Social Business Intelligence

How and where firms can use social media data for performance measurement, an exploratory study

Joeri Heijnen

Master Thesis

Delft University of Technology

Faculty of Technology, Policy and Management

Social Business Intelligence

How and where firms can use social media data for performance measurement, an exploratory study

Master Thesis

Joeri Heijnen

j.heijnen@student.tudelft.nl

1320319

December, 2012

Faculty of Technology, Policy and Management · Delft University of Technology

The work in this thesis was supported by KPMG Advisory N.V. Their cooperation is hereby gratefully acknowledged.

Graduation Committee

Formal Chair
prof. dr. Y.H. (Yao-Hua) Tan
Full professor
Delft University of Technology
Faculty of Technology, Policy and Management
Section Information- and Communication Technology

First Supervisor
dr. ir. G.A. (Mark) de Reuver
Assistant professor
Delft University of Technology
Faculty of Technology, Policy and Management
Section Information- and Communication Technology

Second Supervisor
dr. M.E. (Martijn) Warnier
Assistant professor
Delft University of Technology
Faculty of Technology, Policy and Management
Section Systems Engineering

External Supervisor
ir. M.H. (Han) Horlings AITAP
Manager Business Intelligence
KPMG Advisory N.V.
IT Advisory
Business Intelligence

Abstract

Introduction

Both for individuals and for organisations the first decade of the 21^{st} century is characterised by the social media trend. Social media platforms are increasingly popular, and are amongst others used by individuals to express their opinions. Also firms acknowledge the opportunities offered by social media and are therefore increasingly pursuing to realise their goals through means of social media (Murdough, 2009). The value of the data produced on these platforms lies in the fact that *consumers* – i.e. (potential) clients – produce these data. In addition, the information is created instantly, real-time and by many people. It is therefore not surprisingly that Dey and Haque (2008) state that data generated from online communication acts as "potential gold mines" for discovering knowledge.

Next, firms are increasingly hungry for information that reveals underlying trends and dependencies affecting the firm's performance. Business intelligence systems are used to obtain such insights (Lonnqvist & Pirttimaki, 2006). The demand for (real-time) business intelligence systems and the popularity of social media offer room for synthesis. Systems that are purposed to derive actionable information from social media to support managerial decision-making are referred to as *social business intelligence* systems. Thus far, business intelligence systems particularly derive management information from internal data. With the rise of a new data source – social media platforms – the question rises how a firm should process these external data, what kind of managerial information could be derived from the new data sources, and whether or not each firm is able to apply social business intelligence. In business intelligence, indicators representing the strategy of a firm are established. These indicators are termed 'key-performance indicators'. Consequently, data reflecting the performance of different processes are linked to these key-performance indicators.

Whereas links between social media data and key-performance indicators may leverage the opportunities of social media for firms, a fundamental prerequisite allowing social business intelligence is the existence of user-generated social media content. After all, user-generated content that does not exist can not be analysed. Thus, an organisation is dependent for the generation of content on social media users and needs to determine whether social media data exists before considering to invest in social business intelligence systems. So far, it is not clear which organisational characteristics affect the existence of social media content. In this research, two general characteristics describing a firm are used to investigate the existence of social media data; (i) *industry* type and (ii) *customer relation* type.

Research Objective

On the one hand social media is a new phenomenon and acknowledged as a source of data of which valuable information can be derived. On the other hand, it is unclear which firms are able to collect social media data that is related to their firm and how firms should process these new data in accordance with existing business intelligence processes. Therefore, the objective of this research has been formulated as:

> *The objective of this research is to develop a procedure to utilise social media data for business intelligence, for which the applicability is investigated for firms in different industries and for different customer relations.*

Method

Our sample consists of social media messages related to eighteen different firms, in seven different *industries* performing different *customer relations.* Because the sample firms operate in different industries and execute

different customer relations, it is possible to gain insight in potential differences between the social media messages related to these firms. During a period of two weeks, social media messages from various platforms have been crawled into a local database to allow further analyses. The content in the dataset is sourced from Twitter, Facebook public pages, Flickr, Newssites, Google+ public pages, (Wordpress) Blogs, Picasa, YouTube and Friendfeed. These platforms are popular in Western Europe.

To gain insight in the amount of firm-related social media messages, the *average daily mentions* of firms served as a proxy to compare the volume of messages related to different firms. Next, using a content analysis, a portion of the collected messages have manually been classified into different categories based on the messages' subjects. These categories correspond with generally applied categories of key-performance indicators. As such, the results of the content analysis are directly linked to firms' key-performance indicators, allowing to draw conclusions on the relatedness of social media messages to different key-performance indicators.

Incorporating the new external data source requires traditional business intelligence systems to be adjusted. A social business intelligence procedure should be consistent with these traditional systems, and should additionally consider the challenges involved when processing social media data. As such, the requirements for a social business intelligence procedure have been established based on generally applied business intelligence concepts. Furthermore, the challenges involved in the processing of social media data are discovered by the collection of social media messages for the content analysis. Based on the traditional BI concepts and the challenges discovered in the content analysis, a business intelligence procedure is developed. The procedure is verified by analysing its consistency with existing BI systems and its ability to solve the issues emerging when processing social media data.

Results

The results of this research are twofold. Firstly, we gained insight in the applicability of social business intelligence by investigating the existence and content of firm-related social media messages. Secondly, a procedure to collect, process and analyse social media data for business intelligence purposes has been established.

(ii) Applicability of social business intelligence

The applicability of social business is investigated on two facets. Firstly, the volume of firm-related social media messages is investigated to obtain insight in the amount of data that is available for firms. The volume of firm-related social media content is however not sufficient to draw conclusions on the applicability of social business intelligence. Therefore, the second facet on which the sample data is analysed relates to the content of the social media messages. Especially, the subjects of the messages were analysed.

Volume

The average daily mentions differs from firm to firm. This implies that the applicability of social business intelligence will not be possible for all firms, since not for each firm data is generated. Figure 1 illustrates the average daily mentions of different firms in our sample.

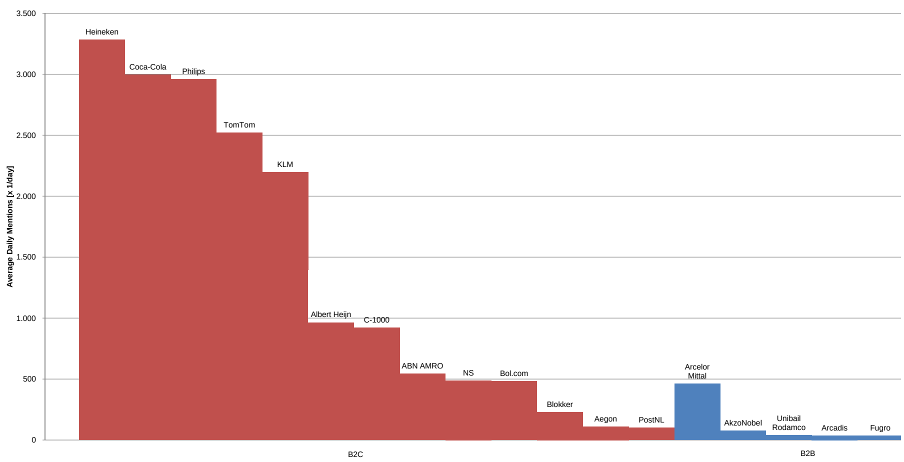


Figure 1: Average Daily Mentions of Firms, Clustered per Customer Relation Type

Figure 1 shows the daily volume of firm-related social media content, in which the firms are clustered on their *customer relation* type and consequently ordered descending. This figure suggests that B2C firms – coloured in red – are more likely to find social media content that is related to their firm than firms performing B2B relations (coloured in blue).

The second dimension on which the volume of firm-related social media content is investigated relates to *industries*. Our sample consists of eighteen different firms active in seven different industries. As a first step to identify possible differences in the volume of daily messages between industries, the firms have been clustered on industry type in figure 2, and have consequently been sorted in descending order.

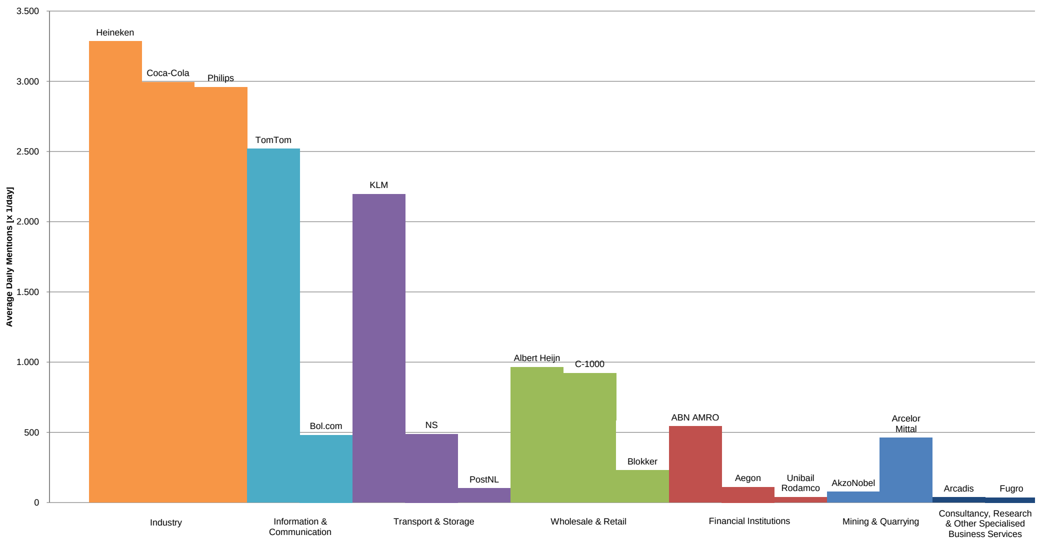


Figure 2: Average Daily Mentions of Firms, Clustered per Industry Type

Figure 2 strongly suggests that there exists a difference in the amount of user-generated content between different industries, with industrial firms being highly mentioned on social media, while consulting firms are the least mentioned.

Subjects
Next to an assessment of the amount of social media posts that are created on the web, this thesis examined the subjects of the social media posts in order to link the messages to firms' key-performance indicators. The social media messages of the firms have been classified into categories based on their subject. These categories are based on ten categories of commonly applied key-performance indicators. Consequently, the collected social media posts of the firms in the sample have manually been classified into one of these categories.

Our analysis shows that the subjects of social media messages differ from firm to firm. The majority of social media messages related to firms (41%) express how the external stakeholders of a firm perceive the company. In this thesis, such posts have been classified as *community* posts. 18% of the social media messages in our dataset contained the name of a firm, but did not contain any valuable information for the firm and have consequently been assigned as *undefined* posts. About 11% of the social media messages relate to *financial results*, which consist of *financial performance discussions* (5%) and *stock related discussions* (6%).

The content analysis of this research suggests that the subjects of social media messages related to B2B firms contain a higher percentage of *short term financial results*, *news* and *professionals* related messages than messages related to B2C firms. Unfortunately for B2B firms, such type of information is yet available internally. Acquiring social media data to gain additional management information is therefore of less value for B2B firms. Next, the analysis indicates that the social media messages related to B2C firms contain a higher percentage of posts related to *customer relations*, *product and service quality* and *product and service innovation* than messages related to B2B firms. It are these types of information that deliver additional value to the firm, since this information is not available at firms internally.

In addition, the content analysis of this research suggests that the subjects of social media posts differ between industries, but that the majority of the subjects in each industry relates to community, i.e. social media posts revealing how the community perceives the company. The results indicate that firms active in the *information & communication*, *financial institutions* and *transport & storage* industries are more subjected to social media messages related to *customer relations*, while firms active in the *mining and quarrying* and *consulting* industries will find messages related to *financial performance*.

(ii) Procedure for social business intelligence

Based on (i) traditional business intelligence frameworks and (ii) the experience we gained in collecting, processing and analysing social media data in the content analysis, a social business intelligence ("SBI") procedure has been developed. Figure 3 schematically shows the social business intelligence procedure.

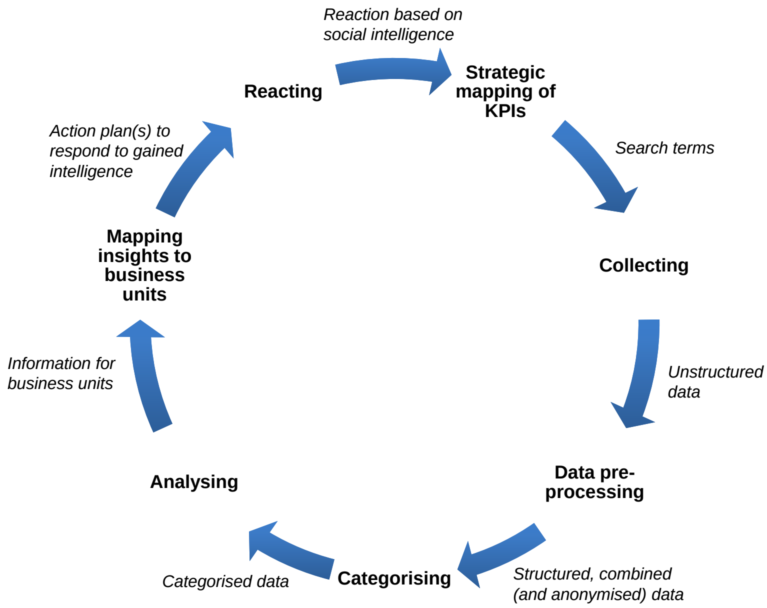


Figure 3: Blueprint: Social Business Intelligence Procedure

Our SBI procedure consists of seven main components, being (i) strategic mapping of KPIs, (ii) collecting, (iii) data pre-processing, (iv) categorising, (v) analysing, (vi) mapping insights to the business units, and (vii) reacting. The seven steps can be interpreted as a cycle, i.e. the output of the last step influences the first step.

The very first step of social business intelligence sets the scene for the objects that are to be collected and analysed. Namely, in the first step the key-performance indicators that are to be measured by social media data are selected. Not each type of KPI is to be measured by social media data since there does simply not exist any related social media data to these types of KPIs. Firms should mainly focus on KPIs related to customer relations, public image and – to a less extent – on product and service innovation when selecting KPIs that are to be measured using social media data.

The second step of the SBI procedure relates to data *collection*. In contradiction to regular BI systems, the data is to be sourced from external parties in social business intelligence. People create firm-related messages on different platforms, of which the vast majority of publicly accessible messages are created on Twitter. The search terms that are used to filter out the content at which the firm is interested should be based on the social KPIs selected in the previous step.

The social media data has been collected from multiple platforms which adhere to their own data format. The different format are to be combined into one uniform database, so that – in a later step – data analysis can be applied on the complete dataset. Furthermore, the firm should select those attributes that are necessary for the analysis, not each platform offers the same richness of attributes to a social media post. In addition, the data should be anonymised to be in compliance with new Regulations regarding data privacy. Finally, spam – i.e. social media posts that do not relate to the firm – should be removed from the collected data.

The *data pre-processing* step resulted in a structured database in which the social media messages from multiple platforms are combined. In the *categorising* step, the messages are clustered on different issues of interest, depending on the firm's subject of interest. E.g., messages related to certain products can be categorised, or one can cluster the messages that are created by people with many followers, etc. Again, the criteria at which the messages are categorised are determined by the selection of the social KPIs in the first step.

So far, the collected data has not provided any insights. It is in this *analysis* step of the procedure where data is transformed into information. The categories that were established in the previous step are analysed in this step. For instance, sentiment analysis can be applied on the categories related to the firm's products in order to acquire intelligence related to customer experiences of the products. However, the most valuable intelligence

is gained when social media data is related to internal data. For instance, the volume of social media messages related to a certain product may be correlated with the sales volume of that product. It is in this phase of the SBI procedure where such relations are explored.

In the first step of the procedure, KPIs have been selected. These KPIs typically relate to a certain function of the firm, and hence have an “owner”. The intelligence gained in the previous step relates to KPIs, and should feed back to the owner of the KPI. Generally, it are the people in the firm that are responsible for the KPI who are the ones that can reason how the KPI is influenced. Therefore, these people are the ones that can draft an action plan in case the KPI needs improvement.

The final step of the social intelligence procedures consists of the execution of the action plans that are developed in collaboration with people from the business lines that are responsible for the respective KPIs. Actions on the gained intelligence may involve revisions of internal processes or strategies, or external interventions such as social media engagement.

Contents

Preface

Social media is a trend in the first decade of this century, and the concept is increasingly incorporated in the daily lives of people. Scepticism towards the new technology is losing support, and companies are aware that the new trend cannot be denied. Though the new phenomenon is gaining attention in the scientific world, social media was not yet part of the curriculum at my faculty. I am grateful that I was offered the opportunity and the confidence to dive in the rather unexplored world of research into social media, and explore the opportunities for companies offered by social business intelligence.

First of all I would like to thank my graduation committee. My first supervisor of Delft University of Technology, Mark de Reuver, critically reviewed my work on a regular basis. I hereby thank Mark for his constructive comments and suggestions for improvements, I experienced our meetings as pleasant and useful. Mark's experience in scientific research and knowledge of ICT and business models contributed to the quality of my thesis. Martijn Warnier supervised my work as second supervisor from the Systems Engineering section. Martijn indicated issues concerned with social media (data) that I did not think of in the first place, for which I am grateful. Harry Bouwman chaired my committee as professor from the ICT section. Harry contributed to this thesis by critically reviewing my work and offering suggestions for improvements, which were mainly related to scientific concepts. Thank you for these comments, the critical notions improved the level of this work.

I would like to express my gratitude to my supervisor at KPMG, Han Horlings. Thanks to weekly meetings with Han I was driven to progress my thesis. I found a sparring partner to discuss especially business intelligence related aspects of my work. Han, thanks for your time, contributions and coaching! In addition I would like to thank all employees and co-interns of KPMG's Business Intelligence department for their interest in my thesis, their contributing opinions and ideas on the subject, for being challenging competitors during the karting event and for the fun at the Amsterdam Parade last summer.

Amstelveen, December 2012

Joeri Heijnen

Chapter 1

Research Problem

1-1 Introduction

More and more, customers are using content sharing sites to express their opinions about almost anything, from soccer matches to financial statements of large corporations. Examples of platforms where these expressions are shared to the world are blogs and forums, social network sites and wikis. In 2008, "75% of internet surfers used social media" (A. M. Kaplan & Haenlein, 2010), and the usage of social media is not limited to teenagers. Members of generation X, now 35–44 years old, are increasingly active on social media sites (A. M. Kaplan & Haenlein, 2010). Anno 2012, people express how they feel, what they do, what they think of, and what they intend to do in over 340 million daily Twitter posts (Twitter, 2012). The value of the information produced on these platforms lies in the fact that *consumers* produce these data. In addition, the information is created instantly, real-time and by many people. Since social media posts are often non-anonymous and directly linked to a person, firm or brand, the content produced on social media platforms can be interpreted as an indicator of people's attitude towards a firm, product or service. The user-generated content is considered as a driver for future sales by Dhar and Chang (2009), hence containing economic value for firms (Ghose & Panagiotos, 2010).

In the first decade of the 21^{st} century, business intelligence ("BI") has evolved to one of the critical processes for organisations to provide useful insight, to support decision-making, and to drive organisational performance (Ramakrishnan, Jones, & Sidorova, 2012). According to Watson and Wixom (2007) BI has become a "strategic initiative and is regarded as an instrument in driving business effectiveness and innovation". For organisations, it is increasingly important to quickly respond to changes in the environment (Gessner & Volonio, 2005). Therefore, BI systems are required to contain a component that allows monitoring the real-time environment. We define such systems as 'real-time BI' systems.

From the above, we can derive two trends in the current business landscape:

(i) an increase in the usage of social media, and,

(ii) an increase in the usage of business intelligence systems.

Trend (i): Social Media in Organisations Organisations are increasingly pursuing to realise their goals through social media (Murdough, 2009). Social media applications support organisations in creating value in many of their activities, e.g. in marketing, services, human resource management and customer relationship management (A. N. Smith, Fischer, & Yongjian, 2012). In addition, firms are able to acquire data from social media at low costs. Dey and Haque (2008) state that data generated from online communication acts as "potential gold mines" for discovering knowledge. It is therefore that this thesis focuses on the extraction of information based on the data created by consumers on social media.

The increased application of social media has serious consequences for an organisation's exposure to the actors in their environment, which include (potential) customers, suppliers and competitors. It seems that the power has been taken from the corporate marketing departments by individual consumers that create, share and discuss online blogs, tweets, Facebook entries, movies, pictures, etc. (Kietzmann, Hermkens, McCarthy, & Silvestre, 2011). With or without permission from the organisation, communication about brands will happen. In an environment where customers gain more and more power, an organisation needs to carefully treat its actions and control its exposure. Therefore, companies empower employees to talk, listen, and respond to what consumers post on social media (A. N. Smith et al., 2012).

Though many organisations acknowledge the opportunities in the application of social media, there also exists a fair degree of uncertainty with respect to allocating marketing effort and budget to social media, and "limited understanding" of the social media platforms (Weinberg & Pehlivan, 2011). Kietzmann et al. (2011) argue that many executives avoid or ignore social media because they do not understand what it is, how to engage with it and learn from it. This thesis contributes to a further understanding of social media and discovers opportunities to leverage the valuable content on these platforms for business purposes.

Trend (ii): Business Intelligence in Organisations Business intelligence systems are applied to obtain a better understanding of underlying trends and dependencies – often coming from the external context – that affect the business (Lonnqvist & Pirttimaki, 2006). Whereas BI systems were initially perceived as tools that were used exclusively to support strategic decision-making, organisations have recently commenced to further exploit the capabilities of BI systems to support wider business activities (Elbashir, Collier, & Davern, 2008).

The scale of recent investments in BI systems reflects the growing importance and highlights the need for more attention in research studies. Elbashir et al. (2008) estimated that global spending on BI systems and related products reached USD 6.1 billion in 2008. A paper by Gartner (2009) predicted that organisations will increase spending on "packaged analytic applications, including corporate performance management ("CPM"), online marketing analytics that optimise processes, not just report on them". Azvine, Cui, and Nauck (2005) predict that in the future, "business intelligence will be available to everyone in the enterprise, and will be embedded in many business systems".

1-2 Problem Statement

The demand for (real-time) business intelligence and the popularity of social media offer room for synthesis. The opportunities offered by linking both concepts are acknowledged in the literature, e.g by Dey and Haque (2008) and Lovejoy, Waters, and Saxton (2012). However, search queries[1] related to the subject of this thesis into the scientific databases ScienceDirect and JStore, and the search engine Google Scholar resulted in the understanding that social media applications for BI purposes are relatively underexposed in the literature. Generally, research in the area of social media is related to marketing activities, sales, promotions, public relations and customer relationship management, e.g. by Dong-Hun (2010); Ratner (2003); Klassen (2009); Kozinets, de Valck, Wojnicki, and Wilner (2010); Kirtis and Karahanb (2011); Hanna, Rohm, and Crittenden (2011); A. M. Kaplan and Haenlein (2012); You, Xia, Liu, and Liu (2012). The focus of the research conducted in the literature is mainly focused on the organisation expressing itself to the outside (social media) world, whereas this thesis focuses on the incoming aspect. A reason for the shallow results discovered in the literature may be the relatively new character of combining social media and business intelligence.

Zeng, Chen, Lusch, and Li (2010) distinguish social media research between social media *analysis* and social media *intelligence*. Social media *analysis* is concerned with "developing and evaluating informatic tools and frameworks to collect, monitor, analyse, summarise, and visualise social media data". Social media *intelligence* – on the other hand – "aims to derive actionable information from social media in context-rich application settings, develop corresponding decision-making or decision-aiding frameworks, and provide architectural designs and solution frameworks for existing and new applications that can benefit from the wisdom of crowds through the web".

Many social media monitoring tools, like Socialmention.com, Radian6, RowFeeder, Trackur, uberVU, SAS Social Media Analytics, Finchline, Sprout Social, etc. mainly reveal the performance of a firm on social media (number of mentions, number of likes, % of positive mentions), and treat the social media component of a firm as a separate business unit executing its own strategy. However, the purpose of business intelligence is to reveal the underlying parameters that determine the performance of the organisation, that is, not limited to solely social media performance. In order to understand the influence of social media content on a firm's performance, a link between the company's key performance indicators ("KPIs") and social media parameters is required because KPIs measure the performance of an organisation with respect to its strategy. Some social media

[1](SOCIAL BUSINESS INTELLIGENCE), (BUSINESS INTELLIGENCE 3.0), (SOCIAL MEDIA) AND (ORGANISATION), (SOCIAL MEDIA) AND (BUSINESS), (SOCIAL MEDIA) AND (BUSINESS INTELLIGENCE), (TWITTER) AND (BUSINESS INTELLIGENCE), (WEB 2.0) AND (BUSINESS INTELLIGENCE), (SOCIAL MEDIA) AND (STRATEGY)

monitoring tools, like Kapow Software and ListenLogic seem – at a glance – to establish this link. Zeng et al. (2010) highlight the need for clearly defined social media performance measures because much of the research is conducted in a setting which aims to support decisions in organisations. We argue that the possibilities of social media for business intelligence purposes reaches further than what is currently offered by the social media analytics tools. This argument is supported by Reinhold and Alt (2011), who state that "existing tools still have a limited functional scope". The key benefits will be gained whenever the KPIs of an organisation are linked to the parameters that are measured by social media tools. Only in that case, one can speak about 'social business intelligence'. This thesis contributes to a transition from social media 'monitoring' towards social 'business intelligence'.

Whereas links between organisational performance and social media content can leverage the opportunities of social media for firms, a fundamental prerequisite allowing social business intelligence is the existence of user-generated social media content. After all, user-generated content that does not exist can not be analysed. Thus, an organisation is dependent for the generation of content on social media users and needs to determine whether social media data exists before considering to invest in social business intelligence systems. However, it is not clear which organisational characteristics affect the existence of social media content. The following section illustrates which factors are to be considered when one tries to categorise the availability of social media content that is related to firms.

Firstly, it is likely that within some *industries* users express their opinions more often than in other industries. We expect that one expresses his or her opinion more often about a product that is purchased on a frequent basis. For example, domestic products are purchased more frequent than a car or a house. Therefore, the consumer industry is probably discussed more often than the real-estate market. Secondly, the *relation with end-users* makes it that people discuss the company on social media, or not. Some firms are more visible for consumers than others. Zhang, Jansen, and Chowdhury (2011) support this factor by concluding that "business engagement on social media relates directly to consumer's engagement with online word-of-mouth communication". When users experience malfunctions in a mobile network they complain at the firm at which they signed the contract, while the firm that delivered the network equipment – which may be responsible for the errors – remains unaffected. This example illustrates that it is necessary to make a distinction between companies in the same industry based on their position regarding consumers. Turban, Lee, King, and Chung (1999) classify e-commerce into either business-to-business ("B2B"), business-to-consumer ("B2C"), consumer-to-consumer ("C2C"), consumer-to-business ("C2B"), non-business e-commerce, or intra-business e-commerce" (as cited in Chen, Jeng, Lee, and Chuang (2008)). We will use this classification to assign an organisation's position regarding consumers since it clearly illustrates how close an organisation acts to the end consumer. As such the network service provider can be positioned as a B2C firm, while the provider of the equipment performs B2B relations.

Next, in the case that there exists social media content, an organisation should be able collect and analyse the data. The unstructured nature of the data, various languages, various data formats, interpretation difficulties, unverified information and privacy issues are aspects that make the usage of social media data for business intelligence different from 'regular' – i.e. internal management information – BI data.

Knowledge Gap From the previous, we can conclude the following. It is unclear in which industries and for which type of customer relations firms can apply social media data for business intelligence. Secondly, there is no understanding how organisations should process social media data in relation with business intelligence. Taking into account the previous, the following knowledge gap is formulated:

> *It is unclear how firms can process social media data for business intelligence, and how the applicability of social media data for business intelligence varies among different industries and different customer relation types.*

1-3 Research Objective

Social media is a new phenomenon, and increasingly popular for both consumers and organisations. Business intelligence is applied in organisations to measure organisational performance and to provide managerial information. The literature agrees that social media posts may contain valuable insights for organisations that managers can use in their decision-making. Hence, the two concepts offer room for synthesis. However, there does not exist a structured procedure that prescribes how organisations should acquire and analyse these social media posts in order to generate managerial information. In addition, it is unknown how (i) different industries and (ii) different customer relations affect the existence of social media data on the web. After all, if (potential) clients do not generate social media posts related to a firm, it will not be possible to derive information from the posts. Therefore, the objective of this thesis is formulated as:

> *The objective of this research is to develop a procedure to utilise social media data for business intelligence, for which the applicability is investigated for firms in different industries and for different relations with end-users.*

As such, insight in (i) the suitability of social media for business intelligence for different organisations and (ii) a procedure prescribing the steps required for social business intelligence is obtained.

Concepts in Research Objective In order to clarify the research objective, the key concepts are listed and explained below.

- *Procedure to utilise social media data for business intelligence*
  A procedure to utilise social media data for business intelligence prescribes which steps are necessary when an organisation applies social media data for the measurement of organisational performance. Within business intelligence procedures, managers endeavour to measure organisational performance based on metrics that reflect the performance of organisational activities. Generally, these activities are performed by different departments. In this thesis we look for performance metrics that are influenced by social media data.
- *Social media data*
  Social media data can be quantitative or qualitative in nature. Examples of quantitative social media data are the number of likes, views or shares of a certain page, the number of followers, friends or retweets through the course of time. Qualitative social media data contains the text of the posts. In this thesis, we investigate how social media data can be used for business intelligence.
- *Business intelligence*
  Business intelligence is a process in which information is derived from data to support decision making. The acquired information is required to measure organisational performance, at which managers can base their decisions. Information may for example relate to trends in the level of inventory of a certain product, or the amount of sales in a certain period.
- *Firm contexts*
  Though there are various ways to define a firm's context, we describe the context of a firm based on two dimensions in this thesis: (i) industry and (ii) relation with end-consumers. We employ this definition of context in this thesis because we are particularly interested in the variations of the applicability of social business intelligence on these two dimensions. Next, a generic classification of a firm's context on these two dimensions allows the conclusions of the research to be applicable at a broad range of firms.

  i. *Industry*
  Organisations can be classified in industries. All organisations in the same industry deliver similar products / services. We apply CBS' (2012) classification to position firms in certain industries. Examples of industries are the telecommunications industry, or the financial industry.

  ii. *Relation with end-users*
  Each organisation has different customers. Generally, a distinction between Business-To-Business ("B2B") and Business-To-Consumer ("B2C") is made to described the relation with an organisation's customer. In B2C relations, the end-user is part of the relation.

1-4 Research Questions

From the research objective, the following main research question is formulated:

> *How can firms use social media data for business intelligence, taking into account the firm's specific industry and relationship with end-users?*

In order to describe the domain of this thesis, the first sub question describes the current state of social media, the role of business intelligence in firms and the developments towards social business intelligence. Therefore, the first sub question is formulated as:

1. *What is the current state of social media in relation with business intelligence?*
   (a) *What are social media?*
   (b) *How are social media generally applied within firms?*
   (c) *How is business intelligence generally applied within firms?*
   (d) *How are key-performance indicators established within firms?*
   (e) *How can key-performance indicators be categorised?*
   (f) *What is social business intelligence?*

The main research objective contains a component in which we reveal in which contexts – i.e. for which *industries* and for which *customer relation type* – firms are able to acquire social media data, and in which not. This objective follows from the fact that firms are dependent on the users of social media whether or not social media data is available. Therefore, the second sub question investigates for which firms social media posts are available, and to what subjects the posts are related. The subjects of social media posts are consequently used to assign social media posts to the KPIs of a firm. The composition of the second sub question is twofold, sub questions 2(a) and 2(b) are quantitative in nature and provide insight in the *volume* of social media posts. On the other hand, 2(c) and 2(d) are qualitative in nature and provide insight in the *content* of the social media posts related to firms. The second sub question is formulated as:

2. *In which firm contexts*[2] *are firms able to acquire social media data for business intelligence?*
   (a) *How does the volume of social media posts related to firms vary between different industries?*
   (b) *How does the volume of social media posts related to firms vary between different relations with end-users?*
   (c) *How do subjects of social media posts related to firms vary between different industries?*
   (d) *How do subjects of social media posts related to firms vary between different relations with end-users?*

Secondly, the research objective contains a component in which we describe how a firm can acquire and process social media data for business intelligence purposes. The third sub question focuses on the development of a procedure to process social media data so that it can be joined up in business intelligence processes. A key requirement of this process is that it should fit within existing business intelligence activities. Therefore, 1(c) investigates how business intelligence is generally applied in organisations, and will result in requirements for a procedure in which social media data is applied for business intelligence. As discussed, social media data differs from data that is generally processed in BI systems. Question 3(a) discusses the potential problems and pitfalls when processing social media data. Consequently, 3(b) provides solutions for these problems. In 3(c), we determine how social media data can be linked to KPIs. Finally, 3(d) describes how a firm can process social media data while following the generally applied BI approach. The third sub question is defined as:

3. *Which processes are required to incorporate social media data into general business intelligence frameworks?*
   (a) *What problems arise when applying social media data for business intelligence?*
   (b) *How can the problems discovered in 3(a) be tackled?*
   (c) *How can social media data be linked to key-performance indicators?*
   (d) *How can social media data be processed in accordance with general business intelligence systems?*

1-5 Coherence of Research Questions

Each research question delivers information that is required to answer another question. The coherence of the research questions is presented in figure 1-1. The arrows represent the output of a research question which, in turn, serve as input to answer an other research question.

[2] In this thesis, we define a firm context based on the firm's *industry* and *customer relation type*.

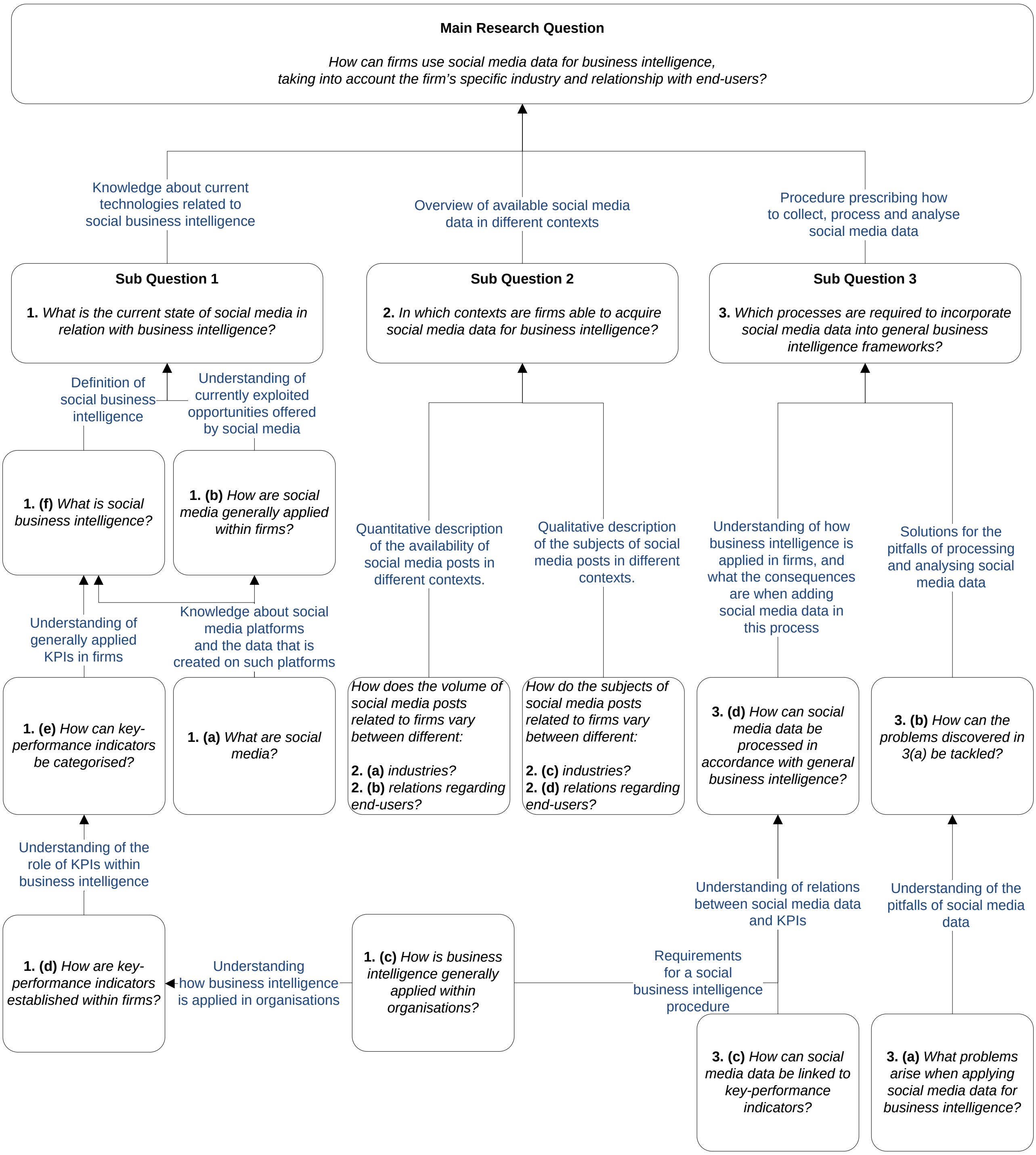


Figure 1-1: Coherence of Research Questions

1-6 Research Method

This section describes the type of research (section 1-6-1), the research method (section 1-6-2) and the approach of the research (section 1-6-3).

1-6-1 Exploratory Research

Exploratory research is conducted for a problem that has not been clearly defined. It relies on reviewing literature and/or data. Often, the results of exploratory research are not usually useful for decision-making by themselves, but they can provide significant insight into a given situation. The goal is to learn "what is going on there?", and to investigate social phenomena without explicit expectations. Mainly, the purposes of exploratory research are exploratory, descriptive and explanatory in nature. This thesis researches an area that is relatively unexplored, and of which the functioning is not clearly documented in theories and frameworks. Therefore, this thesis can be positioned under exploratory research.

1-6-2 Description of Research Methods

The research questions formulated in section 1-4 individually require different research methods in order to be answered. The research consists of a mix of literature studies, consulting experts, and content analysis on the acquired data. All methods and the corresponding requirements for data and other resources are discussed in this section. Table 1-1 schematically lists the corresponding research method for each research question.

Table 1-1: Research Questions versus research Methods

	Literature review	Content analysis	Consulting BI experts
1. What is the current state of social media in relation with business intelligence?			
(a) What are social media?	■		
(b) How are social media generally applied within firms?	■		
(c) How is business intelligence generally applied within firms?	■		
(d) How are key-performance indicators established within firms?	■		
(e) How can key-performance indicators by categorised?	■		
(f) What is social business intelligence?	■		
2. In which contexts are firms able to acquire social media data for business intelligence?			
(a) How does the volume of social media posts related to firms vary between different industries?		■	
(b) How does the volume of social media posts related to firms vary between different relations with end-users?		■	
(c) How do subjects of social media posts related to firms vary between different industries?		■	
(d) How do subjects of social media posts related to firms vary between different relations with end-users?		■	
3. Which processes are required to incorporate social media data into general business intelligence frameworks?			
(a) What problems arise when using social media data for business intelligence?		■	
(b) How can the problems discovered in 3(a) be tackled?		■	
(c) How can social media data be linked to key-performance indicators?		■	
(d) How can social media data be processed in accordance with general business intelligence systems?		■	■

Literature Review

Scientific articles are studied, mainly in the *Journal of Electronic Markets*, *Journal of Information Systems Management*, *Journal of Business Research*, *Journal of New Media & Society*, *Business Horizons*, *Journal of*

Strategic Information Systems and the *Journal of Computer-Mediated Communication*. The literature review was supported by books in the related research context. In addition, reports and white papers by acknowledged consulting firms in the field of information technology have been studied. The novel character of social media and social business intelligence makes it that especially in these reports social business intelligence is mentioned, whereas this term is less visible in the scientific area. These reports often contain examples from innovations and practical experiences. A such, a variate overview will be presented about related research and theories to this thesis.

Consulting Business Intelligence Experts

Firstly, interviewing experts contributes to an understanding of the actual situation of business intelligence in organisations and the potential role of social media in this field. This allows to scope the research in a topic that is actual and relevant. Secondly, a part of the research will describe how business intelligence is applied in organisations. Whereas this is mainly investigated using literature in the field of business intelligence, BI experts can validate the findings. Thirdly, a procedure prescribing how to execute social business intelligence will be develop. Such a procedure is required to be applicable in organisations as an integral part of the existing – regular – BI process.

Content Analysis

Content analysis is appropriate for this research since it offers a systematic method to compare content for a large sample of data. Content analysis is a research technique that can be used to identify what people are sharing on social media. The research technique is described by Stephens (2012) as an "in-depth look at recorded information" and as "a means of analysing texts" by Bos and Tarnai (1999). The sources of these texts can be various, for example newspapers, articles, web sites, or – as in this research – social media posts. Neuendorf (2002) defines content analysis as a "systematic, objective, quantitative analysis of message characteristics". As discussed, this thesis purposes to analyse the characteristics of social media posts, and link these posts to organisational functions. Krippendorff (2004) states that a "content analysis entails a systematic reading of a body of texts", and argues that every content analysis requires the following six questions to be considered:

1. Which data are analysed?
2. How are they defined?
3. What is the population from which they are drawn?
4. What is the context relative to which the data are analysed?
5. What are the boundaries of the analysis?
6. What is the target of the inferences?

Bos and Tarnai (1999) provide a procedure for analysing content, which is schematically shown in figure 1-2. In the first step, the problem is formulated at the theoretical level, research questions are defined and the object of investigation is determined. Secondly, the unit of analysis is defined by establishing categories and determining the sample. The third step consists of pretesting the reliability of the data, and the validation of the categories that were established in step 2. Discovered deficiencies are consequently renovated. In the fourth step of the content analysis procedure the data is collected and analysed. Finally, the results are interpreted and discussed on the basis of the problem.

It is the stepwise approach of Bos and Tarnai (1999) that is applied on the content analysis of this thesis. We will retrieve user-generated content from various social media platforms, store it into a database, and consequently analyse the collected posts. By analysing the social media content, it is possible to classify the nature of the content into categories, and find differences between posts related to different organisations.

1-6-3 Data Collection and Research Approach

Figure 1-3 illustrates the sequence and the links of the research steps in a schematic manner. A sample consisting of several organisations across different industries and with different customer relation types will be established. The selection of the organisations forms the point of departure for the collection of social media data. The

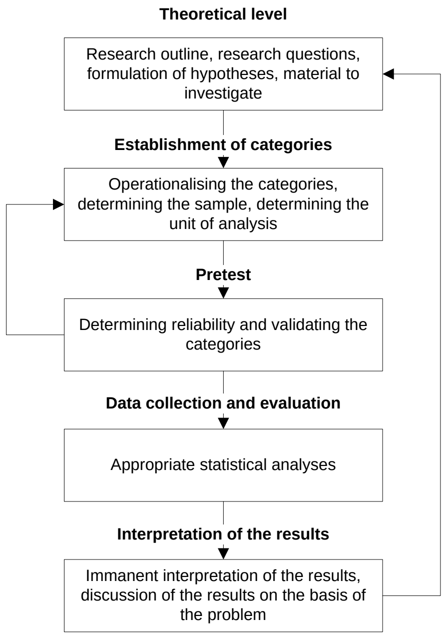


Figure 1-2: A procedure for analysing content (Bos & Tarnai, 1999).

content analysis requires that the social media data is available in a database. Therefore, social media posts need to be loaded from the web into our database. This process is called scraping. The selection of the data will be executed based on keywords corresponding to the selected organisations.

Scraping content from social media platforms results in unstructured data. In addition, the data is expected to be polluted by e.g. spam or by users who apply nicknames related to the search terms used to scrape the content. Therefore, the data needs to be cleaned before commencing the analysis. Once the spam and irrelevant posts are removed from the dataset, the content analysis can start. In this analysis, social media posts are classified in relation to KPI categories based on the subject of the posts. Once the content analysis has been performed for the firms, it is possible to identify differences between the subjects of social media posts across industries and different positions regarding end-users. Consequently, we can draw conclusions on the applicability of social media for business intelligence purposes.

The third research question relates to *how* organisations should execute social business intelligence. For that reason, a procedure prescribing how to execute social business intelligence will be designed. However, not before the requirements of a social business intelligence systems are clear, the framework can be designed. The framework is verified by (i) BI experts and (ii) the fit in the system that is currently executed in general BI systems. Finally, conclusions are drawn regarding the applicability of social business intelligence in firms.

1-7 Project Scope

Business intelligence and social media are broad concepts. In order to describe the focus of the proposed research, this section describes the scope of the research. Firstly, the research is scoped by a focus on a particular process of business intelligence; registering and processing. Next, the research analyses social media activities on a set of platforms, while others are excluded. Finally, some firms are part of our analysis while other are not.

Registering and Processing One possible way to represent BI, is through a cycle. Though many of these cycles exist in the literature, they do not differ much from each other (Pirttimäki & Hannula, 2003). Van Beek (2006) describes BI as a cycle of registering, processing, and reacting on gathered data. Figure 1-4 highlights the focus of this thesis. The gathering of data, ‘getting the data in’, is the most challenging aspect of BI, requiring about 80% of the time and effort (Watson & Wixom, 2007). The fundamental scope of the proposed research will be on this part; the gathering and registering of unstructured data generated on social media, and is highlighted in figure 1-4. One of the core activities related to business intelligence, is the formulation of key performance indicators. Not before these metrics are defined, the registering of data can commence. Therefore, key performance indicators take a central role in this research.

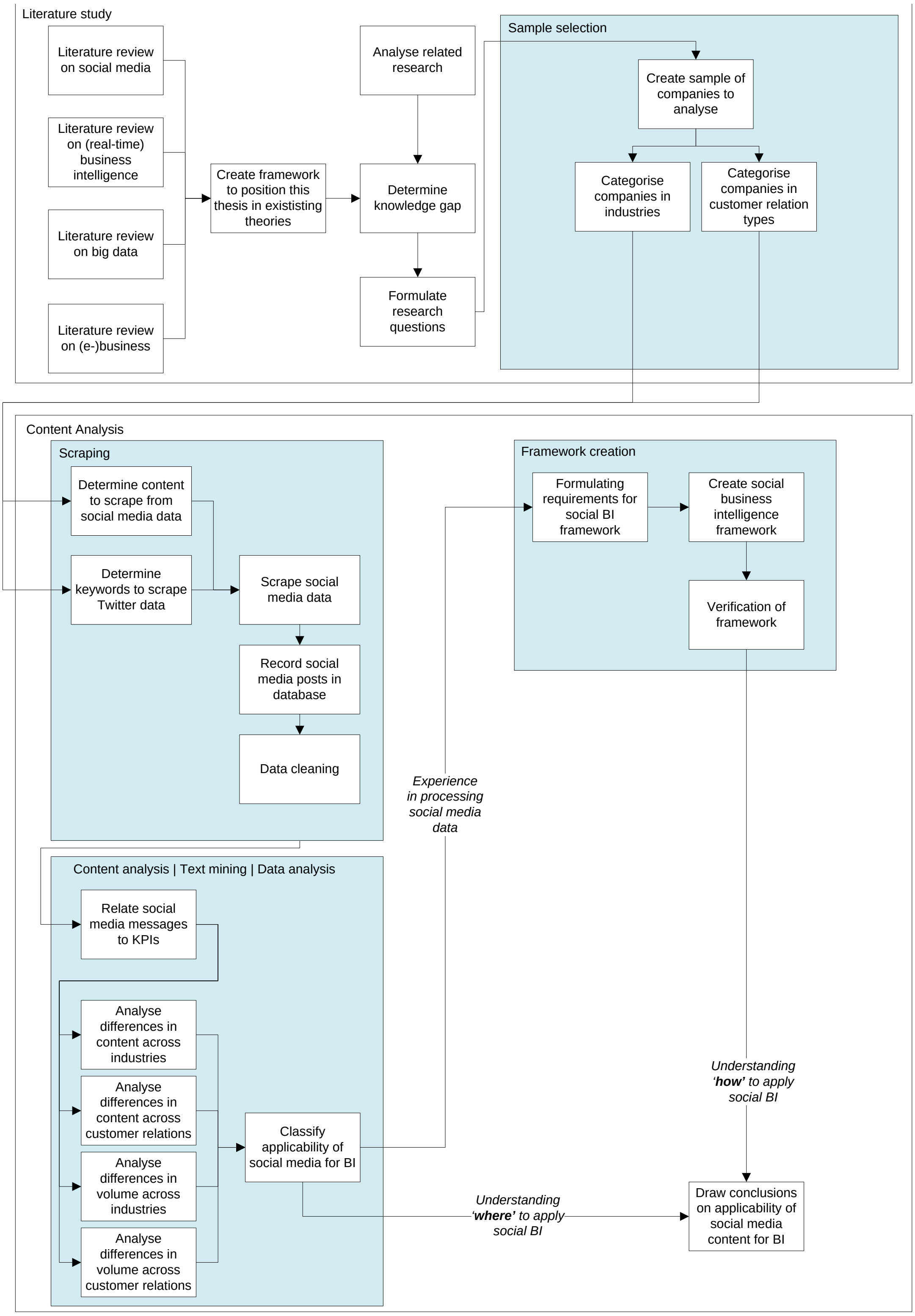


Figure 1-3: Research Approach

Social Media Platforms Many social media platforms are available, and the range of social media platforms is vast and growing (A. N. Smith et al., 2012). These platforms differ in scope, functionality and in culture (Boyd & Ellison, 2007). “Some sites are for general masses, like Twitter, Hi5 and Facebook. Other sites, like LinkedIn, are more focused on professional networks. Media sharing sites such as MySpace, YouTube, and Flickr concentrate on shared videos and photos” (Kietzmann et al., 2011). In addition, there also exist platforms that are explicitly not purposed to be publicly accessible. An example of such a platform is Yammer, which is used by

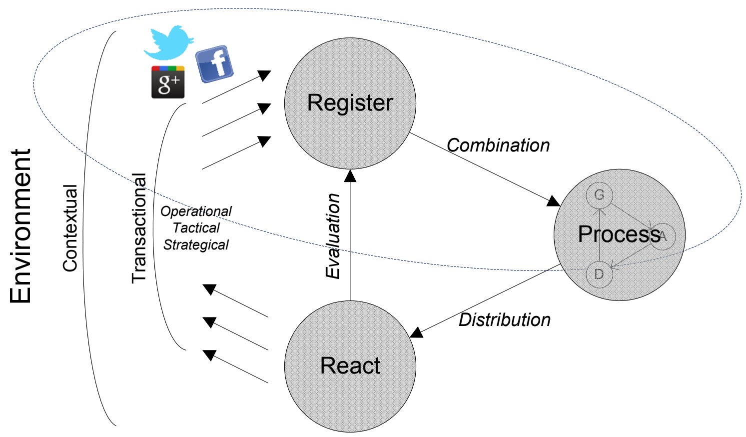


Figure 1-4: Thesis Scope, visualised in the BI cycle (van Beek, 2006)

organisations for internal communication. The focus of this thesis however is on publicly accessible platforms, since the key purpose is to investigate what kind of information firms can derive from publicly accessible social media. When investigating the opportunities of social media data for business intelligence purposes it is valuable to collect data from a great variety of platforms, so that possible differences in the nature of the content can be identified. Our analysis includes 25 platform types that are monitored, among them Facebook's public pages, Twitter, Google+'s public pages, Identi.ca, YouTube, Flickr, Vimeo, Picasa, Wordpress based blogs, Blogger, Typepad, RSS enabled blogs, Yahoo! Answers and Newssites. The different platforms are monitored, implying that each time a post is generated containing the predetermined keywords (e.g. 'Albert Heijn', or 'Heineken'), the post is extracted and saved into a database containing all gathered social media posts. Anno 2012, these platforms are the most popular social media platforms in the Western World. However, there are many other social network sites in the world. E.g Sina Weibo (the Chinese counterpart of Twitter), Qzone (China), Habbo, Badoo (Latin America) and many other platforms are not part of our analysis. The selected social media platforms – as well as the firms – are active in Europe.

Selection of Firms 18 firms have been selected for the analysis, that are active in different industries. The starting point of the sample selection has been the list of firms that are part of the Amsterdam Exchange Index ("AEX"). The main reason for this selection criterion is the fact that these organisations are stock listed, and hence publicise annual reports containing information about strategic initiatives, financial figures, etc. In case the analysis shows inter sector differences – e.g. between two comparable financial institutions – the annual reports may provide company specific information (e.g. amount of employees, attitude towards social media, etc.) clarifying these differences. Whenever a sample containing privately owned companies would have been selected, access to additional information would be limited. In addition, organisations listed in the AEX are generally well-established, visible to the public and regularly subject to news articles. It is therefore expected that these firms are subject of discussion on social media. The sample is further elaborated in section 4-2-2 (page 45).

