Towards a Social Media Quick Scan

Development of an interactive visualization tool for social media information in support of tactical crime analysis

Annelies Brands
1896504
August 2017

Master Project Thesis
Human-Machine Communication
University of Groningen, the Netherlands

Internal supervisor:
Dr. Fokie Cnossen (Artificial Intelligence, University of Groningen)
External supervisor:
Ir. Arnout de Vries (TNO Groningen)

university of groningen / faculty of mathematics and natural sciences / artificial intelligence

Abstract

Detectives perceive an information overload in using social media data in police investigations, because of the abundance of available information, limitations in human processing capacity and problems in human-machine interaction. This results in a possible loss of relevant information that is publically available from the first minute after a crime. The research field of information visualization is especially aimed at making large amounts of data intelligible.

This thesis aimed at working towards a Social Media Quick Scan by developing a prototype of a tool that provides visualizations of social media information. The goal was to investigate how interactive visualizations of social media information can support detectives in their work, focusing especially on objective reasoning and human-machine interaction. First, theory on information visualization was reviewed on how available social media information could best be visualized. Next, interviews with police employees were conducted to determine requirements on data insights and functionality. The results were combined in the design of the tool. A prototype was developed with partially implemented visualizations of data from an example case and evaluated with police employees.

The design was evaluated positively on meeting the requirements and usability, and resulted in a list of suggestions for further development. In this way, this project contributed towards a Social Media Quick Scan that enables detectives to use social media information earlier in police investigations, support reasoning and possibly reduce information overload.

Contents

1 Introduction

1.1 Social media usage

The use of social media has increased dramatically over the past years. 8 out of 10 Dutch people are actively engaging in social media, and among young people the number is even higher (9 out of 10) [CBS, 2015, CBS, 2016] In this thesis we define social media as *"Media where users communicate with each other online and share things like knowledge, photos, movies, experiences, things to sell, opinions, ideas, software, joint work and games, without the need of complicated technical knowledge."* (translated from [De Vries and Smilda, 2014a]) This increased usage resulted in a revolution in tactical crime analysis, because social media have grown into a new information source. One can think of 140 characters on Twitter[1] associated with location data, a 'like' on a Facebook-post[2] or someone posting an online message about suspicious behavior on the other side of the street. For police investigations, this means that more information is quickly available and freely accessible. Research has shown that social media information can provide relevant knowledge in police investigations: 86% of the surveyed law enforcement agencies indicate that social media helped solve crimes in their jurisdiction [International Association of Chiefs of Police, 2015].

1.2 The problem

Unfortunately, some problems are encountered in the use of social media information in police investigations. A study by TNO [De Vries and De Groen, 2012] showed that detectives perceive an information overload when using social media data, because of three main reasons. First of all, the amount of available data is often too high. This makes it hard and time-consuming to interpret all information. A second problem is limitations in human processing capacities. With the importance of objective reasoning in police investigations, determining the relevance and reliability of all the available information became more complex with information from open sources. This can result in an overload of mental steps to take to validate a piece of information. The result is that sources are (sub)consciously ignored or more easily disregarded as irrelevant. The last problem pointed out is the human-machine interaction. Detectives have to work with several technical systems that are not working optimally, which results in loss of time. Some systems are obsolete; others are not optimally integrated for detective work. Next to that, detectives sometimes lack the expertise to work with these technical systems.

The problems described above account for a limited use of available social media information in police investigations. Sometimes the information is used to a limited extent, sometimes the information is only used later on in the investigation, and sometimes this information is not looked for at all. This relates to organizational issues as well, because it can be unclear who is designated to this task. In some cases, the problems result in the loss of potential relevant information that is publically available from the first minute after a crime.

[1] http://www.twitter.com/. See Appendix A for an example of a tweet and Twitter data

[2] http://www.facebook.com/. See Appendix B for an example of a Facebook page and Facebook data

1.3 Introduction to the project

In the past years, TNO[3], a non-profit organization for scientific research, has conducted research on possibilities of giving meaning to all the available data in context. They researched the development of tools that could automate and visualize the data to make it ready to reason with [De Vries and De Groen, 2012, De Vries, 2013]. The present project fits within the "Vraaggestuurde kennisprogramma informatievoorziening" (Demand-driven knowledge program) of the ministry of National Security and Justice and the goals of TNO to expand the knowledge of data visualization and interaction. The project was commissioned by TNO. The subject of application of this project was the Dutch police.

The first hour after a crime, the so-called 'golden hour', is the most crucial to find possible offenders. All relevant information can provide instant directions for the case and indications for the needed personnel and expertise. Although several tools have been developed to monitor and analyze social media in the past years, these focus either on deep and extensive social network analysis (such as Gephi [Bastian and Heymann, 2009] and NodeXL [Ahn et al., 2011]) or visualize information from just one source (such as the first release of Twitcident [Abel et al., 2012].

This project focused on bridging the gap between available social media information and detectives working in the golden hour, by working towards a tool that enables a quick scan of social media, brings information from multiple sources together and provides quick insights in this information. In this way, the available information can be used earlier in the investigation process and can possibly reduce information loss due to information overload or data that has disappeared between the crime and the time of the investigation. A prototype is presented and evaluated with end users. The design of this tool was approached from the interaction design cycle and used theoretic insights from the field of information visualization that aims specifically at making large amounts of information intelligible [Spence, 2001].

1.4 Goal of the project

To work towards a 'Social Media Quick Scan', the goal of this project was to investigate how interactive visualization of social media information could support detectives in their work. To determine this, the following research questions were formulated:

- How can the results of a quick scan of social media best be visualized?
- How can objective reasoning be supported in the visualization? (tunnel vision prevented)
- Following from the previous question: How can the relevance and reliability of User-Generated Content be validated?
- How can the visualization(s) be accessible for people with different technical expertise levels?

[3] http://www.tno.nl/

The first part of this thesis provides a theoretical background on related subjects. Chapter 2 provides a background on police investigations, the use of social media, and a review on relevance and reliability. Chapter 3 reviews earlier research on information visualization and discusses visualization strategies. Several visualization- and social media tools that have been developed earlier are discussed in chapter 4.

The second part discusses all the design aspects of the development of the tool: The used methodology (chapter 5), the process and results of identifying the needs and requirements of the tool (chapter 6), an explanation of the design and design decisions (chapter 7), and the results of the evaluation of the prototype (chapter 8).

Last, the third part of this thesis discusses the results of the design and the evaluation, and the conclusions that could be drawn from this research.

Part I
Theoretical background

2 Social media in police investigations

The goal of this project was to investigate how interactive visualization of social media information could support tactical crime analysis. This chapter discusses the use of social media in police investigations and addresses what is known from theory on information overload and how relevance and reliability of information can be determined. The results serve as guidance for the design of a tool that supports objective reasoning and prevents information overload.

2.1 The basis of police investigation

To properly address the impact of social media on police investigations, it is important to understand how the police investigate crimes. A crime is seen as consisting of three parts: the crime scene, victim(s) and offender(s) [(ACPO), 2006]. Each of these parts can be broken down into different elements. Detectives usually create profiles of these parts. For the crime scene, examples are track investigation, escape routes and environmental analysis. A victim profile can contain information on lifestyle, the relation between victim and offender or personal characteristics. Possible motives, used violence and number of offenders, among others, are mapped for the offender profile [De Vries and Smilda, 2014b]. Detectives create these profiles by collecting information from different sources. Next, the profiles are used to find connections between all information parts. To connect the information, detectives try to answer the 8 'golden Ws' :

1. Who is it about? (offender, victim, witness)
2. What happened? (type of offense)
3. Where did it take place? (location)
4. When did it take place? (time)
5. With what did it take place? (weapon)
6. Why did it happen? (motive)
7. In which way did it happen? (modus operandi)
8. Why can we say this? (reasons for knowledge)

The first seven Ws were introduced by Hans Gross [Gross and Van der Does de Willebois, 1904, Gross, 1893]. The last W is intended to avoid tunnel vision and asks the question how the answers are established in an objective way [De Poot, 2011]. This means that a detective has to base the answers on the real 'facts' and not on subjective interpretations of data or a gut feeling. The aim of police investigations is to answer all of the questions in order to solve a crime case.

2.2 The influence of social media on police investigations

The rise of social media as an information source has impact on answering the 8 W-questions. Van Berlo [van Berlo, 2012] described three characteristics that distinguish social media from other information sources: open, social and user-centered. The information is open because it is available legally and without formal barriers, for every citizen [De Vries and Smilda, 2014c]. Access to this information is either free or available after registration or payment. In this sense it is publicly available, without the police needing a specific warrant. In the context of police work, social media are therefore seen as part of the 'open sources' [Kop et al., 2012]. Secondly, social media are social, because it is produced by interactions between users. Messages on social media are therefore called 'User-Generated Content', or UGC. Last, social media are user-centered, i.e. not information-provider centered.

These characteristics of social media result in quickly available personal information on persons, relations and locations. Together with the fact that 8 out of 10 Dutch people are active on social media [CBS, 2015], it has impact on answering the 8 W-questions. Given the statistics, it is likely that a victim or offender has one or more profiles on social media. This may provide information on the *who*-question. Timelines on Twitter[1] and Facebook[2] can give insights on the *when* of the crime. The victim and offender may have had some sort of contact before the crime took place. This may have left traces online. And witnesses may post texts, images and videos on social media about something they have seen.

Next to the quickly available personal information, social media also influenced the use of specific investigation methods [De Vries and Smilda, 2014c]. On the one hand, it can have a possible negative influence. Ensuring the safety of the crime scene and crime investigation can be complicated when bystanders film the scene and share the movie on social media. On the other hand, a crime analyst has faster access to witness information through online videos. Multiple times, these kinds of social media information have shown to be relevant in criminal investigation. Examples are the 'Facebook-murder'[4], where the assignment for the murder was given on Facebook, or the missing of two brothers Ruben & Julian[5], where Facebook and Twitter served as a medium for a joint search and sharing relevant information and updates. See Appendix A and B for examples of information that can be found on Twitter and Facebook.

2.3 Current use of social media in police investigations

At the Dutch police, investigations take place at several levels: local, district, regional and national. Social media are used to monitor community sentiment, to get in contact with civilians at a local level and to look for information in crime investigations. This thesis focused on the use of social media for crime detection and excluded the monitoring and community parts.

An investigation process starts at the moment someone calls the emergency number and can take several forms, depending on the type of crime. When we look at the detection of crimes, the use of social media investigation can roughly

[4]http://socialmediadna.nl/facebookmoord/
[5]http://socialmediadna.nl/de-vermissing-van-de-broertjes-ruben-en-julian/

be divided into three groups.

First we have the crime analysis specialists. Their function is to make thorough analyses of suspicious organizations and persons, for example motor gangs. They use open source investigations and can make use of advanced software, such as Analyst Notebook[6] or Palantir[7] or Gephi [Bastian and Heymann, 2009]. This kind of investigation is thorough and time-consuming and requires specialist skills.

Second, the Real Time Intelligence Center (RTIC), located at the emergency control room, provides police units with relevant operational information, in order for them to better perform their task [Bos, 2013]. Here, social media is a standard information source next to closed information sources, to find as much relevant information as possible in the first 15 minutes after an emergency. This investigation is focused on the immediate situation to ensure an adequate rapid response and maximum safety of the involved personnel. However, beyond these 15 minutes the involvement of the RTIC is limited.

Third, the regular police detectives are in the process of making social media investigation a standard part of police work [Politie, 2013]. Unfortunately, this is not yet the case in practice, partly because detectives perceive an information overload in their job, and social media usage is not yet common practice for all investigations. [De Vries and De Groen, 2012].

The next section discusses what is known on the causes of information overload, what specific problems detectives encounter within this context and results from earlier research to form points of attention for the design.

2.4 Information overload

We speak of information overload when the processing capacity of a system is exceeded by the amount of input to the system [Milford and Perry, 1977, Toffler, 1984]. In humans, an overload of information that cannot be processed by the brain results in sub-optimal decision making [Gross, 1964]. This problem is increasingly experienced by social media users [Gomez-rodriguez et al., 2009] and is a major limitation in the use of social media in police investigations [De Vries and De Groen, 2012].

[Groen and Rijgersberg, 2012] extended the definition of information overload to the context of police work: *"Information overload occurs when desired information for making a decision on time remains unnoticed due to a too large amount of (less important) information and/or a too limited information processing capacity, despite the availability of the information."* They conducted field research on the information overload perceived by police officers when it comes to social media investigation. The results are listed under the three main variables in the definition of information overload and illustrated in Figure 1: the amount of information, the available response time and the information processing capacity.

First, there is too much information. The amount of social media is still increasing and, as mentioned before, social media information is open and unstructured. This makes it time-consuming and complex to interpret all of the

[6] http://www-03.ibm.com/software/products/nl/analysts-notebook
[7] http://www.palantir.com

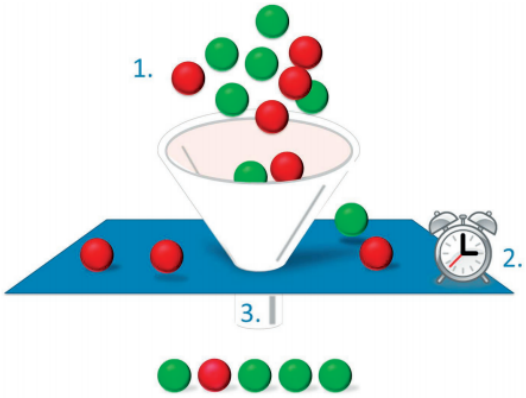


Figure 1: Information overload scheme from [Groen and Rijgersberg, 2012]. (1) represents the amount of information, (2) represents the available response time and (3) represents the information processing capacity.

available information. For example, a short search on Twitter can easily return tens of thousands of tweets to investigate.

Secondly, there is the available response time. Especially in police investigations, the response time is crucial. As mentioned before, the golden hour after a crime is the most critical. Therefore, detectives work under a high time pressure to decide which actions to take.

The third element of information overload is the information processing capacity. De Groen and Rijgersberg listed several factors that limit this capacity. One of them is the encountered problems with the several systems. The technical expertise of detectives varies (mostly depending on age) and not all of the systems are working optimally. They suggest developing software that can integrate and structure multiple sources and (partly) automates the selection and presentation of information. [Simon, 1996] supports the information overload problem by describing a shift in the scarcest resource for people. First this was information, but now it has changed to human attention. Therefore software that is developed for this domain should focus on presenting *"the right information, at the right time, in the right place, in the right way to the right person"*, instead of providing more and more available information (the 5 Rs) [Fischer, 2012]. Social media investigation tools should be developed within the context the users are working in and should be aware of the background knowledge of the users [Fischer and Herrmann, 2011].

Another processing capacity-limiting factor mentioned by [Groen and Rijgersberg, 2012] is the complexity of determining the relevance and reliability of information; two important factors for ensuring objective reasoning. The next section discusses what is known from research on how to determine the relevance and reliability of social media information.

2.5 Objective reasoning: Relevance and reliability of social media information

The detection of crime is completely based on central intelligence [De Vries and Smilda, 2014c], also called 'Intelligence-led Policing' [Ratcliffe, 2003]. Intelligence is *"analysed information and knowledge where decisions and police actions are based on"* [Meesters and Niemeijer, 2000]. Collected data has to be interpreted and integrated into an information piece. In the intelligence process, a distinction is made between data, information, knowledge and intelligence [Kop and Klerks, 2009]. See Figure 2 for a visual representation of the intelligence pyramid.

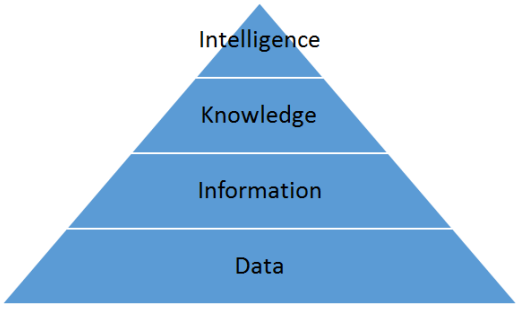


Figure 2: Visual representation of the intelligence pyramid

For example: The number '86' is a datum. This datum becomes information when we know that the number is about the speed of a car, 86 km/h. When this information is placed in a context, it becomes knowledge, for example when we know that this car is driving in a street where the maximum speed is 50 km/h [Kop et al., 2012]. From this and other knowledge, different options for action can be set. The knowledge becomes intelligence when it has been determined as reliable, complete and specific enough that it can be used for decisions and actions [Kop and Klerks, 2009].

Objective reasoning is of the essence in this process. In each step, data is carefully interpreted and given meaning to. The complexity of determining the relevance and reliability of information of social media is one of the limiting factors in the information overload problem. In the next paragraphs, the two are discussed in the context of social media investigation.

2.5.1 Relevance

The first step in the process from information to intelligence is to determine the relevance of information. This means that a detective has to decide whether a piece of information is relevant for the crime case. [Groen and Rijgersberg, 2012] state that this can be hard to determine, especially in the beginning of an investigation, but it is an important step in downsizing the amount of information. Den Hengst [Snel and Tops, 2011] explains that in order for the police to value information, the source of the information and the context of the in-

formation are important. In research on intelligence operations, these qualifiers of information are called 'meta-information'. [Pfautz et al., 2005] studied the impact of meta-information on decision-making in intelligence operations in the military domain. They state that *"for each piece of information delivered to the decision-maker, he or she must make a judgment about the qualities of the data he is receiving"*. The results of their cognitive task analysis show a list of meta-informational factors for different classes of functional in the military domain, with, for example, the type of source, perceived trustworthiness of the source, temporal aspects of the information and accuracy of the content.

When we would translate the list from [Pfautz et al., 2005] to the domain of social media investigation, some factors of this list match the domain. For example: when a Twitter message is found that claims witness information, factors that can determine the relevance of this information could include knowledge on the person that tweeted the information. Questions can be asked, such as: Who is sending this message? When was the message sent? What other information is there that can confirm/contradict this? And what is known about the reliability of this source? The answers to these questions can guide determining the relevance of information.

Recalling the goal of this project, to develop visualization of social media information, the factors described in the previous paragraph must be taken into account in the visualizations. It raises important questions, such as how much meta-information it should show, and when it should be presented. The possible disadvantage of building a tool that adapts the presented information to the user is the 'Filter-Bubble' [Pariser, 2011]: When the computer decides what is relevant and what is not, a detective may lose the overview and context of the information he is dealing with. This could stimulate tunnel vision, which could lead to erroneous decision-making. As mentioned before, objective reasoning is essential in criminal investigation, so the tool should focus on preventing tunnel vision.

2.5.2 Reliability

The relevance of information is closely related to its reliability. Information can be relevant and at the same time unreliable [Groen and Rijgersberg, 2012]. An example is a tweet with information on a plane crash (see figure 3. This information can be very relevant, but is unreliable compared to information from emergency services or news agencies.

Detectives need to keep in mind that decisions are based on information parts they found. Therefore it is very important to assess the reliability of the information. A starting point is to assume that data from social media is false. After all, it is created by humans for several possible reasons. For example, in emergency situations uncertainty and anxiety are the two main reasons that people repeat and invent information that is questionable [Silverman, 2014]. This makes social media information more sensitive to misinterpretation. [Silverman, 2014] describes the importance of verification of information and provides guidelines for investigations. He states that the main questions that should be asked are: 'How do you know that?' and 'How else do you know that?' He provided four verification checks, similar to answering the 8 W-questions:

Figure 3: An example of relevant but initially unreliable information. From: https://twitter.com/jkrums/statuses/1121915133

1. Provenance: Is this the original piece of content?
2. Source: Who uploaded the content?
3. Date: When was the content created?
4. Location: Where was the content created?

