus echt valideren van wat je nu hebt staan.

I: Ja, maar ook benieuwd naar jullie inzichten omtrent data beleid. Misschien zijn er wel dingen die jullie belangrijk vinden maar niet aanwezig zijn in mijn lijst.

R2: Kan je me even helpen wat Alpina group doet?

I: Alpina group bestaat uit Heilbron en Voogd & Voogd voornamelijk. Ze zijn betrokken met verschillende verzekeringen zoals arbodienst, ziekte en nog wat anderen. Ik zit op dit moment in het data teams, dat zich vooral bezig houd met een algemeen platform van beide bedrijven.

R2: Misschien goed om even te zeggen, Heilbron en Voogd & Voogd zijn allebei business partners waarmee *Bedrijf* samen werkt, dus het zijn partners, maar ook even zoeken waar de concurrentiële informatie ligt. Dus mochten we daar tegen aanlopen, moeten we daar alert op zijn. Ik denk niet dat dit zal gebeuren zolang we op data beleid blijven. Maar toch wel even op letten.

I: Oké. Wat ik bij de andere interviews heb gedaan is dat ik eerst vraag aan die respondenten wat zij denken dat belangrijk is, zonder mijn resultaten te laten zien. Dit is zodat jullie niet beïnvloed worden om mijn resultaten te valideren, maar zelf ook komen met wat belangrijk is. Ik heb dit opgedeeld in drie dimensies. Data beleid, AI beleid en een algemene dimensie. De algemene dimensie is wat niet specifiek toebehoord dat data of AI, maar wel componenten die ondersteuning bieden. Een voorbeeld is evaluatie van beleid. Als we beginnen met data beleid, wat vinden jullie belangrijk omtrent data beleid?

R1: Zal ik dat doen?

R2: Ja begin maar jij zit er dichterbij dan ik

R1: ja, ja, dat is zeker waar. Data beleid, de termen zijn sowieso erg breed voor interpretaties. Wat je misschien moet realiseren dat bij *Bedrijf* we verschillende producten en verzekeringen hebben. De data zijn daarin verspreid. Wat we intern hebben is dat, altijd als we samenwerkingen aangaan, we overeenkomsten met elkaar opstellen om die data eventueel uit te wisselen met elkaar. Daar heb je verschillende vormen voor. Een DLO is een standaard data leveringsovereenkomst, waar we met elkaar in gesprek over gaan met een BRM'er, een business risk manager, en ook juristen om te gaan kijken is dit allemaal nodig. En ook zeker weten mag dit wettelijk gezien. Daarnaast ook een PIA, vanuit een privacy technische hoek, mogen we dit doen vanuit de privacy belangen van de klanten. En middels

een aantal van die documenten proberen we zoveel mogelijk te borgen dat dit binnen de belangen van de klant blijft, maar ook dat we ons werk kunnen doen.

I: Oké, en al deze overeenkomsten die je met elkaar sluit, zijn die ook vanuit beleid vastgesteld?

R1: Uhm nee, waar we sowieso met elkaar bezig zijn is om een zo breed mogelijk data science framework op te stellen, voor alle projecten dat framework wordt gehanteerd. Daar moet ik eerlijk zeggen zijn we daarmee aan het werk. Daar zit inderdaad die onderdelen in, dus alle aspecten van waar moet je aan denken als je van start gaat met een data science project of uitleveren van data.

I: Oké, en uitleveren is eigenlijk data versturen naar andere partijen?

R1: Ja, niet letterlijk versturen, maar dat je bij andermans bronnen kunt inkijken.

R2: Een aanvulling van het verbond van verzekeraars, een beleidsdocument hoe je om moet gaan met AI achtige toepassingen. Vanuit beleid geformuleerd. Daar zitten een aantal principes in die we toepassen. Ik weet alleen zo niet de naam uit men hoofd.

I: Oké, jullie zijn eigenlijk die richtlijnen van verbond van verzekeraars pas je dan voornamelijk toe in jullie AI beleid?

R2: Ik zou het eigenlijk vertalen in drie niveaus. Het hoogste niveau is dat van de verbond van verzekeraars, dat beleid. Dat geeft vrij generieke uitgangspunten van hoe je met AI moet omgaan. Dan heb je de PIA, waar je veel meer vanuit het privacy aspect kijkt, van wat doen we hier nu eigenlijk en is dat daadwerkelijk in het belang van de klanten. Dan heb je de DLO, de data leveringsovereenkomst, dat is veel meer de praktische kant van de data, waar we mee omgaan, wat is het, wat doen we ermee. Volgens mij staat er ook wel in wanneer we dingen weggooien, dus dat is veel meer de wat hardere kant, waar slaan we het op, wie heeft er toegang toe, hoe zorgen we ervoor dat niet meer mensen dan nodig, hoe zorgen we ervoor dat er niet meer data in zit dan dat we daadwerkelijk mogen gebruiken, snappen we wat voor data het is en waar het vandaan komt, zodat we geen dingen mee gaan doen die vanuit de PIA niet verstandig zijn. Dus als je niet weet wat de data daadwerkelijk is, is het heel moeilijk om daarna analyses erop te doen. Omdat er stiekem wel een bias op zit, waarvan jij niet eens weet dat die erin zitten. Al dat soort elementen probeer je op dit manier op een procedure af te richten. Je moet je realiseren dat wij naast alle polis data heel veel medische data in ons systeem hebben. Veel meer nog dan een Voogd & Voogd zal hebben. Omdat wij een zorgverzekeraar hebben, maar ook de inkomensbedrijf toegang heeft tot alleen medische data omtrent arbeidsongeschiktheid. Tot aan *Bedrijfs applicatie*, even als voorbeeld, waarin we veel inzicht hebben in het beweegpatroon van mensen, daar zit een bron van data om allemaal aan elkaar te koppelen. Dat is de gave kant. Tegelijkertijd allemaal dingen aan elkaar koppelen, dat is helemaal niet wat de klant per se verwacht dat we dat met de data gedaan hebben. Daar zitten de borging van drie lagen van besluitvorming, dat proberen we daarmee af te vangen.