1-8 Literature Review

In the following section research that relates to this thesis is presented. The literature related to the topic of this thesis has been found using search queries (SOCIAL MEDIA), (SOCIAL BUSINESS INTELLIGENCE), (SOCIAL MEDIA DATA), (SOCIAL MEDIA) AND (DATA EXTRACTION), (SOCIAL MEDIA) AND (CRAWLING), (WEB 2.0) AND (CRAWLING), (TWITTER) AND (BUSINESS INTELLIGENCE) and (WEB 2.0) AND (BUSINESS INTELLIGENCE) in the scientific literature databases ScienceDirect, JStore and the search engine Google Scholar. Existing research related to this thesis have been found in the scientific journals of *Electronic Markets*, *Computer Science*, *Journal of the American Society for Information Science*, *Public Relations*, *Journal of Marketing*, *Public Relations Review*, *Expert Systems with Applications* and the *Journal of Interactive Marketing*. Though not all of these studies are explicitly related to business intelligence, the central theme is the extraction of information from social media sites. We do not limit our review of related work to *one* social media platform. Instead, the presented research consists of a mix in which Twitter, Facebook, Blogs, and Questioning and Answering sites served as the data source.

Jansen, Zhang, Sobel, and Chowdury (2009) analysed 150,000 tweets containing branding comments, sentiments, and opinions. The researchers analysed the content of the tweets, and found that 19% of microblogs' posts contain a mention of a brand. Of these branding microblogs, nearly 20% contained some expression of brand sentiments. Of these, more than 50% were positive and 33% were critical of the company or product. The research concludes that microblogging is an online tool for customer word of mouth communications, and is especially suited for brand management activities.

Zhang et al. (2011) – in their quest to uncover the Twitter community dynamics – studied the "influences of business engagement in online word-of-mouth communication" and investigated "the trajectories of a business' online word-of-mouth message diffusion in the Twitter community". They studied nine-brands on Twitter, and concluded that "business engagement on Twitter enhances consumers' engagement with online word-of-mouth communication". Therefore, the authors argue that "businesses must go beyond simply being aware of or taking into consideration electronic word-of-mouth messages and instead must engage in the communication process as both initiators and active participants. Next, Zhang et al. (2011) found that "retweeting, as an explicit way to show consumers' response to business engagement, only reaches consumers with a second-degree relationship to the brand" and that the "life cycle of a tweet is generally 1.5 to 4 hours at most".

McCorkindale (2010) investigated – based on a content analysis – how the Fortune 50 companies used Facebook. The research studied how many fans an organisation had, what organisational information was included, if they used photos and videos, if they used discussion boards, whether they generated feedback, etc. She found that companies are using Facebook extensively, but that most companies were not using the site to "disseminate news and information about the organisation". Next, the research indicated that the companies should focus more on "relationship-building strategies in order to encourage users to revisit the sites". The content analysis of McCorkindale (2010) revealed that there are several reasons why people post on Facebook pages. "Some were current employees who identified where they worked and for how long, while some were former employees reconnecting with past coworkers. Headhunters posted jobs at competing corporations on the wall, and job seekers posted they were looking for employment. Customers having product problems, especially in the technology field, would post their issues on the wall hoping to find solutions. Journalists also posted on pages requesting interviews".

Agichtein, Castillo, Donato, Gionis, and Mishne (2008) argue that the "quality of user-generated content varies drastically from excellent to abuse and spam", and that the "task of identifying high-quality content in sites based on user contributions – social media sites – becomes increasingly important". Therefore, the authors developed a method to exploit "community feedback to automatically identify high quality content". Agichtein et al. (2008) applied their model on a popular questioning and answering site (Yahoo! Answers). The system of Agichtein et al. (2008) models all user relations, and applies the user ratings on the individual answers. As such, the system determines high-quality content based on the ratings that users assigned to the content.

Guo, Zhang, Tan, and Guo (2012) developed a system that detects popular topics on Twitter. According to the authors, the key technology in mining web text includes the modules "text classification, clustering, topic detection and tracking, opinion tendency identification, and multi-document automatic summarisation". Guo et al. (2012) argue that popular topic detection systems should entail these five modules. However, the nature of tweets – "very short, sparse and spreading rapidly" – is different from regular web text. Therefore, Guo et al. (2012) propose a more "flexible and practical approach based on frequent pattern mining".

Kozinets et al. (2010) qualitatively studied 83 blogs in order to understand how marketing departments influence consumer-to-consumer communications. The authors distinguished the strategies of the marketeers into four categories – evaluation, embracing, endorsement, and explanation.

Araujo and Neijens (2012) researched how top global brands participated in social network sites by investigating which factors influence the presence and level of engagement of these brands on social network sites. The authors reviewed the corporate websites of 129 brands in different markets, targeting different ages of audience, different home markets, different web operations and in different countries. Consequently, the authors determined whether or not the companies refer to their presence at social network sites. The research found that social network site "presence was significantly higher for information technology and telecommunication brands" (Araujo & Neijens, 2012), implying that the presence of firms on social media differs between firms in different industries. Furthermore, Araujo and Neijens (2012) found that "brands targeting younger audiences also engage at higher levels than brands targeting generic audiences" and that the "country in which the brand operates plays a significant role in a brand's likelihood of adopting social network sites". The findings of Araujo and

Neijens (2012) indicate that the applicability of social media for business purposes differs between firms, which support the basis of this thesis.

Dey and Haque (2008) acknowledge that “the data generated from online communication acts as potential gold mines for discovering knowledge”. However, as Dey and Haque (2008) illustrate, “the quality of texts generated from online sources can be extremely poor and noisy” because the “text data typically comprises spelling errors, ad-hoc abbreviations and improper casing, incorrect punctuation and malformed sentences”. It is therefore that text mining techniques based on “pure linguistic strategies fail to extract information from noisy text”. According to Dey and Haque (2008), “statistical techniques on the other hand which though not as successful as the linguistic methods, are more suited to extract information from noisy text. However, lack of appropriate training data often poses as a bottleneck”. Dey and Haque (2008) conclude that – when processing unstructured social media data – “domain related training sets” are required to the clean the text before the text can be processed by Natural Language Processing Tools. With such domain related training sets, the word ‘small’ can be classified as either positive or negative, depending on its context.

Lovejoy et al. (2012) performed a content analysis of the tweets related to 73 non-profit organisations to examine “how these organisations use Twitter to engage stakeholders”. Within that analysis, the researchers looked at “the organisations’ utilisation of tweet frequency, following behaviour, hyperlinks, hashtags, public messages, retweets, and multimedia files”. Lovejoy et al. (2012) conclude that non-profit organisations use social media as a “one-way communication channel”, and not as a platform for “conversation and community building”.

Lee (2012) acknowledge that the “contents of microblogs preserve valuable information”. In his study, Lee (2012) focused on real-world offline events, and the information that was generated on social media sites related to those events. With his system, it is possible to detect real-world events through the content on social media sites. Also Lee (2012) argues that the challenge of automatically classifying social media posts is the informal structure of the text.

Dhar & Chang’s (2009) research is one of the few that studied the relation between social media activity and organisational performance. More specifically, they employed social media data to predict sales in the music industry. Using linear and nonlinear regression, Dhar and Chang (2009) found that “(a) the volume of blog posts about an album is positively correlated with future sales, (b) greater increases in an artist’s Myspace friends week over week have a weaker correlation to higher future sales, (c) traditional factors are still relevant – albums released by major labels and albums with a number of reviews from mainstream sources also tended to have higher future sales”.

Tirunillai and Tellis (2012) studied the relationship between user-generated content and stock market performance of the firm. The authors found that “of all the metrics of UGC, volume of chatter has the strongest positive effect on abnormal returns and trading volume. Whereas negative UGC has a significant negative effect on abnormal returns, positive UGC has no significant effects on these metrics. The volume of chatter and negative chatter have a significant effect on trading volume”. In addition, Tirunillai and Tellis (2012) found that “an increase in off-line advertising significantly increases the volume of chatter and decreases negative chatter”.

From the literature review, we can conclude that there is scientific attention in the research field of social media and the relation with organisational performance. However, no research has been found that investigates the applicability of social media for organisations in (i) different industries and with (ii) different customer relation types. Next, though some studies individual tackle difficulties that are inherent to the usage of social media data, no research has been found that integrally describes how social media should be collected and processed within a firm. As illustrated, the opportunities for social media are beneficial on many aspects. However, managers are also reluctant to allocate budget to social media activities (Weinberg & Pehlivan, 2011) and incorporate social media data in the firm’s BI process, because they do not fully understand what social media intelligence may bring to the firm. In addition, it is unclear which type of firms are subject of discussion on social media and – if they are – how a firm should collect and process these data so that it adds value to the firm.

1-9 Scientific Relevance

The proposed research touches the world of e-business, which implies “the transformation of key business processes through the use of internet technologies” (Chaffey, 2009). The monitoring of opinions, customer thoughts, etc. by electronic means – for instance by social media sites – can be positioned under the denominator ‘e-business’. Many literature exists in which the world of e-business is described. This research contributes to existing models and theories by positioning social media content as an external factor in these theories.

Next, science is built on data. The more data is available to scientists, the "greater the level of transparency and reproducibility and hence the more efficient the scientific process becomes" (Molloy, 2011). Historically, scientific data has not been openly available. In recent years, several scientists advocate the application of open data. The proposed research will be based on publicly accessible data – coming from social media – and will hence contribute to understanding the opportunities and threats of applying public data for scientific purposes.

The literature contains many definitions of business intelligence, and provides theories describing how BI processes internal as well as external data. Data from social media can be positioned under external factors. This thesis positions explicitly adds social media data into the existing theories of business intelligence.

Next, the research will be executed based on the research methodology of content analysis. Though this method is yet widely applied in many research areas, the fact that the source of the content in this research is social media, makes it new. The lessons learned in this research from applying a content analysis on social media data contribute to the research methodology of content analysis.

Finally, the literature of customer relationship management ("CRM") describes how organisations interact with their customers. Recent literature also includes social media solutions into CRM activities. This research reveals how user-generated content varies between industries, departments and the position regarding consumers. As such, the conclusions of this thesis contribute to the applicability of CRM through means of social media.

1-10 Societal Relevance

Many executives avoid or ignore social media because they do not understand what it is and how to engage with it and learn, though they sense that social media is – and will remain – an important "fabric of commerce" (Kietzmann et al., 2011; Weinberg & Pehlivan, 2011). This thesis contributes to a further understanding of social media, and to leverage the opportunities of applying the valuable content on these platforms for organisational efforts. The social media phenomenon is relatively fresh, Facebook was launched in 2004, Twitter in 2006. Because of the novelty, the opportunities for organisations' social BI activities are rather unexplored. This research also contributes to an understanding for firms whether or not social media data can be applied for business intelligence purposes in which context. In addition, we expect that legacy BI vendors – such as SAP, Oracle and IBM – are soon asked by their clients to add a social media component to their BI suite. For these organisations social business intelligence is also a new phenomenon, and social media data can not be directly applied to their existing systems (Reinhold & Alt, 2011). The conclusions of this thesis support BI vendors in the development of social media components within their product range.

1-11 Project Deliverable

This research will reveal two central questions that describe (i) *where* and (ii) *how* firms can derive information from social media data for their decision-making process. Therefore, the project has two deliverables.

Deliverable 1: Where? This deliverable specifies in which contexts a firm can implement social business intelligence. For reasons explained in section 1-2, it is expected that a firm's contexts and aspects determine the applicability of social media data for BI purposes. This deliverable allows organisations to determine whether or not they are suited for the applicability of social media for business intelligence.

Deliverable 2: How? This deliverable consists of a procedure that prescribes how firms can execute a social business intelligence system. The procedure can be considered as a document prescribing how an organisation can execute social business intelligence, and which technical and institutional elements are involved in a social BI system.

Chapter 2

Conceptual Frame of Research

In this chapter, we define and illustrate what business intelligence ("BI") is, how BI is applied in firms, what the most important elements are and how BI is regarded in this thesis. In a later stadium of this research a procedure for processing social media data to support business intelligence will be developed. Such a procedure is required to fit in the current method that firms adhere to in executing BI. Therefore, an understanding of business intelligence within firms is essential. This chapter provides the knowledge of business intelligence that is required when we develop a procedure to collect and process social media data for BI purposes in a later stadium.

Section 2-1 starts by a description of the various definitions of BI, and consequently formulates the perspective on BI that is adhered to throughout this research. In section 2-2 the elements relating to the determination of 'what to measure?' are discussed, a key activity in business intelligence. Section 2-2 explains the relation between a firm's strategy and the firm's performance metrics. As we will see, performance metrics take a central role in BI. In section 2-3, the processing of data is described. Finally, section 2-4 concludes this chapter by a description of how business intelligence is applied within firms.

2-1 Business Intelligence Perspectives

Business intelligence is a process in which data is translated into information that is required for managerial decision-making. The literature contains many definitions of business intelligence. Elbashir et al. (2008) state that "business intelligence systems provide the ability to analyse business information in order to support and improve managerial decision making across a broad range of business activities". Van Beek (2006) defines business intelligence as "a continuous process that helps organisations gathering and registering data, analysing it and consequently applying the resulting information and knowledge in decision-making processes to improve organisational performance". Rouibah and Ould-ali (2002) describe business intelligence as "a strategic approach for systematically targeting, tracking, communicating and transforming relevant weak signs into actionable information on which strategic decision-making is based". Lonnqvist and Pirttimaki (2006) define BI as "an organised and systematic process by which organisations acquire, analyse, and disseminate information from both internal and external information sources significant for their business activities and for decision making". Although these definitions vary slightly from each other, the common aspect is that business intelligence is perceived as a process that translates data into interpretable information that supports managerial decision-making.

Van Beek (2006) visualises business intelligence ("BI") as a cycle, consisting of three main processes (figure 2-1). For the remainder of this thesis, we follow van Beek's – loosely defined – perspective on business intelligence because it captures the various definitions of BI found in the literature. The three main processes – register, process and react – are discussed in the following paragraphs.

Register The BI cycle starts with carefully listening – *registering* – to the environment. Within the environment, a distinction is made between contextual and transactional environments. The contextual environment consists of aspects that (may) have an effect on the organisation. The transactional environment

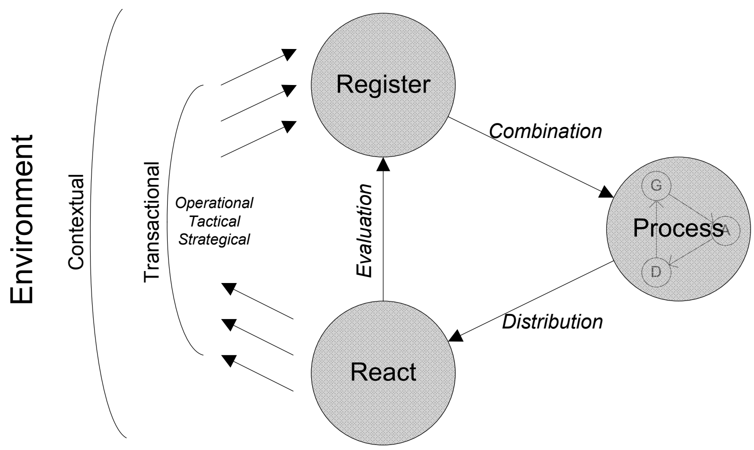


Figure 2-1: Business Intelligence Cycle (van Beek, 2006).

consists on the one hand of actors that have a direct relation with the company, like customers, suppliers, employees and competitors. On the other hand, the transactional environment is made up of institutions affecting the organisation, like new policies or legislation. BI registers signals arising from this environment.

Process Consequently, when data (in whatever format) is registered, it is required to be *processed*. Processing the gathered data will reveal trends and provide valuable information. Van Beek (2006) positions a 'small BI cycle' within this process; data is gathered, analysed, and distributed to the right organisational departments. This part of the BI cycle will be further described in section 2-3.

React Following the results provided by the processing of the data, the company can *react*. Van Beek (2006) argues that a company can react on three levels; operational, tactical or strategical. Consequently, the environment evaluates the companies' changes in interactions, resulting in new signals for the firm's BI cycle.

2-2 Registering the Right Indicators

As illustrated in the previous section, the business intelligence process starts with registering. But, what is it that a firm has to register? Not unless a firm has clearly set what will be registered, the BI cycle can commence. The determination of 'what to measure' is a process on itself, which is described in this section. In the following sections, the necessary steps to determine what a firm has to register are described. Section 2-2-1 describes that firms formulate – or, should formulate – their performance metrics based on their strategy. In section 2-2-2, two widely accepted frameworks that support firms in the formulation of performance metrics are discussed. Section 2-2-3 discusses a framework that illustrates how a firm should design and implement a system for measuring organisational performance. Next, section 2-2-4 describes the various types of performance measures that are applied by firms. Finally, section 2-2-5 describes ten commonly applied categories of key-performance indicators, which take a central role in the business intelligence process.

2-2-1 Strategy and Business Model

Firms align their business model with their strategy. The determination of 'what to measure' initially starts with the firm's strategy. A strategy consists of a mission, values, vision, goals, objectives and plans. The "mission and values define why the organisations exists, what it does, and its guiding principles. The vision combines an overarching purpose with an ideal, future-state competitive positioning. Goals are broad statements that embody what the organisation would like to achieve in three to five years, while objectives represent short-term goals of one to three years" (Eckerson, 2009).

Strategies are translated into business models (Bouwman, Faber, Haaker, Kijl, & de Reuver, 2008). There exists a variety of views on business models in the literature. Osterwalder and Pigneur (2010) state that "a business model describes the rationale of how an organisation creates, delivers, and captures value". Bouwman et al.

(2008) provide an all-embracing definition of a business model for service-oriented organisations by stating that "a business model is a blueprint for a service to be delivered, describing the service definition and the intended value for the target group, the sources of revenue, and providing an architecture for the service delivery, including a description of the resources required, and the organisational and financial arrangements between the involved business actors, including a description of their roles and the division of costs and revenues over the business actors". In other words, a business model describes what a firm does, why it does that, how it does that, with whom it does that and for whom it does that.

By aligning the business model with the firm's strategy, managers question themselves "why are we performing this activity?", and "what does it contribute to?". "It is only through consistency of action that strategies are realised" (Neely, Gregory, & Platts, 1995). Gates (2001) argues that applying value driver maps to analyse a company's vision and drivers of performance helps a company aligning its business model with its strategy. By aligning the business model with the strategy, a firm ensures that it performs those activities that contribute to the intended strategy. It is in this phase of the BI process where managers determine where in the organisation and by which activities value is added to the main objective of the organisation, and why and how the performance of these activities are to be measured. There are many frameworks developed to craft a business model that is based on the firm's strategy, at which – in the end – performance metrics can be formulated. These frameworks are the topic of the following section.

2-2-2 Frameworks Supporting the Formulation of Performance Indicators

Performance measurement is "the process of quantifying the efficiency and effectiveness of action" (Neely et al., 1995). To quantify actions – forming the firm's business model – indicators are required that represent these actions. Such indicators are called performance indicators. Performance indicators should be derived from a firm's strategy (R. S. Kaplan & Norton, 1992, 1993, 1996; Kennerley & Neely, 2003; Tsai & Cheng, 2012; Fortuin, 1988). The previous section revealed that a business models amongst others specifies which activities are required to deliver value for the customer of the firm. Hence, the business model prescribes which activities are to be executed. The activities are consequently required to be measured, so that managers can determine the organisation's performance in accordance with its strategy. These activities are measured by performance indicators, indicating the performance of the individual activities.

But how do we determine the right performance indicators? One of the first widely recognised frameworks (Neely, Bourne, & Kennerley, 2000) that help managers to decide 'what to measure' is the balanced scorecard, developed by R. S. Kaplan and Norton (1992). R. S. Kaplan and Norton (1992) provide a framework that supports managers in formulating performance indicators based on four perspectives. Because the balanced scorecard ensures that managers do not only focus on financial figures, it gives managers a "comprehensive view of the business", which is required in the competitive environment of the firm (R. S. Kaplan & Norton, 1992). Since the balanced scorecard links a company's strategy with concrete actions (R. S. Kaplan & Norton, 1996), it is considered as a tool helping managers to align the company's business model with its strategy. In response to the balanced scorecard, Neely, Adams, and Crowe (2001) developed a scorecard that adopts a multi-actor view in formulating performance metrics, and hence incorporates the perceptions of multiple stakeholders into the performance metrics formulation process. Both frameworks are discussed in the following sections.

The Balanced Scorecard

Because "you get what you measure" (Kennerley & Neely, 2003), R. S. Kaplan and Norton (1992) advocate that a firm should measure those metrics that contribute to the firm's strategy. In addition, R. S. Kaplan and Norton (1992) argue that managers should not only focus on financial figures, but also on other areas representing organisational performance. Therefore, R. S. Kaplan and Norton (1992) developed the "balanced scorecard". The balanced scorecard is not without reason called a 'balanced' scorecard. In essence, it stimulates managers to not only think in financial figures when measuring organisational performance, but also on other areas. The framework distinguishes organisational activities in the (i) customer-, (ii) internal business-, (iii) innovation and learning-, and (iv) financial perspective, which are discussed below:

- The *customer perspective* describes how customers view the firm, and ensures that customer's needs are fulfilled. R. S. Kaplan and Norton (1992) further categorise the customer's concerns into time, quality, performance and service, and cost. Hence, when a manager adopts the customer perspective in formulating performance metrics, he or she will formulate performance metrics that involve these categories.

- The *internal business perspective* describes where the particular firm must excel at and specifies what a "company must do internally to meet its customers' expectations" (R. S. Kaplan & Norton, 1992). It is in this perspective where managers "attempt to identify and measure their company's core competencies, the critical technologies needed to ensure continued market leadership. Companies should decide what processes and competencies they must excel at and specify measures for each" (R. S. Kaplan & Norton, 1992).

- The third perspective of the balanced scorecard – *innovation and learning* – ensures that the organisation continues to improve and create value. Due to competition, "the targets for success keep changing" (R. S. Kaplan & Norton, 1992). It is therefore that companies must design its organisation in a way that it can innovate. It is in this perspective where managers consider the development of new products, entering new markets, etc.

- The *financial perspective* ensures that the shareholders' needs are fulfilled. "Financial measures indicate whether the company's strategy, implementation, and execution are contributing to bottom-line improvement" (R. S. Kaplan & Norton, 1992).

The balanced scorecard set the scene for the development of a variety of other performance measurement frameworks at the beginning of the 1990s.

The Performance Prism

In response to the various types of scorecards that have been developed after Kaplan & Norton's (1992) *balanced scorecard*, Neely et al. (2001) developed a "second generation performance measurement framework" called the *performance prism*. According to Neely and Adams (2005), there were three fundamental reasons why the balanced scorecard was outdated, and why a new framework was required. Firstly, the balanced scorecard solely focuses on the needs of two groups of (internal) stakeholders; shareholders and customers. In today's business environment, firms can no longer consider only these two groups of stakeholders. For example, employees, environmental parties, labour unions, other communities, regulatory bodies, etc. have been fully denied in the balanced scorecard, while these groups truly influence a firm in practice. Second, an organisation's "strategy, processes, and capabilities have to be aligned and integrated with one another" (Neely & Adams, 2005), e.g. with the processes of the firm's suppliers. Third, firms "have to recognise that their relationships are reciprocal – stakeholders have to contribute to organisations as well as receive something from them" (Neely & Adams, 2005). Thus, the key innovation that is captured in the performance prism is the fact that it takes a comprehensive stakeholder orientation, i.e. a multi-actor perspective, whereas former frameworks (such as the balanced scorecard) adopt a mono-actor perspective. It is necessary to adopt a multi-actor perspective, since firms need to have "contributions from their stakeholders – usually capital and credit from investors, loyalty and profit from customers, ideas and skills from employees, materials and services from suppliers, and so on" (Neely & Adams, 2005).

The performance prism consists of five perspectives, whereas the balanced scorecard comprises of four. The central element in all these five performance prism perspectives is the stakeholder aspect.

- **Stakeholder Satisfaction**
  The *stakeholder satisfaction* perspective ensures that managers consider "who the firm's stakeholders are, what they do, and what they need". Whereas the balanced scorecard explicitly focuses on two groups of stakeholders, i.e. on customers through the customer perspective and on shareholders through the financial perspective, the performance prism does not specify stakeholder groups but rather allows for ambiguity. As such, the firm is stimulated to consider all stakeholder groups in its ecosystem, and specify what these groups want.

- **Strategies**
  The second facet of the performance prism focuses on *strategies*. With the identification of the needs in the *stakeholder satisfaction* perspective, strategies can be developed that fulfil the needs of the stakeholders. The key question in this facet is: "What strategies should the organisation adopt to ensure that the wants and the needs of its stakeholders are satisfied?" (Neely & Adams, 2005). Whereas the balanced scorecard method of formulating measures starts with the firm's strategy, Neely and Adams (2005) argue that strategies should be designed in accordance with the needs of the stakeholders.

- **Processes**
  In the third facet of the performance prism, the *processes* are defined that are required to fulfil the strategies defined in the second perspective. It is in this perspective where general business processes, such as product development, demand generation, demand fulfilment, planning, etc. are defined. Neely et al. (2001) stress the importance of measures that reflect the performance of the processes. Thus, it is in this perspective where organisations consider which measures are required to determine the organisational performance.

- **Capabilities**
  Within the fourth aspect of the performance prism, the *capabilities* required to execute the processes are identified. Capabilities consist of a combination of "people, practices, technology, infrastructure, etc.".

- **Stakeholder Contribution**
  Finally, the performance prism "recognises the fact that not only do organisations have to deliver value to their stakeholders, but also that organisations enter into a relationship with their stakeholders which should involve the stakeholders contributing to the organisation" (Neely et al., 2001). It is in the *stakeholder contribution* perspective where firms consider what they want from their stakeholders.

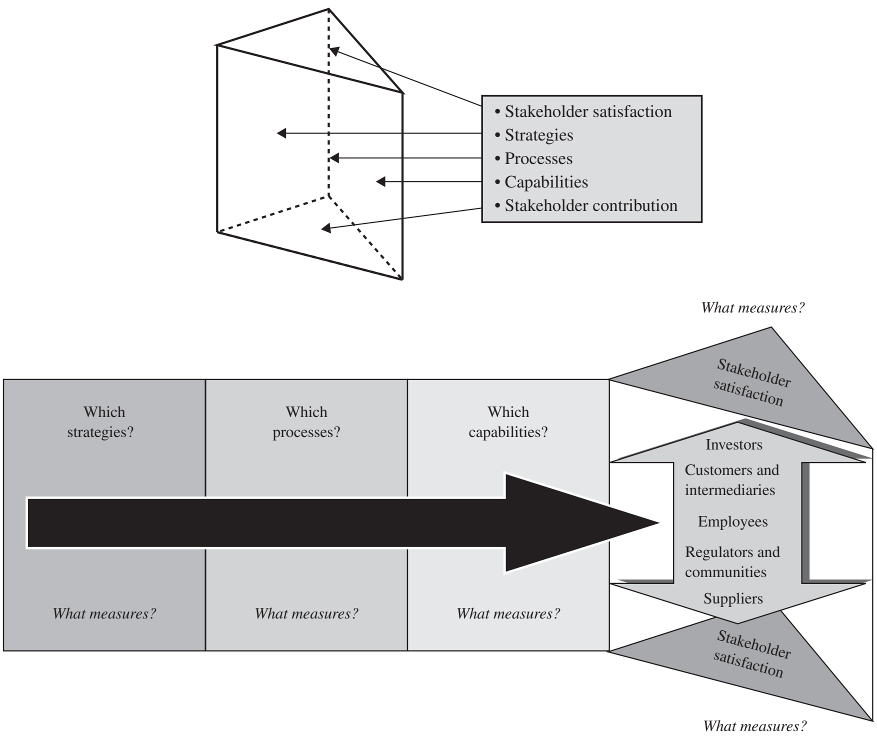


Figure 2-2: Five Facets of the Performance Prism. Adopted from Neely & Adams (2005).

Figure 2-2 presents the *performance prism*. The five areas of the prism represent one of the perspectives. The triangular – outside facing – surfaces represent the two stakeholder perspectives, which are unique for the performance prism. The three rectangular – inside facing – surfaces represent the strategy, processes and capabilities perspectives. The fact that stakeholders are an important element in the performance prism is also highlighted by the substance of the prism, which mentions groups of stakeholders. As argued by Neely and Adams (2005), the performance prism "has been deliberately designed to be highly flexible so that it can provide either a broad or a narrow focus". As a result, the perspectives of the performance may seem vague. The authors explicitly made the perspectives vague so that the framework is broadly applicable.

To conclude, a firm should register those metrics that reflect the performance of the activities that contribute to the firm's strategy. Since organisational activities are tailored to the firm's strategy, and metrics are derived from these activities, there exists a link between performance metrics and an organisation's strategy. Though the alignment of performance metrics and strategy may sound self-evident, many organisations struggle with strategic alignment: even at the healthiest companies, about 25% of the employees are unclear about their company's direction. KPMG (2009) argues that in many organisations, there is "no explicit linkage between the strategy and the information used to manage the business", implying that managers are measuring activities that do not contribute to the firm's strategy. Managing without or the wrong metrics "gives one the feeling of being lost with no hope", and leads to a "lack of management control" (R. Smith, 2006). Once the performance metrics are established and mutual differences in importance are assigned, the BI process can commence. The values of the performance indicators will then reveal the performance of the firm. As illustrated, R. S. Kaplan and Norton (1992) introduced the first framework that considered other metrics than solely financial figures. Next, Neely et al. (2001) developed – in response to the flaws relating to the mono-actor perspective of the *balanced scorecard* – a framework that considers the firm's stakeholders; the *performance prism*. The *performance prism* is deemed as a framework that is – given today's business landscape – better suited for performance metrics formulation than the *balanced scorecard*. Moreover, since the purpose of this thesis is to incorporate social media data – created by multiple actors – in management information for different fields and departments, a multi-actor perspective is required.

2-2-3 Performance Measurement System Design Process

Wisner and Fawcett (1991) – as quoted in Neely et al. (2000) – developed a "process for performance measurement system design". Figure 2-3 shows the nine-step process proposed by Wisner and Fawcett (1991), which clearly illustrates that performance metrics should be derived from a firm's strategy, and that performance indicators should be assigned to functional areas – performing individual activities – of the firm.

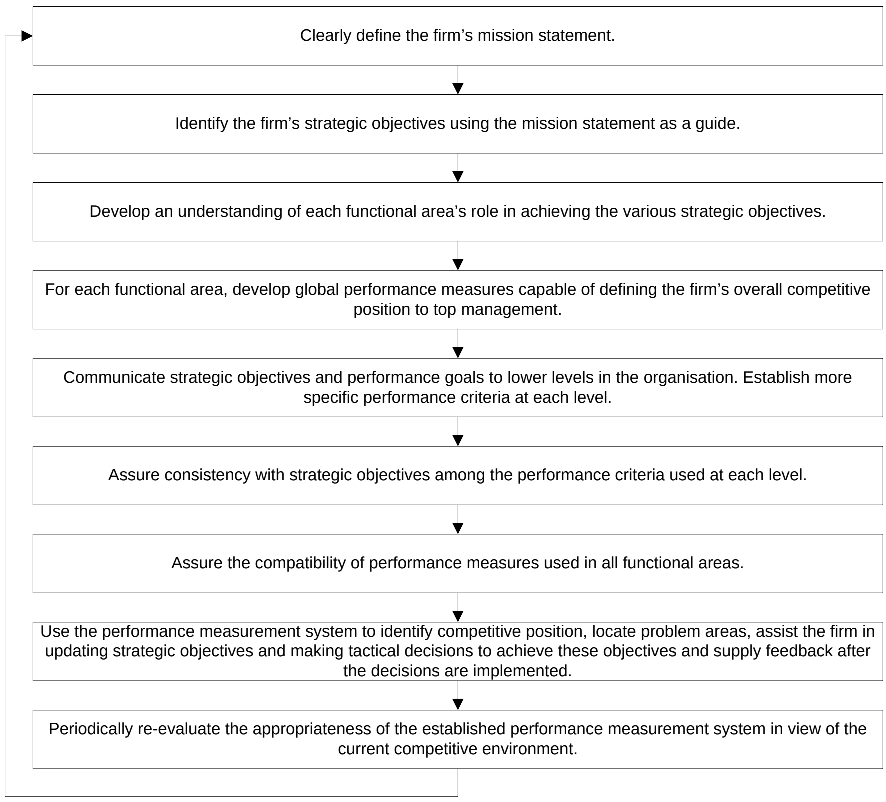


Figure 2-3: Performance measurement system design process (Wisner and Fawcett, 1991).

In a joint research, the business intelligence system vendors SAP, IBM, Corda and Pentaho examined how organisations formulate performance metrics. Also they found that managers determine performance metrics based on the firm's strategy. Next, they concluded that performance metrics should be "tailored to every individual and role in the organisation" (Eckerson, 2009). As a result, departments and individuals consequently understand how their activities contribute to the company's strategy, which is often stated in generic and vague terms. Consequently, employees will focus on those activities that are important, because "what's get measured, gets done" (Kennerley & Neely, 2003). This typical management quote illustrates the imposing consequences that descent from the determination of performance metrics, that is, "what's *not* get measured, gets *not* done". As illustrated by Eckerson (2009), "if the metrics do not accurately translate the company's strategy, the organisation will flounder". It is therefore that determining performance metrics is a critical activity of business intelligence.

2-2-4 Typology of Performance Indicators

Performance indicators are key elements in business intelligence, since they reflect the performance of the activities that contribute to the firm's strategy. In this section, we elaborate more about performance indicators and the types of metrics that exist.

Leading and Lagging Indicators There are two fundamental types of indicators; leading indicators and lagging indicators. *Leading* indicators lead to results, and are also referred to as '(value) drivers'. *Lagging* indicators are the results that measure the output of past activities, and are also known as 'outcomes' (R. Smith, 2006). Leading indicators are used to manage, while lagging indicators measure how well has been managed.

With leading indicators it is possible to respond directly when poor results are found. With lagging indicators, "we get value from knowing how well we performed but have little opportunity to immediately affect under-performance" (R. Smith, 2006). Hence, leading indicators are more powerful, and can be perceived as short-term indicators of an organisations' results. It is therefore that firms manage by leading indicators. Illustratively, table 2-1 lists some examples of leading and lagging indicators. It is noteworthy that among different departments and individuals in organisations there could exist pluriformity in the perception of the type of indicators, "one man's outcome measure can be another man's value driver" (Eckerson, 2009).

Table 2-1: Examples of leading and lagging indicators

Leading indicators	Lagging indicators
New sales today	Revenues
Planned rework today	Cost
Customer cases currently open	Capacity
Contracts in negotiation for Q2	Return on equity
Identified software bugs	Customer satisfaction
	Employee retention
	Margins
	Reliability
	Failures
	Downtime

Quantitative and Qualitative Indicators Another distinction between metrics is the difference between quantitative or qualitative based indicators. Quantitative indicators measure processes by counting, adding, averaging, etc. numbers. Examples of quantitative measures are inventories, number of orders, number of clients, delivery time of goods, sales, other financial figures, etc. In contrast with qualitative indicators, quantitative indicators are relatively easy to measure.

However, there are many other criteria to judge performance than solely on (financial) quantitative indicators (Neely et al., 2000; Eccles, 1991). Other metrics are qualitative in nature and require a proxy to be measured. An example of a qualitative measure is customer satisfaction. The measurement of customer satisfaction results in quantitative data, but is primarily based on subjective interpretation of customers' opinions. Customer satisfaction is therefore traditionally measured by surveys (Peterson & Wilson, 1992). "Traditionally, performance evaluation has depended to a great extent on financial indicators. However, given the current environmental uncertainties, financial indicators can no longer give a complete view of business operations" (Tsai & Cheng, 2012). It is therefore that qualitative measures are as much as important as quantitative measures. The trick is to identify the links between qualitative measures and financial measures. Firms can for instance conduct statistical analyses to correlate qualitative indicators with financial performance. Regression analysis can be applied to identify the key drivers that impact sales, profitability, etc. The performance prism framework allows for the incorporation of qualitative indicators next to quantitative indicators.

Key Performance Indicators To distinguish between performance indicators that are more important than others, some indicators are termed 'key-performance indicators' ("KPIs"). But what is it that makes a performance indicator 'key'? PWC (2007) argues that the performance indicators that are key to a firm are those that are used to manage the business. According to Tsai and Cheng (2012), KPIs "are the groundwork of the performance system which turns the strategic goals of a company into long-term objectives". The addition of the word 'key' to a performance indicator indicates that these metrics are assigned more attention than others. Thus, it are the KPIs that represent processes that are paramount for the success of a firm. Table 2-2 lists the elements that a key-performance indicator should fulfil, it should be specific, measurable, attainable, realistic and time-sensitive ("SMART") (Shahin & Ali Mahbod, 2007).

Table 2-2: Requirements of a key-performance indicator (Shahin & Ali Mahbod, 2007).

Requirement	Description
Specific	KPIs should be detailed and as specific as possible.
Measurable	A KPI should be measurable against a standard of performance and a standard of expectation.
Attainable	The goal of a KPI should not be out of reach. They should be reasonable and attainable.
Realistic	A goal should be realistic taking into account the particular working environment.
Time sensitive	Goals should have a time frame for completion, to monitor the progress.

2-2-5 KPI Categories

The previous sections illustrated that firms manage their business by measuring key-performance indicators, and that these indicators should represent – whether indirectly – the firm's strategy and stakeholders' needs. Because not every firm executes the same strategy and not each firm has the same stakeholder groups, different firms will apply different KPIs for performance measurement (Shahin & Ali Mahbod, 2007). Generally, managers apply value driver maps to determine the performance metrics that correspond with the firm's specific strategy (Gates, 2001). A value driver map is a break-down of the firm's strategy into activities – drivers – that are required to achieve the firm's strategy. On the top level of a value driver map, drivers are generic and for many firms identical. Examples of generic performance metrics are net result, operating result, operating expenditures and operating margin. These high-level, mostly financial metrics, are generally applied within firms. As indicated by R. S. Kaplan and Norton (1993), *all* firms should focus on the four perspectives; financial, customer, internal business and innovation and learning when defining performance metrics. However, the authors also note that "*specific* measures within these categories should be tailored to the firm's strategy" (Ittner, Larcker, & Randall, 2003). Thus, firms with different strategies require different metrics. How can we categorise metrics that are specific for each firm?

Table 2-3: Categories of Key-Performance Indicators (Ittner et al., 2003).

KPI Category		Example KPI
1	Short-term financial results	Annual earnings, return on assets, cost reduction
2	Customer relations	Market share, customer satisfaction, customer retention
3	Employee relations	Employee satisfaction, turnover, workforce capabilities
4	Operational performance	Productivity, safety, cycle time
5	Product and service quality	Defect rates, quality awards
6	Alliances	Joint marketing or product design, joint ventures
7	Supplier relations	On-time delivery, input into product/service design
8	Environmental performance	Government citations, environmental compliance or certification
9	Product and service innovation	New product or service development success, development cycle time
10	Community	Public image, community involvement

In order to categorise the many performance indicators that one can think of, Ittner et al. (2003) reviewed literature in the field of the balanced scorecard, intangible assets, intellectual capital, and value-based management to find the most applied categories of KPIs. Based on the models and frameworks that have been developed in these research areas, Ittner et al. (2003) distinguish ten performance categories, being *short-term financial results*, *customer relations*, *employee relations*, *operational performance*, *product and service quality*, *alliances*, *supplier relations*, *environmental performance*, *product and service innovation*, and *community*. These categories are listed in table 2-3. The final column of the table shows example metrics. The classification of Ittner et al. (2003) clearly takes a multi-actor perspective into account, and is therefore considered as an appropriate classification in line with the performance prism. The search terms (PERFORMANCE INDICATORS CATEGORIES), (KPI CATEGORIES), (KPI CLASSIFICATION), (KEY PERFORMANCE

INDICATORS) AND (CATEGORISATION) in the scientific databases ScienceDirect and JStore, and the scientific search engine Google Scholar did not result in literature containing other KPI classifications than Ittner et al.'s (2003) categories. Business intelligence professionals from KPMG have acknowledged that the ten categories are representative for the KPIs that are actually used by firms in practice.

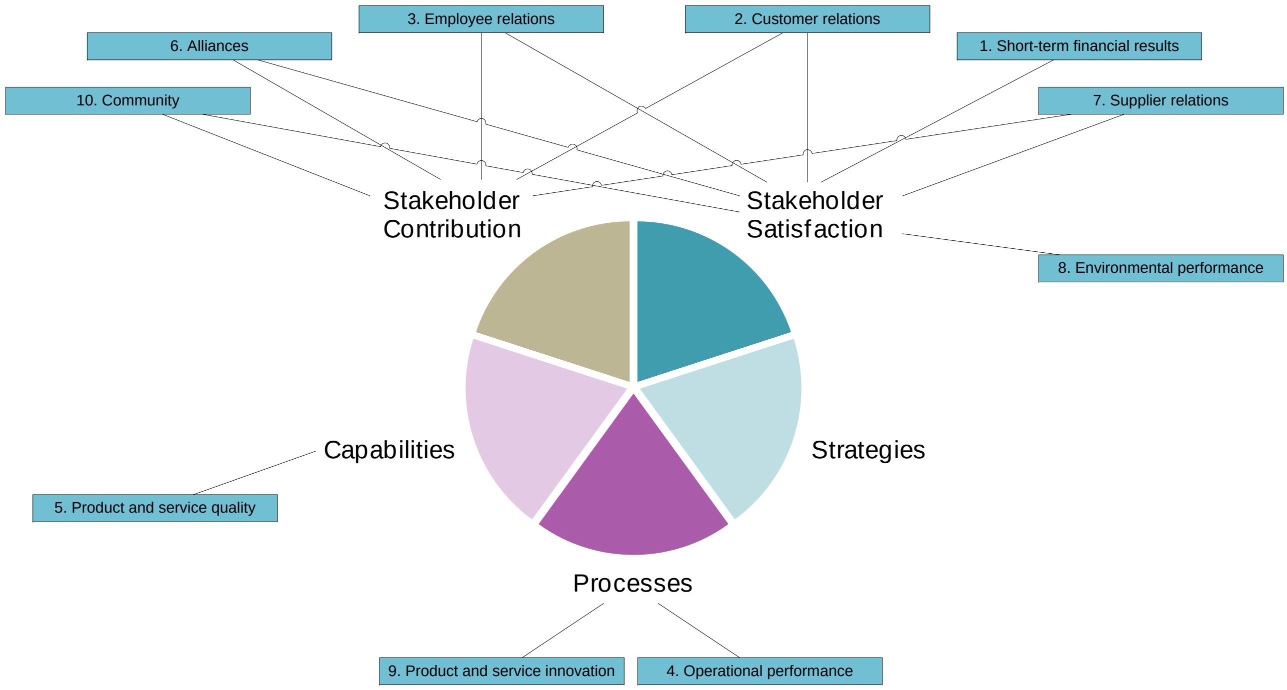


Figure 2-4: The five Performance Prism Perspectives and corresponding Key-Performance Indicator Categories

As described earlier, performance indicators, and especially key-performance indicators should contribute to a firm's strategy. And, as illustrated by R. S. Kaplan and Norton (1992), performance metrics should be established from multiple perspectives, that is, not only financial figures. Furthermore, Neely and Adams (2005) argued that firms should formulate performance metrics from a multi-actor perspective. Therefore, the performance metrics categories established by Ittner et al. (2003) must somehow relate to one of the five *performance prism* perspectives. Figure 2-4 schematically shows the relations between the five performance prism perspectives and the key performance categories. For readability issues, the five perspectives have been visualised in a pie chart, rather than in a prism. For an explanation of the assignment of the KPI categories to the five performance prism perspective, see appendix A. The performance categories of Ittner et al. (2003) allow us to systematically assign social media posts to one of the ten categories. As a result, we can draw conclusions from the applicability of social media data for certain KPI categories. For the remainder of this thesis, we will apply the categorisation of Ittner et al. (2003) to categorise key-performance indicators.

2-3 Processing: From Data to Information

The registering of signals results in raw data which needs to be processed before it represents information. Figure 2-1 showed the business intelligence cycle. In the second phase of BI, registered signals are processed. Van Beek (2006) describes this process as a cycle on itself, which is discussed in this section. It is important to understand the theory underlying the processing of signals when considering to apply social media data for BI purposes, because an organisation usually applies business intelligence already. A social media component should hence be consistent with the existing system(s) and process(es). Van Beek (2006) distinguishes 15 activities making up the processing of gathered data and turning it into information. Figure 2-5 shows the activities in the BI cycle.

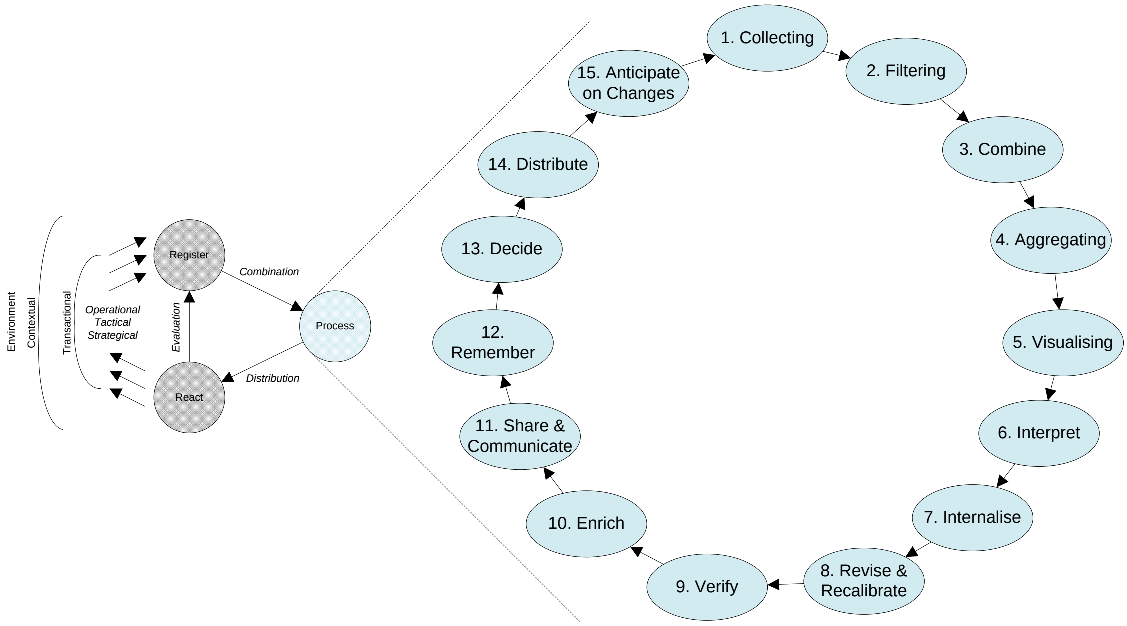


Figure 2-5: Processing data to information. Based on van Beek (2006).

The fifteen steps are discussed in the following section.

1. **Collecting**
   Signals, which are stored as data in different systems are collected in a separate system. When considering social business intelligence, access to different social media platforms – like Twitter and Facebook – is required in order to collect the data. Each social media platform has its own method of storing data, and not each platform is as publicly accessible as the other.

2. **Filtering**
   Only signals that contribute to the deduction of information pass the filtering process. Data that are outdated or of poor quality are removed. Especially when applying social media data for BI purposes, this step requires much effort. The data can consist of spam, polluting the data.

3. **Combine**
   The data that is collected and filtered on separate systems are combined and integrated into one single source, so that analyses are based on one 'version of the truth'.

4. **Aggregating**
   Detailed data are aggregated to a level so that users can quickly understand the data and find information.

5. **Visualising**
   In order to make the data quickly interpretable for the users, the data is visualised.

The first five steps consist of automated activities that convert signals into information. So far, the signals are translated into information that is now interpretable for the users. The next ten steps of the process consist of non-automated activities that involve humans to interpret the information, and act on it.

6. **Interpreting**
   The information generated by the first five steps are interpreted by humans. For example, the automated process generated a graph showing the amount of sales in a given region over a given time span. The meaning of the graph is interpreted by the user.

7. **Internalise**
   In this step, the information derived from the interpreting step is combined with other information in the problem's context. It is in this step that the real underlying trends and explanatory factors are analysed so that the information is embedded in one's cognitive understanding of the system.

8. **Revise & Recalibrate**
   The new information may affect existing information. This step ensures that existing information is revised and adjusted based on the new information.

9. **Verify**
   This step verifies the new information with other mechanisms. For example, a decrease in market share is compared with the companies' turnover development. Whenever turnover increases while market share decreases, it may indicate an increase of the overall market. If such mechanisms contradict, the process of turning the data into information has to be checked for errors.

10. **Enrich**
    In this step, the information – graphs, figures, numbers, etc. – are enriched by textual explanation of the information. A decrease in market share, which is visible in e.g. a pie chart, may be enriched by a textual explanation of two new competitors on the market.

11. **Share & Communicate**
    By sharing and communicating about the information with other members of the organisation, the information is brought under submission of various perceptions and views.

12. **Remember**
    Some information do not require immediate action. However, the information may be relevant whenever future information is acquired. It is therefore important that the information is remembered.

13. **Decide**
    This step involves the reaction on the information. Managers decide how they act on the information, for instance by launching an advertisement campaign, or to sell a part of the organisation.

14. **Distribute**
    The decision in the previous step is generally taken by managers on higher levels of the organisation. The new information and decisions following from that information are distributed to the right persons in the organisation in this step.

15. **Anticipate on Changes**
    The new information may be of a negative character, requiring (structural) organisational change. An organisation should adopt a positive attitude to change according to the new information.

These fifteen steps describe how a signal is generally translated into information at which managers can act. When an organisation intends to implement a (sub) system that extracts signals from social media platforms to derive information, it should be designed according to this method of processing signals.

2-4 Sub Conclusion: How Business Intelligence is Applied

With the rise of a new data source – social media platforms – for firms to access customer perceptions, the question rises how a firm should process these data. The process should in any case correspond with existing business intelligence processes in firms. Therefore, it is essential to understand the general business intelligence process that firms adhere to. This chapter reviewed literature in the field of business intelligence, of which the conclusions are presented in this section.

Though there exist many views on business intelligence, the common aspect is that BI collects and translates data into information that supports managerial decision-making. BI can be regarded as a process of three steps; *registering* data, *processing* the data into information and *reacting* on the conclusions derived from

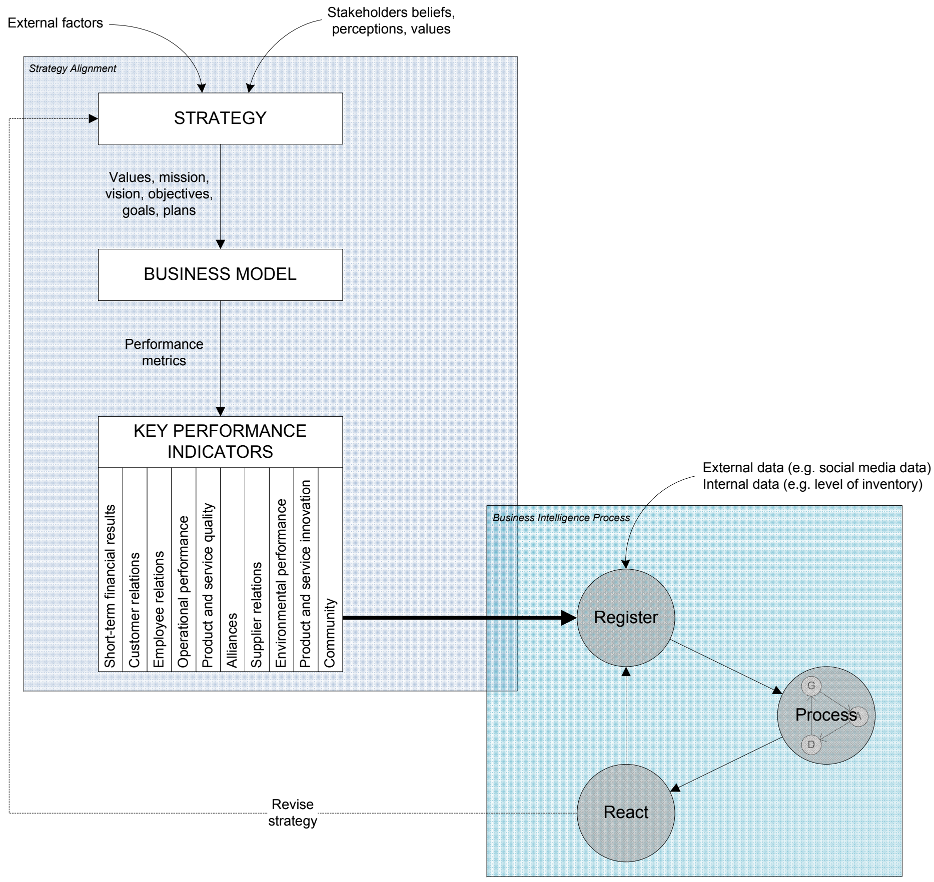


Figure 2-6: Strategy and Business Intelligence

that information. It is essential that firms measure those activities that contribute to its business model and corresponding strategy, because 'what's get measured, get's done'. The key notion that should be concluded from this chapter is that an organisation's strategy should be based on the needs and preferences of the firm's stakeholders, and that a company's strategy drives the values, objectives, goals and plans of company, which, in turn, determine the (key) performance indicators. Therefore, a link between the BI process and an organisation's strategy is required. This link is established by defining key-performance indicators that are based on the firm's business model. In turn, the business model should reflect the organisation's strategy. This perspective on business intelligence is schematically presented in figure 2-6.

