An investigation into materiality assessment practices

Process analysis of and a text mining application for materiality assessments

A master's thesis in Business Informatics

Author: Melchior Keijdener
m.n.c.keijdener@uu.nl

Utrecht University

1st supervisor: Sergio España
s.espana@uu.nl

2nd supervisor: Slinger Roijackers
slinger.jansen@uu.nl

Date: August 11, 2019

This page is intentionally left blank

Abstract

An important aspect of sustainability reporting is determining which topics are material for an organization. These are topics for which it is important, or required, to disclose their strategy on. To determine these, an organization can execute a materiality assessment. This is a set of activities to determine which topics are material to an organization; which aspects should be disclosed, reported and managed. However, few methods for materiality assessment have a quantitative basis, making assessments subject to incompleteness and subjectivity. Furthermore, scientific literature provides little guidance, as most of the work is based on information disclosed in sustainability reports. This makes findings prone to reporting bias.

To discover which activities are executed, a survey of semi-structured interviews is conducted across eleven firms. By mapping multiple materiality assessment methods and comparing them with each other, a set of fourteen general activities is identified, which are combined into five steps which serve as the basis of materiality assessment. This has been supplemented by a systematic literature search of scientific and grey literature.

We found that materiality assessments begin with the identification of topics. These topics can be collected in multiple ways. For example, by analyzing the organization's current sustainability strategy, topics emerge which are already being worked on. Some topics might be hidden in documentation of previous years, which were previously not material topics, but of which their materiality has increased this year. Furthermore, longlists such as the GRI indicator list or a longlist created by an organization itself can serve as a basis from which topics can be selected. After these topics are identified, a plan is created to assess them. This includes the stakeholders whose opinions are important to the firm, which instruments are useful to elicit information from them and metrics to assess materiality. Apart from measuring whether the stakeholders find the topic material, also, for example, the influence of the firm on the topic, or the impact of the topic on the firm can be determined. A third phase is the data collection from these stakeholders on these metrics. We found that surveys, interviews and focus group are common instruments. However, also (social) media provides a worthwhile contribution if there is sufficient tool support to analyze it. Next, these topics are ranked and this ranking is visualized in a materiality matrix. The final step is then to validate this ranking with management. This is done by discussing the results with (senior) management, as it is vital that these topics are tackled on a strategic level.

Additionally, we discovered multiple problems in the industry, which are combined into a research agenda. The three most prominent issues are overcoming a gap in maturity between practitioners. Some organizations have very mature processes, while others do not. However, there is no step-by-step guide on how to start, i.e. which activities are the most important to do first, and how to expand this year by year. A maturity model could provide such a framework. Next, more research is required to determine parameter settings for a materiality assessment. There is no common number of surveys, little knowledge on longlists and how set these up. Finally, there is no flexible, accessible and transparent tool in the industry. Therefore, we have already created a prototype of a tool which is freely available to use[1].

[1] https://github.com/Melchior-Keijdener/OpenMAsses-Python

Acknowledgements

This thesis is the result of eight months of work, in which I've spent hours writing, reading articles, interviewing people and programming a tool. Even though I have been fortunate interchanging these activities with each other, remaining focused for such a long time is no evident task. Thankfully, during this process I have not been alone, as many others have supported me. Therefore, I wish to take this moment to thank everyone.

First of all, I wish to thank my supervisor Sergio España for his contributions. He has been so kind to devoting time to me during this process. Although I have occasionally scared him by walking in without an appointment, probing him into thinking he had forgotten about our meeting, he always made time to discuss new results and steering me in the right direction. I believe he has been the catalyzer, allowing me to raise the level of this research. I am grateful to having such a dedicated supervisor. I would also like to thank Slinger Jansen for his input on my proposal and review of my work.

Next, I wish to thank all my respondents for their time and input. Without your considerate efforts, delivering this work would have been downright impossible. Participating in research such as this is unlikely to create applicable value right away. However, I hope this thesis provides you with new insights, or forms the basis of them, to improve your materiality assessment and sustainability reporting. I consider this work to be the groundwork for much more work to come and would be delighted if you would be willing to work with me, my supervisors or successors again.

Furthermore, I've concluded that the hardest part of writing a thesis is staying motivated. I wish to thank my fiancé for dragging me out of bed in the morning, listening to my struggles and proofreading work which 'made you fall asleep'. I feel very fortunate for receiving your support during this time. Finally, I wish to express my gratitude to all my friends and family for their support. Whether they took time to discuss my work, put up with me shaking my hands vigorously at my screen in frustration due to Windows errors or were simply present during this period. Writing a thesis with you at the same time has helped stay on track and I am happy our study schedules aligned.

- Melchior Keijdener, July 2019

Contents

List of Figures

List of Tables

Chapter 1. Motivation for research into materiality assessment

In sustainability reporting, firms are required to disclose their strategy on certain topics. These topics can entail a large range from the negation of environmental damage to the abidance to gender equality to the increase of transparency in their supply chain. However, not each company requires to disclose their stance on the same issues; a bank cares about ethical investments, whereas a retail organization focuses on a responsible supply chain. Adapting a sustainable strategy does not merely demonstrate a sign of goodwill, but has been shown to create value in both theory and practice (Eccles, Ioannou, & Serafeim, 2016; Forstater et al., 2006).

To determine which topics need to be addressed, the process of materiality assessment identifies, prioritizes and reviews these topics (Forstater et al., 2006). For this, multiple methods exist (Phelps, 2016), although these methods struggle with two issues. One, the existing standards are kept generic to be applicable to different organizations and settings. Therefore, the guidelines on materiality assessment result in little guidance for practitioners. Two, few of these methods are based on quantitative foundation, which makes them prone to subjectivity and incompleteness (Calabrese, Costa, Levialdi, & Menichini, 2016).

Additionally, there is little scientific work on this topic and the work that does is often based on information disclosed in sustainability reports. (Guix, Bonilla-Priego, & Font, 2018; Puroila & Mäkelä, 2019; Saenz, 2019; Sarraj, 2018). This makes this research subject to reporting bias, rendering findings incomplete. Therefore, fieldwork on practitioners of materiality assessment in required.

An exploratory study by Mukhopadhyay (2017) investigated the practices of a single consulting firm. In their materiality assessment method, several pain points were identified. To combat these, a holistic tool was proposed and a simple implementation was created (Jongerius, 2017). However, as this study conducted a single case study, external validity is low and generalizability cannot be claimed, thus more research on this subject is required.

Therefore, the motivation for this thesis is twofold. First, we will investigate the landscape further, identifying methods used for materiality assessment by multiple practitioners. To properly compare these, a conceptual model is required to discuss which aspects in these methods overlap and differ. By mapping these methods and discussing materiality assessment with industry experts, we investigate which problems exist within the industry. Due to the complex nature of materiality assessment (Puroila & Mäkelä, 2019), it is unlikely that there are simple solutions to these problems. Therefore, these problems are aggregated into a research agenda, probing other scholars to tackle them as well. Second, this thesis will continue the development of the tool, mainly by extending it with the ability to assess (social) media; a research challenge marked in the original paper (Mukhopadhyay, 2017).

These contributions, of insight into the industry and of development of the tool, are combined into the following research question:

How can the process of materiality assessment be enhanced?

In this question, enhance is not specifically defined. Rather, it is a term that encompasses the possibility of making the process either more complete, robust or accurate. To find an answer to this question, two more questions with three sub-questions are coined:

- What is the current state of affairs in materiality assessment?
    - What are common practices in materiality assessment?
    - What are common problems in materiality assessment?
    - What are requirements for a tool to support materiality assessment?
- How can a tool contribute to this enhancement?
    - How can a tool support a general materiality assessment?
    - How can (social) media be analyzed through a tool?
    - What are benefits and drawbacks of a tool?

The rationale behind these questions is simple. By answering the first major question, insight into problems within the industry is gained. By answering the second question, we investigate which issues can be alleviated by this tool.

This thesis has an unorthodox structure. Finding the answer to these questions requires different approaches. Furthermore, these questions are tackled in partial collaboration with other researchers. Therefore, each question is answered in a separate paper, each with their own problem description and method, which comprises chapters 2 and 3. For depth into some topics which are not fully elaborated on in the papers, appendices are added which describe the issue in more detail.

Disclaimer on author's contributions

Research is often not executed in isolation. This means that multiple researchers have contributed to this thesis in terms of data collection, interpretation and conclusions. As I do not wish to portray their work as my own, a more detailed of their precise contributions is given below to give credit where credit is due.

Mukhopadhyay (2017) has created an outline for a materiality assessment tool to support this process. This tool consists of a long-list of predefined topics, which counts these topics in pdf files such as sustainability reports. This way, an overview is provided of the topics which are mentioned in the document, which can be used in the business strategy. She is also responsible for the data collection, interpretation and conclusions of respondent 11. My findings on this respondent are based on her documentation of them. This tool has been implemented in an R-environment by Jongerius (2017) using a R-shiny interface to interact with the system, which has been validated by them in the first validation round as described in chapter 3. However, this approach has been somewhat naïve, as the tool did not yet recognize bigrams and did not contain stemming. This means that phrases such as 'low sustainability' and 'no investment' are not yet recognized. Furthermore, terms as sustainability, sustain, sustainable are all counted as different occurrences, rather than as a single topic. Work by Horn & Turk (2018) has aimed to improve the tools' accuracy, which has been partially successful, as the recall was improved but the precision mostly was not. Additionally, as this tool relies on an input of a longlist of topics, Quik (2018) has conducted a crowd sourced enterprise to compose such a longlist. Here, she found 129 sustainability topics that can be recognized, but only for the domain of higher education.

Two bachelor's theses have contributed to this work. Fechner (2019) worked on identifying methods and tools in the literature. His work serves as a basis for methods found in literature, i.e. Deloitte, KPMG, etc. Furthermore, he found multiple tools, which he divided into three classes. One, class one tools, are tools such as: OpenMasses[2], Datamaran[3] and MAT[4]. These tools generally speaking support many activities in the assessment process. Two, tools which support a specific part of this process, such as Ecovadis[5], Sofi[6] and Polecat[7]. These tools often support a few activities, such as a media analysis, but do not provide the user the ability to generate a matrix based on this. Three, general purpose tools such as Microsoft Office and Google. These are tools which are used in the process but do not necessarily support a specific activity. This distinction of classes is adopted by me, although the terminology has been changed slightly. In another bachelor's thesis in this project. De Wit (2019) created base PDDs based on the findings of my interviews (respondents one up until nine). These PDDs were reviewed by me and adapted versions are used for this research.

As the goal of this thesis is partly to combine and publish the results of this research, the term 'we' is used to cover the contributions of all of the previously mentioned researchers. However, the writing of this thesis is all my own work.

[2] https://github.com/sergioespana/openmasses
[3] https://www.datamaran.com/about-us/#technology
[4] http://www.zooid.com.au/tools/
[5] https://www.ecovadis.com/nl/
[6] https://www.thinkstep.com/software/corporate-sustainability
[7] https://www.polecat.com/

Chapter 2. Methods in materiality assessment: an overview and research agenda

1. Introduction

Materiality assessment is the process of identifying, prioritizing and reviewing which topics are important to a firm or organization (Forstater et al., 2006). This is often done by combining opinions of multiple internal and external stakeholders, such as employees, board of directors, suppliers, NGO's or clients. Additionally, media trends, and topics disclosed by industry peers can serve as input for an assessment. This process is of growing importance in business, as it shifts from a matter of compliance to sustainability themes to one of value generation (Forstater et al., 2006).

In general, this value generation can be attributed to three reasons (Forstater et al., 2006). One, performing materiality assessments enables business leaders to anticipate on long-term issues. This allows them to align their strategy with issues that might become more important in the future, hence increasing the sustainability of their strategy. Two, this long-term vision can be utilized by leaders to attract capital from investors, partners and other stakeholders. Three, by acting preemptively on emerging issues; it allows leaders to receive credit and goodwill of policymakers and civil society. This makes it easier for them to deflect criticism, receive rewards and influence regulations. This value creation is not purely theoretical, as companies that adopted strong sustainability policies in 1990 outperform their counterparts today (Eccles et al., 2016).

To perform this process of materiality assessment, multiple companies, for example British Petroleum, Ford Motor company and BT group, have developed materiality assessment frameworks (Forstater et al., 2006). Although these frameworks differ in the details, they all produce a materiality matrix, outlining the material topics to each firm. However, a problem with these frameworks is that few are based on quantitative methods. This makes the assessment process susceptible to subjectivity and incompleteness (Calabrese et al., 2016).

To support these companies, there are some organizations which focus on providing guidelines on materiality. Through a survey among the industry, Phelps (2016) found the most notable are the Global Reporting Initiative's G4 guidelines (GRI)[8], Sustainability Accounting Standards Board's standards (SASB) (*Implementation Guide for companies*, 2009), International Integrated Reporting Council's Framework (The International Integrated Reporting Council (IIRC), 2015) and AccountAbility's Materiality Framework (Forstater et al., 2006). However, to keep the GRI standards generic and applicable to different organizations and settings, the guidelines on materiality assessment are kept high-level. As a result, processes of and metrics used in materiality assessments might differ throughout organizations, which makes alignment and benchmarking difficult.

Additionally, there is little scientific work on this topic and the work that does is often based on information disclosed in sustainability reports (Guix et al., 2018; Puroila & Mäkelä, 2019; Saenz, 2019; Sarraj, 2018). This makes this research subject to reporting bias, rendering findings incomplete. Therefore, fieldwork on practitioners of materiality assessment in required. As a

[8] These have transitioned to GRI standards after July 1, 2018.

result, this paper aims to provide an overview of actual industry practices and provide a research agenda on issues the industry struggles with.

Unlike the standard in academic work, findings from theory and practice are not separated in different sections. This is done because scientific work is rather limited and little clarity would be gained from discussing these findings separately. Therefore, the outline of this paper is as follows. First a method is described on how to discover common practices and problems by the industry in section 2. In section 3, two models are proposed to describe the industry and concepts of materiality assessment. Based on these models, section 4 describes the results. Next, section 5 proposes a research agenda for this topic. Finally, section 6 and 7 respectively discuss limitations and conclusions.

2. Method

This research method is a hybrid between a multivocal literature review and a survey of interviews. This approach is taken due to there being insufficient (scientific) literature, a lack of industry standards and an absence of an agency enforcing any. This results in many practitioners using their own methods. As these are mostly undocumented or even tacit, it is impossible to gain a full understanding of the problem based on literature alone. Therefore, to sketch an overview of the industry, its common practices and its issues, this hybrid approach is required.

2.1. Literature

To determine the type of literature review, a consideration between a multivocal literature review (MLR) and systematic literature review (SLR) is made. Garousi, Felderer, & Mäntylä (2019) provide seven criteria to judge whether a topic is suitable for a MLR or a SLR. Although they acknowledge that these are subjective criteria, they deem a research team capable of assessing their own need. We deemed an MLR suitable due to three criteria. First, the subject is 'complex' and there is not enough formal literature on it. Next, it would provide a synthesis of insights from industrial and academic community and finally, there is a large volume of practitioners showing interest in the topic. For the MLR, 3rd tier and above grey literature (GL) is accepted, as long as the source is a reputable institution, i.e. large international firms with an established track record in the field or specialized institutions on the topic, or it is a materiality report published by the company (Garousi et al., 2019).

Based on a conceptual model in section 3.2, several search terms are used for the search. These are a combination of the terms "materiality assessment" and "materiality analysis" in combination with "methods", "process", "technique", "matrix" and "tool". "Process" and "technique" are used as synonyms for "method". "Tool" is used to identify tools for materiality assessment. The other terms of the conceptual model (in section 3.1) are not combined with materiality assessment, as "goal" and "source" both create an ambiguous query. "Activities" is not combined as they are inherent of "method". A Scopus search[9] is used for this collection of scientific literature. By performing content analysis, only nine unique scientific articles were found to be relevant, i.e. these articles discuss a matter related to the search terms.

The same terms were used for a Google search[10] into the topic and the first fifty links per search query are considered. 489 articles are collected through this search, of which 241 are unique links. For the collection and analysis of this literature, a review tool[11] is used to judge whether the literature is usable and whether it is of sufficient quality. After reviewing 100 links, data saturation was reached and the analysis was stopped. 27 of these links are discarded, because the link did not work (8 links), they discussed an unrelated subject (5 links) or were marketing for a company containing no information on materiality assessment (5 links). The majority of the 73 remaining links are from 2017 and 2018, as shown in Figure 1. 43 links are materiality reports, which are reports from companies discussing their materiality method on their website. However, only seven of these reports contain a novel element, i.e. an element which is unique among the reports. Furthermore, seven scientific articles were found, of which two

[9] Executed on 28-5-2019

[10] Executed on 3-7-2019

[11] Link to tool: https://github.com/Melchior-Keijdener/Systematic-Literature-Review-on-Google

are new in addition the Scopus search. Finally, five links are blogposts, discussing various elements of materiality assessment.

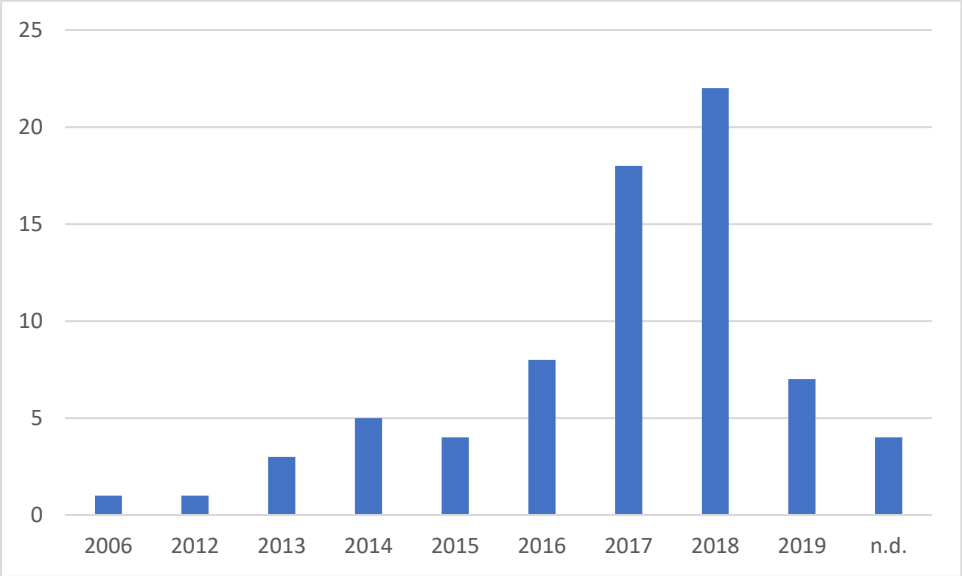


Figure 1. Distribution of links per year

2.2. Interviews

For this phase, twelve experts in the industry are interviewed in ten semi-structured interviews. Additionally, before this series of interviews, exploratory interviews are held (R11). The identification of these experts is an iterative process and uses two approaches. The first approach was to identify experts by searching for the terms 'materiality assessment', 'materiality analysis' or their Dutch equivalents on LinkedIn. The researchers assessed their profiles and discarded candidates who did not have any experience with materiality assessment within the last five years. The second approach was to use the same queries on Google. Companies that discussed their assessment in detail on the website or offered it as a service were contacted and requested to cooperate. The majority of experts were found through referrals from these candidates. The distribution of these respondents is given in Table 1 with regards to their company size, experience and the type of industry they either operate in or provide consulting services to. In this study, two types of interviews are performed: investigation interviews and validation interviews. For the former, the focus of the interview is on mapping their method and their usage of tools and sources. For the latter, the focus of the interview is to validate key findings from the investigation interviews. The interview protocols are enclosed in respectively appendix A and B.

Table 1. Distribution of interviewees

Respondent	Purpose	Company size[12]	Experience	Type of industry
Respondent 1 [R1]	Investigation	Small	4 years	Retail, Supermarkets, retail suppliers and agricultural sector
Respondent 2 [R2]	Investigation	Large	2 years	Any, mostly large clients
Respondent 3 [R3]	Investigation & validation	Micro	5 years	Facility, insurance, government, education, small-medium enterprises in technical sector,
Respondent 4 [R4]	Investigation & validation	Micro	9 years	Retail, environmental/humanitarian organizations
Respondent 5 [R5]	Investigation	Large	1 year	Insurance
Respondent 6 [R6]	Investigation	Micro	3 years	Any, mostly small-medium enterprises
Respondent 7 [R7]	Investigation	Large	3 years	Financial
Respondent 8 [R8]	Investigation	Micro	14 years	Consumer products, financial, heavy industry
Respondent 9 [R9]	Investigation	Large	4 years	Any, mostly large clients
Respondent 10 [R10]	Validation	Large	10 years	Any, mostly large clients
Respondent 11 [R11]	Exploratory	Medium	9 years	Any

In the remainder of this thesis, claims backed by respondents from the investigation interviews are cited as: [R1; R2; Rn]. In case all of the respondents back the claim, the format [R1-R9] is used. For validation interviews, the format [R3*; R4*; R10*] is used. Finally, a note is required on R3 and R10*. These interviews were conducted with two people, instead of one. As they are part of the same company and same interview session, silent agreement between them is assumed.

2.3. Data Analysis

Data from the problem investigation interviews from R1 to R9 is transcribed and analyzed with the program NVivo. For this, a coding tree is created based on the conceptual model in Figure 3 in section 3.2. This coding tree is enclosed in appendix C. The methods described by the respondents are modelled by using a Process Deliverable Diagram (PDD), a technique created by van de Weerd & Brinkkemper (2009) to analyze and compare methods. This technique models the method into activities (actions which are executed) and concepts (outcome/product of these actions). The activities of the resulting collection of PDDs can then be mapped in order to compare the methods to each other, as for example demonstrated by Keijdener, Overbeek, & España (2018). By mapping the methods, activities with a similar nature and goal can be grouped into high-level activities. This is an iterative process, performed by a single researcher, who discusses the findings with another researcher until inter-reviewer agreement is reached. As this interview is aimed at answering and

[12] Based on taxonomy by Eurostat: https://ec.europa.eu/eurostat/web/structural-business-statistics/structural-business-statistics/sme

strengthening specific hypotheses, the data is transcribed and analyzed through content analysis without a coding tree.

3. Describing materiality assessment through models

For clarification, two models are created based on the findings of this study: (1) an overview of the industry which describes the types of practitioners and their perspectives on materiality assessment, and (2) a conceptual model of materiality assessment which describes the general process.

3.1. Perspectives of materiality assessment

Materiality assessment is a term encompassing multiple types of assessments. In general, we found four distinct types of practitioners of materiality assessment: consultants, small and medium enterprises, accountants and major firms, each with their own perspective on materiality assessment. Their relations are illustrated in Figure 2. In this research, six of the respondents are identified as consultants, two of them are identified as a major firm and three are identified as accountants. As no literature on these perspectives is found, this model is a hypothesis based on the findings from the interviews. In general, there are three perspectives on materiality assessment, whose goal all slightly differ.

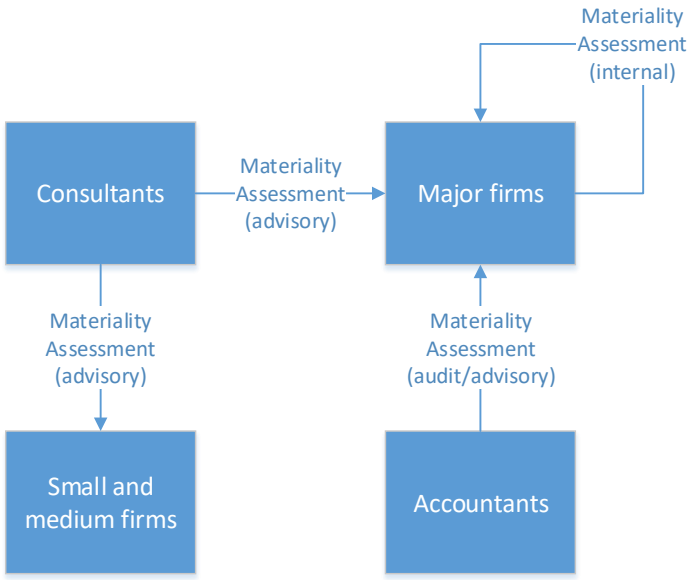


Figure 2. Industry overview: four types of practitioners and three perspectives on materiality assessment.

First, there is the advisory perspective of consultants and accountants. From this perspective, consultants execute a materiality assessment for a specific client as an assignment. However, it must be noted that materiality assessment is often not the core of this assignment, as these are often part of a larger sustainability reporting cycle [R1; R3; R7; R9]. Additionally, it must be noted that these assignments differ heavily from each other, as sometimes consultants need to execute the whole assignment and sometimes the firm already provides them with data, meaning they execute it only partly [R3; R4*; R7].

Second, there is the 'internal perspective of major firms'[13]. From this perspective, a firm needs to execute the materiality assessment from beginning to end. A large focus is on implementing a process to perform this assessment [R7]. For this, they can enlist help from consultants for some small or large parts [R5].

[13] This is also sometimes abbreviated as the 'internal perspective' or 'firm perspective'.

Third, there is the audit perspective, where a materiality assessment is executed as a measure of control; to verify that the outcome of both the accountant and the client matches. This requires a different method than used in the advisory perspective: *"In advisory you really perform the assessment. In an audit, the assessment is already done and you check whether all the steps were properly executed."* [R10*].

Although all respondents agreed with this model during the validation [R3*; R4*; R10*], the distinction between consultants and firms requires a caveat. In practice, it is possible that consultants which have a long-term cooperation with a firm take place in the firm's core-team [R4*]. This creates a hybrid role of the consultant and firm. However, this is more of an organizational change, rather than a merge of perspectives.

3.2. Towards a conceptual model of materiality assessment

To harmonize these different perspectives on materiality assessment, we propose a conceptual model of materiality assessment. Therefore, the concept is split up into multiple smaller concepts, as shown in Figure 3. These concepts are briefly introduced in this section and are discussed in more detail in section 4.

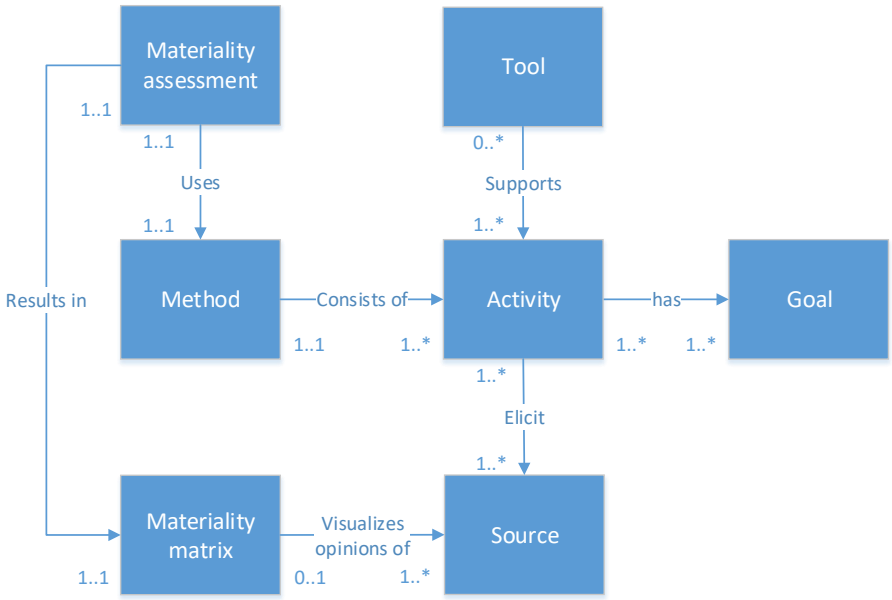


Figure 3. Conceptual model of materiality assessment

The model consists of seven main concepts. First, there is the term itself. Each materiality assessment uses a method. These methods are defined at a high-level, as for example is the case with GRI or IIRC. Each method consists of multiple activities, such as the consultation of stakeholders, media analysis or reporting. Each of these activities have one or multiple goals. For example, the goal of surveying suppliers is to collect data. Similarly, several activities can share the same goal. Data can also be collected by a focus group session with management or an interview with employees. It is also important to note that a single activity can fulfil multiple goals. For example, during an interview with a client, consultants do not only focus on data collection, but can also discuss scoping and strategy [R3]. Furthermore, each activity can or cannot be supported by a tool. Additionally, each activity elicits of one or multiple sources. Revisiting the example above, a survey of suppliers elicits the source suppliers (or external stakeholders). A focus group with management elicits the opinions of management

(or internal stakeholders). Finally, each opinion of these sources can be visualized in a matrix, which is the product of a materiality assessment. By using this general model, we can describe materiality assessment regardless whether it is executed from the advisory, internal or audit perspective.