I: Oké, die drie lagen zijn dus het verbond beleid, het privacy gedeelte, met het belang van de klant en dan als laatste het praktische gedeelte, hoe je het toepast.

R2: Ja

I: Oké, ik schrijf het heel even op hoor. Heeft dat prioriteit? Word het verbond beleid altijd voorop gesteld?

R2: Uh zeker. Nou ja, ik moet wel eerlijk zeggen dat het een vrij generiek beleid is. In de praktijk kun je niet zeggen, ik voldoe of ik voldoe niet. Het is meer een dialoog document op basis waar je met elkaar gaat kijken, doen wij verstandige dingen met elkaar? Het is geen vink lijstje, zo is dat ding ook helemaal

niet opgezet. Wik denk ook niet dat je er op die manier gaat komen met het data science project ben je continue bezig met iets nieuws bedenken, daar bestaat geen vink lijstje voor.

R1: Ja, dus dat zijn een beetje de basis wat er ligt. Wij zijn nu bezig met een data science framework te ontwikkelen, dus uniforme werkwijzen te generen om ook juist dit allemaal te borgen. Dat we aan de juiste ethische kaders wordt voldaan, aan al die voorwaarden. Dat er uniformiteit is in de tooling waar we mee werken. Dus dat dat allemaal zoveel mogelijk geborgd is. Wij zijn een centrale data science afdeling. Maar zo heb je ook in alle business units, dat je allemaal uniform werkt en de standaard hanteren.

I: Oké, dus dat data science framework is dan echt gericht op hoe vertaal je het beleid naar de werknemers.

R1: Ja, dus kijk. Wat je misschien realisteer, het beleid is wat hoger over. Op een gegeven moment ga je met wat concrete projecten aan de gang. En eigenlijk, waar moet je dan allemaal… bijna een soort stappenplan van zorg dat je aan al die bedrijfsformaliteiten voldoet door dat je dit framework hanteert. Dus meer een soort toepassing van het beleid in de praktijk.

I: Oké, interessant. Ik heb in een ander interview gehoord dat ze dat data governance noemen.

R1: Ja, dat is een beetje de overkoepelende term. En data governance gaat ook vaak meer over het beheer van de data zelf. Dit gaat eigenlijk ook over als je AI toepast, wat voor gevolgen heeft dat dan ook voor de klant. Wil je dat allemaal, voldoet dat allemaal aan de privacy voorwaarden. Goverance is dan ook vaak geborgd, als je een toepassing gaat realiseren: wat mag wel, wat niet. Denk aan een stukje profileren, wat mag daar wel en wat niet.

I: Oké. Ik weet niet of dit al gezegd is, maar hoe wordt dit beleid vertaalt naar medewerkers? Is dat dan een training die gedaan wordt?

R1: Uhm, ja ja. Eigenlijk hoe de organisatie, omtrent data science, in elkaar steekt is dat je dus … wij van centraal zetten elkaar neer als centre van excellence. Als er projecten gedaan worden bij de business units, helpen wij mee en proberen wij deze werkwijze mee te geven. Zo wordt dat framework verspreid in de organisatie. Zo worden die standaarden gecreëerd en ervoor gezorgd dat je aan alle juiste voorwaarden voldoet. Maar wat *Respondent 2* net zei, met het algemene beleid, dat is daar allemaal in doorgevoerd.

I: Oké, heel even kijken. Dan heb ik dat opgeschreven. Qua ethisch beleid, u had het net over een ethisch kader. Hoe is dat dan… is dat ook via de training aan te passen? Zoals met het data science framework, is dat in dezelfde manier opgesteld?

R2: Dat is mij niet helemaal duidelijk.

I: Het ethisch kader…. Uhm.

R2: Zo heet dat ook echt. Dat is de naam van het document, het heet geen data science of wat dan ook, analytics beleid ofzo. Het heet het ethisch kader.

I: Het ethisch kader, waar is dat op gebaseerd? Het verbond van verzekeraars?

R2: Wij hanteren inderdaad het documenten van het verbond. Daar hebben we geen apart document voor gemaakt.

I: Ah oké, ik dacht dat jullie het hadden gebaseerd op dit document.

R1: Nee, dat ethische is opgenomen in het framework en één van de vele aspecten. Maar er zit ook gewoon een checklistje in van: Hier moet je aan denken als je opstart.

I: Dit krijgen ze als ze bezig zijn met een project? Iets dat ze naast zich houden?

R1: Ja, wat wij doen is: Eerst voordoen, dan samen doen, dan zelf doen. Dus eigenlijk eerst proberen we samen een project op te zeten met de data science binnen de business unit. Dan houd je het erbij. Dan ga je het samen doen, houd je het er weer bij. Dan zelf doen, dan kunnen ze het er zelf bij pakken.

I: Oké, dan wou ik graag mijn resultaten delen. Ik weet niet of *Respondent 2* mee kan kijken?

R2: Nee, dat lijkt mij heel onverstandig.