Though each firm executes its own specific strategy and will consequently measure organisational performance on specific KPIs, we can classify KPIs into ten categories; *short-term financial results*, *customer relations*, *employee relations*, *operational performance*, *product and service quality*, *alliances*, *supplier relations*, *environmental performance*, *product and service innovation* and *community*. Furthermore, figure 2-6 illustrates that the BI cycle can not commence before the firm has determined 'what it should measure'.

Chapter 3

Research Domain

The purpose of this chapter is to describe the research domain. As illustrated in chapter 1, this thesis intends to develop a procedure that allows firms to process social media data for business intelligence purposes. It is therefore that this chapter explains what social media is, how firms currently apply social media, for which purposes, what social business intelligence is and what developments currently take place in the world of social business intelligence. Since social media is one of the many applications enabled by Web 2.0, we start with a description of Web 2.0.

3-1 Web 2.0

Web 2.0 is the generation of web pages that not only provide information, but additionally allow users to interact with these web pages. In contradiction with the first phase of the web's evolution, Web 2.0 allows anyone to create and share content. The content-creating feature makes it that Web 2.0 is also referred to as "the wisdom web, people-centric web, participative web, and read/write web" (Murugesan, 2007). Web 2.0 allows users to do more than just retrieve information. Whereas the internet was traditionally applied to read, watch, and buy products in Web 1.0, it is increasingly utilised to create, modify, share, and discuss content in the Web 2.0 era.

It is the enabling of the creation of user-generated content that distinguishes Web 2.0 from Web 1.0. O'reilly (2007) – who sees Web 2.0 as "the web as platform" – indicates that the power of Web 2.0 is "collective intelligence", and turns the web into a kind of global brain". The best-known example is probably Wikipedia (launched in 2001), an online encyclopedia created by the Internet users that contains now 23 million (Wikipedia, 2012) articles. The existence of Wikipedia illustrates that people are willing to share their knowledge with others. For example, people that searched for 'Hasbro's Easy-Bake Oven' in Web 1.0 would have found a static web page promoting the product, while in the Web 2.0 era, people also find in the top 5 of search results a warning that the Easy-Bake Oven may lead to serious burns on hands due to a poorly-designed oven door (A. M. Kaplan & Haenlein, 2010). These warnings have been written by users feeling that they had to share their (negative) experiences and product knowledge. Other examples of collective intelligence are ask and answering sites, demonstration videos, and product reviewing sites. Not only knowledge is shared with the rest of the world through Web 2.0 solutions, people also are willing to share their opinions about various topics, their favourite dining places, what they think about the new election candidate, etc. Especially these topics, that we can position under the denominator of opinions and private events, are shared through means of social media.

3-2 Social Media

Social media allow users to connect and share content with each other through Web 2.0 based platforms. The rise of social media in the first decade of the 21^{st} century is a natural consequence of Web 2.0. The relation with Web 2.0 is highlighted by Kaplan & Haenlein's (2010) definition of social media, stating that "social media is a group of internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of user-generated content". However, with this definition we cannot distinguish social media from Web 2.0 sites, since user-generated content is a fundamental element in Web 2.0

anyhow. Why is it that we refer to social media when we talk about Facebook or Twitter, but to Web 2.0 when considering Wikipedia? Kietzmann et al. (2011) state that social media “allow individuals and communities to share, co-create, discuss, and modify user-generated content”. Again, we see the importance of user-generated content when defining social media. But Kietzmann et al.'s (2011) definition contains an important additional component that allows us to distinguish social media from Web 2.0; communities.

Within social media – and especially in social networking sites – users can connect with other users, so that they can share (personal) information. It is the aspect that allows people to connect with each other that distinguishes social media from Web 2.0. The social element, connecting with other users, is much more existent in social media than in Web 2.0. Whereas in Web 2.0, user-generated content is accessible to anyone, in social media people can restrict this accessibility to people they have selected beforehand. Therefore, we define social media as Web 2.0 based applications that allow users to create and share user-generated content with pre-selected users and communities.

3-2-1 Social Media Platforms

The web applications through which users can connect and share content with each other are called social media platforms. There are many social media platforms available, and the range of social media platforms is vast and growing (A. N. Smith et al., 2012; A. M. Kaplan & Haenlein, 2010; Hanna et al., 2011). These platforms differ in scope and functionality. In turn, there is variation in how people use these platforms. “Some sites are for general masses, like Friendster, Hi5 and Facebook. Other sites, like LinkedIn, are more focused on professional networks. Media sharing sites such as MySpace, YouTube, and Flickr concentrate on shared videos and photos (Kietzmann et al., 2011).

In order to support managers in understanding social media, and to select the right platform for the firm's purpose, researchers have tried to classify the differences between the social media platforms. Weinberg and Pehlivan (2011) distinguish social media platforms based on two dimensions; (i) *half-life of information* and (ii) *information depth*. The half-life of information refers to the “longevity of the information in terms of availability / appearance on the screen and interest in a topic. Depth of information refers to the richness of the content, and the number of diversity of perspectives” (Weinberg & Pehlivan, 2011). As such, Weinberg and Pehlivan (2011) positioned popular social media platforms in their framework (figure 3-1). Micro-blogs, such as Twitter, allow users to use a limited number of character in each post and therefore have a shallow information depth. On the other hand, community sites purposed to extensively discuss topics among users have a higher information depth.

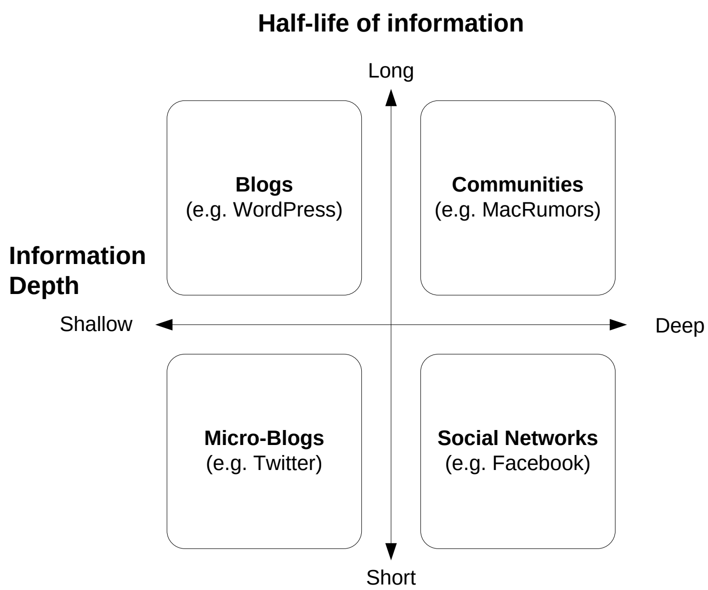


Figure 3-1: Social media by information half-life and information depth (Weinberg & Pehlivan, 2011).

Kietzmann et al. (2011) created a framework that distinguishes social media platforms based on seven building blocks. These blocks are “constructs that allow us to make sense of how different levels of social media functionality can be configured”. The functional building blocks are shortly discussed below, and are applied on Facebook, LinkedIn and Twitter in figure 3-2 (page 30) to illustrate different focus points on different social media platforms.

1. **Identity**
   This block represents the extent to which users reveal their identities in a social media setting. Especially on self-branding platforms, such as LinkedIn, identity is a strong aspect.

2. **Conversations**
   This block represents the extent to which users communicate with other users in a social media setting. Some sites are much more intended to facilitate conversations – like Twitter – than others.

3. **Sharing**
   Sharing represents the extent to which users exchange, distribute, and receive content. Especially on Twitter, people share what they are doing, what they think of, etc.

4. **Presence**
   The presence block represents the extent to which users can know if other users are accessible. It includes knowing where others are, like 'check-ins' at Facebook or Foursquare.

5. **Relationships**
   This block represents the "extent to which users can be related to other users". Related implies "some form of association that leads them to converse, share objects of sociality, meet up, or simply just list each other as a friend or fan".

6. **Reputation**
   Reputation is the extent to which users can identify the standing of others.

7. **Groups**
   The groups functional block represents the extent to which users can form communities and sub-communities.

In the following sections we discuss three important social media platforms that are part of the analysis in this thesis, Twitter, Facebook and Blogs.

Twitter

Twitter is recognised as being *the* site on which users ask information and complain. Twitter is a micro-blogging site, designed to let people post short – 140 character – text updates called 'tweets' to others. Twitter prompts users to answer the question 'what are you doing?', leading to a constantly updated timeline of short messages that range from humor, opinions, musings on life to links and breaking news. Kietzmann et al. (2011) argues that Twitter posts are "mostly short status updates of what users are doing, where they are, how they are feeling, or links to other sites". Participants choose Twitter accounts to 'follow' in their stream, and they each have their own group of 'followers'. Unlike social networks like Facebook and LinkedIn, where a connection is bidirectional, Twitter has an asymmetric network infrastructure of followers. The site was launched in 2006, and broke into the mainstream in 2008 – 2009, when accounts and media attention grew exponentially (Marwick & Boyd, 2011)". In February 2012, Dugan (2012) announced that Twitter had over 500 million users registered. Twitter is an important phenomenon from the standpoint of its incredibly high number of users.

According to Jansen et al. (2009), around 20% of all tweets contain mention of a brand. Of these brand-related tweets, nearly 20% express a brand sentiment, of which 50% were positive, and 33% were critically. In 2010, the number of Twitter followers per firm increased by 241% over the year (Kirtis & Karahanb, 2011). Acknowledged by A. N. Smith et al. (2012), Twitter posts contain more brand-related information than Facebook and YouTube. Since the purpose of this thesis is to contribute to the development of social business intelligence in firms, Twitter is a social media platform that is part of the analysis.

Facebook

Facebook is the absolute number one social networking site. Though it was only founded in 2004, it is ranking second in the most popular websites in the world these days. In July 2012, the website reported to have 955 million monthly active users (Sloan, 2012) who log on at least once every 30 days. Half of these active users

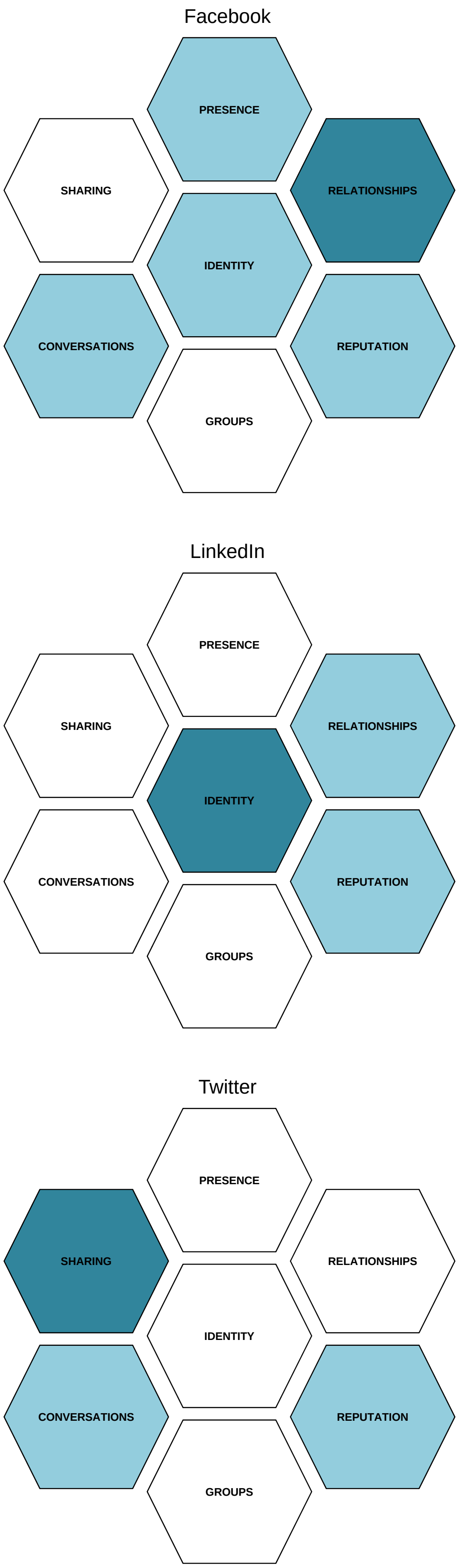


Figure 3-2: Different social media serve different purposes. Based on Kietzmann et al. (2011).

log on every day (Laroche, Habibi, & Richard, 2012). Facebook users can "create profiles featuring personal information, interests, photos, and the like, and can "friend" other site users. They can also participate in a wide range of activities such as writing on friends' walls, commenting on links, participating in forum discussions, and "liking" brands. Facebook allows people to build or maintain social capital, communicate with others, keep up with other peoples' lives, and discover rumours and gossip" (A. N. Smith et al., 2012).

Over 2010, the percentage of corporations active on Facebook increased by 13%, with the number of 'likes' per page rising by 115% globally (Kirtis & Karahanb, 2011). Where Twitter is considered as a platform for companies to communicate instantly with stakeholders, Facebook is suitable for creating communities among stakeholders.

Blogs

While Twitter and Facebook conversations are often unstructured and brief, blogs may be a source for organisations to discover structured customer opinions. After a slow start in the late 1990s, weblogs ("blogs") – such as Blogspot and Wordpress based websites – have become very popular, because they are easy to create and to maintain (Kietzmann et al., 2011). Blogs are often designed as product review sites, where customers can share their experiences with their products. Generally these blogs are publicly accessible and there are no restrictions to the amount of characters per post.

3-2-2 User-Generated Content on Social Media Platforms

Social media platforms exist by the virtue of user-generated content ("UGC"). UGC is content that is "publicly accessible, created outside of professional practices and shows a certain amount of creative effort" (A. M. Kaplan & Haenlein, 2010). UGC can take many forms, such as pictures on Facebook, videos on YouTube, statements on Twitter, product experiences on blogs, etc. Without users that create and share content, social media sites are like an empty shell. Therefore, UGC is considered as the fundamental element underlying social media (A. N. Smith et al., 2012; Boyd & Ellison, 2007). Enabled by Web 2.0, user-generated content has become increasingly popular on the internet since the early 2000s: more and more users participate in content creation, rather than just consumption (Agichtein et al., 2008). In China, the percentage of internet content that is user-produced now exceeds that what is professionally produced (A. N. Smith et al., 2012). According to A. M. Kaplan and Haenlein (2010) it are not only the technical developments of Web 2.0 and an "increased broadband availability and hardware capacity" that has contributed to the popularity of UGC on social media these days, but also "the rise of a generation of 'digital natives' and 'screenagers' with substantial technical knowledge and willingness to engage online".

Though UGC varies in nature between different social media platforms (Kietzmann et al., 2011), A. N. Smith et al. (2012) and Jansen et al. (2009) indicate that much user-generated content – around 20% – on the internet contains a brand name. It is here where the opportunities for firms materialise, firms can inspect the posts that contain a brand name to discover customer opinions related to their brands. The internet's accessibility, reach, and transparency have empowered firms that are interested in consumers opinions (Kozinets et al., 2010). User opinions were not that easy to be gathered before the social media era, while they are now accessible at low costs (Kirtis & Karahanb, 2011).

As illustrated in the previous paragraph, UGC on social media contains opportunities for firms. However, user-generated content also contains disadvantages. Especially issues related to variance, cohesion and verification are at stake when processing user-generated content from social media sites. The three issues are discussed in the following paragraphs.

Variance Firstly, the variance in the quality of UGC is high, "any data can contain information ranging from excellent to spam" (Agichtein et al., 2008). This makes the tasks of filtering and ranking the importance of social media posts more complex than non user-generated content.

Cohesion Secondly, professionals that base decisions on social media content should be aware of the negative effects arising from cohesion; one negative message about a firm – which may not even be true – can snowball over the internet, reaching many people, and may eventually harm the performance of the company. "Cohesion describes the phenomenon that evaluations of cost and benefit associated with prospective behaviour are aligned via strong communication relationships. People thus become more homogeneous as a result of direct contact via social networking links from node to node. This type of social contagion is typically referred to as word-of-mouth" (Takac, Hinz, & Spann, 2011).

Cohesion can work out positive as well as negative for an organisation. The negative characteristic materialises whenever a negative message circulates along the social network. This message is likely to influence the perceptions of the organisation in a negative manner. On the other hand, cohesion may offer opportunities. Whenever an organisation intentionally influences the discussions on social networks in a positive manner, the message is likely to be adopted by a large community.

Verification Thirdly, the providers of the information in the social media world generally spread information without verification, unlike the traditional mass media. Dong-Hun (2010) argues that social media is not yet capable of replacing the traditional media because of the credibility problem. However, Wikipedia is a successful example of a website that is based on trust, and established and maintained by the crowd. Information posted on the website is verified among other users, that directly renovate incorrect information.

The identified threats related to social media data should be considered when developing and implementing a social business intelligence system.

3-2-3 Current Applications of Social Media in Firms

Both large and small organisations are increasingly visible on social media platforms. In addition, managers "sense that social media is and will remain an important fabric of commerce" (Weinberg & Pehlivan, 2011). A Burson-Marsteller research investigated the application of the platforms Twitter, Facebook, YouTube and Corporate blogs, and found that 25% of the firms actively use all four social media platforms, 84% uses at least one of them (Kirtis & Karahanb, 2011). Of the Fortune 2000 companies, 69% currently use social networking sites, while 37% planned to use more of them over the next five years (McCorkindale, 2010).

What does it mean when firms 'use' social media? IBM (2011) researched for which activities firms applied social media, the results are shown in figure 3-3. Although IBM (2011) provides detailed insight in the many social media activities that firms employ, we can conclude that firms generally apply social media to communicate with customers, promote activities, monitor the brand name and inspect customer ideas. These four activities are discussed in the following paragraphs in more detail.

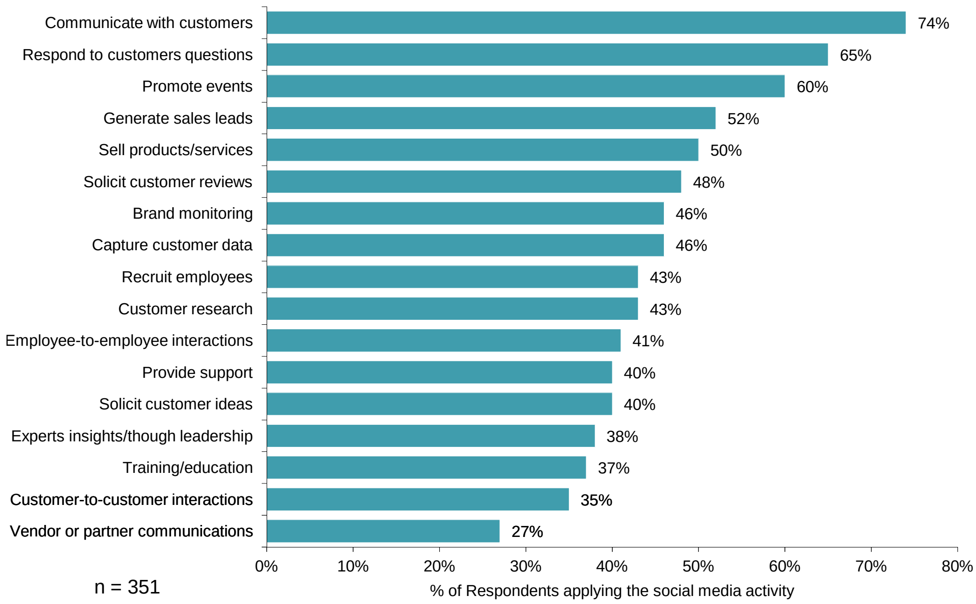


Figure 3-3: Applications of social media (IBM, 2011)

Marketing

Web 2.0, and especially social media, has empowered the 'voice of the customer'. Consumers are no longer merely passive recipients in the marketing exchange process (Hanna et al., 2011). In the past, marketing campaigns were typically developed by companies in-house, without interference of (potential) consumers. Campaigns had the character of 'here is the advert, please absorb it', or 'here is the product, we hope you like it'. This 'we talk, you listen' approach has been replaced by 'you talk, we listen' as a result of the possibilities offered by social media (e.g., Patterson, 2012; Klassen, 2009).

Research indicates that marketing through social media is effective for firms. "70% of the consumers that have used social media websites to take product or brand information, 49% of these consumers made a purchase decision based on the information they pound through social media sites" (Kirtis & Karahanb, 2011).

Firms are also turning into social media marketing to lower the firms' expenditures. The cost reduction aspect of online marketing as compared to traditional marketing is one of the main reasons why companies are nowadays applying social media for marketing purposes (Kirtis & Karahanb, 2011). Cost reduction is mainly achieved by the elimination of the distribution phase, which is required in traditional mass media. In addition, marketing through means of social media is less expensive than the regular channels because most social media applications are free of charge. As such, the biggest expenditures related to the execution of a social media strategy represents the time employees spend on posting messages, responding to comments and blogging. Social media allows marketers to specifically target on client groups, and distinguish between products / services case by case. In comparison with traditional marketing channels, social media shows also on this aspect lower costs. Driven by the global recession, many firms are in a cost-reducing mode. Because of the economic turmoil social media is applied as a survival tool by many firms, so the economic recession has increased the rate of shift change from traditional media to social media (Kietzmann et al., 2011).

For the beneficiary reasons of marketing through social media described in the previous, A. N. Smith et al. (2012) estimate that the percentage of companies using social media for marketing is expected to reach 88% by 2012, up from 42% in 2008.

Customer Relations Management

Firms also use social media for customer relations management ("CRM"), also referred to as social CRM. Thanks to social media, "the nature of public relations and how organisations engage their public has changed a great deal in the past several years" (McCorkindale, 2010). "An environment in which control of the relationship has shifted to the customer, who has the power to influence his or her social network" (IBM, 2011) drives organisations willing to participate in the online conversations.

Social media platforms hold unprecedented opportunities for companies to get closer to customers, allowing firms to communicate directly with customers, for instance to provide support when customers encounter problems with products / services. According to Patterson (2012), firms "have made progress in conversing with their customers".

A recent study by Laroche et al. (2012) indicates that it is beneficiary for firms to establish online communities in which both firms and customers communicate with each other, "brand communities established on social media have effects on customer/product, customer/brand, customer/company and customer/other customer relationships, which in turn have positive effects on brand trust, and trust has positive effects on brand loyalty".

Reputation Management

A failing social media engagement strategy can significantly impact a firm's reputation and sales (Kietzmann et al., 2011). The increased application of social media has serious consequences for an organisation's exposure to its environment. It seems that the power has been taken from the corporate marketing departments by individual consumers that create, share and discuss online blogs, tweets, Facebook entries, movies, pictures, etc. With or without permission, communication about brands will happen. In an environment where customers gain more and more power, organisations need to carefully tread their actions and control its' exposure. Therefore, companies more and more empower employees to talk, listen, and respond to what consumers post on social media (A. N. Smith et al., 2012) in order to control the firm's (online) reputation.

One negative message about an organisation – created by one single person – can snowball over the internet, reaching many people, and may eventually harm the performance of the company. In 2008, Canadian singer Dave Caroll wrote a song about United Airlines' luggage handling employees recklessly throwing his guitar, which caused a break in his guitar. Frustrated by bad customer experience, he uploaded his 'United breaks guitars' song on YouTube. Consequently, United Airlines experienced a 10% drop (Patterson, 2012) in its share value and suffered damage to its reputation. The YouTube clip has been viewed over 12 million times. This is one of the examples that show how powerful the force of social media can be, when a company does not act according to the preferences of the community. As such, social media platforms may be a source of both threats and opportunities for brands experiencing unfavourable exposure (A. N. Smith et al., 2012).

Co-creation & Pro-sumers

Today, consumers "are taking an increasingly active role in co-creating everything from product design to promotional messages" (Berthon, Pitt, McCarthy, & Kates, 2007). This phenomenon is known as co-creation, and more recently termed as "prosuming" (DesAutels, 2011), illustrating that people are not only consumers but at the same time producers. Firms are much more required to perceive consumers as partners in the process of creating products, whereas this was – before the social media era – formerly an activity for solely the firm.

An example of such a process is Lay's recent campaign to decide the new flavour of their potato crisps. In Lay's campaign, consumers were stimulated to contemplate new flavours and to post these ideas on the web. Other users consequently rated the ideas that were send it. The winning flavours have actually been brought to production. Another example related to co-creation is Samsung, which 'listened' closely to the user-generated content on blogs, and, after hearing complaints that the speakers on the side of the TV were too wide for many customers' entertainment cabinets, it redesigned the product (Klassen, 2009).

The co-creation opportunities for firms offered by social media reach even further. A growing number of organisations, among them 3M, AEGON, HCL Technologies, Red Hat and Rite-Solutions have recently experienced with crowdsourcing their strategies (Gast & Zanini, 2012). The organisations offered the public the possibility to provide input in the form of proposals for the company's future directions. The effects resulting from strategy crowdsourcing is twofold. In the first place, the company gathers information from the external environment, including perceptions from important actors that would normally be overlooked. An organisation can consequently craft its strategy with a higher quality. Secondly, the organisation creates "enthusiasm and alignment behind a company's direction" (Gast & Zanini, 2012).

Though the previous sections illustrate that social media is widely applied for different purposes in organisations, many executives eschew or ignore this form of media because they "don't understand what it is, the various forms it can take, and how to engage with it and learn" (Kietzmann et al., 2011). Also A. M. Kaplan and Haenlein (2010) state that the reluctant attitude of some managers towards social media is due to "a lack of understanding regarding what social media are". Many organisations acknowledge the opportunities in the application of social media, while, on the other hand, there also exists a fair degree of uncertainty with respect to allocating effort and budget to social media, and "limited understanding of the distinctions between various social media platforms" (Weinberg & Pehlivan, 2011).

3-3 Social Business Intelligence

Firms should measure the effects of social media activities on organisational performance. As illustrated in chapter 2, the process of business intelligence requires key-performance indicators to be defined so that the performance of the firm can be measured against its strategy. This value-based management approach is generally applied within firms, implying that when a firm pursues to perform social media activities, it should measure the effects of these activities in relation with organisational performance.

Existing social media monitoring tools mainly reveal the performance of the organisation on social media (number of mentions, number of likes, % of positive mentions), and treat the social media component of a firm as a separate business unit executing its own strategy. However, the purpose of business intelligence is to reveal the underlying parameters that determine the performance of the organisation, that is, not limited to solely social media performance. In order to understand the influence of social media content on an organisation's performance, a link between the company's key-performance indicators and clear social media parameters is required.

It is argued that the possibilities of social media for business intelligence purposes reaches further than what is currently offered by the social media analytics tools. The key benefits will be gained whenever the KPIs of an organisation are linked to the parameters that are measured by social media tools. Only in that case, one can speak about 'social business intelligence'. In social business intelligence, the social media activities related to a firm are translated to organisational performance.

3-3-1 The Current State of Social Business Intelligence: Early Adoption

Software developers acknowledge the opportunities generated on social media platforms for firms. With the rise of social media, and the popularity of BI within organisations, software solutions offering social media 'intelligence' are emerging rapidly. As a result, tools for analysing information become widely available at ever-lower prices (Bughin, Chui, & Manyika, 2010), some are even offered for free.

Auditore (2012a) – the former head of SAP's Business Influencer Group and now researcher at Asterias research – investigated the market for social business intelligence and found that the top four emerging SBI platforms consists of Radian 6, Kapow, evolve24 and NetBase. According to Kapow (2009), a provider of business intelligence software, we are at a point in time where social media can be integrated into enterprise business intelligence platforms. Not only small software development firms are on a discovery journey, well-established companies offering total business intelligence solutions are also embracing social media data. For example, SAP collaborates with NetBase to offer social media analytics. IBM's Cognos provides social network capabilities. Oracle recently acquired Involver, Vitrue and Collective Intellect to add social media analytics to their portfolio of services. SAS incorporated social media analytics in its platform, and QVSource allows QlikView users to extend their BI platform with social media intelligence. Table 3-1 lists the top existing, new and emerging vendors of (social) business intelligence solutions.

Table 3-1: Top (Social) Business Intelligence Vendors (Auditore, 2012a).

Legacy BI vendors	New social media BI vendors	Emerging social media BI vendors
1. IBM	1. Google	1. Radian6
2. Oracle	2. SAS	2. Kapow
3. SAS Institute	3. IBM	3. evolve 24
4. SAP		4. NetBase

Emerging Social Media Business Intelligence Vendors

Companies that apply social media in their organisation generally apply a cycle consisting of three steps; (i) monitoring, (ii) analysing, and (iii) engaging (Kapow, 2009; Bryant, 2011) using social media monitoring platforms. The objective of these platforms is to 'listen', in order to *monitor* the brand(s). Generally, "automated scripts monitor a handful of keywords from targeted web sites" (Kapow, 2009). The gathered data in the listening phase is generally continued by mapping customer perceptions, sentiment measuring and an indication of the company's reach respecting social media. In general, these tools are solely based on "simple quantitative counts of how many times a brand has been mentioned" (Patterson, 2012). Some exceptions exist that provide a general mood of the brand, often based on large datasets. These functions are referred to as *analytics* by the software offerers. Clients receive weekly or daily reports containing figures representing the amount of last week mentions on social media platforms, the number of likes, the sentiment related to that, the number of shares, retweets, a distribution of the locations, gender distribution, etc. In addition, the software platforms generally scan all social media platforms continuously and present all relevant content to the user(s), via dashboards and/or automatic generated reports. Companies' managers can consequently *engage* with the social media users via one portal. The nature of the engagement of companies is often related to customer relationship management, e.g. a customer-service department explaining to an individual why his or her credit card is not functioning, or why the company's website is not presented properly in the customer's browser. Other social media posts made by organisations are often marketing related, e.g. an announcement of a new product release or an offer.

The emerging social media intelligence tools – Radian6, Kapow, evolve24 and Netbase – and their features are presented in the following section.

- **Radian 6**
  Salesforce's Radian6 provides social media monitoring tools, social media engagement software and social customer relationship management and marketing software. It provides companies with social analytics comprising of social media metrics and sentiment analysis. Radian6 provides firms with a dashboard illustrating their performance on social media. Consequently, firms can engage in online discussions. Clients of Radian6 include Fujifilm, Commerce Bank, KLM, Pepsi, L'Oreal, Baker Tilly and Activision (Radian6, 2012). Radian6 offers clients different packages with different features, ranging from EUR 750 per month to EUR 12,000 per month.

- **Kapow**
  Kapow offers solutions for accessing, extracting and enriching web data (Kapow, 2009). The software developer illustrates the applicability of public web data for business intelligence, it explicitly mentions that it offers software that structures social media data and turns it into interpretable information. One of the many tools offered by Kapow is the monitoring of social media platforms. Users of Kapow's software include AT&T, Intel, Cisco, Vodafone, Morgan Stanley, P&G, DHL, Barclays, Lenovo and Audi.

- **evolve24**
  evolve24 mines, priorities and scores online conversations so that relevant intelligence is provided to the manager of a firm. The software of evolve24 allows users to create custom dashboards to present those social media metrics that are relevant to the firm. Next, it allows predictive modelling to predict the impact of certain issues, so that the decision process in the organisation is supported by this information (evolve24, 2012).

- **NetBase**
  NetBase allows users to track social media issues related to the topics of interest. It processes billions of social media posts to extract structured insights that enterprises can use to quickly discover market needs and trends, quantify market perceptions about products, services, and companies (Netbase, 2012). SAP collaborates with NetBase to for social media analytics solutions. Hence, SAP users can easily integrate the social media analytics provided by NetBase in their existing BI platforms. Amongst others, Tupperware, Hewlett Packard, Coca-Cola and Kraft Foods are users of NetBase.

Intelligence Provided by Social Media Monitoring Tools

Social media tools, whether they refer to themselves as monitoring, analysis or intelligence tools, offer a variety of insights in the performance of firms on social media. The novelty of social media and the even more unexplored applicability of business intelligence on the new phenomenon makes in that there is little scientific literature available in this field. Instead, both large firms as IBM and small organisations present white papers and blogs in which they describe their view on social media metrics. Many of these documents refer to the same variables that they measure, though they generally adhere to their own 'invented' name. Common social media metrics and intelligence that are provided by the social media monitoring tools are presented below.

- **Volume of Posts**
  The volume of posts measures number of messages or articles that have been created on social media for a specific topic over a given period. The volume of created social media posts containing a firm's name (or product / service name) illustrates to what extent a company is subject of discussion of social media. The volume of social media posts can vary from day to day or even from hour to hour. A sudden upwards deviation in the average volume is a signal for a firm that people are paying attention to the firm, whether positive or negative. Firms can relate this figure to marketing campaigns or other organised events to determine the success of their reach.

- **Engagement**
  Engagement represents the involvement of users in the brand. Often, companies measure engagement by the amount of likes, followers, shares, retweets, etc. However, solely looking at this figure is not enough. After all, it is relatively easy to influence this figure, for instance by organising a lottery in which users can win an iPad. Doeland (2012) distinguishes engagement metrics into *distribution* metrics and *interaction* metrics. Distribution metrics describe how well the organisation is visible to the social media public, while interaction metrics represent how well the public engages in the brand.

- **Sentiment**
  Almost all software tools – even those available for free – offer sentiment analysis, a measure that represents the attitude of the content generated by the social media users. Generally, social media posts are classified as either positive, neutral or negative by linguistic algorithms. These algorithms 'simply' textmine each post associated to the organisation and consequently connect words and phrases like 'great', 'wow', 'good', ':-)', 'super', etc. with a positive attitude. Posts containing words like 'bad', 'dumb', 'worthless', etc. are classified as negative posts. As such, an indication of the sentiment under the social media users is generated. Figure 3-4 illustrates the output of a sentiment analysis as it is provided to a user of a social media monitoring tool, in this case uberVU, one of the popular social media monitoring tools.

  Sentiment analysis is a complex activity. Not only because each language requires its own meta data to classify words and phrases in different languages, but also because most of the sentiment analysis tools

Table 3-2: Examples of Engagement Metrics

Distribution Metrics	Interaction Metrics
Followers	Retweets
Fans	Forwarding
Mentions	Sharing
Reach	Comments
Bookmarks	Likes
Inbound links	Rates
Blog subscribers	Reviews
	Contributors
	Traffic generated
	Time spent on site
	Response time

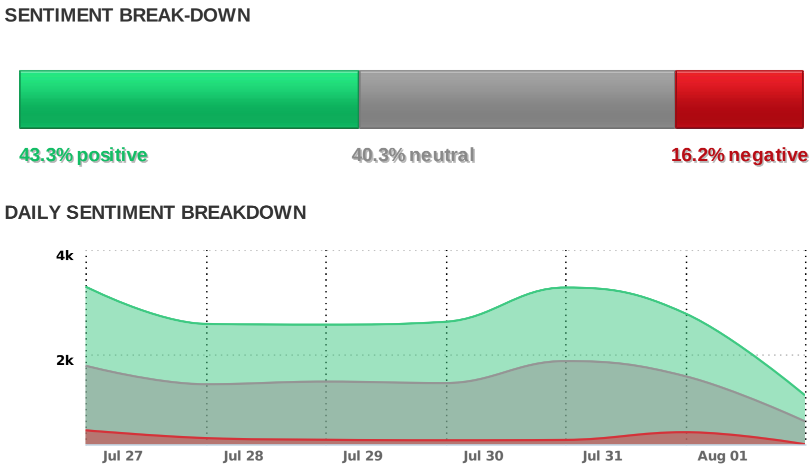


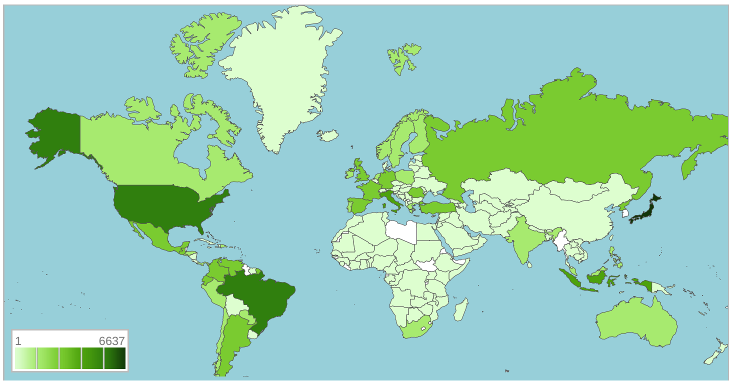
Figure 3-4: Sentiment Analysis Example

use Natural Language Processing techniques. These techniques assume that the underlying text is "clean and correct" (Dey & Haque, 2008), a requirement that is not always present in social media posts. Social media posts comprise spelling errors, ad-hoc abbreviations and improper casing, incorrect punctuation and malformed sentences. These features pollute the outcome of the algorithms. However, "interest in noisy text analytics has increased significantly in the recent past" (Dey & Haque, 2008). The systems that are currently developed also take phrases into account (Agichtein et al., 2008), turning "Wow, the new product of ABC is really great.. NOT!" into a negative sentiment post. As such, the accuracy of sentiment analysis is expected to increase by new methods that are currently developed. Most platforms are commercial and do not disclose full details of their internal feature set.

- **Geography**
  Whenever a person registers itself for a social media platform, he or she is required to fill up some personal information, including the person's residence. Though it is not guaranteed that users provide legitimate personal details, social media monitoring tools use this information to determine the location of where the posts has been made. Next, mobile devices including a GPS component can – if allowed by the user – provide the social media post with more accurate geographic information. As such, social media monitoring tools provide details about the geography of the social media posts of a firm in a given period. Figure 3-5 shows an example of the output of a social media geography analysis.

- **Topic and theme detection**
  Social media monitoring tools provide details in the primary topics and themes that consist in the dataset related to the firm. Generally, a list of the ten most 'trending topics' is presented. Topic and theme detection allows firms to quickly grasp an understanding of the most discussed topics that consist in the social media posts related to the firm.

- **Influencer ranking**



TOP LOCATIONS	
JAPAN	17%
UNITED STATES	11%
BRAZIL	8%
INDONESIA	7%
ITALY	6%
TURKEY	4%
SPAIN	4%
MEXICO	3%
SOUTH KOREA	3%
UNITED KINGDOM	3%

TOP LANGUAGES	
SPANISH	35%
ENGLISH	22%
PORTUGUESE	10%
JAPANESE	8%
ITALIAN	4%

Figure 3-5: Geographic Analysis Example

Almost all social media platforms – and especially social networking sites – provide the possibility to follow other users. As a consequence, the messages that have been created by people with many followers will reach many other users. Social media monitoring tools provide insight in the amount of followers of the people that posted a message containing the firms names. The users with the most followers are considered the strongest influencers.

It would be wise to combine the sentiment of the posts with the posts made by the strongest influencers. A negative message created by a strong influencer is likely to reach many people, which may result in an overall decrease of the sentiment towards the firm. On the contrary, strong influencers posting positive messages may increase the overall sentiment. With this information, web care teams can focus on the people with many followers in order to have the strongest effect on the desired result, which may be an increase in the overall sentiment.

- **Channel distribution**
  In order to understand on which social media platforms firms are subject of discussion, social media monitoring tools provide insight in the distribution of the posts related to the firm across the different platforms. With this insight, firms can decide to focus on those platforms where their firm is subject of discussion. Figure 3-6 shows an example of the output provided by a social media monitoring tool (uberVU) illustrating the distribution of social media posts in a certain period across various channels.

A 2012 research by IBM and SHARE-Unisphere amongst 711 business and IT managers from across the world revealed that “72% of the respondents – firms – are monitoring social media networks, reflecting great awareness of the importance of understanding information flow and engaging social media networks”. The most mentioned business functions employing social media include “sales and marketing (64%), public relations and communications (38%), IT (37%) and customer service (37%)” (Auditore, 2012b). Table 3-3 lists the top business initiatives and the parameters that were measured (Auditore, 2012a). The research also indicated that investments in the area of social business intelligence “continue to trend upward … 60% of the respondents indicated that they expect to increase social media monitoring over the next 1–2 years, while 21% indicated it would be 3–5 years. However, the study shows, managers are unclear about the ultimate usefulness of social media. This reflects that social business intelligence is still in an ‘early adopter’ phase. The study concluded that “social media based business intelligence represents the next great frontier of data management, promising decision makers vast vistas of new knowledge gleaned from exabytes of data generated by customers, employees, and business partners” (Auditore, 2012b).

Figure 3-6: Channel Distribution Example

Table 3-3: Top Business Initiatives for Social Media and Measured Parameters (Auditore, 2012b).

Top business initiatives	Top metrics employed
1. Brand-reputation management	1. Customer satisfaction
2. Marketing communications	2. Overall chatter
3. Customer service	3. Brand experience
4. Customer experience management	4. Advertising campaign performance
5. Sales	
6. CRM	

3-4 EU Legislation on Social Media Data Processing

Social networks have obtained a "poor reputation for protection users' privacy due to a continual flow of media stories discussing privacy problems" (Bonneau & Preibusch, 2010). Examples of such stories include "disclosure of embarrassing personal information to employers and universities, blackmail using photos found online and social scams" (Bonneau & Preibusch, 2010). The European Commission is of the opinion that social networks are a useful tool for staying in touch with friends, family and colleagues, but that these networks also present a risk that personal information, photos and comments might be viewed more widely than people realise. The Commission also states that in some cases this can have financial, reputational, and psychological consequences.

Currently, legislation in the European Union's member states on data privacy is based on the *Data Protection Directive* 95/46/EC. This Directive has been established in 1995, a period in which Web 2.0 and social networks did not exist. The technological developments and the scale of data sharing and collecting have increased in recent years. Given the advances in IT, the Commission deems Directive 95/46/EC outdated. In addition, as with any Directive, all member states have composed national legislation based on the Data Protection Directive, implying that each member state applies its own Data Privacy Policy. E.g, in the Netherlands this resulted in the *Personal Data Protection Act*[1] in 2001. It is therefore that the Commission drafted a proposal in January 2012 for *Regulation on the protection of individuals with regard to the processing of personal data.* This new legislation takes the social media era into consideration, and is directly applicable in all EU member states.

Directive 95/46/EC provides the basis for the definition of personal data, which may be contained in social media messages. Personal data are defined as "any information relating to an identified or identifiable

[1] In Dutch: *Wet Bescherming Persoonsgegevens.*

natural person; an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity"[2]. The processing of personal data is defined as "any operation or set of operations which is performed upon personal data, whether or not by automatic means, such as collection, recording, organisation, storage, adaption or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, blocking, erasure or destruction"[3]. The *Data Protection Directive* is only applicable when the data can be marked as personal data.

The newly proposed Data Protection Regulation adheres to the personal data definition of Directive 95/46/EC. Thus, any data that provides one the possibility to retrace a natural person from that data, is personal data. The Commission introduces the 'right to be forgotten', implying that a social network user can request – if there is to legitimate reason to store it – to remove all data related to the person from their system. Personal data can only be collected after explicit consent of the person that provides the information. Furthermore, providers of social media should adopt the principle of 'privacy by default', implying that the default settings should be those that provide the most privacy. Social media sites should also inform users about how the personal data will be used. The new legislation is expected to come into force in 2014, with penalties up to one million Euro or 2% of the firm's global revenue in case of a breach. In the following paragraph, we discuss how the new legislation affects the possibilities offered by publicly accessible data for firms and what procedures are necessary to be in compliance with the new legislation.

3-4-1 What Firms are allowed to do with Public Data

Firms are allowed to process data whenever these data are not personal data or whenever the creator of the data has given prior consent to process the data. In order to avoid data to be legally labelled as personal, the data should be pre-processed in a way that it is not possible to retrace a natural person from the data. Thus, the data should be made anonymous. E.g., attributes containing the name of the users should be removed. Though it is not guaranteed that persons actually use their official name on social media, it is advised that firms remove those attributes that may contain information allowing one to retrace a natural person from these data. Furthermore, firms can aggregate the data to a level at which the individual message is not considered for their analyses. The 'right to be forgotten' has consequences for the way in which social media data is stored and distributed. With the new Regulation, any person can withdraw his or her information from a social media site. However, social network sites distribute – by means of APIs or trough other ways – the social media messages to third parties. It will be the social network providers that will become responsible to communicate to its third parties that a certain user has requested to delete its content.

In order to avoid suspicions, it is advised that firms intending to process social media data carefully document the steps that they undertake to make the data anonymous, and how the 'right the be forgotten' is enabled in the processing of the social media data. Such procedures are to be designed so that privacy is embedded in the procedure, known as the 'privacy by design' principle in the new Data Protection Regulation.

3-5 Sub Conclusion

Social media platforms are Web 2.0 based applications that allow users to create and share user-generated content with pre-selected users and communities. There exists a variety of social media platforms, some are aimed at relations between the users, while others are developed to share media like photos and videos. Research indicates that of all the user-generated content on the internet, about 20% is brand-related. Users e.g. write their opinion about a new product, complain about a service, discuss new ideas, etc. Therefore, it is interesting to investigate the opportunities for firms to analyse the content that contains their brand name. However, though user-generated social media content may be valuable for a firm, there also exist pitfalls in collecting, analysing and drawing conclusions from these data. Firstly, the variance in the quality of UGC is high; any data can contain information ranging from perfectly true to spam. Secondly, cohesion may lead to homogeneous content; implying that one user adopts the opinion of another. Thirdly, one should be aware of the fact that users post their messages on social media sites without verification.

Firms are increasingly visible on social media. This trend is even amplified by the current global recession, bringing firms in cost-reduction mode. Firms generally apply social media for marketing efforts, customer

[2] Directive 95/46/EC, *Official Journal of the European Communities.* L 281/31, Article 2(a).
[3] Directive 95/46/EC, *Official Journal of the European Communities.* L 281/31, Article 2(b).

relations management, reputation management and to stimulate co-creation. For all these four aspects, social media engagement is a more efficient and inexpensive activity than the traditional channels.

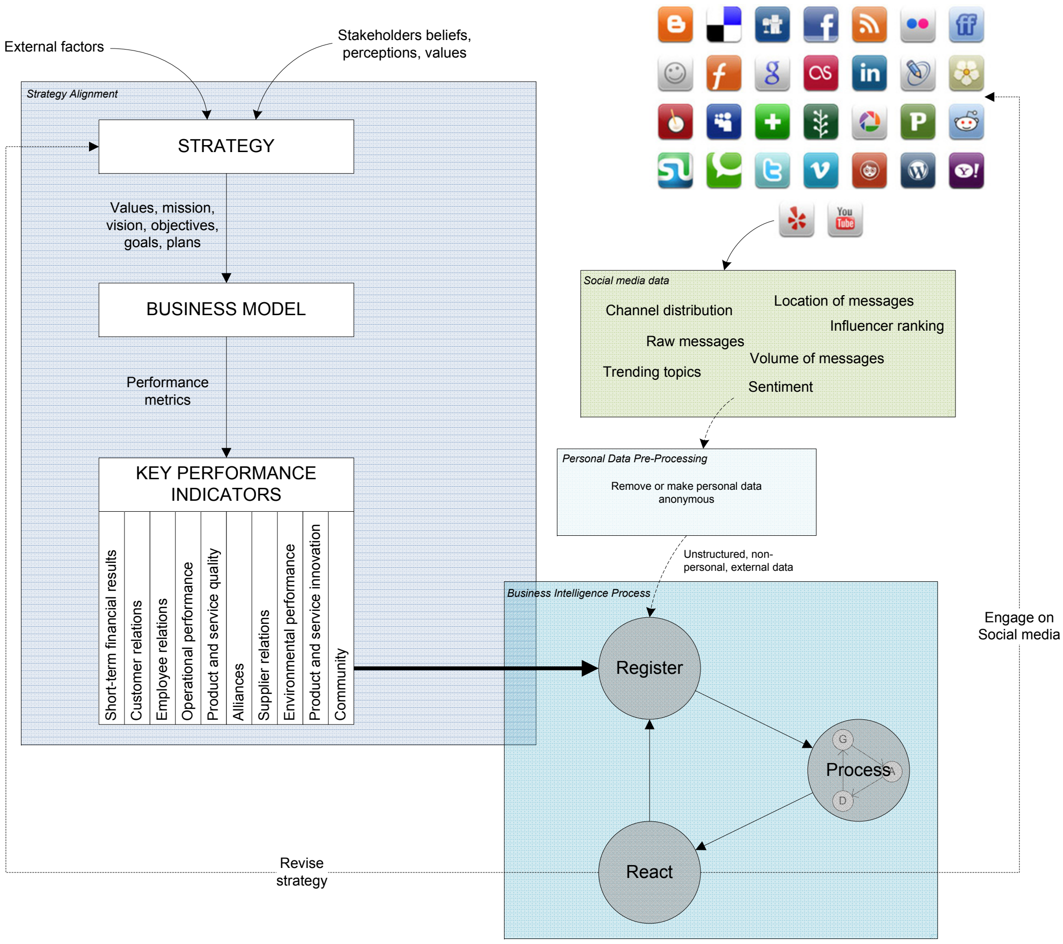


Figure 3-7: Strategy, Business Intelligence Process and Social Media Data

As with any activity that is performed in a professional organisation, the performance of the activity and its contributing value to the firm's overall objective is to be measured. Current social media monitoring tools – which are evolving rapidly – do not offer insight in the effects on organisational performance due to the organisation's social media undertakings. Rather, these tools provide the number of brand mentions on different platforms, the locations of where the posts were made, gender classification, language, (unreliable) sentiment analysis, etc. and thereby treating the firms' social media activities as a separate, isolated, business activity. It is therefore argued that firms need to establish a clear link between social media metrics (such as number of likes, shares, sentiment) and the firm's key performance indicators. Figure 3-7 illustrates the concept of social business intelligence, with the types of social media data flowing in the business intelligence process.

Chapter 4

Content Analysis

This thesis examines the applicability of social business intelligence for firms in different contexts. More specific, the applicability is investigated for firms in (i) different industries and (ii) for different customer relation types. The purpose of this chapter is to reveal differences in social media content related to different firms. As such, social media data related to different firms will be collected from social media platforms, after which these data are analysed on these two dimensions. As described in section 1-6-2, we will apply the content analysis procedure developed by Bos and Tarnai (1999) as a guidance for this analysis. A content analysis "entails a systematic reading of a body of texts" (Krippendorff, 2004), which is required to analyse the social media posts related to different firms. The structure of this chapter corresponds with the five step procedure, as shown in figure 4-1.

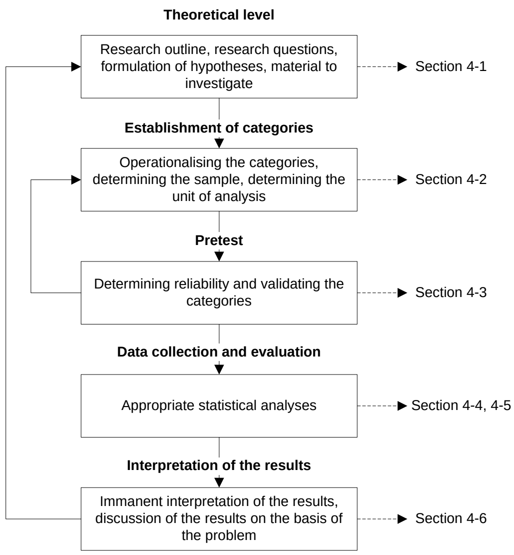


Figure 4-1: Content Analysis (Bos & Tarnai, 1999) and Outline Chapter 4.