4. Results

4.1. The concept of materiality and materiality assessment

There has been quite some discussion on the concept of materiality (Puroila & Mäkelä, 2019), as one respondent noted: *"this has been debated for so long"* [R7]. Therefore, when discussing materiality assessment, it is important to clarify the term materiality first. The origins of the term lie in accounting, where materiality is defined in financial context by the financial accounting standards board (FASB) as: *"The omission or misstatement of an item in a financial report is material if, in light of surrounding circumstances, the magnitude of the item is such that it is probable that the judgment of a reasonable person relying upon the report would have been changed or influenced by the inclusion or correction of the item."* (FASB, 2018, p.9).

The essence of this definition lies in the omission of items that influence decision-making. An item is material if it can possibly influence the decision-making process of management, third-party auditors, tax offices or shareholders. However, this definition does not yet state the temporal aspect of materiality. This has been stated by for example Hoffman & Lydenberg (2013): *"a matter is material if it is of such relevance and importance that it could substantively influence the assessments of providers of financial capital with regard to the organization's ability to create value over the short, medium and long term."* (Hoffman & Lydenberg, 2013, p.2)[14]. These matters of value creation and strategy have also been raised by respondents [R4; R7]. It should be noted that this origin is in accounting is both a blessing and a curse (King, 2013). On one hand, it has heightened the awareness of economic, social and governance issues, spreading non-financial reporting. On the other hand, it has created a stumbling block for companies that do not wish to disclose this information due to legal and regulatory implications, while this would not be the case. This switch to non-financial materiality was also marked a notability during stakeholder consultations [R4*], where stakeholders might confuse it with its accounting counterpart.

Therefore, in the sustainability domain, Forstater et al. (2006) define materiality as: *"Material issues are those things that could make a major difference to an organisation's performance."* (Forstater et al., 2006, p. 9). With this, the aspect of systematicity is added to the materiality criteria. Forstater et al. (2006) discuss the matter of motor companies donating to cancer research. In theory, such a topic can be sponsored for a prolonged period of time. However, it is argued that it is not material as long as it remains unconnected to the business strategy. This has also been pointed out by R9 when discussing the move of Unilever's headquarters towards The Netherlands (in light of its recent media frenzy). Even though this theme has long term consequences for the company, it can be argued to be an incident, rather than a material topic.

Next, an issue raised by two respondents is the degree of influence an organization can exercise over the issue [R5; R6]. They argued that a topic is only material if the organization can influence it. Such a situation can occur when the ideals of a parent and daughter company clash. For example, let there be a parent airline company with a low-cost daughter firm. The parent company might consider CO_2-emissions reduction important and raise the price of tickets to compensate for the emissions. The daughter company finds this issue immaterial, as it contradicts with their strategy: offering tickets at the lowest possible rate. R10* disagreed, providing an example on Coca-Cola. Coca-Cola should still care about obesity and diabetes

[14] Although both FASB and IIRC use respectively item and matter, we use the term topic from now on for the same thing because this term was more commonly used in the interviews. In Dutch: onderwerp.

due to the large amount of sugar they put in their drinks, even though they have no control over the excessive consumption of its consumers leading to diabetes and obesity. R10* also questioned the claim of not having influence, as a company can hire lobbyists to combat these issues somewhere else in the chain. To shed more light on this issue, R3* and R4* mentioned that a topic is material if it has an impact on the company. From this perspective Coca-Cola is impacted by obesity and diabetes, so they should disclose their stance on these topics.

Furthermore, a problem with these definitions is that they do not explicitly include sustainability and business ethics, typically issues discussed in non-financial reporting. Therefore, in her thesis, Phelps (2016) coins the term 'sustainability materiality' as: *"…, materiality is the measurements of how much something matters."* (Phelps, 2016, p.7). Even though she acknowledges that the concept is more complex in its core, as it is dependent on whom, how much and why it matters, this definition is workable. However, this definition forgoes the systematic element. In addition, a note is required on the notion of sustainability as this is often defined too narrowly. Therefore, the Karlskrona Manifesto by Becker et al. (2015) calls for the extension of the term to include social, economic, individual and technical sustainability. We also include business ethics topics such as 'transparency' and 'corruption' under this sustainability tag, where business ethics are defined as: *"rules, standards, codes or principles which provide guidelines for morally right behavior and truthfulness in specific situations."* (Lewis, 1985, p.381). An alternative would be to refer to these topics as economic, social and corporate governance (ESG) or economic, environmental and social (EES) topics.

In short, materiality for non-financial reporting entails topics in sustainability and business ethics. The topic must be of importance in the long term and be of systematic nature. To prevent a page long definition, we define materiality as: 'the measurement of how much a sustainability topic recurrently matters in the long term'.

With this definition of materiality, let the term be combined with assessment. Once again, at first glance, multiple definitions exist, where many scholars and organizations have formulated their own:

"A tool to identify and prioritize issues that fall outside of traditional risk or financial management process, but have long-term implications for the firm." (Murninghan, 2013, p.4)

"[it] identifies which environmental sustainability aspects are material; which aspects should be disclosed, reported or managed." (Phelps, 2016, p.7)

"[it] helps companies identify and prioritize the material topics to report on benefits companies and ensures they are responding to investors' needs." (GRI & RobecoSAM, 2016, p. 36)

In additions to these, other scholars may refer to materiality assessment as materiality analysis (Calabrese, Costa, & Rosati, 2015; Font, Guix, & Jesús Bonilla-Priego, 2016; Muñoz et al., 2012), but do not provide a concrete definition of it. This confusion has spread to the industry, as one respondent noted: *"There is no real definition, GRI has one, which is OK, although it might not really be a definition but more of a process description."* [R1].

However, the core of these definitions is fairly similar. All authors discuss the notion of identification and prioritization of material topics. When asking respondents on a definition for materiality assessment, similar terms were used such as 'determination of topics that matter to your stakeholders (and strategy)' [R1; R2; R9], 'find the relevant issues they have an

impact in' [R6], 'provide insight into the most important topics for the future of the business' [R8]. One respondent compared materiality assessment to the packing of a backpack, where the right equipment should be collected for a hike [R7].

It appears that the focus from the respondents lies on the identification and prioritization of topics. However, that would neglect the reporting cycle which follows these tasks. As Phelps (2016) provided a workable definition, we mostly stick to it, with few adaptions. Therefore, we can state that in essence materiality assessment is 'a set of activities which determine which topics are material to an organization; which aspects should be disclosed, reported and managed.'

Although the concepts are clarified, there are two caveats mentioned by respondents. When discussing materiality topics, these are often composed of multiple topics [R1; R2; R6; R7]. Therefore, to avoid confusion, we propose to use the term 'materiality theme', or 'theme' when discussing the composition of multiple 'materiality topics', as this was coined by R1[15]. However, the precise number of topics per theme is difficult to determine. Respondents raised numbers from three [R1] to twelve [R8]. The second caveat is that the term 'materiality assessment' is not popular in communication, as illustrated by two respondents. *"Materiality assessment is a nice term, but it is not very well known in organizations. Stakeholder consultations are more important to everyone."* [R1]

"I don't call it materiality assessment myself, but I call it the determination of topics for the sustainability report that needs to be reported on, so they know which topics to focus on. That this has a difficult term, that GRI calls it like that, is something they don't care about." [R3]

It is difficult to determine what the implication of this is and further research is required to determine whether this affects the results of materiality assessments. R10* mentioned that as their clients often have an in-house sustainability officer, large clients are able to deal with this jargon. This means that maturity of materiality assessment and sustainability reporting in organizations as a whole might play a role here.

4.2. Discussion on methods in materiality assessment

Data from the interviews shows the usage of very different methods. When discussing the topic of methods, none of the respondents provided a precise name of a method they used [R1-R9]. Instead, they explained their adaptations of a high-level method. The majority mentioned they based it on GRI [R1, R3; R4; R7; R8]. In addition, none of the respondents mentioned explicit methods they were familiar with but did not use themselves [R1-R9]. This shows that there is no golden standard in the industry. An issue which was emphasized by R1 and R5.

The reason for these adaptations is frequently mentioned by respondents, as they replied that GRI is too high-level for them. The method does not concretely describe which stakeholders to interview [R7], how to prioritize issues [R8] or is not suitable to the time or budget [R3]. One respondent also found these methods to be more of a checklist, rather than a useful method [R5]. This results in many different approaches, which depend on the perspective of the user. This was also emphasized by R10*, who stated that in an audit it is not feasible to perform a stakeholder dialogue as the primary goal is to assess whether the materiality assessment is executed properly, while reproducing results is less important.

[15] Given the translation of the Dutch word: thema.

This lack of holistic methods also appears in literature. For example, Calabrese, Costa, Ghiron, & Menichini (2017) discuss multiple multi-criteria decision models, and the advantages and disadvantages of each method. They compared three methods, analytic network process, fuzzy analytic hierarchal process and multi-attribute-group-decision. However, these methods are mostly aimed at ranking topics, rather than identifying. This extended on their previous work, where Calabrese et al. (2016) show a way of ranking the topics through an analytic hierarchal process. They use the GRI indicators as an input and although the method's results are promising, we believe that this does not completely solve of which indicators to use, as the GRI indicators are not customizable to fit each company. It is therefore part of a method, rather than a complete one. Similarly, Whitehead (2016) also proposed a ranking method based on the saliency and risk of a topic for the wine industry in New Zealand. For identification a large set of indicators is combined.

This lack of holistic methods could also be caused by the fact that materiality assessment is often seen as a subprocess of sustainability reporting. In his thesis, Nummert (2014) analyzed the process of sustainability reporting of which materiality analysis is a part. He states that a good selection of key performance indicators (KPI) is important for materiality assessment. He also found that GRI's indicators are too extensive to be accepted. However, he noted that for the first time in the cycle, the GRI list serves as a solid base and it is recommended to use it. Borgert, Donovan, Topple, & Masli (2018) discuss multiple sustainability assessments, of which 'the identification of relevant sustainability issues' is mentioned as a sub-step of a sustainability assessment. They highlight this step, but do not really seem to discuss the prioritization of topics, as they state that a "materiality analysis provides a mechanism to engage with both internal and external stakeholders." (Borgert et al., 2018, p.104). They also state that it is unclear how a materiality analysis is applied and mention that further research is needed into the other steps of the procedural framework.

Another issue raised by respondents is the matter of standards [R1; R5; R8]. There are no standards currently on which steps to follow and how to disclose the information. This makes benchmarking with other organizations difficult and allows for many variations. This is also affected by the fact that materiality assessment is a value-laden and socio-political process (Puroila & Mäkelä, 2019). This was emphasized by R3, who mentioned that some issues are not discussed because the firm has no desire to. For example, an analysis of disclosed topics by peers (other firms in a similar industry) was not accepted due to this fact. Similarly, R4 mentioned that they wished some firms would dare to be more open on certain issues, allowing them to take a stance on a specific topic. Furthermore, Puroila & Mäkelä (2019) analyzed reports and find that most approaches are technic-rational. They identify multiple activities, mentioned by firms in their reports. However, they do not provide a clear process description of in which order these activities are placed.

Some scholars have already raised problems in materiality assessment in general. In their research, Guix, Bonilla-Priego, & Font (2018) analyzed the process of materiality assessment based on sustainability reports. They found it to be a black box, especially in regard to which stakeholders are weighted. This is also emphasized by Puroila & Mäkelä (2019) who conclude that materiality assessment fails to consider problems in equally measuring each aspect.

This black box is caused largely due to lack of disclosure, as not all companies are as transparent about it. Sarraj (2018) investigates materiality in sustainability reports of large companies in the Gulf States. He found that only 37% of the firms performing materiality

assessments do disclose a materiality matrix in 2014 and only 38% disclosed one in 2016. This is conflicting with the findings from the interviews, as all respondents mark the matrix as a product of the assessment. [R1-R9]. Additionally, he analyzed multiple sustainability reports over the course of four years (2013-2016) and found that only 50% of the firms reported on their materiality assessment (some more than others) in 2014, while decreasing to 40% in 2016. Meanwhile, GRI reports numbers of respectively 72% in 2014 and 89% in 2018 that disclose information on their materiality process (Taneva & Stracchi, 2018). This shows that considerable efforts should be made into what parts of the materiality assessment process should be disclosed.

Finally, multiple respondents discussed the interval in which a materiality assessment is conducted, as this is not prescribed by any method. However, most respondents agreed in a range from somewhere between a year and every three years [R2; R3; R4; R5; R8; R9]. More frequently would be futile, less frequently would make the organization lag behind. Although this depends on the type of firm, there is awareness that materiality assessment should become a continuous process which consists of a two-three yearly cycle. Every year however, quick checks are required to align the material issues with societal issues, as some issues might grow quicker than expected.

4.3. General activities in materiality assessment methods

The methods discussed in the interviews are specified with PDDs, where each step in the process is numbered. Activities with a similar nature are mapped into general activities through semantic content analysis. For example, 'pick stakeholders' and 'select stakeholders' are combined into 'determine stakeholders', as 'pick' and 'select' are synonyms. Through this, we identified fourteen general activities, which describe the practices of a materiality assessment. In Figure 4 an overview of how often these general activities were performed by respondents is given. It should be noted that the order of A1 to A14 is based on the general trend among respondents. However, this means that this could differ per respondent. Therefore, for the interested reader, a more detailed overview can be found in appendix D.

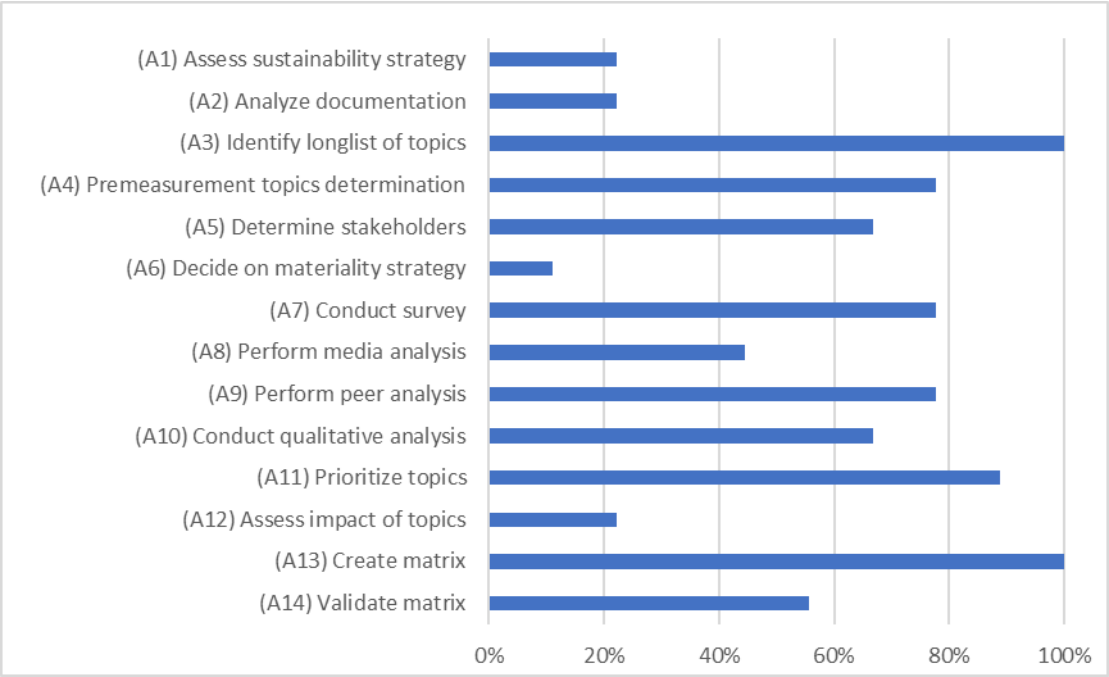


Figure 4. Percentage of respondents executing a general activity in their MA method.

The first activity is the **assessment of the sustainability strategy (A1).** The goal of this activity is twofold. One, the current strategy is investigated for material topics which are already part of it; these are likely to be still material. Two, this activity is conducted to align the materiality assessment with the sustainability strategy and prevent the assessment from being an isolated process. If there is no such strategy, consultants often start with creating one, although this is a different process outside of the scope of materiality. Only 22% of the respondents mentioned they started with assessing the current sustainability strategy (A1) [R4; R8].

The next activity is to **analyze existing documentation (A2).** This is an activity performed in the audit perspective, where documentation on the materiality assessment is analyzed. This documentation describes how the (previous) assessment(s) is executed; the inclusion of stakeholders, which topics were ranked and how this ranking process was performed. This step prevents reperforming the assessment, as the goal is to judge whether the assessment has been executed correctly. 22% of the respondents, the audit companies in our sample, mentioned this [R2; R9]. However, we argue that this activity should also be included in the method of other perspectives as well. R4* and R10* marked aligning this year's materiality assessment with previous years' as a big challenge. By analyzing this documentation from previous years, this can be (partly) incorporated. This is already being done by R7, whom included the fifteen most material topics from last year.

Then, an activity is to **identify a longlist of potential material topics (A3)**. This is an initial collection of all topics which could bear some material relevance. These topics originate from different sources. Some are gained by brainstorms with internal teams or by selecting them from the GRI's indicators list. Others are proposed by the consultant based on their experience with the firm's sector or arise by looking at prominent topics discussed in the media. Furthermore, sustainability reports by peers might discuss topics the firm should too. With all respondents [R1-R9] performing this activity and these many options, there are different approaches. The longlist is often created from scratch, as only a single respondent mentioned

they rely on a pre-existing list [R8]. R8 also conducts a media (A8) and peer analysis (A9) in parallel, to extend this list. R1 and R6 mention they have a pre-existing list they use as a checklist to ensure they have not missed any major topics. For a single respondent, the collection of topics is not among the first steps [R6], as they focus on determining stakeholders (A5) and formulate a materiality strategy (A6) before this activity (A3).

After this, the next activity is to **apply some premeasurement to this longlist (A4)**. This activity encompasses all steps in which the longlist expanded or reduced to a shortlist before consulting stakeholders. This is important, as the longlist gathered in the previous topics often contains too many topics to collect data on; stakeholders would have to provide their opinion on too many topics. 67% of the respondents perform this activity [R1; R2; R6; R7; R8; R9] but use different approaches. R1 and R7 follow up on the previous activity (A3) and narrow down this list based on their expertise. From an audit perspective, where the longlist is already distilled from the documentation. R2 performs a media analysis (A8) before this activity (A4) to identify which topics should be present in the list. Meanwhile, R9 performs a peer analysis (A9) with the same goal.

Then, an activity is performed to **determine the stakeholders (A5)** to elicit information from. This encompasses all steps from identifying these stakeholders, to reaching out to them and to assigning weights to them according to their importance. 67% of the respondents mentioned performing this activity [R1; R3; R4; R6; R7; R8], while the two auditors (R2 and R9) did not. However, this is not surprising as, in an audit, these stakeholders are already known. These stakeholders can be determined based on consultant's experience, where they propose a set of stakeholders which is validated with the client [R3]. However, they can also be identified through a brainstorm or discussion with a client [R1; R4].

The next activity is to **construct a materiality assessment strategy (A6)**. In this activity, a strategy and action plan are formulated. This includes matters such as the number of stakeholders to reach out to and which instruments to deploy. Furthermore, metrics should be determined. For example, magnitude, which illustrates how much impact the organization has on a specific topic. Another example is social maturity[16], which is a metric on how well the topic is known in society. This scale ranges from awareness in small groups, e.g. non-governmental organizations (NGOs) who care about the topic, to the existence of regulations on the topic where the consequence of incompliance is being fined. Only a single respondent [R6] mentioned the construction of such a strategy. However, R5 mentioned the assembling of a team, prior to starting. Furthermore, although R7 did not mention the construction of this strategy, they discussed measures such as speed and velocity. Here, speed is a metric on how quickly the topic is gaining or losing importance. This is an estimation of when the topic will become material or how long it will stay material. Velocity is how large the impact of the topic on the organization will be (which can also be used as a matrix axis as is common practice by R4), i.e. how important it is for the topic to deal with the topic. Furthermore, a notion on R6 is required, who start with applying the theory of change, which is a technique of backwards goal-reasoning. In this technique, they start by identifying long-term goals; creating a more sustainable strategy. Then, they come up with requirements which will fulfill this goal and come up with interventions to fulfill these requirements. Indicators are defined to measure the outcome, i.e. what factors the change causes to be successful. By reasoning backwards, it

[16] Term as used by respondent, not to be confused its counterpart in psychology.

becomes clear which stakeholders are affected, as these appear in the interventions. However, as this theory of change is an instrument to help materiality assessment, it is not classified as a general activity, but as a part of the materiality strategy.

Next, an activity is to **conduct a survey (A7)**, which is often deployed to elicit information from external stakeholders such as suppliers and customers. R5 is an exception here, as they also used a survey to elicit information from their internal stakeholders such as employees. A survey can consist of multiple formats. Some surveys ask their stakeholders to rank certain topics on a scale of one to ten, although a Likert scale of one to five is often preferred. However, surveys can also request the stakeholder to order topics from most important to least important. This is the most common quantitative instrument, as 78% of the respondents mentioned using this activity [R1; R3; R4; R5; R5; R7; R8; R9]. A caveat for surveys is required. Multiple respondents [R3; R4; R5; R8] mention that downsides of a survey include a low response rate, thus jeopardizing the validity of conclusions, and that the lack of control on how the questions are interpreted, which could cause discrepancy between stakeholders.

In the next activity, respondents **perform a media analysis (A8)**, in which they check for topics in the media. This activity is often performed manually, where the team constructs some terms based on experience with the firm and sector, and searches for them online. It should be noted that the media analysis is often not used as an input for the analysis, but rather as a check whether any topics have been missed. This is not a thorough check, but rather a non-exhaustive exploratory search relying on serendipitous discoveries. 44% of the respondents use this systematically in their assessment [R2; R7; R8; R9]. However, the timing of this activity in the process differs per respondent, as R2 performs it before A4, R8 performs it at the same time as A3, R9 performs it after A3 and R7 performs it after A12. The reason for the difference between these timings is unknown, but it could be due to the lack of standards on the usage of media in assessments.

The next activity is where respondents **perform a peer analysis (A9)**. Here, topics are elicited from peer reports of firms in a similar industry. This is often done through reading the section on materiality in sustainability reports and is not used as a quantitative source. 78% of the respondents perform this activity [R2; R4; R5; R6; R7; R8; R9], but the timing differs per respondent. R2, R6 and R7 use it as a check for their own analysis after the topics are already prioritized (A11). R4 performs this at the same time as A7 and A10, as it provides input for the prioritization. R5 used it to briefly compare their own findings with their peers but did not find very useful, because they felt it prevented them from telling their own story. R8 and R9 used this activity for the initial collection of topics.

Furthermore, companies **conduct a qualitative analysis (A10)**, where they elicit information on the importance of topics from (mostly) internal stakeholders. For this, one or multiple instruments, such as workshops, focus groups, brainstorms, interviews and round-table discussions are deployed. These are often executed with internal stakeholders and are aimed at making some sort of prioritization of topics. Although the outcome of this kind of sessions is often qualitative, R3 mentioned that they apply some mechanism to quantify this information. For example, if a specific term is mentioned in an interview, they give it a point, eliminating this qualitative aspect. In other cases, this is done based on discussions [R4], where the placement of the dot on the matrix is based on a 'feeling'. 78% of the companies were found to use some sort of qualitative analysis instrument [R1; R3; R4; R6; R7; R8; R9]. R2 did not, as they mentioned that these instruments are already deployed by the firm itself. Additionally,

R5 did also not use qualitative instruments, as they felt it would not change the outcomes of the survey but would make prioritization more complex.

After all data is collected, the next activity is to **prioritize topics (A11)**. In this activity, topics are ranked from least to most material. There are multiple ways to do this. When quantitative data is available, it is a matter of aggregating this data and computing the scores. However, when there is qualitative data, some other sort of ranking method is required. This can be performed by consultants themselves but is often done in a workshop or discussion session with the client, where the client orders the topics from most to least important. If that is the case, this activity can be combined with A10. All respondents [R1-R9] perform this task. For R3, this activity is not explicitly mentioned, as they quantify the data to create the matrix (A13). The timing of this activity is however varying between respondents, as R2, R6, R7, R8 and R9 already perform this activity before completely finishing with activities A7 to A10. In their cases, one of the previous activities are used as a measure of control; to determine whether this prioritization is somewhat correct. Meanwhile, R1 and R5 create a prioritization after which they move on to the creation of the matrix (A13).

The next activity is to **assess the impact of topics (A12)** on the firm. In this activity, the metrics that are determined in A6 are operationalized into KPIs. Only 22% of the respondents actually incorporate this impact assessment [R6; R7].

Next, respondents **construct a materiality matrix (A13)**, which is a communication instrument to easily demonstrate the most material topics. In A11, each topic has been ranked for a specific source, e.g. topic A has an importance of five to stakeholders, three to media and two to peers. Meanwhile, topic B has an importance of three to stakeholders, one to media and five to peers. To visualize these topics in a matrix, these sources are reduced to two dimensions, where for example scores of stakeholders are mapped to the y-axes, and scores of media and peers are aggregated into the x-axes. A metric from A12 can then be used as a z-axis, but this is optional. All respondents perform this activity [R1-R9], but produce different matrices as further described in section 4.8.

The final activity is to **validate the results (A14)**. This is often done in a discussion with (senior) management. It is important to link the sustainability strategy with the business strategy, as when this step is skipped, the risk of creating an isolated sustainability strategy increases [R4]. Meanwhile, only 78% of the respondents mentioned that they validate the matrix [R1; R2; R3; R4; R6; R7; R9].

4.4. The usage of PDDs

To illustrate how these general activities are created, two PPDs are given in Figure 5 and Figure 6. For each diagram, the mapping to the conceptual model is explained. The complete left part of the diagram is the method, where each activity is labeled with an identifier. Note that this number is unrelated with the order in which these activities are executed. The goal of each activity is modeled as a concept, such as data, ranking or topic. Sources are modeled in the diagram implicitly, as they are often latent in relations between concepts. The matrix is the output of the activity. Finally, tools are not modelled in the diagram, as the usage of many tools would overwhelm the diagram, making it messy and too difficult to interpret. The full collection of PDDs is included in appendix J.