Although Silverman states that experience with ways to answer these questions can speed up the verification process, manual processing of all the verification checks is insufficient. Tools are needed to support the detectives, keeping in mind that there is always a role for the human in this process [Gorissen and Johannink, 2016]. That role is especially important in supporting objective reasoning. Examples of existing utilities are reversed Google Image search[8] or Twitter advanced search[9].

This chapter discussed the influence of social media on police investigations and addressed the information overload problem and points of attention for determining relevance and reliability of information. The next chapter reviews earlier research on information visualization and how social media information could best be visualized using this knowledge.

[8] http://images.google.com/
[9] http://twitter.com/search-advanced

3 Background on visualization

The previous chapter discussed the problems with information overload in social media investigation. These problems can be seen as part of the larger information revolution. Information is constantly available everywhere and it is presented to us all day through various media. Next to that, the growing storage capabilities of computers support the generation and storage of large amounts of data [Cao et al., 2012]. The field of information visualization has evolved as an approach to make complex information comprehensible to people. Visualization in itself has nothing to do with computers, it is a human cognitive activity, but nowadays information visualization is immediately associated with computer data [Spence, 2001].

This project focused on how visualization of social media information can support detectives in their work. The way the data is visualized is essential, because it influences the reasoning and decision making in police work. Decisions are based on what information is presented and not on the information that is not presented.

First, we discuss what is known from research on how to visualize data. Next, we discuss what presentation strategies are known that could guide the design of visualizations of social media information and what usability guidelines were developed.

3.1 Information visualization

An information visualization is a visual user interface to information, with the goal of providing users with information insight [Spence, 2001]. [Ware, 2004] explains why creating interactive visual representations of information supports intelligibility: it exploits the capabilities of the visual system to perceive information and reason about it. Humans are especially good at reasoning about visual information and recognizing underlying patterns [Card et al., 1999]. This is subconsciously made clear in the famous expression "a picture is worth a thousand words" and it confirms why infographics are increasingly used as a substitute for paper manuals [Siricharoen, 2013, Artacho-Ramírez et al., 2008].

The goal of information visualization is to provide insight in information. In the Handbook of Human Factors and Ergonomics from [Salvendy, 2012], North [North, 2012] explains that there are different types of insight: from simple insights (like the minimum of a dataset, the maximum or averages) to complex insights (patterns, relationships, structures, anomalies, etc.). Designers of information visualizations make use of human cognitive capabilities by providing the user tools to gain the desired insights much easier than having to look at all the data together or at data presented in only one way. This can help to find underlying patterns and structures that otherwise wouldn't be visible.

3.2 The visualization process

The process of information visualization can be seen as a pipeline (see Figure 4). This pipeline is quite theoretical and goes into small details, but it is worth of being explained because it provides a solid basis. The process starts at the raw data and the structuring of the data. The data can be structured in a table,

in figures, in a tree or network graph or in a text or document structure. The type of structure depends on the type of data and the desired insights.

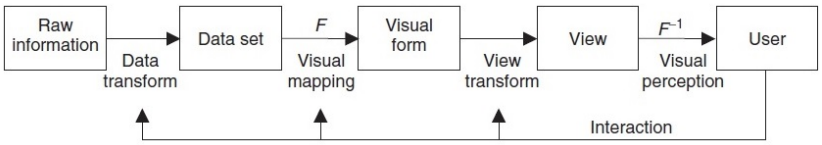


Figure 4: The visualization pipeline, adapted from [Card et al., 1999]

Social media data are often multidimensional data and can be structured in several ways. For example, Twitter data can be structured in a table (see Appendix A) and the relation between friends on Facebook can be structured in a network graph [Hansen et al., 2010]. From the previous chapter, some insight goals for social media information can be imagined, like finding online connections between a victim and suspects or to find specific location information. Specific information on the desired insights is to be determined from the interview results.

3.3 Visual mapping

The next step in the visualization process is mapping the data into visual forms. This ranges from mapping individual data entries to mapping underlying patterns. Figure 5 shows what these basic visual forms look like. Glyphs are points (simple shapes), lines, regions (areas, volumes) and icons (symbols). Attribute values of data are mapped into visual properties of glyphs, like spatial position (x,y,z), size (volume, length), color (hue, gray scale), orientation and shape. Other literature refers to glyphs as 'marks' [Spence, 2001].

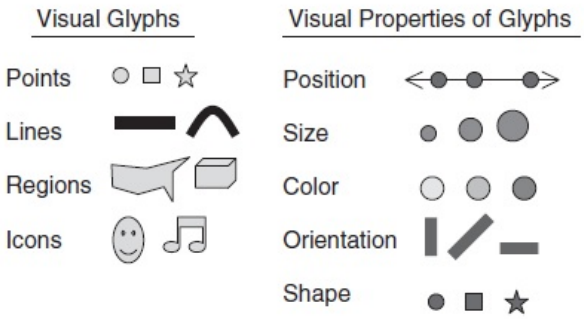


Figure 5: Vocabulary of glyphs and visual properties of glyphs [North, 2012]

The choice of visual encoding of data values and attributes depends on the problem requirements and desired insights. Data attributes with the highest priority should be applied to the most effective visual properties. The visual properties of glyphs in figure 5 are ordered by effectiveness. At the top are spatial position properties. They should be reserved for the most important data attributes, because the human visual system is the best in accurately judging

spatial ratios [North, 2012]. [Wickens and Hollands, 2000] state that data points that are positioned closer to each other are naturally perceived as more related than data points that are further away in the space. This is referred to as the *proximity compatibility principle*. For the remaining properties, Bertin [Bertin, 1983] provided guidance on how the mapping can best support the required tasks. He defined four tasks common to information visualization (see figure 6 for an overview):

- Association: the glyphs can be perceived as similar. Visual properties that can support this are texture, color, orientation and shape.
- Selection: the glyphs can be perceived as different. This task can be supported with size, value, texture and color.
- Order: the glyphs can be perceived as ordered, effectively supported by size, value or texture.
- Quantity: the marks can be perceived as proportional to each other. Only size can effectively support this task.

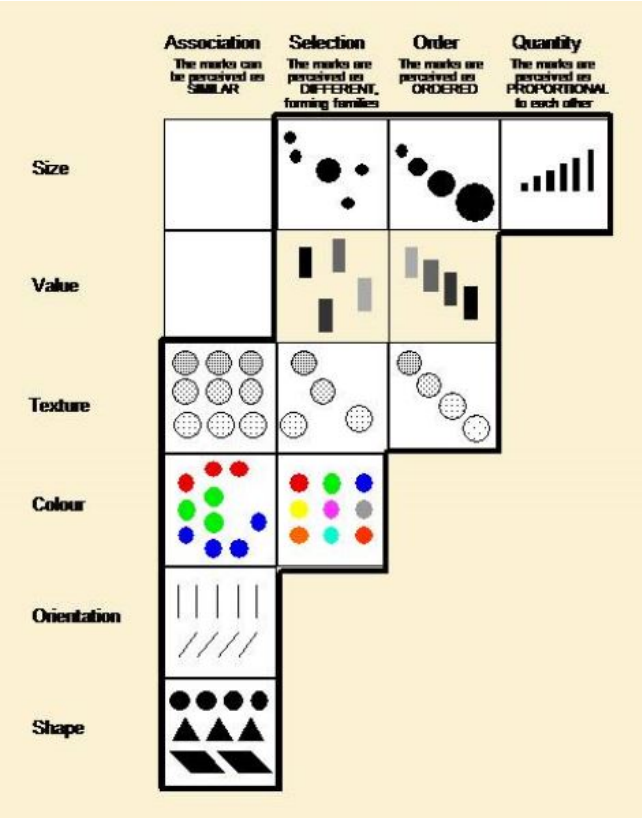


Figure 6: Bertins guidance on visual mapping. Adapted from [Spence, 2001]

Based on this guidance and on empirical evidence, [North, 2012] furthermore provides guidelines for mapping different types of data. For quantitative data,

the user has to be able to estimate order and ratios of data values. Spatial properties are effective for the mapping of this type of data. Color maps and especially rainbow color maps are less effective, because they lack perceptual ordering. For categorical data, the user usually has to be able to distinguish groups. Color and shape are effective for this. For any remaining attributes, the guideline is to apply interaction techniques.

A few notes can be made on the mapping of social media data. A part of that data is about relationships between people, for example friends on Facebook or followers on Twitter. The simplest way of encoding relation between two data entities is drawing a line between them [Spence, 2001]. That is why social media network data is often represented with a network of lines. Specific visualizations of network data are discussed later on in Chapter 4.

The guidelines described above are the low-level basic guidelines for visual mapping. With a large number of data attributes, the visualization becomes more complex, raising issues on available screen space and providing a solid overview of the *presentation* of visual information, which is the next step in the visualization process.

Several strategies were developed to address these issues of presentation: overview strategies, navigation strategies and interaction strategies. The latter serve two purposes: one is to enable users to interactively review remaining data attributes that are not visible at hand, and the other is to generally make contact with the user and to let the user influence the visualization for exploration and clarification. Before we reach the last step of the visualization pipeline, the next three paragraphs discuss strategies for overview, navigation and interaction.

3.4 View: Overview strategies

One of the major limitations of visualizing large amounts of data is the amount of available screen space, the visual scalability issue. There is usually not enough space to display all the data plus the data attributes, and if there were enough, then the individual data points would diminish due to cluttering.

[Shneiderman, 1996] summarizes his solution in the so-called visualization mantra: *"Overview first, zoom and filter, then details on demand."* He suggests to first provide a broad overview, enabling the creation of mental models and overseeing which data are most important. Other advantages of providing an initial overview are direct navigation to parts of the data by selection form the overview and the encouragement of exploration [North, 2012]. Deciding what information is displayed in the overview and what information is only displayed on detail-level is like deciding what products to display in a store's window. The other two parts of the visualization mantra correspond with the navigation- and interaction strategies and will be discussed in the next paragraphs.

To create an overview with as much relevant information as possible, there are generally two approaches: reducing data quantity before visual mapping and miniaturizing the physical size of visual glyphs. According to Manovich [Manovich, 2011], data reduction is one of the key principles of information visualization. One method is to use aggregation. In this method, entities in the data set are formed into groups, creating a smaller new data set. Several design decisions are involved, like choosing what groups are formed, determining new

attribute values and choosing the visual representation [North, 2012]. One way to group entities is to use clustering algorithms. An example is edge clustering, see figure 7 [Cui et al., 2008].

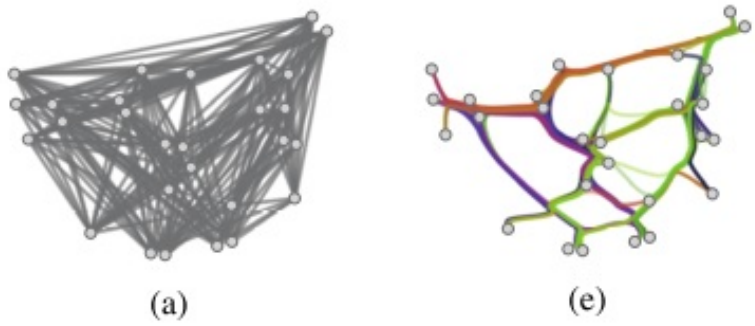


Figure 7: Example of the effect of edge clustering. Adapted from [Cui et al., 2008].

The figure shows a before (a) and after (e) image of an edge clustering process, where each color represents a different group. A geometry-based clustering algorithm was applied on the nodes of a network graph, resulting in a much better overview of the data, and the most important relations are visible immediately. Edge clustering, or bundling, is the most popular data reduction technique for graph visualizations [Liu et al., 2014] and the groupings of entities or data attributes can be based on a various amount of, often domain-specific, algorithms [Hurter et al., 2012,Zinsmaier et al., 2012,Liu et al., 2014,Keim et al., 2008]. Another method to reduce data quantity is to enable filtering through dynamic queries [Ahlberg and Shneiderman, 1994,Ahlberg and Wistrand, 1995]. This will be discussed later on in the section on interaction strategies.

The second approach to creating overviews is miniaturizing the physical size of the visual glyphs. This is argued for by Tufte [Tufte and Graves-Morris, 1983], who explains that in the visual mapping process, the data density on the screen can be increased by maximizing the *data-ink-ratio*. The guideline is to minimize the amount of ink required for the glyphs and to not waste ink on elements that are not data. This saves printing costs, helps providing a comprehensible overview and creates screen space for more data points [North, 2012].

The paragraphs above discussed the most effective overview strategies. As long as the amount of data keeps increasing, the more data reduction should be applied and therefore these techniques will have to keep improving in the future [Liu et al., 2014]. The promising field of *visual analytics* aims at handling massive amounts of data by expanding interactive visualization of data with automatic analysis methods, especially applicable on complex and large datasets [Keim et al., 2008]. They extended the visualization mantra to *"Analyse first – show the important – zoom, filter and analyze further – details on demand."*

The previous chapter discussed that detectives have to be able to determine relevant information to quickly downsize the amount of information. Creating a suitable overview of the data can contribute to this, by presenting all data together and revealing the most salient patterns in the data.

3.5 View: Navigation Strategies

The second part of the visualization mantra described in the previous section is *"zoom and filter"*. The user must be able to zoom in on items of interest and filter out items that are not interesting [Shneiderman, 1996]. After an overview has been created, methods are needed that support the navigation between overview and details. Three of them have evolved as primary navigation design strategies [North, 2012]. They are discussed below.

3.5.1 Zoom + Pan

When the *"Zoom + Pan"* strategy is applied, the visualization starts with an overview and enables the user to zoom in to a level-of-detail of interest. *Zooming* is the *"smooth and continuously increasing magnification of a decreasing fraction of a two-dimensional image under the constraint of a viewing frame of constant size"* [Spence, 2001] and *panning* is the continuous movement of such a frame. Figure 8 shows a well-known example: Google Maps[10].It shows three possible levels of detail. The red pin was placed at the location of broadcasting foundation NOS in the city Hilversum, the Netherlands. The first picture shows the location from an overview of the Netherlands, the second picture shows the location is in the north of the city Hilversum and the last picture shows the exact placement of the building between two streets. From left to right, the view is zoomed in to a higher detail-level through scrolling, and the view is panned in order to present the location in the center of the screen. To find another location, the user can pan across the space on a detailed level, or first zoom out to an overview and then zoom in on the new location of interest.

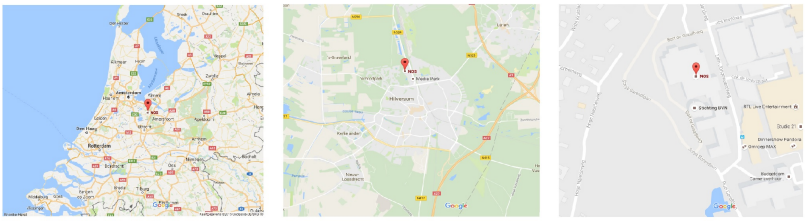


Figure 8: Example of zoom and pan (Google Maps[10])

Spence [Spence, 2001] distinguishes the type of zooming in the example (geometric zoom) from semantic zoom, where zooming in does not only result in a magnified view with details, but also in a discrete transition to a new representation. With Google Maps and other browsing systems, this strategy is efficient in terms of screen space and scalability. There is no clearly defined overview level, which is not needed with these systems. In other cases, there is a potential loss of context when a user is zoomed in.

[10]http://maps.google.com/

3.5.2 Overview + Detail

The second navigation strategy uses multiple views to display overview and detail at the same time. Figure 9 shows the example of the Microsoft Powerpoint [11] editor. On the left, an overview of the slides is presented, and the yellow square (around slide 7) indicates the slide that is currently visible in the detailed view on the right. In contrast to the "Zoom + Pan" strategy, this provides a more stable overview. On the other hand, the multiple views compete for screen space, which is no issue with "Zoom + Pan".

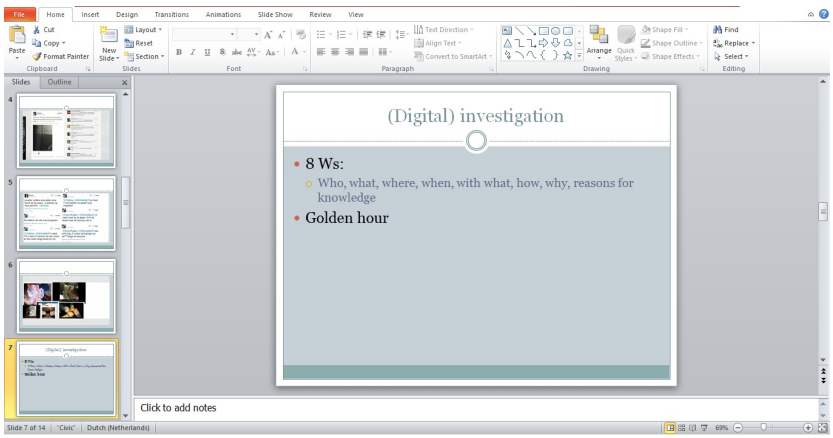


Figure 9: Example of overview and context (Microsoft Powerpoint[11])

Another disadvantage of the "Overview + Detail" strategy is the potential loss of mental connection between views. See for example figure 10. The Drag-Mag technique [Ware and Lewis, 1995] magnifies selected regions, in this case the orange rectangle on a map. How the pink road in the magnified view is connected to the roads on the big map is unclear. This problem is addressed in the third navigation strategy.

3.5.3 Focus + Context

The *"Focus+Context"* strategy displays a magnified part, or focus part, within the context of the overview. This part is enlarged and provides more detailed information than the overview. One can think of a 'fisheye' lens rolling over a map. See for example Figure 11. The focused part is enlarged, and the surrounding parts are diminished through suppression and distortion algorithms [Furnas, 1986].

The fisheye technique can be applied on other data types, like in displaying large menu's [Bederson, 2000]. The advantage of the "Focus + Context"strategy relative to "Overview + Detail" is that the focused view remains connected to the context. In this way, the mental connection between overview and detail is retained. However, the relative distance can be hard to determine in the distorted parts and can be disorienting for the user.

[11] https://products.office.com/nl-nl/powerpoint

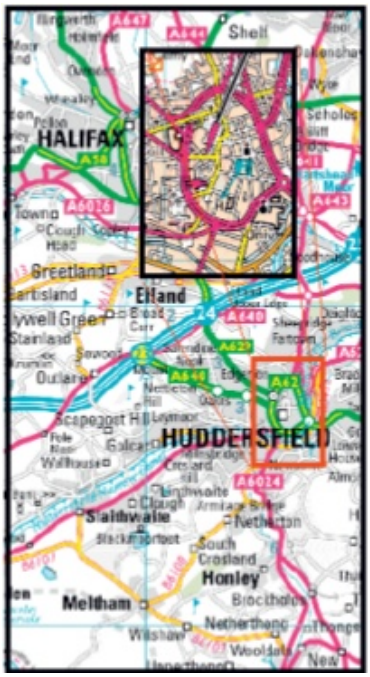


Figure 10: Example of the DragMag technique (from [Ware and Lewis, 1995])

3.6 View: Interaction strategies

As explained before, interaction strategies provide the user options to interact with the data in different parts of the visualization process. This allows for different views of the same data and manual selection of relevant data. After interacting with the data to zoom and filter, usually details can pop-up by clicking on an item [Shneiderman, 1996].

Figure 12 shows Norman's Action Cycle [Norman, 2002]. It shows the psychological steps behind an interaction. The cycle starts with the Gulf of Execution. The user sets a goal to reach some state (e.g. to reach the homepage of a website) and forms this into an intention to do an action (click on the 'Home'-button). The intention is translated into a set of actions that need to be executed to satisfy the intention, and last, the action is actually executed (the 'Home'-button is pressed). The second part of the cycle is the Gulf of Evaluation. The state of the world is perceived (what page does the website show?), interpreted and compared with the user's intention and goal.