Section 4-1 formulates the research questions that are to be answered by the content analysis, and it describes the material that will be investigated. Next, section 4-2 establishes the categories that are to be analysed, it describes the sample firms and the sample period. Thirdly, in section 4-3 a pretest is performed to validate whether it is possible to collect the data and classify it into one of the established categories. If necessary, the categories will be adjusted. Furthermore, section 4-3 describes the categories in more detail and presents the taxonomy – or coherence – of the categories. Fourth, in section 4-4 the data – social media messages related to different firms – is collected and the collection process is evaluated. In section 4-5, the data is analysed and visualised. Section 4-6 interprets the results of the content analysis. Finally, section 4-7 concludes the findings of this chapter.

4-1 Theoretical Level

The first step in the content analysis contains theoretic notions. Hypotheses and the material to be investigated are determined. Hypotheses are established to explicitly specify what will be researched in the analysis. The hypotheses and the material to investigate are described in this section.

4-1-1 Hypotheses Formulation

This thesis answers three sub research questions (see page 4), of which the second question will be answered in this chapter. From the research questions that are to be answered by means of the content analysis, hypotheses are formulated. These hypotheses explicitly state what will be investigated. In total, four hypotheses are formulated that correspond to the second research question of this thesis, i.e. *In which contexts are firms able to acquire social media data for business intelligence?* As discussed in chapter 1 we specify a firm's context on two dimensions. Firstly, a firm's context is described based on its customer relation type, consisting of either B2C or B2B. Secondly, a firm's context is specified by the industry in which it is active. The specification of the firm's context is incorporated in the hypotheses that are established in this section. Two groups of hypotheses are established; volume-related and subject-related hypotheses.

Hypotheses related to the ***Volume*** of Social Media Messages

The first topic of investigation relates to the volume of firm-related social media messages. Particularly, the volume of messages related to firms performing different customer relations are compared. Throughout this thesis, we distinguish two types of customer relations; B2C and B2B. As illustrated in section 1-2, it is expected that firms performing B2C relations are more often subject of discussion on social media than B2B firms. Firms performing B2C relations generally have more customers than B2B firms and are more visible to the end-consumer than B2C firms. We expect that these aspects influence the amount of social media messages related to a firm. The reason why the volume of social media messages related to different firms is important to investigate is the fact that social business intelligence will only be possible for a firm in case that there are actually messages created that are related to the firm. Hence, the first hypothesis is formulated as:

> H_1: *The volume of firm-related social media messages is higher for B2C firms than for B2B firms.*

Secondly, it is expected that the volume of firm-related messages differs among firms active in different industries. As illustrated in section 1-2, the nature of the products and/or services traded in different industries affects the rate at which products/services are sold. Therefore, we expect that firms in some industries are more often subject of discussion on social media than firms in other industries. For instance, retail products are more frequently bought by people than houses. Taken into account this rational reasoning, the second hypothesis is formulated as:

> H_2: *The volume of firm-related social media messages differs between industries.*

Whereas the customer relation types have yet been operationalised by two groups (B2B or B2C), the industry types have not yet been established. In section 4-2, the industry categories will be established based on a generic classification.

Hypotheses related to the ***Subjects*** of Social Media Messages

The third aspect that is investigated in the content analysis relates to the subjects of social media messages. Whereas the first two hypotheses provide insight in the existence of firm-related social media messages, this insight is not sufficient to draw conclusions on the applicability of social business intelligence. The subjects of social media messages are also to be included in the analysis. It is important to research the subjects of social media messages related to firms since the subjects of the messages are used to assign the messages to key-performance indicators. As such, the subjects contained in the messages determine – in combination with the volume – the applicability of social business intelligence for firms.

The subjects of firm-related social media messages are investigated on the same two dimensions as the volume of the messages. The contexts of B2C firms differ from B2B contexts. It is therefore likely that the subjects of the messages related to B2C firms differ from the subjects of B2B related messages. So far, we do not have strong signals that certain subjects are more often discussed in one customer category than in the other. In order to understand which type of firms can find messages related to different KPIs, the third hypothesis investigates whether or not the subjects of firm-related social media messages differ between firms performing different customer relation types. Hence, the third hypothesis is formulated as:

> H_3: *The subjects of firm-related social media messages differ between firms performing B2B and B2C relations.*

For similar reasons that are concerned with the second hypothesis, the subjects of messages related to firms in different industries are expected to vary. For example, it is likely that messages related to user experiences are more frequent created in an industry that creates electronic consumer products as compared to an industry that delivers consulting services. The subjects contained in social media messages – combined with the volume of these messages – affect the opportunities for social business intelligence for firms. Since this thesis examines the opportunities for social business intelligence for different firms, it is necessary to investigate variations in the subjects related to firms in different industries. Consequently, the last hypothesis is formulated as:

> H_4: *The subjects of firm-related social media messages differ between industries.*

The content analysis will be designed according to these hypotheses, and the results of the content analysis allow us to confirm or reject the four hypotheses. As such, we can draw conclusions on the applicability of social business intelligence for firms in different contexts.

4-1-2 Material to Investigate

The second step in the theoretical level consists of a description of the material to be investigated. In this thesis, we investigate social media posts that are related to firms. As described in chapter 3, different social media platforms serve different purposes, leading to different type of posts. A 140 character tweet has a lower information depth than e.g. a product review site. Since this research is exploratory in nature, it is valuable to gain as much understanding as possible from the content posted on different social media platforms. Therefore, the material to investigate is sourced from popular social media sites in Western Europe. The content in the dataset is sourced from Twitter, Facebook's public pages, Flickr, Newssites, Google+ public pages, (Wordpress) Blogs, Picasa, YouTube and Friendfeed.

4-2 Establishment of Categories

In the second step of the content analysis, the categories to be analysed are established and the sample set is determined. Depending on the research, the categories will differ. In this thesis, the categories that are to be analysed consist of firms in different industries and with different customer relations. Next, the content analysis of this thesis requires categories of social media posts to draw conclusions on the applicability of social media posts for business intelligence.

4-2-1 Operationalising the Categories

We examine differences in the volume and subjects of social media messages related to firms on two nominal dimensions; (i) *industries* and (ii) the *customer relation* type. The industry dimension is operationalised by categorising firms in different industries. The relation with end-users dimension is operationalised through means of a distinction between either Business-To-Business firms or Business-To-Consumer firms.

(i) Industry Classification

CBS (2012) – Statistics Netherlands – provides a hierarchical classification of economic activities, called SBI[1]. The European Union also has a classification, called NACE[2], on which SBI is based. SBI allows to classify firms based on their economic activities. SBI distinguishes multiple levels, of which the most aggregate level distinguishes twenty main activities. These activities are listed in table 4-1, and are engaged to classify the firms that have been selected in the analysis of this thesis.

[1] Standard Industry Classification, in Dutch *Standaard Bedrijfsindeling.*
[2] Statistical Classification of Economic Activities in the European Community, in French *Nomenclature statistique des Activités économiques dans la Communauté Européenne.*

Table 4-1: General Classification of Firms (CBS, 2012)

Industry	
A	Agriculture, forestry and fishery
B	Mining and quarrying
C	Industry
D	Production and distribution of and trade in electricity, gas, steam and air
E	Extraction and distribution of water, sewerage, waste management and remediation
F	Construction
G	Wholesale and retail
H	Transport and storage
I	Accommodation, meals and drink provision
J	Information and communication
K	Financial institutions
L	Real estate
M	Consultancy, research and other specialised business services
O	Public administration
P	Education
Q	Health and welfare
R	Culture, sport and recreation
S	Other services
T	Households as employers
U	Extraterritorial organisations and bodies

(ii) Customer Relation Type Classification

As illustrated in chapter 1, the type of customer relations is likely to have an effect on the availability of social media data related to a firm. Therefore, a category describing the type of customer relation is established. Based on Turban et al.'s (1999) classification of e-commerce, we classify firms in either Business-To-Business ("B2B") or Business-To-Consumer ("B2C"). This classification will provide insight in the availability of social media data related to firms based on the customer relation type.

Categories of Key-Performance Indicators

The key purpose of this thesis is to link social media posts to organisational performance. As described in chapter 2, firms measure organisational performance based on key-performance indicators. In order to draw generic conclusions of the applicability of social media data for the purpose of organisational performance, a generic classification of key-performance indicators is required. Section 2-2-5 illustrated that KPIs can be classified into ten categories. We will use these categories for the classification of social media posts based on the subjects of the messages. Table 4-2 shows the categories that are pursued in this research. In section 4-3-1 these categories are described.

4-2-2 Determining the Sample

One of the objectives of this research is to investigate possible differences in the user-generated content on social media related to different firms. In order to spot differences, our sample exists of firms that are active in different industries and take different positions regarding end-users. This section describes the sample and the industries that are part of the sample.

Selection of Firms

Based on the industry classification presented in table 4-1, eighteen firms have been selected. The starting point of the sample selection has been the list of firms that are part of the Amsterdam stock Exchange ("AEX"). The main reason for this selection criterion is the fact that these firms are stock listed, and hence publicise annual reports containing information about strategic initiatives, financial figures, etc. In case the analysis shows inter industry differences – e.g. between two comparable financial institutions – the annual reports may provide company specific information (e.g. amount of employees, attitude towards social media, etc.) clarifying these

Table 4-2: Categories of Social Media Posts

Category	Social media posts ...
Short-term financial results	related to the firm's financial performance
Customer relations	from individuals purposed to contact the firm, or from the firm purposed to contact an individual
Employee relations	related to employees of the firm
Operational performance	related to the firm's productivity, fact-based statements
Product and service quality	related to the experience of products and services
Alliances	related to joint-ventures or other cooperations
Supplier relations	related to the suppliers of the firm
Environmental performance	related to environmental / sustainability compliance
Product and service innovation	related to innovation
Community	revealing the community's perception of the firm (not purposed to contact the firm), chatter
Undefined	that could not be defined in one of the categories
Spam	that are not related to the firm

differences. If a sample containing privately owned companies would have been selected, access to additional information would have been limited. In addition, firms that are listed in the AEX are generally well-established, visible to the public and regularly subject to news articles. It is therefore expected that these firms are subject of discussion on social media. Furthermore, the firm that sponsors this research – KPMG – requested to apply the analysis on this list of corporations. Table 4-3 lists the selected firms, their corresponding industries and main customer relation. For a description of the individual firms and their activities, see appendix B.

In order to design a uniform sample, firms of over-represented industries have been replaced by non AEX firms which are also well-established corporations. The distribution of firms in the different industries is shown in the third column of table 4-3. The firms have also been classified based on their type of customer relation. Though the split between B2B and B2C is hard to make for some firms because they have B2C as well as B2B relations, the motivation for the classification of the firms is based on its main activities. The main activities, on which the industry classification as well as the customer relation classification is based, of each firm are described in appendix B. The final column of table 4-3 shows the distribution of B2B versus B2C firms in the sample.

Table 4-3: Sample

Firm		Industry	CBS	Relation
1	Akzo Nobel	Mining and quarrying	B	B2B
2	ArcelorMittal	Mining and quarrying	B	B2B
3	Unibail-Rodamco	Financial institutions	K	B2B
4	Arcadis	Consultancy, research and other specialised business services	M	B2B
5	Fugro	Consultancy, research and other specialised business services	M	B2B
6	Coca-Cola	Industry	C	B2C
7	Heineken	Industry	C	B2C
8	Philips	Industry	C	B2C
9	Albert Heijn	Wholesale and retail	G	B2C
10	Blokker	Wholesale and retail	G	B2C
11	C-1000	Wholesale and retail	G	B2C
12	KLM	Transport and storage	H	B2C
13	NS	Transport and storage	H	B2C
14	PostNL	Transport and storage	H	B2C
15	Bol.com	Information and communication	J	B2C
16	TomTom	Information and communication	J	B2C
17	ABN AMRO	Financial institutions	K	B2C
18	Aegon	Financial institutions	K	B2C

Description of the Industries

The sample consists of firms active in seven different industries. As indicated, the industry classification is based on CBS' (2012) Standard Industry Classification. A description of the industries that are part of the sample is presented in this section.

1. **Mining and Quarrying**
   The activities of firms in the *mining and quarrying* industry are concerned with the extraction of oil, gas and/or minerals such as sand, gravel and clay.
2. **Industry**
   Industry firms are producers of food, beverages, tobacco, textile, chemical products, pharmaceutical raw materials, metal, electric products, machines, cars, other transport modes, furniture and other products.
3. **Wholesale and Retail**
   The industry *wholesale and retail* consists of firms trading in cars, food, machinery, agricultural products, textile, books, and other consumer products. In addition, firms in this industry operate shops in which consumers can buy their products.
4. **Transport and Storage**
   The activities of firms that are active in the *transport and storage* industry transport persons or products across land, water, air or other transport modes. Next, firms in this industry store products. Also, mail related activities belong to the *transport and storage* industry.
5. **Information and Communication**
   Firms in the *information and communication* industry are publishers of books, papers, magazines, software and computer games. In addition, the production and distribution of films, music and television shows are assigned to the *information and communication* industry. Also telecommunication activities, whether through wires, wireless, satellite or other mediums are assigned to the *information and communication industry.* All software related activities required for telecommunications are part of the *information and communication* industry.
6. **Financial Institutions**
   The industry *financial institutions* consists of banks, financial holdings, investment institutions, insurance companies, pension companies, asset management companies and other financial firms.
7. **Consultancy, Research and Other Specialised Business Services**
   The activities of firms in the *consultancy, research and other specialised business services* relate to advisory services on different domains. Examples of firm types in this industry are law firms, accountancy firms, engineering firms, architects, marketing firms, research firms, etcetera.

4-2-3 Description of the Measuring Period

uberVU, one of the emerging social media monitoring and analysis tools has granted access to their tool for a period of 14 consecutive days, i.e. from Friday 20 July to Thursday 2 August. This tool is further described in section 4-4-2. During the measuring period the eighteen firms have been monitored, resulting in the collection of 224.687 social media posts related to different firms. This amount of messages is deemed sufficient to analyse what the subjects of social media messages related to firms are, and how these subjects differ in volume from each other, which is the purpose of this chapter. During 7 of the 14 days in the period, the 2012 Olympic Games took place. As a consequence, the social media posts of firms that are for some reason – e.g. as a sponsor – involved in the Olympic Games often have the Olympic Games as a subject of the post. It is common for firms to sponsor events. In case the social media messages would have been collected during another period, it is likely that some firms were sponsoring an event as well during the measurement period. Furthermore, the summer holidays took place during the measurement period. Though people undertake other activities during their holidays, and hence may show different activities on social media as well, it is not likely that large corporations – i.e. the ones in our sample – are not subject of discussion during this period. On the contrary, some firms will precisely be mentioned during this period. However, when interpreting the results after the analyses, one should be aware of the fact that the social media messages at which the conclusions are based have been created during a holiday period. Furthermore, conclusions related to the procedure of social business intelligence will not be affected by the fact that the data are created during a holiday period, since the way of collecting, processing and analysing the data will be the same in any period.

4-3 Pretest

In step 3 of the content analysis – see figure 4-1 –, the data collection is tested to ensure that the established categories can be filled with data from the selected sample. To validate the categories that were established to classify social media posts, we pretest the categories that are presented in table 4-2 by classifying the first 100 messages of each respondent. Each of the pretest social media posts have been read and consequently assigned to one of the categories. The sample illustrated that some of the categories were too generic. Therefore, we added sub categories to some of main categories to gain more insight in the nature of the social media posts. The categories of social media posts are discussed in the following section.

4-3-1 Categories of Social Media Posts

The purpose of this thesis is to assign social media posts to key-performance indicators. To achieve this, we apply the classification scheme of Ittner et al. (2003) to distinguish key-performance indicators from each other. This classification scheme distinguishes ten key-performance indicator categories. Basically, we adhere to these ten categories. However, as we have experienced in the pretest, social media posts within one category are too heterogeneous to simply assign the social media post to the high-level classification that distinguishes between ten categories. Therefore, an additional level of detail has been assigned to some of the main categories. This additional level of detail has been established based on the pretest of the categories. Thus, the empirical data has driven the establishment of these categories. The naming of these – more detailed – sub categories represent the nature of the social media posts as well as possible. In the following section we discuss the categories of the social media posts, which are based on the the KPI classification of Ittner et al. (2003).

1. Short-term financial results

The first KPI category consists of indicators representing the (short-term) financial performance of the firm. Financial indicators are typically measured by firms using internal systems. In other words, there is no external influence required to measure these metrics. Though the added value of the information of social media posts related to financial results may be of little value for the firm (management has financial results earlier available than the firm's environment), it is wise to classify these social media posts nevertheless to provide an as complete overview as possible of the type of social media posts that are available for firms. Typical examples of financial indicators are the number of sales in a certain period, amount of debt on a certain moment, operating expenditures in a certain period, etc. Social media posts that can be classified into this category relate to the firm's financial performance. As we have experienced in the pretest, social media posts related to financial aspects of a firm are either related to discussions of the performance of a firm, or related to the firm's share prices. Accordingly, these two categories have been added as sub categories. These two sub categories are discussed in the following sections.

1.1 Financial performance discussions

Social media are used to discuss the financial performance of a firm. Often, posts related to financial performance contain a factual statement of the firm's performance, which is sometimes followed by an opinion of the creator of the post. In addition, these type of social media posts often contain a hyperlink to a website at which the financial performance is further analysed. The following two example posts that consist in our sample illustrate the type of posts that are classified as *financial performance discussions* posts:

> *"Akzo Nobel Q2 Profit Takes 21.5% Hit on Restructuring Charges http://t.co/b1518IDQ".*

> *"STEEL RESULTS: #ArcelorMittal Flat Carbon #Europe reports Q2 earnings fall http://t.co/0nPq7VjM #steel".*

The existence of financial performance related discussions in our sample data is due to the selection of the firms in our sample. As discussed in section 4-2-2, the sample firms are based on stock listed companies. These companies are public limited liability firms, hence required to publicise their financial performance on a regular basis by law. Since the financial performance is publicly available, it makes that these firms' financial performance are subject of discussion on social media. In case that our sample would have consisted of limited companies, which are not required to publicise their financial position, we would probably not have found social media posts related to financial performance in the dataset.

1.2 Shares related discussions

Another substantial part of the social media posts related to the financial aspects of a firm has the firm's shares (prices) as subject. Often, these posts are made by analysts specialised in stock markets. The following two social media posts illustrate discussions related to the shares of a firm:

> *"TomTom: Ster van de week: TomTom maakte vorige week bekend navigatieproducten en -diensten te gaan leveren... http://t.co/8QRPxJmj #beleggen".*

> *"Stijgers 1: Wereldhave (4,05%) ; 2: TomTom (2,43%) ; 3: Ballast Nedam (2,25%) ; 4: VastNed Retail (1,94%) ; 5: Heineken (1,76%)".*

The character of the second example post is typically found in the dataset, it shows the top five funds of that day. These type of posts are generated automatically by computers, often referred to as 'bots'. For similar reasons with the financial performance discussions posts, we expect that shares related discussions are particularly found in our dataset because the firms in our sample are stock listed.

2. Customer relations

The second KPI category defined by Ittner et al. (2003) consists of performance indicators related to customer relations. As indicated in section 1-2, many researches acknowledge the opportunities for customer relationship management using social media. Not surprisingly, social media posts in our test sample could be assigned under the umbrella of *customer relations*. In our opinion, this category required a higher level of detail to distinguish the social media posts made by the firm's own web team from the posts that were directed towards the firm's web team. Seven categories have been added under the customer relations category to distinguish the nature of the social media posts made by or directed to the web care teams. From a BI perspective it is valuable to gain insight in the type of social media posts that are made by web care teams, since these posts may influence other factors like customer satisfaction, sales, or the costs related to customer relations.

The social media posts related to customer relations typically show a conversation between a firm and a customer. Social media posts related to customer relations represent a direction that is either from the customer to the firm, or from the firm to the customer. Social media posts from customer to customer containing a firm's name are categorised in a different category, which will be discussed later. The identification of social media posts from customer to firm, or the other way around, is relatively easy because people begin their message with the name of the receiver, preceded with an "@". As such, social media posts starting with e.g. @ABN AMRO have been classified as a customer to firm post. Social media posts that were made by a webcare team could also be easily recognised, because the creator of the message generally contains the name of the firm. To illustrate the direction of the customer relations posts, the naming of the categories represent the direction of the post.

2.1 Customer questioning the Firm

Social media are used by customers to ask questions. The nature of the questions differ, some people ask specific questions about a product or service, how to use it or how it's made while other questions are very broad and relate to the company's strategy or position in the market. Social media posts that were made by customers, directed towards a firm and illustrating a question have been classified under the category *questioning customer*. Figure 4-2 illustrates the direction of these posts; from customer to firm.

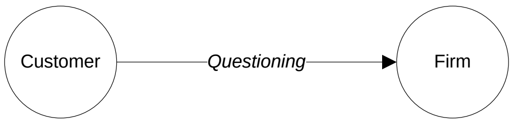


Figure 4-2: Questioning customer

Two example posts of the category *questioning customer* that have been found are illustratively shown below.

> *"@albertheijn Wat zijn de ramadanproducten? Ken je me die ff tweeten? :) alvast bedankt."*

> *"@KLM Hi, booked flights with you via @lastminute.com, wondering how we check in online? Saying that option isn't available?"*

Obviously, customer relationship teams that are active on the Web monitor social media posts in which customer ask questions that are related to the firm, and consequently respond to these questions. Often, a *questioning customer* post is followed by an *explaining firm* post.

2.2 Firm explaining the Customer

Clearly, one of the purposes of a web care team is to help customers with problems they experience. Many social media posts that are made by web care teams contain an explanation of problems or questions that customers posted on social media platforms. These posts have been classified as *explaining firm*, illustrating that the social media post has been written by a firm's web care team to explain a certain customer something in response to an earlier post made by the customer. Figure 4-3 illustrates the direction of the social media posts that have been categorised as *explaining firm*.

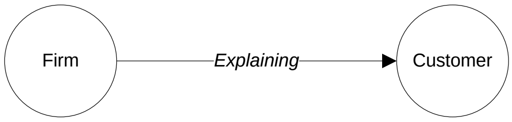


Figure 4-3: Explaining firm

Two example posts of the category *explaining firm* are shown below:

> *"@kiimberley94 Dag Kim. Als een bedrag dmv een automische incasso is afgeschreven, wordt het bedrag met max 2 werkdagen teruggeboekt. Elvira."*

> *"@mepe176 Bij geldautomaten met een Maestro-logo is dit zeker mogelijk. Sommige winkels bieden ook deze mogelijkheid. Suzanne."*

2.3 Customer complaining to the Firm

Customers employ social media as a means to complain. Plenty examples that have reached the newspapers in recent years exist. It is therefore not surprising that our data set shows social media posts that represent a complaint. Customers complain about product experiences, how they have been treated in their complaints procedure, etc. Social media posts that have been made by customers and illustrating a complaint have been classified into the category *complaining customer*. Figure 4-4 illustrates the direction of these social media posts; from customer to firm.

Figure 4-4: Complaining customer

Two example posts illustrating *complaining customer* posts are shown below:

> *"@ABNAMRO Blijkbaar is de enige manier om met jullie een probleem op te lossen om NIET TE BETALEN. Want telefonisch sta je in de kou! HELP"*

> *"@PostNL @PostNLWebcare de zoveelste keer dus dat de bezorgers de aangetekende stukken niet laten tekenen ...."*

It is the purpose of a firm's web care team to respond to the complaining posts. Therefore, a *complaining customer* post is often followed by an *understanding firm* or *complaining firm* post.

2.4 Firm showing feeling of understanding to the Customer

Next, as the sample data shows, a firm's web care team is also purposed to show a customer a feeling of understanding of his or her experienced problem. The social media posts that were made by a firm's web care team and represent a feeling of understanding with the customer's complaint have been classified into the category *understanding firm*. Figure 4-5 schematically shows the direction of these posts. Understanding firm posts differ from explaining posts since understanding firm posts do not offer the customer a solution to the experienced problem, but rather show a feeling of understanding.

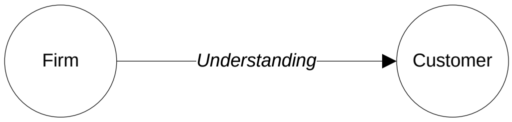


Figure 4-5: Understanding firm

Two example posts that were found in the test sample and clearly illustrate a sense of understanding of the customer's experienced problems are illustrated below.

> *"@normanwillems Dag Norman, vervelend te horen dat je reis niet doorgaat. Als je belt met 0900-0024 kunnen we je verder helpen. Margot."*

> *"@noni1967 Dag Nanette, dat is erg vervelend om te horen. Ik hoop dat het snel verwerkt wordt. Margot."*

2.5 Customer thanking the Firm

As discussed, web care teams are amongst others purposed to help customers with problems that they experience, thereby replacing the traditional telephone help desks. The social media posts in our sample clearly show a conversation, where a firm replies to a social media post made by a customer. Once a customer has been assisted by a company's web care team, some customers take the effort to thank a firm for their assistance. Social media posts that have been made by customers that are directed towards a firm and illustrating gratitude towards the firm, have been classified into the category named *thanking customer*. Figure 4-6 illustrates the direction of *thanking customer* posts; from customer to firm.

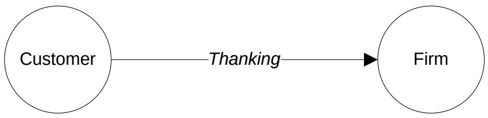


Figure 4-6: Thanking customer

Two example posts illustrating *thanking customer* posts are shown below:

> *"@ABNAMRO Ok, thanks voor de snelle reactie en fijne dag! :-)."*

> *"@albertheijn Oke bedankt! Dan ga ik Valkeniersplein proberen :)."*

2.6 Firm thanking the Customer

The sixth sub category that was added to the customer relations umbrella has been assigned the name *thanking firm*. As we experienced, web care teams often thank the customer for mentioning a deficiency of a product / service, or the web care teams thanks a customer for a compliment made on the side of the customer. Figure 4-7 illustrates the direction of these type of posts; from firm to customer.

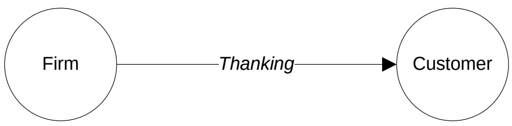


Figure 4-7: Thanking firm

Below, two example posts that were found in the sample and clearly present that the web care team is grateful towards the customer's earlier post are shown.

“@Birdy_Fly Bedankt voor de tip Tom, ik zal deze door gaan zetten naar de betreffende afdeling. Fijn weekend. Martijn.”

“@LindaWestenberg Bedankt Linda! Ik wens je een fijne dag toe. Elvira.”

2.7 Firm informing many Customers

Finally, firms also use social media to inform their customers on certain topics. Social media posts made by a firm and purposed to inform customers on a certain topic are classified into the category named *informing firm*. Whereas social media posts of the category *explaining firm* also inform customers, *informing firm* posts differ because they are directed to anyone. Figure 4-8 illustrates the “one-to-many” relation of the *informing firm* posts. Thus, *informing firm* posts are not specifically directed towards an individual and hence – as opposed to *explaining firm* posts – do not start with an “@”.

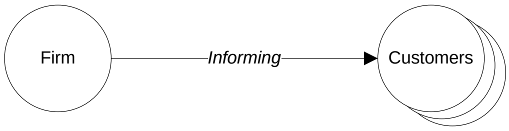


Figure 4-8: Informing firm

Two example social media posts found in the sample and classified as *informing firm* are shown below:

“#NS Deventer-Zutphen (overwegstoring) Tussen Zutphen-Deventer geen treinen door overwegstoring.Extra reistijd 30/60 min.Tot +/- 14:30... .”

“#NS Zwolle-Amersfoort: defecte trein: Tussen Amersfoort en Zwolle minder treinen.. Extra reistijd 15 - 30 min. (Tot +/- 22:00).”

3. Employee relations

The third category of key-performance indicators comprises indicators related to employee relations. In this thesis we seek for social media posts that relate to KPIs. Social media posts that are related to employee relations will be classified under this umbrella. However, as the pretest illustrated, there exists a variety in the social media posts that could be assigned to the employee relations category. Therefore, two sub categories have been established; *recruitment* and *employee posts*. These findings are in line with McCorkindale (2010).

3.1 Recruitment

The first sub category related to employee relations contains social media messages that involve employee recruitment processes. Our sample shows messages where people write about vacancies in a firm, students asking companies for an internship place, human resource managers wishing new employees a good start at their first day of work in the company, etcetera. Social media messages that are related to a company’s recruitment process have been classified into the *recruitment* category. Illustratively, two example posts of the category *recruitment* are shown below.

“#nieuwe #vacature: Medewerker Verkoopklaar / Rotterdam / Albert Heijn #VCW #banen http://t.co/bEDnvCFw.”

“Job opportunity: Deputy Program Manager - Trenchless Tech at ARCADIS - Washington D.C. Metro Area #jobs http://t.co/RRSw5RtV.”

3.2 Employee posts

The second sub category related to employee relations contains social media messages made by the firm’s employees. As our dataset illustrates, employees use social media to share their work experiences or indicate that they are working at the firm. Social media messages that have been made by employees have been classified under the category *employee posts*. Illustratively, two posts existing in our data set that have been classified as *employee posts* are shown below.

"Officially 10 years working @ ABN Amro Bank... Unofficially 13 years.. ."

"@PostNL Overuren niet betaald, geen bevestiging van gevraagde vakantie en fietsdeclaratie wordt niet uitbetaald. Lekker motiverend! #postnl."

4. Operational performance

The fourth main category of key-performance indicators relates to operational performance. Social media messages that relate to the operational performance of a firm are classified into this category. As discussed, category 2.3 contains social media messages in which users complain about the firm's product / or service. It is possible that customers complain about the firm's operational performance, for instance about the delivery time of a product. However, messages that have been classified into the *operational performance* category reflect facts, while customer-to-firm posts classified as *complaining customer* (category 2.3) are more subjective in nature and directed towards a firm. Below, two posts in our sample that have been classified as *operational performance* posts are shown.

"RT @ Webwereld: Derde keer in korte tijd storing ABN Amro http://t.co/v8Y953kC."

"Vrijdag 27-7 kaart verstuurd uit Assen, 27-7 gestempeld in Zwolle. 31-7 al aangekomen in Tynaarlo. Bravo @postnl. Niet gek voor 15 km."

5. Product and service quality

The fifth KPI category that firms apply contains indicators related to product and service quality. As our sample illustrates, customers share their product and/or service experiences through social media. Social media messages that represent product and service experiences of customers have been classified in the category *product and service quality.* Two posts existing in our sample and classified as *product and service quality* posts are shown below.

"Ik haat die Heineken met draaidop, altijd snij ik m'n hand er mee open."

"Ik had een albert heijn tas toen ik thuis kwam waren myn handen helemaal blauw."

6. Alliances

The sixth main category of key-performance indicators relates to the firm's alliances. Social media messages that are related to the firm's partnerships / alliances have been classified into the *alliances* category. Two posts that have been classified as *alliances* posts are shown below.

"In what has to be one of the strangest collaborations ever, military scientists from the UK's Defence Science and Technology Laboratory (DSTL) have been working with global paint and coating company AkzoNobel to develop an anti-chemical weapon paint that can absorb harmful chemicals from enemy... http://inhabitat.com/uk-military-develops-paint-that-absorbs-fallout-from-chemical-attacks."

"Op weg naar #Atrium MC om te spreken met nieuw bestuur #vereniging #artsassistenten en samenwerking met #ABNAMRO."

7. Supplier relations

The seventh main category of KPIs involves indicators related to a firm's supplier relationships. Since this thesis seeks to link social media data to KPIs, social media messages related to this category of KPIs have been classified into the *supplier relations* category. Apparently, suppliers of a company post social media messages indicating that they supply the firm with products / services. Illustratively, two social media messages existing in our sample which have been classified as *supplier relations* posts are shown below.

"Onze koks fietsen door het Holland Heineken House, alles loop op rolletjes! #hhh2012 http://t.co/DrbbbXwl"

"Bezig met een nieuwe klus, het #vormgeven van een #advertentie deze keer voor de #albertheijn #AH Valkeniersplein te #Breda."

8. Environmental performance

The next main category of key-performance indicators involves indicators reflecting the firm's environmental performance. Social media messages relating to the environmental performance of firms have been classified into this category. The dataset shows social media messages in which consumers discuss the environmental responsibility of the company. Below, two posts existing in our dataset that have been classified as *environmental performance* posts are shown.

> *"We kunnen heel veel bijdragen aan ontwikkelingen op het gebied van duurzaamheid... aldus Albert Heijn. Ze verkopen uien uit Australië #AH."*

> *"@GreenpeaceNL - olijfolie v @albertheijn zit tegenw. in plastic "samen meer doen voor het milieu" -> is t idd beter? http://t.co/2vLBda61."*

9. Product and service innovation

The ninth main category of key-performance indicators involves indicators related to product and service innovation. As illustrated in section 3-2-3, social media are used by consumers to share product experiences and to suggest innovations. The process of co-creation and prosuming is shown in our dataset as well. Social media messages that reflect people's attitude to new products or services or contain suggestions for innovations have been classified as *product and service innovations*, of which two posts are illustratively shown below.

> *"Best Reviews - Philips Sonicare HX6732/02 HealthyWhite R732 Rechargeable Electric Toothbrush - http://maxtodaystore.info/today-p... ."*

> *"Moe worden van #ABN-AMRO bank geld overmaken steeds weer die achterlijke reader nodig , neem een voorbeeld aan tan- codes van #ING!!"*

10. Community

The tenth and final main category of key-performance indicators related to indicators related to the firm's community. Social media posts belonging to the category *community* reveal how the community, that is, external actors, perceive the firm. Many social media posts in the pretest could be assigned under the community category. However, to provide more detail in the type of social media posts related to the community category, we established five sub categories to the *community* category. These sub categories are discussed in the following sections.

10.1 Promotion

Some social media posts in our sample were created by firms themselves, and are hence not perceived as user-generated content from a firm point of view. Social media messages that were made by firms themselves and were purposed to promote the firm to the environment, were assigned to the sub category named *promotion*. The following two posts illustrate *promotion* activities of a firm:

> *"What do you want to do more of in retirement? Travel, spend time with family, pursue hobbies, or more education? Check out the results from the quick poll here!"*

> *"Albert Heijn - Kom op 18,19,25 en 26 aug. naar de Open Dagen van onze boeren en telers. http://t.co/Saiv2aWY http://t.co/bmP1CePw."*

10.2 News

The test sample illustrated that many social media posts are (simply) notifications of news articles. We positioned these posts under the category *community* since *news* messages determine the firm's exposure to the firm's community. Posts belonging to the sub category *news* are written by professionals, which are mostly journalists promoting their news article. Below, two examples of *news* categorised social media posts are shown.

> *"En verder in de serie vakantiebaantjes vandaag: Gerrit Zalm, waarmee verdiende de baas van #abnamro zijn eerste centjes? #BNR."*

> *"@huizenprijzen: Han de Jong (Chief Economist ABN AMRO) : "Er is weinig in dit leven zo gevaarlijk als schuld" : http://t.co/VOTreLdg via @youtube #schuld."*

10.3 Public image

Thirdly, the sample test illustrated that individuals share their attitude towards a firm through means of social media. We classify these posts into a sub category named *public image*. Social media posts classified into the *public image* category are not directed towards a firm, or, not purposed to get in contact with the firm. Rather, social media messages assigned in the *public image* category represent discussions and "chatter" amongst the social media users, in which the topic of discussion is the firm or its products / services. *Public image* posts are written by non-professionals, while – as we will see in the next sub category – posts created by professionals are assigned into a separate sub category. Below, two example posts that were found in the sample set and represent the public's image towards the firm are shown.

> *"Even kijken of blokker kruimeltje de film heeft liggen want ik heb m alleen nog maar op video band!"*

> *"Je moet staatsbanken ABNAMRO en ING 15% betalen als je rood staat op je betaalrekening en je krijgt 2% als je + staat #Schurkenbanken."*

10.4 Professionals

Fourth, our pretest indicated that there are social media messages created by professionals. Messages that have been created on social media by external professionals – not from the company – and talking about the firm have been classified into the category labelled *professionals*. Two example posts of this category are:

> *"Presentatie @ jaccooudhof van #KPMGmkb op de "Kengetallenbijeenkomst" van @ FullFinance @ ABNAMRO en NOVAK"*

> *"Sarah Harding interviews Arcadis at the A&WMA conference in San Antonio."*

10.5 Distributors

Finally, the data showed that social media are also used to promote products. However, the firm that produces does not have to be necessarily the one that promotes the product. Our dataset contains social media posts made by distributors of the product. These posts have been classified into the category labelled *distributors*. Two example posts of *distributors* are shown below.

> *"Macco Akzo Nobel Pai DWP24 Liquid Nails Drywall Construction Adhesive: Specially formulated latex product for in... http://t.co/Vk3PdXLR"*

> *"Best Offer - Philips Norelco AT830 PowerTouch Rechargeable Cordless Razor, Gray/Silver/Black - http://maxtodaystore.info/weekly-..."*

Undefined

Based on the 26 social media post categories (including main categories) that have been established in the previous sections, we can not classify *all* social media posts. As discussed before, social media data is unstructured and the interpretation of a social media post is not always easy. Therefore, messages that – despite of the mentioning of the firm's name in the post – could not be assigned to one of the categories have been assigned into the *undefined* category. Two examples of posts that were *undefined* are:

> *"Volg ons (Unicum) @ ABN AMRO bij zuidplein om 13 : 00 !!!!! RT RT."*

> *"Nu naar fietsenwinkel, blokker en c1000 met pap en mam!"*

Spam

Unfortunately, the search queries that were used to scrape the social media content did still result in the collection of data that is totally unrelated to the firm. This is due to the fact that people's names or IDs are similar to the firm's name. Social media posts that were totally unrelated to the firm have been classified as *spam*, of which two examples are shown below:

> *"@Klm_babe Okay well maybe sometime next week then :)"*

> *"@Jack_Heineken meen je dat nou? -_- :p een korte broek aan naar de zaak :p"*

4-3-2 Revised Taxonomy of Categories

The categories of the social media posts are based on the KPI classification scheme of Ittner et al. (2003), which allows classification of performance metrics into one of the ten categories. Our addition of sub categories does not affect the structure of Ittner et al.'s (2003) classification scheme, but rather adds a layer of detail to the categories. An overview of the revised taxonomy – after addition of the sub categories – and a short description of the social media posts of the corresponding categories is presented in table 4-4. Figure 4-9 schematically shows the taxonomy of the key-performance categories and the social media post categories.

Table 4-4: Taxonomy of Categories of Social Media Posts

KPI Category		Social media posts ...
1.	**Short-term financial results**	**related to the firm's financial performance**
1.1	Financial performance discussions	related to the firm's financial performance
1.2	Stock related discussions	made by professionals/individuals analysing the firm's stock price
2.	**Customer relations**	**from individuals purposed to contact the firm**
2.1	Questioning customer	posts from a customer asking a question to the firm
2.2	Explaining firm	from the firm purposed to explain the customer something
2.3	Complaining customer	from a customer complaining about the firm / firm's products or services
2.4	Understanding firm	from the firm purposed to show the customer shared understanding
2.5	Thanking customer	from individuals purposed to thank the firm
2.6	Thanking firm	from the firm purposed to thank the customer for an earlier post
2.7	Informing firm	from the firm purposed to inform customers (not responding to an individual)
3.	**Employee relations**	**related to employees of the firm**
3.1	Recruitment	related to recruitment of new employees
3.2	Employee posts	made the firm's employees
4.	**Operational performance**	**related to the firm's productivity**
5.	**Product and service quality**	**related to the experience of products and services**
6.	**Alliances**	**related to joint-ventures or other cooperations**
7.	**Supplier relations**	**related to the suppliers of the firm**
8.	**Environmental performance**	**related to environmental compliance**
9.	**Product and service innovation**	**related to innovation**
10.	**Community**	**revealing the community's perception of the firm (not purposed to contact the firm)**
10.1	Promotion	made by the firm for promotion activities
10.2	News	made by external professionals (journalism)
10.3	Public image	made by non-professionals, individuals ('chatter')
10.4	Professionals	made by professionals talking about the firm
10.5	Distributors	made by distributors of the firm's product/service
	Undefined	**that could not be defined in one of the categories**
	Spam	**that are not related to the firm**

4-4 Data Collection and Evaluation

The fourth step of the content analysis comprises the data collection and analysis. The purpose of this section is to collect social media posts related to the firms of the sample, and to analyse these data to identify differences in the content related to different firms. Moreover, the experiences that we encounter in the data collection

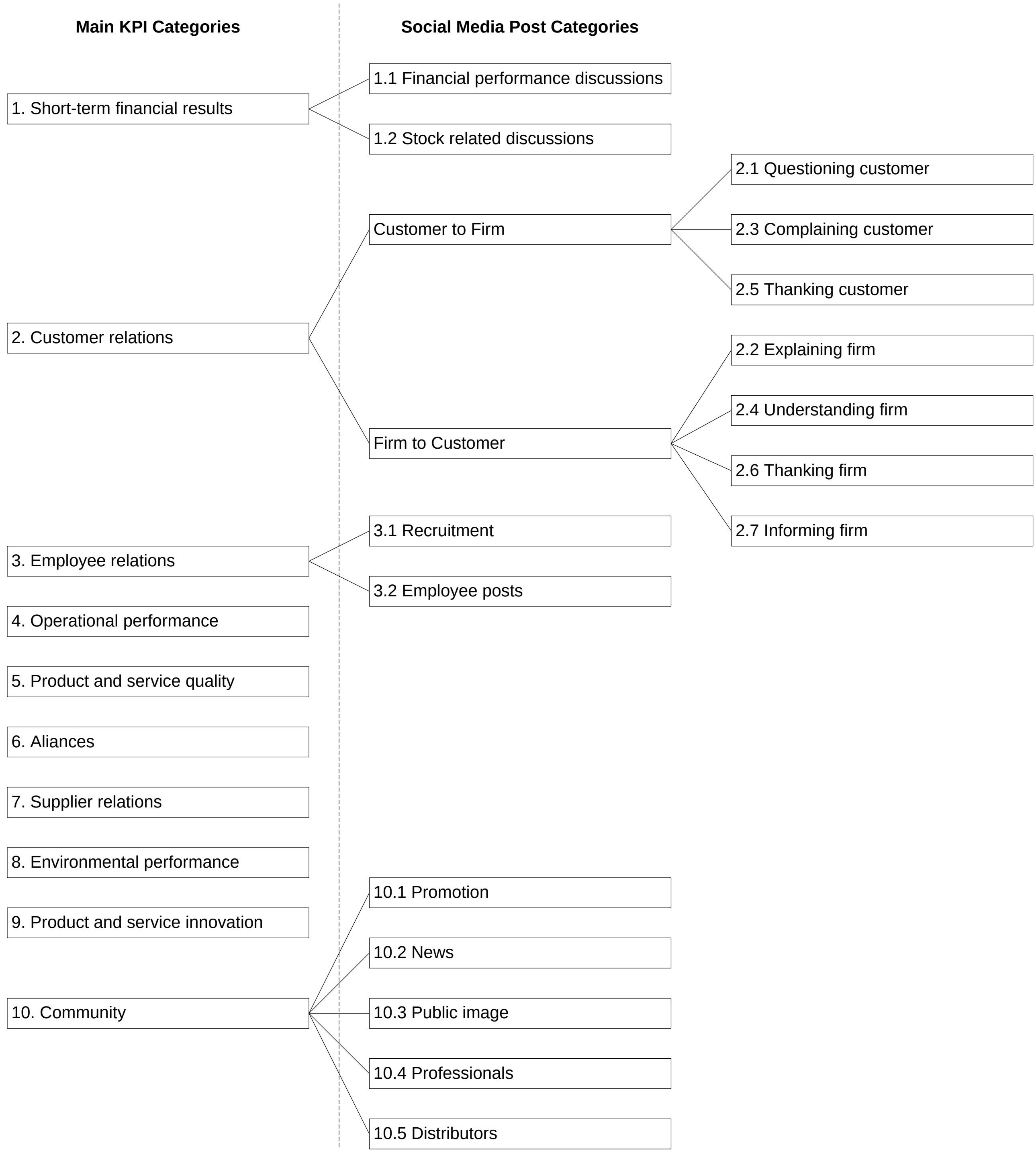


Figure 4-9: Taxonomy of Social Media Post Categories
The figure illustrates the ten main categories of key-performance indicators that have been found in the literature. Additionally, sub categories have been established at which the social media messages could be assigned. These additional categories have been constructed based on the empirical data.

phase as well in the data analysis phase will serve as a baseline in formulating requirements for a social business intelligence procedure that we develop in a later stage of this thesis.

Watson and Wixom (2007) illustrate that data collection requires about 80% of the "time and effort" related to business intelligence, and that data collection is responsible for "50% of the unexpected costs" in BI projects. Social media platforms are new sources for firms to collect data. The experiences from the collection of social media data for this research contain valuable lessons learned for firms willing to utilise social media data for business intelligence. Therefore, we pay much attention to describing the steps that are necessary to collect social media data.

In section 4-4-1 we discuss the search queries that are used to filter out those social media posts that relate to the firms in our sample. Section 4-4-2 describes how the social media posts have been extracted from the web, and how these posts have been placed in a database allowing to be analysed. Although we will use proper search terms, it is expected that many social media posts contain unrelated information. Therefore, the data will be cleaned in section 4-4-3. Once the data is cleaned, section 4-5 provides descriptive statistics about the amount of social media posts available for firms in different industries and for different positions regarding end-users.

4-4-1 Search Terms

As on the regular web, search terms are used on social media to filter out information that is the subject of interest. In social media, and especially on Twitter, users place a # ('hashtag') before a word to indicate the subject of the particular social media post. Hashtags can be considered as meta data tags indicating the subject of the social media post. When each user uses the same hashtags about a certain topic, it becomes easy to track the stream of social media posts related to that subject. We will use the strength of hashtags to filter out the social media posts that are related to the firms in our sample. Another widely used symbol to indicate that a social media post is direct to a person, or a firm, is the @ ('at symbol'). As with the hashtag, this symbol is positioned before one's (nick)name to illustrate that a post is direct towards this person or organisation. We will use the at symbol in our search terms, because social media posts containing this symbol are directed to a receiver, the firm.

Additionally, because not everyone and not each social media platform adheres accurately to the usage of hashtags, we add a search term containing the name of the firm without a hashtag to the search terms. Next, because some firms have name that can be written in multiple forms, we also search on different names. An overview of the search terms used to filter out the social media posts that are related to the firms in our sample is presented in table 4-5.

Table 4-5: Firms and Corresponding Search Terms

Firm		Search Terms
1	ABN AMRO	#abnamro, #abn amro, @abnamro, @abn amro, abnamro, abn amro
2	Aegon	#aegon, @aegon, aegon
3	Akzo Nobel	#akzonobel, #akzo nobel, @akzonobel, @akzo nobel, akzo nobel, akzonobel
4	Albert Heijn	#albertheijn, #albert heijn, @albertheijn, @albert heijn, albertheijn, albert heijn
5	Arcadis	#arcadis, @arcadis, arcadis
6	ArcelorMittal	#arcelor mittal, #arcelormittal, @arcelor mittal, @arcelormittal, arcelor mittal, arcelormittal
7	Blokker	#blokker, @blokker, blokker
8	Bol.com	#bol.com, @bol.com, bol.com
9	C-1000	#c1000, @c1000, c1000
10	Coca-Cola	#coca-cola, #cocacola, @coca-cola, @cocacola, coca-cola, cocacola
11	Fugro	#fugro, @fugro, fugro
12	Heineken	#heineken, @heineken, heineken
13	KLM	#klm, @klm, klm
14	NS	#ns, @ns, ns
15	Philips	#philips, @philips, philips
16	PostNL	#postNL, @postNL, postNL
17	TomTom	#tomtom, @tomtom, tomtom
18	Unibail-Rodamco	#unibail-rodamco, @unibail-rodamco, unibail-rodamco

4-4-2 Scraping Social Media Content

Web scraping – also known as web crawling – is the excavation of data from web pages into a local structured database, so that these data can be analysed (Huang, Li, Li, & Yan, 2012). Figure 4-10 visualises this process. The web scraper is provided with keywords, so that it can detect those particular web pages or social media posts related to the topic of interest. All content that fulfils the keywords are consequently stored into whatever form the person prefers, which is often a database, a web page, another application or – as in this thesis – a spreadsheet. Scraping web pages allows one to extract that particular information from the web that one is the topic of interest, and consequently process the data for its own purpose.

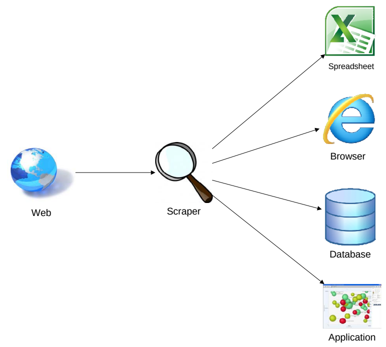


Figure 4-10: Web Scraper
Content from the web (e.g. a social network site) is filtered, extracted and stored to into different types of media, e.g. in a database, or a spreadsheet.

Software tools purposed to monitor social media are emerging rapidly. uberVU is one of such tools that are available on the market. In addition to social media monitoring uberVU allows the extraction of social media posts into comma-separated value ("CSV") format, and is therefore regarded as a social media platform scraper. It is this aspect of the tool that was decisive for the selection of a social media extraction tool that we used to scrape the content. uberVU was established in 2008, their software is amongst others used by NBC, Microsoft, Audi, Nestle, T-Mobile, Thomas Cook, 3M, PayPal, BASF and The World Bank. The software indexes multiple social network platforms, including Facebook, Twitter, YouTube, Flickr, Vimeo, Picasa. In addition, traditional media like news sites and (Wordpress) blogs are monitored. Consequently, the software presents metrics including number of mentions over the last period, number of likes, number of shares, platform distribution, sentiment of the online posts, gender distribution, language of the posts and the countries where the posts originated.

The eighteen selected firms have been monitored for a period of 14 days using search queries based on keywords containing the name of the firms. Please see table 4-5 for the list of keywords used to filter out content from the social media platforms. As such, all posts that were publicly available have been scraped from the social media websites. uberVU allowed the exportation of maximum 10.000 posts in CSV format per request. Therefore, the firms were subjected to a request on a daily basis. As a consequence, the search request on day t contained content that existed yet in the search request of day $t-1$. Figure 4-11 illustrates the overlap in the scraped content. Before the individual daily search requests were consolidated, the records that existed were removed. The CSV exportation has been executed on a daily basis, and each addition to the database (except for the first) resulted in the notion that there existed yet certain posts in the database. Therefore, we can conclude that the social media messages in our database provide a complete overview of the messages created in the measuring period and related to the sample firms.

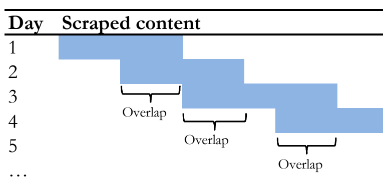


Figure 4-11: Overlap in Scraped Content
The figure illustrates that the daily search runs for new social media messages resulted in content that did yet exist in the database. Messages that existed in the database were removed.

When scraping the social media posts in the database, the following attributes of the posts were recorded: date, platform, username, content, language, sentiment, gender, followers, profile, country, region, city and URL to the post. The URL is the only attribute that is unique for a post, and is used to determine whether or not a post existed yet in the database, before it was recorded. It is common that social media messages are copied from one platform to the other. With our scraping method, identical social media messages created on two platforms are regarded as two individual posts, and therefore consist two times in our database. Table 4-6 shows a cross-cut of one of the monitored firms, showing one scraped post and the corresponding attributes. As illustrated, the scraper is not able to determine all attributes of a posts. For the example post presented in table 4-6, the tool was unable to determine the sentiment while is it very easy for a Dutch speaking person to determine that the content is obviously negative. Probably, this is due to the fact that the content is written in Dutch, while meta data required to determine the sentiment of the post about this language is not (yet) available. Also the country and the region are unknown, which is due to the fact that the user that has written the posts did not agree to share his or her location with to the social media platform.

Table 4-6: Example of Scraped Data

Attribute	**Example**
Date	23-7-2012
Platform	twitter
Username	Anne
Content	@ABNAMRO De internet site doet het nog steeds niet... Ik kan dus geen geld overmaken nu. Dat is mijn probleem nu.
Language	dutch
Sentiment	unknown
Gender	f
Followers	S
Country	unknown
Region	unknown
Profile	http://twitter.com/AnneXD_
URL	http://twitter.com/AnneXD_/statuses/227776320254382080

4-4-3 Data Cleaning

Before the social media data is ready for analysis, it requires cleaning. Though we used proper search terms, not all these posts actually relate to the firms. The search terms used to monitor the selected companies resulted in posts that did not have any relation with the selected companies. For instance, Twitter user names like @klm_klm_klm, @KLM_350, @KLM_2013, @KLM_babe and @klm_luvsya existed in the dataset belonging to KLM, though these Twitter accounts do not have any relation with KLM (the company). This so-called 'noise' has been classified as spam. The existence of noise in datasets is especially applicable on social media data. Therefore, any organisation that using social media data should filter out the valuable content from the noise.