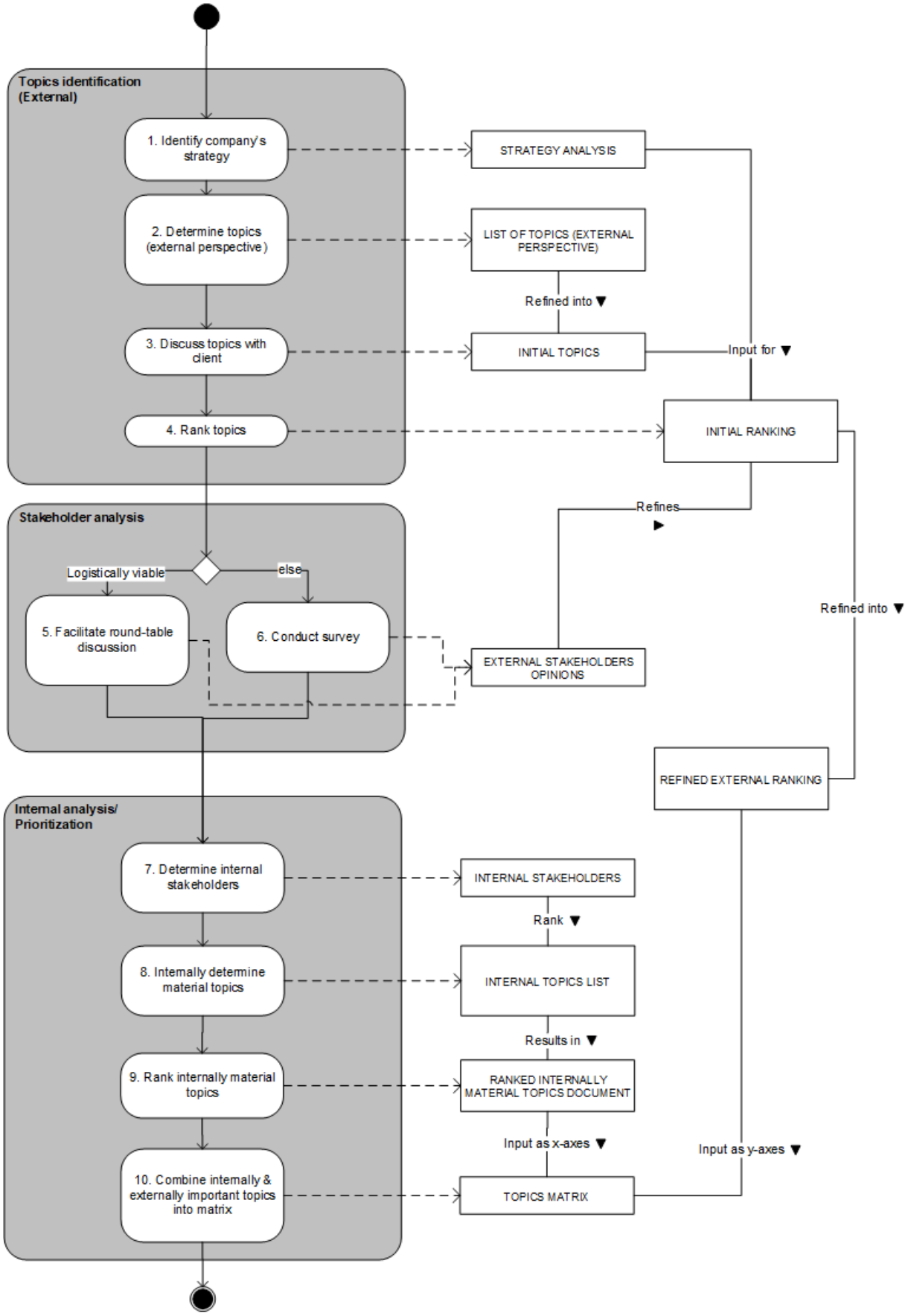


Figure 5. PDD of respondent 8.

Starting with Figure 5, R8 has a clear distinction between the external and internal perspective, starting with the external one. This makes it also difficult to map to general activities, as half of the general activity is performed for the external perspective and the other half is performed for the internal one. Starting with the first activity, it assesses the strategy of the company. This is mapped to GA1[17]. Then, a longlist of topics is determined, through peer and media analysis. This is a combination of multiple GA's. Therefore, this activity is mapped to GA3, GA8 and GA9. Next, in a discussion with the client, this is discussed, where this list of topics is narrowed down. This is some premeasurement of topics (GA4). In activity 4, an initial prioritization is given to the topics (GA11). Then, activity 5 and 6 collect data from the client, either through a survey or through a round-table discussion. A round-table is preferred, however, sometimes it is not logistically viable. This is a qualitative activity, therefore mapped to GA10. Meanwhile the survey in activity 6 is mapped to GA7. Moving on with activity 7, internal stakeholders are defined, which is mapped to GA5. Activity 8 is the process of internally determining these topics, with e.g. management. Therefore, this is as mapped to A3. These are then ranked with the internal stakeholders (GA11) after which in activity 10 the two prioritizations of A4 and A9 are combined into the matrix (GA13).

R3 in Figure 6 on the other hand does not have this clear distinction between the external and internal process. R3 starts with listing all the topics that could be material, which is mapped to GA1. Next, they already provide some weights, to narrow down this list (GA2). Then, in activity 3 and 4 they determine and present stakeholders so they end up with a list of them (GA4). Next, they execute a number of data collection techniques. Not all of these techniques are necessarily executed, but this can contain surveys (GA7), website analysis and interviews. The latter two are grouped, in combination with activity 8 and mapped to GA8. Then, the matrix is created in activity 9 (GA13) and validated in activity 11 (GA14). Finally, they provide an additional activity 11. This is not mapped to any general activity, as others did not mention this step.

[17] Abbreviated as GA instead of A for clarity between general activity and activities in the PDDs.

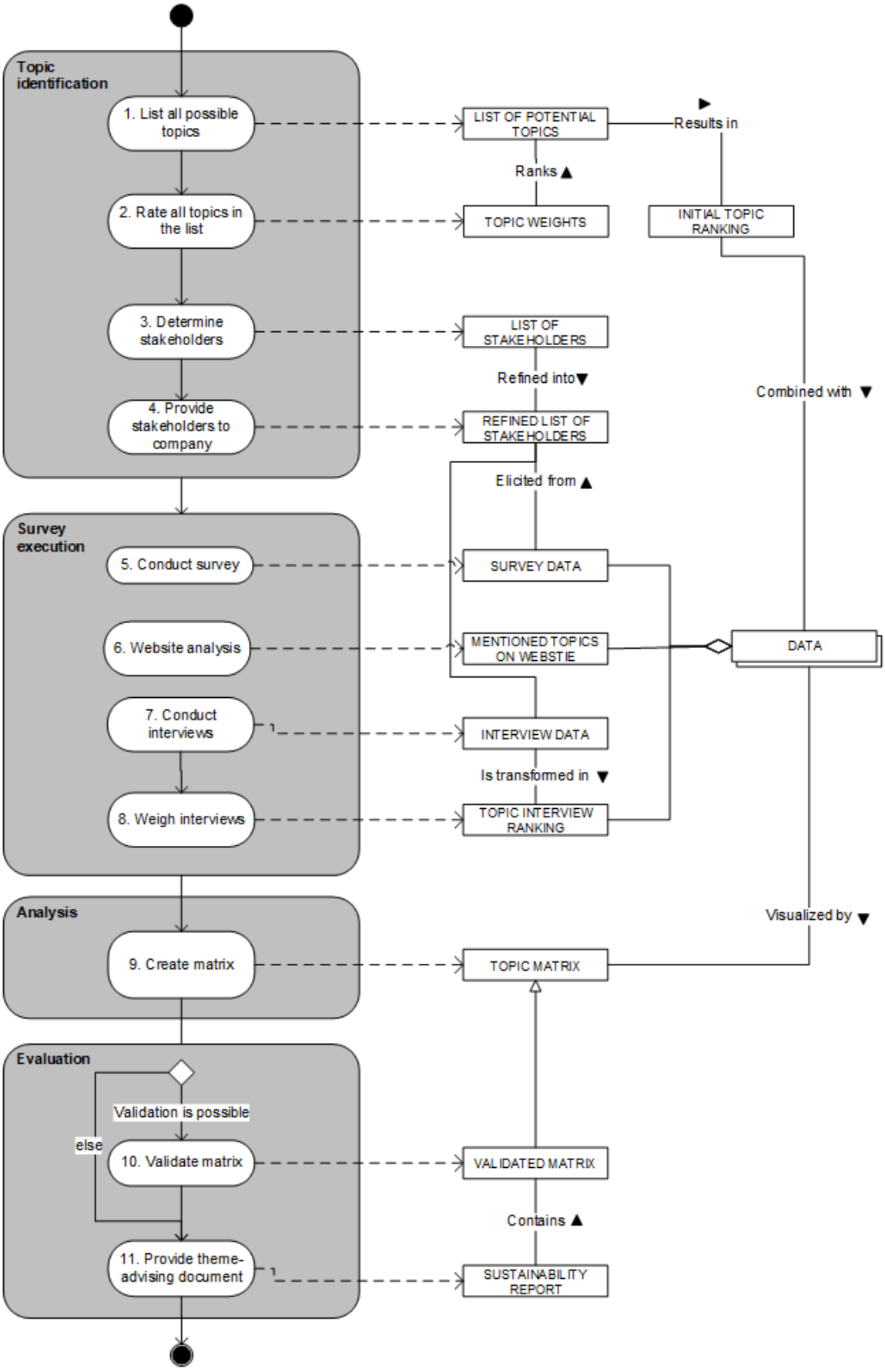


Figure 6. PDD of respondent 3.

4.5. Goals of activities

As mentioned in section 4.2, whether these activities are executed in an assignment is subject to change. In Table 2, an overview is given for which activities are performed by which respondent. Each respondent has been marked with its perspective, either a C for consultant, I for internal or A for auditor. All activities marked with a * are revisited later in the process. Activities which are combined with a + are executed in more or less parallel.

Table 2. Activity order per respondent.

R1 (C)	R2 (A)	R3 (C)	R4 (C)	R5 (I)	R6 (C)	R7 (I)	R8 (C)	R9 (C/A)
A3	A2	A3	A1	A6	A6*	A3 + A9*	A1	A2
A4	A3	A4	A3	A3	A5	A4	A3 + A8 + A9	A9
A5	A4	A5	A5	A7	A6*	A11*	A4	A3
A7	A8	A7	A10*	A9	A3 + A10	A5	A11*	A8
A11	A11	A10	A7	A11	A4	A7 + A10	A10*	A4
A10	A9	A13	A9	A13	A11*	A11*	A7	A11
A14	A14	A14	A10* + A11		A9	A12	A5	A10
A13	A13		A11		A14	A9*	A3*	A7
			A13		A12	A8	A11*	A13
			A14		A11*	A13	A13	A14
					A13	A14		

The table above shows how varied the approaches to materiality assessment are in terms of order. This results in a question on what the minimal set of activities required is to perform a materiality assessment. We can create this set based on the goals of the activities, where activities with a similar goal are grouped into a higher-level group, which allows for a comparison with the other high-level methods. In general, five goals are distilled from the fourteen activities. These goals are similar to steps which should be executed in a materiality assessment.

1. Topic identification: assess sustainability strategy, analyze documentation, identify longlist of topics, premeasurement topics determination
2. Planning: determine stakeholders, prioritize & pick stakeholders, decide on strategy,
3. Data collection from stakeholders: conduct survey, perform media analysis, perform peer analysis, conduct qualitative analysis
4. Ranking: prioritize topics, assess impact of topics, create matrix
5. Validation: validate matrix

At this high level, these steps can be compared to other guides from institutions. In the systematic literature review, to validate these steps in practice, they were compared some of the 43 materiality reports. These five steps can be identified in these reports, but a gap in terms of execution appears between companies. Some examples are given of this.

In terms of topic identification there is a difference in the number of topics included on the list. For example, Heijmans starts with a longlist of 203 topics from internal documents. Meanwhile, Q-park uses a longlist but does not mention how many topics it contains. Other firms do not discuss the use of a longlist at all.

An example of planning is demonstrated by NIBC, who show which stakeholders they identified and which instruments are used on them. UPL just shows five stakeholders, while H&M had ten, so this differs per firm as well.

Data collection also depends per company where the amount of collected data can differ. Arkema collected more than 2200 answers to their questionnaire in their data collection. In contrast, Swiss Re does not use a survey, but relied on dialogue with their stakeholders.

In ranking, Autogrill and Enel do not discuss how they actually rank the topics, while H&M provides an almost complete formula. Furthermore, there is a difference in how detailed the topics are described. Credit-Suisse provided detailed descriptions of their topics during the ranking, while others such as Orica just provides the material terms. These differences clearly show a difference in the maturity of organizations.

Finally, Unilever provides details on validation of the assessment, while others such as Huhtamaki do not.

As it is clear that each company has its own aspects to report on, there is no definitive method either. Some organizations included last year's assessment and some are monitoring their topics. Meanwhile, a GRI report provides typical materiality assessment steps (Taneva & Stracchi, 2018), which also focuses on the steps beyond materiality. Additionally, Datamaran provides seven tips (Yumasheva, 2018), where they recommend benchmarking against leaders and laggards in the same industry. However, as shown in previous paragraph, with such a big difference in activities, determining these is difficult. Similarly, some organizations are oriented towards their outer stakeholders, while there is little internal support for the project or vice versa [R4*].

The examples above illustrate that benchmarking in the industry is difficult and applicable methods are scarce. We argue that a potential cause for this caused by the lack of a taxonomy for firms practicing materiality assessment. Even though it might be obvious that two thousand respondents on a survey are better than two hundred, the ideal balance between quantitative and qualitative methods is unknown. This makes the classification as leaders and laggards at best subjective and at worst impossible.

A potential solution for this would be to develop a materiality maturity model. This is a way of comparing methods and applying a classification for benchmarking (Khoshgoftar & Osman, 2009). Such a model can identify organizational strengths and weaknesses and organizes them accordingly. This allows organizations to standardize, measure, control and continuously improve the process (Khoshgoftar & Osman, 2009).

4.6. Usage of (software) tools in materiality assessment

Little scientific work has been performed on materiality assessment tools and the work that does often discusses methods instead of tools (according to our classification). Recently, Wu et al. (2018) discusses four tools in materiality assessment: SASB materiality map, GRI's sustainability topics [list], sustainability disclosure database and social life cycle perspective. The latter is more of a method, demonstrating that the notion of tools and methods is coupled. In this study, they found that two tools do not obtain the same material topics. They also reckon that the GRI database is a good start for the identification of material topics during preliminary assessment, which is comparable to using a longlist of topics. Additionally, a tool such as the Materiality Assessment Tools (MAT) exists but does not support dimensions other

than internal and external stakeholders. Furthermore, Datamaran is a tool which supports many of the general activities and the only holistic tool we found.

In general, the software tools can be divided into three classes. First, there are the holistic materiality assessment tools. These tools support many and ideally all of the activities in materiality assessment. Second, there are activity-specific tools. This set of tools support a single activity, such as survey conduction or media analysis. Third, there are general support tools, such as Microsoft Office. These are all-purpose tools which are used for reporting and data analysis but have no specific link to an activity.

In practice, we found that many organizations use software tools to collect data and process results. However, few of these used tools are tailored towards materiality assessments. Most often, tools are activity-specific, for example by deploying a tool to conduct surveys [R1; R3; R4; R7; R8]. In some cases, this was supplemented by (social) media analysis tools, where the status of the company is monitored [R2; R7; R8; R9]. Three out of nine respondents used Datamaran on a regular basis [R4; R7; R8]. However, R4 stopped using it after a while due to the price tag. The use of activity-specific tools Polecat and Quid was mentioned by R2. R9 mentioned the use a risk analysis tool[18]. In this class, another widespread tool used by all respondents is a survey tool, whether that is their own or an easily accessible one such as Google forms or SurveyMonkey. General purpose tools are used by every respondent, as Google and Excel are basic tools widely spread to each respondent.

4.7. Sources of information in materiality assessment

There are many potential sources to draw from during materiality assessment. To provide insight into which sources are used, Table 3 shows the distribution of sources per respondent and the instruments used to elicit them. When a source is elicited, but not systematically for every project, it is marked with a ~.

Table 3. Elicitation instrument on sources per respondent

Rn	Internal stakeholders	External stakeholders	News	Website	Social media	Peer reports	Pre-existing longlist
R1	Int + FG	Survey					MA
R2			LexisNexis + Quid + Google search		Polecat	MA	
R3	Int	Survey	~	MA			
R4	Int + FG	Survey	~	MA	~	MA	
R5	Survey	Survey					
R6	Int + Brainstorm	Survey		MA		MA	MA
R7	Survey + Int	Survey	DM		DM	DM	DM
R8	Int + FG	Survey	DM		DM	DM	DM
R9	Int	Survey	Google search + RAT		RAT	MA	

To save space, some terms are abbreviated, of which the full names are enclosed in Table 4.

[18] Name undisclosed to ensure confidentiality.

Table 4. Abbreviations for table terms

Abbreviation	Full
Int	Interviews
DM	Datamaran
MA	Manual analysis
FG	Focus groups
RAT	Risk analysis tool
~	Used, but not explicitly in assignments.

Table 3 requires several noteworthy mentions. First, R2 is the only respondent that does not use internal and external stakeholders. This has been discussed in the interview, where they noted that due to their role as an auditor, these sources are utilized by their client, not by them. If they were to use these sources, their independence could be at risk [R2]. This is a contrast with R9, who approached the topic from both the roles as an auditor as well as an advisor.

Second, R5 is the only respondent to use purely survey data from internal and external stakeholders, while other parties use qualitative instruments for internal stakeholders. However, R5 did mention that their process is not very mature yet and they are looking to expand their process. For externals, surveys are the only instrument used.

Third, news is a source used by 44% of the respondents [R2; R7; R8; R9], however a caveat is required. R3 and R4 mentioned using news but read it to stay-up to date. They do not have a method to translate this into results for the materiality assessment. Those that use news, use advanced tools to analyze them, although do not blindly rely on them. These tools are often checked by a manual Google Search. This practice was also used by R10*. Another caveat is that maturity of the company seems to play a role here. Three of the four respondents using news systematically are large firms, whereas smaller firms do not. This is something which is also raised by R6, as they see value in analyzing news, but lack the financial resources to do so.

Fourth, there is the analysis of the website. This is clearly a practice performed from the advisory perspective [R3; R4; R6], where consultants analyze the website of their clients to find whether some topics are already mentioned there. These consultants perform this practice manually.

Fifth, social media is not utilized by everyone. In general, there is a divide between large and smaller companies again, where larger companies are more likely to use it. A special note is required on R4, who mentioned that sometimes they do use social media. However, they relied on the social media monitoring team of their client, meaning they have no method or tool of their own. All parties that utilize social media use a tool for this [R2; R7; R8; R9]. This can be either Datamaran, or a combination of tools with a Google search. In this Google search, the name of the company is combined with terms relevant to the industry the company is in. For example, for an oil company, terms to Google on are 'spill', 'leakage', 'safety', 'CO_2'. However, R2 feels that more could be done to use it more systematically and robustly.

Sixth, the usage of peer reports is common practice in materiality assessment [R2; R4; R5; R6; R7; R8; R9]. However, this source is also used for reference, rather than input [R2; R6]. 89% of the respondents acknowledged however that they found peers an interesting source to include, as it is suspicious when a single firm does not report on topics its peers report on.

Additionally, R4 mentioned that a peer analysis can identify opportunities, as topics which no peers report on can be used to take a stance on, taking the lead in driving change on a specific topic.

Finally, 44% of the respondents use a pre-defined longlist [R1; R6; R7; R8]. However, R1 and R6 noted that this is a tool for reference, rather than input. R7 and R8 rely on the longlist of Datamaran, which is used in the tool to analyze the news and social media on those topics. This is important, as all respondents do create a list of topics at some point to discuss with their clients. However, this list is never written down and made explicit, for numerous reasons ranging from a specific choice to prevent bias [R4] to not being worth it [R9]. R5 mentioned that their consultant provided a long-list, but they disregarded it as they felt it constricted them too much in reporting on the topics they find important.

Furthermore, it must be noted that there are more sources than these seven. For example, R7 mentioned the usage of fourteen different sources, including speed and velocity, in which they analyzed the growth of material topics and the likely impact it would hit them at. The rationale to not include these is twofold. One, R7 is the only one to mention them. Two, according to our classification, these are not sources, but metrics constructed based on sources.

4.8. Materiality matrix

The product of materiality assessments is often a matrix [R1-R9]. This matrix consists of two axes, where each topic is plotted. These axes are often split to create four quarters, where in general the topics in quadrant two are material, illustrated in Figure 7 (Saenz, 2019).

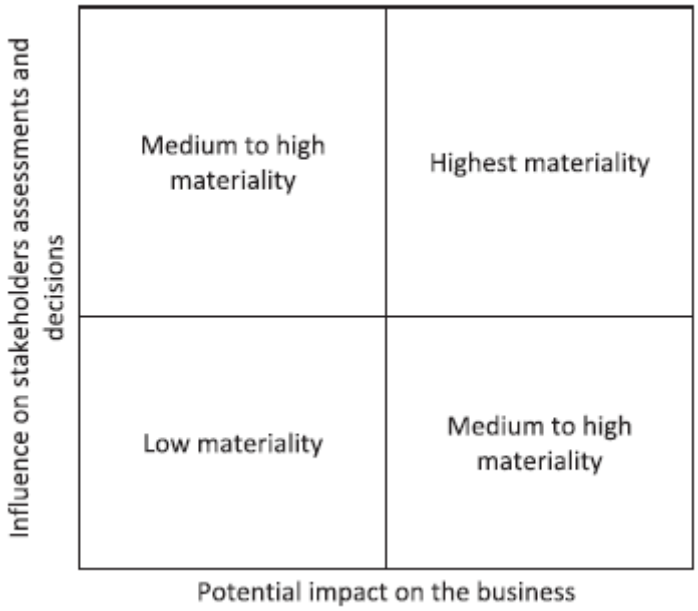
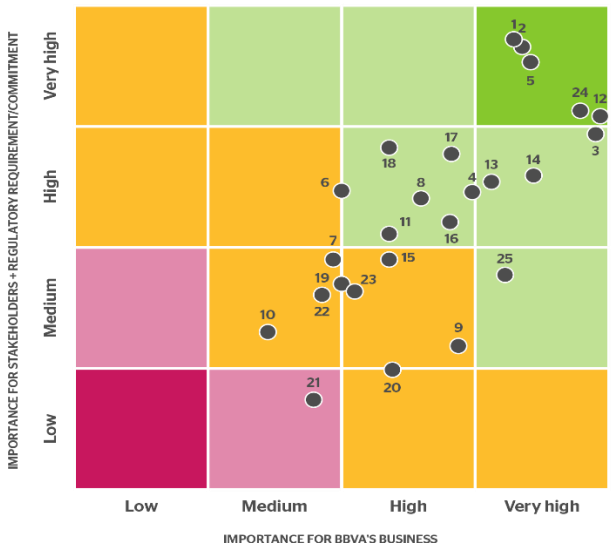
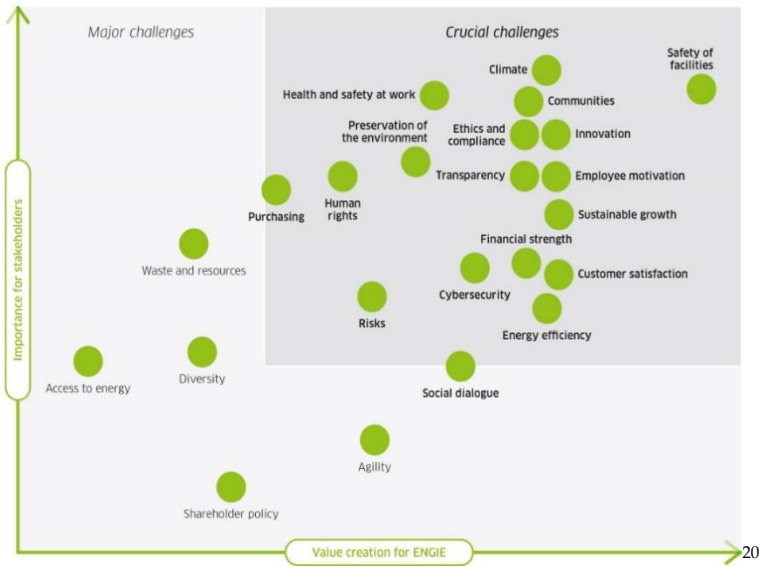


Figure 7. General principle of materiality matrix (source: Saenz, 2019)

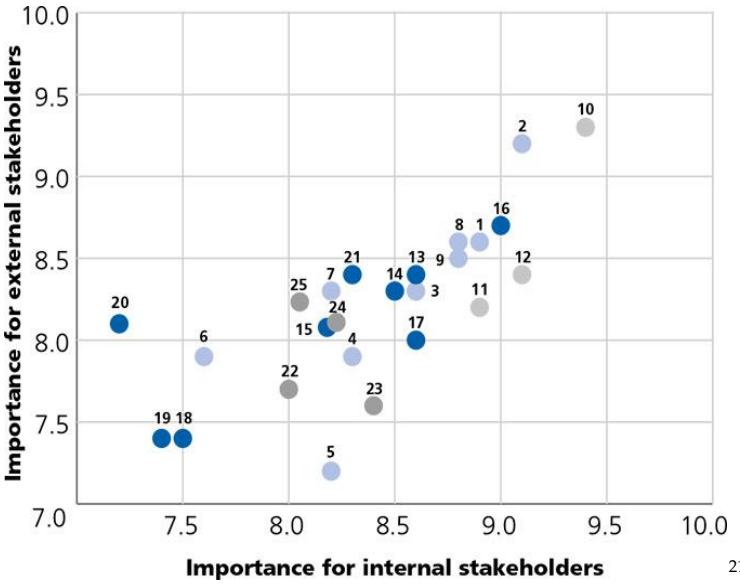
However, there is no uniform way to visualize a matrix. Instead of quarters, curves can be used as a decision boundary between (im)material topics. Topics can be represented as a bubble of different sizes, with different colors and different meanings. Furthermore, the axes can change. These differences are illustrated in Figure 8.



19



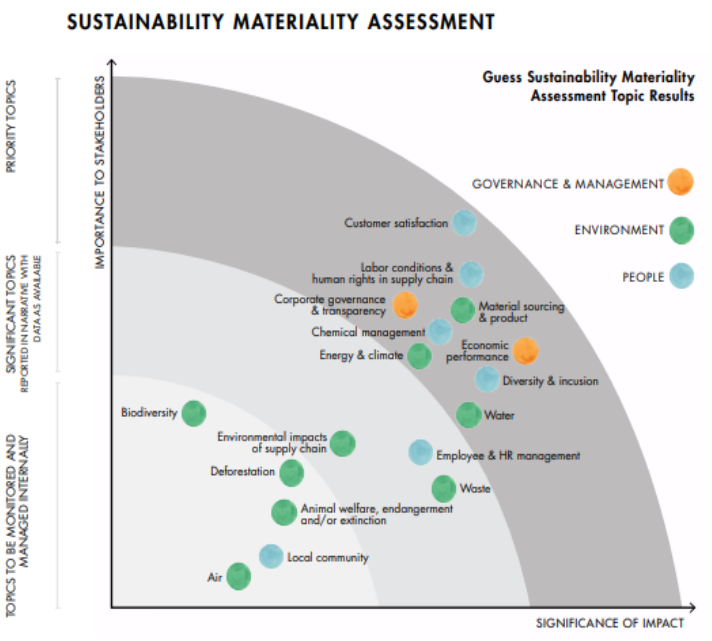
20



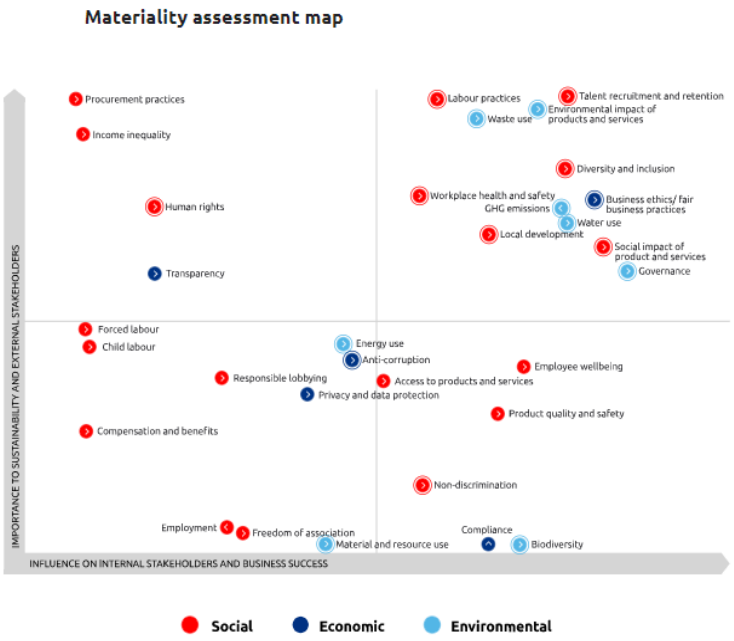
21

22

[19] BBVA, materiality report, 2015
[20] Engie, materiality report, 2017
[21] Michelin, materiality report, 2018
[22] Pernod-Ricard, materiality report, 2019



23



24

Figure 8. Examples of materiality matrices

These matrices have some similarities and differences, and these examples illustrate how diverse the application of matrices is. For example, some visualize a z-axis through color-coding of the bubbles, while some do not and when some do, the colors do not bear the same meaning. Michelin and Pernod-Ricard use color to group their topics into themes. Meanwhile, Guess and Swire-pacific color the topic by respectively ESG and EES categories. Furthermore, all six matrices have a different x-axis, where Michelin uses the most traditional one. The y-axis is however more commonly phrased as the impact/importance on/for (external) stakeholders. Finally, it is noteworthy on how the topics are grouped. BBVA uses sixteen fields, whereas Swire-pacific uses the same principle for four. Engie uses only two fields; topics that require immediate attention and those require attention later. Pernod-Ricard and Guess use curves, which by the look of the eye seem arbitrary.