I: Ja oké, dan zal ik er een beetje bij praten, dan kunt u ook wat meekrijgen. Kunt u het zien? Nou ik heb hier dus een voerzicht van verschillende componenten. Hier ziet u drie componenten onder data policy component, data security, data lifecycle en Roles & responsibilities. Dat zijn eigenlijk de drie hoofdcomponenten die ik gevonden heb die specifieke in data beleid horen. Bij AI beleid heb je explicability, dat is eigenlijk de mate in hoeverre AI uitleg baar is, maar ook transparant. Dan heb je AI-principles, dat is eigenlijk het ethisch gedeelte. Zorg je ervoor dat AI niet discrimineert, en zorg je ervoor dat iedereen eerlijk behandeld en zelf discriminatorische factoren weghaalt met AI. Do no harm. Dan heb je hier de technical guidelines, dat is rood want het komt in de literatuur naar voren, maar niet in de praktijk. Technical guidelines is eigenlijk hoe je die ethische dingen toepast, en hoe vertaal je die in dingen die mensen kunnen doen. AI moet eerlijk zijn, maar hoe zorg je ervoor dat AI eerlijk is. General AI is eigenlijk een soort algemene dingen die ingaat op…. Eigenlijk dingen die je in je achterhoofd moet houden. Als we kijken naar de in-depth, zie je dat AI afhankelijk moet zijn van de organisatie, en continu moet veranderen. Dit zijn niet snel dingen die je concreet kan opnemen in je beleid. Dan heb je AI-models, dat zijn specifieke richtlijnen op AI modellen hoe je hiermee moet omgaan. Dan heb je ook nog privacy, specifiek in AI. Echter als zie je privacy ook weer terugkomen in algemeen, en ik heb dan ook ondervonden dat privacy overal wel weer terugkomt. Want privacy is altijd belangrijk. Dan heb je een paar algemene componenten, dat zijn eigenlijk componenten die je gebruikt als ondersteuning van je beleid. Risk assessments, design & Strategy dit gaat voornamelijk over het verbinden van je beleid met je strategie. Employee is vooral te verbinden met jullie toepassing met hoe vertaal je je beleid naar de medewerker, hoe zorg je ervoor dat de medewerker hierdoor beïnvloed wordt, hoe zorg je ervoor dat hij of zij zen gedrag gaan aanpassen. Ook heb je nog ethiek. Dan heb ik een in-depth gemaakt, waar ik gekeken heb naar waar bestaat een component uit. Zoals voor data security heb je: Privacy, integrity, confidentiality, availability, data classification en access. Dat zijn eigenlijk alle componenten van data security, die volgens de literatuur en praktijk aanwezig moeten zijn in data beleid. En dat is dan voor elk component gedaan. Mijn vraag is, en dat kan overweldigend zijn voor jullie, specifieke voor *respondent 1*, zien jullie dingen die opvallen?

R2: Kan je even teruggaan naar het kleinere overview? Deze is erg groot en ik wil daar even rustig overeen kijken. Uhm, even kijken. Ja, wat ik … nou wat ik lastig vind is dat de woorden best wel generiek zijn. Je kunt heel veel dingen, zoals data security, kan super veel daar onder vallen. Ik wil wel als je dit met ons deelt wil ik er wel rustig overeen kijken, zeker in die in-depth comparison. Maar om dat nu zo te zeggen is het een te generiek beeld. Dan moet je eigenlijk die in-depth hebben, maar daar moet ik even rustig naar kijken.

I: Ja oké. Inderdaad, dat is misschien wel handig, dan zou *Respondent 2* er misschien ook naar kijken.

R2: Ja, dat lijkt me prima, als je daar tijd voor hebt om naar te kijken. Dan krijg je misschien een betere reactie dan zo vanuit de heup geschoten. De manier waarop je hebt uitgelegd klinkt alsof je er over na

hebt gedacht, het klinkt als een logische structuur. Maar om dan iets te vinden wat ontbreekt is dan wel een lastige vraag. Misschien kan het ons ook wel helpen in het opzetten van een framework.

I: Ja, dat is dan ook deel van mijn onderzoek. Dat ik eigenlijk een soort framework opstel waar anderen bedrijven in de verzekeringswereld ook wat aan hebben, niet alleen Alpina. Daarom houd ik de externe interviews om te kijken wat andere bedrijven belangrijk vinden.

R1: Ja, dat klinkt als een mooi streven. Ik ben wel eigenlijk benieuwd, dit zijn allemaal termen eigenlijk. Om van deze termen om naar echt beleid, om naar een echte lijst te komen, ga je die stap ook nog maken? Of is dit je eindproduct?

I: Nou mijn eindproduct is…. Ik heb afgelopen woensdag met mijn leraren gepraat over wat ik oplever. Wij zaten te twijfelen of ik een checklist maak, zo van: check check check. Of dat je dit vooral gebruikt als een inhoudsopgave.

R1: Ja, zelf, puur als ik … kijk alle termen ken ik natuurlijk en ik zou ook wel kunnen bedenken wat je ermee bedoelt. Maar dan moet je nog wel die vertaalslag maken met, denk hier hier hier aan. Dus als ik zelf aan een framework denk, denk ik aan: Nou data security, denk hier hier hier aan. Hier moet je keuzes opmaken in je beleid of framework. Dus dat, om direct al even een puntje te geven. Om dat direct naar een toepassing te brengen, zou dat voor mij veel toegevoegde waarde hebben.

I: Ja, ik denk echter wel dat dat moeilijker is. Soms, zoals ik al merk, iedereen ziet deze termen als iets anders. Er is al een groot verschil tussen wat mensen onder data lifecycle verstaan en wat ze in hun beveiligingsbeleid. En ik denk als je een algemeen framework opstelt dat dat voor bedrijven heel verschillend is, want elk bedrijf heeft andere prioritieten. Wat ik wel ondervonden is dat in de verzekeringswereld ze veel meer gefocust zijn op strategie en hoe je dat eigenlijk vertaalt in beleid.

R1: Ja, dat is wel lastig hoor. Maar dan wordt het wel sneller generiek toegepast, als je die volgende stap toepast. Ik moet naar een volgende meeting, heel veel succes met het afronden van dit, en stuur het maar op, dan ga ik er nog even naar kijken.

I: Ja top, dan stuur ik het naar de mail van *respondent 2*, als *respondent 2* het dan kan doorsturen naar u?

R2: Yes, doe ik. Helemaal goed, ik ga ook door, succes.

I: Bedankt nog en alvast een fijne dag.

Appendix 13: transformation codes interview 4

Table 49 shows what interviewee 2 has mentioned regarding the components.