4-5 Descriptive Statistical Analysis

This section describes the statistics that are acquired by the collection of the social media messages. More specifically, three topics are discussed in the following section. First, the distribution of the sources of the social media messages are presented (section 4-5-1). This distribution will reveal – taken into account the publicly accessible social media posts – which social media platforms are mostly used by customers to discuss firms and firms' products and services, hence providing firms insights in 'where to look' for firm-related social media content. Second, the volume of firm-related social media posts are examined (section 4-5-2). The volume of these messages is analysed per industry. As such, we gain insight in the amount of user-generated content that is created in different industries. The average daily mentions are also analysed per customer relation type. The final topic that is discussed in this section describes the statistics of the classified social media posts into categories (section 4-5-3). These categories are related to different sort of KPIs. Therefore this analysis will reveal which sort of KPIs are likely to be influenced by social media activities.

4-5-1 Channel Distribution

The social media messages have been collected from a variety of social media platforms. Figure 4-12 presents the distribution of all collected posts along the social media channels. As can be concluded, the largest share

(83%) of the collected posts have been created on Twitter. These findings are in line with A. N. Smith et al. (2012). While Facebook and other social media channels are responsible for a much smaller portion of the posts according to this dataset, one should place a note to these data. Facebook profiles may namely be set unattainable for non-friends. Therefore, the scraper – like any other web scraper – was unable to extract data from private profiles. Although one can argue that this distribution may provide an incomplete view of the situation, it is representative for a real situation in which a firm would collect social media posts from the popular sites because it holds also for a firm that it cannot access private social media profiles of people.

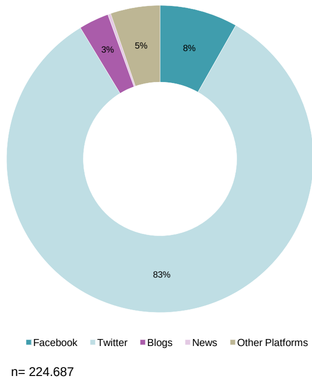


Figure 4-12: Social Media Channel Distribution
Illustrates the sources of the social media messages in our sample.

The channel distribution differs per firm, as shown in table 4-7. For each firm, the table shows from which channels the messages have been collected. As can be concluded, it holds for all firms that the majority of the publicly accessible messages are created on Twitter. The table shows one remarkable value, the collected social media posts of ABN AMRO are for 55% created from Picasa, a photo sharing platform. This figure can be perceived as a one-time event, because ABN AMRO has uploaded pictures from marketing events (KLM Open, World Tennis Tournament) that the firm organises to their Picasa profile. The web scraper recognised each individual picture as a separate social media post. Appendix C shows the channel distribution on a firm to firm basis graphically.

Table 4-7: Social Media Channel Distribution
Shows the absolute number of messages that have been collected from the various platforms. Furthermore, it shows the percentages of the platforms from which the messages have been collected.

Platform	**Facebook**		**Twitter**		**Blogs**		**News**		**Other**		**Total**
	Abs	%	Abs	%	Abs	%	Abs	%	Abs	%	Abs
ABN AMRO	124	2%	3.000	42%	70	1%	15	0%	3.858	55%	7.067
Aegon	110	8%	1.173	81%	79	5%	20	1%	67	5%	1.449
Akzo Nobel	30	3%	806	87%	43	5%	25	3%	18	2%	922
Albert Heijn	328	3%	11.116	96%	77	1%	1	0%	59	1%	11.581
Arcadis	8	2%	422	93%	9	2%	10	2%	6	1%	455
ArcelorMittal	439	8%	4.569	83%	296	5%	89	2%	139	3%	5.532
Blokker	155	6%	2.526	91%	71	3%	3	0%	14	1%	2.769
Bol.com	472	8%	5.124	89%	115	2%	-	0%	71	1%	5.782
C-1000	362	3%	10.583	96%	81	1%	5	0%	33	0%	11.064
Coca-Cola	1.653	5%	29.347	89%	999	3%	69	0%	885	3%	32.953
Fugro	6	1%	385	90%	20	5%	15	4%	2	0%	428
Heineken	5.726	15%	32.332	82%	494	1%	122	0%	751	2%	39.425
KLM	2.316	9%	22.601	86%	617	2%	90	0%	740	3%	26.364
NS	703	12%	4.970	85%	103	2%	-	0%	87	1%	5.863
Philips	4.641	12%	26.260	68%	3.404	9%	138	0%	4.007	10%	38.450
PostNL	77	6%	1.207	91%	27	2%	-	0%	12	1%	1.323
TomTom	1.308	4%	29.787	91%	630	2%	67	0%	956	3%	32.748
Unibail-Rodamco	3	1%	487	95%	9	2%	12	2%	1	0%	512
Total	18.461	8%	186.695	83%	7.144	3%	681	0%	11.706	5%	224.687

4-5-2 Volume of Firm-Related Social Media Messages

As described in chapter 1, this thesis analyses the availability of user-generated social media content on two dimensions. These dimensions are *customer relation* type and *industry* type. The availability of user-generated content has empirically been measured during a period of time. In this thesis, the variable called *average daily mentions* serves as a measure to describe the amount – or volume – of generated firm-related user-generated social media content.

The eighteen firms have been monitored for a period of two weeks. Each time that a social media post that contained one of the firms' names and that was publicly accessible has been downloaded. Some firms were mentioned more extensive than others. As such, it is possible to gain insight in the number of social media messages that are daily generated on the web per firm. Table 4-8 shows for each firm how many posts have been collected. The final column shows how many times – on average – the firm has been mentioned on a daily basis. The average daily mentions for each firm i have been calculated based on formula 4-1.

$$Average\_Daily\_Mentions_i = \frac{Total\_Collected\_Posts_i}{Measured\_Days_i} = \frac{Total\_Collected\_Posts_i}{MAX\_Date_i - MIN\_Date_i} \tag{4-1}$$

Table 4-8: Average Daily Mentions per Firm
The second column of the table indicates the total amount of messages that have been collected in relation with the firm. In the third column, this number is divided by the number of days at which messages have been found, hence representing the average daily mentions of the firms.

Firm		Collected posts	Average daily mentions
1	ABN AMRO	7.067	544
2	Aegon	1.449	111
3	Akzo Nobel	922	77
4	Albert Heijn	11.581	965
5	Arcadis	455	38
6	ArcelorMittal	5.532	461
7	Blokker	2.769	231
8	Bol.com	5.782	482
9	C-1000	11.064	922
10	Coca-Cola	32.953	2.996
11	Fugro	428	36
12	Heineken	39.425	3.285
13	KLM	26.364	2.197
14	NS	5.863	489
15	Philips	38.450	2.958
16	PostNL	1.323	102
17	TomTom	32.748	2.519
18	Unibail-Rodamco	512	39
Σ		224.687	

Volume per Firm

As can be concluded from the final column in table 4-8, the average daily mentions differs from firm to firm. From this table, we can conclude that the available user-generated content differs from firm to firm, and that the applicability of social media data for business intelligence purposes will not be possible for all firms, since not for each firm UGC is generated. Figure 4-13 illustrates the average daily mentions of different firms in our sample. The figure as been ordered from highly mentioned firms to less mentioned firms. In the following paragraph, the volume of social media messages is investigated from a customer relation type perspective.

Volume per Customer Relation Type

Our sample consists of a mix of firms that pursue a B2C or a B2B relation. One of the objectives of this thesis is to investigate whether and to what extent B2C firms are more often subject of discussion on social media than B2B firms. With the collected data we can analyse this topic.

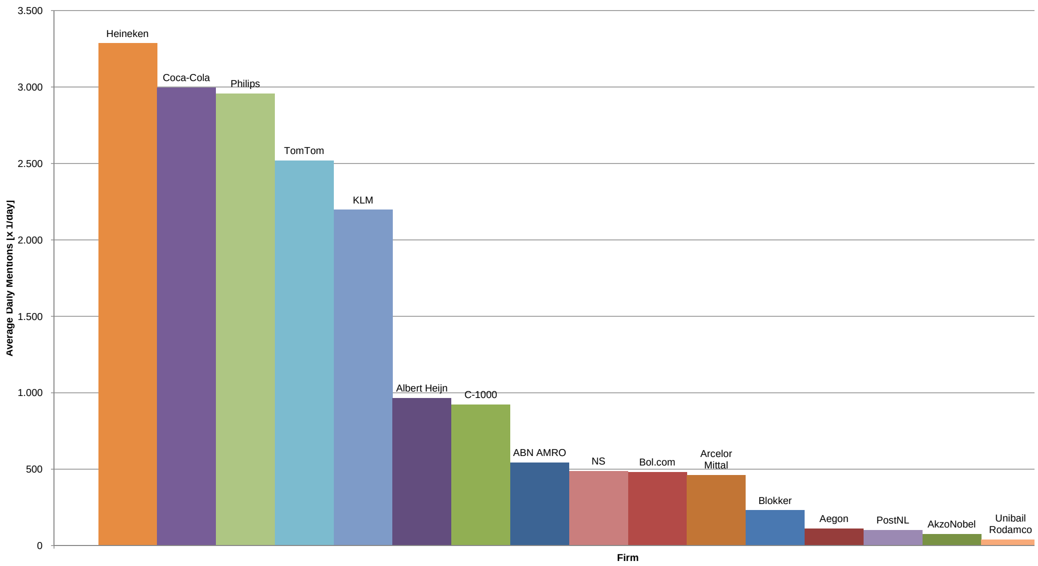


Figure 4-13: Average Daily Mentions of Firms
Bar chart illustrating the variation in the volume of firm-related social media posts. Firms have been ordered descending.

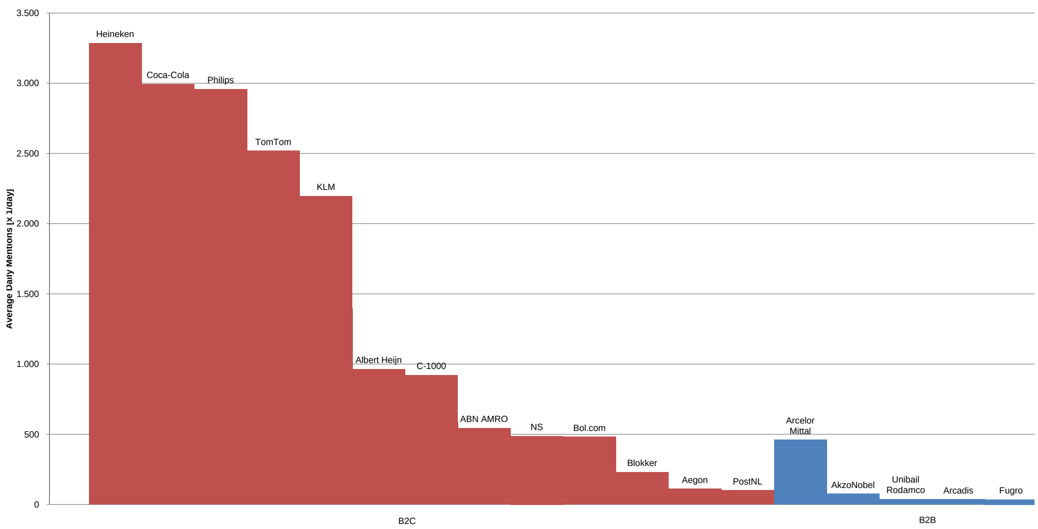


Figure 4-14: Average Daily Mentions of Firms
An overview of the average daily produced firm-related social media messages. Firms have been clustered based on their customer relation type.

Figure 4-14 shows the daily volume of firm-related social media content, in which the firms are clustered on their relation type and consequently ordered descending. This figure suggests that B2C firms – coloured in red – are more likely to find social media content that is related to their firm than firms performing B2B relations (coloured in blue). Figure 4-14 shows one remarkable value; the average daily mentions of *ArcelorMittal*. When analysing the content of the messages related to this firm, the explanation is discovered. *ArcelorMittal* has constructed the belvedere for the Olympic Games, called the ArcelorMittal Orbit. During the measurement period, the tower has been opened for the public, leading to discussions on social media.

In table 4-9, the average daily mentions of B2C firms have been consolidated, as are the B2B firms. Thus, the final column of table 4-9 presents an average of an average. Hence, the values are normalised and thereby eliminating the fact that the number of respondents differs between the two groups. The first hypothesis formulated at the beginning of this chapter was:

H_1: *The volume of firm-related social media messages is higher for B2C firms than for B2B firms.*

Table 4-9: Average Daily Mentions
The table consolidates the messages of all firms operating the same customer relation type, i.e. B2C or B2B. The final column illustrates the average daily mentions of an individual firm operating either a B2C or B2B relation.

Customer relation	Collected posts	Average daily mentions per firm
B2C	216.838	1.369
B2B	7.849	130
Σ	224.687	

With the figures presented in table 4-9, a bar chart is created in order to draw conclusions with respect to the first hypothesis. Figure 4-15 depicts the average daily volume of firm-related social media messages, consolidated per customer relation type.

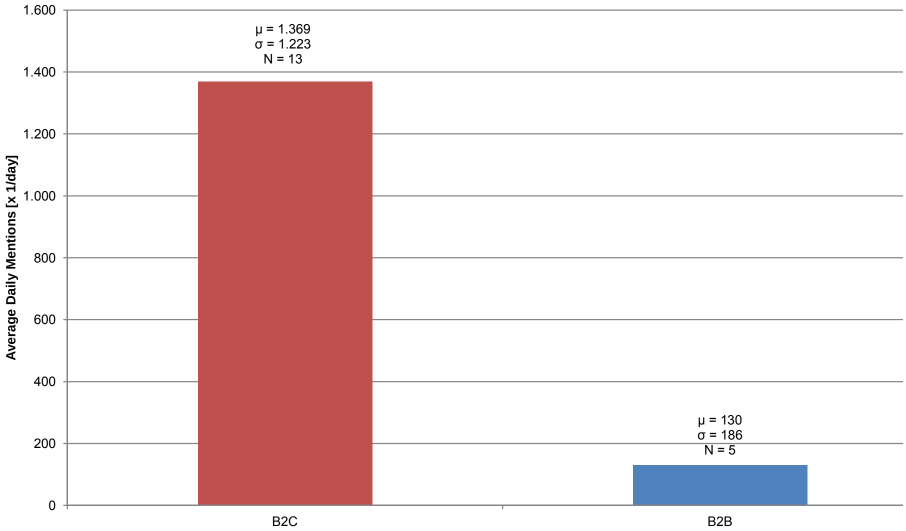


Figure 4-15: Average Daily Mentions of Firms
Bar chart illustrating the average daily volume of firm-related social media messages. Firms have been consolidated per customer relation type.

Figure 4-15 strongly suggests that B2C firms are far more often subject of discussion on social media sites than B2B firms. Thus, the results of our content analysis strongly suggest that the first hypothesis is to be accepted, implying that the volume of firm-related social media messages differs for performing B2B or B2C relations, with B2C firms being highly more mentioned on social media than B2B firms.

Volume per Industry Type

The second dimension on which the volume of firm-related social media content is investigated relates to *industries*. Our sample consists of eighteen firms active in seven different industries, see table 4-3 for an overview. As a first step to identify possible differences in the volume of daily messages between industries, the firms have been clustered on industry type in figure 4-16, and have consequently been sorted in descending order.

Figure 4-16 suggests that there exists a difference in the amount of user-generated content between different industries. Therefore, the different volumes are consolidated per industry, and analysed in the following paragraphs. Furthermore, figure 4-16 reveals that while an industry average may be lower than the average of an other industry, an individual firm may still be mentioned higher than a firm of an other industry. For example, ABN AMRO is mentioned more often than PostNL, while the industry financial institutions is on average less mentioned than the transport & storage industry. These insights suggest that there are company specific aspects that also influence the amount of messages that are created in relation to the firms, i.e. the industry type is not the only aspect influencing the volume of firm-related messages.

We examine the availability of user-generated content in the different industries by comparing the *average daily mentions* of the different industries with each other. Table 4-10 presents the number of mentions of firms,

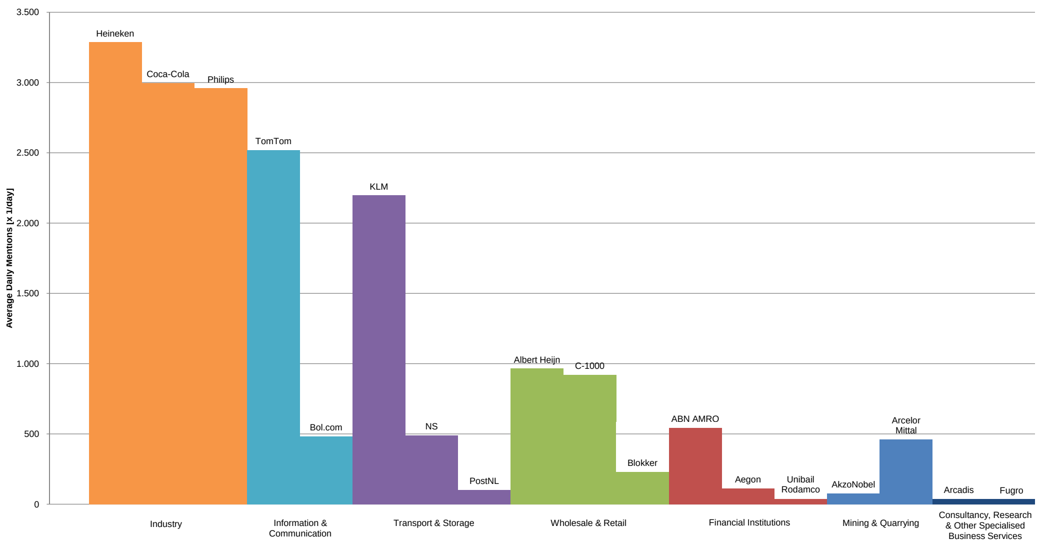


Figure 4-16: Average Daily Mentions of Firms
Bar chart illustrating the average daily volume of firm-related messages. Firm have been clustered over the industries and consequently ordered descending.

consolidated across different industries. The second column of the table shows the total amount of social media posts that have been collected in the corresponding industry. The third column shows the average daily mentions of a firm in the corresponding industry. The values in the third column thus represent an average of an average, thereby eliminating the fact that the number of firms – respondents – differs per industry type. Figure 4-17 presents these values. This figure suggests that the existence of firm-related UGC differs among industry type, implying that differs per industry whether or not there exists user-generated content on social media.

Table 4-10: Average Daily Mentions, Consolidated per Industry

Industry	Total collected posts	Average daily mentions per firm
Mining and quarrying	44.904	269
Industry	72.378	3.080
Wholesale and retail	25.414	706
Transport and storage	33.550	929
Information and communication	38.530	1.500
Financial institutions	9.028	231
Consultancy, research and other specialised business services	883	37
Σ	224.687	

Figure 4-16, 4-17 and table 4-10 provided insight in the variations in the volume of firm-related social media messages across different industries. With this insight, we can examine the second hypothesis of this chapter, which was formulated as:

> H_2: *The volume of firm-related social media messages differs between industries.*

The results of our analysis illustrate variations in the volume of firm-related social media messages, which suggest – based on our sample – that there exists a variation in the daily volume of social media messages that are created. When ordered descending, the *industry* firms are mentioned mostly, followed by *information and communication*, *transport and storage*, *wholesale and retail*, *mining and quarrying*, *financial institutions* and *consultancy, research and other specialised business services* being the least mentioned on social media. Thus, the results of this analysis indicate that it matters in which industry a firm is active whether or not the firm will be subject on social media. It is therefore that – taken into account our sample – we accept the second hypothesis, implying that the volume of firm-related social media messages differs between industries. However, the categories on the two dimensions that are researched in this thesis are not fully independent. For example,

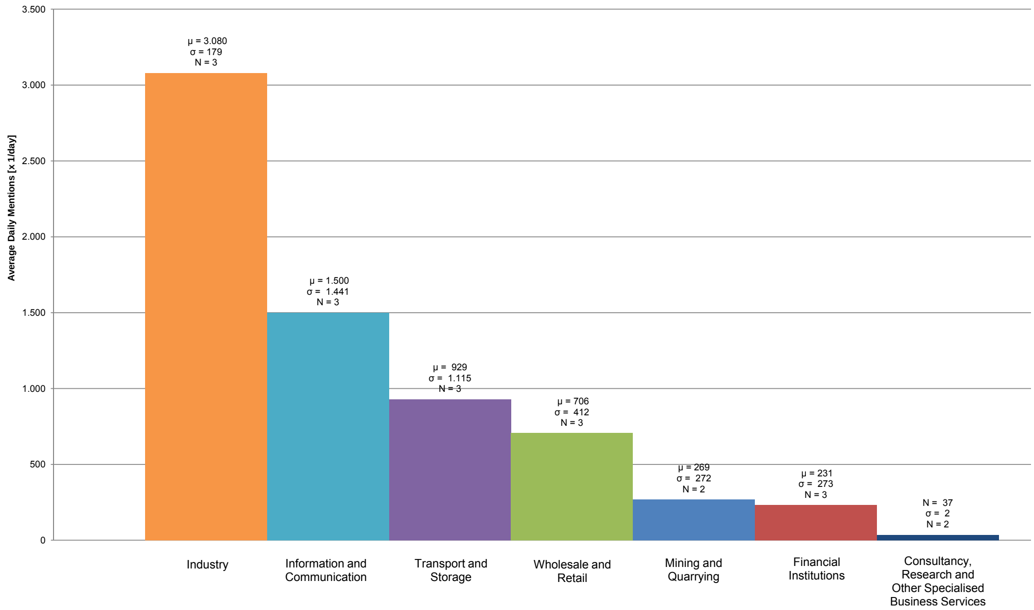


Figure 4-17: Average Daily Mentions of Firms
Bar chart illustrating the volume of daily mentions for firms in different industries. The social media messages related to firms in the same industry have been consolidated.

the category *consultancy, research and other specialised business services* solely consists of *B2B* firms. This issue is discussed later in this thesis.

4-5-3 Subjects of Social Media Posts

Next to an assessment of the amount of social media posts that are created on the web, this thesis examines the subjects of the social media posts in order to link the messages to key-performance indicators. The social media messages of the sample firms have been classified into one of the categories that have been established in section 4-3-1. These categories are based on ten categories of commonly applied key-performance indicators. Consequently, the collected social media posts of the firms in the sample have been classified into one of these categories. The results of this activity are documented in appendix B from firm to firm. In this section, the subjects of social media posts are analysed. First, the subjects of social media messages are discussed from firm to firm. Next, the social media posts are analysed from the two dimensions which are the perspectives of this thesis. Consequently, we analyse whether or not the customer relation type influences the types of subjects that are contained in social media messages. Finally the same analysis is executed, only this time from an industry perspective.

Subjects per Firm

Figure 4-18 shows the results of the classification process, in which all firms are displayed. The percentages in the figure indicate how many of the classified posts are assigned to the corresponding KPI category. The colours of the bars represent the main categories of subjects of social media posts. When purely looking at the colours, it becomes clear that some firms' social media posts contain much *financial result* ('orange') posts, while others contain a high portion of *customer relations* ('red') posts. Furthermore, we see the existence of *community* ('blue') posts in each firm. For reasons of readability, under-represented categories of subjects have been grouped under a category called *other* ('grey'). Figure 4-18 shows two remarkable values of the 'other' category. For KLM, these messages mainly have been classified as being spam. The three letters are used by other people on social media as well, for instance because these are the initials of the person. Also C-1000 shows a remarkably high percentage of 'other' posts. A closer look at C-1000's messages reveals that many of these posts can not be classified into one of the ten categories and are hence classified as 'undefined' posts. Mainly, these posts contain expressions of people who use the C-1000 stores as a point of reference to meet each other. A full overview is presented in appendix B.

Figure 4-18 illustrates that the subjects of social media posts related to firms differs from firm to firm. In

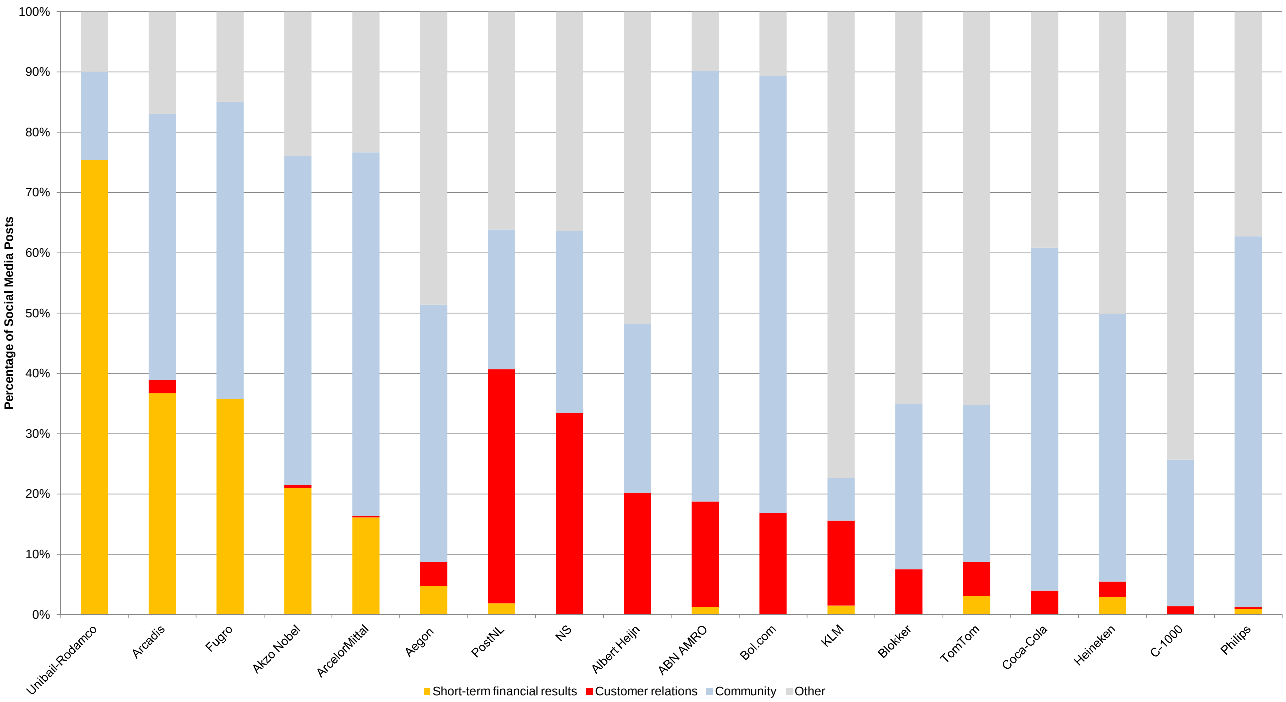


Figure 4-18: Social Media Posts Subject Classification
Stacked bar chart illustrating the percentages of social media posts assigned to different categories based on the posts' subjects. Small percentages of categories have been merged under 'other'.

the following paragraphs, we investigate whether or not it is likely that the factors *customer relation type* and *industry* type affect the type of subjects of firm-related social media posts.

Subjects per Customer Relation Type

In this paragraph we assess whether the subjects of social media messages differ for different *customer relation types*. Illustratively, we examine amongst other whether or not the percentage of *product and service quality* related messages differs for firms performing a B2B or a B2C relation. As a first step to analyse differences in subjects across B2B and B2C firms, the individual firms have been grouped into their respective customer relation type, and the percentages of the subjects have been plotted in figure 4-19. Again, for readability issues, under-represented categories of social media subjects have been grouped under an *other* category.

The third hypothesis that is examined in this chapter was formulated as:

> H_3: *The subjects of firm-related social media messages differ between firms performing B2B and B2C relations.*

As can be concluded from figure 4-19, social media posts related to B2B firms contain a high percentage of posts related to *financial results* ('orange'), while this percentage is under-represented for B2C firms. Such information is not of any additional value for a firm, since these posts contain information that is yet available at the firm. On the contrary, social media messages related to B2C firms contain a high portion of *customer relations* ('red') related posts in comparison with B2B firms. For both type of firms it holds that a high portion of the social media messages reveal the *communities'* perceptions of the firm ('blue' bars). However, B2B firms' *community* related social media posts are created by *professionals*, while in the *community* messages related to B2C firms, these messages are created by consumers. For a detailed overview of the percentages of subjects related to each firm, please see table B-1 in appendix B. These insights suggest that the subjects of social media messages related to firms that pursue different customer relation types vary, and that the third hypothesis is to be accepted. Thus, firms performing different customer relation types will find different subjects in their firm-related social media messages.

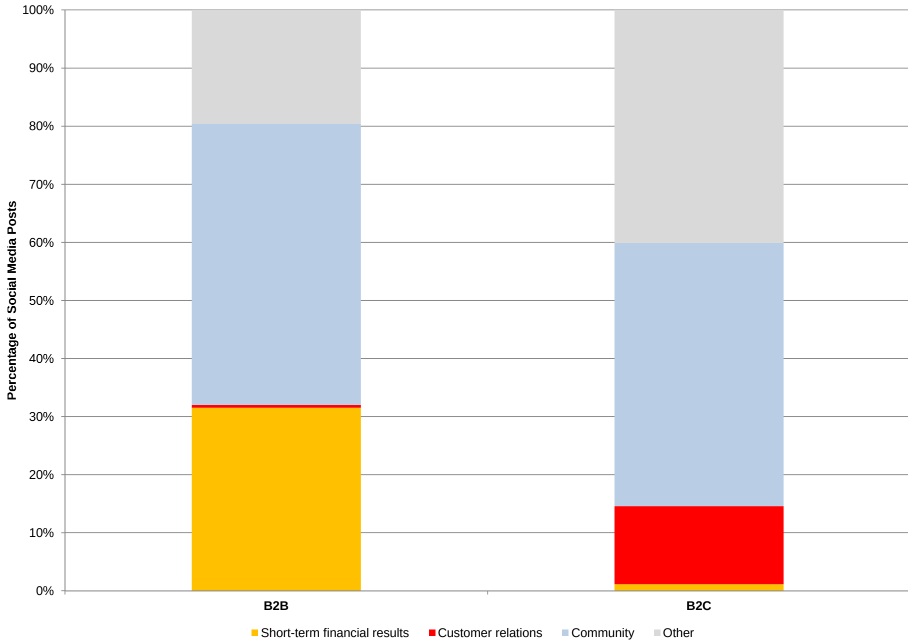


Figure 4-19: Social Media Posts Subject Classification, Consolidated per Customer Relation Type
Stacked bar chart illustrating the portion of social media messages related to different subjects. Firms have been consolidated per customer relation type. Categories that are under-represented are merged under 'other' posts.

Subjects per Industry Type

In the previous paragraph the availability of user-generated social media content related to different social media post categories has been investigated across the customer type dimension. This section performs the same analysis, only this time the industry dimension serves as the distinguishing factor of the firm types. In the appendix, figure D-1 (page 113) lists – in detail – for each industry the average amount of social media posts related to the different categories of social media posts.

The fourth hypothesis that is examined in this chapter was formulated as:

> H_4: *The subjects of firm-related social media messages differ between industries.*

Figure 4-20 depicts the portions of social media messages of the different subjects across the seven industries. Under-represented subjects of social media messages have been merged in the *other* category. When looking at the colours of the bars, differences are seen – again – in *customer relations* ('red') type of posts and *financial results* ('orange') posts. E.g. the posts related to wholesale and retail firms contain a higher portion of *customer relation* posts than *financial results* posts. The contrary is seen in mining and quarrying, and consultancy firms. Thus, our results suggest that the subjects of social media messages differ per industry type, implying that the fourth hypothesis is to be accepted.

4-6 Interpretation of the Results

In the fifth step of the content analysis, the results are interpreted into meaningful conclusions. In the beginning of this chapter, the hypotheses that are to be examined by the content analysis have been formulated. These hypotheses relate to two aspects: (i) *volume* of social media posts, and (ii) *subjects* of social media posts. The content analysis has been designed and executed in a manner to examine these hypotheses, of which the results are discussed in this section.

4-6-1 Volume of Social Media Posts related to Firms

The collecting process of social media messages related to the firms in our sample resulted in different amounts of messages for different firms. E.g. for Heineken, 39.425 messages have been collected while only 428 posts have

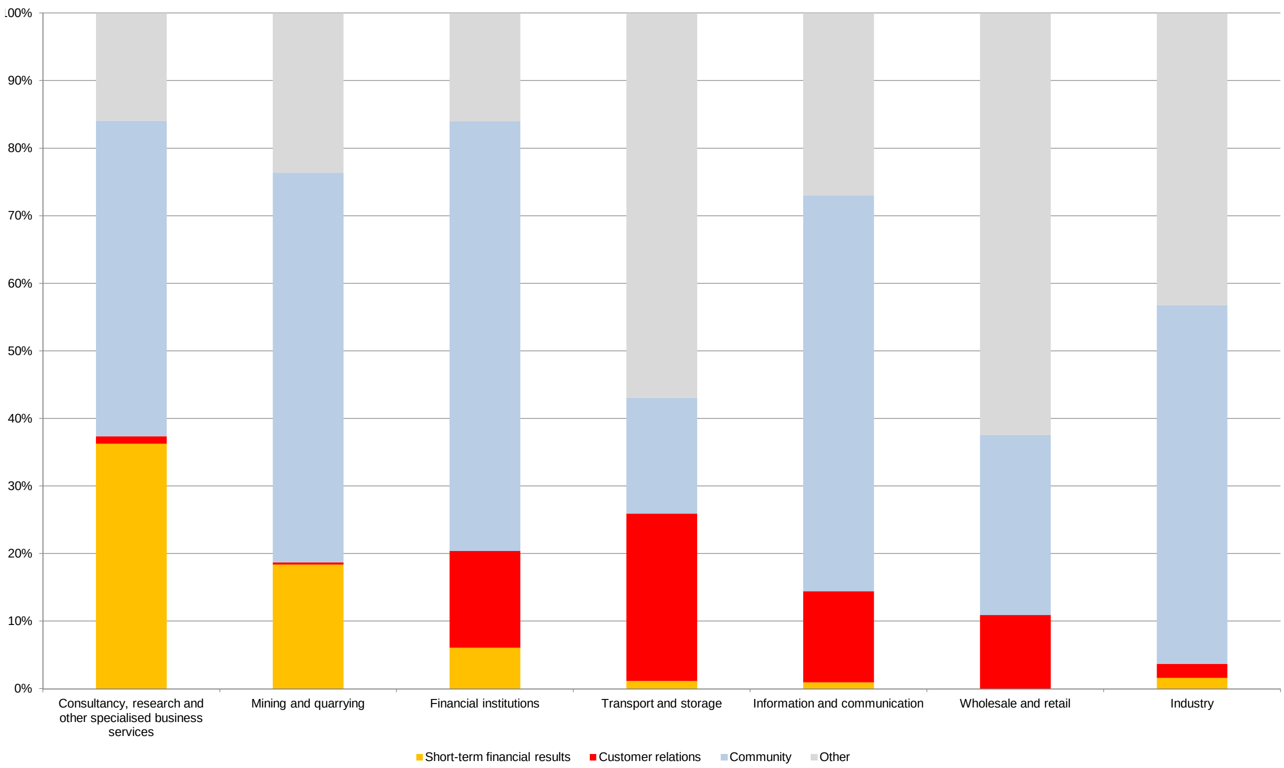


Figure 4-20: Social Media Posts Subject Classification
The figure consolidates social media messages of firms in the same industries. The figure illustrates that the subjects of social media messages related to firms differs per industry type.

been found that are related to Fugro. We constructed a variable labelled *average daily mentions* by dividing the total amount of collected posts for a firm by the amount of days that the firm has been monitored. The variable *average daily mentions* is deemed as the variable that reflects the volume of social media posts. In our sample, the mean *average daily mentions* amounts 1.025 messages per firm per day.

The *average daily mentions* differs from firm to firm. In order to draw generic conclusions, i.e. not firm specific, the firms were firstly grouped into either business-to-business or busines-to-consumer firms. As the analysis showed, firms conducting a B2C relation will find more social media posts that are related to them than firms active in the B2B sector. In our sample, we find an average of 130 daily mentions of a B2B firm, while a B2C firm is daily – on average – 1.369 times mentioned. Secondly, the firms have been grouped into seven industries. Again, the *average daily mentions* have been analysed and differences between the volumes are illustrated.

4-6-2 Subjects of Social Media Posts related to Firms

The second aspect of this thesis' content analysis analyses the subjects of the social media posts. As discussed, the subjects of the social media messages are to be linked to the firm's key-performance indicators. The collected social media messages have been classified into one of the 28 categories of social media messages that have been established. Figure 4-18 (page 67) illustrates that subjects of social media messages differ from firm to firm.

In order to draw generic conclusions of the subjects of the social media messages related to different firms, the firms have firstly been grouped based on their customer relation type. We can conclude that the subjects of social media messages related to B2B firms contain a higher percentage of *short term financial results*, *news* and *professionals* related messages than messages related to B2C firms. Next, the analysis indicates that the social media messages related to B2C firms contain a higher percentage of posts related to *customer relations*, *product and service quality* and *product and service innovation* than messages related to B2B firms. Secondly, firms have been grouped into seven different industries. In the same way as with the analysis of the volume of social messages, we find that the subjects of social media messages differ among the firms participating in the different industries of our sample. The majority of social media messages related to firms (41%) express how the external stakeholders of a firm perceive the company. In this thesis, such posts have been classified as *community* posts. 18% of the social media messages in our dataset contained the name of a firm, but did not contain any valuable information for the firm and have consequently been assigned as *undefined* posts. About 11% of the social media messages relate to *financial results*, which consist of *financial performance discussions* (5%) and *stock related discussions* (6%). Table 4-11 lists the interpretation of the results of the content analysis in a summarily manner.

Table 4-11: Conclusion of Content Analysis

	Volume of Social Media Messages	Subjects of Social Media Messages
Customer Relation	B2C firms are more often subject of discussion on social media than B2B firms.	B2C firms related social media messages are more often subjected to *customer relations*, *product and service quality* and *product and service innovation* than B2B related firms. On the other hand, B2B firms' related messages are more often subjected to *financial results*, *news* and *professionals* discussing the firm. However, the information contained in the social media posts of B2B firms is often yet available to the firm, hence not offering added value to the firm's richness of management information.
Industry	Our analysis shows a variation in the volume of social media messages across different industries.	Our analysis indicates that there is a difference in the subjects of social media posts related to firms in different industries.

4-7 Sub Conclusion: Social Media Posts that relate to KPI Categories and the Performance Prism Perspectives

Firstly, our analysis indicates that there exists a difference in the volume of firm-related social media messages that are daily generated. These differences are indicated when comparing firms with firms, but also when we compare between B2B and B2C firms and when a comparison between different industries is made. With respect to the volume of firm-related social media content, we can state that especially B2C firms are able to collect social media data for business intelligence purposes because it are these firms that are subject of interest on social media sites.

Secondly, in our analysis social media messages have been assigned to different categories of KPIs. The assigning of messages to KPI categories was based on the subject of the messages. As the analysis indicated, it differs from firm to firm which kind of KPIs are candidates to be measured using social media data. When taking a customer relation type perspective, the analysis indicates that KPIs related to the *community*, i.e. the metrics that reflect the attitude of external stakeholders towards the firms, are the ones that are particularly suited to be measured using social media data. Social media messages related to *community* metrics provide a firm with insight that cannot be generated with internal systems, the information contained in these messages are created by individuals discussing the firm and/or the firm's products / services. Additionally, we see that a substantial part of the social media messages related to B2C firms are related to *customer relations* metrics. These messages contain questions and/or complaints of customers and are purposed to get in contact with the firm. Several firms embrace these messages by establishing a web care team that actively responds to customers writing messages purposed to contact the firm. As regards to B2B firms, we see that a high percentage of the social media posts relate to *short-term financial results*. Regrettably, the information in these messages are also available to the firm without the existence of social media messages. Most likely, the firm is aware of the information in these messages before it is available on social media. Next, we see that the percentage of *professionals* posts (a sub category of *community*) is higher for B2B firms than for B2C firms. These messages contain valuable information for the firm, such as market analyses and the position of the firm in that situation, or forecasts for macro economic developments and the effects on the firm and/or the firm's ecosystem. However, we have to bear in mind that the volume of social media messages related to B2B firms is much lower than for B2C firms.

In chapter 2, a framework has been presented that illustrates the relation between the five *performance prism* perspectives and ten categories of key-performance indicators (page 23). As indicated, this thesis examines possibilities to measure operational performance – i.e. the key-performance indicators of a firm – by means of

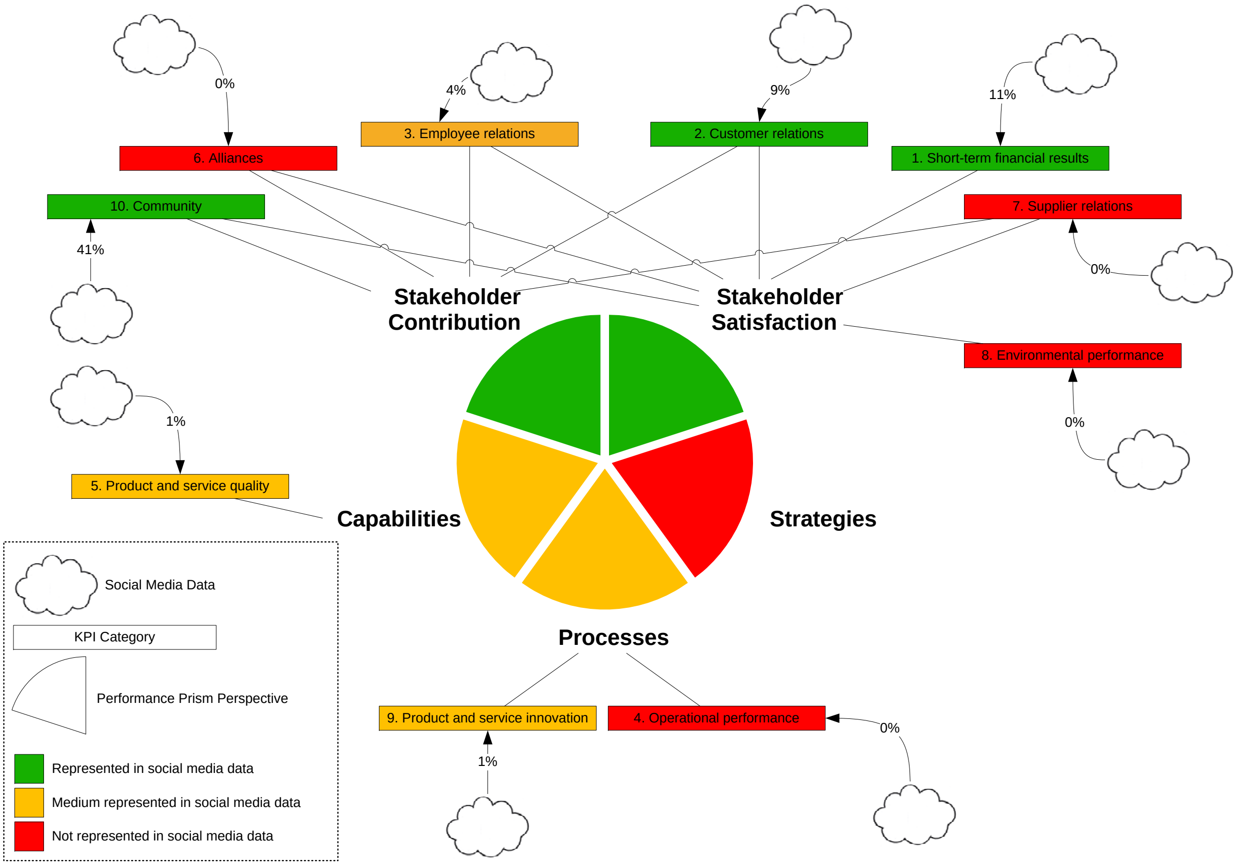


Figure 4-21: Social Media Data related to Key-Performance Indicators
Figure illustrates the link between social media data, key-performance indicators and performance prism perspectives. The colours indicate for which KPI categories - and hence for which performance prism perspectives - social media data can be found that relates to these categories.

social media data. With the knowledge we gained with a content analysis of a sample of social media messages related to firms, we can conclude which categories of KPIs are candidates to be measured using social media data. Figure 4-21 schematically shows for which type of KPIs there exists social media data that is related to these KPIs. KPI categories that were not represented in the sample – i.e. smaller than 1% of the total messages that have been classified – have been coloured red, implying that the respective KPI category is under-represented in social media data and hence not able to be measured by means of social intelligence. These under-represented KPI categories are related to *alliance* metrics, *supplier relations* metrics, *environmental performance* metrics and *operational performance* metrics. Next, categories of KPIs that were somehow represented in the analysed messages – i.e. between 1% and 5% of all classified messages – have been coloured orange. As can be concluded, these KPI categories relate to *employee relations*, *product and service quality* and *product and service innovation*. Finally, categories of KPIs that were – in comparison with the other categories – highly represented in the sample data have been coloured in green. Formally, KPI categories of which more than 5% of the sample data could be assigned to the respective category have been coloured green.

Consequently, since section 2-4 yet established links between the ten KPI categories and the five performance prism perspectives, we can draw conclusions on the applicability of social business intelligence for the different performance prism perspectives. Corresponding to the colours of the KPI categories, the five performance prism perspectives have been coloured red, yellow or green. The colours represent the applicability of social intelligence for the different perspectives. As can be concluded, performance metrics in the domains *stakeholder contribution* and *stakeholder satisfaction* are especially suited to be measured using social media data.

The results of the content analysis showed that the applicability of social business intelligence differs from firm to firm. Especially firms B2C firms are likely to find firm-related messages. Whereas figure 4-21 shows the overall percentages of social media messages related to the different KPI categories, these figures vary from firm to firm and are thus higher for B2C firms. For a detailed overview, see appendix D).

Chapter 5

Blueprint of a Social Business Intelligence Procedure

Chapter 4 illustrated that firm-related social media messages contain information that can be linked to a firm's key-performance indicators. However, chapter 4 also showed that not all categories of KPIs can be linked to the content created on social media, simply because the user-generated social media content does not relate to all categories of key-performance indicators. For those KPIs that *are* related to the subjects of the social media messages, a procedure is required that prescribes how a firm should acquire and process these social media messages for business intelligence purposes. A blueprint for a procedure in which social media data are collected and processed in a way that corresponds with companies' general business intelligence processes is developed in this chapter. We refer to such a procedure as a *social business intelligence procedure*. Given the insights gained in chapter 4, we state that the following chapter is only relevant for certain firms; firms that are mentioned on social media. Firms that are unable to find related social media data should not invest in the development of social business intelligence procedures.

Section 5-1 starts this chapter by formulating the requirements of a social business intelligence procedure. Next, section 5-2 provides the blueprint of the procedure, and discusses the necessary steps that form the procedure. In section 5-3 the procedure is verified. In section 5-4, the real-time aspect of social business intelligence is discussed. In section 5-5 traditional business intelligence procedures are compared with social business intelligence. Finally, section 5-6 concludes the findings of this chapter.

5-1 Requirements Formulation

Based on the business intelligence concepts that are discussed in chapter 2, the possibilities of social media monitoring tools that are discussed in chapter 3, and the experience that we gained in chapter 4 by performing a content analysis on the social media messages related to different firms, nineteen requirements for a business intelligence procedure have been formulated. Section 5-1-1 describes these requirements. In section 5-1-2 the formulated requirements are verified on consistency with general business intelligence procedures.

5-1-1 Description of Requirements

Table 5-1 lists the requirements for a social business intelligence procedure. These requirements are discussed in the following sections. A social business intelligence procedure should:

1. **Have access to social media platforms**
   Obviously, to acquire intelligence from social media messages, a firm should have access to the platforms where these messages are produced.

2. **Identify the social media platforms at which the firm is discussed**
   The fundamental purpose of social intelligence is to acquire insight in the the perception of the firm's

Table 5-1: Requirements

	A social business intelligence procedure should ...
1	Have access to social media platforms
2	Identify the social media platforms at which the firm is discussed
3	Identify the volume of social media messages related to the firm
4	Remove the spam from social media messages that initially seemed to relate to the firm
5	Anonymise personal data
6	Identify who the people are that discuss the firm on social media
7	Identify what the subjects of the social media messages related to the firm are
8	Determine whether the information contained in the social media messages related to the firm offers additional value
9	(Automatically) Classify the social media messages related to a firm into categories
10	Relate the (categories of) subjects of the social media messages to the firm's key-performance indicators
11	Determine the firm's social reputation
12	Determine the social reputation of the firm's product(s)
13	Determine relations between social media metrics and the firm's (social) key-performance indicators
14	Update the status of the social media metrics and the values of the KPIs constantly
15	Present the slope of the relations between social media metrics and KPIs on a time chart
16	Interpret the gained intelligence and position it into the firm's developments
17	Assign the gained intelligence to the right persons in a firm
18	Allow a firm to engage on social media platforms
19	Regularly update the search terms to anticipate on changes

external stakeholders – including (potential) customers, competitors and parter firms – towards the firm and/or the firm's products/services. As illustrated in chapter 4, the majority of social media messages that are related to firms, and publicly accessible, are written on Twitter. However, the distribution of the platforms where the firm is discussed may vary from firm to firm. Therefore, before starting the monitoring of social media messages, a firm should investigate where – i.e. on which platforms – the firm is subject of discussion.

An overview of the platform distribution provides a firm insight into which social media platforms the firm should focus, engage or advertise. Though a firm may be subject of discussion on multiple platforms, it does not imply that the firm is required to monitor these platforms individually. The social media monitoring tools offer the possibility to monitor and engage on multiple social media platforms through *one* dashboard.

3. **Identify the volume of social media messages related to the firm**
The fact that a firm is subject of discussion on social media is of less value whenever there are little messages available for a firm to analyse. Furthermore, the existence of more social media messages related to a firm offer opportunities to identify correlations between the amount of these messages and the firm's KPIs. Such analyses are of less value when there are little social media messages available. As chapter 4 revealed, the volume of social media messages differs from firm to firm. Especially business-to-consumer firms are discussed on social media, implying that these firms have the opportunity to acquire social intelligence. A constant monitoring of the amount of social media messages related to the firms allows for the detection of sudden deviations, illustrating that there is "something going on", which may require attention from the firm's management.

4. **Remove the spam from social media messages that initially seemed to relate to the firm**
As experienced in chapter 4, many social media messages contain the name of the firm in the post, though they do not relate to the firm. Especially firms carrying a commonly used name or an abbreviation (such as KLM), are likely to receive many spam messages in their social media messages. Though it may help to use specific user names for the firm's web care team (e.g. @KLM_WebCare), the drawback of such a name is that the firm will not detect all firm related messages since users will nevertheless use the generic name in their posts. Spam related messages are to be removed from the dataset since they do not contain any value for the firm.

5. **Anonymise personal data**

As illustrated in section 3-4, the European Commission has drafted new Regulation on processing personal data. As a consequence, firms are not naturally allowed to process data that allows one to retrace a natural person from that data. In order to be in compliance with the expected new Regulation, firms intending to collect and process social media data should anonymise the personal data.

6. **Identify who the people are that discuss the firm on social media**
Though A. M. Kaplan and Haenlein (2010) argue that the usage of social media is diversifying in terms of the users' age, it is wise to determine who the people are that discuss the firm on social media. It is out of scope of this thesis to describe how different customer groups (e.g. different generations, men / women, different cultures) should be treated, but whenever a firm decides to engage into the social media conversations, it should be aware of the people that make up their social media environment. Furthermore a firm can decide that it does not consider the people that produce the social media messages as critical customers, and therefore does not undertake any action.

7. **Identify what the subjects of the social media messages related to the firm are**
Whereas the second requirement of a social business intelligence procedure ensures that a firm has insight in the amount of messages that are produced and containing the firm's name, it is also valuable for a firm to have insight in *what* it is that social media users discuss in relation with the company. The identification of the subjects of social media posts forms the basis for the translation of social media posts to key-performance indicators. Furthermore, the identification of subjects – combined with the volume of messages – provides a firm with insight in the topics that are "trending", i.e. popular topics at the moment. Trending topics related to firms can serve as a measure describing what people consider as important, and which may be action points for the firm.