In practice, the common axes are based on GRI: importance to internal stakeholders versus importance to external stakeholders. This is common practice as one respondent stated: *"That is pretty standard, I don't know [why]."*, *"we use these two [dimensions] because GRI prescribes them."* [R1]. Variations mentioned for the external axis are impact on society [R2;R6], impact of the world on your organization [R4]. Internal can also be replaced by impact on organization[R5], importance to stakeholder [R2; R4], internal perspective [R8]. The use of these alternatives alters the semantics little. Only a single respondent [R6] mentioned the use of a bubble diagram, in which they used a z-axis. This bubble can represent the legitimacy, dependency, impact or urgency of the topic.

Puroila & Mäkelä (2018) analyzed 29 sustainability reports and describe issues on the purpose of materiality assessments. They state that often the focus lies on creating short-term financial value and do not truly capture the nature of materiality. A particular issue is the fact that the matrix only represents two dimensions and fails to show traceability on how these dimensions were constructed and which sources are combined. Furthermore, they argue that materiality

[23] Guess, sustainability report, 2017
[24] Swire-pacific, materiality report, 2017

assessment is a controversial process, as more powerful stakeholders exert more influence over which topics are highly ranked. Additionally, the matrix does not provide a unified understanding among stakeholders as it is too simple to capture the complexity of the process (Puroila & Mäkelä, 2019). Even though they propose an alternative, firms are often locked in in their ways of working that might therefore not be easy to change (Khalil, 2013).

5. Towards a research agenda on materiality

In this paper we have discussed the practices of industry experts in materiality assessments, where several problematic issues are mentioned. Therefore, to conclude, we propose a research agenda of multiple issues that can be tackled by research into this field.

Many organizations differ in maturity from each other. In section 4.5, the difference between many ideal materiality assessments and the activities taken in practice is shown. To move from practice towards the ideal scenario, we propose to investigate the use of a materiality model. This model can show step-by-step which activities are required to take to improve an organization's materiality assessment process. Currently, no such model exists, even though it could provide a benefit to organizations [R3*; R4*; R10*]. A maturity model on integrated reporting exists, but views materiality as a guiding principle; it does not provide details on the measurement of the assessment (Deloitte & MVONederland, 2015). Meanwhile, another maturity model on integrated reporting measures materiality as part of maturity. However, this part is scored based on how well it is discussed in the report (Mazars, 2015).

Furthermore, the industry suffers from lack of accessible and holistic tools. As illustrated in sections 4.6 and 4.7, there is no tool to support all the activities in materiality assessment. Furthermore, many practitioners focus on internal and external stakeholders and do not utilize potential of social media and traditional news outlets. Therefore, we propose to develop a tool which supports these activities, from data collection to matrix generation. An important criterion of this tool is flexibility. As shown in section 4.8, matrices differ from each other. Therefore, it is important that the matrix is customizable to suit the consultants' way of working. Additionally, as also mentioned in the previous section, transparency is a requirement (Puroila & Mäkelä, 2018). Therefore, the tool must provide clear information on how axes are constructed and sources are combined.

Another issue raised by respondents and also illustrated in section 4.5 is that there is little knowledge on the operationalization of the activities [R10*]; who of the stakeholders are more important and what weight should be assigned to them, how many stakeholders are needed, how many topics a longlist should contain, who creates and maintains the longlist, whether topics should be aggregated into themes or not and how to divide this. More research is needed to discover the current industry practices, identify good practices and offer situation-dependent guidelines.

Next, there is gain in developing the ideal materiality assessment. An organizational challenge is to link risk assessments from the risk department with the materiality assessment in the sustainability department [R7]. Another trend is moving towards continuous assessment, where topics are more closely monitored during the assessment cycle [R8]. A way of doing this is by defining KPIs, but more research into how this monitoring can be applied is required. Furthermore, materiality assessment mainly leans on opinions of stakeholders. Only two respondents [R6; R7] mentioned the usage of other metrics such as impact, magnitude or speed. More research is needed into how to apply these extra metrics; what metrics there are, their benefits and drawbacks and how to measure them.

Finally, there is a set of smaller issues. First, three perspectives are identified in section 3.1. However, it is unknown how these perspectives precisely influence materiality assessment and whether different methods, standards and guidelines are required for each of them. Further research should investigate how these perspectives influence materiality assessment.

Second, the lack of standards in materiality assessment methods are an issue [R5]. At the time of writing, there are no standards on e.g. the number of stakeholders, sources and measures are required. This results in many parties executing different methods, making comparisons between them virtually impossible. This has also been emphasized by Puroila & Mäkelä (2019), who state that each source should be treated equally; being passed through the same filters and with the same settings. Therefore, research should focus on providing a framework for this standards, which ones are required and what they should look like. Third, section 4.7 shows that usage of surveys for the elicitation from external stakeholders is a common practice. However, multiple respondents mentioned that they disliked the instrument, as questions are often ambiguous and there is a low response rate [R3; R4; R8]. This results in opaque results, where there is little control of whether the result shows what the participant actually meant. Therefore, research should investigate how this response rate can either be increased, or whether there are other large-scale instruments to elicit this information.

6. Discussions

As every research, this one has its limitations. However, when discussing limitations in qualitative research, there is a an overwhelming number of possible limiting criteria in qualitative research (Whittemore, Chase, & Mandle, 2001). Therefore, some decision on which criteria to use is required. As this paper follows the design science methodology by Wieringa (2014)[25], its validity criteria are selected and discussed.

First, the industry model (Figure 2) is a hypothesis, based on findings from the interviews and not grounded in other scientific work. Although this model is validated with three respondents, R3* could not classify themselves unambiguously within the model.

The conceptual model describing materiality assessment (Figure 3) has a foundation in a simple process model. Practitioners mentioned that it would make sense to create a link between goal and matrix [R3*; R4*]. This is not added, as this would imply that every goal of activities is related to the matrix. Additionally, it can be debated that the concepts of topic and theme should be part of the model, while they currently are not. Similarly, this model does not discuss value of the materiality assessment [R3*]. These concepts are not incorporated into the model as its relations with other concepts would be arbitrary guesswork. Furthermore, as the nature of this study was exploratory, this model was created based on data from interviews. To rigorously discuss materiality assessment, these concepts also need to be investigated in interviews. As this was not possible, adding these concepts would lead to incompleteness. However, we acknowledge that an extension of the model has potential.

Next, the PDDs are modelled by a single researcher based on the interview data, after which another researcher adapted these models based on the data. Especially in the advisory perspective, the materiality assessment method is subject to change, depending on the assignment. Therefore, it is difficult to model a general method of an advisory firm and mark which activities they precisely perform. It is therefore possible that other researchers would create different PDDs. However, this impact is deemed to be small, because the general activities over all respondents are unlikely to change. Whether data is collected through a survey, workshop or interview does not change the goal of data collection. The fact that they sometimes perform a media analysis and other times do not also does not change this. A way to combat this would be to validate the PDDs in a session with the respondents. However, this would put a large strain on the respondents, leading to stakeholder fatigue.

Similarly, the activities from the PPDs are classified into general activities by a single researcher. Although the results are discussed with other researchers and agreement was reached, others could come up with different labels and granularities. However, this impact is considered to be limited as it does not change the semantics of the general activities.

A note is required on the sample of companies. During the problem investigation interviews, we were unaware of the different perspectives and parties in the industry. This results in an unbalanced ratio between consultants, major firms, accountants and SMEs. Especially the internal perspective is less discussed, as only two firms are identified as such. Furthermore, no SMEs were part of the sample. Although in an ideal scenario, more of these types of firms are interviewed, this is not deemed to be very problematic. This paper mainly focuses on identifying the range of the activities instead of the average. Furthermore, R7 mentioned that

[25] A popular method for research, as it collected 423 citations within five years after publishing.

their firm is considered to be a frontrunner on materiality assessment, meaning that most activities were found. Additionally, although only three auditors are interviewed, auditors are a smaller type of practitioners in the industry.

Furthermore, we believe that the chance of all respondents not disclosing a step is slim. Therefore, we believe our general activities are applicable to the complete industry. However, as the respondents are mainly based in Western Europe, some reservation with regards other countries is required.

Additionally, interviews are conducted in a combination of Dutch and English. Some interviews are then translated into English before they are analyzed (R1, R2, R3, R4). Although this translation was done to the best of our ability, it might change terminology used by respondents. However, we do not believe a change of semantics occurred along the way and therefore others would arrive to the same conclusion.

Finally, data triangulation is not always possible. Although Guix et al. (2018) identified similar activities based on their study of reports, where the (social) media is a source which was not discussed. We urge other researchers to analyze practices of the industry.

7. Conclusions

This paper contains several theoretical contributions. One, it discovered several perspectives on materiality assessment and identified four types of practitioners. These perspectives influence the method that is used for materiality assessment, even though the precise degree to which it is, is unknown.

Two, it has proposed a conceptual model to discuss materiality assessment, regardless of the perspective. With this model, activities, goals, tools and sources are identified which are applied in materiality assessment. We believe that these models pave the way for a structured academic discussion and systematic research into the topic.

Three, we identified fourteen activities which are classified into five goals. We found that many sources of information are not utilized, because it is not cost-effective. Currently, there are no accessible tools for the elicitation of these sources. This is combined with other pain points in a research agenda, which discusses four large and three smaller issues that require more research to help the industry. We identified a need for more research on a maturity model for materiality assessment, the need for a flexible, accessible and transparent tool, the operationalization of materiality assessment methods and a means to integrate risk and sustainability departments for materiality assessments. Furthermore, the lack of standards in methods, the impact of the perspectives on the methods and an alternative to the survey instrument are topics of interest in future research.

For practitioners in the industry, this paper provides a guide of activities. As the one of the few researches to perform extensive field research, we recommend an order and process for the industry to implement, having identified the atomic steps performed by advisors, firms and auditors. This provides practitioners with a reference model to mirror their own methods. This fourteen-step framework is implementable and is incorporable with the proposed research agenda.

Chapter 3. A tool for materiality assessment

1. Introduction

In sustainability reporting, organizations report on their strategy to deal with sustainability and societal issues. This is no simple task, as the nature of organizations influences their most important topics. For example, while safety on oil rigs is highly important to oil companies, it is not important to a bakery. To identify and prioritize these issues, a materiality assessment is required. This is defined as a set of activities which determine which topics are material to an organization; which aspects should be disclosed, reported and managed (Keijdener, 2019). However, the implementation of such an assessment is often opaque, hardly supported by quantitative methods, which makes the assessment subject to incompleteness and subjectivity (Calabrese et al., 2016).

Per Directive EU law (2014/95/EU) businesses are required to disclose topics such as environmental protection, social responsibility or respect for human rights. Information on these topics originates from multiple sources, such as internal and external stakeholders, peers, news and social media. Although these sources contain useful information, they are often not utilized in practice (Keijdener, 2019). One of the reasons for this might be that many of these sources are elicited manually, leading to a heavy strain on employees and falling victim redundant work cycles. This results in the task being cost-ineffective.

Meanwhile, an example of how this can lead to incompleteness, is illustrated by the rise of the MeToo movement at the end of October 2017. Although gaining fame in the movie industry, the movement quickly spread to other industries and sparked a large debate on social media and in news outlets. However, a quick analysis of sustainability reports found that among some major organizations[26], none reported on the topic of sexual predatory behavior in 2018. Similarly, while LGBT-rights are often mentioned on social media and conventional media (Chen, 2018), Dell is the only organization reporting on the topic in its report.

This situation illustrates two problems. First, recent topics might not be included in the report. Even though firms require time to respond to these situations with policy measures (Zhou & Chen, 2011), including discussions of these sources could raise awareness within the firm (Kietzmann, Hermkens, McCarthy, & Silvestre, 2011). This allows the policy-loop to be shortened. Second, topics originating from outside the sector might not be included in the materiality assessment. Some topics are more relevant to a specific sector than others. This means that some firms may fall subject to bias, where a topic is not ranked, because it is not identified. However, it might be important to the firm or its stakeholders. Therefore, a tool to deal elicit these sources is required, a tool which is flexible, transparent and accessible (Keijdener, 2019).

To the best of our knowledge, such a tool does not yet exist. Therefore, this paper proposes a tool which can elicit these sources of information. For this description, the paper is structured as follows. In section 2, some related tools are discussed which support only a part of this process. Then, in section 3, the functionalities of the proposed tool are described with a basic walkthrough. Section 4 discusses the results from validating a prototype with industry

[26] PWC Hungary, Royal Dutch Shell, Unilever, Dell, Leiden University and Friesland Campina

practitioners. Next, in section 5, some limitations are given. Finally, in section 6, the research is concluded and future work on the tool is mapped out.

2. Related work

Scientific work on materiality assessment tools is limited. Multiple 'tools' are in fact methods, instead of software tools. Wu et al. (2018) discuss four tools in materiality assessment: SASB materiality map, GRI's sustainability topics [list], sustainability disclosure database and social life cycle perspective (which is more a method). This shows that the notion of tools and methods are coupled. They found that there is a low consistency in obtaining most material topics from two approaches. They reckon that the GRI database might be a good start for the identification on material issues during the preliminary assessment. Keijdener (2019) found that in practice many organizations use software tools to collect data and process results. However, few of these used tools are tailored towards materiality assessments. Most often, tools were used to support a part of the process, for example through deploying a tool to conduct surveys. In some cases, this was supplemented by (social) media analysis tools, where the status of the company is monitored. Additionally, a tool such as the Materiality Assessment Tools (MAT) exists but does not support other dimensions than internal and external stakeholders. Furthermore, Datamaran is a tool which supports many of these dimensions. However, it was found to be expensive and provided little customizability (Keijdener, 2019).

3. A tool for materiality assessment

This tool is a hybrid between R (3.5+) and Python (3.7+). It requires some R packages and Python modules to install, but can be run through opening the Project in R. It is currently only available on Windows X64 builds, with an experimental version on X32. The tool is freely available on our GitHub[27]. It contains the following features:

- *Reading of pdf files:* To analyze peer reports, the tool contains a pdf converter which converts pdf files into plain text. This is mainly used for converting peer reports, although all documentation from pdf can be converted and used in the analysis.
- *Collection of (social) media:* In the tool, the user can enter a number of hashtags, twitter usernames, or search terms. These are converted to different formats to search through multiple APIs on platforms such as Twitter, Reddit and news platforms, i.e. New York Times, Google News and Yahoo News. This function returns a set of text files with its findings on Twitter, Reddit and news platforms.
- *Tokenization, stemming and keyword tagging:* The tool supports several NLP techniques to analyze the data. The tool cleans the text files, removing all punctuation and symbols. It then removes stop words, or words which bear no meaning, from the data. The data is then tokenized and stemming is applied to each token. The tool then uses keyword tagging, taking the keywords from a pre-entered list, counting how often a specific term is mentioned in the data. Multiple scoring criteria are used, as explained further in Appendix G.
- *Construction of word clouds, document term matrices and a materiality matrix:* To provide insight into this data, the tool supports three kinds of visualizations. First, a word cloud is created based on the whole data set. Then a document term matrix shows which keywords are tagged and how often. Finally, a materiality matrix is created based on these keywords and their scores.
- *Topic detection through Latent Dirichlet Allocation (LDA):* A topic modeler is included to identify latent topics in news text. This model returns fifteen words which are often co-occurring in the news articles. For more background on the workings of this model, Appendix E, discusses the basic technique of document term models which are required is required for the LDA. For those who are unfamiliar with the LDA Appendix F: Probabilistic models and Latent Dirichlet Allocation, explains how the algorithm works in further detail.

[27] https://github.com/sergioespana/openmasses

Performing a basic analysis

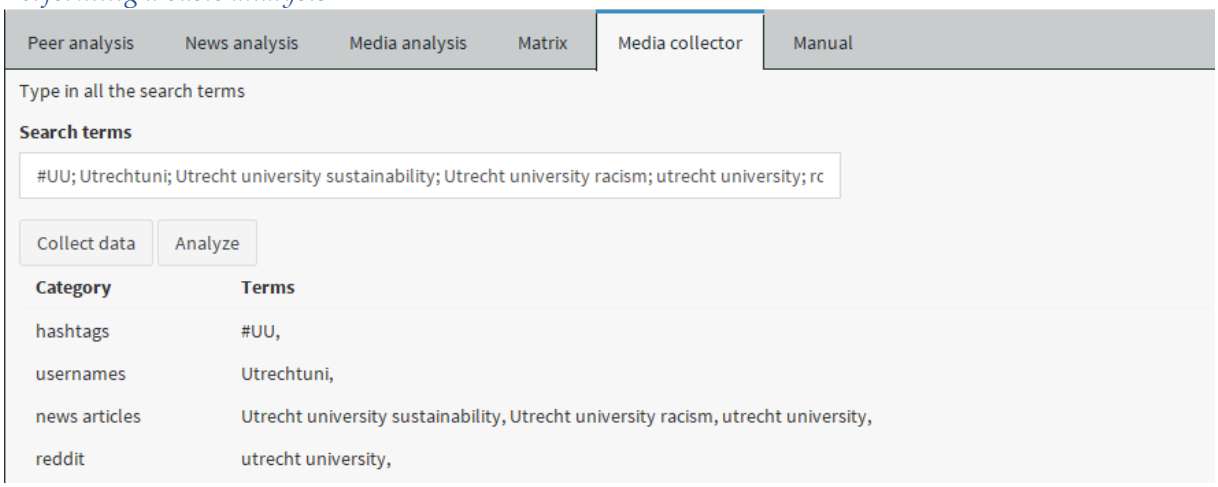


Figure 9. The media collection screen

1) As demonstrated in Figure 9. We start by performing an analysis on Utrecht University, as an input, several terms of interest are entered. The search terms are automatically placed in the right platform, but manual override with the keys such as rdt:, nws:, htg: and usn: is possible for respectively reddit, news, twitter hashtags and twitter usernames. For example, it is interesting to see what the reddit community discusses on the applications, by clicking collect data, the pipeline starts collecting data on these terms.
2) When the collection of the media is done, the output files are uploaded. In addition, the sustainability reports of Leiden University and Tilburg University are added, along with a longlist of topics we wish to investigate.
3) A document term matrix is created based on keyword tagging, to understand how often a specific term is mentioned in each peer report. The same is done for the media files. A dummy one is shown in Figure 10.
4) A matrix is created based on the score of these terms. In general, a higher score means that the issue is more important. These are all plotted on a matrix. An option is given to change this scoring scheme and to adjust the weights for each stakeholder. For example, we find that Twitter is not very important, so we adjust the weight to 0.5. Peers are more important, so we adjust the weight to 2. In Figure 11. the movement of topics is illustrated by the colored dots.

Topics	Score	news.txt	reddit.txt	twitter.txt
fair competition	25	25	0	0
gender gap	0	0	0	0
innovation	719	640	3	76
circular economy	48	40	0	8
working conditions	173	169	0	4
training & education	11	11	0	0
political neutrality	32	29	3	0
contribution to communities	20	20	0	0
human rights protection	328	316	0	12

Figure 10. Example of document term matrix

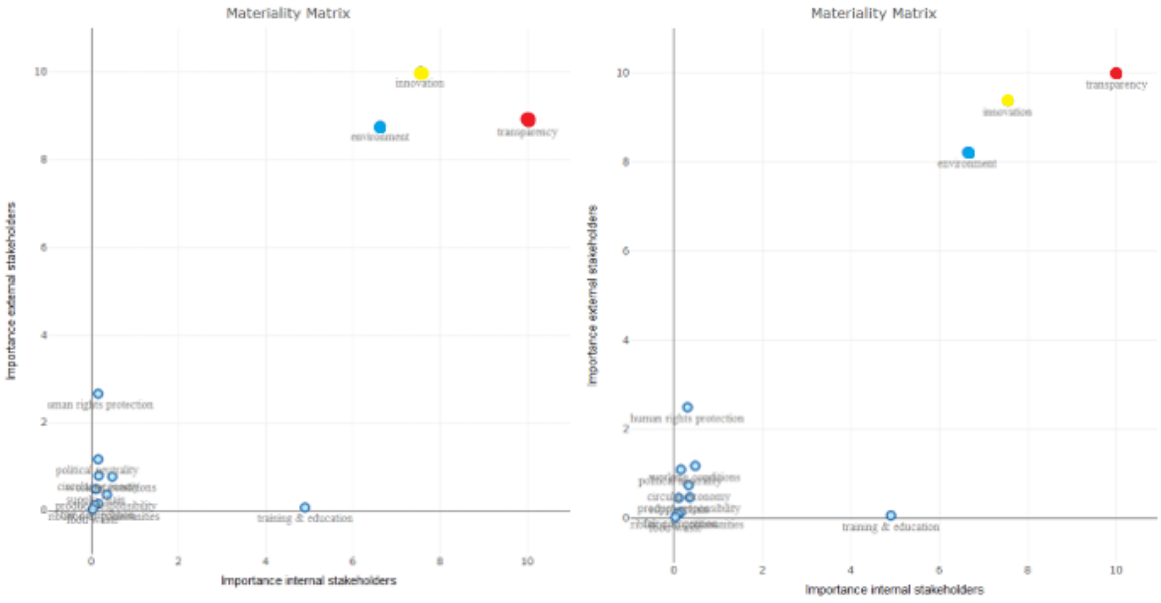


Figure 11. Materiality matrix example

A full screenshot guide is included in Appendix H. Its corresponding architecture is further explained in Appendix I.

4. Validation of the tool

In total, five expert assessments were executed at different times of development. The first validation has been executed after the tool was ready for peer analysis. With two advisors at a consultancy company, an experiment has been conducted. For this, both experts performed a materiality assessment on the same case, but one used the tool while the other one did not. It was found that the expert using the tool was able to execute the assessment roughly five times faster than his colleague. This gain was mainly contributed to analyzing the reports through the tool, which saved the consultant much time reading them in detail.

Next, a session with two auditors of an accountancy company was performed, which accepted the idea of the tool. The first comment was on the use of word clouds, as it was immediately recognized as something familiar. Their second note was on the use of the longlist. They mentioned that using a longlist is somewhat extensive and they would prefer to skip this step. Therefore, they requested a topic detector which could work independently on the data. An alternative might be to provide server-side longlists for them to use and adapt. Finally, they disagreed with each other on whether they would use the tool or not. From an assurance perspective, one found it useful, but the other had his doubts on whether this would not take too much time. From the advisory perspective, both were positive towards using it in assignments. From this conversation, it became clear that a big benefit would be saving time.

Finally, two session were performed with consultants with a semi-structured interview. In this case, a materiality assessment is mimicked of two sectors they have chosen to demonstrate some results. The validation protocol in enclosed in appendix B. Both experts showed a positive reaction to the tool. They saw potential in using the tool and acknowledged it would help them performing materiality assessments. They showed understanding of the tool's concepts and did not feel the tool was too difficult to use. Especially the creation of the matrix was deemed useful, allowing them to play around with the data. When asked about potential improvements, three points emerged. One emphasized that transparency is key and that the tool should provide information on where the data is collected from, which search terms were used and which descriptions are used in the longlist. Two requested the addition of additional languages and, three, survey results integration.

5. Discussions

This tool is currently subject to iterative development and some improvements can still be made. The first issue is with the longlist, as the reliability of the tool's output relies on the accuracy of the descriptions. As there is no publicly available longlist of topics yet, users are required to formulate their own topics and descriptions; all subject to different interpretations and quality. While the authors have one for the education section, this cannot be transferred to other sectors without adjustments. Secondly, the availability of news is limited to several months at best due to implementation constraints in the prototype. For future versions, it is important to invest in more extensive news collectors, such as a connection to Lexis Nexis. Thirdly, the current tool is not easily accessible, as it requires the manual installation of software and requires some understanding of R and Python to use all the functionalities. Therefore, a deployable version is currently being developed in Python. Finally, although the media collection works well, the speed of the collection is slow. To create a substantial body of evidence, roughly ten minutes is used per term. Even though the impact of this is deemed fairly minimal, as this is a background operation, a more efficient collection algorithm could boost overall performance.

6. Conclusions & future work

In this paper, we demonstrated the need for and the use of a materiality assessment tool. A framework for such a tool is designed and the key ideas are elaborated upon. This is the first demonstration of a larger scale research project, where more options are explored. More work is being done to collect a set of sustainability terms which can serve as an input for the longlist. Additionally, more experiments are needed to use a topic detection algorithm to collect these from the news. Furthermore, some work is required on determining a best practice for parameters of the media collection. A question remains on how many terms a search is required to get a comprehensive overview of the organizations. Similarly, as it is unfeasible to collect all the (social) media articles on a specific topic, some number should be determined to serve as a guideline. The current implementation is a minimum viable product, designed for research and testing. However, as the initial response towards the tool was positive, a next iteration is being worked on. The final point of interest would be to perform a semantic analysis on the descriptions. Right now, terms such as women and female are not related, while the input of either terms should cover the other one as well. Research should be conducted on how to link these types of terms

References

Becker, C., Chitchyan, R., Duboc, L., Easterbrook, S., Penzenstadler, B., Seyff, N., & Venters, C. C. (2015). *Sustainability Design and Software: The Karlskrona Manifesto*. *Proceedings - International Conference on Software Engineering* (Vol. 2). https://doi.org/10.1109/ICSE.2015.179

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, *3*(3/1/2003), 993--1022. https://doi.org/10.1162/jmlr.2003.3.4-5.993

Borgert, T., Donovan, J. D., Topple, C., & Masli, E. K. (2018). Initiating sustainability assessments: Insights from practice on a procedural perspective. *Environmental Impact Assessment Review*, *72*, 99–107. https://doi.org/10.1016/J.EIAR.2018.05.012

Calabrese, A., Costa, R., Ghiron, N. L., & Menichini, T. (2017). *MATERIALITY ANALYSIS IN SUSTAINABILITY REPORTING: A METHOD FOR MAKING IT WORK IN PRACTICE*. *European Journal of Sustainable Development* (Vol. 6). Center. Retrieved from http://www.ecsdev.org/ojs/index.php/ejsd/article/view/535

Calabrese, A., Costa, R., Levialdi, N., & Menichini, T. (2016). A fuzzy analytic hierarchy process method to support materiality assessment in sustainability reporting. *Journal of Cleaner Production*, *121*, 248–264. https://doi.org/10.1016/j.jclepro.2015.12.005

Calabrese, A., Costa, R., & Rosati, F. (2015). A feedback-based model for CSR assessment and materiality analysis. *Accounting Forum*, *39*(4), 312–327. https://doi.org/10.1016/j.accfor.2015.06.002

Chen, Y. A. (2018). *Media Coverage and Social Changes: Examining Valence of Portrayal of the LGBT Community from 2000 to 2014 in Two U.S. Magazines*. *Intercultural Communication Studies XXVII*. Retrieved from https://web.uri.edu/iaics/files/Y.-Anthony-Chen.pdf

Deloitte, & MVONederland. (2015). *Integrated Reporting as a driver for Integrated Thinking?* Retrieved from https://www2.deloitte.com/content/dam/Deloitte/nl/Documents/risk/deloitte-nl-risk-integrated-reporting-a-driver-for-integrated-thinking.pdf

Eccles, R., Ioannou, I., & Serafeim, G. (2016). *Is sustainability now the key to corporate success?* Retrieved from http://csr-raadgivning.dk/wp-content/uploads/2017/02/Harvard-Is-sustainability-now-the-key-to-corporate-success-January-2016.pdf

FASB. (2018). *Conceptual Framework for Financial Reporting*. Retrieved from https://www.fasb.org/jsp/FASB/Document_C/DocumentPage?cid=1176171111614&acceptedDisclaimer=true

Fechner, W. F. L. (2019). *Examining the state of the art of Materiality Assessment methods and tools ; what is the state of the practice*. Utrecht University.