Interaction can take many different forms. The most suitable type of interaction design depends on the intentions of the user to interact. This could be learning, exploring, seeking for something specific, opportunistic interaction or involuntary interaction [Spence, 2001]. The different interaction forms can be divided into three modes: passive, continuous and stepped interaction [Spence, 2001].

Passive interaction is the activity that occupies the biggest part of the time a user utilizes a visualization tool. It includes eye-gaze movements and browsing to explore the screen, to see what's there and form an internal model of the

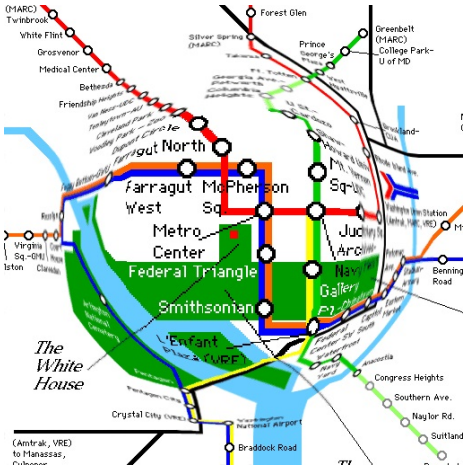


Figure 11: fisheye lens on the map of Washington DC. (From https://www.cs.umd.edu/class/fall2002/cmsc838s/tichi/map.gif)

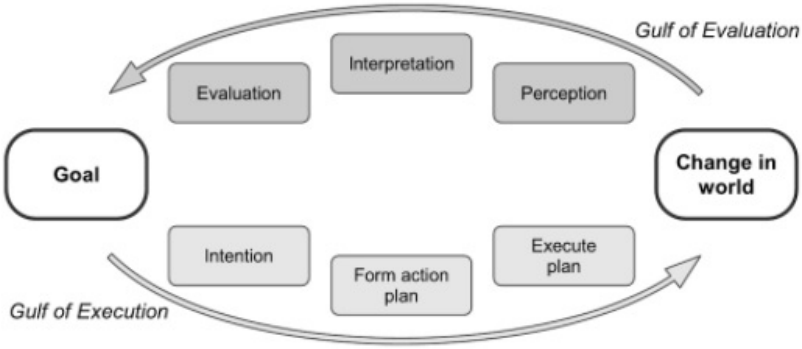


Figure 12: Norman's action cycle. From [Norman, 2002]

information space.

Continuous interaction occurs in highly responsive systems. The visualization is continuously changed on interaction. One can think of dragging a slider from left to right, where the linked visualization immediately changes on sliding.

Stepped interaction is interaction in discrete steps. For example: the user looks at a visualization of data, clicks on a button to pop-up details and evaluates the outcome. Later on, the user might perform a similar action. The difference between stepped interaction and continuous interaction is the frequency of the iteration of the action cycle. Below, three popular interaction techniques are discussed.

3.6.1 Filtering

Filtering is the process of searching a dataset and selecting a subset of the available data, thereby filtering out data that is of no interest at that point. The use of interactive filtering helps in reducing data quantity and focusing

on the information that is relevant for the user. One popular technique is the use of dynamic queries [Ahlberg and Wistrand, 1995]. This type of filtering directly manipulates the data that is visualized. This also enables the user to explore relationships between the data, because filtering for a property can show whether that property is present in the dataset.

3.6.2 History keeping

History keeping is an interaction strategy where a user's history of interaction can be tracked [Heer et al., 2008]. An advantage of this is that a user can go back to previous visualizations that he chose and he can keep the overview of the path of filters and selections that led to the current view. This is also called a 'path breadcrumb' [Spence, 2001]. The other type is the 'location breadcrumb', which provides the user with awareness of their location within the visualization. A well-known example is a line at the top of a web page that shows the location of the page within the structure of the website (see figure 13).

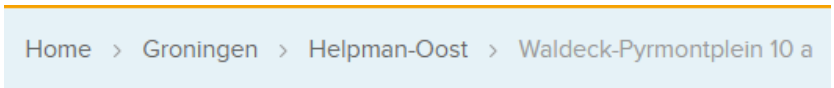


Figure 13: Example of a location breadcrumb on a website. Adapted from http://www.funda.nl/

3.6.3 Brushing + Linking

Brushing and linking is an often used interaction technique, because it supports different types of visualization on the same screen. The idea is that information is related interactively among multiple views [Baldonado et al., 2000]. Entities selected in one view are automatically highlighted in the other view. See figure 14 for an example. The figure shows the brushing and linking of histograms and a geographic map. The histograms represent different population properties. Selecting a percentage of poverty in the histogram automatically causes the individual entities to highlight on the map. The user gets an idea of the location of this group of poverty. Different visualizations can be used at the same time and this helps the user to see relationships.

[Baldonado et al., 2000] added guidelines to the design and usage of multiple views in information visualization. They state that multiple views should only be used when there is a diversity of attributes, the goal should be to bring out correlations; Multiple views should be used minimally and apparent to the user, and the user's attention should be focused on the right view at the right time. That last guideline marks the bridge to the last part of the visualization process: the user.

3.7 The user

At the end of the visualization pipeline (Figure 4), the visualized information is presented to the user. The end goal of information visualization is to make data intelligible for the *user*'s insight goals. The desired insights are different for every design, and sometimes for every user. Therefore, and especially with

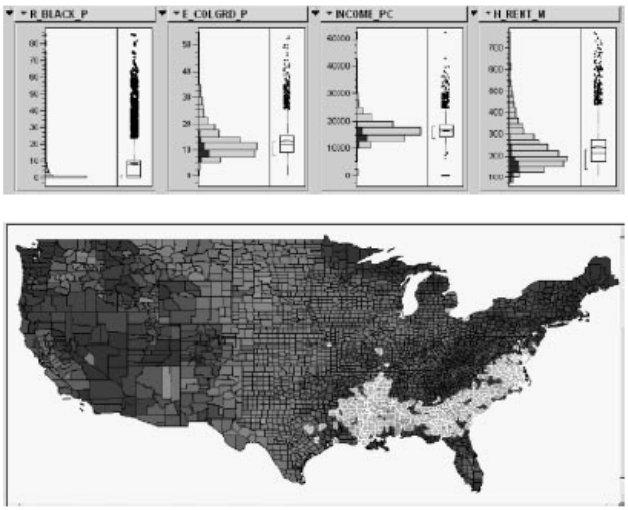


Figure 14: Example of brushing and linking. From [North et al., 2002]

large amounts of data, a visualization tool must be adapted to its context of use, as was discussed earlier in chapter 2. The visualizations should present *the right information, at the right time, in the right place, in the right way to the right person* [Fischer, 2012]. To reach this goal, the designers of visualizations strongly focus on usability. According to [Rogers et al., 2011], usability is *"ensuring that interactive products are easy to learn, effective, to use, and enjoyable from the user's perspective."* It is divided into several goals:

- Effective to use (Effectiveness)
- Efficient to use (Efficiency)
- Safe to use (safety)
- Have good utility (Utility)
- Easy to learn (Learnability)
- Easy to remember how to use (Memorability)

[Nielsen, 1994] set 10 usability heuristics that should be taken into account in design. They are commonly used, and served as usability guidelines for the design of the visualizations;:

1. Visibility of the system status
2. Match between system and the real world
3. User control and freedom
4. Consistency and standards
5. Error prevention
6. Recognition rather than recall

7. Flexibility and efficiency of use
8. Aesthetic and minimalist design
9. Help users recognize, diagnose, and recover from errors
10. Help and documentation

It is necessary for widely used tools to provide accessibility of visualization tools to all kinds of users, regardless of their (dis)advantages, backgrounds or expertise levels. But this remains a challenge for designers. [Plaisant, 2005].

This chapter discussed the theoretical background on information visualization, and reviewed different (re)presentation techniques and focus points. In the next chapter, earlier developed visualization tools are discussed and what can be learned from them for the current study.

4 Review of available tools

The previous chapter discussed how information can best be visualized. It reviewed the visualization process and strategies that are developed to address visualization challenges. For our domain, social media investigation at the police, the specific visualization challenges were reducing information overload, support objective reasoning, and being accessible to users with different expertise levels. In the past years, several tools have been developed to collect, monitor and/or analyze open source data, including social media data. This chapter reviews a variety of these tools and discusses what can be learned from them for our study.

4.1 Data collection

While the focus of this thesis is on visualization of data and not on data collection, data has to be collected before it can be visualized and it is therefore important to discuss different ways of data collection. Most visualization tools include their own data collection technology or assume a given dataset. A few interesting tools have been developed that specialize in the collection of open source data on the web. They are discussed below.

[Pouchard et al., 2009] introduced ORCAT, a tool for systematic collection of open source data. It can automatically build custom collections of web data, and structures it in the ORCAT SQL-database. Users can browse through the data and it updates the content automatically. Two notable properties are the option to structure data based on location information (see figure 15 for an example), and data items are displayed in the context of their source, which can contribute to correct interpretation. Disadvantages are that analyst tools have to be connected for further analysis of the data and that the tool does not support further visualizations of the data rather than displaying all data in a table or on a map.

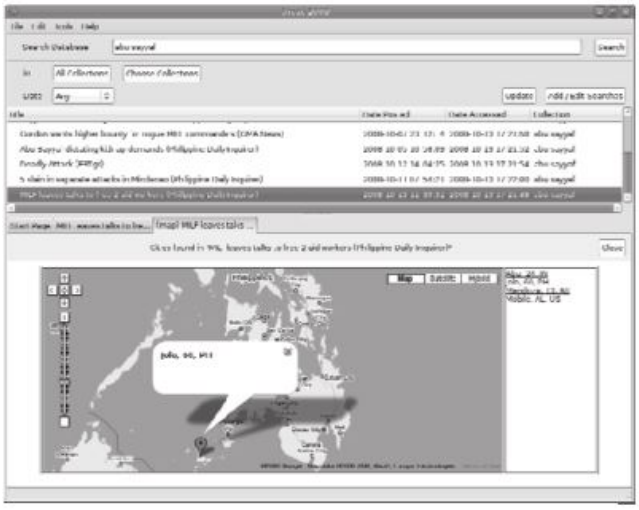


Figure 15: Example of location information extracted from a website. From [Pouchard et al., 2009]

Kaptein et al [Kaptein et al., 2013, Kaptein et al., 2014] have developed Needle Custom Search. This tool supports recall-oriented search tasks on the web (instead of focusing on precision) and provides options to cluster and rank results, based on semantic annotations. See figure 16 for an example search result. This tool is especially applicable to users who search in User-Generated Content and are willing to use multiple search queries to find results. The tool provides entity type annotations (like persons, organizations or locations), Part-of-Speech annotations (the type of language) and temporal annotations. The time stamps and dates for the temporal annotations are extracted with Heideltime [Strötgen and Gertz, 2013], a publically available tagger of temporal data.

Figure 16: The top results of a Needle Custom Search on 'hooligans'. Adapted from [Kaptein et al., 2014]

Needle Custom Search and ORCAT are both promising tools to collect data from User-Generated Content on the web. Especially Needle Custom Search aims at bypassing the Filter Bubble [Pariser, 2011] and presenting data without bias. The output has to be connected to other tools for further analysis and visualization.

As mentioned in earlier chapters, social media data often includes relational data, like followers and mentions on Twitter[1] or a network of friends from Facebook[2]. This kind of data is referred to as social network data, where the structure is built upon individuals, nodes, that are connected to each other through a type of relation, like common interest or friendship [Boertjes et al., 2011]. Social Network Analysis (SNA) is aimed at understanding and interpreting characteristics of those relations that are of interest [McGloin and Kirk, 2010]. Data from social network is often structured in node-link diagrams, such as a hierarchical tree or a graph network without a common starting point (see figure 17 for examples). Popular tools to store the data are SQL, Microsoft Excel[12] and Neo4j[13].

After the data is stored, the relations of interest can be visualized with SNA tools, depending on the research questions and the nature of the links [McGloin and Kirk, 2010]. The required insights derived from the interviews will guide the required visualizations of the network data. SNA tools are reviewed in section 4.3. The next section first discusses general visualization tools.

[12]https://products.office.com/nl-nl/excel
[13]https://neo4j.com/; [Miller, 2013]

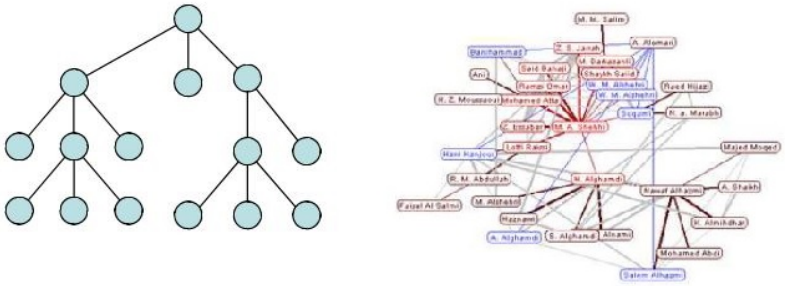


Figure 17: Examples of types of data structures from social networks, with on the left a tree network and on the right a graph network. Adapted from [Boertjes et al., 2011].

4.2 General visualization tools

Several general toolkits have been developed to help a user to visualize their data. An example is InfoVis [Fekete, 2004]. This toolkit supports multiple data structures and visualizations. See figure 18 for examples). The supported data structures currently include tables, trees and graphs. Visualizations that are supported include, scatter plots, time series, treemaps, parallell coordinates, node-link diagrams and adjacency matrices. In all visualizations, dynamic labelling and fisheye lenses can be applied. Disadvantages are that no further animation (interaction) is possible and that the tool cannot be extended to other data structures and visualizations.

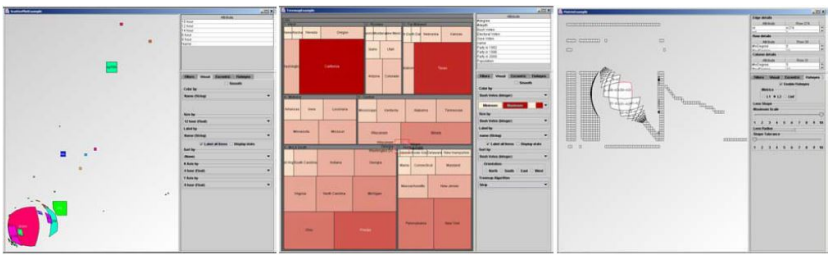


Figure 18: Examples from a scatterplot, treemap and graph visualization, built with InfoVis. Adapted from [Fekete, 2004].

Prefuse [Heer et al., 2005] is a more extended toolkit that offers the user the option to build custom visualizations upon a standard set of visualizations (See figure 19 for an example). In this way, the tool is better extendable, because the user can build widgets that can be set on and off.

[Bostock and Heer, 2009] argue that toolkits like the two described above remain limited in their options. They also state that completely drawing and designing visualization manually would cost too much time and would be too hard. Therefore they developed a toolkit that provides a combination: Small, beautifully designed building blocks that can be built into fully customizable web-based visualizations.

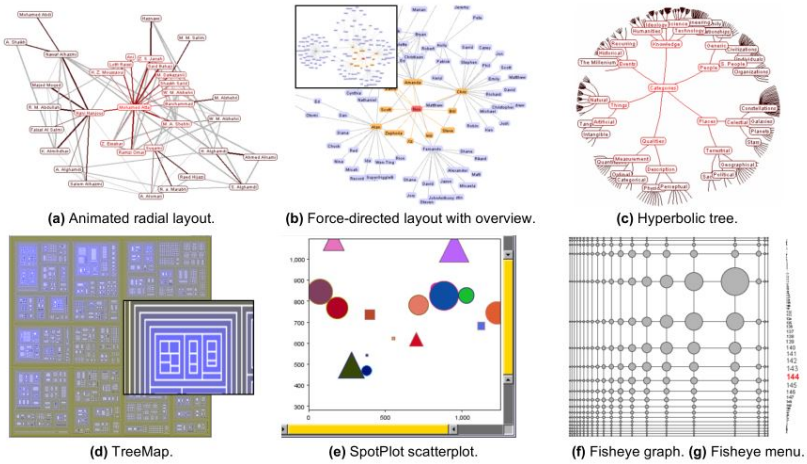


Figure 19: Examples from visualization options in the Prefuse toolkit. Adapted from [Heer et al., 2005].

Regarding web-based visualizations, several Javascript libaries are available. like ChartJS[14], VisJS[15] or D3[16]. These are libraries for creating interactive visualizations on the web. They all provide features to quickly produce bar charts, line charts and area charts. Next to that, VisJS and D3 can be used for creating network graphs and VisJS also provides the creation of a fully customizable timeline. The difference between D3 and the other two libraries is that D3 supports a direct manipulation of web elements, as opposed to scene graph abstractions. The performance of these libraries for visualizations on the web with large data sets is unknown but provides speed challenges on the server.

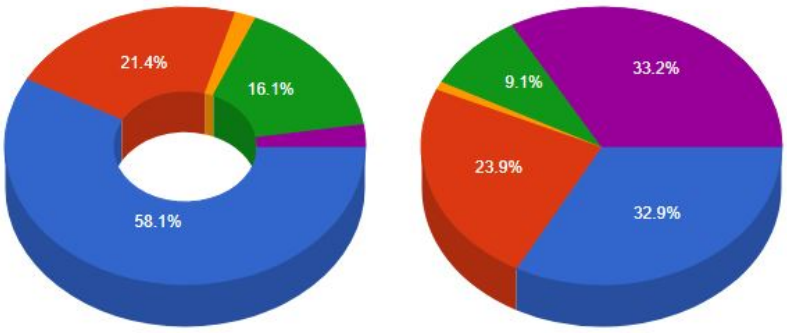


Figure 20: Example from a donut visualization, created with D3. From http://bl.ocks.org/NPashaP/9994181 (date accessed: 30/07/2017)

The paragraphs above discussed general visualization toolkits and web-based

[14]http://chartjs.org
[15]http://visjs.org
[16]http://d3js.org

visualization libraries. Important for our domain was that visualizations could support objective reasoning and would be accessible for people with different expertise levels. The toolkits provide inspiration for what visualizations are compatible with different types of data, but the tools remain limited and require expertise to work with them. The web libraries offer even more visualization options for the data. Unfortunately, they are aimed at users with a strong programming background and therefore not suitable for direct usage for our end users. But they could be integrated in the Social Media Quick Scan tool to base the visualizations on.

The next and last section of this chapter discusses visualization tools that have been developed especially for social media data.

4.3 Social media visualization tools

Regarding visualization tools that focus on social media data, several categories of tools have been developed. These tools focus on specific sources, visualizations or types of data. This sections lists a few of the developed tools to provide a glance on the social media tools as a background overview.

The first category that can be appointed is the group of tools that visualize data from one specific social media source. Examples of such tools are Twinder (search and monitor Twitter data) [Tao et al., 2011], Twitcident [Abel et al., 2012], that can extract Twitter data for visual analytics, and Mentionmapp[17], that displays communucation between Twitter users via mention tweets.

Examples of tools that provide dashboards for monitoring social media are Coosto[18] and Hootsuite[19] and many more. In general, there is a tendency towards using dashboards as an overview of data on a website. See figure 21 for an example.

Figure 21: Example of a dashboard with different visualizations. Source:https://upload.wikimedia.org/wikipedia/commons/0/08/Dashboard_der_ERP-Software_weclapp.png; date accessed: 30/07/2017

[17]http://mentionmapp.com
[18]http://www.coosto.com/
[19]http://www.hootsuite.com/

Regarding timeline tools, the Javascript library VisJS was mentioned before to create timeline visualizations. Next to that, TimelineJS[20] is especially focused on creating timelines from Excel data. Next to that, Storify[21] is an interesting tool that allows the user to build a timeline story with data from social media sources only.