Table 49: Overview mentioned topics interviewee 4

Dimension	Components	Priority
Data policy components	Data retention	-
	Data usage	-
AI-policy components	AI-Principles	-
General policy components	Data science framework	-
	Different levels of policy	-
	Ethics based on the national organization	-
	Communicate framework to employees	-

Context of table 49:

The second external party that was interviewed also works at an insurance organization. These are also seen as experts in the data field within insurance. This party also consisted of two interviewees who worked together, one of them was working in a data team. This interview differs from the rest since it was shorter (half an hour) and thus fewer insights could be gathered from this interview. Because it was shorter, the results from the research were only shortly discussed and most of the time was spent on the first part of the interview; asking what they thought was important within data & AI policies. This means there was almost no feedback on the results of the research. However, the interviewee did not have time for a follow-up within the researcher's time frame.

Interviewee 3 mentioned in the interview their use of a data science framework (Appendix 12). This is a framework that includes all aspects an employee needs to think of before they start a data science project or when they extradite data. This framework can be seen as a checklist or roadmap that should be followed, it this is done, all formalities within this company are fulfilled. Within this framework are all kinds of different aspects, they mentioned data retention and data usage as examples (Appendix 12). The way this data science framework is communicated to other employees is by using a three-step approach. The first step is showing employees how the framework works. After this, they use the framework together. Finally, the employee has to use the framework themselves, this way they learn how to use it. (Appendix 12).

They also noted that the AI policy they have is heavily inspired by the policy document that has been published by one of the national organizations within the insurance industry (Appendix 12). This document is about how organizations should handle AI-related applications. However, the interviewees did not know the name (Appendix 12). This is also the case for general ethics, this is based on the Verbond van Verzekeraars (VvV)'s guidelines.

The interviewees also stated that their organization uses a three-level approach regarding the data policy. The highest level is the policy from the Verbond van Verzekeraars (VvV), which is a national organization within the insurance industry. This is a very generic policy, with which you cannot say, in practice, that you fully comply, since it is mostly guidelines (Appendix 12). The second level is regarding privacy, where the organization should ask themselves, what are we doing, and is it in the interests of the customers? The last level is the practical side, which dives deeper into questions such as: "What data are we using? What should we do with it?" etc. The data science framework is the translation of this practical side. This framework is used to form uniform tooling and working methods for employees (Appendix 12).

To better analyze the input from the interviews with the analyses, table 49 is reorganized. The result can be seen in table 50. Table 51 explains the reasoning behind this transformation.

Table 50: Reorganized components list interviewee 4

Dimension	Components		Priority
Data policy components	Data lifecycle	Data retention	-
		Data usage	-
AI-policy components	AI-Principles	AI-Principles	-
General policy components	Employee	Communicate framework to employees	-
	General policy information	Data science framework	-
		Different levels of policy	-
	Ethics	Ethics based on national organization	-

Table 51: Explanation of transformation process interviewee 4

Dimension	Components	Falls under:	Priority	Explanation
Data policy components	Data retention	Data lifecycle	-	Data retention is a sub-component of the data lifecycle
	Data usage	Data lifecycle	-	Data usage is a sub-component of the data lifecycle
AI-policy components	AI-Principles	AI-Principles	-	Self-explanatory
General policy components	Data science framework	General policy information	-	The data science framework is a method to integrate policy into the culture and behaviors of employees.
	Different levels of policy	General policy information	-	That a policy can have different levels is information that should be in the back of policymakers' heads when they draft a policy. These components are put under general policy information.
	Ethics based on the national organization	Ethics	-	The ethical standards are based on guidelines from a national organization, this is part of the ethics component since it revolves around ethics.
	Communicate framework to employees	Employee	-	Communicating the framework to employees is similar to communication to employees, which is a sub-component of the employee component. Although this component could also be placed under the integration of policy into culture and behavior.

Appendix 14: In-depth overview of the comparison

Tables 52 and 53 show the components and sub-components that have been mentioned in each analysis and both interviews.

Table 52: Overview sub-components from literature, practice, and the interviews part 1

	Literature	Practice	Interview 1	Interview 2	Interview 3	Interview 4
Component:	*Sub-component (code):*	*Sub-component (code):*	*Sub-component (code):*	*Sub-component (code):*	*Sub-component (code):*	*Sub-component (code):*
Data security	Data security	Data security	Data security	Data security	Data security	
	Data privacy	Privacy	Privacy			
	Data integrity	Integrity				
	Data confidentiality	Confidentiality				
	Data availability	Availability				
	Data classification	Data classification	Data classification	Data classification		
	Data access	Data access	Data access		Data access	
Data lifecycle	Data lifecycle	Data process		Data management	Data flows	
	Data storage	Data storage	Data storage		Data storage	
	Data production/creation					
	Data collection/acquisition					
	Data retention	Data retention, deletion, and archiving			Data retention	Data retention
	Data usage (description)				Define data usage	Data usage
	Data quality		Data quality		Data quality	
	Metadata		Metadata		Metadata	
			Data documentation			
			Data lineage			
					Integrate content between systems	
					Managing performance	
					The process around data-based decision making	
Roles & Responsibilities	Data roles & responsibilities	Roles & Responsibilities	Data roles & responsibilities		Role model	
		Connecting roles and responsibilities digitally				
	Data ownership		Data ownership	Process ownership		
					Audit roles and responsibilities	
Explicability	Explainability	Explainability	Explainability			
	Explicability	Explicability				
	Accountability	Accountability				
	Transparency	Transparency				
AI-Principles		AI-principles				AI-Principles
	Beneficence	Beneficence				
	Non-maleficence	Non-maleficence	Non-maleficence			
	Autonomy	Autonomy				
	Justice	Justice				
	Fairness	Fairness				
General AI	AI depends on the organization					
	Continuous change					
			Align AI-technology with goal organization			
Technical guidelines	Technical details					
	Connecting principles with practice					
Privacy	Privacy					
	Communication of privacy within AI usage					
	Consent of usage data					
			Data minimalization			