Font, X., Guix, M., & Jesús Bonilla-Priego, M. (2016). Corporate social responsibility in cruising: Using materiality analysis to create shared value. *Tourism Management*, *53*, 175–186. https://doi.org/10.1016/j.tourman.2015.10.007

Forstater, M., Zadek, S., Evans, D., Knight, A., Sillanpää, M., Tuppen, C., & Warris, A.-M.

(2006). *The Materiality Report Aligning Strategy, Performance and Reporting*. Retrieved from http://www.accountability.org/about-us/publications/materiality.html

Garousi, V., Felderer, M., & Mäntylä, M. V. (2019). Guidelines for including grey literature and conducting multivocal literature reviews in software engineering. *Information and Software Technology*, *106*, 101–121. https://doi.org/10.1016/J.INFSOF.2018.09.006

GRI, & RobecoSAM. (2016). Defining what Matters: Do companies and investors agree on what is material? Retrieved from https://www.globalreporting.org/resourcelibrary/GRI-DefiningMateriality2016.pdf

Guix, M., Bonilla-Priego, M. J., & Font, X. (2018). The process of sustainability reporting in international hotel groups: an analysis of stakeholder inclusiveness, materiality and responsiveness. *Journal of Sustainable Tourism*, *26*(7), 1063–1084. https://doi.org/10.1080/09669582.2017.1410164

Hamborg, F., Meuschke, N., Breitinger, C., & Gipp, B. (2017). news-please: A Generic News Crawler and Extractor. In *Proceedings of the 15th International Symposium of Information Science* (pp. 218–223). Retrieved from https://www.gipp.com/wp-content/papercite-data/pdf/hamborg2017.pdf

Harris, Z. S. (1954). Distributional Structure. *Distributional Structure, WORD*, *10*(3), 146–162. https://doi.org/10.1080/00437956.1954.11659520

Hiemstra, D. (2009). Information Retrieval Models. In *Information Retrieval* (pp. 1–19). Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470033647.ch1

Hoffman, M., & Lydenberg, S. (2013). *Materiality background paper for*. Retrieved from https://integratedreporting.org/wp-content/uploads/2013/03/IR-Background-Paper-Materiality.pdf

Horn, R., & Turk, B. (2018). *Text analytics on crowd-sourced taxonomies*. Utrecht University.

Implementation Guide for companies. (2009). *University of Glasgow*. https://doi.org/10.1002/pro.456

Jones, K. S. (1973). Index term weighting. *Information Storage and Retrieval*, *9*(11), 619–633. https://doi.org/10.1016/0020-0271(73)90043-0

Jongerius, C. (2017). *Master Business Informatics Materiality Assessment Tool*. Utrecht University.

Keijdener, M. N. C. (2019). *Methods in materiality assessment: an overview and research agenda (under review)*.

Keijdener, M. N. C., Overbeek, S. J., & España, S. (2018). Scalability factors in an ICT4D context : A literature review. *Proceedings of the 5th International Symposium "Perspectives on ICT4D" (P-ICT4D 2018) Co-Located with 10th ACM Web Science Conference (WebSci'18), Amsterdam, the Netherlands, May 27, 2018*, *2120*. Retrieved from https://dspace.library.uu.nl/handle/1874/369707

Khalil, E. L. (2013). Lock-in institutions and efficiency. *Journal of Economic Behavior & Organization*, *88*, 27–36. https://doi.org/10.1016/J.JEBO.2011.10.017

Khoshgoftar, M., & Osman, O. (2009). Comparison of maturity models. In *2009 2nd IEEE International Conference on Computer Science and Information Technology* (pp. 297–301). IEEE.

https://doi.org/10.1109/ICCSIT.2009.5234402

Kietzmann, J. H., Hermkens, K., McCarthy, I. P., & Silvestre, B. S. (2011). Social media? Get serious! Understanding the functional building blocks of social media. *Business Horizons*, *54*(3), 241–251. https://doi.org/10.1016/J.BUSHOR.2011.01.005

King, B. (2013). Materiality assessments: The missing link for sustainability strategy | GreenBiz. Retrieved May 30, 2019, from https://www.greenbiz.com/blog/2013/09/10/materiality-assessments-missing-link-sustainability-strategy

Lewis, P. V. (1985). Defining "Business Ethics": Like Nailing Jello to a Wall. *Journal of Business Ethics*, *4*, 377–383. Retrieved from https://link.springer.com/content/pdf/10.1007/BF02388590.pdf

Luhn, H. P. (1957). A Statistical Approach to Mechanized Encoding and Searching of Literary Information. *IBM Journal of Research and Development*, *1*(4), 309–317. https://doi.org/10.1147/rd.14.0309

Mazars. (2015). INTEGRATED REPORTING: HOW FAR HAVE WE COME? A look at insurers annual reports. https://doi.org/10.13140/2.1.1814.6566

Mukhopadhyay, R. (2017). *Navigating the Sustainability Reporting landscape – Method Specification and tool for Materiality Assessment*. Utrecht University. Retrieved from https://dspace.library.uu.nl/handle/1874/357904

Muñoz, M. J., Fernandez, M. A., Rivera, J. M., Soriano, L., Escrig, E., & Ferrero, I. (2012). Materiality Analysis for CSR Reporting in Spanish SMEs. *International Journal of Management, Knowledge and Learning*, *2*(2), 231–250. Retrieved from www.issbs.si/press/ISSN/2232-5697/1_231-250.pdf

Murninghan, M. (2013). *Redefining Materiality II: Why it Matters, Who's Involved, and What It Means for Corporate Leaders and Boards*. Retrieved from https://www.accountability.org/wp-content/uploads/2017/02/Redefining-Materiality-2.pdf

Nummert, B. (2014). Sustainability Reporting and Organizational Change Management in Companies - A survey-based Analysis of the Relationship between Sustainability Reporting and Organizational Change. Retrieved from https://dspace.library.uu.nl/handle/1874/297364

Phelps, D. L. (2016). *Analysis of Materiality Assessment Methods*. Duke University. Retrieved from https://dukespace.lib.duke.edu/dspace/bitstream/handle/10161/11818/Phelps MP - Analysis of Materiality Assessment Methods 2016-04-15.pdf;sequence=1

Puroila, J., & Mäkelä, H. (2018). Materiality in sustainability reporting: An Illusion of Consensus and Objectivity? *Academy of Management Proceedings*, *2018*(1), 16174. https://doi.org/10.5465/AMBPP.2018.16174abstract

Puroila, J., & Mäkelä, H. (2019). Matter of opinion. *Accounting, Auditing & Accountability Journal*, AAAJ-11-2016-2788. https://doi.org/10.1108/AAAJ-11-2016-2788

Quik, C. (2018). *Crowd-sourcing a long-list of sustainability topics for the materiality assessment of*

universities. Utrecht Unives.

Ramos, J. E. (2003). Using TF-IDF to Determine Word Relevance in Document Queries. Retrieved from https://dl.acm.org/citation.cfm?id=1096906

Saenz, C. (2019). Creating shared value using materiality analysis: Strategies from the mining industry. *Corporate Social Responsibility and Environmental Management*, (February), 1–10. https://doi.org/10.1002/csr.1751

Sarraj, D. (2018). Examining Materiality in Sustainability Reporting: Evidence from GCC Countries. Retrieved from https://openrepository.aut.ac.nz/handle/10292/11586

Singhal, A. (2001). Modern Information Retrieval: A Brief Overview. *IEEE Computer Society Technical Committee on Data Engineering*. Retrieved from http://trec.nist.gov

Spärck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, *28*(5), 11–21. https://doi.org/10.1108/00220410410560573

Steyvers, M., & Griffiths, T. (2007). Probabilistic Topic Models. In W. K. T. Landauer, D McNamara, S. Dennis (Ed.), *Latent Semantic Analysis: A road to meaning* (pp. 427–448). Laurence Erlbaum Associates Publishers. Retrieved from http://173.236.226.255/tom/papers/SteyversGriffiths.pdf

Syed, S. A. S. (2019). *Topic Discovery from Textual Data: Machine Learning and Natural Language Processing for Knowledge Discovery in the Fisheries Domain*. https://doi.org/10.4018/978-1-59904-373-9.ch006

Syed, S., & Spruit, M. (2017). Full-Text or Abstract? Examining Topic Coherence Scores Using Latent Dirichlet Allocation. In *2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA)* (pp. 165–174). IEEE. https://doi.org/10.1109/DSAA.2017.61

Szekely, N., & Vom Brocke, J. (2017). What can we learn from corporate sustainability reporting? Deriving propositions for research and practice from over 9,500 corporate sustainability reports published between 1999 and 2015 using topic modelling technique. *PLoS ONE*, *12*(4), 1–27. https://doi.org/10.1371/journal.pone.0174807

Taneva, R., & Stracchi, C. (2018). *Webinar 6 of the six-part GRI Standards In Practice Series The Materiality Principle: A Deep Dive*. Retrieved from https://www.globalreporting.org/SiteCollectionDocuments/2018/CSE/Materiality Webinar 29 Nov 2018.pdf

The International Integrated Reporting Council (IIRC). (2015). *the International<Ir> Framework*. https://doi.org/10.2139/ssrn.2378899

van de Weerd, I., & Brinkkemper, S. (2009). Meta-Modeling for Situational Analysis and Design Methods. In *Handbook of Research on Modern Systems Analysis and Design Technologies and Applications* (pp. 35–54). IGI Global. https://doi.org/10.4018/978-1-59904-887-1.ch003

Whitehead, J. (2016). Prioritizing Sustainability Indicators: Using Materiality Analysis to Guide Sustainability Assessment and Strategy. https://doi.org/10.1002/bse.1928

Whittemore, R., Chase, S. K., & Mandle, C. L. (2001). *Pearls, Pith, and Provocation Validity in Qualitative Research*. Retrieved from https://journals.sagepub.com/doi/pdf/10.1177/104973201129119299

Wieringa, R. (2014). *Design Science Methodology for information systems and software engineering*. Springer Heidelberg New York Dordrecht London. Retrieved from https://link.springer.com/content/pdf/10.1007/978-3-662-43839-8.pdf

Yumasheva, E. (2018). The Seven Tips to the Perfect Materiality Analysis | Datamaran. Retrieved June 21, 2019, from https://www.datamaran.com/blog/seven-tips-perfect-materiality-analysis/

Zhou, Z., & Chen, Z. (2011). Formation mechanism of knowledge rigidity in firms. *Journal of Knowledge Management*, *15*(5), 820–835. https://doi.org/10.1108/13673271111174348

Appendix A: Interview investigation protocol

Thank you for making the time for this interview. The goal of this interview is to investigate materiality assessment practices in the areas of sustainability and business ethics. To later process the data, and if you agree with this, I will record this session. Do you have any questions before we start?

For starters, I would like to ask you some questions about your role in the organisation.

- What industry sector(s) does your company operate in? (e.g. consultancy)
- What is the name of your job function (or role)?
- How many years of experience do you have in such job function?
- This research is about materiality assessment. Before continuing, I would like to know your definition of materiality assessment. We do not expect from you an academic definition, but rather an informal one, based on your understanding of such practice. (After s/he talks freely, provide our definition and determine whether s/he agrees with it, wants to clarify or add anything)
- Your experience with materiality assessment:
    - To what extent are you involved with materiality assessment practices?
    - Is it part of your duties of your job function?
    - How many times per year do you perform materiality assessment?
- The clients of materiality assessment practices:
    - Do you perform materiality assessment for external clients or only for your own company? (We refer to the organisation for which the material topics are analysed)
    - To what industry sectors do these clients typically belong? (More than one is possible; when many, just the 5 most frequent ones are enough)
    - How often do these clients require a materiality assessment? (Frequency; e.g. once every 2 years)
- Methods for materiality assessment:
    - What method(s) do you apply when performing materiality assessment? That is, are there any guides that you use to do the materiality assessment?
    - Are these methods your own or defined externally? It would be nice to have the name of the handbook, a URL where the method specification can be found, etc. If the method is internal, we would like to receive the handbook or, at least, a comprehensive written or verbal description, in order to later create a process deliverable diagram (PDD).
    - Do you apply the method as described in the handbook or do you adapt it to your situation and needs? If we already have a PDD, go through it and ascertain that they follow it as it is specified; when variations are found, then record them with notes.
    - What are the criteria that you use to decide which sustainability and business ethics topics are material for your client organisation (e.g. to score, prioritise or rank them)? First, let the interviewee speak freely. Then see whether s/he uses any of these: importance for the internal stakeholders, importance for the external stakeholders, recency or topicality, impact, risk, etc. It might be the case that s/he combines several of these criteria during the scoring or prioritisation of topics; if so, we would like to know how s/he proceeds.

    - How do you report on the results of the materiality assessment? E.g. a section in a sustainability report containing text, tables, a materiality matrix, etc. In case s/he indeed uses matrices, we would like to know what are the dimensions of the matrix (i.e. the labels in the vertical and horizontal axes).
- Tools for materiality assessment:
    - What tools do you use to support materiality assessment? A tool can be anything that you use during the materiality assessment project to quantify the data (including Excel or Word, surveys, interviews, when used to prioritise as list of topics).
    - For each of the tools…
        - Has this tool been developed in-house or it is an external tool?
        - If it is an internal tool, what features does it have? If it is an Excel, we would appreciate a copy (unless you value is as a trade secret). We are developing tools that are open-source and are happy to share them with you.
        - If it is an external tool, we would like to have the name and some URL or reference where we can find more information about it.
        - Is there any important feature that you miss in this tool? What do you think would be a useful addition to such tool?
- On the use of long-lists:
    - As an instrument to facilitate identifying sustainability and business ethics topics that might be material for the client organisation, do you use any pre-compiled list of topics? We will refer to such list as long-list, since it typically contains a large number of topics that need to be filtered in some way in order to identify those that deserve discussion with the stakeholders. A long-list pre-exists and can be reused in many materiality assessment projects. There can be a single long-list, or you could have one per sector or type of client organisation. Do you have something like this?
    - Did you (or someone in your company) create it or is it external (e.g. GRI Standards)
        - if it is your own (internal):
            - We would like to receive the list (unless you value it as a trade secret).
        - If it is external:
            - We would like to have a URL or a reference where we can find the long-list you use;
            - Did you modify the external long-list in order to adapt it to your needs? How? E.g. are there topics that are not in the list that you add to it?
- On the analysis of external documentation:
    - Peer reports
        - As part of your materiality assessment practices, do you read reports/documents from peer companies in order to identify the topics they typically report about? These could be sustainability or non-financial reports, CSR websites, etc.
        - What is your motivation for reading peer reports?

        - When and how do you do this? Indicate the moment in time, in relation to the method s/he applies (e.g. make a note in the PDD).
        - Would you appreciate a tool feature that automatically scans a given set of documents (e.g. peer reports) to analyse how frequently each topic in list is mentioned?
    - Other external documentation:
        - Do you read other type of external documentation, such as standards, white papers, regulations?
        - What is your motivation for reading this external documentation?
        - When and how do you do this? Indicate the moment in time, in relation to the method s/he applies (e.g. make a note in the PDD).
        - Would you appreciate a tool feature that automatically scans news sites to analyse how frequently each topic in list is mentioned in the news?
    - News:
        - Do you read news or use your knowledge about the current state of affairs of your country or the world during the materiality assessment project? (For the purpose of materiality assessment, of course, not for leisure)
        - What is your motivation for reading news?
        - What news sites do you read for the purpose of materiality assessment? (unless you consider this sensitive information and prefer not to share this, of course)
        - Would you appreciate a tool feature that automatically scans news sites to analyse how frequently each topic in list is mentioned in the news?
        - What news sites would you like the tool to scan?
        - Why, how and when would you use this feature? Indicate the moment in time, in relation to the method s/he applies (e.g. make a note in the PDD)
    - Social media:
        - Do you read social media during the materiality assessment project? (For the purpose of materiality assessment, of course, not for leisure)
        - What is your motivation for reading social media?
        - What social media platforms do you read?
        - Would you appreciate a tool feature that automatically scans a social media platform to analyse how frequently each topic in list is mentioned in the news?
        - What social media platforms would you like the tool to scan?
        - Why, how and when would you use this feature? Indicate the moment in time, in relation to the method s/he applies (e.g. make a note in the PDD)
    - In case s/he does not read any external documents: why do you not read such external documentation? Maybe it is not deemed useful, of maybe it is not cost-effective but an automated feature would help.
    - Is there any other source of textual information that you deem relevant in the context of materiality assessment?
- Is there any other important issue related to materiality assessment or our research that you would like to discuss?

- Can we contact you later to ask any additional questions that might arise or validate some of the findings?
- Would you like to receive the results of our research?
- We plan to write a scientific paper. We will treat the information you have given us anonymously, so it cannot be traced to you, your company or your clients.
    - However, can we mention the name of your company when indicating who we have approached? Later, the results will still be presented anonymously and untraceable to you or the company. Otherwise, we will give your company a fictional name.

Thank you for your time.

Appendix B: Interview validation protocol

Questions for validation

- Given the conceptual model of materiality assessment, do you feel this describes the situation?
- How often do you use the term materiality assessment?
    - Why?
    - Is this problematic?
- Do you start with an internal or external perspective in materiality assessment?
- What are problems you run into on this topic? Think of:
    - (Lack of) standards
    - Unfamiliarity with the terminology
    - Data collection
    - Data processing
- What is the difference between audit and advisory in the materiality assessment process?
    - Given the conceptual model of the industry, do you feel this describes the situation?
- Is a matter material if you do not exert any influence in the matter?
    - Think of Coca-Cola and obesity or diabetes
- When is a matter material?
    - When it is recurrent?
    - When it is systematic? i.e. there is a link with the strategy
- Is there is big difference in maturity between your clients?
    - What is your opinion on a maturity model? Would this be beneficial?

Questions after demo

- Would you see yourself using this tool?
- Are you able to use this tool?
- What are features you are missing?

Appendix C: Coding tree

- Activities: when a specific activity is described. E.g. conduction of surveys, interviews, workshops.
    - Goals: when a reason for the activity is given. E.g. interviews to discuss materiality matrix. Can overlap with activity as it is a sub node.
- Additional findings: when issues which were raised by respondents when asked if they had questions/ wished to discuss additional matters.
- Background: when background information is given which does not impact the materiality assessment process but provides context of the respondent's persona and firm. E.g. organization's description, experience.
- Materiality general: when issues on materiality are discussed on a high-level. E.g. definitions.
- Matrix: when discussing all matters related to the matrix. E.g. axes, weighing of sources, aggregation of these sources in the matrix, use of visualizations.
- Process: when discussing the whole process in general. E.g. whether they use GRI, IIRC, or any other method without mentioning the specific activities.
- Sources: when the usage of certain sources are mentioned. E.g. the use of news, social media.
- Tools: when the names of tools which are used in the process is described. E.g. the use of Excel, Datamaran.

Appendix D: Activities in materiality assessment

An important note on the table in general is that most processes are not this linear or structured. In a single interview, multiple steps can be taken at once. Especially contact moments with a client are spread throughout the process, which means that some things might shift during the process. In addition, not every client requires the same process with the same extensiveness. An overview of this is given in Table 5. Activities are marked with a ~ when the respondent sometimes does the activity but did not mention it explicitly. Activities marked with a * are split or executed in parallel as a single activity can have multiple goals, therefore be part of multiple general activities. Activities between brackets () are performed by another company.

Table 5. Overview of activities

General Activity	R1	R2	R3	R4	R5	R6	R7	R8	R9
(A1) Assess sustainability strategy				1				1	
(A2) Analyze documentation		1					~		1+2
(A3) Identify longlist of topics	1	2+3	1	2+3+4	2	8*	1	2*+8	4
(A4) Premeasurement topics determination	2	4	2			9	2+3+4	3	7
(A5) Determine stakeholders	3		3+4	5+6		4+5	6	7	
(A6) Decide on materiality strategy			~		1	1+2+3 +6+7			
(A7) Conduct survey	4		5	11	3+4+5		7*	6	10
(A8) Perform media analysis		5					12	2*	5+6
(A9) Perform peer analysis		7+8		10	6	11	1*+11	2*	3
(A10) Conduct qualitative analysis	6		6+7+8	7+8+9 +12*+13*		8*	7*	5	9
(A11) Prioritize topics	5	6		12*+13*	7	10+16	5+8	4+9	8
(A12) Assess impact of topics						13+14 +15	9+10		
(A13) Create matrix	8+9	11	9	14	8	17	13	10	11
(A14) Validate results[28]	7	9+10	10	15		12	(14)		12

[28] Can be either results as in the matrix or a prioritized list.

Appendix D: Activities in materiality assessment

Below, each activity is mapped to a general activity. The format is behind each activity is [Rx.y], where x is the number of the respondent and y is the number of the activity in the PDD. A collection of all PDDs is enclosed in Appendix J: Collection of respondents' PDDs

- Assess company strategy (A1)
  Investigate which topics a company should care about based on their strategy. Clarify what the goal and scope of the sustainability report is.
  - Analyze current strategy [R4.1]
  - Identify company's strategy [R8.1]

- Analyze documentation (A2)
  Investigate the existing documentation in a company.
  - Consult with client [R2.1]
  - Goal & background investigation [R9.1] + Document analysis [R9.2]

- Identify longlist of topics (A3)
  The process of creating a longlist of initial topics.
  - Consultation with client [R1.1]
  - Perform extended analysis of new client [R2.2] + Identify possibly relevant topics [R2.3]
  - List all possible topics [R3.1]
  - Identify possible themes [R4.2] + Identify pain points [R4.3] + Identify opportunities [R4.4]
  - Brainstorm session on possible topics [R5.2]
  - Brainstorm session on picked topics [R6.8]
  - Create list of all available topics [R7.1]
  - Determine topics (external perspective) [R8.2] + Internally determine material topics [R8.8]
  - Identify possible topics [R9.4]

- Premeasurement topics determination (A4)
  The process of determining which topics are going to be measured.
  - Pick topics [R1.2]
  - Test for cherry-picking [R2.4]
  - Rate all topics in the list [R3.2]
  - Validate topics with longlist [R6.9]
  - Meet with team to discuss potential topics [R7.2] + Pick 20/25 usable topics [R7.3] + Risk identification [R7.4]
  - Discuss topics with client [R8.3]
  - Update topic list [R9.7]

- Determine stakeholders (A5)
  Determining which stakeholders are relevant.
  - Pick stakeholders with client [R1.3]
  - Determine stakeholders [R3.3] + Provide stakeholders to company [R3.4]
  - Identify stakeholders [R4.5] + Prioritize stakeholders [R4.6]
  - Brainstorm about potential stakeholders [R6.4] + Prioritize stakeholders [R6.5]

  - Identify internal & external stakeholders [R7.6]
  - Determine internal stakeholders [R8.7]

- Decide on strategy (A6)
  Process of assembling a team, decide on metrics.
  - Assemble internal team [R5.1]
  - Determine core of company [R6.1] + Apply theory of change on organization [R6.2] + Define indicators for monitoring [R6.3] + Present & pick framework [R6.6] + Pick variables [R6.7]

- Conduct survey (A7)
  The entire process of creating, sampling and conducting a survey.
  - Conduct survey [R1.4]
  - Conduct survey [R3.5]
  - Conduct survey [R4.11]
  - Make survey on potential topics [R5.3] + Check survey [R5.4] + Conduct survey internally & amongst all stakeholders [R5.5]
  - Interview/surveys among stakeholders on topics [R7.7]
  - Conduct survey [R8.6]
  - Conduct ranking by external stakeholders [R9.10]

- Perform media analysis (A8)
  All steps which are part of a media analysis.
  - Analyze media [R2.5]
  - Media analysis [R7.12]
  - Determine topics (external perspective) [R8.2]
  - Media analysis with tool [R9.5] + Google search for articles on meaningful topics [R9.6]

- Perform peer analysis (A9)
  All steps which are part of a peer analysis.
  - Identify peers [R2.7] + Peer analysis [R2.8]
  - Peer analysis [R4.10]
  - Perform peer analysis [R5.6]
  - Perform peer analysis [R6.11]
  - Create list of available topics [R7.11] + Leverage data on topics from other companies reports [R7.11]
  - Determine topics (external perspective) [R8.2]
  - Analyze competitors [R9.3]

- Conduct qualitative analysis (A10)
  Use of instruments such as workshops, focus groups, interviews to collect data.
  - Interview internal stakeholders [R1.6]
  - Website analysis [R3.6] + Conduct interviews [R3.7] + Weigh interviews [R3.8]
  - Identify wishes [R4.7] + Identify expectations [R4.8] + Identify requirements [R4.9] + Conduct interviews [R4.12] + Discuss themes with stakeholders in focus group [R4.13]

    - Brainstorm session on select topics [R6.8]
    - Interview/surveys among stakeholders on topics [R7.7]
    - Facilitate round-table discussion [R8.5] +
    - Conduct ranking by internal stakeholders [R9.9]

- Prioritize topics (A11)
  All steps that conduct some sort of rankings on topics.
    - Rank topics [R1.5]
    - Topic ranking [R2.6]
    - Conduct interviews [R4.12] + Discuss themes with stakeholders in focus group [R4.13]
    - Prioritize possible topics using survey results [R5.7]
    - Prioritize topics [R6.10] + Finalize prioritization of topics [R6.16]
    - Order topics [R7.5] + Adjust list of topics [R7.8]
    - Rank topics [R8.4] + Rank internally material topics [R8.9]
    - Rank topics [R9.8]

- Assess impact of topics (A12)
  Process of assessing extra metrics.
    - Analyze outcome of stakeholder ratings [R6.13] + Analyze magnitude [R6.14] + analyze social maturity related to topics [R6.15]
    - Measure velocity of topics [R7.9] + Measure speed of topics [R7.10]

- Create matrix based on results (A13)
  The process of creating a visual representation of the results.
    - Combine results in matrix [R1.8] + Define themes [R1.9]
    - Create final topic matrix [R2.11]
    - Create matrix [R3.9]
    - Set up matrix outline [R4.14]
    - Create topic matrix [R5.8]
    - Create matrix [R6.17]
    - Make topics matrix [R7.13]
    - Combine internally & externally important topics into matrix [R8.10]
    - Create matrix [R9.11]

- Validate results (A14)
  All steps to validate whether findings are correct.
    - Conduct focus group [R1.7]
    - Match with GRI indicators [R2.9] + Client discussion on analysis results [R2.10]
    - Validate matrix [R3.10]
    - Check matrix with senior management [R4.15]
    - Discuss first draft with client [R6.12]
    - Assurance [R7.14]
    - Validate with stakeholders [R9.12]

Appendix E: Background on (inverse) term frequency

The basis of many text mining, and information retrieval approaches are based on the bag of words model (Harris, 1954). This model, or its variations are used by more complex models today. To illustrate the use of the model, consider the following four strings of words or 'documents from d1-d4' below.

- The cat eats
- The fish swims with fish in the sea
- The cat does not like the sea
- Because the sea is water

In this model, each word is categorized separately into a 'bag of words', where the term frequency for every word is computed. The term frequency is defined as the number of times the term occurs in the document. This results in Table 6.

Table 6. Example of bag of words model

	the	cat	eats	fish	swims	with	in	sea	does	not	like	because	is	water
d1	1	1	1											
d2	2			2	1	1	1	1						
d3	2	1						1	1	1	1			
d4	1							1				1	1	1

In terms of topic detection, one way could be to use the term frequency to rate the importance of the topics (Luhn, 1957). However, this is generally not a good approach, as in this case the most important topic would be 'the', while the words of 'fish' and 'sea' might be more interesting. Therefore, this does not provide a satisfying result and a more complex method is required. This lead to the extension of this model with the inverse document frequency (or idf) (Spärck Jones, 1972). Originally defined as term specificity, this approach weights the terms based on how often they occur in the entire corpus. The idf is defined as:

$$idf(t, D) = \log\left(\frac{D}{d \in D : t \in d}\right)$$

Where D is the number of documents in the corpus and $d \in D : t \in d$ the number of times that the term t appears is any document d over all documents D. The easiest way of thinking about this is that a term is scored higher when it occurs in few documents (e.g. 'fish', 'swims', 'water') and lower when it occurs in many documents (e.g. 'the', 'cat', 'sea').