Regarding mapping tools, Google Maps[10] is a popular website. Next to that, Wikimapia[22] and OpenStreetMap[23] were developed to let a user create custom maps. Tools like Fluid Views [Dörk et al., 2012] provide a mapping tool with integrated dynamic querying and semantic zooming.

Several SNA tools have been developed to visualize relational data. Examples are NodeXl [Hansen et al., 2010] that can be built upon Excel data and Gephi [Bastian and Heymann, 2009] that can deal with large amounts of data. Other examples of tools can be found in [Heer and Boyd, 2005, Heer et al., 2008, Perer and Shneiderman, 2008].

This chapter reviewed a variety of visualization tools and provided an overview of the landscape of the tools that have been developed in the past years. Which tools/visualizations can be used as an inspiration and which can be integrated in the tool that is designed for this study will become clear after the identification of needs from the interviews. From chapters 2, 3 and 4, we reviewed the most important subjects regarding social media in police investigations and information visualization. These provide the theoretical basis of this thesis. The next chapters discuss the design and the design process towards developing the prototype of the Social Media Quick Scan.

[20] http://timeline.knightlab.com/
[21] http://storify.com/
[22] http://wikimapia.org
[23] http://www.openstreetmap.nl/

Part II
Design

5 Methodology

The goal of this project was to investigate how interactive visualizations of social media information could support detectives in their work. Four research questions were connected to this goal, regarding how social media information can best be visualized, supporting objective reasoning, supporting the validation of the relevance and reliability of the information and being accessible for people with different expertise levels. Chapters 2, 3 and 4 reviewed earlier research on the related topics. This chapter discusses the methodology that was used in this project for development of the prototype and provides an outline of the following chapters regarding the second part of this thesis: the design.

The stated research questions were addressed by the interaction design approach. This type of design is a user-centered approach to system development [Rogers et al., 2011]. It is aimed at finding the most optimal way to present all of the information from the system to the users of the system. Figure 22 shows the interaction design lifecycle.

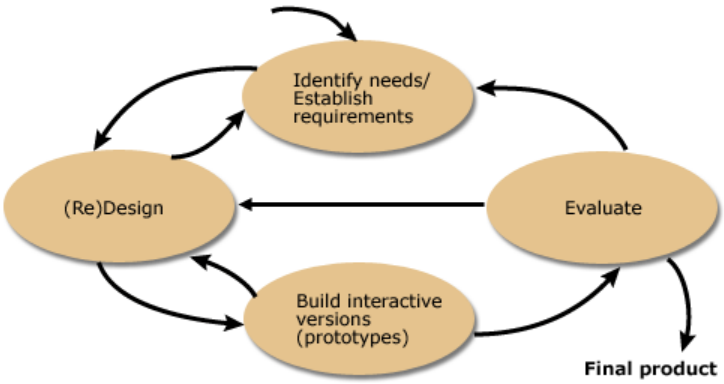


Figure 22: The interaction lifecycle. From [Rogers et al., 2011]

First, we needed to identify the needs of the potential users. The aim was to understand as much as possible about the work of the users, the context they are working in and the users themselves, to guide the development in a direction of supporting the work goals of the users. 11 interviews were conducted with employees at the RTIC and with social media experts and detectives at the police department (9 male, 2 female). We wanted to talk to as many different people as possible, especially to be able to answer the fourth research question. At the RTIC, four people were interviewed of different sex, age, and expertise level with social media. The other interviewees were people from different police departments who are all working with social media. Appendix C shows an overview of the people that were interviewed. The data was gathered via interviews on location. There are several advantages for using the interview on location as data gathering technique [Rogers et al., 2011]. Compared to digi-

tal or paper questionnaires, a face-to-face interview stimulates more elaborate answering of the questions and facilitates the possibility of interaction with the interviewer. An interview that takes place in the user's work environment can result in easier talking about their activities and the context can help them remember specific actions and problems. Naturalistic observation is an even more suitable technique to gather information on the work tasks of the user and their context of work, but due to legal constraints at the police department this was not an option for the present study. Therefore the data gathering took place via interviews only. One of the interviews was conducted with two people at once (due to time constraints). This resulted in an interaction between person's perspectives.

A general interview protocol was made by combining questions regarding the potential 'needs' for the system and questions regarding the four research questions (See Appendix D for a list of all the questions). They were divided into four categories corresponding to the research questions and a general category, asking for identity information and their general opinion about their work and the use of social media. Not all questions were applicable in every interview, because of the different departments the interviews were conducted at.

Next, the interview results were translated into insight goals for the visualizations and requirements for the Social Media Quick Scan tool by designing use cases. Sommerville [Sommerville, 2004] describes user requirements from the field of software engineering. User requirements are "high-level descriptions, typically formulated in natural language". They are "factors or conditions necessary for a user to achieve results" (ISO 9241-210, 2010a, p.17f). This kind of requirements is distinguished from functional requirements. Rogers, Sharp & Preece [Rogers et al., 2011] distinguish the types of requirements differently: Functional requirements, system functionalities and non-functional requirements, describing quality aspects. The last one, non-functional requirements, can contain user needs and usability requirements, for example. The requirements were translated into a use case diagram. The goal of the use case diagram was to describe the requirements in more detail and focus on the tasks the user has to perform [Spath et al., 2012].

The interview results and requirements are discussed in Chapter 6.

After the first step in the interaction lifecycle (see figure 22), the insight goals and requirements were translated into the design of a Social Media Quick Scan visualization tool. The interviewees appointed *threats* to be a type of offense where often relevant information was found on social media. One such case was chosen to build the design on: A possible hostage situation at a Dutch broadcasting station, where a gunman entered the studio and demanded air time on the national news, threatening with several bombs spread across the country. Social media data of the case was collected. Next, low-fidelity mock-ups were designed with Pencil[24] to draw the outline of the design. These were evaluated in meetings with the project supervisors. After this, the design was developed into a web-based prototype of the tool with partly implemented visualizations.

The design and design decisions are discussed in Chapter 7.

The end product of this study was an interactive prototype of the Social

[24]http://pencil.evolus.vn/

Media Quick Scan tool. It was evaluated with possible end users. This was done via a semi-structured interview on location. 9 interviews were conducted, mostly with the same people as the first interviews, except for some RTIC employees (See Appendix C for a list of the interviewed people). As research shows that around 75% of the usability problems can be found through evaluating the system with only 5 people [Nielsen, 1992], 9 evaluations is a sufficient number to draw conclusions on the usability of the design. First, the interviewees were given 10 minutes to look at the design, explore it independently and comment on everything they encountered in the exploration. The participants were encouraged to think aloud during their usage of the tool. All comments and questions were noted. In the next part they were given 5 information-seeking tasks. After that, opinions were asked on specific elements of the design. And last, a list of statements was filled in. This list included statements on reaching the insight goals, the research questions, and a standardized system usability questionnaire (SUS). The first questionnaire contained 17 statements on the insight goals and the research questions. The participants had to fill in a score between 1 (strongly disagree) to 5 (strongly agree). The goal was to gather specific appraisal of the design regarding the insight goals and the research questions. The statements were all formulated positively and participants commented on their answers while filling in this questionnaire. It was filled in by 10 participants, one extra participant who sat next to an interviewee the whole interview.

The second questionnaire was the official System Usability Scale[25]. It is a scale with 10 items, also a Likert-scale, where participants had to rate their disagreement or agreement on the statements (From 1(strongly disagree) to 5 (strongly agree)). It measures efficiency, effectiveness and satisfaction of a system and was created by John Brooke [Brooke, 1996]. 10 participants filled in this questionnaire. The items were scored accordingly and calculated to a usability score[26]. Per participant, the item's score contribution was scored between 0-4 by subtracting one point from statements 1, 3, 5, 7 and 9 and subtracting 5 minus the score on statements 2, 4, 6, 8 and 10. The contributions were summed and multiplied by 2.5 to obtain the overall value of System Usability on a scale of 0 to 100. See Appendix E for the full interview protocol.

The design was presented on a laptop with a mouse and it was launched via a local server. In three interviews, the internet connection was disrupted a couple of times. When this happened, the interviewer had to describe what was shown in some embedded visualizations because it wouldn't load.

The results of the evaluation are discussed in chapter 8.

To reach a final product, the interaction lifecycle is usually run through several times, where each design phase is evaluated multiple times. Feedback on every round leads to an improved design and an improved product, until the goal of the development is reached. The final result of this project is placed at the end of a first complete loop of the cycle. More loops are needed to further implement and improve the system. Therefore, this project served as the first steps towards the development of a Social Media Quick Scan tool.

The next chapter discusses the results of the first step of the cycle: Identifying needs and establishing requirements.

[25]Digital Equipment Corporation, 1986.

[26]See following link for the scoring method: https://www.usability.gov/how-to-and-tools/resources/templates/system-usability-scale-sus.html

6 Identification of needs & Establishment of requirements

The previous chapter discussed the used methodology in this project. This chapter discusses the results of the conducted interviews and a translation of these results to goals for insight in the social media data, and the requirements for the design of the Social Media Quick Scan tool. These are the results of the first step of the interaction lifecycle (figure 22), with the general goal of understanding as much as possible about the context of the users and their desired insight goals. Grouped by their function, the interviewees will be referred to as 'RTIC employee', 'social media expert' or 'detective'.

As mentioned in the previous chapter, the interview questions were divided into categories corresponding to the four research questions and a general category regarding context of social media use (see Appendix D for the complete interview protocol). The results of this last category are discussed first, followed by the results on the questions regarding the four research questions. Each section starts with the goal of the questions of the category, followed by the results.

After the interview results are discussed, the results were translated into insight goals, user requirements and a use case diagram. These are discussed in section 6.6.

6.1 Current use of social media as an information source

The goal of the first set of questions was to find out as much as possible about the current use of social media information in police investigations. We asked the interviewees on how they would indicate their experience level, what they think is the current situation regarding the use of social media information, and how they make use of social media in their work.

8 out of the 11 interviewed people claimed to be experienced in using social media in police investigations, of which two people claimed to be an expert. Two RTIC employees indicated to have a medium expertise level and one detective stated to make little use of social media in police investigations.

In the emergency room, social media are only consulted when there is time left and the RTIC employee estimates that social media investigation could result relevant information. Outside the emergency room, RTIC employees consider the use of social media as a standard part of their job in the *golden hour*. The template of the Situation Rapport (SitRap) contains a chapter 'social media' where the employee is supposed to add relevant information found on social media. Standardized search strings are used to search social media and next to that, different social media sources are searched manually, depending on the type of case. Here, a difference in expertise level was found between the employees. Two RTIC employees indicated that they perform little manual searches outside of the standardized available search strings, while two other employees indicated to perform extensive social media investigation, both specializing in different social media sources.

The social media experts and detectives all argued that there is a limited

use of a social media investigation at the regular police departments. In most cases, the internet is consulted later on in the investigation process, and next to that examples are noted where no social media investigation takes place. They claimed that this is due to capability and capacity problems. All share the vision that social media investigation should be a standard part of any investigation. Two experts and one detective stated that every police officer, regardless of their specific function, should have basic knowledge on how to perform social media investigation. Three experts and one detective voted for open source education for at least one person per team. One detective pointed out that age has a big influence on the knowledge level of social media.

6.1.1 Process of social media investigation at the Real-Time Intelligence Center (RTIC)

The following section results from the interviews at the RTIC and is an extended description of the process that takes place in the golden hour.

When someone calls 1-1-2 in the North of the Netherlands, it reaches the emergency room in Drachten. An employee from the RTIC listens to the conversation with the centralist and immediately starts to collect information from different closed sources. If it is expected that relevant information can be found on social media, the employee also consults these online information sources. This all happens within the first 5-10 minutes after the phone call and is purely aimed at providing highly relevant information to police officers on the street.

If the case requires immediate extensive investigation, the *'golden hour'* starts. RTIC-employees go to their team room to form a Situation Rapport (SitRap). If the situation is upscaled to a TGO (Team Grootschalige Opsporing) or a SGBO (Staf Grootschalig Bijzonder Optreden), the SitRap serves as a report of the information part of the scheduled meetings. Otherwise, the SitRap will be closed after an hour and will be submitted as a report to the detective that takes over the case.

As stated before, a short scan on social media is a standard part of the SitRap-format. A distinction is made between information about victims, suspects, and other information (such as online witnesses). The employees make use of a secured computer, the iRN-computer (internet Research Network). This computer enables the employee to store the information safely and protects them against some forms of tracing by others. The employees make use of programs like Hootsuite[19] and Tweetdeck[27]. If relevant information is found, a screenshot is made and sent by secure e-mail to the computer where the SitRap is formed. Next to that, the iRN-computer saves all the source information of the current website, for administration and evidence logging.

6.2 Visualization of social media information (RQ 1)

The second set of questions regarded the first research question on how social media information could best be visualized. The goal was to inventory insight goals and requirements for the tool. Specific questions were asked on what social media sources are most relevant for police investigations and what type of content is pointed out to be potentially relevant. Next to that, all interviewees

[27] http://tweetdeck.twitter.com/

were asked on their opinion on how social media information could best be visualized and what should be focus points for the design.

The following social media sites and open sources were indicated to be often accessed and designated as relevant sources. They are sorted by number of times mentioned:

- Twitter[1] (10x)
- Facebook[2] (8x)
- Forums and news sites (3x)
- YouTube[28] (4x)
- Google Maps[10] (3x)
- Google+[29] (2x)
- Instagram[30] (1x)

The following types of tools were indicated to be used to search in social media sources:

- Social media search tools, such as Hootsuite[19], Twitcident[31]
- Mapping tools, such as Google Maps[10], Mentionmapp[32]
- Verification tools, such as reverse Google Image search[8]

When a social media investigation is performed, different types of information are often found out to be relevant for the case. Most of the times, social media returned relevant information about persons (victim, suspect or caller). Detectives mentioned profiling to be the main reason to look at social media information. Of special interest are the networks of friends. Interviewees stated looking for information to create a relation scheme, and looking for messages of specific contact between two persons. This could provide information on the relation schemes but also on a possible motive. Furthermore, social media are used to create a visual map with information of the environment of the crime scene. This includes research on interesting buildings in the environment of the crime scene. The last type of information, where mostly RTIC employees and social media experts are interested in, is information gathered from online discussions between people about incidents. This could provide relevant information on the impact of an incident but also information from possible online witnesses.

The types of offense where social media information turns out relevant more often than others where appointed as the following: threats (mentioned 3x), missing persons (mentioned 3x), fraud and violence incidents.

[28] http://www.youtube.com/
[29] http://plus.google.com/
[30] http://www.instagram.com/
[31] http://www.wis.ewi.tudelft.nl/twitcident/ [Abel et al., 2012]
[32] http://mentionmapp.com/

The interviewees were asked about their opinion on the visualization of social media information. One of the main struggle points with social media investigation mentioned was that the searches and analyses are performed mostly manually and it is pointed out to be time consuming. Visualization of the information could help if it would make the search process easier and if it would save time in making connections between information parts. Several of the interviewees wished that some search processes became more automatic. Benefits are seen in creating of a visual overview of the information in order for them to see connections easier. More than half of the interviewees are positive about a larger role of the computer in visualization and automation, but several interviewees mention the importance of focusing on objective reasoning. The visualizations should support human reasoning, and not take over interpretation. Risks are seen in the possible lack of transparency and context of information when a computer would automate too much.

Several focal areas for visualization of social media information were designated. 8 people indicated the importance of showing the source and context of the information, to support objective reasoning. A visualization tool should be able to combine different sources (preferably open and closed). Multiple people mentioned a desire to be able to interact with the visualizations, on the one hand to keep control of the reasoning and on the other hand to play with different detail levels and connections. Other mentioned focal areas for the visualizations were the use of screenshots and pictures, and the divisibility of relevant information with colleagues.

Regarding the hardware context of use, the RTIC-employees make use of PCs only and will continue to do so. Other interviewees preferred the use of their PC for a visualization tool, but pointed out the idea of the use of an iPad 'on the road' where the relevant information can be collected and visualized during field work.

6.3 Objective reasoning (RQ2)

The third set of questions regarded how objective reasoning could best be ensured in the visualizations. The goals were to map the points of attention users would point out for ensuring objective reasoning in general and in social media investigation specific.

All people at the police are trained in objective reasoning. This shows in approximately identical answers on how objective reasoning is best ensured. It is important to keep all the options open all the time and to be careful with interpretations. The source and context of social media information are therefore key information parts that need to be clear. These parts are automatically clear with information from closed sources, but social media information is much more uncertain. In practice, the interviewees try to capture the route to the information as detailed as possible. When a piece of relevant information is found, the previous and next messages are also stored for context information. The role of visualization could be the visual representation of the information in context with insight in the source and route to the information. The combination of social media information with information from closed sources could aid in preventing tunnel vision because more information is available. Risks are seen if the visualization interpreted too much of the information or automatically

provided wrong context information.

6.4 Relevance and reliability of User-Generated Content (RQ3)

The goals of the fourth set of questions were to gather the user's opinion on how relevance and reliability of information should be determined and what role would be provided for the computer in automatically determining these.

The main points of attention to determine relevance and reliability of information that were mentioned regarded checking open sources with police sources and looking at the context in which the information was found. When the same content is found multiple times or on multiple sources, than the reliability of the information is increased. Next to that, the reliability increases if the source has provided reliable information before. This was stated by more than half of the interviewees. Two people mentioned to verify the sources by its metadata and two persons mentioned the option to directly contact the person who posted the message to verify the information. All people stated that the most important thing in this context is to remain critical and always assume that information is false until proven otherwise.

The results of the interviews showed that there is little confidence in the computer regarding the determination of reliability of information. Computer tools and programs are seen as background information that can be used to support human reasoning. Several interviewees refer to their 'gut feeling' as a leading factor in the determination of relevance and reliability. The respondents predominantly prefer for the computer to play a small roll in this process, although two people stated to see opportunities for the computer in suggesting relevance of information. One of those two persons stated the desire for a computer tool to suggest and trigger with new information, like a pop-up saying: "Have you looked at this yet? It might be interesting."

6.5 End users and usability (RQ4)

The last category of interview questions regarded issues to address for usability. The goal of these questions was to find out which properties the potential end users think that a visualization tool should have in order for it to support them in their work, which was the main goal for the tool.

All the interviewed people saw possible applications for people with their function. The RTIC employees state that a visualization tool is best applicable for RTIC employees, detectives desire the social media visualization tool applied for detectives and the social media experts state that police officers at every level could benefit from such a tool.

For the design of a usable tool, several points of attention were pointed out. They are listed below, sorted by number of times mentioned:

- A new tool should contribute to the investigation by speeding up the search process (5x). This should be done by:
  - Automating the search in different social media sources (3x)

  - Combining different sources (3x)

- The visualizations should be simple to work with, with a short learning curve and the usage of visuals rather than text. (3x)
- A visualization tool should allow for interaction with the users (2x)
- One person suggested for a visualization tool to be voice-driven

6.6 Establishment of requirements

The previous sections discussed the interview results that provided the specific needs for the development of a visualization tool for social media information. In this section, the results are translated into requirements that guide the design of the tool.

From the interview results and theoretical analysis some user requirements and functional requirements were identified. From the results section on visualization of social media information, a list of sources was derived with which users work frequently. The user should be able to analyze and work with these types of data in the visualization tool. From the results on which information could be relevant and the focal points on objective reasoning, the following user requirements were set to guide the design of the visualizations:

- Quickly access data from multiple sources
- Connect data from different sources
- Collaborate and share relevant information with colleagues
- Interact with the data
- Insight goals for the visualizations:
  - Find factual information on persons
  - Look into the network of friends
  - Create a visual overview of the environment of the crime scene
  - Insight in the online impact of a crime
  - Select and export relevant information
  - Show the source and context of the information
  - View data in a timeline

When these requirements are specified into more specific system properties, the results indicate that the tool should be web-based, compatible with the iRN systems and possibly compatible with tablets and phones, next to PCs. The tool should be able to interactively integrate data from multiple open sources and support collaboration and sharing information outside of the tool.