Table 53: Overview sub-components from literature, practice, and the interviews part 2

AI-models		AI-decision making Data quality input Monitoring AI-models	Continuous change			
General guidelines	Goal & Scope Definitions Stakeholders Compliance/audit	Goal & Scope Definitions Stakeholders' roles & responsibilities Audit & Compliance Maintenance policy Documentation Procedures	Stakeholder roles & responsibilities Procedures	Goal & Scope Responsibilities stakeholders Audit & Compliance Maintenance policy/ evaluation	Define terms Audit & Compliance Key activities organization	
Liability	Risk Liability	Risk assessment		Risk assessment	Identify risks	
Design & Strategy	Strategy Design Architecture	Strategy Design Architecture		Architecture	Design Architecture Identify data needs and requirements	
Employee	Communication to employees Education & Training Reinforcement of policy	Communication Education & Training Reinforcement	Communication Education & Training	Communication Education & Training	Reinforcement	Endow framework
Ethics	Ethical standards	Ethics procedures and methods	Ethical standards			Ethics based on the national organization
General Policy information	Context depended Limit restrictions policy Integration policy into culture and behavior Best practices	Access to laws and regulations Managing change Exceptions Best practices	Priorities components change	Priorities components change	Context dependency Priorities components is context depended	Data Science framework Different levels of policy
Privacy		Privacy Communication data usage & privacy	Translating privacy laws Data leak procedure	Translating privacy laws Data leak procedure	Data reduction	
General security			Development policy regarding applications Access policies General information security Security requirements applications			
Retention periods			Retention periods			

Mentioned and the same
Mentioned in another component
Different sub-component
New sub-component / missing in other analysis
Not mentioned

Appendix 15: In-depth explanation of prioritization

Table 54 shows the comparison between the literature prioritization and the prioritization of components that the interviewees had. Since interviewees 3 and 4 did not give a priority indication, this was marked with a "- ". It also includes the final priority, which is the priority each component has based on the three inputs. Each priority has a weight attached to it ranging from one to four, where one is the W and four is the M. When there are two priorities given, such as "M -> S", the average will be taken. For the "M->S" this will be 3.5. Table 54 also shows the weight of the priority between parentheses. These weights are added up and divided by the number of inputs that have been given. The answers were rounded to whole numbers and based on their number (1=W, 2=C, 3=S, and 4=M).

Table 54: Prioritization from literature and interviews, with the final priority at the end

Component	**Literature**	**1**	**2**	**3**	**4**	**Final**
Data security	M (4)	M -> S (3.5)	M (4)	-	-	M (4)
Data lifecycle	M (4)	S -> M (3.5)	M (4)	-	-	M (4)
Roles & Responsibilities	M (4)	M (4)	M (4)	-	-	M (4)
Explicability	M (4)	M (4)	M (4)	-	-	M (4)
AI-models	C (2)	S (3)	C (2)	-	-	C (2)
General guidelines	C (2)	M (4)	S (3)	-	-	S (3)
Liability	W (1)	W (1)	M (4)	-	-	C (2)
Design & Strategy	W (1)	W (1)	M -> C (3)	-	-	C (2)
Employee	S (3)	M (4)	M (4)	-	-	M (4)
Ethics	M (4)	M (4)	M (4)	-	-	M (4)
Background information	W (1)	W (1)	W (1)	-	-	W (1)
Privacy	S (3)	M (4)	M (4)	-	-	M (4)

The AI-model component now includes the previously known "General AI" component. Their prioritization is combined. This can be done relatively easily since it is the same. The priority of the "liability" and "Design & Strategy" components are, according to the calculation, both a Could-have. However, interviewee 2 did emphasize the importance. Therefore, even though it has a lower priority, it should still be accounted for.

It should be noted that the priority of these components changes for each organization. The list of priorities might be more biased towards a transformational phase since interviewees from Alpina are currently in that phase and prioritize components that are connected to this transformation. Appendix 17 has an in-depth overview that shows all the sub-components of each component.

Appendix 16: In-depth comparison between the list and Alpina's coverage

The final list of components, together with the prioritization can be used to determine what Alpina is missing from its policies and what Alpina should prioritize. Table 55 shows the list of the components and sub-components that should be in a data policy; the colors indicate whether these sub-components appear in Alpina's policies.

Table 55: Differences between the final components list and what Alpina has (Data policy components)

	Practice	
Dimensions	*Component:*	*Sub-component (code):*
Data policy components:	Data security	Data security
		Privacy
		Integrity
		Confidentiality
		Availability
		Data classification
		Data access
	Data lifecycle	Data process
		Data storage
		Data production/creation
		Data collection/acquisition
		Data retention, deletion, and archiving
		Data usage (description)
		Data quality
		Metadata
		Data documentation
		Data lineage
		Integrate content between systems
		Managing performance
		The process around data based-decision making
	Roles & Responsibilities	Roles & Responsibilities
		Connecting roles & responsibilities digitally
		Process ownership
		Data ownership
		Audit roles and responsibilities

	Exists in the policies
	Exists, but slightly different
	Not within the same context
	Does not exist

As can be seen from table 55, data security and all its sub-components are present in the policies from Alpina. This means that all these sub-components have been mentioned in the right context within the policy documents that have been analyzed.