Both these methods were combined by Jones in 1973, to create the term frequency – inverse document frequency, where the tf is multiplied with the idf for each term:

$$tf - idf = tf * idf$$

This method takes the best of both worlds; the highest-ranking term frequency and the sparsity in occurrence of these terms. An example of the scoring in these methods are shown in Table 7.

Table 7. Overview of td-idf computations

Tf	Idf	Tf-idf
$\boldsymbol{tf}\ (\boldsymbol{'the'}, \boldsymbol{d1}) = \mathbf{1}$	$idf('the', D) = \log\left(\frac{4}{4}\right) = 0$	$tf - idf('the', d1, D) = 1 * 0 = 0$
$\boldsymbol{tf}(\boldsymbol{'the'}, \boldsymbol{d2}) = \mathbf{2}$		$tf - idf(the', d2, D) = 2 * 0 = 0$
$\boldsymbol{tf}(\boldsymbol{'fish'}, \boldsymbol{d1}) = \mathbf{0}$	$id('fish', D) = \log\left(\frac{4}{1}\right) = 0.6$	$tf - idf('fish', d1, D) = 0 * 0.0 = 0$
$\boldsymbol{tf}(\boldsymbol{'fish'}, \boldsymbol{d2}) = \mathbf{2}$		$tf - idf\ ('fish', d2, D) = 2 * 0.6 = 1.2$

Computing this for every term, results in the following matrix shown in Table 8.

Table 8. Example of inverse document term matrix

	the	cat	eats	fish	swims	with	in	sea	does	not	like	because	Is	water
d1	0	0.3	0	0	0	0	0	0	0	0	0	0	0	0
d2	0	0	0	1.2	0.6	0.6	0.6	0.12	0	0	0	0	0	0
d3	0	0.3	0	0	0	0	0	0.12	0.6	0.6	0.6	0	0	0
d4	0	0	0	0	0	0	0	0.12	0	0	0	0.6	0.6	0.6

There are multiple problems however with this approach, as illustrated by the table above. When using only this model, the best topic would be 'fish', a term which provides little insight into the topic of the documents. Furthermore, synonyms such as a bank would not be recognized by this technique (Ramos, 2003). Therefore, more sophisticated methods are required.

Appendix F: Probabilistic models and Latent Dirichlet Allocation

Early IR systems were often founded in Boolean logic, using complex combinations of ANDs, ORs and NOTs (Singhal, 2001). The downside to this approach is often in the lack of relevance ranking, a feature which most IR systems nowadays commonly possess. Nowadays, most algorithms rank the document by the assignment of a numeric score, where three types of models are most commonly used: vector space models, inference network models and probabilistic models.

In a vector space model, text is represented by a vector of terms (Singhal, 2001). This means that for every word the model creates a dimension. This highly dimensional model then uses a vector to connect each sentence through the position of the words in the vector space. The similarities between the queries can be computed by calculating the similarity between the vectors in the vector space. A downside to this model is that is assumes that all terms are independent of each other.

In inference network models, a network of the documents and their corresponding terms is created, along with the queries and information need. It functions as a Bayesian network, where it uses the rule of joint probability to reason its way downwards (Hiemstra, 2009).

Probabilistic models are based on the idea that each document is relevant to the query, yet some documents are more likely to be more relevant than other documents (Singhal, 2001). Therefore, the model should return the documents ranked by the highest probability of relevance to the user. This type of approach is dominant today and many topic detection methods build on this principle (S. A. S. Syed, 2019). Therefore, this type is chosen to use in the research.

In their paper Blei, Ng, & Jordan (2003) address the issues of the tf-idf as: *"the approach also provides a relatively small amount of reduction in description length and reveals little in the way of inter- or intradocument statistical structure."* (Blei et al., 2003, pp. 994). They state that the most common dimensionality reduction technique is the (probabilistic) latent semantic indexing (LSI), which in turn lead to the development of the Latent Dirichlet Allocation (LDA). This section discusses only the basic idea behind the LDA. For the interested reader, further reading can be done in the original paper by Blei et al. (2003).

Applying this algorithm to detect topics has been proven worthwhile in the research by Szekely & Vom Brocke (2017), who used the technique to analyze 9514 sustainability reports, ranging from over a twenty year period of 1995-2015, distilling 10 themes with multiple topics per theme. The algorithm defined 42 topics, which could be grouped into ten themes. This is done through manual interpretation.

The LDA is a probabilistic generative model. It assumes that documents are created by a prior determined set of probabilities concerning a specific topic (20% A, 30% B, 50% C) and reasons backwards. In a sense, it aims to capture the creation process of the document. It is based on the likelihood of occurring topics and assumes that topics are described by similar set of words, therefore it identifies groups of words regularly occurring in the documents (Steyvers & Griffiths, 2007).

In practice, the algorithm provides three outputs. One, clusters of words, where every word can belong to multiple clusters. This, in essence, shows of which words a topic exists. Next,

there is the frequency of words per topic. For example, how often each word appears within the topic 'sports'. Finally, there is a distribution of topics per document, where it shows that this document discusses the topic 'politics' and the topic 'sports'. Especially these first two are of importance as the cluster of words can be labeled as a topic and serve as an input for the materiality assessment process and the frequency can be used to measure a certain importance of the topic.

The LDA is best represented by Plate notation as shown in the following Figure 12 (Blei et al., 2003).

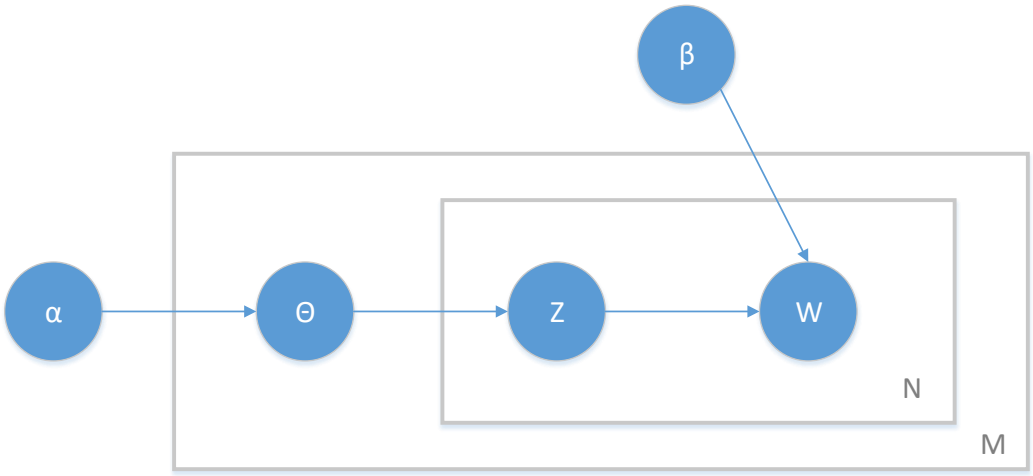


Figure 12. Plate notation of LDA (Blei et al., 2003)

In this figure, M is a collection of documents, of which each document consists of a collection of words called N. Θ is the topic distribution in these documents M. This is influenced by an external parameter called one of the Dirichlet priors, α. α indicates how many topics a document contains, where a high α means that the document contains many topics, while a low α indicate a few topics. Next, for each N, there is a topic z which is described by the words w. On w, the other Dirichlet prior β determines the degree of mixture of the words, where a high β provides a high mixture of words and a low β indicates a low mixture of words. In this way, both α and β can be used to input domain knowledge into the algorithm.

Limitations to this approach is that the number of topics should be known beforehand and topic distribution cannot capture the correlations between the topics. However, to overcome this, multiple models can be trained with a number of topics ranging from one to twenty, where a coherence measure is used to compare the models and select the best performing one.

Appendix G: Ranking methods in the tool

This section discusses the scoring algorithms used in the tool. As the goal of materiality assessment is to derive some matrix from the results, the end-result should be two dimensions. Therefore, some dimensionality reduction is required when using internal and external stakeholders, peer reviews, social media and news outlets. For this, three approaches are used.

First, is to simply average the scores of each source into a single score. This assumes an equal weight of sources and is implemented as the standard approach for dimensionality reduction.

A second approach to reduce its dimensions is to provide the user with a textbox for each axis. Each textbox can be filled with a dimension and can attribute a weight to the dimension. The user is free to compose each textbox to its own liking, setting the weights somewhere between zero and one. These boxes then add each dimension to form a final score, plotting it on its axis. This is principle illustrated in Figure 13.

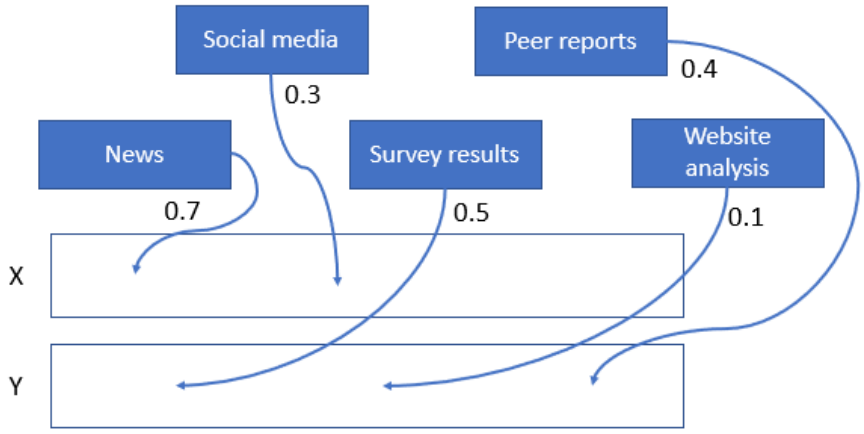


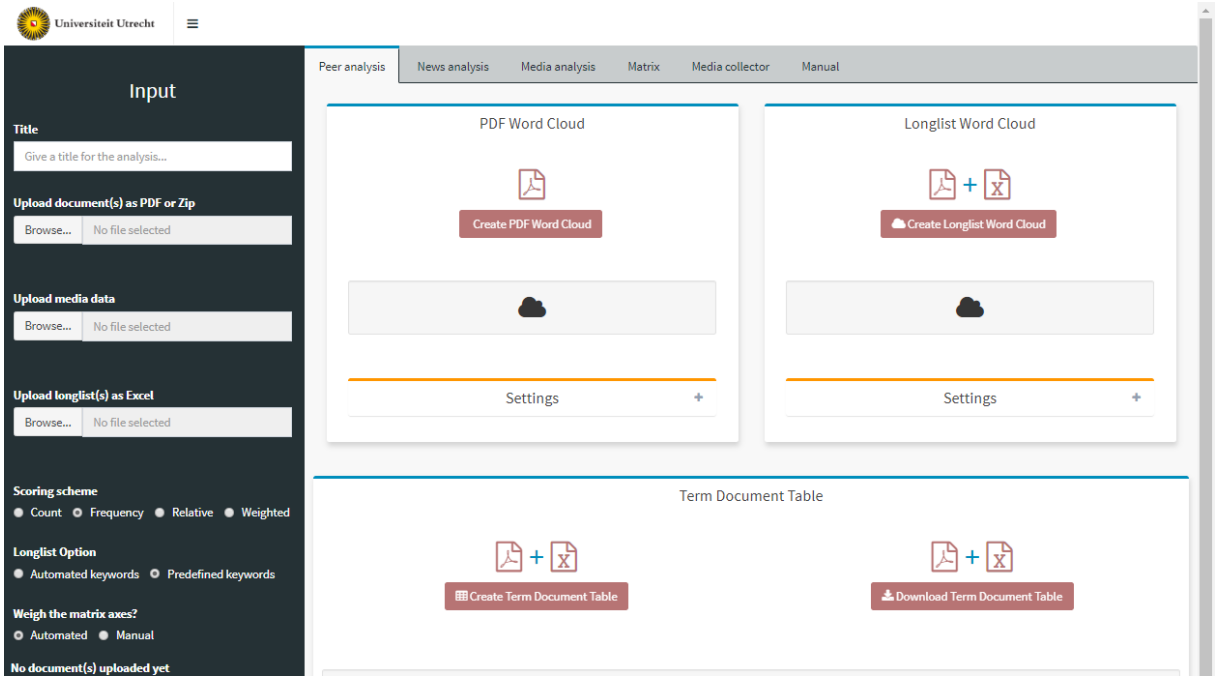
Figure 13. Example of weight by formula approach

A pro of this approach is the fact that it is visually appealing to users, where dragging a box into the axis is an intuitive solution. This also makes it easy for users to leave out a dimension they do not find important or does not provide data of sufficient quality to use. This dragging option has not been implemented, however the user can assign a weight to each source as they see fit.

The third and most advanced option for dimension aggregation is to use an analytic hierarchy process (AHP). This method relies on the idea of relative weighting between the dimension in which the user mentions how dimensions should be weighed against each other. The user provides input for this, providing that dimension A is x times as important as dimension B, and B is y times important as C and A is z times as important as C. This is also a downside, as the number of weights is $\frac{N^2-N}{2}$, growing to even more if sub-criteria are also considered. This could place an unwanted strain on the user, while a universally defined weight might not suit the wishes of the user.

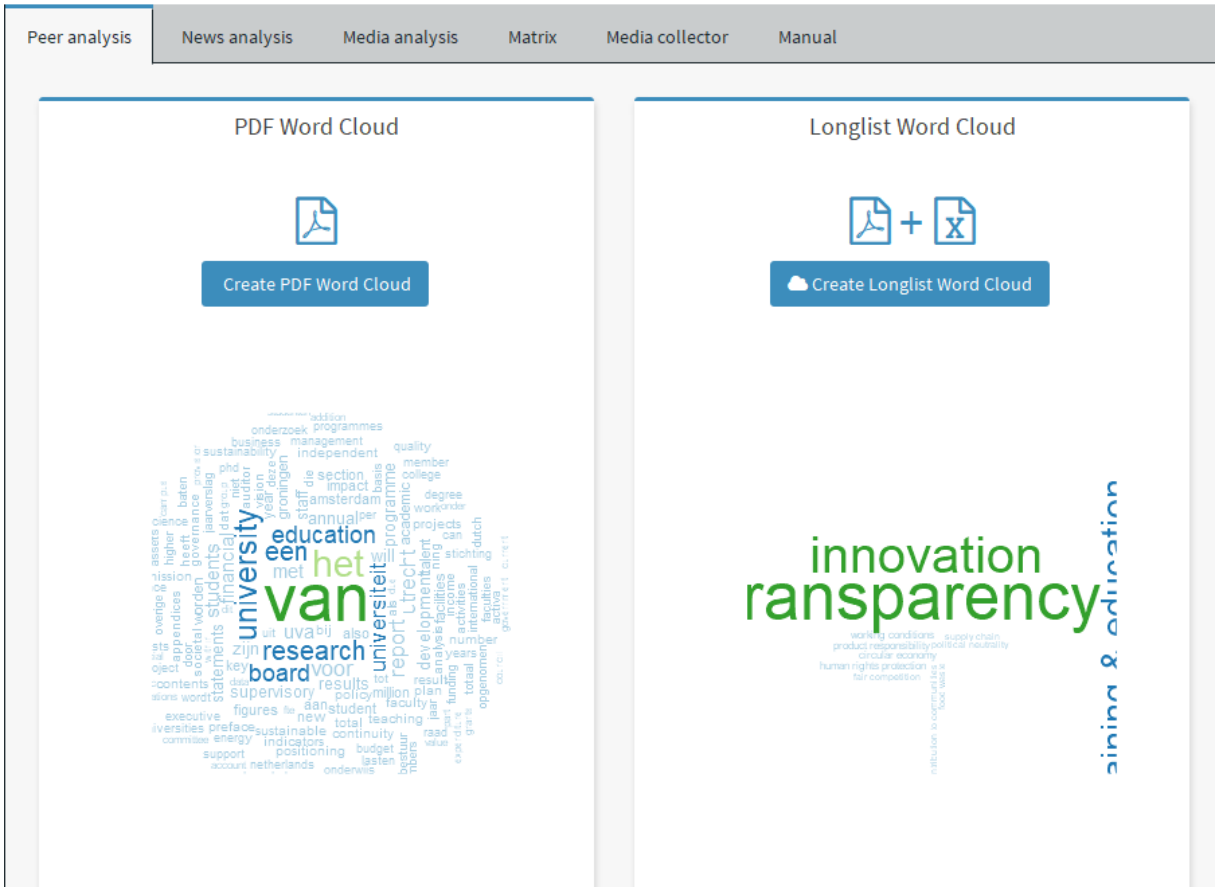
Appendix H: Screenshot guide of OpenMAsses

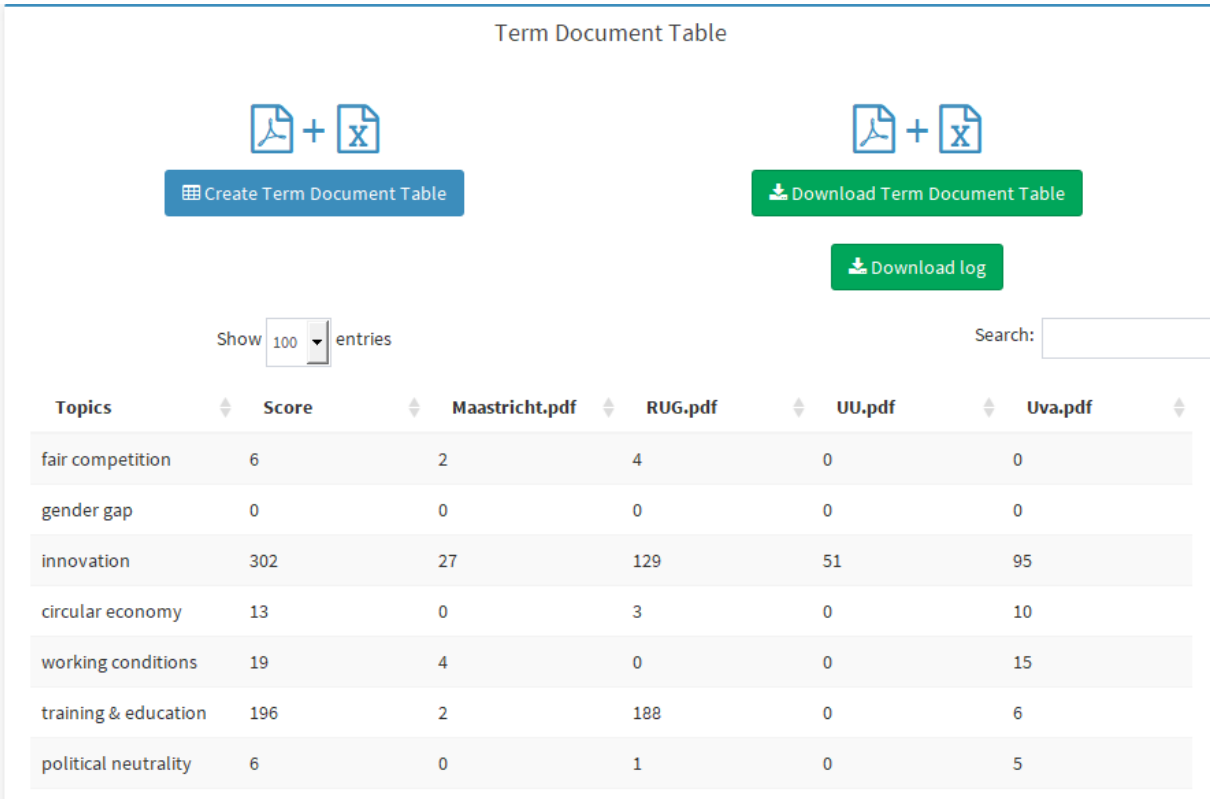
This guide describes all the main panels of the application. It opens with the first panel, on which the user can upload pdfs and media files (in .txt format). The user can also upload the longlist and select a scoring method, type of longlist and determine how the weigh the matrix axes. When these files are uploaded, the icons turn blue.



The user can create word clouds on this data, both in the peer analysis, media analysis and news analysis tabs. Below the word clouds document term matrices can be created, which show how often a term is used in a specific source.

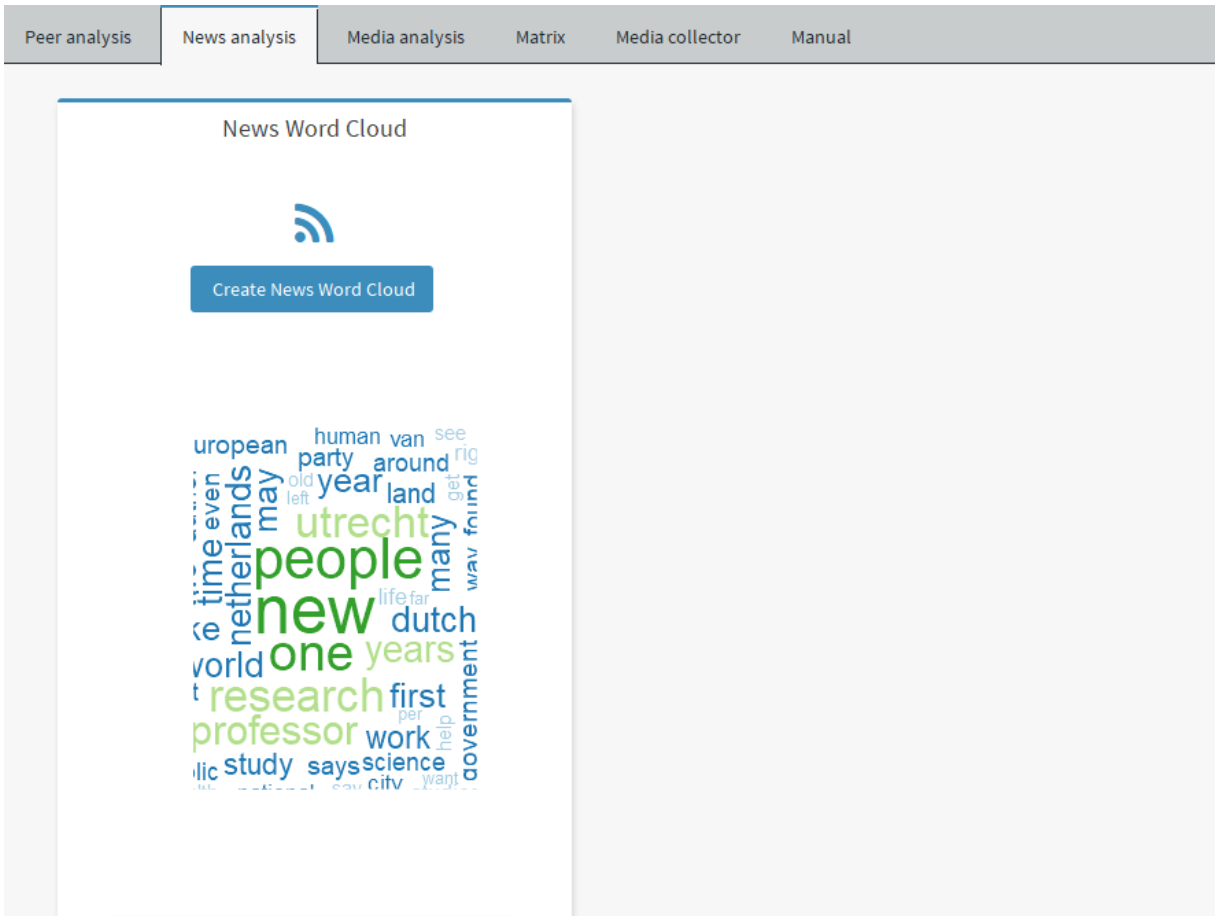
Appendix H: Screenshot guide of OpenMAsses



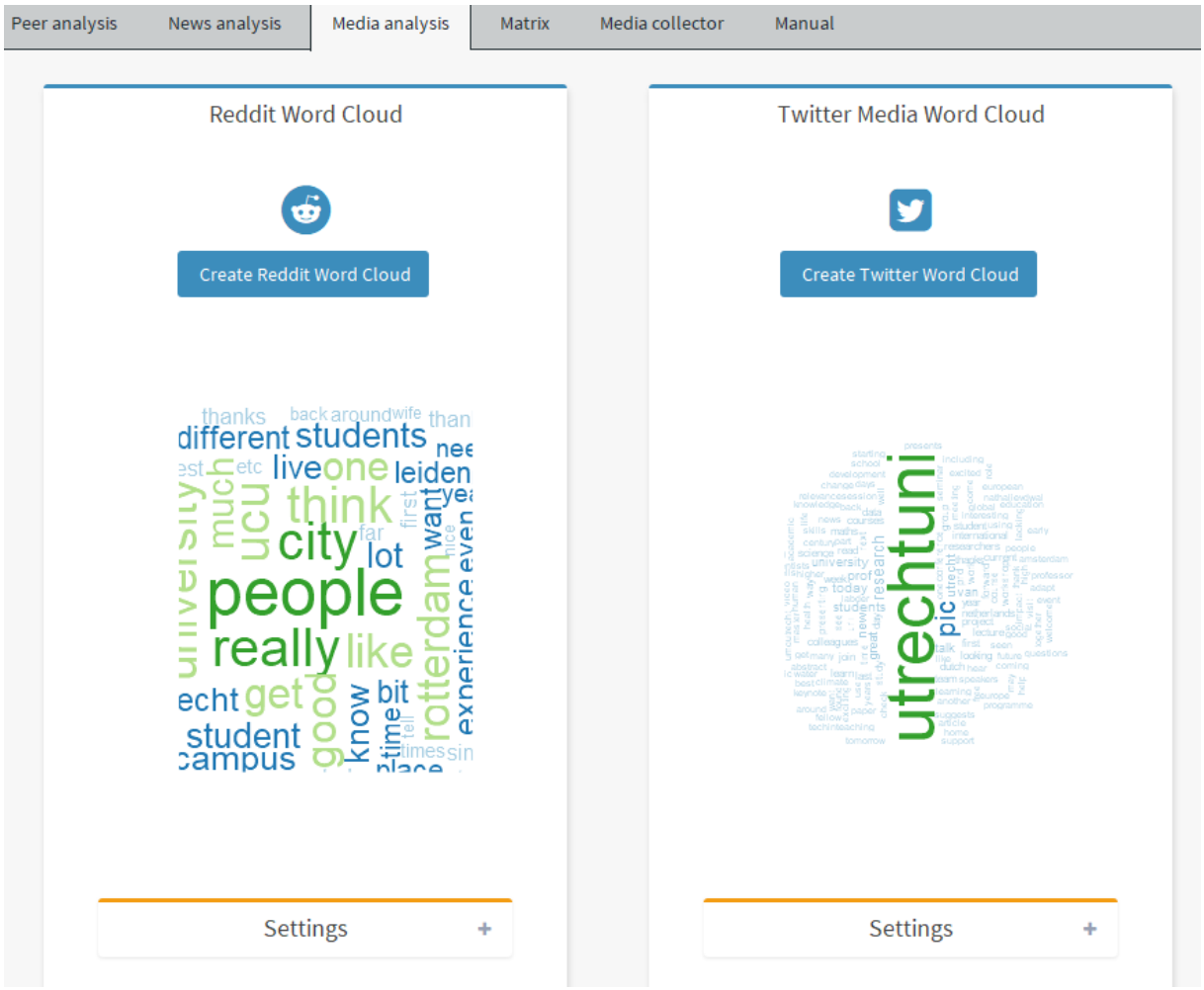


Topics	Score	Maastricht.pdf	RUG.pdf	UU.pdf	Uva.pdf
fair competition	6	2	4	0	0
gender gap	0	0	0	0	0
innovation	302	27	129	51	95
circular economy	13	0	3	0	10
working conditions	19	4	0	0	15
training & education	196	2	188	0	6
political neutrality	6	0	1	0	5