The requirements were translated into a use case diagram, with the goal to describe the requirements in more detail and focus on the tasks the user has to perform [Spath et al., 2012]. From this point, the Social Media Quick Scan tool is referred to as the 'system'.

The use case diagram (See Figure 23) shows the system, with several functionalities, and three other elements outside the system. On the left, an actor

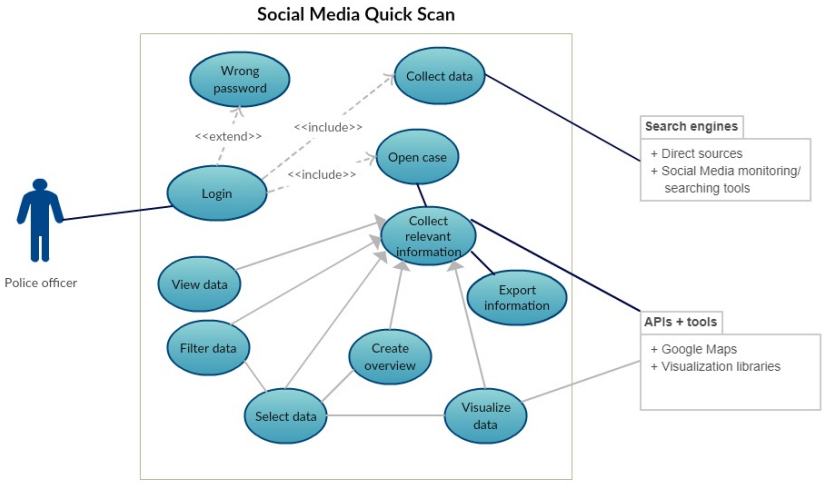


Figure 23: The use case diagram of the Social Media Quick Scan tool

called 'police officer' is shown. This could be any of the possible end users and will be referred to as the 'user' from now on. The other two are external systems that will be connected to the systems. The first is a collection of possible search engines. The system can place data requests through search engines connected to a specific source (such as Twitter Search[9]) or use a search tool (such as the search tools discussed in Chapter 4). The second external system box is the collection of possible Application Programming Interfaces (APIs) and tools that could be connected to the system for visualization purposes. Examples were discussed in Chapter 4 and can vary per case.

The tasks a user has to be able to work with the system are one of the following five:

1. Log into the system
2. Start a new data search
3. Open an existing case
4. Collect relevant information
5. Export the information to a SitRap or other format

The collection of relevant information is of most interest in this project, and it can be broken down into five subtasks:

1. View the data in the original context
2. Filter the data in different ways to highlight interesting information
3. Select data that is interesting to the user
4. Visualize data in different ways to find connections and insights
5. Create an overview of the selected data

These subtasks are related as the tasks affect each other's results and the user can perform these tasks in multiple orders and in a combined manner. For the detailed use cases of every task, see Appendix F. The use case diagram and the described requirements provide the basic guideline for the design of the system.

This chapter discussed the identified needs from the interview results and the translation into requirements for the system. The next chapter discusses the design of the visualization tool.

7 Design

The previous chapter discussed the results from the interviews and described a translation from the results to requirements of the tool. The next steps in the interaction design lifecycle (See figure 22) were to build the design for the tool and build an interactive prototype. This chapter discusses the design decisions and describes all parts of the design and the prototype. First, the concept of the design is discussed, together with the wireframes. Next, a detailed description of the example case is given. And last, the design of the prototype is discussed, in general and specific.

7.1 Concept and wireframes

The design of the tool was based on the set requirements from the previous chapter, and can be described by three main characteristics: a web-based tool, with a pinboard-like concept, combining data from multiple sources. These characteristics are laid out in the next paragraphs.

First, the general concept of the tool was chosen to be in the form of a web page or *web application*. Regarding the nature of the content of the tasks the user will be performing, working with data from social media sources, this is a suitable choice to be the basis of the tool. In this way, the tool can easily be connected to multiple web sources, facilitate online collaboration and be compatible with different systems, such as a PC, tablet or smartphone. Furthermore, this implies no need of installing software before usage and a web server is the only place where it is operated and where data has to be stored. Users are used to working with web pages, so providing the design with web page items like a logo, a menu with items and pages belonging to the menu items will support the requirement of a usable tool with a small learning curve.

Secondly, the concept is based on a *pinboard*. This relates to the familiar idea of a detective connecting the facts on a physical board. A popular example of the pinboard concept is the Pinterest website[33]. On this web site, the user has the ability to search and filter in data, and select the pieces that are relevant or interesting. The selected data is stored to a board. The board shows the overview of the selected data. This concept is applicable to our tool, because it allows the user to interact with the data in several ways and select the data or visualizations that are relevant. The selection on the board looks like an interactive SitRap (the social media part), and it can be extracted to a SitRap or other extensions. It provides an immediate overview of the social media information that is selected and determined to be relevant, and allows for multiple users to work with the board.

The last main characteristic of the design is the concept of *combining data from multiple sources*. As discussed earlier in the theoretical background, a major disadvantage of earlier developed social media analysis tools was that they mostly focus on visualizing data from just one source. The design is built upon the concept of being able to integrate data from multiple sources. To support detectives in their work, the requirement was set to build a tool that integrates other social media tools that they already work with, like using output from

[33]http://www.pinterest.com/

social media search engines and build it to possibly embed several other tools, like Google Maps[10] and networking tools.

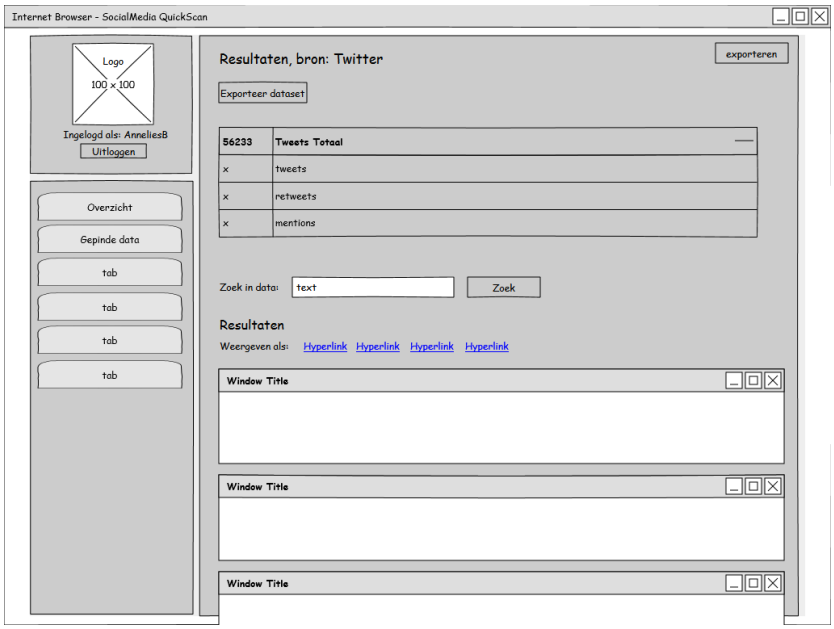


Figure 24: One of the wireframes created in the first design phase. The wireframe shows an example of a page with Twitter results.

Figure 24 shows a low-fidelity wireframe that was developed. It shows the result of the first design of the concept. On the left, space was drawn for a logo an a menu with links to different pages, among which an overview page and a page with 'pinned data'. On the right, an example page is show that displays panels with Twitter search results, the option to search in the data, and space for different visualizations. More wireframes like this were drawn. They are listed in Appendix G.

The design was developed into a high-fidelity prototype. The prototype was built upon a single case. As described in Chapter 5, the case of the possible hostage situation at the Dutch broadcasting station was chosen. The next section describes the timeline and details of the events and the collected data for this case.

7.2 Case description

On January 29, 2015, a man entered the building of NOS, the Dutch public broadcasting station. He wielded a gun, which later appeared to be a fake one, and demanded air time on the national news. He handed a letter to the receptionists containing his specific demands and a threat to detonate 8 heavy explosives with radioactive material and launch a cyber attack if the demands wouldn't be met. He also mentioned to be accompanied by 5 other persons. A security officer brought him to an empty studio, where he was captured by the

police later on. He appeared not to be a terrorist and was working alone. He was identified as Tarik Z., a chemistry student from Pijnacker in the Netherlands. He was treated as a 'confused person' and was charged for hostage.

The events caused a disruption in the broadcast of the evening news for over an hour. Meanwhile, a stream of Twitter messages occurred, with people discussing the cause of the disruption of the broadcast. After it became known what was happening at the NOS, people started a discussion on Twitter who would have done it. Several rumors were spread [Jong and Dückers, 2016]. Afterwards, the Dutch Institute for Safety- and Crisis management (COT) published an evaluation report [Kaptein et al., 2015] discussing the role of the emergency services and the municipality. Based on their reconstruction, the following timeline was set of the events that occurred during the gunman crisis:

- 7:40 pm: The gunman enters the NOS building through the front door, hands a letter with demands to the receptionist and shows his gun
- 7:53 pm: The gunman is taken to studio 10
- 8:00 pm: No broadcast of the evening news at this time. A test image is shown with the text: "Due to circumstances, no broadcast is available at this time"
- 8:07 pm: The ANP press bureau sends out a press alarm, releasing information on evacuation of the NOS building and the appearance of a gunman demanding air time
- 8:14 pm: The gunman is arrested by the police
- 8:20 pm: The commercial channel RTL brings out the new of the arrest; it becomes known to the world
- 8:38 pm: The first picture of the gunman and his letter appear online.
- 9:05 pm: The broadcast of the news is resumed from another location. Video images of the gunman are shown repeatedly.
- 9:31 pm: The gunman is recognized by old classmates and his name, Tarik Z., is mentioned for the first time on Twitter
- 10:30 pm: A press conference is held by the police and the mayor, where the name of the gunman is officially confirmed.

Interesting about this case is that during the crisis, people started discussing the events and began a common search for the identity of the gunman. Information was shared, and little by little a lot of personal information became known to the public, well before the press conference took place. The amount of online messages provided information about the impact of the crisis on the Dutch people, and the content of the messages provided information for the police on personal information and possible henchmen and possible information on the origin of the weapon. For a short time, the events were taken as a national crisis situation, also resulting in the spreading of rumors on the alleged gunman. Two rumors that turned out to be false were the rumor that the same had taken place at the Belgian broadcasting channel VRT at the same time. The other rumor was that Tarik's parents were killed during the MH17 plane crash. Jong

& Dückers [Jong and Dückers, 2016] discussed the spreading and dissemination of these rumors in their article. They collected data from Twitter. The same dataset was used as the twitter data for this project. Next to that, data from other social media sources was collected manually. Together, these data sets were used for the design of the visualizations. The next section discusses the design of the prototype that was built upon the data of this case.

7.3 Development of prototype

From the design concept, an interactive prototype was implemented by building a website. The goal was to give users the idea of what the tool would look like and what functionalities the tool would have. It is mostly a 'clickable' prototype, where screenshots of the functionalities and visualizations were shown. Some parts were already implemented, where the user could actually filter and search in data and interact with the visualizations. First, the technical implementations are described, followed by a description of the general layout and design decisions regarding usability. Next, the layout and functionalities are discussed per page.

7.3.1 Implementation

The static content was implemented with HTML (Hypertext Markup Language). It was used form the structure of the pages and the layout of the panels on the screen.

CSS (Cascading Style Sheets) were used for the styling of the pages and the layout of the buttons and the panels, combined with the Bootstrap plugin[34]. The Bootstrap framework consists of a set of CSS with set up styling options for the HTML elements, providing a consistent layout and responsiveness of the elements to other screens, such as tablets and smartphones. The Font-Awesome[35] and Glyphicon[36] libraries were used for the icons.

The interactive parts were implemented with Javascript and the JQuery plugin[37], a Javascript library with functions for event handlers. This was implemented to respond to the user's clicks, hovers etc. on the elements of the web page.

For the visualizations, different libraries were used: D3JS[38] for loading and setting up the data, MetisMenu[39] for the menu, MorrisJS[40] and RaphaelJS[41] for the donut chart and the tweets-per-minute figure, DataTables with extensions[42] for the interactive Twitter data table, GoogleMaps API for drawing Twitter markers on a maps, and TimelineJS[43] for rendering the horizontal timeline. The choices of which library was used for what visualization were based on which visualization seemed the most useful for every data type, and which insights the

[34]http://getbootstrap.com/
[35]http://fontawesome.io/
[36]incorporated in Bootstrap
[37]http://jquery.com/
[38]http://d3js.org/
[39]https://github.com/onokumus/metismenu
[40]morrisjs.github.io/morris.js/
[41]dmitrybaranovskiy.github.io/raphael/
[42]http://datatables.nl/ Extensions: AutoFill, Buttons, HTML5 export, ColReorder, FixedColumns, Responsive, RowReorder and Select
[43]http://timeline.knightlab.com/

user should get in the data. Based on this, the library was chosen that could visualize that data part in the best way.

7.3.2 General design

The focus of the design was on overview, calmness and usability. From the wireframes, one major change was made in the general strucure of the pages. Eventually, the design was chosen to focus around two main pages: the data overview page and the SitRap-board page. For this, no large menu bar is needed on the left of the screen, because all other pages can easily be accessed from the two main pages. Therefore, the prototype included a menu bar on top of every page, resulting in more screen space for the page content.

The *menu* at the top of every page displays the logo on the left, that always directs the user to the Data-Overview page on click. Next to the logo, the name of the displayed case is shown. Below that, also on the left, two buttons are displayed that link to the two main pages: 'Data Overview' and 'SitRap Boards'. If one of those two pages is displayed, the button changes to a darker color. On the right, two drop-down buttons are displayed with profile options (log-out, change case, settings) and a list of notifications of the last changes the user had made with the tool. All pages have the same menu on top and every page consists of different panels with content.

The *color* of the menu is gray with dark blue buttons for the contrast. The main color of the panels is blue for a consistent and calm appearance. Furthermore, mainly four basic colors were used for consistency and a clean layout: blue, red, yellow and green.

At every page with displayed data, the user has the option to *pin* the selected information to a board. This is displayed with a green button with the logo of a drawing pin. Upon selecting this option, the user can annotate the data to a W-question, relate it to other data and add comments. Screenshots of the design of this pop-up can be viewed in Appendix H at Figures 52 and 53.

Design priorities regarding goals of insight and usability
From the insight goals and usability requirements, the following principles were set for the design:

- Embed data directly when it is possible, to shorten the path from the tool to the original content. This supports viewing the data in its original context as much as possible and this support objective reasoning and reasoning about the relevance and reliability of the data.
- Always show the source of the information and meta-information when it is available. This is also for viewing the data in its context and provide easy linking to the original content.
- The user can always go back to the Data-Overview page by clicking on the logo on the top-left of the menu. This logo is always visible and provides consistency and prevents the user from getting lost in the tool.
- In the pop-ups, we chose to use multiple closing buttons so the user can always know how to get out (cross in the top-right corner, 'close' button, and click outside of the pop-up)

- If comments are available on some data, the comments are only shown on interaction. In this way, the user can look at the data without the interpretations and opinion of others, before looking at the added comments.

This section discussed the general design of the tool. Several pages were developed. The next section shows screenshots of the pages with descriptions of the functionalities and specific design decisions.

7.3.3 Design of the pages

Several pages were designed where the user can select, filter and interact with the data, creating visualizations and overviews. The actions the user can perform correspond with the tasks set up in the use case diagram. This section discusses the design of the specific pages, starting at where the user would start: the login-page.

Login

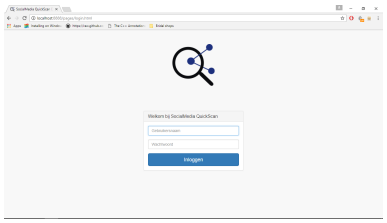
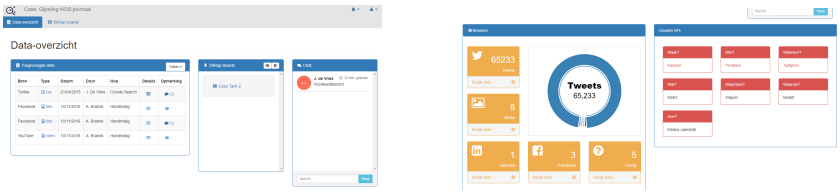


Figure 25: Screenshot of the login page. See Figure 46 in Appendix H for larger screenshot.

The prototype starts with a login page (see Figure 25), where the user has to enter a username and password before entering the content of the tool. After loggin in, the user is directed to the Data-Overview page.

Data-Overview page



(a) Upper part of the dashboard (b) Lower part of the dashboard

Figure 26: Screenshots of the Data-Overview page. See Figures 47 and 48 in Appendix H for larger screenshots.

The first page that is displayed after logging in is the Data-Overview page (See Figure 26). This page displays a dashboard with an overview of all the added data and several links to other pages where the data can be viewed in specific ways. 5 panels with different content are displayed.

The *Added Data* panel displays a table with details of the added data sets and data entries. The user can gain insight in the details on click, like the date and time the data was added, added comments by colleagues, or the used search queries. Here, the *Overview+Detail* strategy is applied. Furthermore, a button is displayed which provides the user options to add new data to the case.

The *Sitrap Boards* panel displays an overview of the available boards. The boards serve as a button that link the user directly to the corresponding page.

The *Chat* panel provides the user to leave general comments for colleagues. The user can enter a message in the input field and clicking 'Send' will display the message in the chat panel.

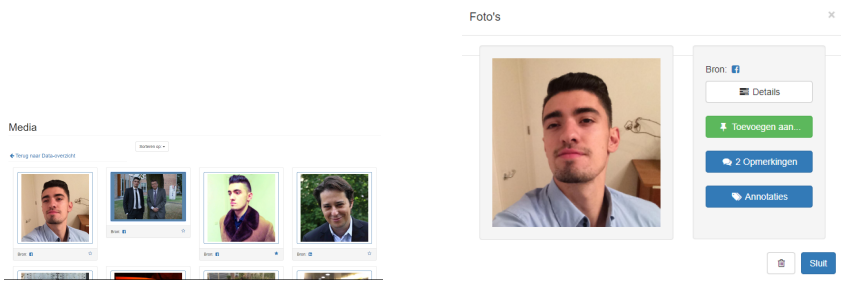
The added data can be viewed in different ways at different pages. These pages are sorted by two subjects: Pages with data from specific sources and pages with data grouped by one of the Ws.These two subjects are represented in the two panels at the lower part of the dashboard (See Figure 26b).

The *Sources* panel shows a donut chart at the right. This chart displays the distribution of the amount of data on the different sources. Clicking on the slices will show the source and the amount of content corresponding to the slice. Surrounding the charts, different smaller panels are displayed, one for every source. The panel shows the name of the source, the corresponding logo, and the amount of content from the source. At the bottom of each panel is a button that links to the page of the source. Three of these source-pages were implemented: Facebook, Media and Twitter. These pages are described later on.

The *Golden Ws* panel contains contains seven smaller panels, corresponding to the seven Ws. Data that is tagged to one of these Ws will be shown on the page corresponding to it. The panels link to the pages. Three of the pages were implemented: "Where", "Who", and "When". These pages are described later on.

The user can start looking into the data by clicking on one of the links to the pages that display data from a specific source. The next paragraph describes the three corresponding pages that were implemented.