However, this is not the case for the data lifecycle, they have several lacking sub-components. Data production/creation is about data that is created within the company (Alhassan et al., 2016). A procedure for this should be in place that describes what is done with data that is created and produced in Alpina. Data usage is another sub-component that is not covered entirely, which is about describing how data is used within Alpina. This is important since confusion about how data is and should be used can lead to unethical practices (Parsons, 2013; Abraham et al., 2019). Data quality is also a big sub-component that consists of multiple other sub-components according to interviewee 3

(Appendix 9) and the DAMA-DMBOK framework (International, DAMA, 2017). Alpina does not currently cover this component in the right context, they only touched upon it in the context of AI models. Metadata is also not covered, which is data about other data (Abraham et al., 2019). Metadata is used to describe data (Riley, 2017), which can be titles or names. This influences data analysts, which use metadata. A policy should specify how metadata is defined, captured, and used. This is similar to the sub-component "Process around data based-decision making", which is requiring a description of the process behind data based-decision making. The data-based decision-making is done based on the data analysis. To support the data analyst, Alpina should clarify these but can also include processes to support the decision-making process. Data documentation and data lineage are both not covered and were mentioned by interviewee 1 (Appendix 5). Alpina currently lacks the required data documentation and cannot see what the lineage of their data is. Without this insight into the origin of data, classifications of data can be incorrect. This in turn can lead to incorrect usage of data, which can lead to even more problems. Therefore, the data documentation and data lineage must be in order. Systems should also be connected and their content should be integrated according to the DAMA-DMBOK framework (International, DAMA, 2017). Another component that has been mentioned by the DAMA-DMBOK framework (International, DAMA, 2017) is managing performance, which Alpina also does not cover in its policies. This is about managing the performance of data within the lifecycle, but also the performance of data transactions. A process for managing the performance of data should be included in the policy.

The roles & responsibilities component also has some sub-components that are missing. The process and data ownership are both not discussed in the policy documents but are specifically mentioned in the interviews as important. It should be noted that both interviewees stated that they are working on this, but for now, this is still missing. Interviewee 2 (Appendix 7) states that these are important since the ownership is necessary to comply with the law. However, Alpina is lacking employees that are appointed as accountable for processes and data. The roles & responsibilities sub-component is different since the policy documents mainly discuss this sub-component in the context of what the stakeholders' roles & responsibilities are. This leans more towards the general guidelines sub-component, stakeholders' roles & responsibilities. The roles & responsibilities within the data policy dimension are more focused on data and the roles & responsibilities revolving around data. Alpina is also missing an audit mechanism for checking the roles and responsibilities, which is mentioned by interviewee 3 (Appendix 9).

Table 56 shows the list of components and sub-components for the AI policy. The AI-models component is missing one sub-component, which is "Aligning AI with the organization". AI models and their technology should be aligned with the goals of an organization. This is important to include, otherwise, AI is not used effectively and might not contribute to the goals of the organization.

Table 56: Differences between the final components list and what Alpina has (AI-policy components)

Dimensions	*Component:*	*Sub-component (code):*
AI policy components:	Explicability	Explainability
		Explicability
		Accountability
		Transparency
	AI-models	AI-decision making
		Data quality input
		Monitoring AI-models
		Continuous change
		Aligning AI with organization

The last dimension, supplementary policy components, its components, and their sub-components can be seen in table 57. The table includes, just like the other tables, a color-coding that indicates whether sub-components are present in the policies of Alpina.

Table 57: Differences between the final components list and what Alpina has (Supplementary policy

Dimensions	*Component:*	*Sub-component (code):*
Supplementary policy components	General guidelines	Goal & Scope
		Definitions
		Stakeholders' roles & responsibilities
		Audit & Compliance
		Maintenance policy & evaluation
		Documentation
		Procedures
		Key activities organization
	Liability	Risk assessment
	Design & Strategy	Strategy
		Design
		Architecture
		Identify data needs and requirements
	Employee	Communication
		Connect principles with practice
		Education & Training
		Reinforcement
		Access to laws and regulations
	Ethics	Connect principles with practice
		Ethical standards, procedures, and methods
		Integration policy into culture and behavior
		Ethical-principles
		Translation of national organization guidelines
	Privacy	Privacy
		Communication data usage & privacy
		Data leak procedure

As can be seen in table 57, Alpina has some holes to fill regarding this dimension. Regarding supplementary guidelines, Alpina is only missing a description of its key activities. This is important to include since this gives the context of an organization. This context can be used to determine or argue for the priority of certain elements. They can include this within the introduction section of a policy document.

Liability and privacy are both completely green, so Alpina covers these components. However, the data leak procedure within the privacy component is a special case. This was not found within the practice, but during the time this research was written, one has been made. Because it is added later, after the policy analysis had been done, it has been decided to color this green.

Design & Strategy is also mostly covered by Alpina, but they still lack the “identifying data needs and requirements” component. This is an important sub-component, which should be done before a policy is made. Alpina might have done this but did not include it within their policy document. For this research, it is assumed Alpina did not do this. Therefore, it is recommended that the policymakers identify the data needs and requirements within Alpina, which can differ for each department. The policymakers can then use the data needs and requirements to determine what measures are needed to meet these data needs. This can also influence other policies.

Alpina has most of the sub-components within the employee component, but one still needs to be adjusted. One of these is the connection of principles with practice. Ethical principles should be translated into more technical guidelines which can be implemented. This helps the employees, but especially the engineers in the process of changing the system to be compliant with the principles. However, Alpina has a limited amount of this in place, which is why the sub-component is colored blue.

Regarding ethics, a lot of its sub-components are colored blue. This is because Alpina does have something in place regarding these sub-components, but it can be improved, which is the case for the "Ethical standards, procedure, and methods" component. While Alpina does have some procedures and methods in place, they do not translate to the practice in a way that can easily be used by engineers. A sub-component that is relevant to answering the research question is 'the integration of policy in culture and behavior' sub-component. This sub-component is about how a policy can be integrated into the culture and behavior of an organization. Because education & training are methods of integrating policy into culture and behavior, it is chosen to color this sub-component from red to blue. However, Alpina does not yet have enough in place regarding integrating specific policy components in the culture and behavior. Interviewee 4 (Appendix 12) mentioned the use of a data science framework, which can be used to comply with policy and organization formalities. The ethical principles component is also colored blue since Alpina only covers the AI principles and not the data principles. The last sub-component, "Connecting principles with practice" is also discussed in the employee component. It has been chosen to also include this sub-component within the ethics component since, without this, the ethics could stumble across implementation problems.

Appendix 17: In-depth overview of components and sub-components

Table 58 shows a final overview of the components and their sub-components.