Appendix H: Screenshot guide of OpenMAsses

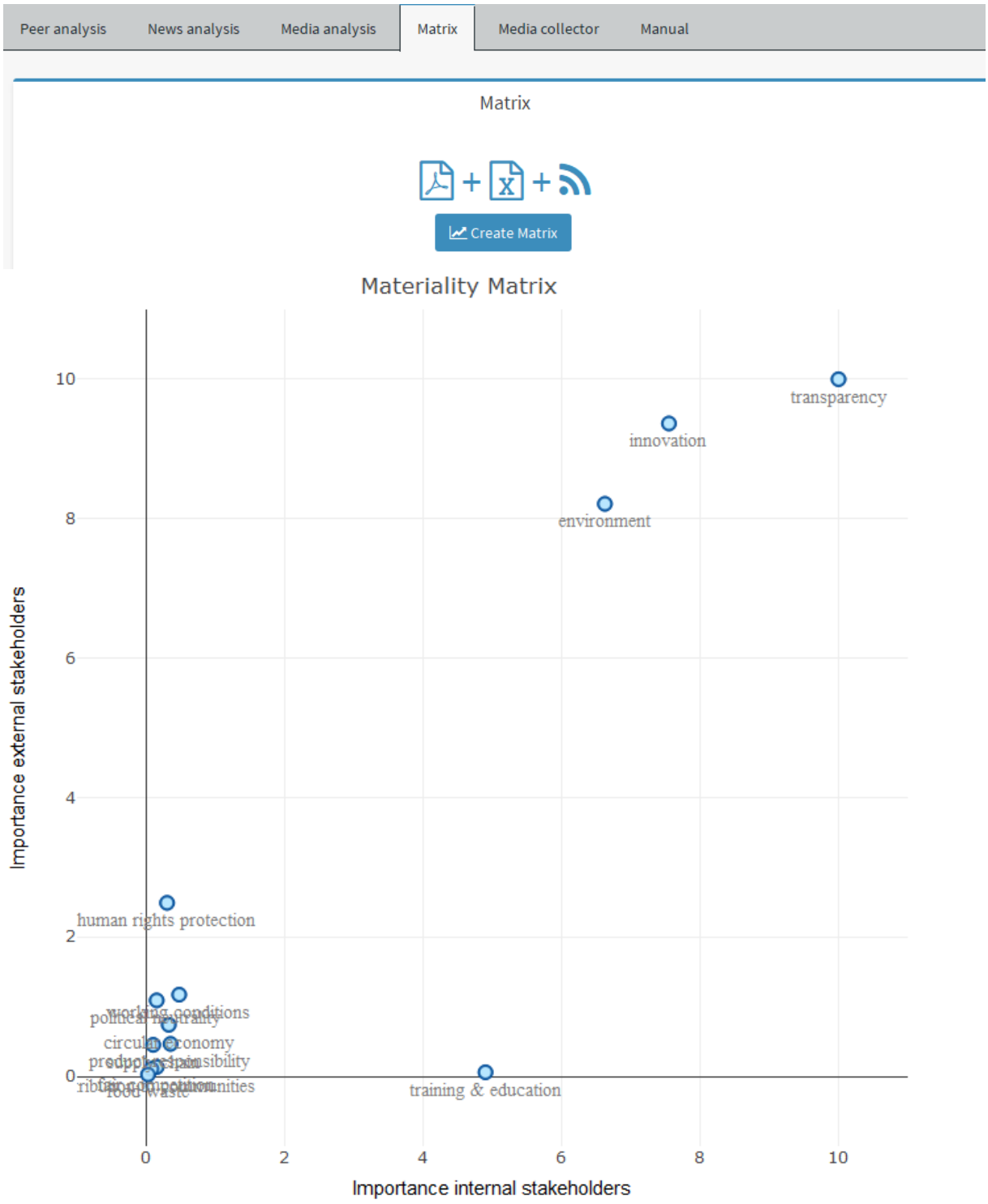


Appendix H: Screenshot guide of OpenMAsses

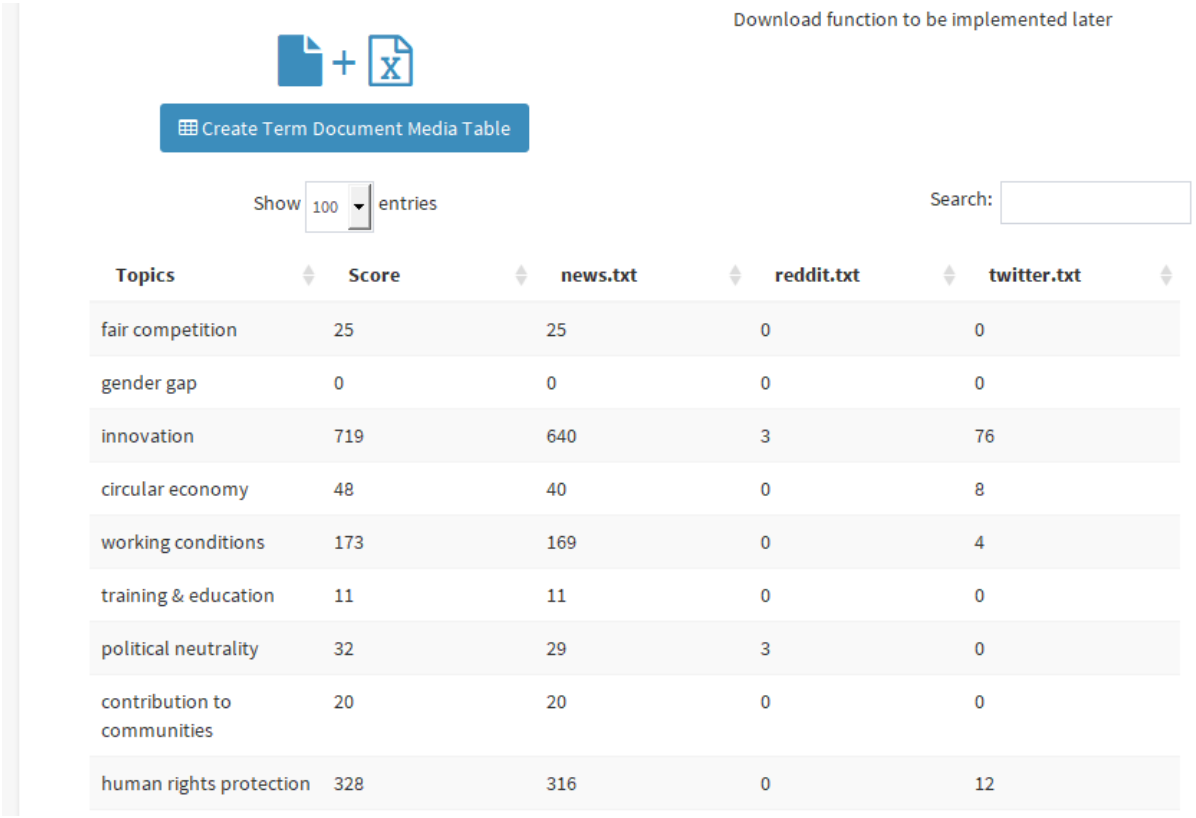


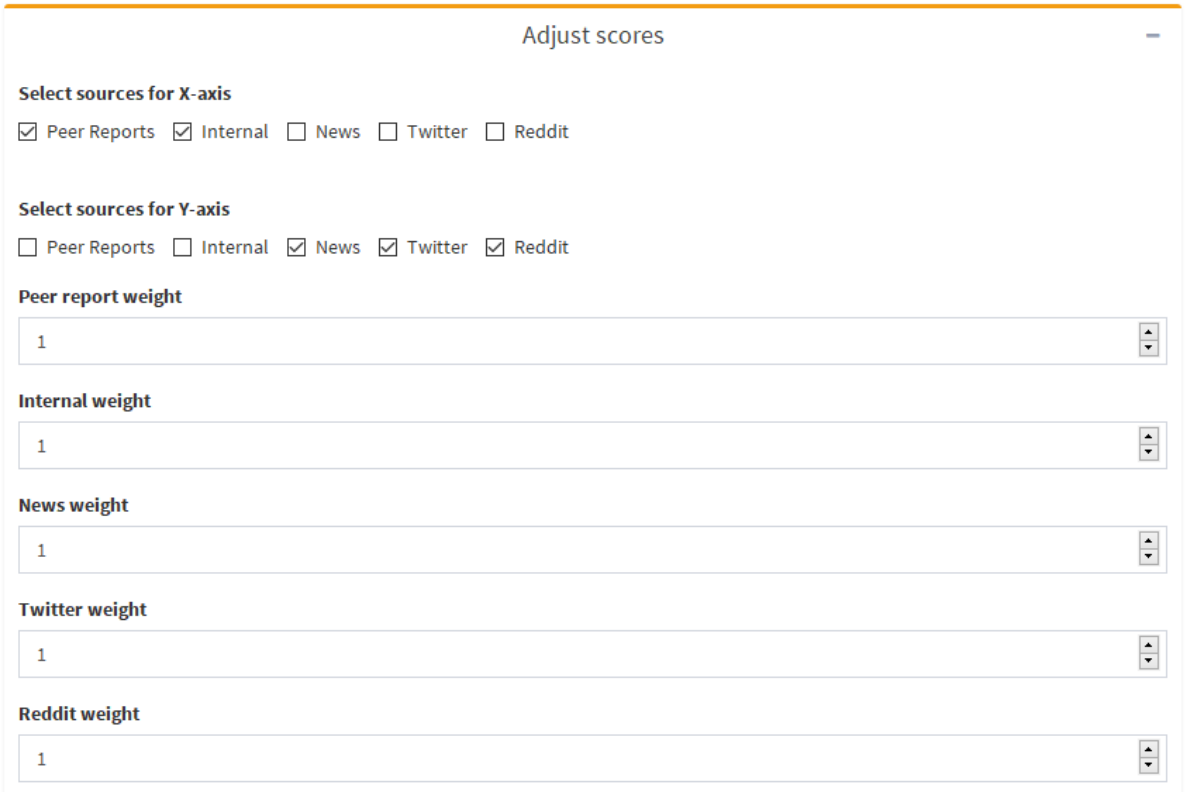
In the matrix tab the user can create materiality matrix. The user can tweak the weights of each source and adjust the score of each topic by a source if they feel that it does not accurately describe the content.

Appendix H: Screenshot guide of OpenMAsses



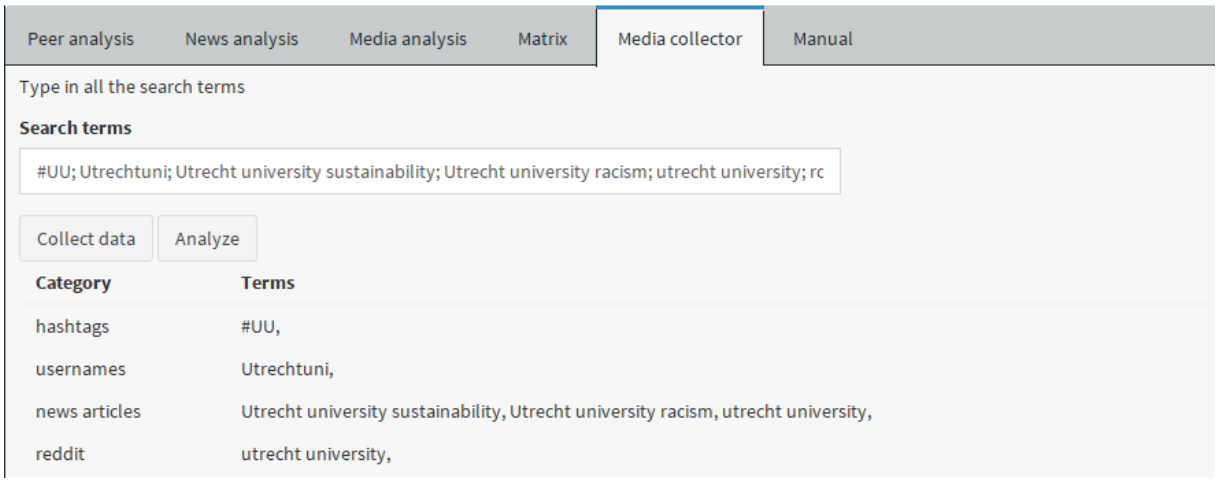
Appendix H: Screenshot guide of OpenMAsses



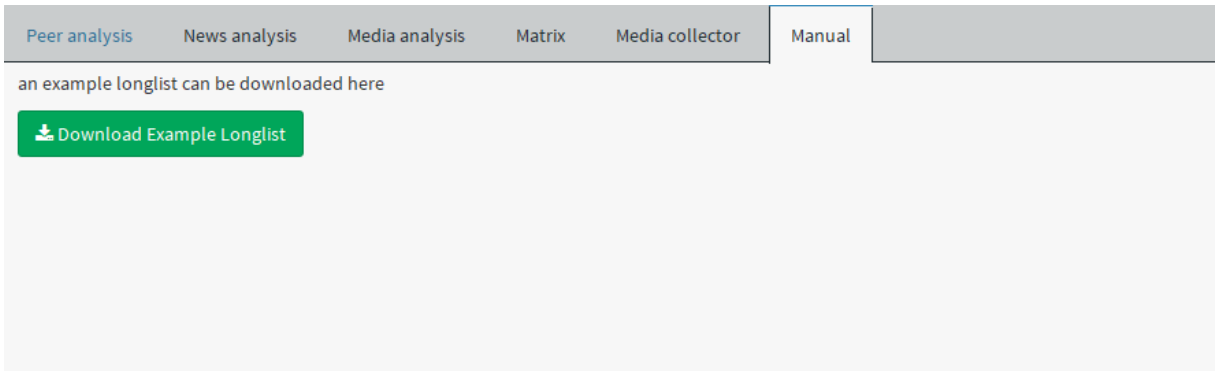


	Topics	Peer reports	Internal	News	Twitter	Reddit	Show
1		0.2	0.0	0.27	0.00	0.00	☑
2		0.0	0.0	0.00	0.00	0.00	☐
3		7.6	0.0	6.92	1.88	10.00	☑
4		0.3	0.0	0.43	0.00	1.05	☑
5		0.5	0.0	1.83	0.00	0.53	☑
6		4.9	0.0	0.12	0.00	0.00	☑
7		0.2	0.0	0.31	1.88	0.00	☑
8		0.1	0.0	0.22	0.00	0.00	☑
9		0.3	0.0	3.42	0.00	1.58	☑
10		0.4	0.0	0.54	0.00	0.39	☑
11		6.6	0.0	10.00	1.88	4.61	☑
12		0.1	0.0	0.91	0.00	0.00	☑
13		0.0	0.0	0.00	0.00	0.00	☐
14		0.0	0.0	0.00	0.00	0.00	☐
15		10.0	0.0	6.78	10.00	3.29	☑
16		0.0	0.0	0.05	0.00	0.00	☑

In the media collector tab, the user is able to enter search terms for the media collector. By pressing collect, the program then collects all the media data and writes it to text files.



In the manual tab, an example of a longlist is given.



Appendix I: Architecture documentation OpenMAsses

This section discusses the architecture of the OpenMasses tool. An overview is given in Figure 14, where after each component is discussed in detail. The light blue architecture is implemented in Pyhton, the dark blue in R. The square boxes are in- and outputs of the process.

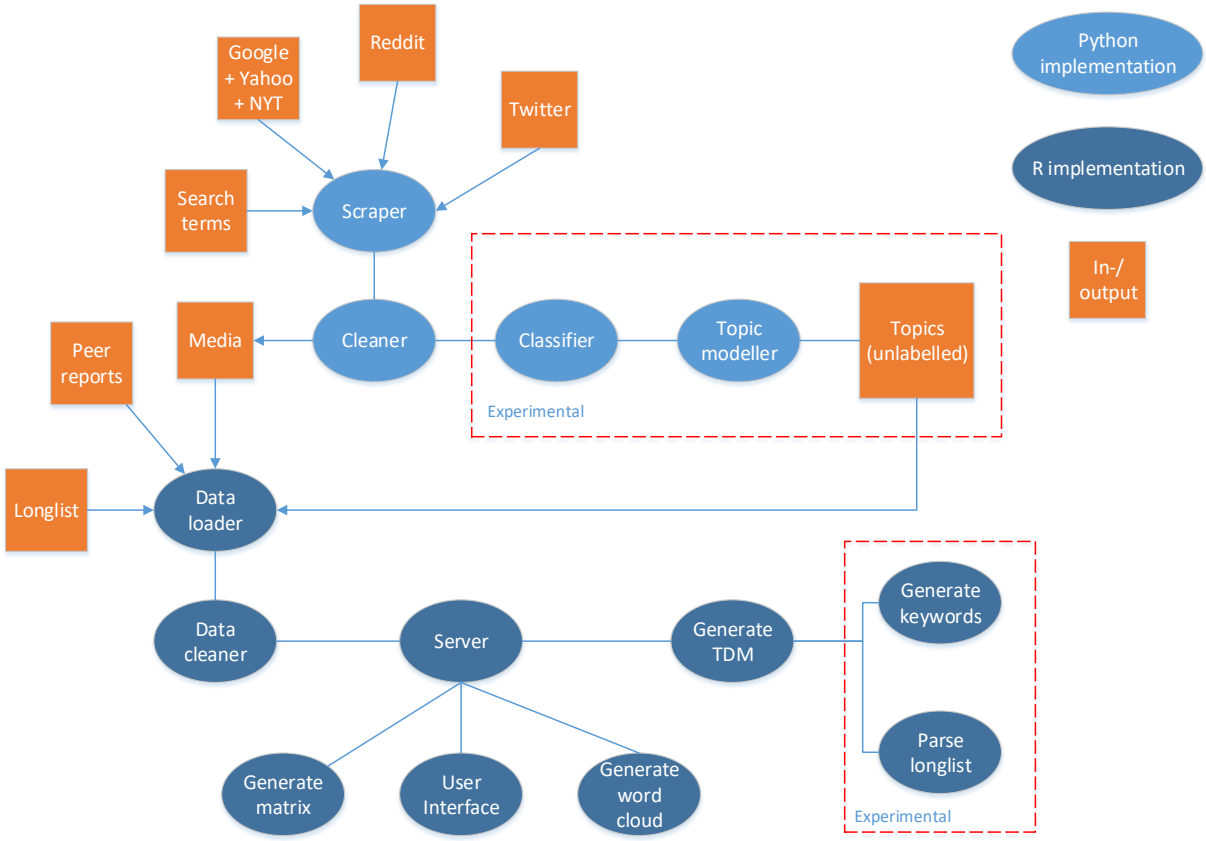


Figure 14. Architecture overview of OpenMAsses

Scraper

The scraper describes the spider object, which collects the data from the web. This is object relies on the modules requests, Tweepy, Praw[29], GetOldTweets[30], Beatifulsoup and Newsplease (Hamborg, Meuschke, Breitinger, & Gipp, 2017). The scraper consists of three parts: news, twitter and reddit.

For the news, a request is sent to Google News, Yahoo News. With beautifulsoup the hyperlinks to news articles are collected and stored in a text file. In addition, the New York Times API is used to collect search the Nyt as well. The url_links are added to the text file. Then, the newsplease module is used to access the content within each url.

For twitter, tweepy is used to collect tweets which contain a certain hashtag through the twitter API. For the search on usernames, the API was not deemed sufficient as contains a time constraint of roughly a week. Therefore, the GetOldTweets module was used for getting tweets from a specific username.

[29] https://praw.readthedocs.io/en/latest/

[30] https://github.com/Jefferson-Henrique/GetOldTweets-python

For reddit, the tool relies on the reddit API through Python Reddit API Wrapper (Praw). This module searches for the search terms on reddit and collects all the comments for each thread.

Cleaner

This cleaner uses a tokenizer and filters the stop words from the data. Odd characters are removed.

Classifier

This is not yet implemented, but its aim is to filter out some articles which are not discussing material issues.

Topic Modeler

The topic modeler uses the gensim implementation of the LDA to train the model. First, terms occurring more than five times are filtered from the data. Similarly, terms which occur in more than 20% of the articles are also removed. The model is run 25 times, using the C_v coherence measure to determine the best model, as it is superior to alternatives (S. Syed & Spruit, 2017).

Data loader

The data loader contains two functions. The load_documents function reads pdf files and transforms these into a list with the content of the pdfs. The load_longlist function reads the longlist and transforms it for further use.

Data cleaner

The data cleaner contains two functions where clean_text transforms this data into usable data, it performs stemming on this data, removes punctuation and stop words. Form_term_table cleans the input in the media collector and organizes the search terms in a table for visualization.

Server

The server is the main container of the program. It functions as a controller, responding to the user's input. It calls the functions described in other sections accordingly.

User Interface

The user interface contains all the code describing the user interface of the program. This is shown in more detail in Appendix H.

Generate matrix

This part contains three functions. The prepare_plotmatrix creates a document term matrix with all the sources. The generate_plotmatrix normalizes all the results between the zero and ten and fills in the document term matrix with the corresponding value. It then uses the plot_ly library to visualize the results and create an interactive matrix. Thirdly, save_TDM is a function to save an image of the plot.

Generate word cloud

This section contains three functions. In prepare_wordcloud the right documents are selected and a term document matrix is created on the input. Prepare_wordcloud_longlist does the same, but then gets the frequency of the topics, so that the biggest topics are displayed. Generate_wordcloud then serves as a plot function which creates the actual word cloud.

Generate TDM

This section contains six functions. Create_TDM is used to match the descriptions in the text. It calls a function combine_terms which groups multiple topics with the same name into a single one. It then calls count_terms to create a vector of with a value for each description, for each given input. This vector contains the number of times a description of a term matches the text. Add_scoreNew then applies the desired scoring scheme. Create_TDMOld serves the same purpose as Create_TDM but is an old implementation. This is also the case for the functions get_frequency (count_terms) and add_score. These were left in because the new functions do not cover all the program's options but require more time to execute.

Generate keywords

This section contains two functions. The function ozp_generate_keywords_nouns_adjectives creates keywords based on the nouns. The function ozp_generate_keywords creates a set of descriptions based on the list of descriptions.

Parse longlist

This section contains a single function, ozp_parse_longlist which transform the longlist to match other functions within the program. It groups multiple descriptions which are separated by a ';' into a list.