Media



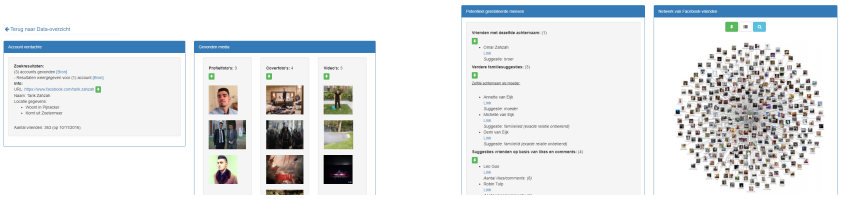
(a) Screenshot of the Media page

(b) Screenshot of a pop-up at the media page

Figure 27: Screenshots of the Media page. See Figures 63 and 64 in Appendix H for larger screenshots.

When the user navigates to the Media page, an overview of the available media is shown in thumbnails, accompanied by two icons: an icon of the source that links directly to the original content on the social media source, and a star icon that allows the user to indicate relevance. The option is displayed to sort the data on demand. Clicking on a thumbnail displays a pop-up (see Figure 27b) with detailed information and options for annotations and commenting.

Facebook



(a) Screenshot of the upper part of the Facebook page

(b) Screenshot of the lower part of the Facebook page

Figure 28: Screenshots of the Facebook page. See Figures 54 and 55 in Appendix H for larger screenshots.

When the user navigates from the Data Overview to the Facebook page, four panels are displayed with different Facebook data.

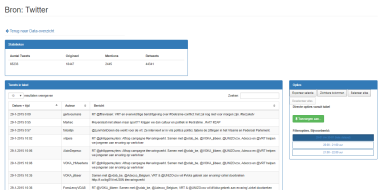
The first panel (See Figure 28a) shows details of the Facebook search on the suspect's name. The user can navigate to the source or pin this data to a board. Next to that, the second panel shows the media that was found on the Facebook Page, grouped by the type of photo. The user can pin a photo to a board or click on a photo to navigate to the original picture at the source.

The lower part of the page (See Figure 28b) displays a panel on the left with suggestions of Facebook friends that are likely to be friends or family in real life. The suggestions are based on the amount of media and messages a Facebook-friend liked on the page of the suspect. At this moment, this was counted manually. The lower right panel displays a network of the Facebook friends in a graph. At the moment, a static screenshot of a graph visualization is shown. When the tool is fully implemented, one can imagine this panel would show an embedded version of the tool or another graph visualization tool to display an interactive visualization where the user can zoom, filter and select data in the network of friends.

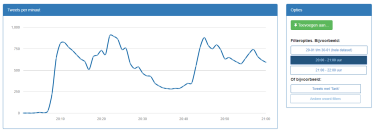
Twitter

When the user navigates to the Twitter page (See Figure 29), several panels with visualizations are displayed among each other, of which most are actually implemented. This page shows the Twitter data that was scraped with Coosto. The user can interact with each visualization either directly or via the filter buttons that are displayed next to the visualizations.

After the panel with the statistics, the data is presented in a *table*, generated with the DataTables-plugin. The user can choose the amount of tweets that are shown on one page and next to that is an input field where the user can search in the tweets. The columns with date+time, author and content of the tweets



(a) Screenshot of the upper part of the Twitter page



(b) Screenshot of the middle part of the Twitter page

(c) Screenshot of the lower part of the Twitter page

Figure 29: Screenshots of the Twitter page. See Figures 56, 57 and 58 in Appendix H for larger screenshots.

are shown at hand; the other columns are visible on interaction. The user can drag the columns in the order he wants them to be displayed. At the bottom of the panel, below the table, buttons are shown to navigate through the pages. The user can filter and (de)select tweets in the table. Other table options are shown in the panel on the right. The selected tweets can be exported in 5 ways: Copy to clipboard, print, and immediate download in a pdf-,excel-, or csv-file. The 'Column visibility'-button lets the user interact with the table by letting the user choose which column are visible in the table. Next to the initially displayed columns, the user can select 'url' and number of followers. The buttons next to that allow the user to select and deselect all the tweets. Furthermore, the 'Pin'-button is displayed where the user can add selected data to the board. Last, options are displayed to filter the tweets on time and date.

Below that, the amount of tweets is displayed over *time* (See Figure 29b), rendered with the Morris library. The user can filter on time or words, and hovering the figure displays details on demand.

The next panel shows an embedded Google Maps-*map* where the tweets with location information are plotted.The tool checks which tweets contain a geo-location and plots them on the map with the Google Maps API. The user can zoom in and out on the map. When the user clicks on a marker, a pop-up is displayed with information regarding the tweet. If the tweet still exists (which is checked the Twitter oEmbed service), the embedded tweet is shown. The user can click on the tweet to open a new tab with the tweet in its original context on the Twitter website. If the tweet doesn't exist anymore the tweet and author information are displayed in plain text. This is possible, because all the data was scraped from Twitter and stored in the data set.Below that, the address is displayed. The address is obtained by sending the coordinates to the Google Maps server and storing the responded address in jSON format. The user can click on the link next to the address to open a new tab where the address is immediately displayed in Google Streetview. Next to the map-panel,

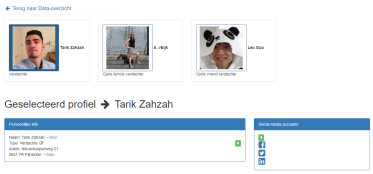
a panel with options is shown. The idea is that the user can filter in the map on a specific time or location or word.

At this page, a suitable option would have been to apply the *Brushing+Linking* strategy. Since all visualizations are based on the same data set, clicking a filter button could apply the filter to all visualizations. In the current design, several filter options were created for each separate visualizations. Considering the amount of visualizations on this page, the user has to scroll down to see them all, the choice was made to let a click on a filter button only apply the filter to the corresponding visualization to avoid confusing and loss over overview of what the consequences are for the other visualizations.

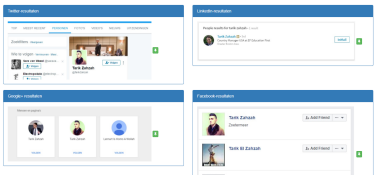
The Twitter page is the last implemented page from the pages with data from the social media sources. The next set of pages correspond to the golden Ws.

Who

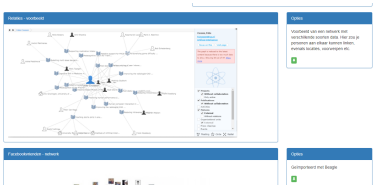
The second way to view the data from the Data-Overview page are to view the data that was annotated with one of the golden Ws. A prototype of three of these pages was implemented. The first one is the *Who*-page.



(a) Screenshot of the upper part of the Profiles page



(b) Screenshot of the middle part of the Profiles page



(c) Screenshot of the lower part of the Profiles page

Figure 30: Screenshots of the Profiles page. See Figures 59, 60 and 61 in Appendix H for larger screenshots.

When the user navigates to this page from the Data Overview, a page is displayed (see Figure 30) that show information that is connected to a specific profile by attaching selections of data from the 'Source' pages (Twitter, Facebook, etcetera). Examples of profiles could be all information on the suspect or on relations of the suspect. An example page was created of the first one, the profile of the suspect. All parts of the page contain a Pin-button.

The panel with *Personal information* (See Figure 30a) shows personal details that were found on social media sources, with a link to the source.

The *Social media accounts* panel (See Figure 30a) displays buttons with icons of the found social media accounts. Clicking on the buttons links the user

to that specific page.

The *Media* panel shows the attached media.

The *Twitter results*-, *LinkedIn results*-, *Google+ results*-, and *Facebook results* panels (See Figure 30b) show the search results on different social media sites. The user can click on them and open a new tab where the results are shown on the original website. From there, the user can search the social media site.

The *Relations* panel (See Figure 30c) shows an example of how different types of information of a person could be related to each other. This should be implemented as an interactive tool to link persons and locations and other types of information to each other, and add it to the board.

Last, the panel with *Network of friends* is displayed at the bottom of the page and shows exactly the same graph as the one on the Facebook page.

Where

Figure 31: Screenshot of the Where page. See Figure 62 in Appendix H for larger screenshot.

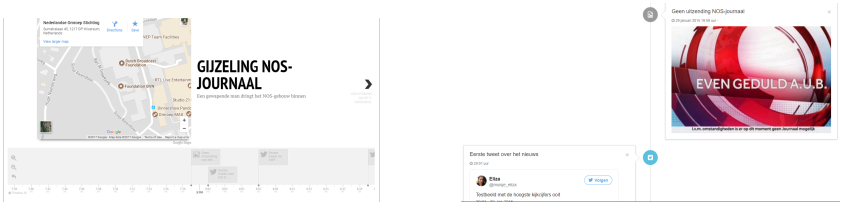
Whenever the user finds interesting information regarding locations, the user can attach the information to the *Where*-page with the Pin-button. Four example maps are displayed. The first panel shows an embedded Google Streetview map of the location of residence of the suspect. The second panel shows an embedded Google Streetview map of the location of the crime scene. The third panel shows a screenshot of a map with plotted Tweets with geo-location in a specific hour. The fourth panel shows a screenshot of a map with the location information of a specific tweet of interest. At the top of every panel, two buttons are displayed: One button that pops up details on the added information, including added comments, and the Pin-button, which allows the user to add a map of interest to the board.

When

The *When*-page sshows two examples of the presentation of selected information in a timeline.

The *Horizontal timeline*(See Figure 32a) was created with TimelineJS[20] and embedded on this page. The selected data was placed in a spreadsheet from which the plugin generated the timeline. The user can click through the timeline with the buttons on the left and the right. The tweets and maps are embedded and therefore clickable to be able to direct the user to the original content. Below the items, a larger timeline is shown where all the items are placed in.

The same selection of information is displayed in a *vertical timeline*(See Figure 32b). The type of source is displayed with the icon on the timeline and



(a) Screenshot of the horizontal timeline

(b) Screenshot of the vertical timeline

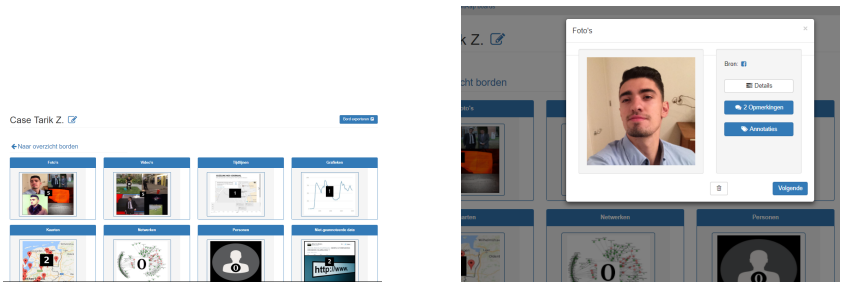
Figure 32: Screenshots of the 'When' page. See Figures 65 and 66 in Appendix H for larger screenshots.

each parts shows date and time, embeds the data when possible and provides the option to delete the part from the timeline.

Both timelines can be added to the board. The goal was to create different options of timelines and to ask the user for its preferences.

SitRap Boards

The previous paragraphs described all pages that were designed for the user to select, filter and interact with the data in different ways. At every step, the user has the option to add a selection of the data or a complete visualization to the pinboard, the *SitRap Board*. The goal was to let the user create an overview of the selected data which can be exported or shared with colleagues. One sample SitRap board was created (See Figure 33). The user can navigate to this page directly by clicking on the button in the *SitRap Boards*-panel in the Data Overview, or navigating to an overview of the boards by clicking on the large button in the top menu and then clicking on the button of the board.



(a) Screenshot of the SitRap Board

(b) Screenshot of a pop-up on the SitRap Board

Figure 33: Screenshots of the SitRap Board. See Figures 50 and 51 in Appendix H for larger screenshots.

Below the menu banner, the title of the board is displayed with an edit button next to it. Clicking on the button opens a pop-up with options for editing the name and description of the box and a delete button.

In the board below, several panels are displayed with data that was selected by the user. At every panel, thumbnails of the content are shown together with the amount of data that is attached to it. The first two panels contain the

attached media. Clicking on the photos opens a pop-up (See Figure 33b) that shows the first photo, and the options to click through the attached photos with 'next' and 'previous' buttons. Next to every picture or video, the following buttons are displayed: the logo of the social media source which acts as a url to go the web page where the picture was originally found, a 'Details' button that shows metadata, date added and person who added it on click, a 'Comments' button that shows added comments on click and the option to add a new comments, an 'Annotations' button that shows the tags added by the user, and last a 'delete' button, to delete the picture or video from the board (not from the tool).

The other panels respectively show specific added visualizations (Figures, maps, timelines), profiles and a panel with 'not-annotated data', data that wasn't tagged by the user to one of the other panels.

On the top-right, an *Export* button is displayed. Clicking this button opens a pop-up with export option.

7.4 Summary

This chapter described the design of the Social Media Quick Scan tool. First, the concept of the tool was described. It consists of three main characteristics: A *web-based* tool, with a *pinboard-like* concept, *combining data from multiple sources*. Low-fidelity mock-ups were designed and these were developed into prototype built upon the case of a possible hostage situation. This case was described in sectoin 7.2. After that, the design of the prototype was described by first listing the tools and plugins that were used for programming the tool, followed by an explanation of the general design decisions. Last, all elements of the pages were described.

This prototype is the result of the strategies and guidelines from the theoretical background, inspired by the earlier developed tools and guided by the insight goals and requirements that were translated from the interviews.

After the prototype was finished, it was evaluated with the intended end users. The next chapter discussed the results from these evaluation.

8 Evaluation

The previous chapter discussed the results of the design process and the properties of the design of the Social Media Quick Scan tool. The goal was to build interactive visualizations of social media information to support detectives in their work. The design was built upon theory on visualization of information and user requirements drawn from several interviews at the police. The final design was a web-based interactive prototype. This chapter discusses the results of the evaluation of the prototype. The goal of the evaluation was to investigate if and how the prototype met the stated requirements. This is the last step in the interaction design cycle (See Figure 22). Details of the interview method were described in Chapter 5 and the full interview protocol can be found in Appendix E.

The results of the interviews are discussed in the next sections and are grouped into the four parts of the interview. Each section starts with the goal of the questions of the category, followed by the results.

8.1 First impression of the design

The goal of the first part of the interview was to collect the comments the participants would have upon the first view of the design. In general, this resulted in a few first comments on the layout of the tool, next a lot of questions and thoughts on the functionalities, followed by several suggestions.

Most participants first mentioned the clean appearance of the layout and the feeling of having a good overview of the available data. The panel with the overview of the sources and the panel with the Ws immediately caught the eye and were indicated as "handy" and "beautifully arranged" by multiple participants. Upon further exploration, multiple users mentioned that the page with the maps was very interesting, beautiful and quickly insightful. Next to that, much interest was shown in the list with Friend suggestions on the page with the Facebook data. Two participants mentioned they liked the timelines and one person exclaimed: "This is easy to work with, this is no rocket science".

The participants were asked to think aloud on what they thought the buttons would return or what the elements would represent. Most of the times, the functionalities corresponded to the participants expectations. However, some things were unclear to participants or were expected but not present in the tool. Four participants asked the question whether it was possible to add data from other sources. The overview of the sources now displayed five different sources with their own panel, and it was unclear to some whether this were all the sources from which data was added and it was unclear to some whether these sources where the only option to add data from or not.

Furthermore, some questions were asked on the table of tweets. At the top right of the table, an input field is shown where the user can filter tweets. Five comments on this were made during the search for information when participants performed the exercises. Two participants mentioned that this search functionality was not visible immediately. And three participants wanted to press the enter key or click on a search button after filling in a word in the input field. These functionalities are not available in the present design, because

the tweets are automatically filtered upon entering a word in the input field. This was confusing to some participants, because it felt natural to press enter to apply a search or filter.

Another element that raised some questions was the 'Added data'-panel on the page with the data overview. Two people mentioned it was unclear to them which information was shown in the table, and two participants pointed out it was not clear how the data was added to the tool. Other people mentioned that the purpose of information displayed in the panel was not immediately clear, but they expected it would show details on the added data. Two participants expected that this data would be added real-time. This is not the case in the initial concept, but is an interesting option.

The last thing about which some questions were asked was how the tool will deal with multiple cases.

Three things were appointed by multiple participants as missing in the tool. The first one is that the search query of the tweets is shown behind the details button on the overview page, but is nowhere to find on the Twitter page. Four participants mentioned that this is important information to show on top of the page, because, as they state it, "everything you find or don't find depends on your search query". The second missing thing that was pointed out is an option to show only original tweets and discard retweets and/or mentions. Three participants mentioned this as a key point in reducing the amount of data from twitter and the first step to find Twitter messages with relevant content. The third element that was missing was a link between the parts of the donut chart on the overview page and the links surrounding the chart. It was expected that clicking on a slice would direct the user to the corresponding page, but this wasn't the case. Next to these three things, one person pointed out to be missing metadata of the photos and videos that were displayed. He mentioned that this is important information for verifying the reliability of the media. Another person mentioned the combination of the filter buttons next to the table of tweets, the figure with tweets per minute and the map. He expected that clicking on a filter would apply that filter to all visualizations on the page. This is not the case in this design, where the filter buttons only apply to the visualization next to the button.

After the exploration time ended, participants had already made a list of suggestions for additional elements of the tool or changing existing elements. Three of these suggestions were mentioned multiple times. The first suggestion was to move the chat panel to a pop-up on the top or bottom of the page. Comparisons were made to the chat function on the Facebook page or to the chat function in several banking apps. The second suggestions was to add specific filter options to the Twitter visualizations regarding witness statements, such as "I saw..." or "I heard..." or "I recognize...". This would speed up the user's search for relevant online witnesses. The third suggestion was to add a sorting option on all data pages that sorts the data on the number of comments that were added to it. Multiple participants mentioned the desire to check the previous work of their colleagues and would therefore like to be able to show that data at the top of the page. A full list of the rest of the suggestions can be found in Appendix I.

8.2 Information-seeking tasks

After the first exploration, participants were given 5 short tasks to find a specific piece of information in the tool. The goal was to observe what participants thought would be the most logical way to reach the sought-after information. This could provide information on the usability and workflow of the tool. The next sections describe the routes that participants chose, and discuss observations that stood out. In all the tasks, the participants started at the page with the data overview.

In the first task, participants had to find the author of the tweet that first mentioned the name 'Tarik'. 8 out of 9 participants (eventually) found that the author of that tweet was 'SuperdaveNL' [44]. The fastest way to get there was to click on the Twitter panel, type in the name 'Tarik' in the input field next to the table, and find that the second tweet contained the requested information. One participant chose the first tweet, which contained the word 'Tarik' in the author's name. Three participants first clicked on the 'Tarik'-filter button next to the 'tweets-per-minute' figure, before changing their attention to the table to find the answer.

The second task was to find the words that were used in the Twitter search. All participants quickly found the search terms in the details of the Twitter search, in the 'Added data' table, but two participants first looked at the top of the Twitter page. Note that most participants found the search terms in the first exploration, although several pointed out that the search terms should also have been displayed at the top of the Twitter page.

In the third task, participants had to find the time at which the most tweets were sent between 20:00 and 21:00 pm. All participants chose '20:22' as their answer and it was found in the 'tweets-per-minute' figure after clicking on the right filter button. Two persons first went to the Timeline page, expecting to find a Twitter timeline in there, but this was not the case.

The fourth task was to find the 'not annotated data' that was added to the SitRap board 'Tarik Z.' 7 out of 9 participants found the information either by clicking directly on the link to the board in the SitRap-boards panel (5 participants) or first turning to the overview page of the SitRap boards and next clicking on the link to the board (2 participants). 2 participants opened the page with the SitRap board, but stopped the task after one minute, claiming they had no idea where to find the requested information.