8. **Determine whether the information contained in the social media messages related to the firm offers additional value**
The content analysis of a set of social media messages related to firms revealed that there are also messages that do contain the firm's name, but neither do contain information that is of any value for the firm. We have classified posts of no value as *undefined* posts.

   Furthermore chapter 4 illustrated that there are also messages that contain information that must be available to the firm without analysing the social media. Especially for B2B firms, many social media posts contain information about the financial performance or share prices of the firm. Generally, a publication or press article has been the source of these messages. These messages do not contain information that is not available in the firm yet, and are therefore considered of less value for the firm.

9. **(Automatically) Classify the social media messages related to a firm into categories**
The unstructured character of social media posts makes it that these messages have to be preprocessed before an analysis can commence. Classifying the messages into categories, e.g. into categories of subjects (as we have done in chapter 4), categories of languages, categories of men and women, categories of many or less followers, etc. allows a firm to structurally analyse the messages and derive that particular information that the firm is interested in.

   The unstructured nature and the large amount of messages that are generated in relation with some firms makes it that social media data can be termed as "big data". It is therefore desired that the classification process of the social media messages into categories runs automatically. Automatic classifiers are existing solutions to this problems, and are also available for text. These classifiers require so-called "training sets" in order to establish criteria at which a piece of text is either classified in e.g. category A or in B. As we have experienced, the subjects of social media messages differ from firm to firm. Two social media posts containing the word "Senseo" and "TomTom One XL" are both related to a product, but do not contain the same words. Therefore, training sets should be established for specific firms. Our classification can be used to train classifiers for the firms that participated in the sample of this thesis.

10. **Relate the (categories of) subjects of the social media messages to the firm's key-performance indicators**
As we have showed in chapter 4, it is possible to classify social media posts into categories that are related to KPIs. The subjects of the social media posts serve as the basis to assign a certain social media message to a certain key-performance indicator. For example, when a firm manages by a KPI representing the customer satisfaction towards a certain product, it can use the social media messages related to that product as a measure to determine the satisfaction level. As we have seen in the previous chapter, TV commercials are also subject of discussion on social media. A firm may determine the success of such a campaign by counting the messages that relate to the commercial.

11. **Determine the firm's social reputation**
A social business intelligence procedure should determine the firm's reputation on social media. The volume of messages related to the firm is not of any value whenever there is no insight in the nature of these messages since it matters whether or not these messages are positive or negative. Social media monitoring tools offer the possibility to determine the sentiment of a social media post. Generally, posts are classified as either positive, neutral or negative. As section 3-3-1 illustrated, the sentiment analyses may not always be as reliable. However, as we expect, sentiment analysis tools will be improved and able to determine the sentiment of the most more accurate. The firm's social reputation – e.g. measured by the percentage of positive posts related to the firm – is an interesting indicator that may reveal correlations with other KPIs, such as sales.

12. **Determine the social reputation of the firm's product(s) / service(s)**
Whereas it is necessary to determine the firm's social reputation, a firm may be interested in the reputation of a particular product or service that it provides. Again, sentiment analysis is required for theses posts. The social reputation of products – e.g. measured by the percentage of positive posts related to that product or service – may reveal correlations with the sales or amount of returns of that product.

13. **Determine relations between social media metrics and the firm's (social) key-performance indicators**
One of the fundamental purposes of business intelligence is to identify which activities of a firm deliver value. In order to determine which social media metrics actually relate to the firm's, the social business intelligence procedure should contain a step in which the relations between social media metrics and the firm's KPIs are determined. An example of such a relation may be the amount of positive messages about product x and the sales in a certain period of product x.

14. **Update the status of the social media metrics and the values of the KPIs constantly**
In order to develop real-time business intelligence, the system should automatically monitor the social media metrics. "This will only be satisfied whenever the right KPIs are defined before the metrics are monitored" (Azvine et al., 2005). As such, the firm gets insight in the values of the social media metrics and the values of the KPIs.

15. **Present the slope of the relations between social media metrics and KPIs on a time chart**
Whereas the previous requirement ensures that the information that is derived from social media is presented in real-time, this requirement ensures that the slope of the values are presented in a way so that deviations over time are easily recognised. Sudden events may trigger social media metrics to fluctuate, these events are able to be notified when the values are presented in a time chart.

16. **Interpret the gained intelligence and position it into the firm's developments**
Whereas the derived intelligence may reveal relations of social media metrics and KPIs and provide insight in the external stakeholders' perceptions of the firm, one should always position this intelligence in the light of developments of the firm.

17. **Assign the gained intelligence to the right persons in a firm**
When it turns out that certain KPIs are influenced by social media metrics, and these KPIs are not performing sufficiently, the acquired intelligence should be communicated to the responsible departments in the firm. The departments can provide clarifying factors for the under performing KPIs, and can take the acquired intelligence (e.g. related to the feature of a certain products) into their decision-making process.

18. **Allow a firm to engage on social media platforms**
A social business intelligence procedure should allow firms to engage with the users on social media. As the content analysis in chapter 4 illustrated, many firms engage in the social media discussions. Though we cannot verify this statement, it is expected that there will be generated more user-generated content whenever a firm actively participates on social media. We will elaborate about this statement in the further research section (section 6-6).

19. **Regularly update the search terms to anticipate on changes**
An up-and-running social business intelligence procedure has been started by search terms that are related to the firm. Since a firm is always in development, it will launch new products, services and employees will come and go. Therefore, the search terms should be updated whenever there are events that influence the required search terms. For example, whereas Microsoft's search terms include "Windows 7", it should add "Windows 8" to these search terms by the time it launches – or pre-launches – this new product.

5-1-2 Requirements Check on Business Intelligence Concepts

Chapter 2 described the business intelligence concept as it is applied within firms. Especially section 2-3 elaborated about the activities that make up the business intelligence process. Van Beek (2006) argues that a BI process consists of three main tasks, being (i) registering, (ii) processing and (iii) reacting. Additionally, the processing task consists of 15 sub tasks required to process the registered data. In total, 17 (1+15+1) tasks can be distinguished that are required for a business intelligence process. We verify the requirements for a social business intelligence procedure – that have been established in section 5-1-1 by controlling whether each of the BI steps are represented by at least one of the requirements that we have established.

As can be seen concluded from table 5-2 (page 77), each activity is represented by at least one requirement. This allows us to conclude that the requirements of the social business intelligence procedure are consistent with existing BI procedures.

Table 5-2: Business Intelligence Activities and Social Business Intelligence Requirements

	Social Business Intelligence Requirements	Register	Collect	Filter	Combine	Aggregate	Visualise	Interpret	Internalise	Revise	Verify	Enrich	Share	Remember	Decide	Distribute	Anticipate	React
1	Access to social media platforms	✓	✓															
2	Identify the social media platforms at which the firm is discussed																	
3	Identify the volume of social media messages related to the firm			✓														
4	Remove the spam from social media messages that initially seemed to relate to the firm			✓														
5	Anonymise personal data																	
6	Identify who the people are that discuss the firm on social media			✓														
7	Identify what the subjects of the social media messages related to the firm are			✓														
8	Determine whether the information contained in the social media messages related to the firm offers additional value										✓							
9	(Automatically) Classify the social media messages related to a firm into categories			✓														
10	Relate the (categories of) subjects of the social media messages to the firm's key-performance indicators	✓			✓													
11	Determine the firm's social reputation							✓										
12	Determine the social reputation of the firm's product(s)							✓										
13	Determine relations between social media metrics and the firm's (social) key-performance indicators				✓	✓												
14	Update the status of the social media metrics and the values of the KPIs constantly													✓				
15	Present the slope of the relations between social media metrics and KPIs on a time chart						✓											
16	Interpret the gained intelligence and position it into the firm's developments							✓				✓						
17	Assign the gained intelligence to the right persons in a firm								✓				✓		✓	✓		
18	Allow a firm to engage on social media platforms																	✓
19	Regularly update the search terms to anticipate on changes									✓							✓	

5-2 Social Business Intelligence Procedure

A blueprint for a social business intelligence (“SBI”) procedure has been developed, in which all requirements of section 5-1 are taken care of. An aggregate overview of the procedure is presented in figure 5-1.

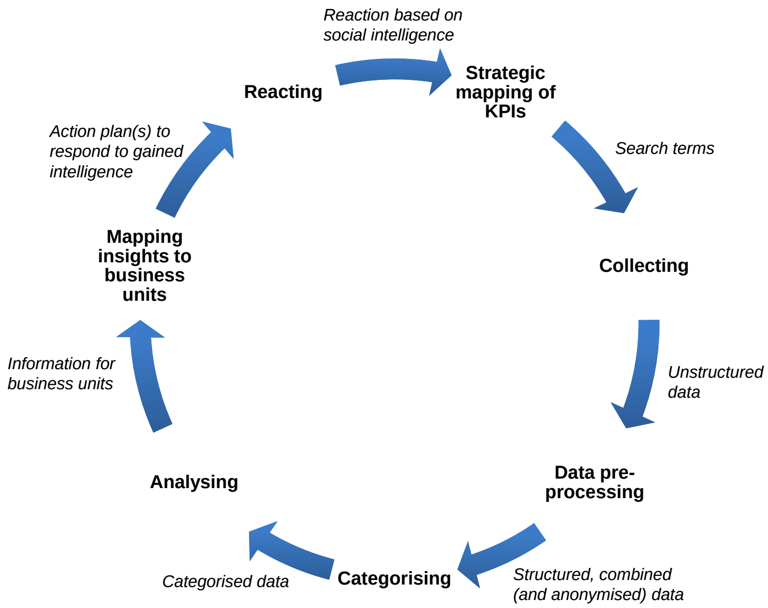


Figure 5-1: Blueprint: Social Business Intelligence Procedure

The SBI procedure consists of seven main activities that are related to each other, as figure 5-1 illustrates. Each of the main activities are further exemplified in the following sections.

5-2-1 Strategic mapping of KPIs

Firms deduct key-performance indicators from their strategy. This process is further elaborated in section 2-2. The KPIs that a firm eventually established are to be measured. As the content analysis of this thesis revealed, some KPIs are not appropriate to be measured by social media because there does not exist any content that related to these KPIs. Other KPIs are best measured by internal systems, and some KPIs are properly measured by social media. From the list of KPIs that a firm uses, a selection can be made of indicators that are to be measured by social media. Illustratively, figure 5-2 highlights the KPIs that are to be measured by social media. These KPIs form the starting point of the social business intelligence procedure, since it are these KPIs for which social media data is to be collected and analysed.

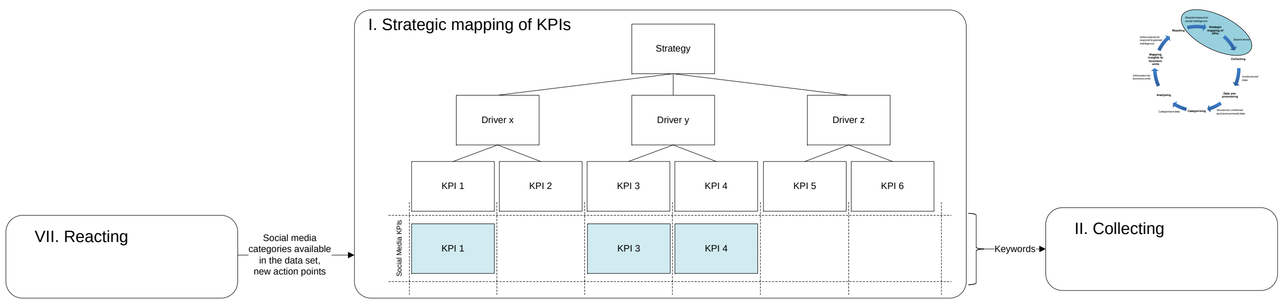


Figure 5-2: Blueprint: Social Business Intelligence Procedure (Strategic Mapping of KPIs)

The KPIs selected to be measured by social data determine the categories that are to be analysed – i.e. the subjects of social media messages – and hence the keywords that are to be used in the collecting process. On the other hand, the available social media data determines whether or not it is possible to measure the KPI by social media data. After all, a KPI for which no related social media data exist, can not be measured by social data. Thus, there exists an interaction between on the one hand what a firm wants to measure by social media data, and on the other hand what a firm is possible to measure using social media data. As we have seen in chapter 4, not every KPI is subject of discussion on social media.

5-2-2 Collecting

After the first step, in which the KPIs that are to be measured by social media data have been selected, the data is to be *collected*. The step is schematically represented in figure 5-3

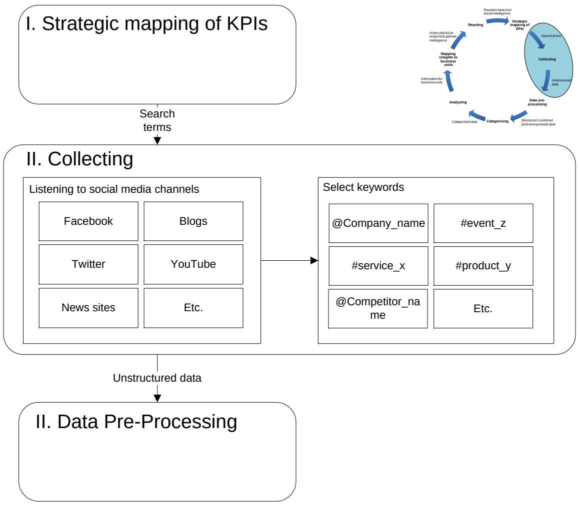


Figure 5-3: Blueprint: Social Business Intelligence Procedure (Collecting)

Keywords related to the firm, the firm's products/services and the selected KPIs are used to "listen" to multiple social media channels at which the firm could be mentioned. The content analysis of this thesis revealed that it differs per firm on which social media platform the firm is discussed. It is therefore that the first step involving social media platforms consists of the determination of the platforms at which the firm is discussed. As we have experienced in chapter 4, search queries related to firms will result in unstructured data from multiple social media platforms. These unstructured data are to be *pre-processed*, which is the next step in the SBI procedure.

5-2-3 Data Pre-Processing

The third step in the social business intelligence procedure consists of *pre-processing* the collected data. In contrast to 'regular' BI data, social media data is unstructured, sourced from multiple platforms, containing spam and personal data, and is therefore required to be pre-processed. Figure 5-4 illustrates this process.

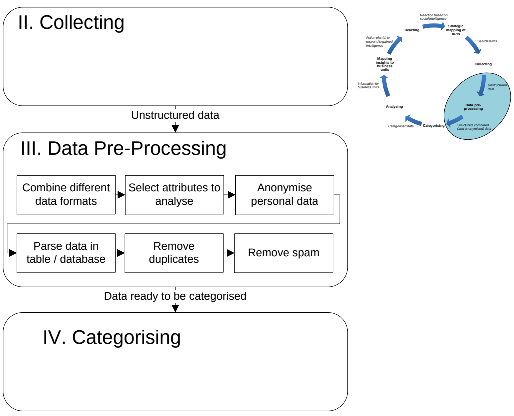


Figure 5-4: Blueprint: Social Business Intelligence Procedure (Data Pre-Processing)

The collected data consists of social media messages that are sourced from multiple sources in different formats, such as CSV, JSON, XML, etc. Each data source may employ its own structure of social media messages, and not each platform may contain the same richness in attributes as the other. For instance the Twitter API offers developers the opportunity to extract so called geotags – geographic coordinates of the origination of the Tweet – while other social media platforms do not offer this attribute to the messages. Each social media post should be parsed – structured – into one and the same data format. Next, as we have experienced in the scraping process of chapter 4, multiple search queries will lead to multiple messages yet available in the database. Therefore, only social media messages that do not exist in the table should be added. The final step in the *data pre-processing* step consists of the removal of spam. After the data *pre-processing* has been completed, the data is structured, clean and ready to be *categorised*.

5-2-4 Categorising

The third step in the SBI procedure consists of *categorising* the social media posts. The purpose of this step is to divide the messages into clustered categories at which the firm is interested. The aspects at which the social media posts are categorised may vary. Figure 5-5 schematically shows the third step of the SBI procedure.

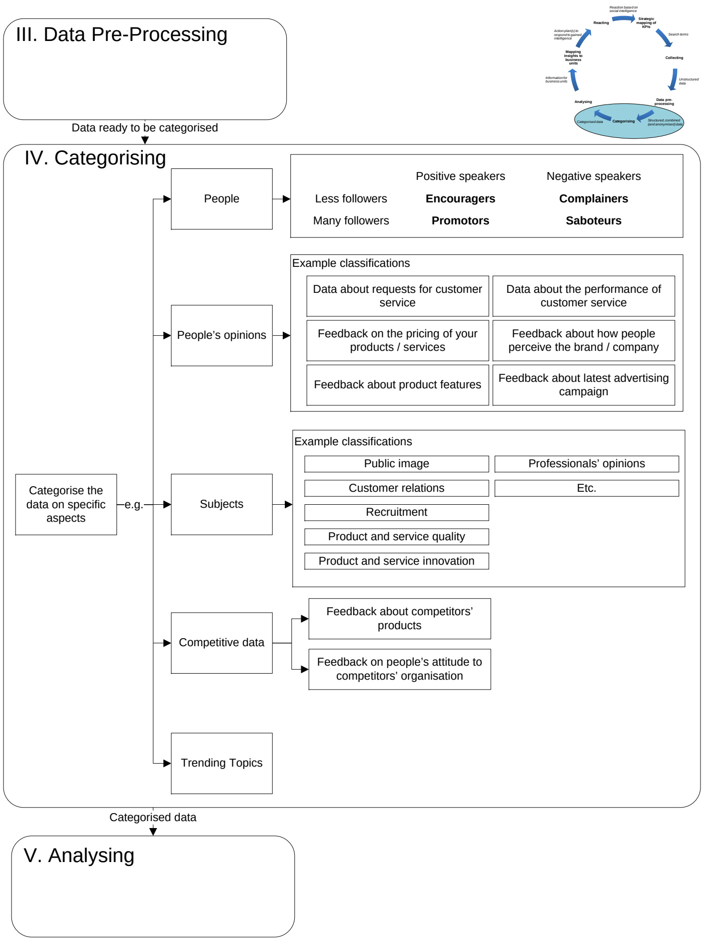


Figure 5-5: Blueprint: Social Business Intelligence Procedure (Categorising)

One can decide to analyse the people that create the messages, and group these people in e.g. people with many/less followers or friends, or into people that write/negative positive about the firm. We have labelled the four categories of people. Encouragers are the people with less followers though speak positive about the firm or its products. Complainers are the people with less followers and write negative about the firm. People with many followers who speak positive about the firm have been labelled as promoters, while people with many followers writing negative have been termed saboteurs. An analysis of the people provides the firm with intelligence about the power of the people that write about the firm, and may form the starting point of a social media engagement strategy.

Another aspect at which social media messages may be classified is based on their subjects. Our content analysis of chapter 4 also categorised social media messages based on their subjects. The subjects that were represented in our dataset related to public image, customer relations, recruitment, product and service quality, product and service innovation, professionals' opinions, etc. By classifying posts into categories based on subjects, it becomes possible to link the volume of messages related to a certain subject to the companies' corresponding KPIs. For instance, public image posts – which may be additionally classified as positive, neutral or negative – are related to a customer satisfaction KPI. There are plenty of other categories that one can think of to categorise social media messages, but to link the firm's KPIs to social media data, one should classify the messages based on their subjects.

Whereas the data on social media is generally publicly accessible, it is possible for a firm to perform the same analysis based on search queries related to competitors and competitors' products. As such, a competitive analysis provides the company intelligence about their position with respect to the market average.

Furthermore, word counts can be used to determine so called trending topics; topics that are over-represented in the social media messages related to the firm. Trending topics, or a top 10 of the words that are most frequently used in the social media messages, provide a firm insight in the topics that are discussed on social media in relation with their firm.

5-2-5 Analysing

Once the social media data has been structured and cleaned, the *analysis* of these data can commence in step 5 of the SBI procedure. It is in this step of the procedure where a translation is made from data to information. This step is visualised in figure 5-6. Depending on the matter of interest, a firm can analyse a variety of data and relations. It would be wise to at least plot the conversation volume – or amount of social media messages related to the firm – against the different social media channels to determine where the conversations related to the firm take place. Next, whenever a category has been established in step 4 in which all messages related to a certain product or product feature have been grouped, it is possible to determine the attitude of the public to the product by applying sentiment analysis on these data. Such analysis provides the firm with insight in the the products or product features that are to be improved. Furthermore, a comparable analysis on competitors' social media data will show the firm's position pertaining to the competitors and competitors' products.

The most valuable intelligence will be gained when the firm combines the social media metrics – such as amount of mentions, sentiment, messages originating from a certain region, etc. – with the companies' KPIs, such as sales volume, market share, customer satisfaction and the amount of customers. The slopes that will be gained when these metrics are together plotted on a time chart may reveal relations. The right part of figure 5-6 illustrates such graphs. A correlation analysis may confirm these relations. The intelligence that is gained in the analysis phase may reveal that certain social media metrics are under performing, and that these social media metrics influence key-performance indicators of the company. Consequently, a firm may undertake actions to improve these metrics.

5-2-6 Mapping insights to Business Units

Key-performance indicators are related to different departments in a firm, and the managers of these departments may clarify the under-performance of the metrics and they may suggest actions to improve the KPIs. As figure 5-7 illustrates, the intelligence provided by step 5 should be communicated to the responsible business units. Especially when under-performing KPIs are discovered.

For example, insights related to products should be communicated to the firm's research & development department, customer satisfaction intelligence to the firm's customer relations management department, etc. It are the employees of the responsible departments who posses the knowledge and experience to reason why a KPI is under-performing, and – in collaboration with social media experts – are the ones who may develop an action procedure to improve the indicator.

5-2-7 Reacting

The final step in the SBI procedure comprises the execution of action plans required to improve under-performing KPIs by means of social media. Illustratively, figure 5-8 shows two type of actions that may result from the social business intelligence procedure. A firm can for instance decide to review its products (features) based on complaints and suggestions that the SBI procedure provided. Or, a firm may decide to intervene in social media discussions, for instance because customer satisfaction turned out to be low, and – at the same time – the customer service of the firm turned out to be insufficient.

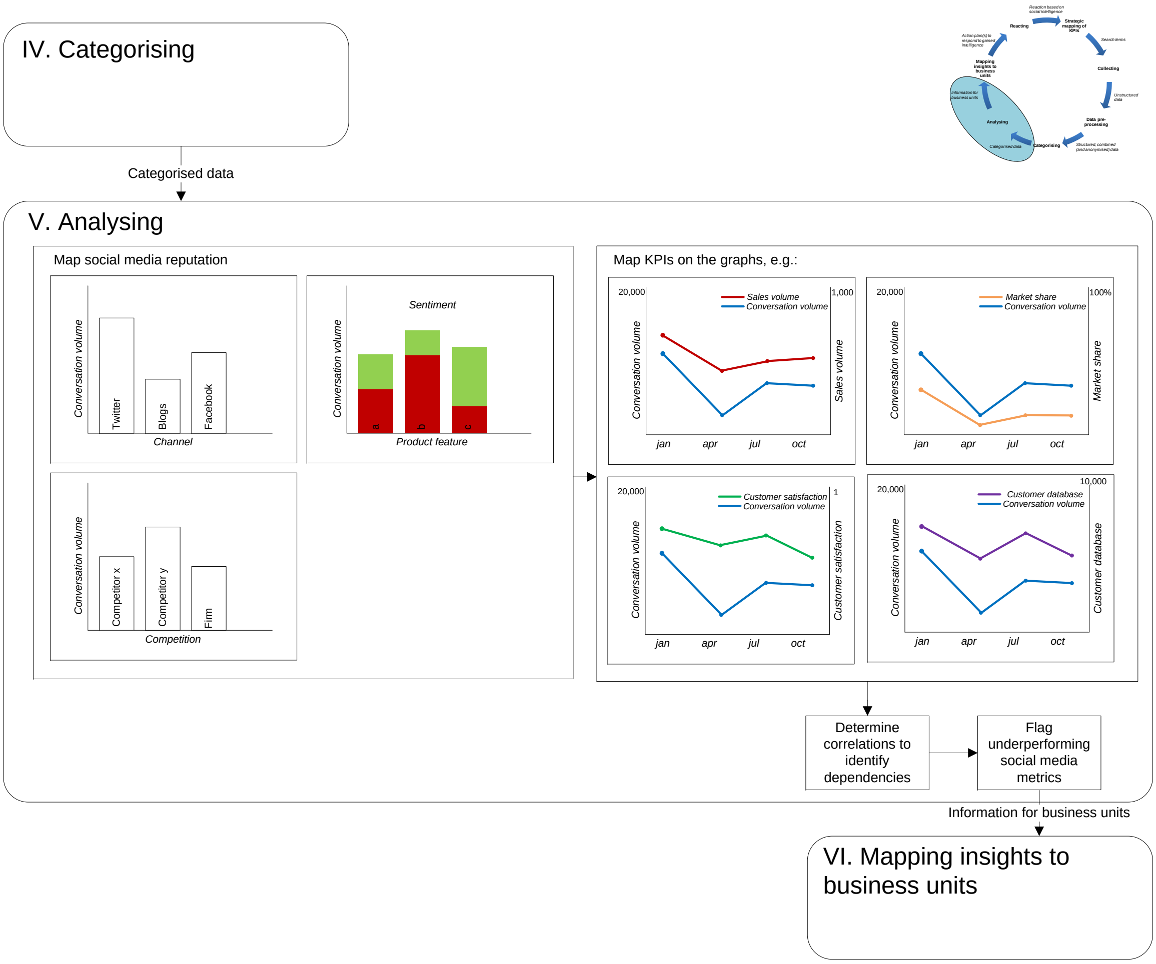


Figure 5-6: Blueprint: Social Business Intelligence Procedure (Analysing)

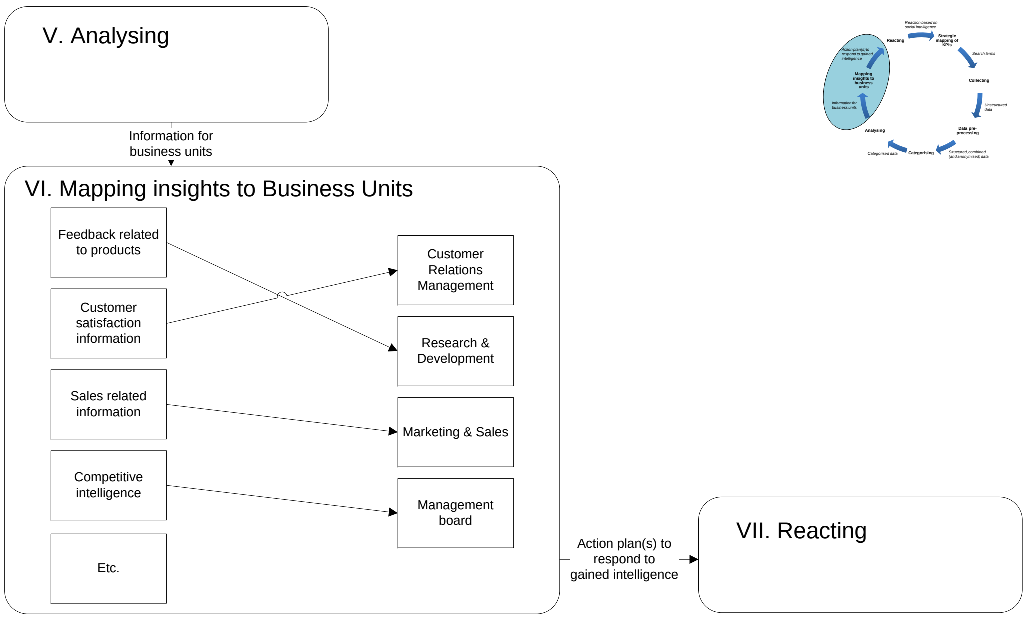


Figure 5-7: Blueprint: Social Business Intelligence Procedure (Mapping Insights to Business Units)

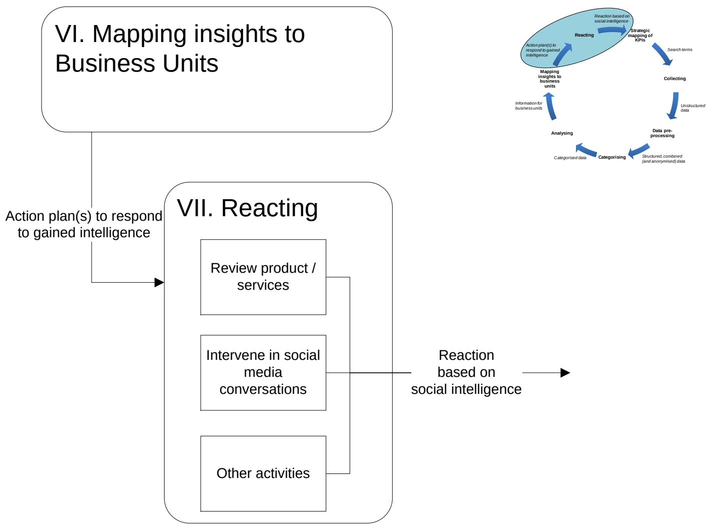


Figure 5-8: Blueprint: Social Business Intelligence Procedure (Reacting)

5-3 Verification of Procedure

The verification of the developed social business intelligence procedure is verified in this section. We will test whether the requirements established in section 5-1 are fulfilled. For each individual requirement an activity is searched for that fulfils the requirement. If all requirements are fulfilled by at least one activity, we can conclude that the social business intelligence procedure is verified in accordance with the requirements. Table 5-3 (page 84) lists the seven main components of the SBI procedure in the columns, and the eighteen requirements in the rows of the table. For each requirement, the activity that serves this requirement has been checked. As can be concluded, each of the eighteen requirements are at least fulfilled by one of the main components. Therefore, we can conclude that the procedure is in accordance with the requirements, which are in turn in accordance with the activities required for general business intelligence.

Table 5-3: Verification Matrix

	SBI Components	Strategic mapping of KPIs	Collecting	Data pre-processing	Categorising	Analysing	Mapping insights to Business Units	Reacting
	Requirements for a social business intelligence procedure	I	II	III	IV	V	VI	VII
1	Access to social media platforms		✓					
2	Identify the social media platforms at which the firm is discussed		✓					
3	Identify the volume of social media messages related to the firm		✓					
4	Remove the spam from social media messages that initially seemed to relate to the firm			✓				
5	Anonymise personal data			✓				
6	Identify who the people are that discuss the firm on social media				✓			
7	Identify what the subjects of the social media messages related to the firm are				✓			
8	Determine whether the information contained in the social media messages related to the firm offers additional value	✓			✓			
9	(Automatically) Classify the social media messages related to a firm into categories				✓			
10	Relate the (categories of) subjects of the social media messages to the firm's key-performance indicators	✓						
11	Determine the firm's social reputation					✓		
12	Determine the social reputation of the firm's product(s)					✓		
13	Determine relations between social media metrics and the firm's (social) key-performance indicators					✓		
14	Update the status of the social media metrics and the values of the KPIs constantly					✓		
15	Present the slope of the relations between social media metrics and KPIs on a time chart					✓		
16	Interpret the gained intelligence and position it into the firm's developments					✓	✓	
17	Assign the gained intelligence to the right persons in a firm						✓	
18	Allow a firm to engage on social media platforms							✓
19	Regularly update the search terms to anticipate on changes	✓	✓					

5-4 Real-Time Social Business Intelligence

In the introduction of this thesis, the concept of real-time business intelligence has been introduced. One of the aspects that makes social media data valuable is the fact that it is created real-time and that these data are directly available. The real-time aspect of social media data is one of the main reasons why this thesis has been executed in the first place. In this section we pay attention to the speed of the social business intelligence procedure as it is presented in section 5-2.

In our analysis, a scan for new social media messages has been executed on a daily basis. However, it would be valuable for a firm to be informed directly whenever social media activity related to the firm deviates from its steady state. For example, an increase in the average amount of hourly firm-related messages may indicate an event of which the firm should be aware from a risk management perspective. Deviations in the volume of social media messages are relatively easy to detect, since detection systems simply count the amount of messages that has been generated in the past period and compare this amount with the average amount. A scan to detect variations in the volume of firm-related messages should be executed periodically, the results are then almost immediately available. Figure 5-9 shows an illustrative example of a comparison between the average volume and the actual volume of today. Such a graph would announce a firm that it is suddenly more frequent subject of discussion on social media than in normally. The commercial tool that has been used in this thesis to analyse social media automatically refreshes the firm-related messages, comparable to the streams offered by Twitter.

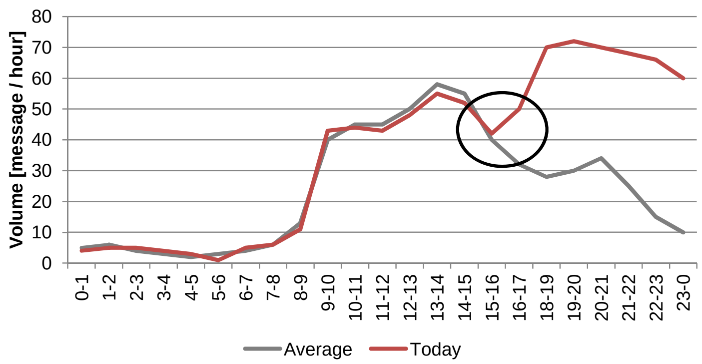


Figure 5-9: Illustrative example: Variations in volume of firm-related social media messages

However, insight in deviations in the volume of firm-related social media messages is not sufficient to speak of social business intelligence. As the procedure in the previous section illustrated, a process of pre-processing, categorising, analysing and mapping insights to business units is required after a firm has determined which key-performance indicators are to be measured by social media data is required. All these steps are to be automated by one critical mechanism; text classification. A tool that is able to automatically classify social media posts into pre-determined categories – e.g. subjects – is a prerequisite for real-time social business intelligence. As we have seen in the content analysis of this thesis, manually classifying social media posts is a time-consuming process and therefore not advised. The further development of automatic text-classification tools and the incorporation of such tools in social media monitoring packages is advised. The programming of each automatic text classifier requires a training set of texts that have manually been classified. Our dataset can be used for such purposes. Furthermore, the automated classification is to be linked with internal data. So far, business intelligence tools like SAP, Oracle, QlikView, etc. manage and process internal data, whereas dedicated tools like uberVU, NetBase, Radian6, etc. are used to process external – social – data. Thus, data from different systems is required to measure the influence of social media metrics on organisational performance. Such systems are to be developed.

In section 3-3-1 we concluded that the current state of social business intelligence can be termed 'early adoption'. With the experiences gained in chapter 4 and the required features mentioned in the previous paragraph, the next (research and development) steps required for social business intelligence systems are determined. These steps are further elaborated in section 6-6 'further research'.

5-5 Social Business Intelligence versus Business Intelligence

The newly developed social business intelligence differs from 'traditional' business intelligence methods. In this section, the most important differences concerned with social business intelligence are discussed. Table 5-4 lists the main differences between the two concepts, which are consequently elaborated.

Table 5-4: 'Traditional' and Social Business Intelligence compared

'Traditional' Business Intelligence	Social Business Intelligence
Data is structured	Data is unstructured
Only the values of the data automatically fluctuate in the course of time	The nature of the data may alter in the course of time
Data are mainly numerical	Data are mainly textual
Relations between data and KPIs are obvious	Relations between data and KPIs are fuzzy
Few sources of data	Multiple sources of data
Data sources are internal	Data sources are external
One data format	Multiple data formats
Origins of the data are known	Origins of the data are unknown
Data represents known subjects (products in stock, sales per month, etc.)	It is beforehand not clear what is contained in the data (complaints, suggestions, etc.)
	Spam
	Data may contain personal data for which no explicit consent of data processing is provided

The far most important aspect of social media data is the fact that it is unstructured. When collecting social media data, one big mishmash of textual data is found containing different subjects in different languages. On the contrary, 'normal' BI systems source data from structured sources in which pre-determined variables are stored. As a consequence, the collecting (also referred to as extracting) process will deliver structured data in any case. In other words, in normal BI systems it is known beforehand what will be measured, whereas the subjects of social media messages differ from firm to firm.

Additionally, the nature of the social media data may alter in the course of time. Due to the fact that anyone who has access to social media is able to formulate new subjects, it is not unlikely that the subjects contained in social media messages may vary. This aspect of social business intelligence is not found in normal BI processes. In normal BI the to be measured variables are pre-determined, the values of these variables will alter in the course of time. However, the nature of the data will not change. E.g. a metric measuring the amount of products that are currently in stock will not suddenly start measuring the number of employees or any other variable. However, subjects contained in social media messages may change. Therefore, social business intelligence systems are required to cope with fluctuations in the nature of the data.

Apart from (simple) word counting mechanisms, whether purposed to show the volume of firm-related messages or to show trending topics, social media data is textual in nature. Fundamental data analysis methods and algorithms underlying traditional business intelligence are developed to process numerical data. Therefore, a translation step is necessary from raw textual data to numerical data before such analyses can be executed on social media data. In our social business intelligence procedure this step is provided by categorising the textual data. The categorising step ensures that equivalent social media posts are grouped – i.e. structured – so that the number of messages in each group serve as the basis for numerical analyses. E.g. if all messages containing the subject 'product x' have been grouped in one category, the number of the messages in this group are ready to serve as the input for further (numerical) analyses.

In this thesis we have assigned social media posts to key-performance indicators based on the subjects contained in these messages. Whereas in traditional business intelligence relations between data and KPIs are obvious and linear, these relations are less evident in social business intelligence. For instance, the volume of messages related to a certain product may influence the sales of that product. However, this relation is not self-evident since a causal relation between the two variables is not guaranteed. Other factors – such as lotteries – may influence the chatter volume, but this does not necessarily underwrite the intention of people to buy the product. The fact remains that these relations may still exist, and it is therefore that monitoring and analysing social media data in relation with the firm's KPIs may reveal valuable intelligence. As indicated in chapter 4, KPIs related to customer relations and the perceptions of stakeholders can be measured using social media data.

Whereas in traditional BI data is stored in relatively few (structured) data sources, there are many social media platforms from which data is to be sourced. These sources differ from traditional BI sourcing systems since they are in the external environment of a firm, implying that the data formats and other institutional aspects are determined by the social media platforms. Furthermore, the platforms can change the format in which the data is delivered. Sourcing data from multiple social media platforms means combining different formats. This step requires more effort than in traditional BI systems. In addition, the attributes that are passed to a firm when it crawls social media data differs per platform.

The creators of the data – social media users – are unknown to the firm in social business intelligence. As such, it can be hard to determine the trustworthiness of the data. A user can post whatever he or she wants on the web, without ensuring that the message actually reflects his or her opinion or intention. However, there are ample examples of self-regulating platforms on the web, of which Wikipedia is probably the most famous. Contributions of users to Wikipedia that are incorrect are automatically corrected by other users with good intentions. Moreover, since messages are grouped in our procedure, it is rather easy to target the popular topics (which require attention) and determine the trustworthiness. Next, we expect that natural language processing tools will be improved, allowing the detection of cynicism and other difficulties concerned with social media data.

In traditional BI systems it is beforehand crystal clear what will be measured, e.g. the time to assemble a product from five components, which clearly affects the operating expenditures of a company through the cost of workers. In social business intelligence, the contents in the data are not clear beforehand. As we have seen in chapter 4, subjects of social media messages differ from firm to firm. Thus, not each KPI that a firm is willing to be measured by social media data can actually be measured by these data. It is therefore that the contents of the social media data determine what can be measured. Each company can be willing to measure KPIs by social media data, however without existence of any data, this will not be possible. In traditional business intelligence, a firm is much less dependent on external stakeholders for the possibilities of BI.

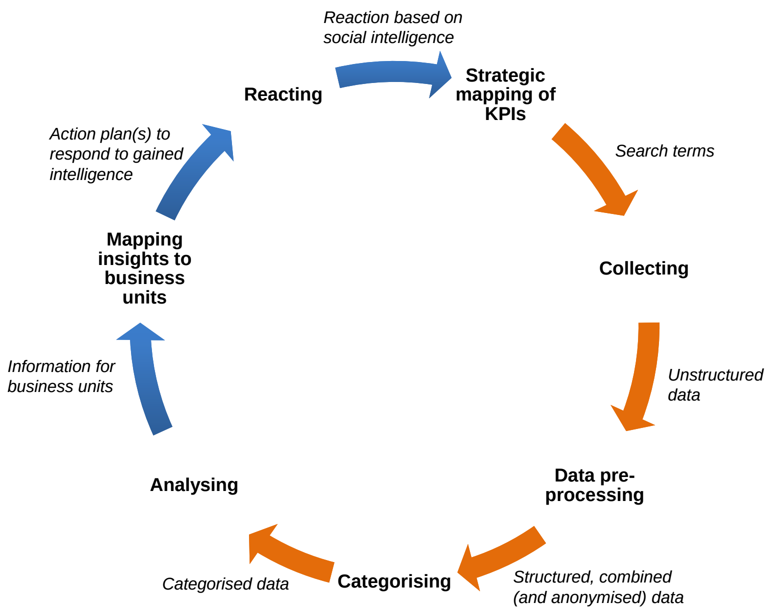


Figure 5-10: Specific Steps in Social Business Intelligence

When recalling the cycle visualising the social business intelligence procedure developed in section 5-2, the specific steps required in social business intelligence can be highlighted. Figure 5-10 shows the social business intelligence cycle, in which the specific social BI steps are highlighted in orange. In the orange steps, a different method is required as compared to 'traditional' BI. Since our procedure is based on existing business intelligence methods, the procedure also shows overlap. The steps that are relatively equal to standard BI are coloured in blue. As can be concluded from figure 5-10, different activities are mainly required in the collecting and processing steps of business intelligence. It is in these steps where the data is converted from unstructured to structured data that is ready for analysis.

5-6 Sub Conclusion

A social business intelligence procedure ("SBI") should fit within the general 'way of executing' business intelligence, since it is not possible that social media metrics measure *all* key-performance indicators as well as firm's internal BI systems will do. Therefore, SBI is considered as an additional component to business intelligence, rather than a replacing procedure. However, as the content analysis of chapter 4 illustrated, there are certain categories of KPIs that are influenced by – or at least related to – a substantial amount of social media messages that are related to the firm. For these type of KPIs, which differ from firm to firm, a procedure has been developed that prescribes the necessary steps to acquire, process and finally gain intelligence from firm related social media messages. The procedure is based on general BI concepts, existing technologic social media analysis solutions and the experience gained in the execution of a content analysis into the social media messages related to eighteen different firms.

A SBI procedure consists of seven main components, being (i) strategic mapping of KPIs, (ii) collecting, (iii) data pre-processing, (iv) categorising, (v) analysing, (vi) mapping insights to the business units, and (vii) reacting. The seven steps can be interpreted as a cycle, i.e. the output of the last step influences the first step.

(i) Strategic mapping of KPIs The very first step of social business intelligence sets the scene for the objects that are to be collected and analysed. Namely, in the first step the key-performance indicators that are to be measured by social media data are selected. As we have seen in chapter 4, not each type of KPI is to be measured by social media data since there does simply not exist any related social media data to these types of KPIs. Firms should mainly focus on KPIs related to customer relations, public image and – to a less extent – on product and service innovation when selecting KPIs that are to be measured using social media data. Whenever a firm has selected the social KPIs, it can start *collecting* the appropriate data.

(ii) Collecting The second step of the SBI procedure related to data collection. In contradiction to regular BI systems, the data is to be sourced from external parties in social business intelligence. People create firm-related messages on different platforms, of which the vast majority of publicly accessible messages are created on Twitter. The search terms that are used to filter out the content at which the firm is interested should be based on the social KPIs selected in the previous step.

(iii) Data pre-processing The social media data has been collected from multiple platforms which adhere to their own data format. The different format are to be combined into one uniform database, so that – in a later step – data analysis can be applied on the complete dataset. Furthermore, the firm should select those attributes that are necessary for the analysis, not each platform offers the same richness of attributes to a social media post. In addition, the data should be anonymised to be in compliance with new Regulations regarding data privacy. Finally, spam – i.e. social media posts that do not relate to the firm – should be removed from the collected data.

(iv) Categorising The *data pre-processing* step resulted in a structured database in which the social media messages from multiple platforms are combined. In the *categorising* step, the messages are clustered on different issues of interest, depending on the firm's subject of interest. E.g., messages related to certain products can be categorised, or one can cluster the messages that are created by people with many followers, etc. Again, the criteria at which the messages are categorised are determined by the selection of the social KPIs in the first step.

(v) Analysing So far, the collected data has not provided any insights. It is in this step of the social business intelligence procedure where data is transformed into information. The categories that were established in the previous step are analysed in this step. For instance, sentiment analysis can be applied on the categories related to the firm's products in order to acquire intelligence related to customer experiences of the products. However, the most valuable intelligence is gained when social media data is related to internal data. For instance, the volume of social media messages related to a certain product may be correlated with the sales volume of that product. It is in this phase of the SBI procedure where such relations are explored.

(vi) Mapping insights to business units In the first step of the procedure, KPIs have been selected. These KPIs typically relate to a certain function of the firm, and hence have an 'owner'. The intelligence gained in the previous step relates to KPIs, and should feed back to the owner of the KPI. Generally, it are the people in the firm that are responsible for the KPI who are the ones that can reason how the KPI is influenced. Therefore, these people are the ones that can draft an action plan in case the KPI needs improvement.

(vii) Reacting The final step of the social intelligence procedures consists of the execution of the action plans that are developed in collaboration with people from the business lines that are responsible for the respective KPIs. Actions on the gained intelligence may involve revisions of internal processes or strategies, or external interventions such as social media engagement.

The developed social business intelligence procedure is based on general business intelligence processes, and the requirements of the social BI procedure have been derived from general BI processes. For the verification of the SBI procedure, all requirements have been checked on fulfilment by systematically tracking which requirement is fulfilled by which activity.

Chapter 6

Conclusions & Discussion

First of all, this chapter presents the conclusions of the research in section 6-1. In section 6-2 the contributions of this work to existing and future research are discussed. Section 6-3 proceeds by discussing the implications of the findings in this thesis for practice. Section 6-4 consequently reflects on the thesis and the research process. In section 6-5 the research is critically reviewed, and limitations are discussed. Finally, section 6-6 provides suggestions for future research related to the subject of this thesis; social business intelligence.

6-1 Conclusions

Firms are increasingly using social media, while at the same time business intelligence systems are increasingly applied for performance measurement of business activities. Though these two concepts offer room for synthesis, it also raises questions related to the applicability and opportunities offered by combining social media and business intelligence. So far, it is not clear which firms are able to find firm-related social media data and if they are able, how these data should be incorporated in the business intelligence processes of firms. As one of the first researches into the opportunities of leveraging social media data for BI purposes, this this was purposed to draw generic conclusions on the applicability of social business intelligence by distinguishing firms on generic aspects. Firms were distinguished on *customer relation* type – either B2C or B2B – and on *industry* type. Therefore, the main research question of this thesis has been formulated as:

> *How can firms use social media data for business intelligence, taking into account the firm's specific industry and relationship with end-users?*

The main research question has been divided into three sub questions, which are answered in the following sections. The first sub question was defined as:

1. *What is the current state of social media in relation with business intelligence?*

Social media is a natural consequence of Web 2.0, and can be defined as Web 2.0 based applications allowing users to create and share user-generated content with pre-selected users and/or communities. The applications through which users are active are known as social media platforms. In 2012, there are many social media platforms available, which differ in scope and functionality. Each platform adheres to its own policy regarding data crawling and data format.

Social media is a topic on the agenda of many firms in 2012. Though many firms acknowledge the opportunities of social media, there also exists a degree of reluctance from managers towards social media. Research indicates that executives who avoid social media do not understand what social media is, how to engage with it and learn from it. On the other hand, firms that *do* embrace the world of social media particularly perform activities in the field of *marketing*, *customer relations management*, *reputation management* and *co-creation / prosuming activities* through the various social media platforms.

Business intelligence – irrespective of the variables to be measured – can be perceived as a cycle consisting of three main steps; (i) register, (ii) process and (iii) react. Before a BI cycle can commence, it has to be determined 'what to measure'. The variables that are to be registered are generally aligned with a firm's strategy and corresponding business model, termed key-performance indicators ("KPIs"). The three BI steps are required when a firm intends to apply business intelligence on the firm's social media activities. However, social media data differs from "regular" business information. Unlike internal business data, social media data is created by non-professionals and stored into a variety of databases that are owned by external parties who employ their own database structure and access limitations. Therefore, a different BI approach is required for social media data.

Firms employ different key-performance indicators. Especially lower level KPIs are firm specific, while top level KPIs are generic and employed by many firms. Based on Adam & Neely's (2001) generic performance prism perspectives, ten categories of key-performance indicators have been established. The ten categories are defined as *(short-term) financial results*, *customer relations*, *employee relations*, *operational performance*, *product and service quality*, *alliances*, *supplier relations*, *environmental performance*, *product and service innovation* and *community.* It are these categories of KPIs for which related social media data has been searched for.

In social business intelligence, a firm analyses the activities on social media related to the firm and determines the effect of these activities on the firm's performance. Existing social media monitoring tools – which are becoming increasingly available on the market – mainly reveal the performance of the firm on social media as a separate component of the firm. The intelligence that such monitoring tools provide relate to the *volume of posts*, *engagement of users*, *sentiment*, *geography*, *topics and themes in the social media messages*, *influencer ranking*, *channel distribution*, etc. However, the purpose of business intelligence is to reveal the underlying parameters that determine the firm's performance, that is, not limited to solely social media performance. In order to understand the influence of social media activities on the firm's performance, a link between the company's key-performance indicators and social media parameters is required. In social business intelligence, such links are required.

The second sub question was formulated as follows:

2. *In which contexts are firms able to acquire social media data for business intelligence?*

In this research, the context of a firm has been described based on two generic dimensions. Firstly, firms were distinguished from each other based on the industry in which they operate. Secondly, firms' contexts were described by distinguishing different customer relations types; i.e. B2B or B2C relations. The volume of messages that contain the name of a firm differs from firm to firm. E.g. in our sample 39.425 messages related to Heineken have been collected, while during the same period only 428 messages related to Fugro have been found. Our analysis indicates that there exists variation in the volume of firm-related social media posts across different industries. Firms classified as *industrials*, *information & communication* were more frequently subject of discussion on social media than *consulting* or *mining & quarrying* firms. Our analysis also illustrates that there exists variation in the volume of social media posts across B2B and B2C firms. B2C firms are far more often subject of discussion on social media than B2B firms.

Apart from an assessment of the volume of social media content related to firms, we analysed the subjects of the messages in order to gain an understanding of the type of information contained in the social media messages. Our analysis shows that the subjects of social media messages differ from firm to firm. The majority of social media messages related to firms (41%) express how the external stakeholders of a firm perceive the company. In this thesis, such posts have been classified as *community* posts. 18% of the social media messages in our dataset contained the name of a firm, but did not contain any valuable information for the firm and have consequently been assigned as *undefined* posts. About 11% of the social media messages relate to *financial results*, which consist of *financial performance discussions* (5%) and *stock related discussions* (6%).

The content analysis of this research suggests that the subjects of social media messages related to B2B firms contain a higher percentage of *short term financial results*, *news* and *professionals* related messages than messages related to B2C firms. Unfortunately for B2B firms, such type of information is yet available internally. Acquiring social media data to gain additional management information is therefore of less value for B2B firms. Next, the analysis indicates that the social media messages related to B2C firms contain a higher percentage of posts related to *customer relations*, *product and service quality* and *product and service innovation* than messages related to B2B firms. It are these types of information that deliver additional value to the firm, since this information is not available at firms internally.

In addition, the content analysis of this research suggests that the subjects of social media posts differ between industries, but that the majority of the subjects in each industry relates to community, i.e. social media posts revealing how the community perceives the company. The results indicate that firms active in the *information & communication*, *financial institutions* and *transport & storage* industries are more subjected to social media messages related to *customer relations*, while firms active in the *mining and quarrying* and *consulting* industries will find messages related to *financial performance*.

As indicated in chapter 1, this thesis describes a firm's context based on two dimensions; *customer relation type* and *industry type*. By distinguishing firms based on customer relation type, we can state that B2C firms are able to acquire (i) a high volume of social media messages related to their firm and (ii) social media messages that contain information that is not yet available in the internal information systems of the firm, hence enriching the business intelligence. Next, when distinguishing firms based on their industry we conclude that the volume of firm related messages differs between firms. Additionally, the analyses suggest that the subjects of the social media messages differ from industry to industry. However, in our sample, there exists interaction between the customer relation type and the industries. The differences in volume and subjects are more visible when distinguishing between B2B and B2C firms rather than distinguishing between industries. Table 6-1 summarises these conclusions.