Table 58: Final in-depth overview of dimensions, components, and their sub-components

Dimension:	***Component:***	***Sub-component:***
Data policy components	Data security	Data security
		Privacy
		Integrity
		Confidentiality
		Availability
		Data classification
		Data access
	Data lifecycle	Data process
		Data storage
		Data production/creation
		Data collection/acquisition
		Data retention, deletion, and archiving
		Data usage (description)
		Data quality
		Metadata
		Data documentation
		Data lineage
		Integrate content between systems
		Managing performance
		The process around data-based decision making
	Roles & Responsibilities	Roles & Responsibilities
		Connecting roles and responsibilities digitally
		Process ownership
		Data ownership
		Audit roles and responsibilities
AI policy components	Explicability	Explicability
		Explainability
		Accountability
		Transparency
	AI-models	AI-decision making
		Data quality input
		Monitoring AI-models
		Continuous change
		Aligning AI with organization
	General guidelines	Goal & Scope
		Definitions
		Stakeholders' roles & responsibilities
		Audit & Compliance
		Maintenance policy & evaluation
		Documentation
		Procedures

Supplementary policy components		Key activities organization
	Liability	Risk assessment
	Design & Strategy	Design
		Strategy
		Architecture
		Identify data needs and requirements
	Employee	Communication
		Connect principles with practice
		Education & Training
		Reinforcement
		Access to laws and regulations
	Ethics	Connect principles with practice
		Ethical standards, procedures, and methods
		Integration policy into culture and behavior
		Ethical principles
		Translation of national organization guidelines
	Privacy	Privacy
		Communication data usage & privacy
		Data leak procedure

References

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

Aggarwal, M., Gingras, C., & Deber, R. (2021). Artificial Intelligence in Healthcare from a Policy Perspective. *Multiple Perspectives on Artificial Intelligence in Healthcare*, 53–64. https://doi.org/10.1007/978-3-030-67303-1_5

Al-Badi, A., Tarhini, A., & Khan, A. I. (2018). Exploring Big Data Governance Frameworks. *Procedia Computer Science*, *141*, 271–277. https://doi.org/10.1016/j.procs.2018.10.181

Alhassan, I., Sammon, D., & Daly, M. (2016). Data governance activities: An analysis of the literature. *Journal of Decision Systems*, *25*(sup1), 64–75. https://doi.org/10.1080/12460125.2016.1187397

Alhassan, I., Sammon, D., & Daly, M. (2019). Critical success factors for data governance: A theory building approach. Information Systems Management, 36(2), 98–110. https://doi.org/10.1080/10580530.2019.1589670

Alpina Group. (2021). *Alpina group*. https://alpinagroup.com/#intro

Autoriteit Persoonsgegevens. (2020). *Belastingdienst/toeslagen de verwerking van de nationaliteit van aanvragers van kinderopvangtoeslag*. https://www.autoriteitpersoonsgegevens.nl/sites/default/files/atoms/files/onderzoek_belastingdienst_kinderopvangtoeslag.pdf

Babbie, R. (2012). Qualitative interviewing. In *The practice of social research* (pp. 310–315). Cengage Learning.

Blackburn, S., Galvin, J., LaBerge, L., & Williams, E. (2022, April 12). Strategy for a digital world. McKinsey & Company. Retrieved June 21, 2022, from https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/strategy-for-a-digital-world

Cai, L., & Zhu, Y. (2015). The challenges of data quality and data quality assessment in the big data era. Data Science Journal, 14(0), 2. https://doi.org/10.5334/dsj-2015-002

Cheng, G., Li, Y., Gao, Z., & Liu, X. (2017). Cloud data governance maturity model. 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS). https://doi.org/10.1109/icsess.2017.8342968

Conner, M., & Norman, P. (2015). Predicting and changing health behavior (3rd edition). Amsterdam University Press.

Delacroix, S., & Wagner, B. (2021). Constructing a mutually supportive interface between ethics and regulation. *Computer Law & Security Review*, *40*, 105520. https://doi.org/10.1016/j.clsr.2020.105520

Floridi, L. (2019). Establishing the rules for building trustworthy AI. *Nature Machine Intelligence*, *1*(6), 261–262. https://doi.org/10.1038/s42256-019-0055-y

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, *28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Grewal, P. D. S. (2014). A critical conceptual analysis of definitions of artificial intelligence as applicable to computer engineering. IOSR Journal of Computer Engineering, 16(2), 09–13. https://doi.org/10.9790/0661-16210913

Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines*, *30*(1), 99–120. https://doi.org/10.1007/s11023-020-09517-8

Heilbron. (2020, October 20). *Heilbron*. Over Heilbron. https://www.heilbron.nl/over-heilbron/

Hrynaszkiewicz, I., Simons, N., Hussain, A., Grant, R., & Goudie, S. (2020). Developing a Research Data Policy Framework for All Journals and Publishers. *Data Science Journal*, *19*(1), 5. https://doi.org/10.5334/dsj-2020-005

Hudaib, A., Masadeh, R., Qasem, M. H., & Alzaqebah, A. (2018). Requirements prioritization techniques comparison. Modern Applied Science, 12(2), 62. https://doi.org/10.5539/mas.v12n2p62

Hugh Heclo, H. (1972). Review article: Policy analysis. *British Journal of Political Science*, *2*(1), 83–108. https://doi.org/10.1017/s0007123400008449

International, DAMA. (2017). DAMA-DMBOK: Data management body of knowledge: 2nd edition (Second ed.). Technics Publications.