In the last task, participants were asked to find the person on top of the list of suggestions of Facebook friends that are likely to be friends in the real world. Five participants clicked on the panel to go to the Facebook page, scrolled down to the panel of "potentially related people" and chose "Leo Guo" from the top of the list. Two participants followed the same path but chose "Omar Zahzah", who is on top of the suggested family members and at the top of the panel. Two participants navigated to the page with the profiles and pointed out "Leo Guo", because there is a panel with a picture of him and the subscript "optional friend of suspect". The intended answer was the first of the three, although the other two results seem logical to be found.

It took most of the participants less than a minute to find the requested information (most of the tasks under half a minute). Almost all participants

[44] https://twitter.com/SuperdaveNL/status/560895780547227648

found the requested information. Some routes were taken by all, some routes by a few participants. The overall observation was that the design allowed the participants to quickly find the requested information. The fact that three participants clicked on the 'Tarik'-filter button in the first task is interesting and raises more questions on the filter options on this page.

8.3 Generalization to other cases

Participants were asked what other information they would look for in this case. Multiple participants indicated the desire to find more connections between the Facebook friends, for example by gaining insight in mutual interests and memberships of Facebook groups. Next to that, multiple participants wanted to look for information in news sites and online forums. One person wanted to do more online research on the weapon. This data was not incorporated in the design.

Next, participants were asked which elements should be different in the design for it to be generalizable to all kinds of cases. The design was built upon one example case, so we wanted to know if and how the tool would be applicable to other cases. 7 participants indicated that this design was already quite extensive and quite general, because of the W-questions to which the data can be attached and the option to add data from different kinds of open sources. Points of attention that were indicated for generalization were to add more options to the maps, perhaps implement it with Globetrotter and to provide theme-based set ups for SitRap boards.

8.4 Opinions on specific elements of the design

After the exploration and the information seeking tasks, several general and specific opinions of the design and suggestions were collected. Participants were asked another set of more specific questions regarding the elements of the design to collect more elaborate feedback. The next sections discuss the opinions on specific elements of the design. The answers that were given multiple times are discussed; the individual comments are listed in Appendix I.

Regarding the *colors*, 6 participants mentioned that they were nice and clear. 4 participants indicate that the use of the colors caused a clean layout. 2 participants mentioned that the reddish color of the 'W'-panels drew unnecessary attention to them. The suggestion is made to change their color and use red for alerts and important updates. Although we know from literature that the color red is often saved for information that attention needs to be drawn to immediately, we chose a soft red color in combination with the other three basic colors yellow, green and blue for a calm layout. The fact that two participants suggested to do adapt this color might be an indication that this design choice has to be reconsidered.

Regarding the *"Data overview" page*, participants suggested other options for the placement of the panels. Most participants suggested to put the 'sources' panel and the 'W' panel on top instead of the other three panels that were displayed at the top. Others suggest deleting the 'SitRap-boards' panel since there is another button at the top of the page that links to the boards. Next to that, participants suggest to display updates of new data in case of the tool being implemented with a real-time factor. Another suggestion that might answer to

all the suggestions above is to make the panels draggable. In that way, a user can change the placement of the panels to personal preferences.

7 participants agreed that the *layout of the SitRap board* is fine the way it is designed now. The only suggestion that was mentioned two times is the same as for the 'Data overview' board: make the panels draggable for personal optimization.

The interviewees also commented on the pop-up page where the user can *add selected data* to a SitRap board or attach it to a 'W'-page. Four people indicated it was fine this way and two people suggested a button to share a selected data entry with colleagues immediately. Another suggestion was to add a 'verified' button when the information was verified as reliable by a colleague. One interesting single comment from a participant was that he didn't understand the purpose of this pop-up, because he assumed that data is automatically attached to the 'W'-pages.

7 participants indicated that the *placement of the buttons* was fine the way it was. Suggestions on the buttons were to add a button to export the chat messages and to create different icons for different SitRap boards.

Regarding the *"Timeline" page*, the opinions differed a great deal. Participants were asked their preference for a horizontal or a vertical timeline. Half of the participants preferred the horizontal timeline, because "we read from left to right", it is "better to keep an overview" and "it is the best way to present a timeline in a briefing". The other half of the participants preferred the vertical timeline, because "it is easy to scroll through the data", it is "better to keep an overview" and "it works for extensive research". The answers indicate that participants have different applications in mind for the timeline. The best solution might be, as suggested by two participants, to present the user with different layout options for the timeline, including a horizontal one and a vertical one.

8.5 Application of the Social Media Quick Scan tool

The goal of the last questions (before the questionnaires) was to collect opinions of the potential end users on the application of the tool in their field of work.

The results show that all participants see great potential for this tool, for several reasons. Four people see a perfect application of this tool in a builder of 'open source SitRaps', that could be used at the RTIC and in TGOs[45] and SGBOs[46]. Four people describe the same potential, but see better application in the detective departments. Both groups describe the same characteristics of the tool that make it promising: the tool creates overview of the data, makes order in chaos, supports the user in finding relevant information in the data and is a building platform for making presentations and reports. This great potential is seen under the conditions that the tool will be integrated with other systems that are currently in use.

[45]TGO = Team Grootschalige Opsporing
[46]Staf Grootschalig Bijzonder Optreden

8.6 Results of the questionnaires

8.6.1 First questionnaire

The final part of the interview consisted of two questionnaires. Details on the choice of statements were discussed in Chapter 5. To get an idea of the results on the statements, Table 1 shows a list of the statements and the mean scores (the raw scores). The statements were translated from the original Dutch version, that can be found in Appendix E. Some participants commented on their scoring on specific items. These comments are listed in Appendix I.

Item	Mean score
1. This tool would support me in the detective process	4,45
2. This tool would help me to determine the reliability of information	3,60
3. This tool would help me to determine the relevance of information	3,95
4. This tool would support me in objective reasoning	3,70
5. I think I control the data	4,05
6. I think I have an overview of the data	4,50
7. I think I can easily gain insight in the data	4,25
8. The option to tag data helps me to filter the right information	4,15
9. This tool could support me to reach from data to intelligence	3,95
10. This tool could support me in answering the 7 W-questions	4,05
11. I think I can use relevant data for control after using this tool	4,05
12. After using this tool I can easily share the results with my colleagues	4,05
13. I have clear insight in who of my colleagues added which data	3,40
14. It would be useful if this tool could be used on the streets (on mobile devices)	3,00
15. I would like to have this tool available for use on my mobile devices	3,80
16. I can imagine that this tool supports an earlier use of social media in the investigation process	4,35
17. I would recommend my boss to buy this tool / put this tool into service	3,90

Table 1: Mean score (min 1 - max 5) on the items of the questionnaire

The first four items regarded the research questions. The scores for the items were high. All participants agreed that the detective process is supported with this tool and that the tools supports determining the relevance of information. The support of determining reliability of information and objective reasoning were rated a little lower, but still between 'neutral' and 'agree'.

Items 5 to 8 regarded the general opinion on having insight in the data with which the user is working. The results indicated that participants rated this very positively: The mean results are between 'agree' and 'strongly agree'. This suggests that the visualizations corresponded well with the desired insights.

Items 9 to 11 were statements on the usage of the tool in the intelligence process and detective process. Corresponding to the first statement, this was appreciated as an average 'agree' on the statements.

The divisibility of information in the tool was checked with item 12 and 13. Because of the clear export function, most participants thought it would be easy to share the relevant information after using the tool. Item 13 was rated variably. Several participants mentioned that they could see the work of their colleagues by looking at the names of the people who added data and looking at the comments that their colleagues could have added, while other participants mentioned they had no idea it was an option to collaborate on the boards and that it was unclear to them who had attached which data to the boards.

The opinions differed a great deal on items 14 and 15, which regarded the applicability of the tool on mobile devices. 6 participants stated to have doubts on the usage of this tool on the streets. The most concerns regarded the potential struggles to work with the tool and edit data on a small screen and the fact that there is little time for extensive social media investigation when agents are on the road or on the streets fulfilling other duties. Two people mentioned that this tool would be easier to use on a tablet than on a phone because of the screen size. Five people mentioned that besides their negative opinion on using the tool on the streets with mobile devices, they do see great benefits in the option to see the results of the tool, by being able to view the SitRap report on their mobile devices or receive specific results from colleagues on the phone.

The last set of items regarded the applicability of the tool in their work in general. Item 16 was averagely rated between 'agree' and 'strongly agree'. The last item, regarding the recommendation to buy this tool, was rated with a 4 (agree) or 5 (strongly agree) by all participants, except one. This participant commented that he disagreed for now because he would like to wait for the actual application to be able to judge his real appreciation of the tool. But, as he mentioned, "I'm sure there is great potential in the development of this tool."

8.6.2 SUS questionnaire

The results of the SUS questionnaire showed a mean SUS score of 80,1. An SUS score above 68 is considered as above average usability of a system. Except for one participant, all individual SUS scores were above this average.This means that the results indicate that the participants appreciated the design of the tool in terms of usability.

Part III
Discussion

9 Discussion

The goal of this research was to investigate how visualizations of social media information could support detectives in their work. An interactive prototype design of a visualization tool was developed and evaluated with intended end users. The results are interpreted and discussed in the four sections below, corresponding to the four set research questions. After that, implications for the research field are discussed.

9.1 Visualization of social media information

The first research question was how social media information can best be visualized. One of the reasons that social media was used less than possible in police investigations was an overload of information from social media. Chapter 3 discussed the process of information visualization and different visualization strategies for overview, navigation and interaction. It stressed the importance of letting the insight goals guide the choice of visualizations. The different strategies were applied in the design. The results of the evaluation of the design showed that the chosen visualizations would suit the desired insights in the data. The participants in general experienced overview and control over the data and indicated to be overall satisfied with the provided options for interaction. Therefore we can conclude that the chosen graphs, figures, tables, timelines and maps are generally suitable for visualization of social media information for application at police investigations.

On the Twitter page, the brushing and linking interaction technique was partly applied to the tables and figures, but not to all. Clicking on a filter button on the right of the table does not apply to other tables, for example. This resulted in some confusion in the information-seeking tasks. Future research should determine whether this should be applied to all visualizations on the page or not.

Furthermore, it remains an open question how the connection of different types of data can best be visualized. The concept now entails manual tagging of data that can be attached to one of the seven Ws and a board, and it should somehow be possible to be integrated in a graph of relations. An example of such a graph was embedded as a static picture. It is an example of a graph with different types of data from a staff member at the University of Groningen and was implemented by Pure[47]. Further research is needed on how the tagged data can easily be visualized in a relatively simple graph.

9.2 Objective reasoning and determining relevance & reliability of information

The second research question was how objective reasoning could be supported in the visualizations and following from that question: how can the relevance

[47] https://www.elsevier.com/solutions/pure

and reliability of the content best be validated?

Results from the interviews and theory showed that objective reasoning can best be supported by interpreting data as less as possible. Information was determined as more reliable when multiple sources provided the same content or multiple messages with the same content were found in the same source, and when the source was verified as reliable. This could also best be supported by providing the users with as much context information as possible. This was implemented in the design by displaying the source of every piece of information, and metadata if it was available. Whenever possible, links were provided to direct the user to the original context of the content. These design decisions were positively appreciated.

An important notion is that a visualization tool can support objective reasoning to a limited extent, because the main reasoning and decision making will eventually remain at the user. Whether information is relevant depends on the decision making and on the case. A list of types of potential relevant information in social media investigations was collected from the interviews. The design aimed at providing the user options to gain insight in these types of information. These results from the evaluation indicated that the concept of the tool could support determining relevance of information. Functionalities that were indicated as supportive for this were the different insight options, the option to add data from different sources, and the option to attach data to the W-questions and hence give some meaning to the data.

Future challenges lie in the integration of intelligent algorithms that on the one hand automatically connect data and suggest connections and interpretations, but on the other hand keep supporting the user in objective reasoning. An interesting field in development is uncertainty visualization, that specializes in displaying certainty and uncertainty in such a way that the interpretation remains fully at the side of the user [Bisantz et al., 2009]. Furthermore, the integration with intelligent algorithms could add towards predictive policing, that crime events can be predicted based on data from previous crimes.

9.3 Application for users with different expertise levels

The last research question was how the visualizations could be accessible for people with different technical expertise levels.

As mentioned in the introduction and in chapter 4, one of the main pitfalls of existing visualization tools was that these tools are aimed at extensive time-consuming analyses and/or require technical (and programming) skills to be able to work with them. This research aimed at providing a *quick* scan of social media information that may be used by people at different departments of the police, not only the social media experts. Therefore, existing visualization tools were embedded and the layout of the design was chosen to be simple and clear. The results of the evaluation supported these design decisions. It was mentioned as one of the strong points of the design, and the words “clear”, “overview” and “easy”, were mentioned several times.

Future usability evaluations would gain better validity when more subjects are evaluated and when the interviewer would not be the same person as the person who programmed the tool.

9.4 Challenges for future development

This research ends at a first complete loop of the interaction lifecycle (Figure 22). To further implement the tool, the lifecycle has to be run through several times, evaluating the implementations with the end users in each phase. The end product of this study was a prototype and a qualitative evaluation. From the evaluation, several design suggestions for enhancements of the visualizations were already discussed in the previous sections and in Appendix I.

Regarding the technical functionalities, future work should focus on setting up a database on the server-side of the web-application and on privacy and security of the data. For police investigations, it is important that the data is recorded (before someone deletes an online message for example) and secured for evidence. Furthermore, a challenge would be to integrate the open source data with data from police sources. Ideally, the tool would be able to combine all available relevant data, also non-open source data.

Regarding the concept of the tool, challenges remain in deciding whether the tool should add data real-time. The design was set up statically, assuming given data sets that included recorded data from social media sources. A potential disadvantage of this is that relevant information can be missed that becomes available after the data collection. This could be solved by adding data real time, for example through the use of web crawlers. However, if the main focus of the tool remains on providing a quick scan of social media data and use this data for a report that can be used early in an investigation process, there might be no need for real-time data adding, and social media monitoring and social media investigation can remain separated.

9.5 Impact on the research field

This project adds to the research on developing visualization software. Some new insights have been developed, such as what visualization concepts are suitable for integrating multiple sources. Furthermore, we know have more knowledge on how visualization tools can be accessible to users with different technical backgrounds, especially non-programmers. With this knowledge, the usability of future visualization tools can be enhanced.

Regarding police investigations, more knowledge was gained on information overload in the police domain, the use of social media in police investigations and what technology is needed to support this. High potential is seen for this tool to support detectives in their work. Furthermore, applications are seen at the Real-Time Intelligence Center. Some of the solutions to the suboptimal use of social media information lie in the communication between these two departments. With this tool, a bridge can be built across a quick social media scan at the RTIC and detectives working on a police investigation, when the RTIC would share the interactive SitRap (created with the tool) to a detective that can have earlier access to relevant information on the streets.

10 Conclusion

This thesis resulted in the first big steps towards a Social Media Quick Scan. The encountered problems with social media usage in the detection of crimes were investigated and translated to requirements for a visualization tool and a design and prototype were developed and positively evaluated. Great potential is seen in the further development of this tool and application at several police departments.

We hope that this project could contribute to the process towards supporting detectives to use of available social media information earlier in police investigations.

Acknowledgements

First, I want to thank my supervisor Fokie Cnossen for her support and feedback. Throughout the project, she continuously stimulated me to think and write at a high scientific level.

I would also like to thank Arnout de Vries, my supervisor from TNO. He facilitated every contact I needed and was a great inspiration. In every conversation we had, he motivated me by always focusing on the need for this tool and the potential impact on the work of the end users.

I would like to thank the employees from the Dutch police for their time and feedback. Thank you for the pleasant conversations and all the cups of coffee.

I am grateful for all the mental support from my family and friends during this project. Special thanks go to my fiancé Pim. Thank you for standing next to me at every stage of the project, patiently listening to all my analyses and endless thoughts.

Promotional video

In collaboration with TNO, a short YouTube video was made to show the intended functionalities and usage of the Social Media Quick Scan tool. The video can be found at https://www.youtube.com/watch?v=Vn9uOuqW8L8

11 References

References

[Abel et al., 2012] Abel, F., Hauff, C., Houben, G., Tao, K., and Stronkman, R. (2012). Twitcident: fighting fire with information from social web streams. *Proceedings of the 21st International Conference on World Wide Web.*, pages 305–308.

[(ACPO), 2006] (ACPO), A. o. C. P. O. (2006). *Murder Investigation Manual.* Centrex, Bedfordshire.

[Ahlberg and Shneiderman, 1994] Ahlberg, C. and Shneiderman, B. (1994). Visual Information Seeking: Tight Coupling of Dynamic Query Filters with Starfield Displays. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '94*, pages 313–317.

[Ahlberg and Wistrand, 1995] Ahlberg, C. and Wistrand, E. (1995). IVEE: An Information Visualizationand Exploration Environment. In *Proceedings of the IEEE Symposium on Information Visualization*, pages 66–73, Washington DC. IEEE Cmoputer Society.

[Ahn et al., 2011] Ahn, J. W., Taieb-Maimon, M., Sopan, A., Plaisant, C., and Shneiderman, B. (2011). Temporal visualization of social network dynamics: Prototypes for nation of neighbors. *Lecture Notes in Computer Science*, pages 309–316.

[Artacho-Ramírez et al., 2008] Artacho-Ramírez, M. A., Diego-Mas, J. A., and Alcaide-Marzal, J. (2008). Influence of the mode of graphical representation on the perception of product aesthetic and emotional features: An exploratory study. *International Journal of Industrial Ergonomics*, 38:942–952.

[Baldonado et al., 2000] Baldonado, M. Q., Woodruff, A., and Kuchinsky, A. (2000). Guidelines for using multiple views in information visualization. *Proceedings of the working conference on Advanced visual interfaces (AVI)*, pages 110–119.

[Bastian and Heymann, 2009] Bastian, M. and Heymann, S. (2009). Gephi : An Open Source Software for Exploring and Manipulating Networks. *ICWSM*, 8:361–362.

[Bederson, 2000] Bederson, B. B. (2000). Fisheye menus. *Proceedings of the 13th annual ACM symposium on User interface software and technology - UIST '00*, pages 217–225.

[Bertin, 1983] Bertin, J. (1983). *Semiology of graphics: Diagrams, networks, maps.* The University of Wisconsin Press, Madison.

[Bisantz et al., 2009] Bisantz, A. M., Stone, R. T., Pfautz, J., Fouse, A., Farry, M., Roth, E., Nagy, A. L., and Thomas, G. (2009). Visual Representations of Meta-Information. *Journal of Cognitive Engineering and Decision Making*, 3(1):67–91.

[Boertjes et al., 2011] Boertjes, E., Kotterink, B., and Jager, E. (2011). Visualization of Social Networks —TNO report 35510. Technical report, TNO.

[Bos, 2013] Bos, J. (2013). Nationale Politie op koers? Technical report.

[Bostock and Heer, 2009] Bostock, M. and Heer, J. (2009). Protovis : A Graphical Toolkit for Visualization. *IEEE transactions on visualization and computer graphics*.

[Brooke, 1996] Brooke, J. (1996). SUS: A "quick and dirty" usability scale. In Jordan, P., Thomas, B., Weerdmeester, B., and McClelland, A., editors, *Usability Evaluation in Industry*. Taylor and Francis, London.

[Cao et al., 2012] Cao, N., Lin, Y. R., Sun, X., Lazer, D., Liu, S., and Qu, H. (2012). Whisper: Tracing the spatiotemporal process of information diffusion in real time. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2649–2658.

[Card et al., 1999] Card, S., Mackinlay, J., and Shneiderman, B. (1999). *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann, San Francisco.

[CBS, 2015] CBS (2015). *ICT: kennis en economie 2015*. Centraal Bureau voor de Statistiek, Den Haag.