Appendix J: Collection of respondents' PDDs

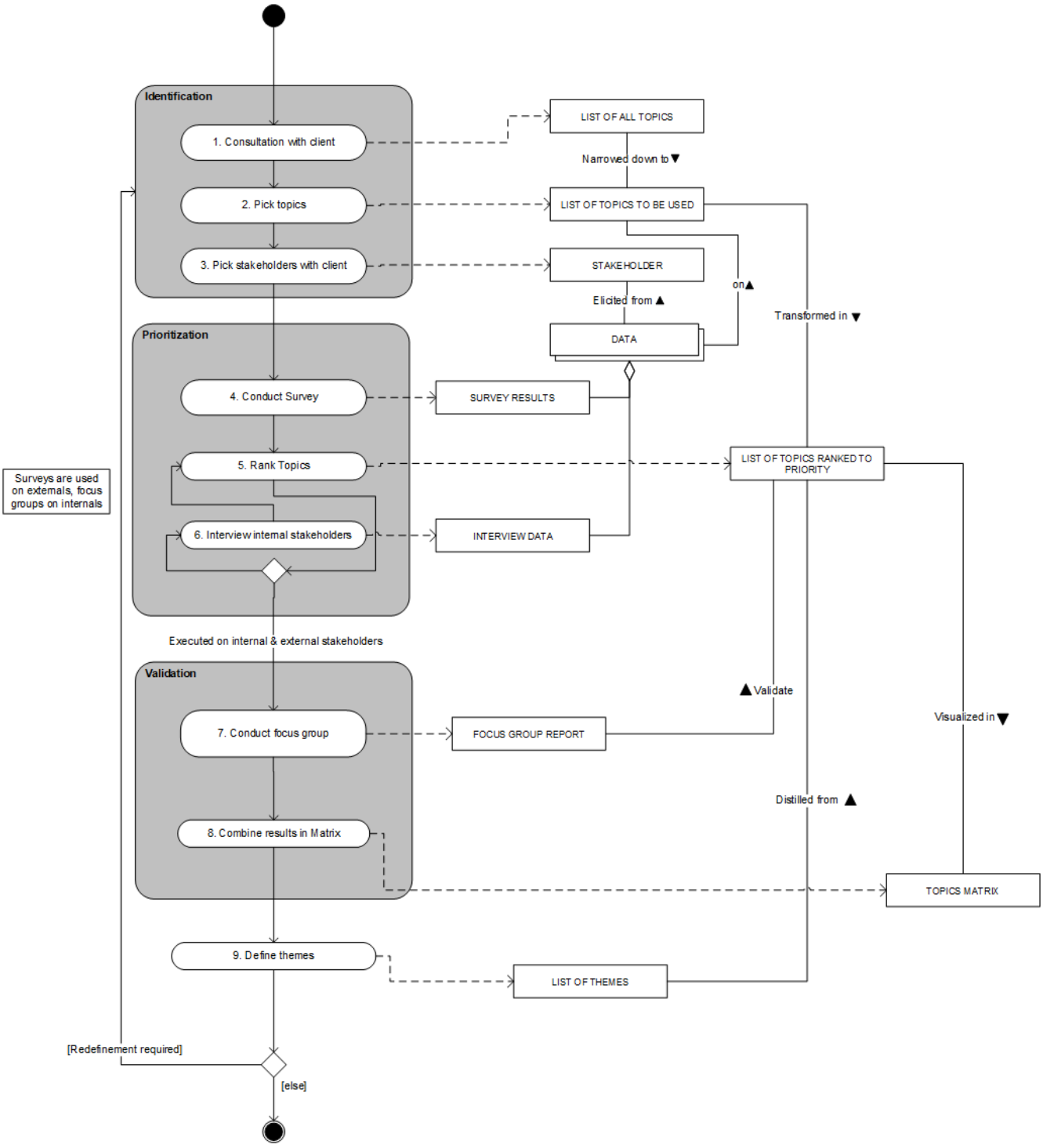


Figure 15. PDD of R1

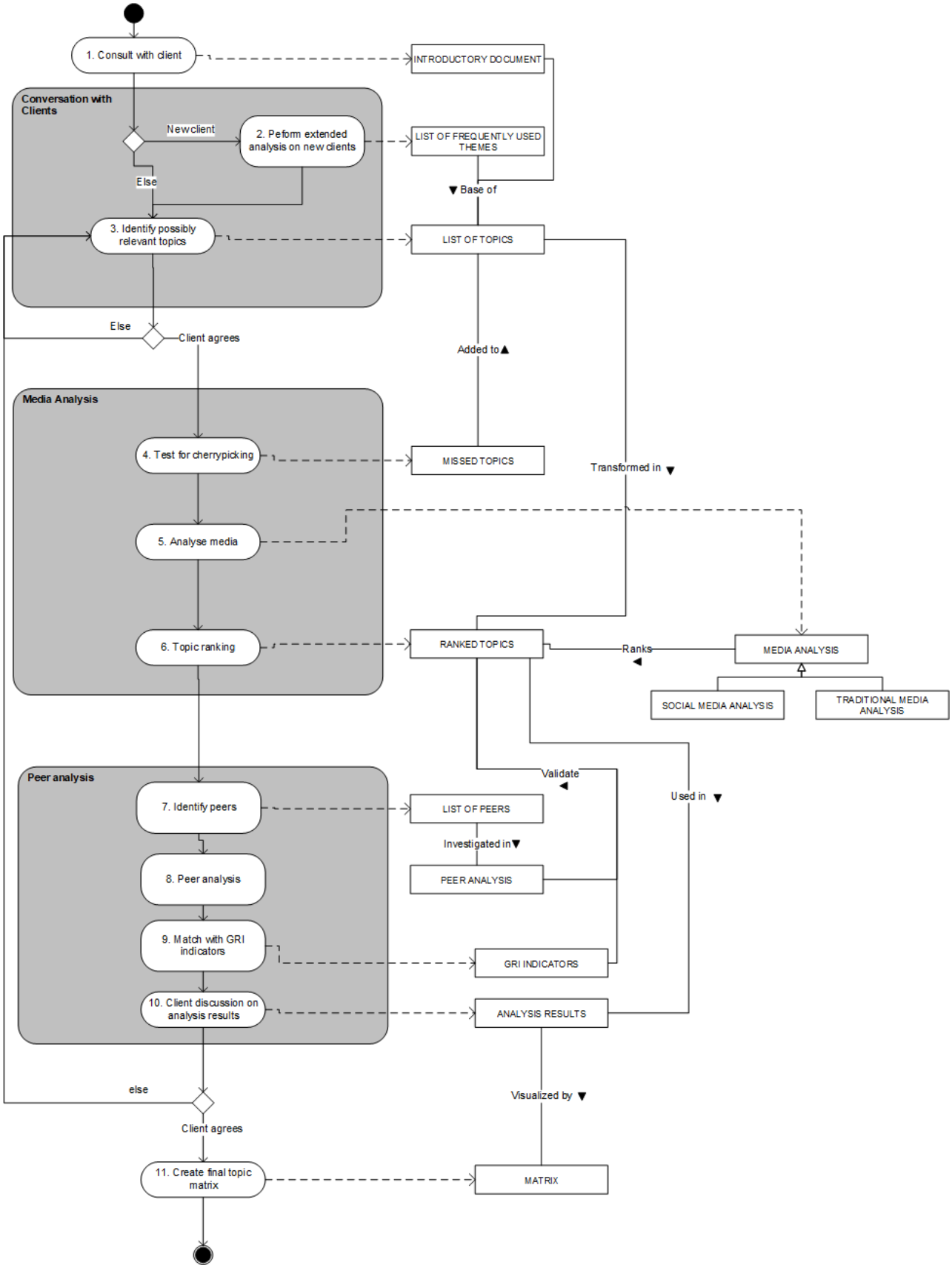


Figure 16. PDD of R2

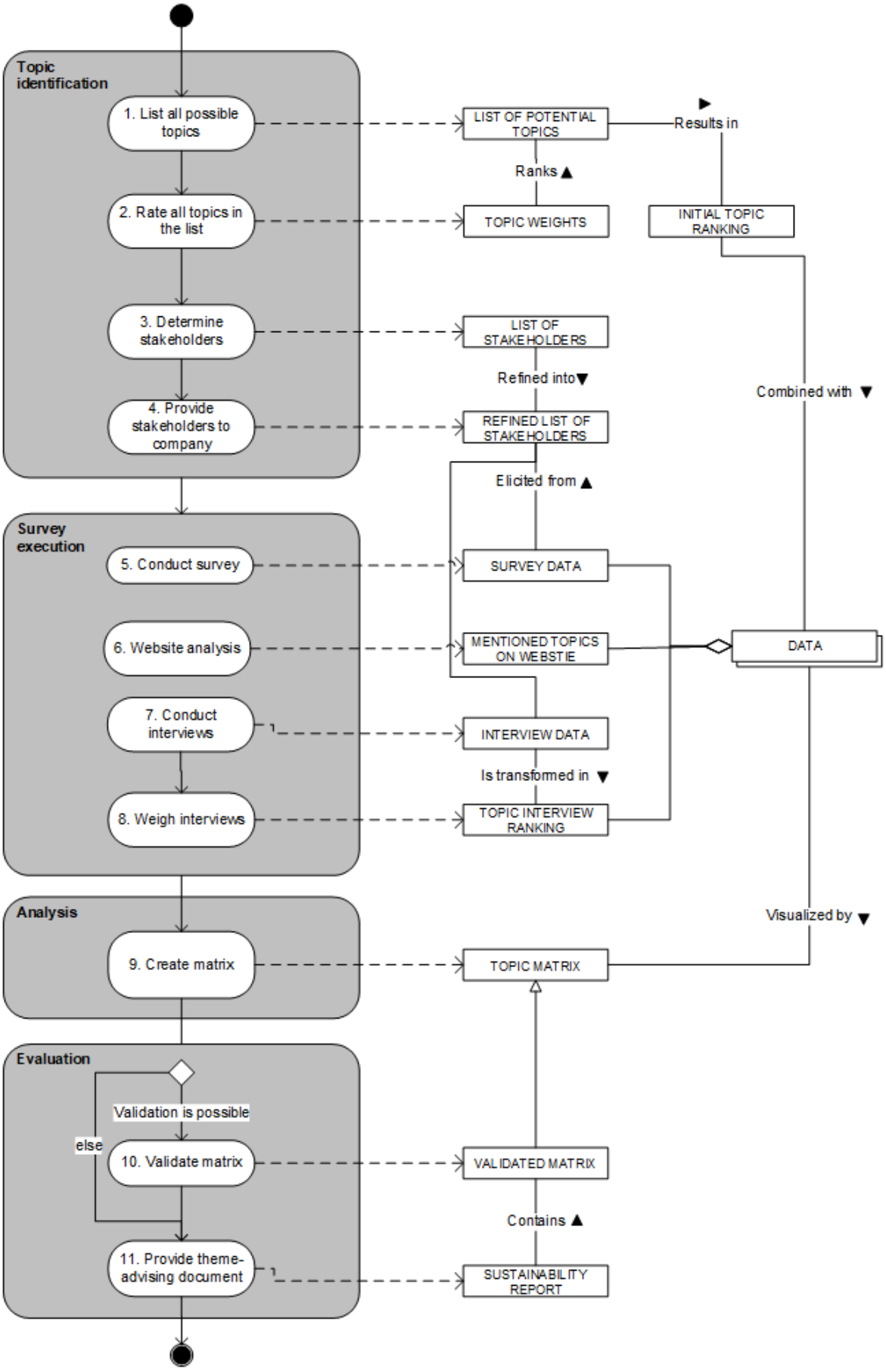


Figure 17. PDD of R3

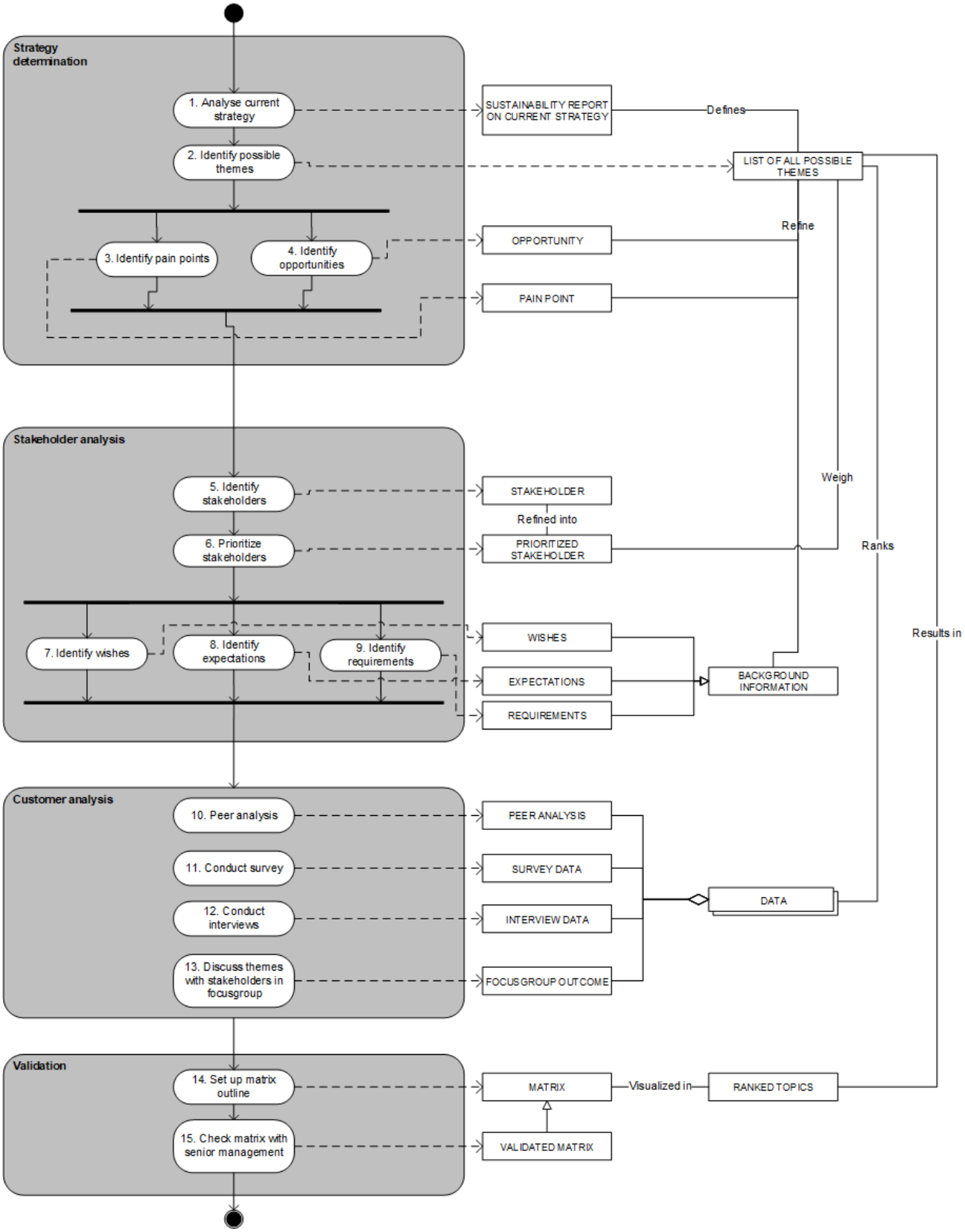


Figure 18. PDD of R4

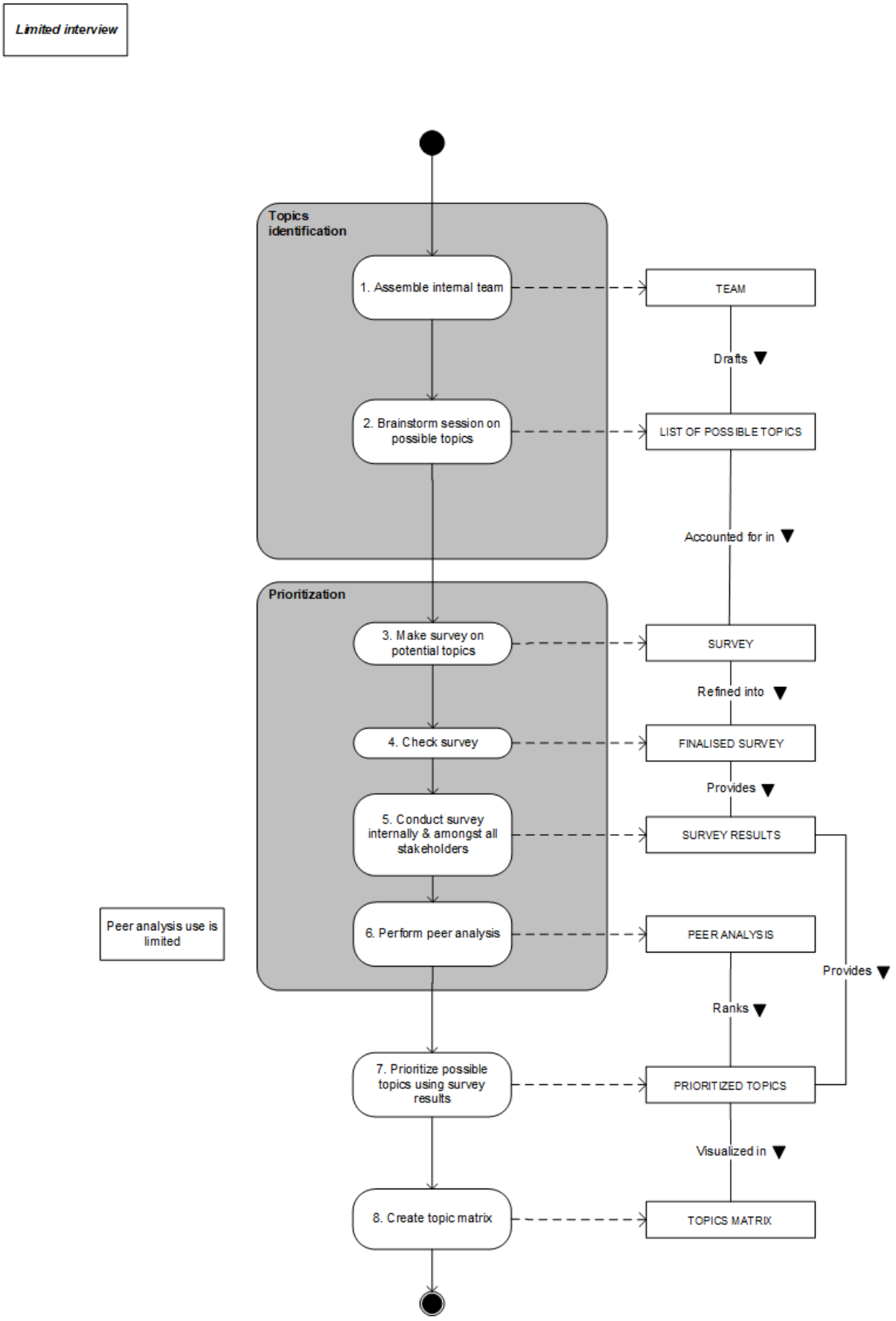


Figure 19. PDD of R5

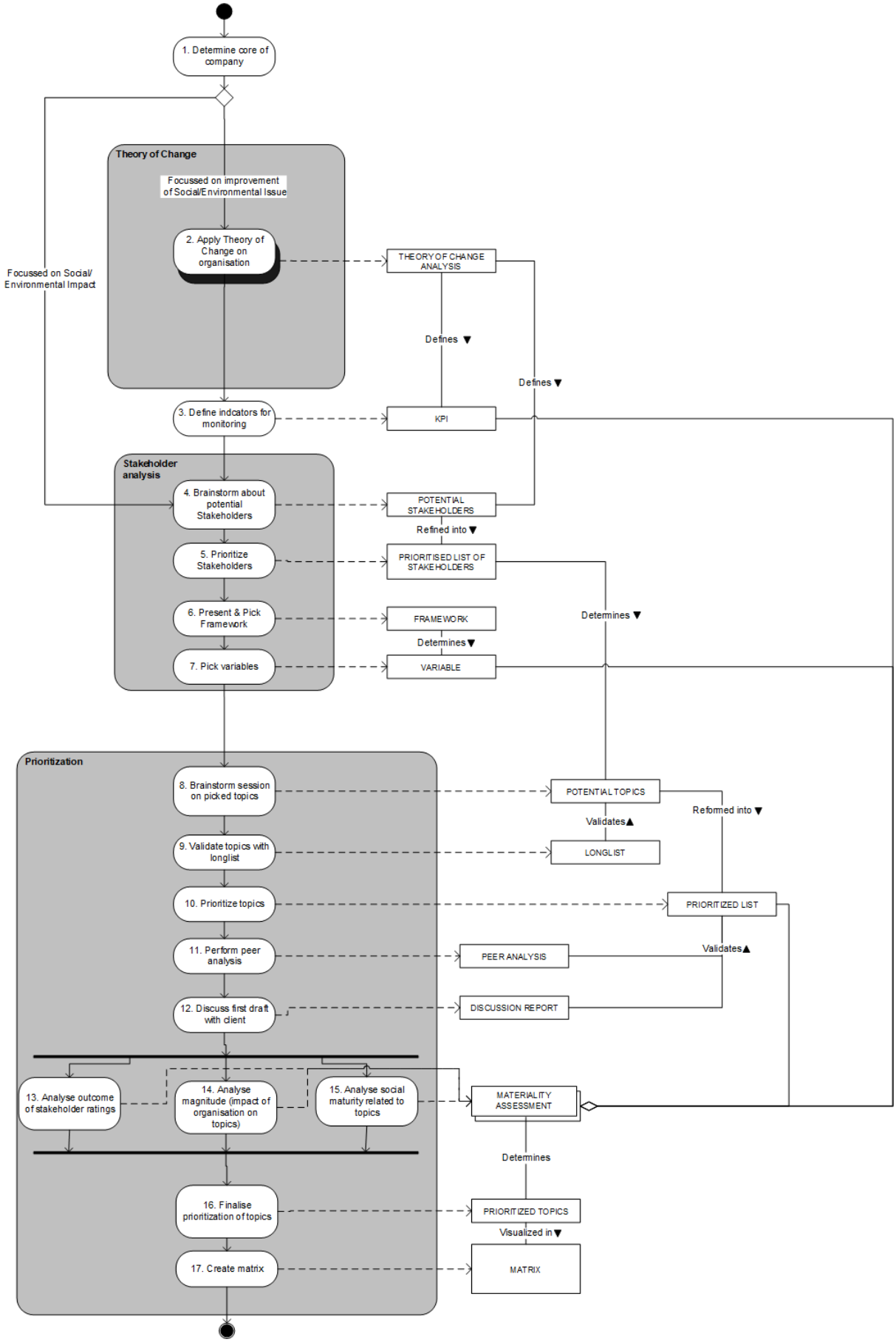


Figure 20. PDD of R6

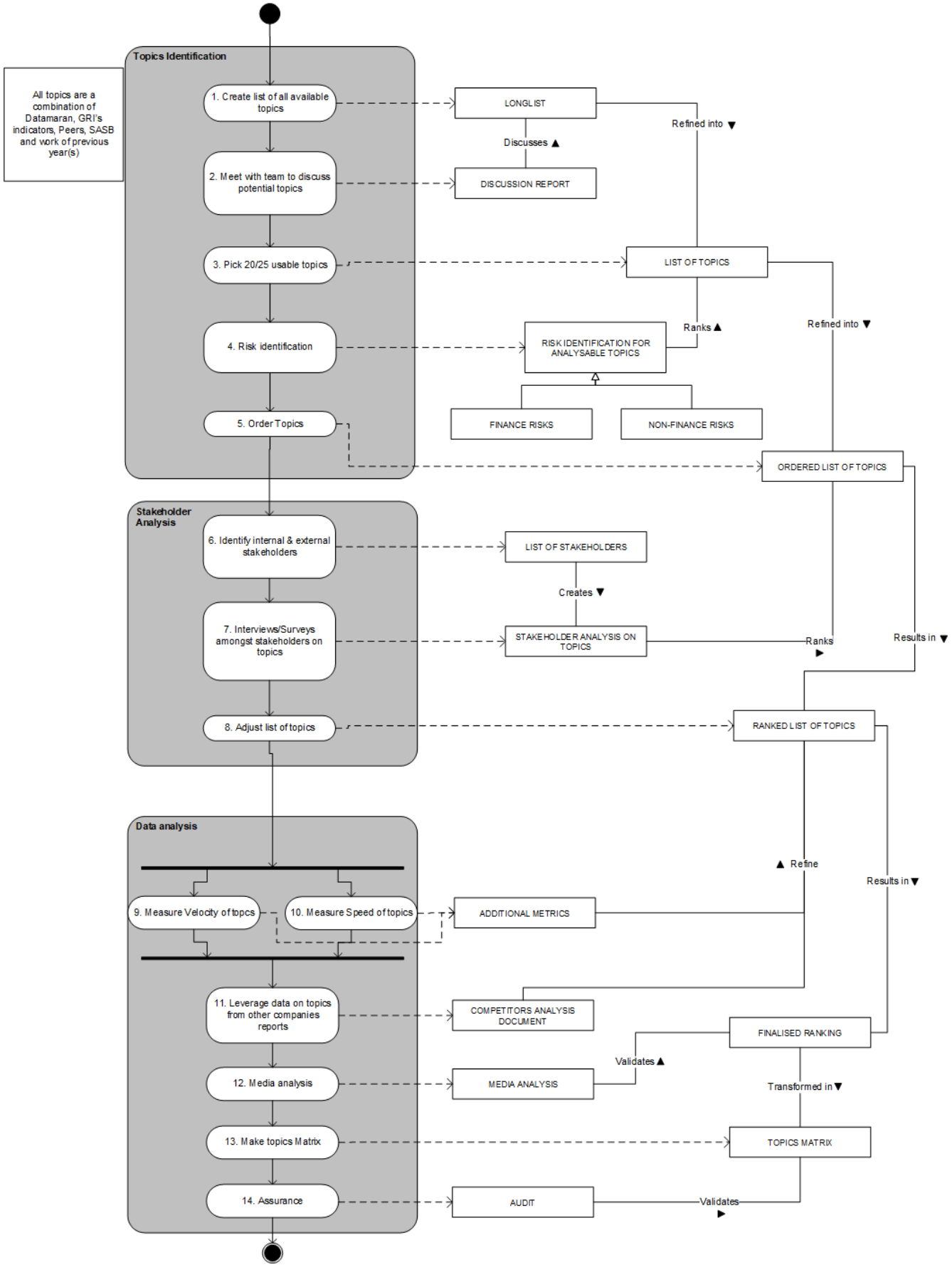


Figure 21. PDD of R7

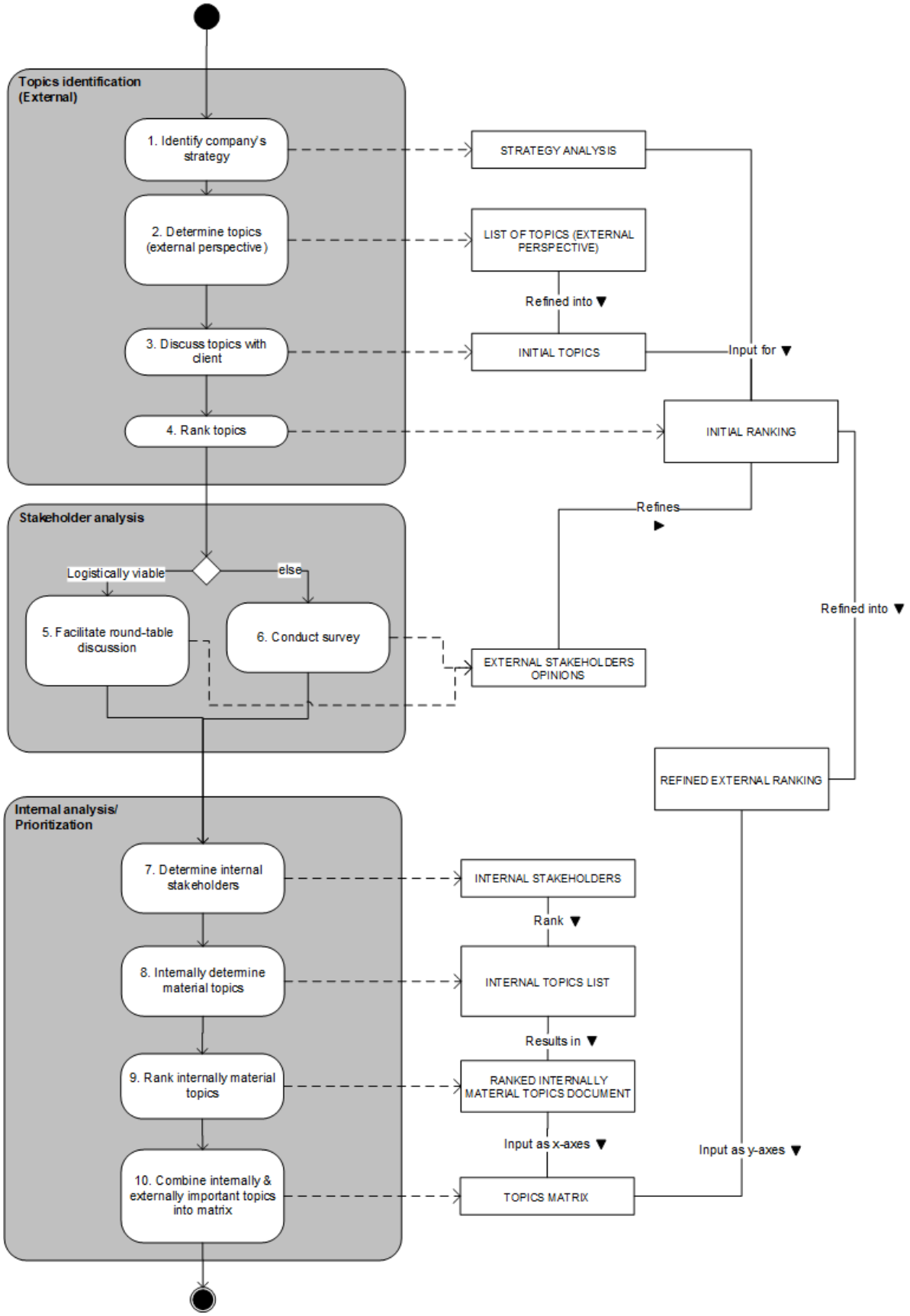


Figure 22. PDD of R8.

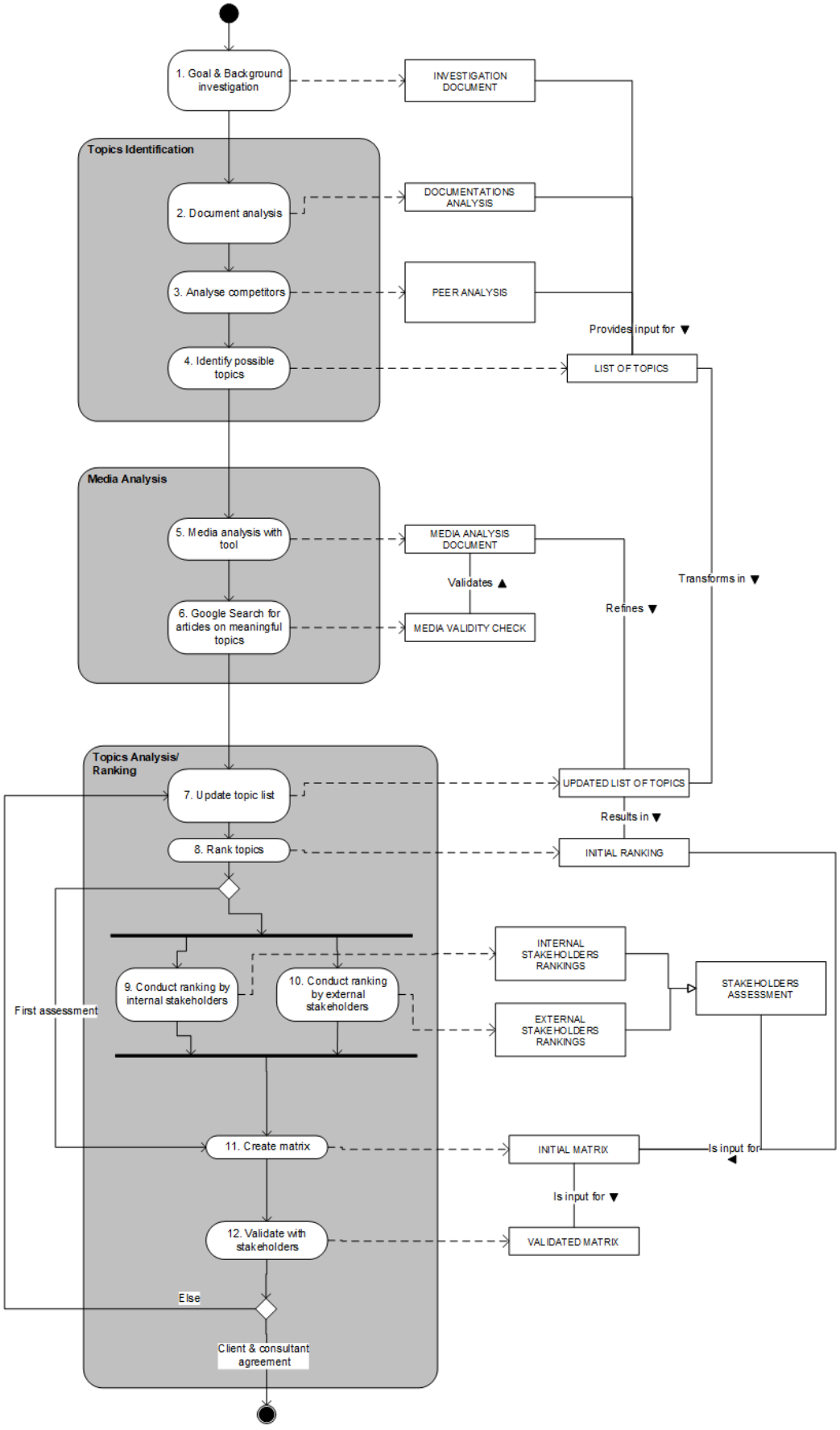


Figure 23. PDD of R9.