Jiankang, W., Jiuling, X., Qianwen, L., & Kun, L. (2011). Knowledge management maturity models: A systemic comparison. 2011 International Conference on Information Management, Innovation Management and Industrial Engineering. https://doi.org/10.1109/iciii.2011.420

Kim, H. Y., & Cho, J. S. (2018). Data governance framework for big data implementation with NPS case analysis in Korea. Journal of Business & Retail Management Research, 12(03). https://doi.org/10.24052/jbrmr/v12is03/art-04

Klievink, B., Neuroni, A., Fraefel, M., & Zuiderwijk, A. (2017). Digital Strategies in Action. *Proceedings of the 18th Annual International Conference on Digital Government Research*. https://doi.org/10.1145/3085228.3085270

Longhurst, R. (Ed.). (2016). Semi-structured interviews and focus groups. In *Key methods in geography* (3rd Revised edition, pp. 143–148). SAGE Publications. https://books.google.nl/books?hl=nl&lr=&id=7hcFDAAAQBAJ&oi=fnd&pg=PA143&dq=structured+interviews&ots=TDOJyo7Q9w&sig=b3Nl11dO6Lyadeik7XkhtAxK3PY#v=onepage&q=structured%20interviews&f=false

Meehan, E. J. (1985). Policy: Constructing a definition. *Policy Sciences*, *18*(4), 291–311. https://doi.org/10.1007/bf00135916

Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2019). From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Science and Engineering Ethics*, *26*(4), 2141–2168. https://doi.org/10.1007/s11948-019-00165-5

Mullins, M., Holland, C. P., & Cunneen, M. (2021). Creating ethics guidelines for artificial intelligence and big data analytics customers: The case of the consumer European insurance market. *Patterns*, *2*(10), 100362. https://doi.org/10.1016/j.patter.2021.100362

Parsons, M. A. (2013). Data Policy. *Data Science Journal*, *12*(0), GRDI43–GRDI50. https://doi.org/10.2481/dsj.grdi-008

Rainey, L. B., & Tolk, A. (2015). *Modeling and simulation support for system of systems engineering applications*. Wiley. https://doi.org/10.1002/9781118501757

Raymond, A. H. (2013). Data management regulation: Your company needs an up-to-date data/information management policy. *Business Horizons*, *56*(4), 513–520. https://doi.org/10.1016/j.bushor.2013.03.001

Richardson, J., & Hoffman-Kim, D. (2010). The importance of defining 'Data' in data management policies. Science and Engineering Ethics, 16(4), 749–751. https://doi.org/10.1007/s11948-010-9231-5

Riley, J. (2017). Understanding metadata. National Information Standards Organization (NICO). https://www.fidgeo.de/wp-content/uploads/2016/07/2017_01-NISO-understanding-metadata.pdf

Siau, K., & Wang, W. (2020). Artificial Intelligence (AI) Ethics. *Journal of Database Management*, *31*(2), 74–87. https://doi.org/10.4018/jdm.2020040105

Siddiqa, A., Hashem, I. A. T., Yaqoob, I., Marjani, M., Shamshirband, S., Gani, A., & Nasaruddin, F. (2016). A survey of big data management: Taxonomy and state-of-the-art. *Journal of Network and Computer Applications*, *71*, 151–166. https://doi.org/10.1016/j.jnca.2016.04.008

Stahl, B. C., & Wright, D. (2018). Ethics and Privacy in AI and Big Data: Implementing Responsible Research and Innovation. *IEEE Security & Privacy*, *16*(3), 26–33. https://doi.org/10.1109/msp.2018.2701164

Swain, P. A. (2011). Social policy for social change. *Australian Social Work*, *64*(3), 408–409. https://doi.org/10.1080/0312407x.2011.611611

van Mil, J. W. F., & Henman, M. (2016). Terminology, the importance of defining. International Journal of Clinical Pharmacy. https://doi.org/10.1007/s11096-016-0294-5

van Zeeland, H., & Ringersma, J. (2017). The Development of a Research Data Policy at Wageningen University & Research: Best Practices as a Framework. *LIBER QUARTERLY*, *27*(1), 153–170. https://doi.org/10.18352/lq.10215

Vance, A., Siponen, M., & Pahnila, S. (2012). Motivating IS security compliance: Insights from habit and protection motivation theory. Information & Management, 49(3–4), 190–198. https://doi.org/10.1016/j.im.2012.04.002

Veiga, A. D., & Eloff, J. H. P. (2007). An Information Security Governance Framework. *Information Systems Management*, *24*(4), 361–372. https://doi.org/10.1080/10580530701586136

Verbond van Verzekeraars. (2022). Werken bij het verbond. Retrieved June 2, 2022, from https://www.verzekeraars.nl/over-het-verbond/werken-bij-het-verbond#:%7E:text=Wat%20is%20het%20Verbond%20van,zijn%20lid%20van%20het%20Verbond.

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

Voogd & Voogd. (2022). *Voogd & Voogd*. Over ons Voogd & Voogd. https://www.voogd.com/over-ons/

Wolfswinkel, J. F., Furtmueller, E., & Wilderom, C. P. M. (2013). Using grounded theory as a method for rigorously reviewing literature. *European Journal of Information Systems*, *22*(1), 45–55. https://doi.org/10.1057/ejis.2011.51

Yang, R. (2014). Comparing policies. *Comparative Education Research*, 285–308. https://doi.org/10.1007/978-3-319-05594-7_10

Yin, R. K. (2018). Case study research and applications (6th Revised edition). SAGE Publications.

Zang, Y. (2016). Unstructured interviews. In B. M. Wildemuth (Ed.), *Applications of social research methods to questions in information and library science, 2nd edition* (2nd ed., pp. 249–250). Libraries Unlimited. https://books.google.nl/books?hl=nl&lr=&id=uv98DQAAQBAJ&oi=fnd&pg=PA239&dq=unstructured+interviews&ots=VVNII9VZik&sig=x5mm2Ljy2A0qmTonevKJLtvzAsQ#v=onepage&q=unstructured%20interviews&f=false

Zhou, L., & Nunes, M. B. (2016). Formulating a framework for desktop research in Chinese information systems. *Handbook of Research on Innovations in Information Retrieval, Analysis, and Management*, 307–325. https://doi.org/10.4018/978-1-4666-8833-9.ch011