Table 6-1: Conclusion of Content Analysis

	Volume of Social Media Messages	**Subjects of Social Media Messages**
Customer Relation	B2C firms are more often subject of discussion on social media than B2B firms.	B2C firms related social media messages are more often subjected to *customer relations*, *product and service quality* and *product and service innovation* than B2B related firms. On the other hand, B2B firms' related messages are more often subjected to *financial results*, *news* and *professionals* discussing the firm. However, the information contained in the social media posts of B2B firms is often yet available to the firm, hence not offering added value to the firm's richness of management information.
Industry	Our analysis shows a variation in the volume of social media messages across different industries.	Our analysis indicates that there is a difference in the subjects of social media posts related to firms in different industries.

The third sub question of this thesis has been formulated as:

3. *Which processes are required to incorporate social media data into general business intelligence frameworks?*

'Social' data differs from internally generated and collected data on the following aspects. Firstly, social data has not been verified before it is published. Anyone can create a social media message, it will not be verified before it is available to the world. Social media messages may contain jokes or cynicism, making it hard for firms to interpret what the writer of the message actually means. Secondly, social media data is unstructured. The data are sourced from multiple sources in different formats and languages. Each source may employ its own structure of social media messages, and not each platform may contain the same richness in attributes as the other. Thirdly, the unstructured nature of the data and the huge amount of data that is generated makes it that social media data can be labelled as 'big data', implying that the problems of big data may also be applicable on social media data. Fourth, as with many data on the web social media messages contain spam, which is to be removed before commencing an analysis of the messages. Fifth, while internal data may represent evident relations (e.g. between the number of employees and the firm's revenues), relations between social media metrics and key-performance indicators are less evident. When a firm intends to add a social component to its existing business intelligence, these aspects should be considered.

Structuring social media data is an important activity required to make an analysis on such data. Firstly, the sourced data has to adopt one and the same data structure. Whereas each social media platform will deliver data to its own favour – e.g. by CSV, JSON, XML, or other formats – the data should be parsed into one common data format. Secondly, dividing the messages into categories makes the dataset ready for interpretation. Many types of categories are possible. A classification based on the people reveals which users are actively engaged with the firm, which users have much power (in terms of followers and friends), which users speak positive about the firm, etc. A classification based on the subjects of the social media messages reveals which topics are considered important by the social media users, and, more important a classification based on subjects allows a firm to link social media messages to the firms' key-performance indicators. For instance, public image posts – which may be additionally classified as positive, neutral or negative – are related to a customer satisfaction KPI. Though there are plenty of other categories to classify social media messages, to link the firm's KPIs to social media data the messages should be categorised based on their subjects.

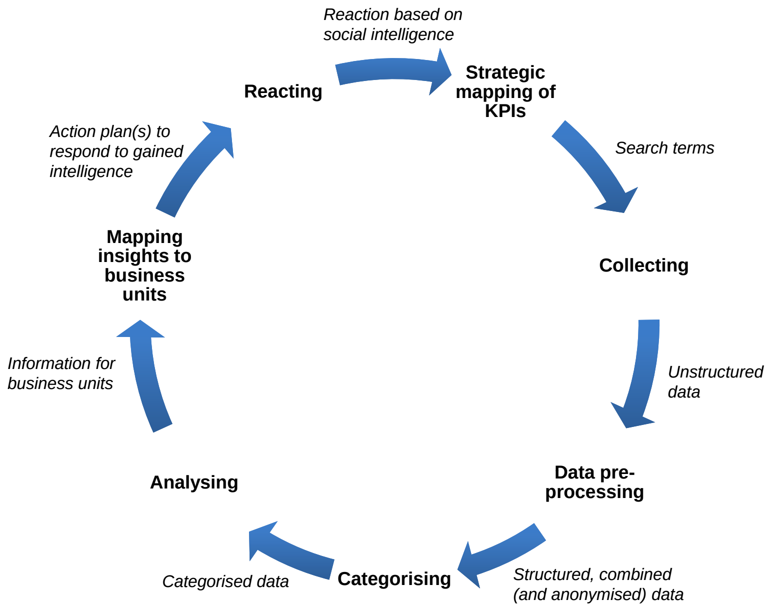


Figure 6-1: Blueprint: Social Business Intelligence Procedure

A social business intelligence ("SBI") procedure prescribes how a firm should collect and process social media data to gain intelligence, at which the firm can consequently base their decision-making. A SBI procedure consists of seven main components, being (i) strategic mapping of KPIs, (ii) collecting, (iii) data pre-processing, (iv) categorising, (v) analysing, (vi) mapping insights to the business units, and (vii) reacting. The seven steps can be interpreted as a cycle, i.e. the output of the last step influences the first step. Figure 6-1 schematically shows the social business intelligence procedure.

At the start of this thesis, the objective has been formulated as:

> *The objective of this research is to develop a procedure to utilise social media data for business intelligence, for which the applicability is investigated for firms in different industries and for different relations with end-users.*

Taking into account the answers on the research questions, we state that the objective of this thesis has been achieved. A social business procedure has been developed, verified on consistency with general business intelligence processes and tailored to the challenges arising from processing social media data. In addition, the applicability of this procedure is investigated. The results of our study indicate that especially firms performing B2C relations are able to execute social business intelligence, because (i) these firms are subject of discussion on social media, hence firm-related social media exists for these firms and (ii) the information contained in B2C related messages offer additional information for the firm. Furthermore, the results of this study also indicate that there exists a difference in the volume of firm-related content between different industries, in which *industrials* and *information & communication* firms are more frequent subject of discussion on social media than *consulting* and *mining & quarrying* firms.

6-2 Contributions to Research

Social media is a hot topic in the academic world. Existing research in the field of social media is often aimed at marketing efforts or other activities in which the firm expresses or should express itself to the outside world. This thesis focused on the incoming information from a firm point of view, i.e. the extraction of information from social media to support decision-making. The fact that this thesis investigated the applicability of social media data for organisational decision-making, makes it that this thesis touches the world of business intelligence. In our opinion, this aspect distinguishes this thesis from other research.

Future research aimed at deriving information – in whatever form – from social media for business purposes, should be aware that the type of firm affects the applicability of such activities, and that one should not draw generic conclusions applicable to all firms. As this research indicates, the existence of and subjects contained in firm-related social media messages differs between firms. Furthermore, the method of this research is partly based on traditional content analysis, in which a new type of data has been analysed. In the following section, our experiences of applying a content analysing on social media data are shared.

6-2-1 Methodological Innovation

The content analysis methodology is not new. Krippendorff (2004) refers to propaganda analysis during World War I as one of the earlies structured approaches of analysing texts. However, the fact that this thesis performed a content analysis on social media messages makes it that part of the research in this thesis is innovative. The structured approach that was executed was based on earlier work from Bos and Tarnai (1999), whom created a framework to execute a content analysis. Though their framework did not speak of social media platforms or data, we found that their research framework – with adjustments – is also applicable on social media data. The adjustments of – or additions to – the framework relate to the data collection and preparation steps of the content analysis procedure. Whereas textual data is considered structured in the framework of Bos and Tarnai (1999), a content analysis on social media data requires a data structuring process before commencing the categorisation process. The strength of the framework lies in the fact that it is generic, hence applicable in many domains. With the experience of the execution of the content analysis on social media data in this thesis, we recommend using the framework of Bos and Tarnai (1999) for future social media content analyses. Figure 6-2 shows the additional steps required when applying a content analysis on social media. These steps are highlighted in blue, and are established based on our experiences in the execution of the content analysis on social media data.

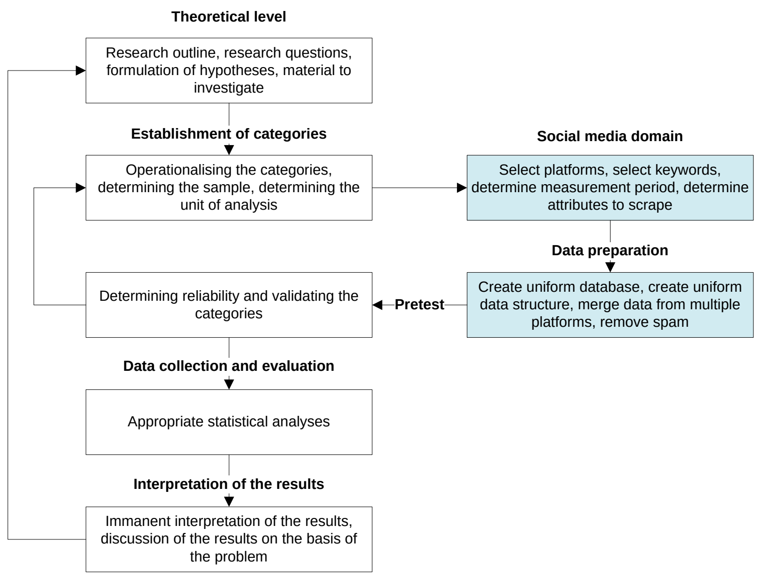


Figure 6-2: Adjusted (Social Media) Content Analysis Procedure

The additional steps required when applying a content analysis take place after the second step. After the research questions, the formulation of hypotheses, the determination of the categories that are to be analysed and the sample establishment, the social media domain comes in. In this step, the researcher needs to determine from which platforms the data is to be sourced. Each platform adheres to its own data format, and not each platform provides access to all posts. Furthermore, each platform has its own focus, implying that different people are active on different platform. Next, the keywords are to be determined. Comparable to search engines, social media scrapers scan for keywords in the many posts created on the web. A researcher may decide to base its keywords on user names, hashtags (subject of message, assigned by the creator of the message), the researcher may decide to scrape all posts in a certain area, or during a time. Other elements to select the messages that are to be analysed are also possible. Next, the attributes to scrape are to be determined. Each social media post exists of various attributes, e.g. user name, time, location, content, hyperlink, etc. It differs per platform which attributes are shared.

Next, the data is to be prepared. In case that the research exists of data sourced from multiple platforms, the data is to be merged and structured into a uniform database. In addition, the data is likely to contain spam. These messages are to be removed before commencing the analysis. The framework then proceeds in the original steps of Bos and Tarnai (1999).

6-3 Implications for Practice

The findings in this research have implications for (consulting) firms willing to use social media data for business intelligence. First, the findings of this research indicate that B2B firms are less likely to find social media data. Furthermore, if a B2B firm will find messages related to the firm, these messages are likely to contain information that is yet available to the firm. Hence, the applicability of social business intelligence for B2B firms is limited. On the contrary, B2C firms are often subject of discussion, and the messages related to B2C firms contain information that is not yet available to the firm internally. Taken into account the results of this research, the opportunities and promises of social media found in many reports and white papers are mainly applicable to B2C firms. Therefore, firms and firms offering consultancy on the domain of social business intelligence, should be aware that the opportunities of social business intelligence are limited to B2C firms.

In addition, a stepwise procedure for social business intelligence has been developed. Such a procedure was necessary to be developed since the new data source – social media platforms – differs from the systems at which normally data is collected and stored. This procedure is applicable on firms for which social media data is available.

6-4 Reflection

In this section, a reflection on the research process is presented. First, the developed Twitter scraper is discussed. Next, additional research steps that were executed whenever there was more time available are presented. Finally, a detailed stepwise approach of our data collection process is presented.

6-4-1 Twitter Scraper

During the early stages of this Master thesis project, a software tool has been developed that scrapes messages created on Twitter. The tool has been written in PHP language and is designed to work with MySQL databases, which we managed using phpMyAdmin. At the same time that the tool was up and running, access to one of the commercial social media monitoring tools (uberVU) was granted to the author of this thesis, for which we are grateful. Clients of uberVU pay a monthly fee of at least $1.000 to access the software. The tool allowed us to scan a variety of social media platforms, whereas our own tool solely scraped Twitter. Therefore, the decision was made to use the commercial off-the-shelf software rather than our own to collect the data. In addition, uberVU offers additional features like sentiment analysis, influencer ranking, location of the message, etc. that are not available in our tool. However, our own scraper – though it solely scrapes tweets – may serve research projects that are aimed at tweets.

6-4-2 If I had More Time

This research has been executed during a period of six months, i.e. from July 2012 to December 2012. The limited time available for this research has implications for both the depth and the breadth of the research, hence on the conclusions and the applicability of the conclusions. In case that the research time would be longer, we would have surveyed more firms so that the analysed categories (B2B/B2C and industries) would have consisted of more respondents, allowing for statistical analyses. Whereas the conclusions of this research are exploratory, the statistical testing of differences between groups of firms would allow for the generalisation of the statements. Furthermore, adding more respondents to the sample leads to more industries being represented in the sample, so that the conclusions of the research are also applicable on other industries. Next, we would have analysed more messages in the content analysis. The social media messages in this research have been collected during a period of two weeks. In case the messages would have been collected during a longer period of time, e.g. six months, the sample would have existed of more messages. As such, it would be possible to gain insight in the 'steady volume' of daily messages, and, more interestingly, deviations in the steady volume. Deviations might be due to the announcement of financial figures, marketing events, etc. However, to perform a content analysis on *many* social media messages, an automatic classifier is required. Manually classifying the posts, as we did in this research, would then take too much time. To train such an automatic classifier, the manually classified posts in this research are suited. Furthermore, if we had more time, we would have investigated whether or not other factors, such as the number of employees, revenues, market capitalisation, etc. also affect the number of social media messages that are related to a firm. With respect to the social business intelligence procedure that has been developed, we would have validated the framework by pilot projects and the involvement of business intelligence experts.

6-4-3 Stepwise Description of Data Collection Process

In this section we share our method of the data collection process. These steps are also incorporated in the adapted content analysis framework in section 6-2-1.

- *Determine keywords / search terms*
  After establishing the sample, the search terms required to filter out the related social media messages have been determined. In this thesis we searched for the firm names. However, it is also possible to search more specific, e.g. on the name of a product or a specific event.
- *Use search terms in social media monitoring tool*
  The search terms were consequently used in the social media monitoring tool.
- *Export the search results into a database*
  This functionality is not offered by each social media monitoring tool, but vital for the data collection process. Tools that do not offer the exportation of search results in whatever format are not suited for further analyses on the data because most analysis software requires the data to be stored on a local machine.
- *Structure the database*
  Depending on the output of the export process, the database is to be structured. The social media monitoring tools used in this thesis exported the messages into comma-separated values, which could easily be loaded into MS Excel. The richness of attributes offered by the social media monitoring tools determines the complexity of the structuring of the database.
- *Daily search for new results*
  In order to collect a large dataset, daily runs for new messages were executed. The social media monitoring tools used in this thesis offered to possibility to export up to 10,000 messages per search run. Pre-testing the collection process illustrated that this constraint was sufficient to get a complete picture of the firm-related messages created on the social media platforms by daily searching for new results. Whenever a tool has been used that offered a lower exporting capacity, e.g. 1,000 messages, the frequency of searching for new results would have been higher.
- *Verify that the new search results do not yet exist in the database*
  A daily run for firm-related messages resulted in the collection of duplicates, i.e. messages that were already collected yesterday. These messages have been identified based on their unique URL that was contained as an attribute to each message using LOOKUP functions in MS Excel. More specifically, the value of each URL was LOOKED UP in the existing spreadsheet. The textual format of these lookup values required some computer power, but our 4GB RAM / i5 machine turned out to have sufficient power for these calculations. Whenever a message did found a match, this meant that the message did yet exist. The messages that did yet exist were not added to the database.
- *Start analyses*
  The data collection process resulted in a structured database in a format so that MS Excel could handle the data for analyses.

Whereas the above steps are described in detail, the underlying ideas will be applicable on each social media data collection process.

6-5 Limitations

This research and consequently its outcomes have limitations that should be taken into consideration when adopting the conclusions of this thesis. The limitation are discussed in this section.

The first aspect of the research limitations, or aspects to consider when interpreting the conclusions related to the *population* from which the social media messages are drawn. This aspect has been assigned by Krippendorff (2004) as an important issue to consider when performing a content analysis. Not everyone uses social media, and even less people actually create content on the platforms. Therefore, firms that analyse social media data should be aware that these data do not represent the full (potential) client base. It is very likely that social media users have other preferences than non-social media users. A firm should always place the conclusions from social business intelligence in the light of their complete client base before it makes a decision to undertake actions, because the actions may only serve those needs of the ones that engaged on social media. Nevertheless, anno 2012 social media is relatively young. The user groups – e.g. age groups or countries – that use social media may increase in the coming years. We expect that social media will be further embedded in the lives of people that grow up in the social media era.

Secondly, each social media platform has its own privacy policy. This implies that a user either has the possibility to determine whether or not it shares its messages to the public, or that the platform determines the publicly available messages. As a consequence, only messages that were *publicly available* have been analysed in our dataset. It is likely that people who have not publicised their social media messages also discuss firms or firms' products / services, these messages are not available in our dataset. Still, the dataset is representative for firms conducting social business intelligence, since they will not get access to private messages either.

Third, not all social media *platforms* have been part of our analysis. Platforms such as Sina Weibo – the Chinese counterpart of Twitter –, Qzone, Renren, Habbo are not part of our analysis. However, the platforms that did exist in our sample are the ones used in the Western world. Therefore we state that the conclusions of our report are valid for Western world firms.

Fourth, the volume of messages, likes, shares, retweets, etc. can easily be influenced by a firm, though this does not necessarily mean that the user is actually engaged with the firm. For instance, a firm may decide to raffle an iPad or organise other lotteries. People can participate in the lottery by e.g. sharing a promotion message or by 'liking' the firm's page. Whereas such activities certainly lead to an increase in the number of likes, shares, etc., the underlying reason why people pay attention to the firm is for the price, and not necessarily the engagement in the company or its products. We refer to content that is generated according to such mechanisms as *biased chatter*, and doubt if such activities actually lead to an improvement of the firm's KPIs, e.g. the number of sales.

Fifth, one of the conclusions of this research is that the *volume* and the *subjects* of social media messages differ among *industries*. Due to time restrictions, we have not been able to analyse the social media messages of more than eighteen firms. As a result, each industry consisted of two or three firms, which we deem a *small sample*. It is therefore that the conclusions of this research are to be perceived as exploratory rather than confirming hypotheses.

Sixth, our research grouped firms based on the industry type in which they are active. With respect to the volume of firm-related messages, intra group difference have been spotted. These notable findings indicate that the industry aspect is not the only determining factor influencing the volume of firm-related messages. Other factors, such as (world-wide) brand awareness or the size of the company are likely to influence the amount of firm-related messages that are daily generated. These company specific aspects have deliberately not been taken into account since the purpose of this thesis is to draw generic conclusions on the applicability of social business intelligence.

Next, this thesis analysed the applicability of social business intelligence on two dimensions; industry type and customer relation type. We did not correct for *interaction* effects between these two groups. It is e.g. likely that there exist more B2B firms in the consulting industry. However, the results of this study are exploratory and – from the insights we gained – future research containing larger samples should correct for such interaction effects.

Next, this thesis mainly focused on the *business perspective* of social business intelligence. Less attention has been paid to the technical perspective. For example, the question "What kind of database is best to store the unstructured data that is captured in text form?" is unanswered. Though the developed social business procedure prescribes which components are required to collect and analyse social media data in relation with the firm's performance, the technical requirements related to these components are underexposed.

Finally, the social business intelligence procedure that has been developed in this thesis is compliant with general business intelligence concepts that are adhered to in firms. However, the *procedure* has – due to time restrictions – not been validated, that is, tested on a real case. Nevertheless, the individual components of the procedure are tested. The general BI steps are yet used by firms, and the collection and classification processes of social media messages have been performed in this thesis.

6-6 Future Research

During the execution of this research, ideas for future research related to this thesis have been devised. These ideas are presented in this section. The suggested researches build further on the conclusions of this thesis.

6-6-1 Classifier

We have manually classified social media posts in categories. With these manually classified posts, it is possible to create an automatic classification process. In automatic classifying, a classifier will be "trained" so that it recognises which words and phrases relate to a certain category. The classified messages in this research can serve as the training set for an automatic classifier. As we have experienced, and which is also argued by Gianfortoni, Adamson, and Rosé (2011), classification of social media posts, e.g. by gender, age, political affiliation and sentiment analysis is difficult, and even more problems arise when models trained in one domain are applied in another domain. Therefore, a social media classifier should not be used generally on each domain. We even argue that each firm requires its own classifier, only because the product names of firms differ.

In the development of the social business intelligence procedure we argued that the categorising process is one of the important steps in structurally analysing social media messages. This process is even more challenged by the increase of user-generated social media content showing big data characteristics. From a social business intelligence view, it is desired that research in the field of automatic text classifying – tailored to firms – proceeds.

6-6-2 Social Media Posts Categories

A part of this research required the establishment of social media posts categories. Whereas the starting point of the establishment of the categories in this thesis was based on former research, the social media messages in the dataset have driven the establishment of additional categories. Future research in the domain of social media, and more specifically the classification of social media messages can use the categories that were established in this thesis.

6-6-3 The Real Source

Many messages are forwarded – retweetet and shared – from users to others. Thereby, messages do not stay within one social media platform. It is interesting to investigate which platforms contains the most initial creations of information. As such, firms can manage their reputation by actively following those platforms that create the most initial messages, before the message goes viral and may harm the firm's reputation.

6-6-4 Case Study: Relations of Social Media Metrics and Key-Performance Indicators

The social business intelligence procedure that has been developed in this thesis contains a component in which social media messages are assigned to key-performance indicators based on the subject of the messages. A research in which the relations between the social media messages and the actual values of various KPIs of a firm are investigated would reveal the strength of these relations. Thereby, the KPIs that have been assigned in this thesis as being able to be measured using social media data could be used for such an analysis.

Appendix A

Performance Prism Perspectives and Key-Performance Indicators Categories

Table A-1 on the next page assigns each KPI category defined by Ittner et al. (2003) to a performance prism perspective defined by Neely et al. (2001). Based on their subjects, social media posts will be assigned to KPI categories in this thesis. With the assignment of KPI categories to performance prism perspectives, we can derive conclusions of the existence of social media data related to different performance prism perspectives.

Table A-1: Assigning Key-Performance Indicator Categories to Performance Prism Perspectives

KPI Category	Performance Prism Perspective	Elucidation
Short-term financial results	Stakeholder satisfaction	Shareholders are the actors that are interested in the return on their investment.
Customer relations	Stakeholder satisfaction / contribution	On the one hand, KPIs related to customer relations can represent the customer satisfaction. On the other hand, customers may also contribute to the firm, e.g. by payments and/or co-creation activities.
Employee relations	Stakeholder satisfaction / contribution	Employees are the actors interested in getting awarded for their contributing value to the firm. Therefore, KPIs related to customer relations can involve both perspectives.
Operational performance	Processes	KPIs related to operational performance reflect the performance of business processes, generally in time of volume, speed, reliability, etc.
Product & service quality	Capabilities	KPIs related to product and service quality represent how capable a firm is in performing its activities.
Alliances	Stakeholder satisfaction / contribution	Metrics related to the firm's alliances are on the one hand purposed to satisfy the participating parties and on the other purposed to measure the contributing value of the alliance to the firm.
Supplier relations	Stakeholder satisfaction / contribution	KPIs related to supplier relations are purposed to measure the contributing value of the suppliers' products/services to the firm, or the metrics can specify the satisfaction of the customers (e.g. in terms of the price paid for the product).
Environmental performance	Stakeholder satisfaction	A firm's activities may affect the environment. Groups representing the environment may not be satisfied whenever the firm's activities affect the environment. KPIs reflecting environmental performance hence relate to stakeholder satisfaction.
Product & service innovation	Processes	One of the key business process relates to the development of new products and services.
Community	Stakeholder satisfaction / contribution	Metrics related to the firm's community relate to the public image of the company. As such, these metrics involve stakeholders.

Appendix B

Classification of Social Media Posts

In section 4-3-1, categories for social media posts have been established. Consequently, the collected social media posts of the firms in the sample have been classified into one of these categories. The results are discussed in this chapter. The results will be discussed from firm to firm. However, we will refer to figure B-1 – which is depicted below – when discussion the individual firms. This figure contains a heat-map of the KPI categories per firm, indicating which category is presented the most (green) and which the least (red) in the sample. The percentages that are contained in the heat-map are visualised in a stacked bar chart in figure B-2.

KPI Category		ABN AMRO	Aegon	Akzo Nobel	Albert Heijn	Arcadis	ArcelorMittal	Blokker	Bol.com	C-1000	Coca-Cola	Fugro	Heineken	KLM	NS	Philips	PostNL	TomTom	Unibail-Rodamco	Average
1.	**Short-term financial results**	**1%**	**5%**	**21%**	**0%**	**37%**	**16%**	**0%**	**0%**	**0%**	**0%**	**36%**	**3%**	**1%**	**0%**	**1%**	**2%**	**3%**	**75%**	**11%**
1.1	Financial performance discussions	0%	0%	1%	0%	32%	8%	0%	0%	0%	0%	0%	0%	1%	0%	1%	0%	0%	47%	5%
1.2	Stock related discussions	1%	5%	20%	0%	5%	8%	0%	0%	0%	0%	36%	3%	0%	0%	0%	2%	3%	28%	6%
2.	**Customer relations**	**17%**	**4%**	**0%**	**20%**	**2%**	**0%**	**7%**	**17%**	**1%**	**4%**	**0%**	**2%**	**14%**	**33%**	**0%**	**39%**	**6%**	**0%**	**9%**
2.1	Explaining firm	4%	0%	0%	5%	0%	0%	0%	3%	0%	1%	0%	1%	4%	4%	0%	2%	2%	0%	1%
2.2	Understanding firm	2%	0%	0%	4%	0%	0%	0%	4%	0%	1%	0%	0%	7%	4%	0%	1%	2%	0%	1%
2.3	Thanking firm	1%	0%	0%	1%	1%	0%	0%	1%	0%	1%	0%	0%	1%	1%	0%	0%	1%	0%	0%
2.4	Informing firm	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	11%	0%	0%	0%	0%	1%
2.5	Questioning customer	3%	1%	0%	3%	0%	0%	5%	4%	0%	1%	0%	1%	1%	4%	0%	5%	1%	0%	2%
2.6	Complaining customer	4%	2%	0%	6%	0%	0%	3%	4%	1%	0%	0%	1%	1%	9%	0%	30%	0%	0%	3%
2.7	Thanking customer	2%	1%	0%	2%	0%	0%	0%	1%	0%	0%	0%	0%	1%	1%	0%	1%	0%	0%	0%
3.	**Employee relations**	**1%**	**4%**	**10%**	**5%**	**5%**	**2%**	**8%**	**0%**	**8%**	**1%**	**10%**	**0%**	**0%**	**3%**	**2%**	**21%**	**0%**	**1%**	**4%**
3.1	Recruitment	1%	2%	9%	1%	4%	2%	1%	0%	0%	1%	8%	0%	0%	3%	2%	7%	0%	1%	2%
3.2	Employee posts	1%	2%	1%	4%	1%	0%	7%	0%	8%	0%	1%	0%	0%	0%	0%	14%	0%	0%	2%
4.	**Operational performance**	**1%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**1%**	**0%**	**0%**	**0%**	**0%**	**0%**
5.	**Product and service quality**	**1%**	**0%**	**0%**	**1%**	**0%**	**0%**	**1%**	**0%**	**0%**	**2%**	**0%**	**1%**	**0%**	**1%**	**7%**	**1%**	**7%**	**0%**	**1%**
6.	**Alliances**	**0%**	**2%**	**0%**	**0%**	**0%**	**0%**	**0%**	**1%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**3%**	**0%**	**0%**	**0%**
7.	**Supplier relations**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**1%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**
8.	**Environmental performance**	**0%**	**0%**	**0%**	**2%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**
9	**Product and service innovation**	**2%**	**4%**	**1%**	**1%**	**0%**	**0%**	**0%**	**0%**	**0%**	**1%**	**0%**	**1%**	**0%**	**1%**	**1%**	**0%**	**8%**	**0%**	**1%**
10	**Community**	**71%**	**43%**	**55%**	**28%**	**44%**	**60%**	**27%**	**72%**	**25%**	**57%**	**49%**	**44%**	**7%**	**30%**	**61%**	**23%**	**26%**	**15%**	**41%**
10.1	Promotion	56%	2%	1%	1%	8%	2%	0%	20%	2%	2%	0%	3%	1%	2%	1%	0%	2%	1%	6%
10.2	News	3%	4%	21%	2%	5%	6%	0%	0%	0%	1%	1%	4%	0%	1%	0%	1%	0%	1%	3%
10.3	Public image	10%	36%	11%	26%	18%	32%	26%	50%	23%	50%	3%	33%	4%	27%	6%	20%	22%	3%	22%
10.4	Professionals	3%	1%	3%	0%	12%	20%	0%	2%	0%	2%	45%	3%	2%	0%	0%	2%	0%	11%	6%
10.5	Distributors	0%	0%	18%	0%	0%	0%	1%	0%	0%	1%	0%	2%	0%	0%	54%	0%	2%	0%	4%
	Undefined	**4%**	**16%**	**10%**	**43%**	**11%**	**18%**	**37%**	**4%**	**50%**	**23%**	**4%**	**28%**	**2%**	**2%**	**1%**	**11%**	**49%**	**9%**	**18%**
	Spam	**0%**	**22%**	**2%**	**0%**	**1%**	**3%**	**19%**	**6%**	**15%**	**13%**	**1%**	**18%**	**75%**	**28%**	**27%**	**0%**	**2%**	**0%**	**13%**
Classified Posts		7.067	1.449	922	2.848	455	1.097	1.788	2.574	1.651	1.078	428	2.050	2.498	1.441	1.623	1.013	1.100	512	

Figure B-1: Social Media Posts Classification

ABN AMRO

ABN AMRO Group N.V. is a Dutch bank with 6.8 million clients and around 25.000 employees. The firm organises multiple marketing events each year, of which the pictures were uploaded to the firm's Picasa profile during our sample period. This declares the high percentage of social media posts made on this platform, as can be seen in table C-1. If we would neglect the Picasa posts, which are somehow irregular posts and moreover made by the firm on its own, we would find that 90% of the social media posts have been sourced from Twitter, which is in line with the other firms.

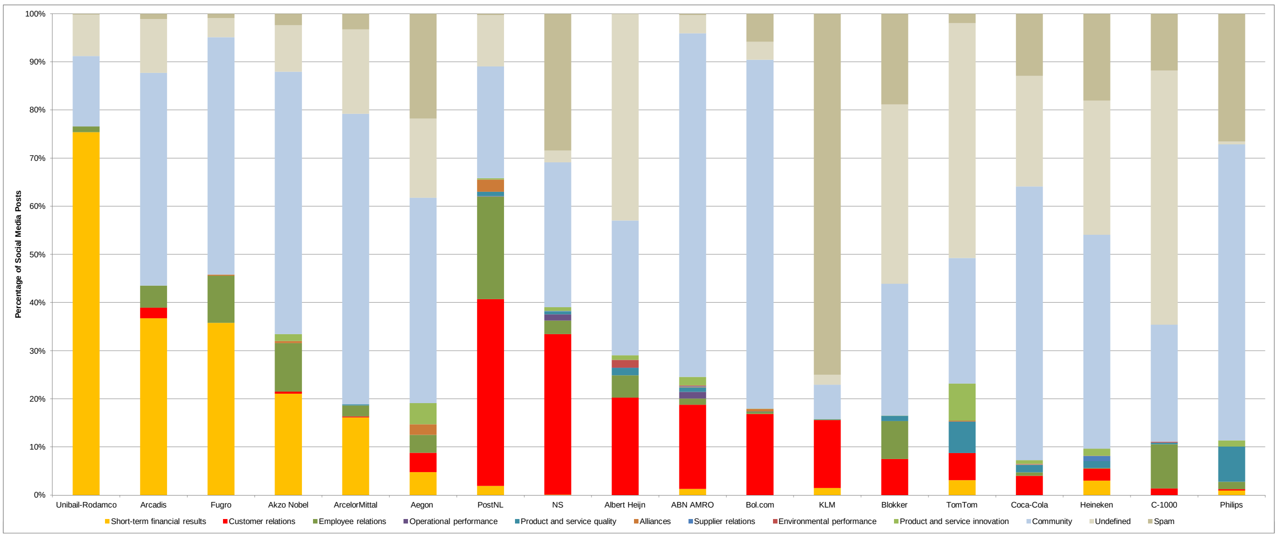


Figure B-2: Social Media Posts Classification

We have classified all 7.067 collected social media posts in the ABN AMRO data set. The results are shown in figure B-1. The columns show the social media post categories as a percentage of the total classified posts of that firm. For ABN AMRO, we can conclude that the majority of the social media posts (71%) are related to the *community* category, which is due to the large share of *promotion* type of posts. As explained, this figure is high because of the Picasa posts made by the firm in the sample period. Secondly, we see that many posts in the ABN AMRO data set are related to *customer relations*. ABN AMRO operates a web care team that actively monitors the social media sites for customers that complain or ask questions. Here, we see that the web care team of ABN AMRO partly replaces the traditional telephone help desk. The table also shows that a substantial part of the posts are not able to be classified into one of the categories. These posts particularly relate to people using ABN AMRO offices as a point of recognition to meet or illustrate where they are. For example, posts contain appointments like "let's meet in front of the ABN AMRO office before we go into town". Whereas these messages definitely contain the name of the firm, they do not contain any relevant information for the firm that can be marked as "social intelligence".

Aegon

Aegon N.V. provides insurances, pensions and asset management products to over 47 million clients in the world. As can be concluded from table C-1, the majority (81%) of the collected posts have Twitter as a source. The channel distribution of Aegon is in line with the distribution of other firms.

Over the sample period, 1.449 posts have been collected containing the word "Aegon". All these posts have been classified in the categories that were established in section 4-3-1. Figure B-1 shows the results of the classification of Aegon's social media posts. As can be concluded, the majority (43%) of the social media posts are related to the KPI category involving *community* related indicators. More specifically, the majority of the *community* related posts are classified as *public image* posts. Public image posts contain the firm's name, indicating that people are talking about the firm, but the posts are not purposed to get in contact with the firm. When analysing the *public image* posts more detailed, we find that many of these posts relate to Aegon's sponsoring activities. The firm is for example sponsor of the Dutch soccer club Ajax, it sponsors a tennis centre called Aegon Arena and it sponsors the Dutch rowing team. The public image social media posts related to sponsoring activities may serve as a measure to determine the exposure of the sponsoring activities. A substantial part of Aegon's social media messages have been classified as *spam*. Spam messages contain the name of the firm but do not relate the firm. The dataset of Aegon revealed that a person named "Aegon The Conqueror" showed up often in posts. Aegon The Conqueror is a character in a popular TV series called Game of Thrones. These kind of *spam* messages, that show up in the dataset because the name of the firm is commonly used for the naming of other entities or people, do not contain any information that may be valuable for the company.

Akzo Nobel

Akzo Nobel N.V. is a Dutch firm active in paint, lacquer, coatings and other specialised chemical products. The company operates in 80 countries and employs around 55.000 people. As can be concluded from table C-1, the 922 posts that have been collected are for the majority sourced from Twitter.

Figure B-1 shows the results of the classification process of the 922 social media messages that have been collected. As can be concluded, around 21% of the social media messages that are related to Akzo Nobel refer to *news* articles. Around 18% of the social media posts are made by *distributors* of Akzo Nobel's products who promote their products on retail websites like Amazon. Another substantial part – 20% – of the messages relates to *stock related discussions*. A remarkable phenomenon in the classification of Akzo Nobel's social media messages is the fact that there are almost no messages related to *customer relations*, whereas we have seen customer related messages in the social media posts of other companies.

Albert Heijn

Albert Heijn is a Dutch supermarket, which is a subsidiary of Royal Ahold N.V. Albert Heijn operates around 850 stores in the Netherlands, and is with 34% market share the market leader in the Netherlands. The company also operates stores in Belgium, Germany and Curacao. Table C-1 illustrates that almost all of the collected social media messages containing Albert Heijn have been sourced from Twitter.

As figure B-1 shows, a substantial part (20%) of the social media messages of Albert Heijn involve *customer relations* management. On the one hand customers are responsible for many complaints (6% of the classified posts), while Albert Heijn's web care team actively responses to these messages by either showing an *understanding* (4% of the classified posts) of the customer's complaint, or even *explaining* (5% of the classified posts) the customer something that they asked. Here, we clearly see that the firm's help desk moves to social media. Another substantial part (26%) of social media messages involving Albert Heijn relate to the firm's *public image*. These messages contain customer's opinions about e.g. the latest Albert Heijn commercial or about the products that they bought in the store.

Arcadis

Arcadis N.V. is a Dutch engineering consultancy, offering solutions in the field of infrastructure, civilised areas and environmental projects. The company is active in more than 70 countries and employing around 18.000 people. During the period of monitoring the companies, we collected 455 posts related to Arcadis. This figure is substantially lower than the number of posts that we collected from other firms. Of the 455 social media posts that have been collected and are related to Arcadis, around 93% has been derived from Twitter.

Shown by figure B-1 on page 101, 32% of the collected messages involve Arcadis' *financial performance discussions*. These posts particularly involve professionals discussing the financial performance of the company and talking about future projections of the company's financial position. Next, 18% of the Arcadis' posts have been classified as *public image* posts. These posts particularly refer to publicised articles or news messages that have been spread by the company. Presumably, the company has spread these articles / news messages to gain exposure. The firm could use the amount of *public image* messages that refer to the articles as a measure to determine the exposure as a result of the publicised articles.

ArcelorMittal

ArcelorMittal is the world's largest steel producer, active in 27 countries and employing 320.000 people. During the period of monitoring, we collected 5.532 social media posts related to ArcelorMittal. Of these posts, around 83% have been derived from Twitter. Other sources involved blogs (5%), news sites (2%) and other platforms.

As can be seen in figure B-1, social media messages related to ArcelorMittal do not involve *customer relations*. Rather, the social media messages relate to the firm's *public image* (32% of the posts) and *professionals* (20% of the posts) talking about the company. The posts classified as *public image* involve marketing activities of the firm. Especially the "ArcelorMittal Orbit", a steel tower constructed by the firm on the 2012 London Olympics, was subject of discussion. This category of social media messages is the only one that involves non-professional people, because all other social media messages related to ArcelorMittal involve professionals. 20% of the sample posts have been classified as *professionals*, in which professionals discuss joint-ventures or the industries' outlook. Finally, a substantial amount of the posts (16%) relate to the companies *financial results*.

Blokker

Blokker is a Dutch store selling products related to household. The Blokker stores are subsidiaries of Blokker Holding B.V., which operates over 2.900 stores in 11 countries, thereby employing 25.000 people. In total, 2.769 social media messages related to Blokker have been collected, of which the majority (91%) has been derived from Twitter. Please see table C-1 on page 108 for a distribution of the sources of the social media posts.

Figure B-1 (page 101) shows that 27% of the classified social media posts are related to the *community* category. The *community* category comprises messages that reveal the community's perception of the firm. Of these 27% *community* related posts, the vast majority consists of social media posts that have been classified as *public image* posts. *Public image* posts are messages that are made by individuals and contain the firm's name, without explicitly seeking contact with the firm. In the case of Blokker, many messages involve statements of people announcing to their followers that they are planning to visit one of the stores, or people referring to articles that are sold by the firm. As can be concluded, customer also ask questions to the firm and also complain about the firm. However, no messages of the company have been found in the sample that respond to these messages.

Bol.com

Bol.com is an online web-shop selling a variety of products, such as books, DVDs, games, blu-rays, electronics, computers, etcetera. Since the foundation of the company in 1999, it has shown solely growth percentages of 18% y-o-y and above in terms of revenues. As of 2012, Bol.com is a subsidiary of Royal Ahold N.V. We have collected 5.782 social media posts that are related to Bol.com, of which 89% has been derived from Twitter.

Of the classified posts related to Bol.com, around 50% are related to the firm's *public image*. For Bol.com, these *public image* posts particularly involve people who share to their followers their recent purchase of a product through Bol.com or people illustrating to other people that a certain product can be bought at Bol.com. Next, a substantial amount of the posts consists of *promotion* activities. These posts are made by the firm or by firm's selling products through Bol.com's website for marketing purposes. Finally, 17% of the Bol.com classified social media posts relate to *customer relations*. We clearly see that people ask questions or complain, and that Bol.com's web care team is consequently responsible for the social media posts that have been classified as either *explaining firm* (3%) or *understanding firm* (4%).

C1000

C1000 is a Dutch supermarket organisation with a market share of 11,5%, employing around 7.000 people and operating 425 stores in the Netherlands. In the future, many C1000 stores will be turned into Jumbo stores as Jumbo Supermarkten acquired C1000 in 2012. Allmost all of the 5.782 social media posts that have been collected in relation with C1000 have been derived from Twitter.

The majority of the social media posts (50%) of C1000 have been classified as *undefined*, implying that these messages cannot be classified into one of the categories that have been established. When taking a closer look at the *undefined* messages, we see that many users refer to C1000 as a location to meet each other, or users share that they are heading for or just returned from C1000. These messages do not contain any management information, and are therefore classified as *undefined*. The other portion of the classified social media messages related to C1000 are classified as *public image* posts. These messages contain statements of customers sharing their followers what they have bought or seen at a C1000 store. Next, marketing campaigns are discussed by people.

Coca-Cola

Coca-Cola is one of the many drinks offered by The Coca-Cola Company, selling Coca-Cola all over the world (except for North-Korea and Cuba). As expected, Coca-Cola is one of the company's in our sample that delivered the most messages because it is one of the most famous brands of the world. During the monitoring period, 32.953 messages containing the word Coca-Cola have been collected. In line with other companies in our sample, 89% of the collected Coca-Cola social media messages have been derived from Twitter.

As figure B-1 illustrates, half of the classified posts of Coca-Cola can be positioned under the category labelled *public image*. These posts contain perceptions of customers to the company, to marketing campaigns of the firm or are made by people talking about the firm's sponsorships. Next, a substantial part of the classified social media posts of coca-cola are classified as *undefined*. These messages to not contain any valuable management information, and contain for instance statements of people that they are drinking Coca-Cola right now, or that they wish that they were drinking one now.

Fugro

Fugro N.V. is a Dutch company that collects and interprets data related to the earth's surface. The company provides advice to firms active in the oil- and gas industry, the mining industry and the construction industry. Fugro is active in over 50 countries, operating 275 offices and employing around 14.000 employees. During the monitoring period, merely 428 social media posts related to Fugro have been collected. Of all firms in our sample, there is no firm with less search results. Though the sample of Fugro is small, it shows a channel distribution that is comparable to the other firms in our sample; around 90% of the derived post are sourced from Twitter.

As indicated by figure B-1 on page 101, the vast majority (45%) of the social media posts related to Fugro are classified as posts made by *professionals*. These posts consists of professionals talking about new vessels that Fugro either ordered or received, or how macro trends are effecting the market in which Fugro operates. Furthermore, automated messages creating tools post a message each time that a Fugro vessel leaves or arrives at a harbour. These posts are also classified as *professionals*. Another substantial part of the social media posts related to Fugro have been assigned to the category labelled *stock related discussions*. Unfortunately the two categories that are responsible for the majority of the social media post categories do not offer any information to the company that is not available at the company internally.

Heineken

Heineken N.V. is a Dutch multinational providing beer and other drinks. The company is active in 178 countries, employing 70.000 people. During our period of monitoring 39.425 social media posts have been collected by the search terms related to Heineken, making Heineken the firm with the most mentions of our sample. As can be concluded from table C-1, 82% of these social media posts have been sourced from Twitter, while Facebook is responsible for 15% of the messages.

Of the 2.050 posts that have been classified, 33% have been classified as *public image* posts (see table B-1). Many of these posts relate to Heineken commercials seen by people on the television, or other ways that Heineken pursues to expose itself such as the Holland Heineken House at the 2012 London Olympics. The number of public image posts can serve as a measure to determine the success of the desired exposure by these kind of marketing events. Another substantial part of Heineken's posts are considered as spam, because they refer to other entities, people naming themselves Heineken on the web or people that are actually named Heineken.

KLM

KLM – Koninklijke Luchtvaart Maatschappij – N.V. is a Dutch airliner that operates 116 airplanes across the globe. The firm has three subsidiaries – KLM Cityhopper, Martinair and Transavia.com – while the parent company is Air France-KLM. The search terms used to filter out the social media posts related to KLM (see table 4-5 on page 58) resulted in 26.364 messages which have been scraped. 86% of these messages have been sourced from Twitter, while Facebook is responsible for 9% of the posts. These figures are in line with the channel distribution of other firms in the sample.

10% of the collected posts – i.e. 2.498 posts – have been classified into one of the social media categories that have been established in section 4-3-1. Figure B-1 shows that an astonishing 75% of these posts have been classified as being *spam*. A closer look at the *spam* classified posts shows that the letters K, L and M are used by many people in their username, e.g. @Klm_babe, @DaOne_KLM, @klm_klm_klm, @klm_nico, and @miyu_klm. Probably, the initials of these people correspond with the name of the firm. However, the dataset of KLM also shows a substantial amount of posts classified as *customer relations*. KLM operates a webcare team that actively monitors the messages directed to KLM, at which the employees of the webcare team consequently respond. Again – as we have seen with ABN AMRO and Albert Heijn – we see that the traditional customer help-desk is (partly) moving to the social media.

NS

Nederlandse Spoorwegen N.V. ("NS") is a Dutch railway company operating the main rail network in the Netherlands. During the monitoring period, 5.863 social media posts related to NS have been collected. 85% of the NS posts have been sourced from Twitter, while Facebook is responsible for 12% of these posts.

Figure B-1 shows that 33% of the classified posts are related to *customer relations*. Especially the category *informing firm* is over-represented in the dataset, this is due to the firm that uses social media to inform customers that certain tracks of the network are subject to delays. These posts do not contain any information that is not available internally, because the nature of the direction of these messages is outgoing; from firm to customers. Next, 9% of the classified posts are *complaining customers*, while 4% of the posts are *questioning customers*. These posts contain information that may not be available to the firm. The firm operates a web care team that answers questions and shows understanding for the experienced problems (4% and 4% of the posts respectively).

Philips

Koninklijke Philips Electronics N.V. is a Dutch electronics firm active in more than 60 countries and employing 122.000 people. The firm is organised into three main divisions: Philips Consumer Lifestyle, Philips Healthcare and Philips Lighting. Philips is the largest manufacturer of lighting in the world. During the monitoring period 32.748 posts have been collected using the Philips search queries of table 4-5, corresponding to almost 3.000 daily posts. Philips is the second largest firm in our sample in terms of collected social media messages. 68% of the messages have been sourced from Twitter, 12% from Facebook, 9% from Blogs and 10% from other platforms including Friendfeed and YouTube.

As indicated by figure B-1, the vast majority (54%) of the social media messages related to Philips are classified as *distributors* posts. The *distributors* posts are made by professionals that are selling Philips products to consumers. Often, distributors use Amazon.com as a site to sell the products, while they use social media to announce the public their offers. Philips is a common surname. As a result, many posts in the Philips sample have been classified as *spam* as they do contain the name Philips, but do not relate to the firm. 7% of the classified posts are labelled as *product and service quality* posts. These posts comprise product- reviews and experiences of users, containing valuable information for R&D related activities.

PostNL

PostNL N.V. is a mail and parcel company operating in the Netherlands, Germany, Italy and the United Kingdom. In total, 1.323 social media messages have been collected using the search terms related to PostNL. 91% of these posts have been collected from Twitter, which is in line with other firms in the sample.

More than in any other dataset of the firms, 30% of PostNL's social media messages have been classified as *complaining customers*. The messages contain statements of customers who are complaining about the service of the firm, about broken parcels, late deliveries, etc. The firm operates a web care team, though it does only respond to a limited amount of complaining and questioning customers. 20% of the PostNL posts have been classified as *public image* posts, posts made by individuals talking about the firm. A surprising figure in PostNL's social media classification overview is the high percentage – 14% – of *employee posts*. Apparently, PostNL's employees – and especially postmen – share that they are working at the firm.

TomTom

TomTom N.V. is a Dutch producer of automotive navigation systems. TomTom is Europe's leading manufacturer of navigation systems. The firm employs around 3.500 people. During the monitoring period, 32.748 social media messages have been collected using the search queries related to TomTom. In line with other firms, 91% of these messages have been derived from Twitter.

Of all classified TomTom messages, 49% has been labelled *undefined*, implying that these messages could not be assigned to any of the other categories. These messages relate to the firm, although people use TomTom in their message, though these messages do not contain any valuable information for the firm. The high percentage of *undefined* posts is due to the fact that people use the word TomTom as a term for navigation systems in generally, or to refer to anyone who is navigating. Apparently, TomTom has become a word in the general vocabulary used by the society, though the relation to the firm TomTom is not always present. 7% of the posts are related to *product and service quality*, in which users share experiences of the usage of TomTom's products. Another 8% of the posts are classified as *product and service innovation* posts, in which users either make innovative suggestions for future products, or share their opinion towards new products / services. These two categories contain valuable information for R&D departments, on the one hand to measure the success of existing products and on the other hand to develop new products.

Unibail-Rodamco

Unibail-Rodamco is a firm specialised in commercial property investments. It is the largest commercial real estate company in Europe, managing three types of assets; shopping centers, convention centers and office properties. Unibail-Rodamco employs around 1.500 people. Only 512 social media messages related to Unibail-Rodamco have been collected, i.e. 39 daily posts on average. 95% of these posts have been sourced from Twitter.

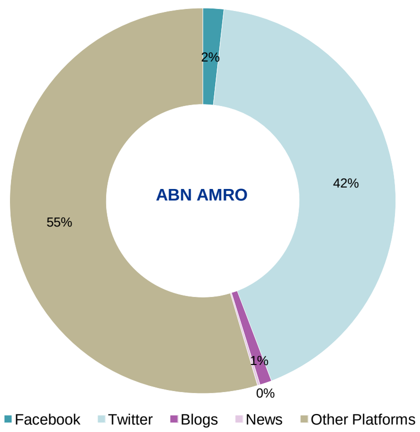
More than in any other firm in our sample, as illustrated by figure B-1 75% of the posts related to Unibail-Rodamco relate to *financial results*. These posts are either related to *financial performance discussions* of the firm or to *stock related discussions*. Furthermore, 11% of the classified posts are messages classified as *professionals*; people writing about the firm from a professional point of view.

Appendix C

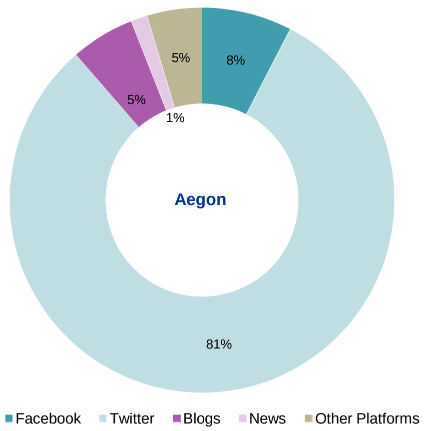
Social Media Platform Distribution

Table C-1: Social Media Channel Distribution

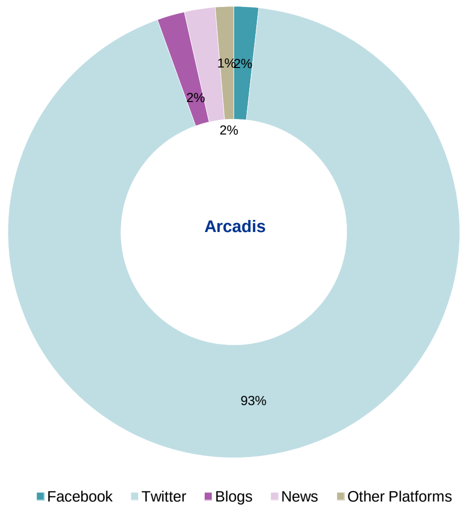
Platform	**Facebook**		**Twitter**		**Blogs**		**News**		**Other**		**Total**
	Abs	%	Abs	%	Abs	%	Abs	%	Abs	%	Abs
ABN AMRO	124	2%	3.000	42%	70	1%	15	0%	3.858	55%	7.067
Aegon	110	8%	1.173	81%	79	5%	20	1%	67	5%	1.449
Akzo Nobel	30	3%	806	87%	43	5%	25	3%	18	2%	922
Albert Heijn	328	3%	11.116	96%	77	1%	1	0%	59	1%	11.581
Arcadis	8	2%	422	93%	9	2%	10	2%	6	1%	455
ArcelorMittal	439	8%	4.569	83%	296	5%	89	2%	139	3%	5.532
Blokker	155	6%	2.526	91%	71	3%	3	0%	14	1%	2.769
Bol.com	472	8%	5.124	89%	115	2%	-	0%	71	1%	5.782
C1000	362	3%	10.583	96%	81	1%	5	0%	33	0%	11.064
Coca-Cola	1.653	5%	29.347	89%	999	3%	69	0%	885	3%	32.953
Fugro	6	1%	385	90%	20	5%	15	4%	2	0%	428
Heineken	5.726	15%	32.332	82%	494	1%	122	0%	751	2%	39.425
KLM	2.316	9%	22.601	86%	617	2%	90	0%	740	3%	26.364
NS	703	12%	4.970	85%	103	2%	-	0%	87	1%	5.863
Philips	4.641	12%	26.260	68%	3.404	9%	138	0%	4.007	10%	38.450
PostNL	77	6%	1.207	91%	27	2%	-	0%	12	1%	1.323
TomTom	1.308	4%	29.787	91%	630	2%	67	0%	956	3%	32.748
Unibail-Rodamco	3	1%	487	95%	9	2%	12	2%	1	0%	512
Total	18.461	8%	186.695	83%	7.144	3%	681	0%	11.706	5%	224.687



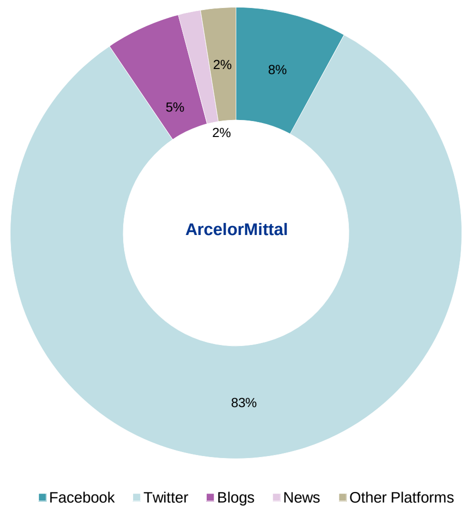
n= 7.067



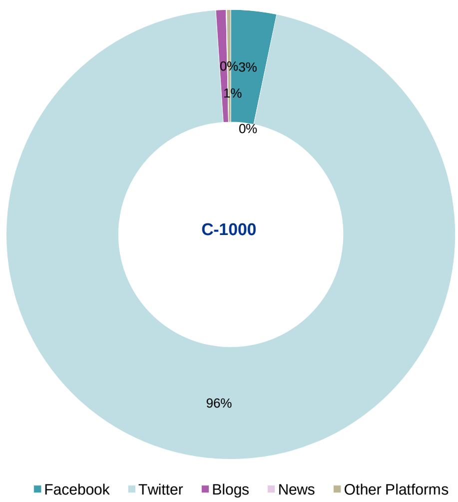
n= 1.449



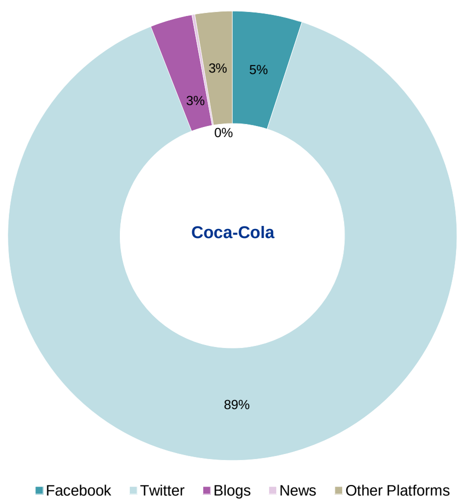
n= 455



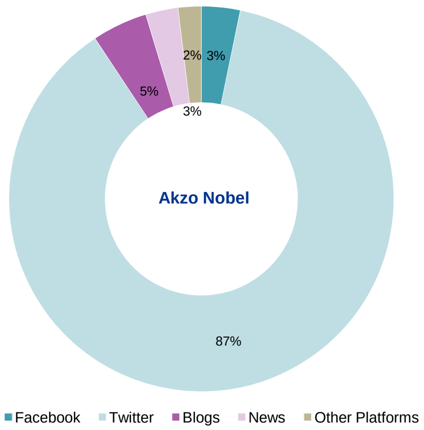
n= 5.532



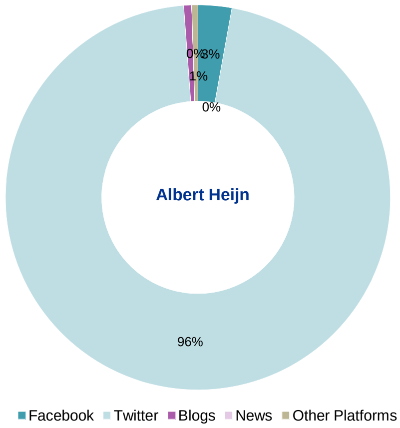
n= 11.064



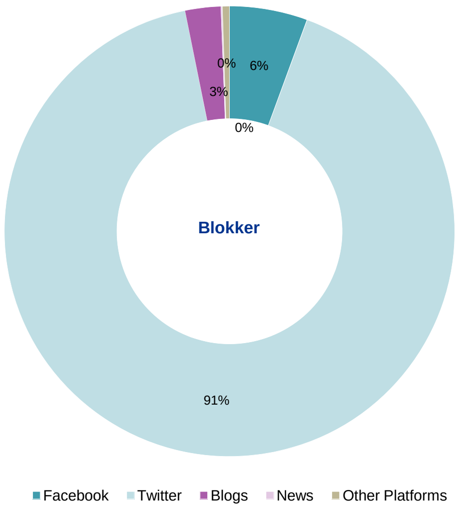
n= 32.953



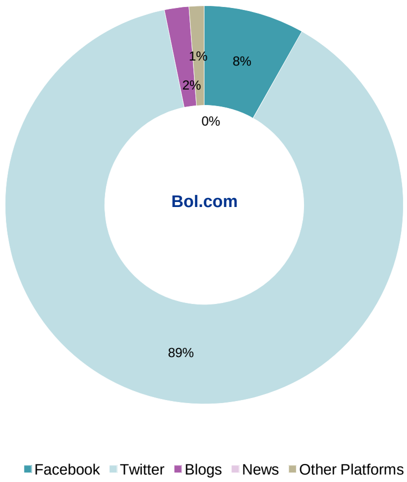
n= 922



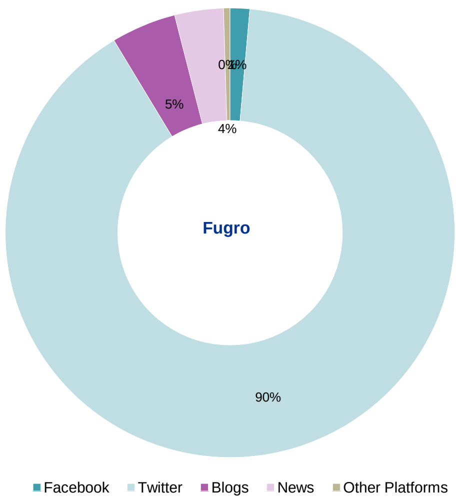
n= 11.581



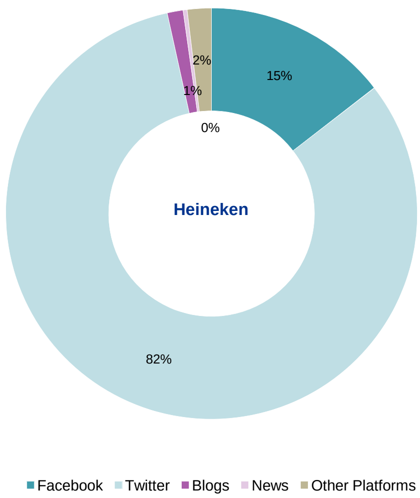
n= 2.769



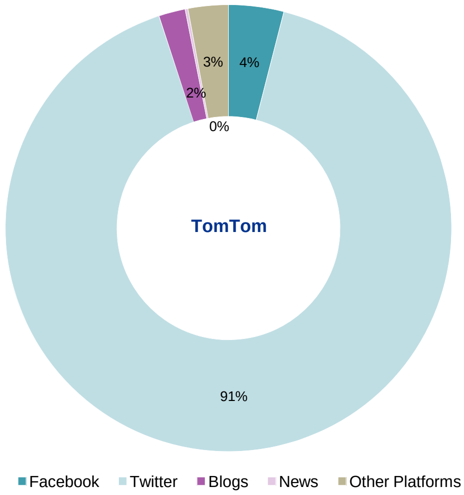
n= 5.782



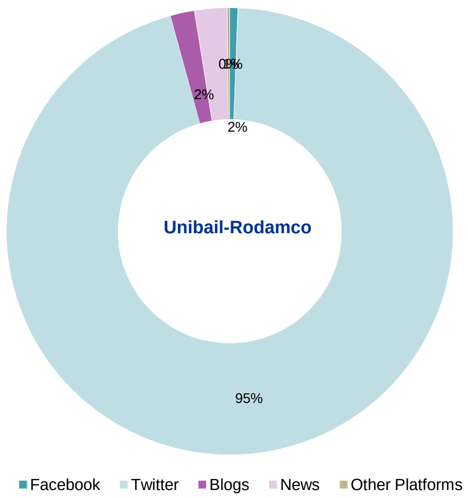
n= 428



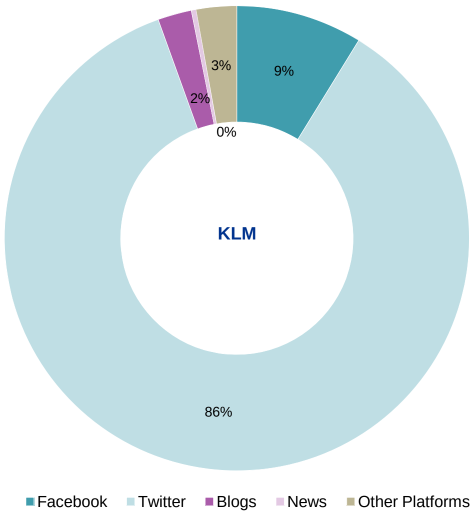
n= 39.425



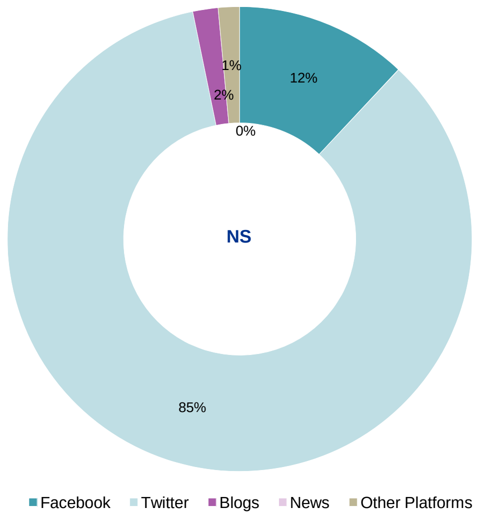
n= 32.748



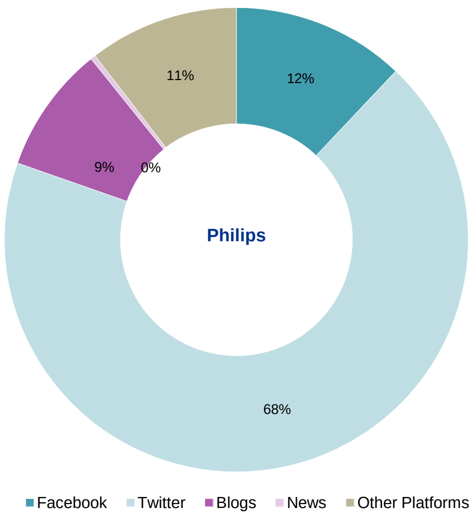
n= 512



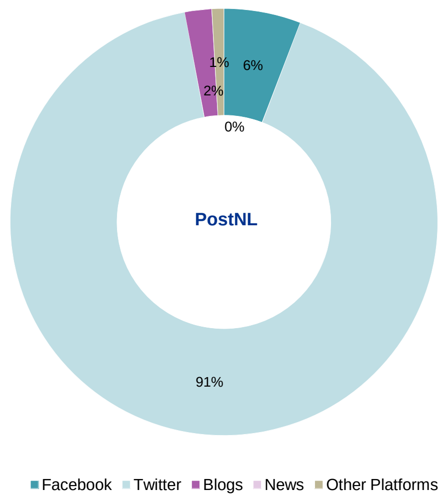
n= 26.364



n= 5.863



n= 38.450



n= 1.323

Appendix D

Descriptive Statistics of Social Media Post Categories

Social Media Categories across Different Customer Relation Types

Table D-1: Social Media Post Categories: Across Customer Relation Type (Descriptives)

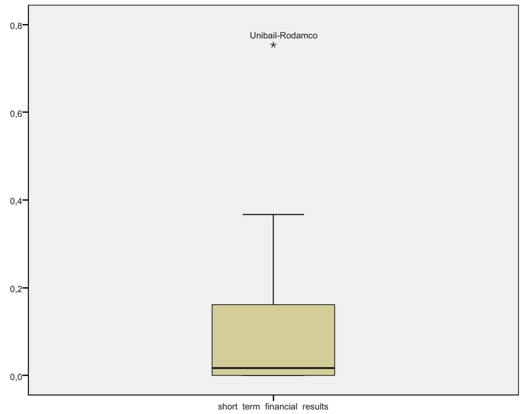
Customer Relation Type	B2C			B2B			Total		
	μ	N	σ	μ	N	σ	μ	N	σ
short_term_financial_results	1,27%	13	1,54%	37,00%	5	23,27%	11,20%	18	20,01%
financial_performance_discussions	0,20%	13	0,44%	17,70%	5	21,04%	5,06%	18	13,01%
stock_related_discussions	1,07%	13	1,56%	19,31%	5	13,10%	6,13%	18	10,62%
customer_relations	12,76%	13	12,34%	0,56%	5	0,93%	9,37%	18	11,80%
explaining_firm	1,92%	13	1,87%	0,07%	5	0,10%	1,41%	18	1,79%
understanding_firm	1,95%	13	2,19%	0,07%	5	0,10%	1,43%	18	2,04%
thanking_firm	0,43%	13	0,41%	0,26%	5	0,59%	0,39%	18	0,45%
informing_firm	0,83%	13	2,98%	0,00%	5	0,00%	0,60%	18	2,54%
questioning_firm	2,15%	13	1,79%	0,08%	5	0,09%	1,57%	18	1,78%
complaining_customer	4,73%	13	8,16%	0,02%	5	0,04%	3,42%	18	7,19%
thanking_customer	0,60%	13	0,64%	0,07%	5	0,10%	0,45%	18	0,59%
employee_relations	4,08%	13	5,91%	5,59%	5	4,17%	4,50%	18	5,41%
recruitment	1,34%	13	1,91%	4,85%	5	3,78%	2,31%	18	2,93%
employee_posts	2,74%	13	4,43%	0,66%	5	0,51%	2,16%	18	3,86%
operational_performance	0,22%	13	0,50%	0,00%	5	0,00%	0,16%	18	0,43%
product_and_service_quality	1,77%	13	2,51%	0,05%	5	0,12%	1,29%	18	2,26%
alliances	0,44%	13	0,88%	0,13%	5	0,20%	0,35%	18	0,76%
supplier_relations	0,11%	13	0,28%	0,00%	5	0,00%	0,08%	18	0,24%
environmental_performance	0,15%	13	0,43%	0,00%	5	0,00%	0,11%	18	0,37%
product_and_service_innovation	1,52%	13	2,21%	0,28%	5	0,63%	1,17%	18	1,97%
community	39,81%	13	20,50%	44,61%	5	17,79%	41,14%	18	19,39%
promotion	7,08%	13	15,50%	2,46%	5	3,15%	5,79%	18	13,28%
news	1,22%	13	1,46%	6,82%	5	8,21%	2,78%	18	4,90%
public_image	25,50%	13	14,61%	13,61%	5	12,12%	22,20%	18	14,67%
professionals	1,27%	13	1,21%	18,14%	5	15,98%	5,96%	18	11,02%
distributors	4,73%	13	15,36%	3,58%	5	8,00%	4,41%	18	13,48%
undefined	20,74%	13	18,88%	10,19%	5	4,90%	17,81%	18	16,77%
spam	17,13%	13	20,00%	1,58%	5	1,24%	12,81%	18	18,28%

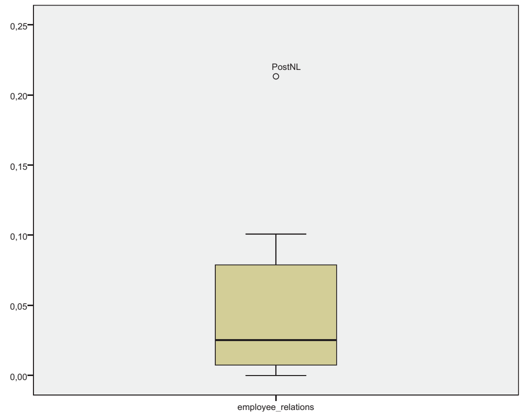
Social Media Categories across Different Industries

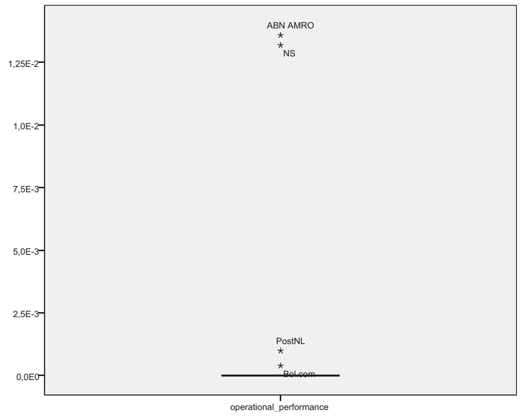
Industry	Mining and Quarrying			Industry			Wholesale and Retail			Transport and Storage			Information and Communication			Financial Institutions			Consultancy, Research and Other Specialised Business Services			Total		
Social Media Post Category	**μ**	**N**	**σ**	**μ**	**N**	**σ**	**μ**	**N**	**σ**	**μ**	**N**	**σ**	**μ**	**N**	**σ**	**μ**	**N**	**σ**	**μ**	**N**	**σ**	**μ**	**N**	**σ**
short_term_financial_results	18,59%	2	3,47%	1,33%	3	1,51%	0,00%	3	0,00%	1,14%	3	0,95%	1,54%	2	2,18%	27,15%	3	41,82%	36,23%	2	0,68%	11,20%	18	20,01%
financial_performance_discussions	4,58%	2	5,25%	0,25%	3	0,43%	0,00%	3	0,00%	0,52%	3	0,84%	0,18%	2	0,26%	15,82%	3	27,40%	15,93%	2	22,53%	5,06%	18	13,01%
stock_related_discussions	14,01%	2	8,72%	1,08%	3	1,65%	0,00%	3	0,00%	0,63%	3	1,08%	1,36%	2	1,93%	11,33%	3	14,48%	20,29%	2	21,86%	6,13%	18	10,62%
customer_relations	0,31%	2	0,18%	2,27%	3	1,83%	9,60%	3	9,74%	28,76%	3	12,99%	11,25%	2	7,94%	7,16%	3	9,16%	1,10%	2	1,55%	9,37%	18	11,80%
explaining_firm	0,05%	2	0,08%	0,41%	3	0,36%	1,56%	3	2,70%	3,42%	3	1,44%	2,12%	2	0,68%	1,54%	3	2,42%	0,11%	2	0,16%	1,41%	18	1,79%
understanding_firm	0,05%	2	0,08%	0,29%	3	0,47%	1,32%	3	2,29%	3,90%	3	3,17%	2,97%	2	1,24%	0,96%	3	1,27%	0,11%	2	0,16%	1,43%	18	2,04%
thanking_firm	0,00%	2	0,00%	0,31%	3	0,54%	0,29%	3	0,51%	0,59%	3	0,53%	0,56%	2	0,03%	0,31%	3	0,37%	0,66%	2	0,93%	0,39%	18	0,45%
informing_firm	0,00%	2	0,00%	0,00%	3	0,00%	0,00%	3	0,00%	3,59%	3	6,21%	0,00%	2	0,00%	0,00%	3	0,00%	0,00%	2	0,00%	0,60%	18	2,54%
questioning_firm	0,10%	2	0,01%	0,74%	3	0,58%	2,73%	3	2,14%	3,09%	3	2,30%	2,55%	2	2,32%	1,05%	3	1,36%	0,11%	2	0,16%	1,57%	18	1,78%
complaining_customer	0,05%	2	0,06%	0,39%	3	0,21%	3,12%	3	2,54%	13,55%	3	15,17%	2,31%	2	2,62%	1,89%	3	1,90%	0,00%	2	0,00%	3,42%	18	7,19%
thanking_customer	0,05%	2	0,08%	0,13%	3	0,14%	0,58%	3	0,90%	0,63%	3	0,12%	0,74%	2	1,04%	0,77%	3	0,90%	0,11%	2	0,16%	0,45%	18	0,59%
employee_relations	6,18%	2	5,52%	0,86%	3	0,78%	6,98%	3	2,05%	8,06%	3	11,57%	0,17%	2	0,25%	2,07%	3	1,44%	7,21%	2	3,68%	4,50%	18	5,41%
recruitment	5,58%	2	5,31%	0,81%	3	0,85%	0,74%	3	0,49%	3,27%	3	3,56%	0,17%	2	0,25%	1,23%	3	0,60%	5,96%	2	3,46%	2,31%	18	2,93%
employee_posts	0,61%	2	0,21%	0,05%	3	0,08%	6,23%	3	2,37%	4,78%	3	8,17%	0,00%	2	0,00%	0,80%	3	0,96%	1,03%	2	0,53%	2,16%	18	3,86%
operational_performance	0,00%	2	0,00%	0,00%	3	0,00%	0,00%	3	0,00%	0,47%	3	0,73%	0,02%	2	0,03%	0,45%	3	0,78%	0,00%	2	0,00%	0,16%	18	0,43%
product_and_service_quality	0,14%	2	0,19%	3,67%	3	3,73%	0,95%	3	0,59%	0,53%	3	0,41%	3,31%	2	4,57%	0,31%	3	0,53%	0,00%	2	0,00%	1,29%	18	2,26%
alliances	0,22%	2	0,31%	0,03%	3	0,05%	0,00%	3	0,00%	0,86%	3	1,48%	0,38%	2	0,28%	0,76%	3	1,25%	0,12%	2	0,17%	0,35%	18	0,76%
supplier_relations	0,00%	2	0,00%	0,34%	3	0,59%	0,05%	3	0,08%	0,02%	3	0,04%	0,00%	2	0,00%	0,06%	3	0,10%	0,00%	2	0,00%	0,08%	18	0,24%
environmental_performance	0,00%	2	0,00%	0,05%	3	0,05%	0,55%	3	0,90%	0,00%	3	0,00%	0,00%	2	0,00%	0,05%	3	0,09%	0,00%	2	0,00%	0,11%	18	0,37%
product_and_service_innovation	0,71%	2	1,00%	1,21%	3	0,40%	0,35%	3	0,55%	0,37%	3	0,37%	3,88%	2	5,43%	2,07%	3	2,22%	0,00%	2	0,00%	1,17%	18	1,97%
community	57,45%	2	4,09%	54,64%	3	9,29%	26,86%	3	1,47%	20,15%	3	11,80%	49,28%	2	32,83%	42,90%	3	28,38%	46,74%	2	3,62%	41,14%	18	19,39%
promotion	1,79%	2	0,69%	1,82%	3	0,96%	0,87%	3	0,84%	1,33%	3	0,91%	11,39%	2	12,63%	19,26%	3	31,48%	4,07%	2	5,43%	5,79%	18	13,28%
news	13,42%	2	10,47%	1,70%	3	2,06%	0,65%	3	1,02%	0,75%	3	0,39%	0,00%	2	0,00%	2,40%	3	1,65%	3,33%	2	3,06%	2,78%	18	4,90%
public_image	21,69%	2	14,57%	29,65%	3	22,56%	24,79%	3	1,48%	16,82%	3	11,77%	35,87%	2	20,15%	16,32%	3	17,58%	10,87%	2	10,74%	22,20%	18	14,67%
professionals	11,60%	2	11,96%	2,07%	3	1,45%	0,11%	3	0,11%	1,25%	3	0,85%	1,03%	2	1,46%	4,92%	3	4,99%	28,47%	2	22,85%	5,96%	18	11,02%
distributors	8,95%	2	12,65%	19,39%	3	31,52%	0,45%	3	0,77%	0,00%	3	0,00%	1,00%	2	1,41%	0,00%	3	0,00%	0,00%	2	0,00%	4,41%	18	13,48%
undefined	13,58%	2	5,55%	17,09%	3	14,65%	43,52%	3	6,59%	5,06%	3	4,86%	26,25%	2	31,85%	9,59%	3	6,39%	7,59%	2	5,12%	17,81%	18	16,77%
spam	2,83%	2	0,63%	18,51%	3	5,84%	11,15%	3	9,84%	34,59%	3	37,74%	3,91%	2	2,71%	7,43%	3	12,45%	1,02%	2	0,12%	12,81%	18	18,28%

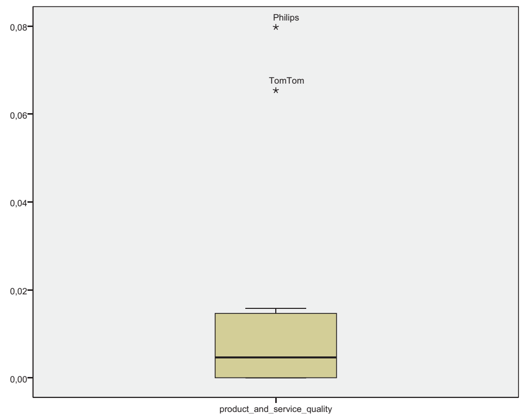
Figure D-1: Social Media Post Categories: Across Industries (Descriptives)

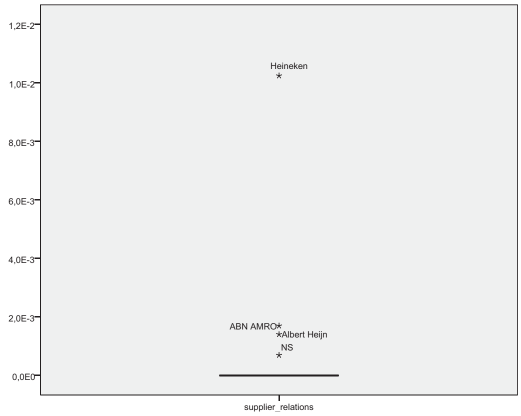
Boxplots of Social Media Categories across Firms

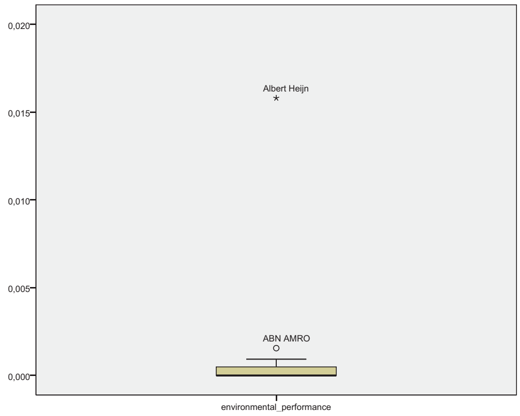


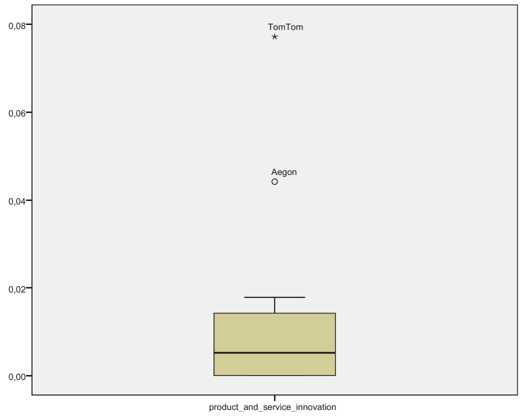


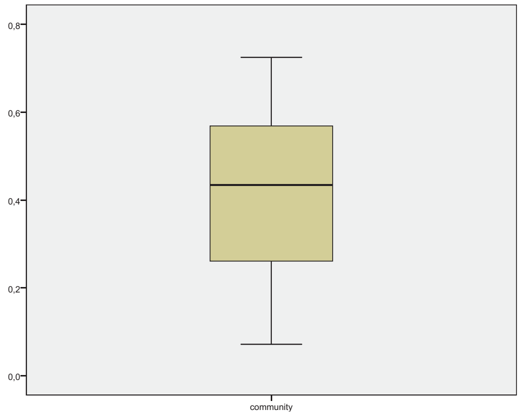


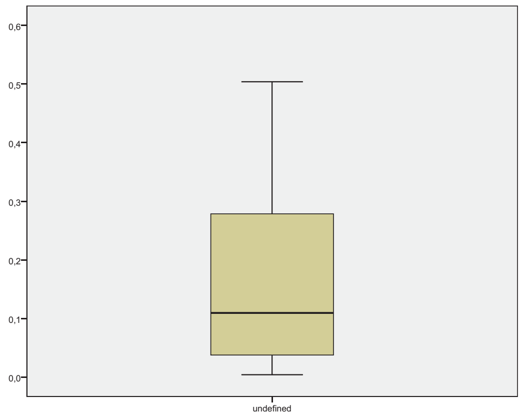


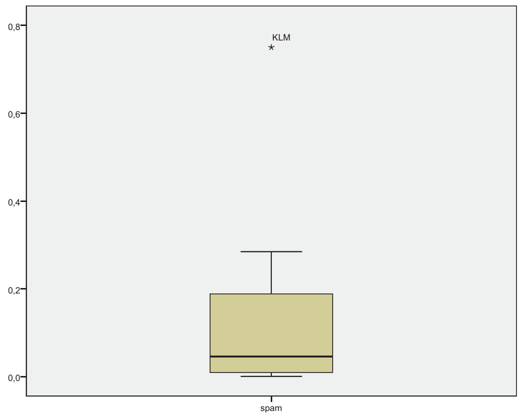

1,2E-2
1,0E-2
8,0E-3
6,0E-3
4,0E-3
2,0E-3
0,0E0
Heineken
ABN AMRO
Albert Heijn
NS
supplier_relations


0,020
0,015
0,010
0,005
0,000
Albert Heijn
ABN AMRO
environmental_performance


0,08
0,06
0,04
0,02
0,00
TomTom
Aegon
product_and_service_innovation


0,8
0,6
0,4
0,2
0,0
community


0,6
0,5
0,4
0,3
0,2
0,1
0,0
undefined


0,8
0,6
0,4
0,2
0,0
KLM
spam

Appendix E

Corporate Engagement

The next page lists the user names of firms that have been found in our dataset. These user names have been used to assign social media messages in the category 'firm-to-customer'.

ID	ABN AMRO	# Posts	Aegon	# Posts	Akzo Nobel	# Posts	Albert Heijn	# Posts	Arcadis	# Posts	ArcelorMittal	# Posts	Blokker	# Posts	Bol.com	# Posts	C-1000	# Posts
1	abnamro_tegelen	2	Aegon Religare Life Insurance	1	AkzoNobel_A&AC	2	Albert Heijn Nuland	3	ARCADIS GS	1	ArcelorMittal USA	1	Blokker	2	bol.com baby	49	C1000 Koetsier	1
2	ABN AMRO Commercial	1	AEGON Seguros	1	AkzoNobel Graduates	2	Albert Heijn Stiens	3	ARCADIS LFN	1	arcelormittal_d	0	Blokker	2	Bol.com FilmTV	47	C1000 Noordman	1
3	ABN AMRO Echt	2	AEGON UK	1	akzonobelaero	1	Albert Heijn Strijp	1	ARCADIS Shelter	4	ArcelorMittalDofasco	0	blokkerholten	1	bol.com games	9	C1000 Puts	1
4	abnamroecht	2	aegon_nl	3	AkzoNobelAero	0	albertheijn	1	arcadis_lfn	0	ArcelorMittal	2			Bol.com Muziek	44	C1000_Franeker	1
5	ABNAMROMEPPEL	547	AEGONNederland	9	Grads@AkzoNobel	1	AlbertHeijnXLTilburg	265	ARCADISBelgique	8	ArcelorMittal Orbit	1			Bol.com Partner	111	c1000beekman	1
6	abnamronijmegen	3.851	AEGONReligare	15	AkzoNobel	1	Albert Bezorgservice	1	ARCADIS	27					Bol.com service	275	c1000dalfsen	2
7	ABN AMRO	1	AEGON	1	AkzoNobel Aerospace	1	Albert Heijn	1	ARCADIS Corp Comm	1					Bol.com acties	26	C1000 Beekman	1
8	ABN AMRO NV	1	AEGON Nederland	44			Albert Heijn 1618	1							bol.com	54	C1000 Kees Elzinga	6
9	ABN AMRO Markets	7	aegonseguros	1			AH Rapportstraat	1							bol_com	12	c1000days	2
10	abnamroturbo	1	AEGON Classic	4											bol_com_acties	6	c1000mulder	1
11	ABNAMROYERSEKEOPEN	1													bol_com_api	1	c1000nijenering	1
12	ABN AMRO EconBureau	21													bol_com_baby	2	c1000offerein	2
13															bol_com_banen	2	c1000panningen	1
14															bol_com_games	2	c1000puts	2
15															bol_com_koken	48	c1000wahle	1
16															bol_com_muziek	1	C1000Wiegmans	6
17															bol_com_partner	2		
18															werkenbijbol.com	8		
Own posts		**4.437**		**80**		**8**		**277**		**42**		**4**		**5**		**699**		**30**
% Total sample		63%		6%		1%		2%		9%		0%		0%		9%		0%

ID	Coca-Cola	#Posts	Fugro	#Posts	Heineken	#Posts	KLM	#Posts	NS	#Posts	Philips	#Posts	PostNL	#Posts	TomTom	#Posts	Unibail Rodamco	#Posts
1	Coca-Cola GB	1	Fugro Seacore	1	Heineken Music Hall	19	KLM Airlines	2	NS online	6	Philips Arena	2	OR PostNL	13	tomtombenelux	3	Unibail Rodamco	3
2	Coca-Cola 24x7	1			Heineken NL	4	KLM Belgium	1	NS Station Zwolle	2	Philips Argentina	1	PostNL	1	TomTom	70		
3	Coca-Cola Argentina	3			Heineken tweets	8	KLM Canada	3	NS Stations	61	Philips AVENT UK	32	PostNL Pride	1	TomTom Brasil	4		
4	Coca-Cola Boi Da Don	4			heineken_house	0	KLM Coupon Codes	1	NS Unofficial	77	Philips Brasil	1	PostNL Vandaag	2	TomTom Deutschland	2		
5	Coca-Cola Brasil	3			heineken_tweets	0	KLM Deutschland	9	NS Vertraging	3	Philips Corp Comms	1	PostNL Webcare	27	Tomtom English	1		
6	Coca-Cola Canada	47			Heineken Eetcafe	2	KLM ECUADOR	1744	NSvertraging	1	Philips Ent Lighting	8	PostNLRecruitment	2	TomTom France	7		
7	CocaColaBelgium	2			Heineken House	6	Air France-KLM	2	NS	132	PhilipsLighting	3			TomTom Italia	3		
8	Coca-Cola	12					Royal Dutch Airlines	1	NS Hispeed	22	Philips	1			TomTom Official	15		
9	Coca-Cola Clothing	2					KLM fares	26	nsvertragingnl	5	Philips Fast	4			TomTom Portugal	3		
10	Coca-Cola Colombia	4					KLM Ireland	2	NS_Wijchen	2	Philips Grooming	2			TomTom South Africa	6		
11	Coca-Cola CR	1					KLM ITALIA	5	ns_zwolle	1	Philips Healthcare	27			TomTom Spain	7		
12	Coca-Cola Ecuador	1					KLM Kazakhstan	6			Philips Home Lighting	3			TomTom Switzerland	4		
13	Coca-Cola ElSalvador	1					KLM Mexico	9			Philips In Lighting	3						
14	Coca-Cola España	2					KLM RUSSIA	1			Philips India	2						
15	CocaCola FM COLOMBIA	3					KLM Spain	34			Philips Indonesia	101						
16	Coca-Cola FM ECUADOR	2					KLM UK	16			Philips Japan	9						
17	Coca-Cola Freestyle	2					KLM zakelijk	1			Philips Jobs NA	9						
18	Coca-Cola GB	2					klm_es	1			Philips Jobs NA	9						
19	Coca-Cola Guatemala	1					KLM_JP	10			Philips Lighting	12						
20	Coca-Cola HBC Italia	1					klm_pt	1			Philips Spain	4						
21	Coca-Cola Honduras	1					klm_uk	1			Philips Support UK	1						
22	Coca-Cola Indonesia	3					klm_us	1			Philips Thailand	10						
23	Coca-Cola Korea	105					klmfares	3			Philips Turkiye	26						
24	Coca-Cola Nicaragua	1					klmitalia	1			Philips UK Graduates	46						
25	Coca-Cola Panamá	1					klmmalaysia	1			Philips_Nederland	6						
26	Coca-Cola Venezuela	11					klmnorway	1			PhilipsCare NL	4						
27	Coca-Cola Web	3					KLMRoyalDutchAirline	1			philipsjapan	2						
28	CocaColaBuffalo	1									philipsjobsna	3						
29	CocaColaIsrael	39									philipslight	2						
30	CocaColaTv	1									PhilipsLighting	1						
31											philipsnljobs	1						
32											PhilipsSoundProject	34						
33											philipsspain	2						
34											PhilipsTurkiye	0						
35											philipsukgrads	18						
Own posts		**261**		**0**		**39**		**1884**		**312**		**390**		**46**		**125**		3
% Total sample		1%		0%		0%		7%		5%		1%		3%		0%		1%

Figure E-1: Social Media Posts Classification

Bibliography

Agichtein, E., Castillo, C., Donato, D., Gionis, A., & Mishne, G. (2008). Finding high-quality content in social media. In *Proceedings of the international conference on web search and web data mining* (pp. 183–194). Association for Computing Machinery.

Araujo, T., & Neijens, P. (2012). Friend me: Which factors influence top global brands participation in social network sites. *Internet Research*, *22*, 626 – 640.

Auditore, P. J. (2012a). *The market for social media and business intelligence tools.* Asteria Research. Retrieved July 31, 2012, from http://sandhill.com/article/the-market-for-social-media-and-business-intelligence-tools/

Auditore, P. J. (2012b). *Share-unisphere research study: Adoption of social media bi on the rise.* SHARE-Unisphere. Retrieved July 31, 2012, from http://www.dbta.com/Articles/Editorial/Trends-and-Applications/SHARE-Unisphere-Research-Study-Adoption-of-Social-Media-BI-on-the-Rise-83119.aspx

Azvine, B., Cui, Z., & Nauck, D. D. (2005). Towards real-time business intelligence. *BT Technology Journal*, *23*, 214–225.

Berthon, P. R., Pitt, L. F., McCarthy, I., & Kates, S. M. (2007). When customers get clever: Managerial approaches to dealing with creative consumers. *Business Horizons*, *50*(1), 39–47.

Bonneau, J., & Preibusch, S. (2010). The privacy jungle: On the market for data protection in social networks. In T. Moore, D. Pym, & C. Ioannidis (Eds.), *Economics of information security and privacy* (pp. 121–167). Springer US.

Bos, W., & Tarnai, C. (1999). Content analysis in empirical social research. *International Journal of Educational Research*, *31*(8), 659–671.

Bouwman, H., Faber, E., Haaker, T., Kijl, B., & de Reuver, M. (2008). Conceptualizing the stof model. In H. Bouwman, H. de Vos, & T. Haaker (Eds.), *Mobile service innovation and business models* (pp. 31 – 69). Berlin, Heidelberg: Springer.

Boyd, D. M., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, *13*(1), 210–230.

Bryant, L. (2011). *Moving from social media monitoring to social business intelligence.* Headshift, Dachis Group. Retrieved July 25, 2012, from http://www.headshift.com/our-blog/2011/06/15/from-social-media-monitoring-to-social-business-intelligence/

Bughin, J., Chui, M., & Manyika, J. (2010). *Clouds, big data, and smart assets: Ten tech-enabled business trends to watch.* McKinsey & Co. Quarterly. (White Paper).

CBS. (2012). Standaard bedrijfsindeling 2008. versie 2012. *Centraal Bureau voor de Statistiek*.

Chaffey, D. (2009). *E-business and e-commerce management (4th edition).* Upper Saddle River, NJ, USA: Prentice-Hall, Inc.

Chen, D.-N., Jeng, B., Lee, W.-P., & Chuang, C.-H. (2008). An agent-based model for consumer-to-business electronic commerce. *Expert Systems with Applications*, *34*(1), 469–481.

DesAutels, P. (2011). Ugis: Understanding the nature of user-generated information systems. *Business Horizons*, *54*(3), 185 - 192.

Dey, L., & Haque, S. K. M. (2008). Opinion mining from noisy text data. In *Proceedings of the second workshop on analytics for noisy unstructured text data* (pp. 83–90). New York, NY, USA: ACM.

Dhar, V., & Chang, E. A. (2009). Does chatter matter? the impact of user-generated content on music sales. *Journal of Interactive Marketing*, *23*(4), 300–307.

Doeland, D. (2012). *Social media & internet key performance indicators.* Retrieved August 10, 2012, from http://ddmca.nl/diensten/social-media-and-internet-monitoring/social-media-key-performance-indicators/

Dong-Hun, L. (2010). Growing popularity of social media and business strategy. *Korean Consumer & Society.*, 112–117.

Dugan, L. (2012). *Twitter to surpass 500 million registered users on wednesday.* Retrieved August 15, 2012, from http://www.mediabistro.com/alltwitter/500-million-registered-users_b18842

Eccles, R. (1991). The performance measurement manifesto. *Harvard Business School*, *69*, 131–137.

Eckerson, W. W. (2009). *Performance management strategies. how to create and deploy effective metrics.* The Data Warehousing Institute. (White Paper).

Elbashir, M. Z., Collier, P. A., & Davern, M. J. (2008). Measuring the effects of business intelligence systems: The relationship between business process and organizational performance. *International Journal of Accounting Information Systems*, *9*(3), 135–153.

evolve24. (2012). *Website.* Retrieved September 26, 2012, from http://www.evolve24.com/

Fortuin, L. (1988). Performance indicators – why, where and how? *European Journal of Operational Research*, *34*, 1–9.

Gartner. (2009). *Gartner reveals five business intelligence predictions for 2009 and beyond.* Retrieved June 17, 2012, from http://www.gartner.com/it/page.jsp?id=856714

Gast, A., & Zanini, M. (2012). The social side of strategy. *McKinsey Quarterly*.

Gates, S. (2001). Aligning strategic performance measures and results. *The Conference Board. Balanced Scorecard Interest Group*.

Gessner, G. H., & Volonio, L. (2005). Quick response improves returns on business intelligence investments. *Information Systems Management*, *22*(3), 66–74.

Ghose, A., & Panagiotos, I. (2010). The economining project at nyu: Studying the economic value of user-generated content on the internet. *Journal of Revenue and Pricing Management*, *8*, 241–246.

Gianfortoni, P., Adamson, D., & Rosé, C. P. (2011, July). Modeling of stylistic variation in social media with stretchy patterns. In *Proceedings of the first workshop on algorithms and resources for modelling of dialects and language varieties* (pp. 49–59). Edinburgh, Scotland: Association for Computational Linguistics.

Guo, J., Zhang, P., Tan, J., & Guo, L. (2012). Mining hot topics from twitter streams. *Procedia Computer Science*, *9*, 2008–2011.

Hanna, R., Rohm, A., & Crittenden, V. L. (2011). We're all connected: The power of the social media ecosystem. *Business Horizons*, *54*(3), 265–273.

Huang, Q., Li, Q., Li, H., & Yan, Z. (2012). An approach to incremental deep web crawling based on incremental harvest model. *Procedia Engineering*, *29*(0), 1081–1087.

IBM. (2011). *From social media to social crm. reinventing the customer relationship.* IBM Institute for Business Value analysis. (White Paper).

Ittner, C., Larcker, D., & Randall, T. (2003). Performance implications of strategic performance measurement in financial services firms. *Accounting, Organizations and Society*, *28*, 715–741.

Jansen, B. J., Zhang, M., Sobel, K., & Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth. *Journal of the American Society for Information Science*, *60*(11), 2169–2188.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! the challenges and opportunities of social media. *Business Horizons*, *53*(1), 59–68.

Kaplan, A. M., & Haenlein, M. (2012). The britney spears universe: Social media and viral marketing at its best. *Business Horizons*, *55*(1), 27–31.

Kaplan, R. S., & Norton, D. P. (1992). The balanced scorecard – measures that drive performance. *Harvard Business Review*(5), 71–79.

Kaplan, R. S., & Norton, D. P. (1993). Putting the balanced scorecard to work. *Harvard Business Review*, *71*(5), 134–147.

Kaplan, R. S., & Norton, D. P. (1996). Using the balanced scorecard as a strategic management system. *Harvard Business Review*(5), 1–13.

Kapow. (2009). *Five key lessons for converting social media data into business intelligence.* Kapow Software (White Paper).

Kennerley, M., & Neely, A. (2003). Measuring performance in a changing business environment. *International Journal of Operations and Production Management*, *23*(2), 213–229.

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.

Kirtis, A. K., & Karahanb, F. (2011). To be or not to be in social media arena as the most cost-efficient marketing strategy after the global recession. *Procedia Social and Behavioral Sciences*, *24*, 260–268.

Klassen, A. (2009). Forget twitter; your best online marketing tool is the humble product review. *Ad Age*, *1*.

Kozinets, R. V., de Valck, K., Wojnicki, A. C., & Wilner, S. J. (2010). Networked narratives: Understanding word-of-mouth marketing in online communities. *Journal of Marketing*, *74*(2), 71–89.

KPMG. (2009). *Does your business intelligence tell you the whole story?*

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology.* (2nd, Ed.). Thousand Oaks, CA: Sage.

Laroche, M., Habibi, M. R., & Richard, M.-O. (2012). To be or not to be in social media: How brand loyalty is affected by social media? *International Journal of Information Management*(0), -.

Lee, C.-H. (2012). Unsupervised and supervised learning to evaluate event relatedness based on content mining from social-media streams. *Expert Systems with Applications*, *39*, 13338–13356.

Lonnqvist, A., & Pirttimaki, V. (2006). The measurement of business intelligence. *Information Systems Management*, *23*(1), 32–40.

Lovejoy, K., Waters, R., & Saxton, G. D. (2012). Engaging stakeholders through twitter: How nonprofit organizations are getting more out of 140 characters or less. *Public Relations Review*, *38*, 313–318.

Marwick, A. E., & Boyd, D. (2011). I tweet honestly, i tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, *13*(1), 114–133.

McCorkindale, T. (2010). Can you see the writing on my wall? a content analysis of the fortune 50's facebook social networking sites. *Public Relations*, *4*, 1–13.

Molloy, J. C. (2011). The open knowledge foundation: Open data means better science. *Public Libary of Science*, *9*(12).

Murdough, C. (2009). Social media measurement: It's not impossible. *Journal of Interactive Advertising*, *10*, 94–99.

Murugesan, S. (2007). Understanding web 2.0. *IT Professional*, *8*, 34–41.

Neely, A., & Adams, C. (2005). Performance prism. In E. in Chief: Kimberly Kempf-Leonard (Ed.), *Encyclopedia of social measurement* (pp. 41–48). New York: Elsevier.

Neely, A., Adams, C., & Crowe, P. (2001). The performance prism in practice. In *Measuring business excellence* (Vol. 5, pp. 6–12). MCB University Press.

Neely, A., Bourne, B., & Kennerley, M. (2000). Performance measurement system design: Developing and testing a process-based approach. *International Journal of Operations and Production Management*, *20*, 1119–1145.

Neely, A., Gregory, M., & Platts, K. (1995). Performance measurement system dsign: A literature review and research agenda. *International Journal of Operations and Production Management*, *15*, 80 – 116.

Netbase. (2012). *Website.* Retrieved September 26, 2012, from http://www.netbase.com/our-network/partners/sap-social-media-analytics/

Neuendorf, K. A. (2002). *The content analysis guidebook.* Thousand Oaks, CA: Sage.

O'reilly, T. (2007). What is web 2.0: Design patterns and business models for the next generation of software. *Communications & Strategies*, *65*, 17–37.

Osterwalder, A., & Pigneur, Y. (2010). *Business model generation.* Hoboken, New Jersey: John Wiley & Sons, Ltd.

Patterson, A. (2012). Social-networkers of the world, unite and take over: A meta-introspective perspective on the facebook brand. *Journal of Business Research*, *65*(4), 527–534.

Peterson, R. A., & Wilson, W. R. (1992). Measuring customer satisfaction: Fact and artifact. *Journal of the Academy of Marketing Science*, *20*, 61–71.

Pirttimäki, V., & Hannula, M. (2003). Process models of business intelligence. *Frontiers of E-Business Research*, 250–260.

PWC. (2007). *Guide to key performance indicators. communicating the measures that matter.* Price Waterhouse Coopers. (White Paper).

Radian6. (2012). *Website.* Retrieved September 26, 2012, from http://www.radian6.com/

Ramakrishnan, T., Jones, M. C., & Sidorova, A. (2012). Factors influencing business intelligence (bi) data collection strategies: An empirical investigation. *Decision Support Systems*, *52*(2), 486–496.

Ratner, B. (2003). *Statistical modeling and analysis for database marketing: effective techniques for mining big data.* Chapman & Hall / CRC Press.

Reinhold, O., & Alt, R. (2011). Analytical social crm: Concept and tool support. *BLED 2011 Proceedings*, *Paper 50*.

Rouibah, K., & Ould-ali, S. (2002). Puzzle: a concept and prototype for linking business intelligence to business strategy. *The Journal of Strategic Information Systems*, *11*(2), 133–152.

Shahin, A., & Ali Mahbod, M. (2007). Prioritization of key performance indicators: An integration of analytical hierarchy process and goal setting. *International Journal of Productivity and Performance Management*,

56, 226 – 240.
Sloan, P. (2012). *Facebook earnings match estimates; 955 million users.* Retrieved August 15, 2012, from http://news.cnet.com/8301-1023_3-57480865-93/facebook-earnings-match-estimates-955-million-users/
Smith, A. N., Fischer, E., & Yongjian, C. (2012). How does brand-related user-generated content differ across youtube, facebook, and twitter? *Journal of Interactive Marketing*, *26*(2), 102–113.
Smith, R. (2006). *Key performance indicators. the basics of maintenance and reliability.* Chapter 6.
Stephens, W. (2012). *Research and social media: Why & how to do a content analysis.* Retrieved July 20, 2012, from http://www.accountplanningblog.com/2012/05/research-and-social-media-why-how-to-do.html
Takac, C., Hinz, O., & Spann, M. (2011). The social embeddedness of decision making: opportunities and challenges. *Electronic Markets*, *21*, 185-195.
Tirunillai, S., & Tellis, G. (2012). Does chatter really matter? dynamics of user-generated content and stock performance. *Marketing Science*, *31*, 198–215.
Tsai, Y., & Cheng, Y. (2012). Analyzing key performance indicators (kpis) for e-commerce and internet marketing of elderly products: A review. *Archives of Gerontology and Geriatrics*, *55*(1), 126 - 132.
Turban, E., Lee, J., King, D., & Chung, H. (1999). *Electronic commerce: A managerial perspective.* Prentice-Hall, Inc., New Jersey, USA.
Twitter. (2012). *Twitter turns six.* Retrieved June 17, 2012, from http://blog.twitter.com/2012/03/twitter-turns-six.html
Van Beek, D. (2006). *De intelligente organisatie. prestatieverbetering en organisatie-ontwikkeling met business intelligence.* Tutein Nolthenius: 's Hertogenbosch.
Watson, H., & Wixom, B. (2007). The current state of business intelligence. *Computer*, *40*(9), 96–99.
Weinberg, B. D., & Pehlivan, E. (2011). Social spending: Managing the social media mix. *Business Horizons*, *54*(3), 275–282.
Wikipedia. (2012). *Wikipedia.* Retrieved September 13, 2012, from http://en.wikipedia.org/wiki/Wikipedia
Wisner, J., & Fawcett, S. (1991). Linking firm strategy to operating decisions through performance measurement. *Production and Inventory Management Journal*, *Third Quarter*, 5 – 11.
You, W., Xia, M., Liu, L., & Liu, D. (2012). Customer knowledge discovery from online reviews. *Electronic Markets*, 1–12.
Zeng, D., Chen, H., Lusch, R., & Li, S.-H. (2010). Social media analytics and intelligence. *Intelligent Systems, IEEE*, *25*(6), 13–16.
Zhang, M., Jansen, B. J., & Chowdhury, A. (2011). Business engagement on twitter: a path analysis. *Electronic Markets*, *21*(3), 161–175.