[CBS, 2016] CBS (2016). *ICT: kennis en economie 2016*. Centraal Bureau voor de Statistiek, Den Haag.

[Cui et al., 2008] Cui, W., Zhou, H., Qu, H., Wong, P. C., and Li, X. (2008). Geometry-based edge clustering for graph visualization. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1277–1284.

[De Poot, 2011] De Poot, C. (2011). Wetenschap op de plaats delict. Technical report, Hogeschool van Amsterdam / Politieacademie.

[De Vries, 2013] De Vries, A. (2013). Informatie die stuurt! - Resultaten 2013. Technical report, TNO.

[De Vries and De Groen, 2012] De Vries, A. and De Groen, L. (2012). Informatie aan het werk! - Resultaten 2012. Technical report, TNO.

[De Vries and Smilda, 2014a] De Vries, A. and Smilda, F. (2014a). De online (r)evolutie. In *Social media: het nieuwe DNA*, chapter 2, page 50. Reed Business Education, 1st edition.

[De Vries and Smilda, 2014b] De Vries, A. and Smilda, F. (2014b). Impact op de 8 W's van opsporing. In *Social media: het nieuwe DNA*, chapter 3. Reed Business Education, Amsterdam, 1st edition.

[De Vries and Smilda, 2014c] De Vries, A. and Smilda, F. (2014c). Opsporingsmethodieken en social media. In *Social media: het nieuwe DNA*, chapter 4. Reed Business Education, Amsterdam, 1st edition.

[Dörk et al., 2012] Dörk, M., Carpendale, S., and Williamson, C. (2012). Fluid Views: a zoomable search environment. *Proceedings of the International Working Conference on Advanced Visual Interfaces*, pages 233–240.

[Fekete, 2004] Fekete, J.-d. (2004). The InfoVis Toolkit. *IEEE Symposium on Information Visualization.*, pages 167–174.

[Fischer, 2012] Fischer, G. (2012). Context-Aware Systems: The 'Right' Information, at the 'Right' Time, in the 'Right' Place, in the 'Right' Way, to the 'Right' Person. *Proceedings of the International Working Conference on Advanced Visual Interfaces - AVI '12*, pages 287–294.

[Fischer and Herrmann, 2011] Fischer, G. and Herrmann, T. (2011). Socio-Technical Systems: A Meta-Design Perspective. *International Journal of Sociotechnology and Knowledge Development*, 3(1):1–33.

[Furnas, 1986] Furnas, G. W. (1986). Generalized fisheye views. *ACM Conference on Human Factors in Computing Systems*, pages 16–23.

[Gomez-rodriguez et al., 2009] Gomez-rodriguez, M., Gummadi, K. P., and Sch, B. (2009). Quantifying Information Overload in Social Media and Its Impact on Social Contagions. pages 170–179.

[Gorissen and Johannink, 2016] Gorissen, I. and Johannink, R. (2016). Monitoring En Analyse Informatie Op Sociale En Online Media. Technical report.

[Groen and Rijgersberg, 2012] Groen, L. D. and Rijgersberg, T. (2012). Informatie-overload in de praktijk. Technical report, TNO.

[Gross, 1964] Gross, B. (1964). *The managing of organizations: The administrative struggle.* Free Press of Glencoe, New York.

[Gross, 1893] Gross, H. (1893). *Handbuch für Untersuchungsrichter als System der Kriminalistik.* J. Schweizer Verlag, Berlin.

[Gross and Van der Does de Willebois, 1904] Gross, H. and Van der Does de Willebois, W. (1904). *De nasporing van het strafbaar feit: een leiddraad voor ambtenaren en beambten van justitie en politie.* Veerman, Heusden.

[Hansen et al., 2010] Hansen, D., Dunne, C., and Shneiderman, B. (2010). Analyzing Social Media Networks with NodeXL. In *Human-Computer Interaction Lab 27th Annual Symposium*, pages 2–3.

[Heer and Boyd, 2005] Heer, J. and Boyd, D. (2005). Vizster: Visualizing online social networks. *Proceedings - IEEE Symposium on Information Visualization, INFO VIS*, pages 33–40.

[Heer et al., 2005] Heer, J., Card, S. K., and Landay, J. A. (2005). Prefuse: a Toolkit for Interactive Information Visualization. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '05*, pages 421–430.

[Heer et al., 2008] Heer, J., Mackinlay, J. D., Stolte, C., and Agrawala, M. (2008). Graphical histories for visualization: Supporting analysis, communication, and evaluation. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1189–1196.

[Hurter et al., 2012] Hurter, C., Ersoy, O., Telea, A., Hurter, C., Ersoy, O., and Telea, A. (2012). Graph bundling by Kernel Density Estimation. *EUROVIS 2012, Eurographics Conference on Visualization*, 31(3):865–874.

[International Association of Chiefs of Police, 2015] International Association of Chiefs of Police (2015). 2015 Social Media Survey Results.

[Jong and Dückers, 2016] Jong, W. and Dückers, M. L. A. (2016). Self-correcting mechanisms and echo-effects in social media: An analysis of the "gunman in the newsroom" crisis. *Computers in Human Behavior*, 59:334–341.

[Kaptein et al., 2015] Kaptein, N., Kaouass, A., and Zannoni, M. (2015). Het nieuws gegijzeld. Technical report, COT.

[Kaptein et al., 2014] Kaptein, R., Koot, G., and Huis, M. A. A. (2014). Needle Custom Search Recall-Oriented Search on the Web Using Semantic Annotations. In *Advances in information retrieval*, pages 750–753. Springer International Publishing.

[Kaptein et al., 2013] Kaptein, R., Van den Broek, E. L., Koot, G., and a.a. Huis in 't Veld, M. (2013). Recall oriented search on the web using semantic annotations. *Proceedings of the sixth international workshop on Exploiting semantic annotations in information retrieval - ESAIR '13*, pages 45–48.

[Keim et al., 2008] Keim, D. a., Mansmann, F., and Ziegler, H. (2008). Challenges in Visual Data Analysis. *Information Visualization IV*, pages 9–16.

[Kop and Klerks, 2009] Kop, N. and Klerks, P. (2009). *Intelligencegestuurd politiewerk*. Apeldoorn: Politieacademie.

[Kop et al., 2012] Kop, N., Van der Wal, R., and Snel, G. (2012). *Opsporing Belicht*. Boom Lemma Uitgevers, Den Haag, 2nd edition.

[Liu et al., 2014] Liu, S., Cui, W., Wu, Y., and Liu, M. (2014). A Survey on Information Visualization: Recent Advances and Challenges. *The Visual Computer*, pages 1–21.

[Manovich, 2011] Manovich, L. (2011). What is visualisation? *Visual Studies*, 26(1):36–49.

[McGloin and Kirk, 2010] McGloin, J. M. and Kirk, D. S. (2010). An Overview of Social Network Analysis. *Journal of Criminal Justice Education*, 21(2):169–181.

[Meesters and Niemeijer, 2000] Meesters, P. and Niemeijer, B. (2000). Criminaliteitsbeeldanalyse: Problemen en mogelijkheden. *Criminaliteitsanalyse in Nederland*, pages 293–306.

[Milford and Perry, 1977] Milford, J. and Perry, R. (1977). A methodological study of overload. *Journal of General Psychology*, (97):131–137.

[Miller, 2013] Miller, J. (2013). Graph Database Applications and Concepts with Neo4. *Southern Association for Information Systems Conference*, pages 141–147.

[Nielsen, 1992] Nielsen, J. (1992). Finding Usability Problems through Heuristic Evaluation. In *Proceedings of CHI'92*, pages 373–800.

[Nielsen, 1994] Nielsen, J. (1994). Enhancing the explanatory power of usability heuristics. *CHI '94: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 152–158.

[Norman, 2002] Norman, D. (2002). *The Design of Everyday Things.* Basic Books.

[North, 2012] North, C. (2012). Information Visualization. In Salvendy, G., editor, *Handbook of Human Factors and Ergonomics*, chapter 43, pages 1209–1236. John Wiley & Sons, 4th edition.

[North et al., 2002] North, C., Conklin, N., and Saini, V. (2002). Visualization schemas for flexible information visualization. *IEEE Symposium on Information Visualization*, pages 15–22.

[Pariser, 2011] Pariser, E. (2011). *The Filter Bubble: What the internet is hiding from you.* Penguin UK.

[Perer and Shneiderman, 2008] Perer, A. and Shneiderman, B. (2008). Integrating Statistics and Visualization: Case Studies of Gaining Clarity during Exploratory Data Analysis. *CHI, Human Factors in Computing Systems*, page 265.

[Pfautz et al., 2005] Pfautz, J., Fouse, a., Fichtl, T., Roth, E., Bisantz, a., and Madden, S. (2005). The Impact of Meta-Information on Decision-Making in Intelligence Operations. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 49(3):214–218.

[Plaisant, 2005] Plaisant, C. (2005). Information visualization and the challenge of universal access. *Exploring Geovisualization*, pages 53–82.

[Politie, 2013] Politie (2013). Sociale media vast onderdeel van politiewerk. https://www.politie.nl/nieuws/2013/maart/27/00-sociale-media-vast-onderdeel-van-politiewerk.html. Date accessed: 2017-07-25.

[Pouchard et al., 2009] Pouchard, L. C., Dobson, J. M., and Trien, J. P. (2009). A Framework for the Systematic Collection of Open Source Intelligence. *Association for the Advancement of Artificial Intelligence (www.aaai.org)*, pages 102–107.

[Ratcliffe, 2003] Ratcliffe, J. (2003). Intelligence-led policing. *Australian Institute of Criminology.*

[Rogers et al., 2011] Rogers, Y., Sharp, H., and Preece, J. (2011). *Interaction Design: Beyond Human-computer Interaction.* John Wiley & Sons.

[Salvendy, 2012] Salvendy, G. (2012). *Handbook of Human Factors an Ergonomics.* John Wiley & Sons, Hoboken, New Jersey, 3rd edition.

[Shneiderman, 1996] Shneiderman, B. (1996). The Eyes Have It : A task by data type taxonomy for information visualizations. In *IEEE Symposium on Visual Languages*, pages 336–343.

[Silverman, 2014] Silverman, C. (2014). *Verification Handbook: An ultimate guide on digital age sourcing for emergency coverage.* European Journalism Centre.

[Simon, 1996] Simon, H. (1996). *The Sciences of the Artificial.* The MIT Press, Cambridge, 3rd edition.

[Siricharoen, 2013] Siricharoen, W. V. (2013). Infogragraphics: The New communication Tools in Digital Age. In *The International Conference on E-Technologies and Business on the Web*, pages 169–174.

[Snel and Tops, 2011] Snel, G. and Tops, P. (2011). *Een wereld te winnen.* Politieacademie, Apeldoorn.

[Sommerville, 2004] Sommerville, I. (2004). *Software Engineering.* Pearson Education, Harlow, England, 7th edition.

[Spath et al., 2012] Spath, D., Hermann, F., Peissner, M., and Sproll, S. (2012). User Requirements Collection and Analysis. In *Handbook of Human Factors and Ergonomics*, pages 1313 – 1322.

[Spence, 2001] Spence, R. (2001). *Information visualization.* Springer, New York: Addison-Wesley.

[Strötgen and Gertz, 2013] Strötgen, J. and Gertz, M. (2013). Multilingual and cross-domain temporal tagging. *Language Resources and Evaluation*, 47(2):269–298.

[Tao et al., 2011] Tao, K., Abel, F., Hauff, C., and Houben, G.-j. (2011). Twinder : A Search Engine for Twitter Streams.

[Toffler, 1984] Toffler, A. (1984). *Future Shock.* Random House, New York.

[Tufte and Graves-Morris, 1983] Tufte, E. R. and Graves-Morris, P. R. (1983). *The visual display of quantitative information.* CT: Graphics press, Cheshire, 2nd edition.

[van Berlo, 2012] van Berlo, D. (2012). *Wij, de overheid. Cocreatie in de netwerksamenleving.* Programma Ambtenaar 2.0.

[Ware, 2004] Ware, C. (2004). *Information visualization: Perception for Design.* Morgan Kaufmann, San Francisco.

[Ware and Lewis, 1995] Ware, C. and Lewis, M. (1995). The DragMag Image Magnifier. *Proceedings of the 13th annual ACM conference on Human factors in computing systems (CHI '95)*, pages 407–408.

[Wickens and Hollands, 2000] Wickens, C. D. and Hollands, J. G. (2000). *Engineering Psychology and Human Performance.* Prentice Hall, Upper Saddel River, New Jersey, 3rd edition.

[Zinsmaier et al., 2012] Zinsmaier, M., Brandes, U., Deussen, O., and Strobelt, H. (2012). Interactive level-of-detail rendering of large graphs. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2486–2495.

A Twitter example

Figure 34 shows an example of a tweet and what information can be found in it. Table 2 shows an example of twitter data in a table structure.

Figure 34: Example of a tweet

Date + Time	**URL**	**Author**	**Followers**	**Message**
29-1-2015 20:01	https://twitter.com/muisje_eliza/status/560875508481929219	muisje_eliza	3492	Testbeeld met de hoogste kijkcijfers ooit

Table 2: Example of Twitter data structured in a table

B Facebook example

Figure 35 shows an example of a Facebook page and what information can be found on it. Figure 36 shows an example of data extracted from the Facebook page. A selection of the list of Facebook friends is shown.

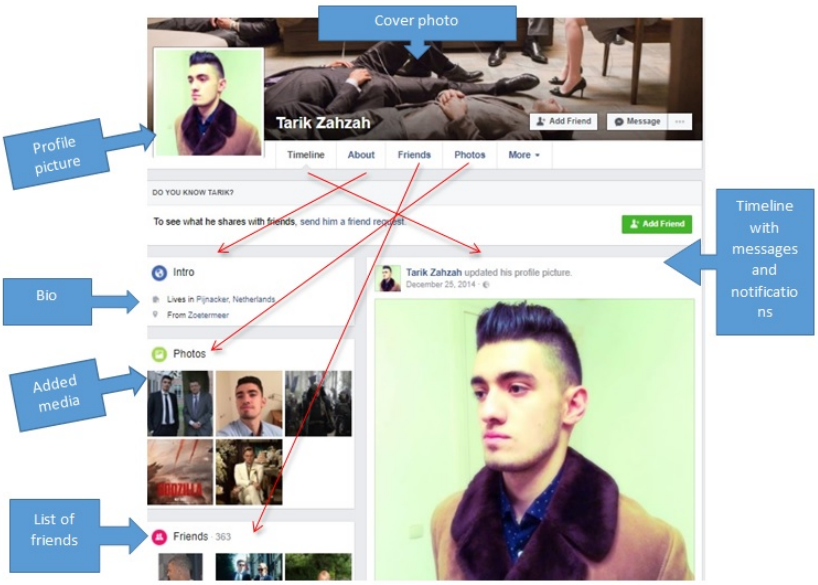


Figure 35: Example of a Facebook page

Object Type	Label	Identification Number	Name / Alias	URL	Online Identifier	Social Media Profile Description	Link: Friend of
Facebook Profile	Frederik Weulen Kranenberg	USER ID 1665722147 FACEBO	Frederik Weulen Kranenberg	https://www.facebook.com/1665722147	1665722147	Universiteit Leiden / Leiden University	tarik.zahzah (Facebook Profile)
Facebook Profile	jelle.slump	USER ID 100001510230485 FA	Jelle Slump	https://www.facebook.com/jelle.slump	100001510230485, jelle.slump		tarik.zahzah (Facebook Profile)
Facebook Profile	yarno.wortel	USER ID 100002194075745 FA	Yarno Wortel	https://www.facebook.com/yarno.wortel	100002194075745, yarno.wortel		tarik.zahzah (Facebook Profile)
Facebook Profile	bob.kallman.7	USER ID 1163527205 FACEBO	Bob Kallman	https://www.facebook.com/bob.kallman.7	1163527205, bob.kallman.7	Works At Leiden University	tarik.zahzah (Facebook Profile)
Facebook Profile	martijn.vandergoes	USER ID 100002514570507 FA	Martijn van der Goes	https://www.facebook.com/martijn.vandergoes	100002514570507, martijn.vandergoes	Leiden University	tarik.zahzah (Facebook Profile)
Facebook Profile	ward.dejong.3	USER ID 100002326607358 FA	Ward de Jong	https://www.facebook.com/ward.dejong.3	100002326607358, ward.dejong.3	TU Delft	tarik.zahzah (Facebook Profile)
Facebook Profile	Sven Wijngaarden	USER ID 100006377385304 FA	Sven Wijngaarden	https://www.facebook.com/100006377385304	100006377385304	Leiden University	tarik.zahzah (Facebook Profile)

Figure 36: Example of data of Facebook friends

C List of interviewed people

C.1 Identification of needs (first set of interviews)

Function	**Number of interviewees**	**Sex**
RTIC-centralist	4	2 male(m) + 2 female
Specialist Open Sources	1	m
Senior Cop Intelligence / Specialist Open Sources	2	m
Head of Business Intelligence & Quality	1	m
Tactical Detective Force	1	m
Tactical Detective Force & Operational Specialist Digital Forensics	1	m
Open Source Intelligence (OSINT) & Operational Specialist Digital Forensics	1	m

Table 3: Overview of interviewees in the first set of interviews.

C.2 Evaluation of the prototype (second set of interviews)

Function	**Number of interviewees**	**Sex**
RTIC-centralist	3	m
Specialist Open Sources	1	m
Senior Cop Intelligence / Specialist Open Sources	2	m
Tactical Detective Force	1	m
Tactical Detective Force & Operational Specialist Digital Forensics	1	m
Open Source Intelligence (OSINT) & Operational Specialist Digital Forensics	1	m

Table 4: Overview of interviewees in the second set of interviews.

D First Interview protocol

Inleiding / Algemeen

- Wat is uw functie?
- Hoe zien uw dagelijkse werkzaamheden eruit?
- Hoe lang werkt u hier, wat is uw achtergrond van daarvoor en hoe zou u uw ervaringsniveau inschatten als het gaat om onderzoek op social media?
- Wat is volgens u de huidige situatie in de opsporing als het gaat om het gebruik van sociale media bij onderzoek?
- Zou daar iets in veranderd/verbeterd moeten worden? Zo ja, wat?

Deelvraag 1 (social media + visualisatie)

- Hoe maakt u gebruik van social media als informatiebron?
- Welke informatie is vaak het meest relevant? Wat voor informatie gaat u naar op zoek?
- Kunt u mij omschrijven hoe het proces bij u in zijn werk gaat als u in het gouden uur een SitRap gaat opstellen? Wat zijn de globale stappen in dit proces? [alleen van toepassing bij RTIC-medewerkers]
- Welke bronnen denkt u dat het meest relevant zijn? (social media)
- Kunt u een delicttype aanwijzen waar social media als informatiebron vaak relevante informatie oplevert of op zou kunnen leveren?
- Hoe denkt u dat visualisatie van deze informatie zou kunnen ondersteunen in de opsporing?
- Wat zouden voor u hierbij de belangrijke aandachtspunten zijn?
- Hoe ziet u de rol van de computer als intelligent systeem dat informatie kan analyseren en suggesties voor verbanden kan geven?

Deelvraag 2 (tunnelvisie)

- Hoe wordt objectief redeneren (W van Wetenschap) het beste gewaarborgd volgens u?
- Zijn er op dit punt specifieke aandachspunten bij social media? Zo ja, welke?
- Hoe brengt u dat in de praktijk? D.w.z.: Hoe maakt u nu inzichtelijk hoe/waar/op welke manier de informatie op internet gevonden is?
- Wat zou visualisatie van social media informatie voor eigenschappen moeten hebben om dit te kunnen waarborgen?
- Waar ziet u risico's?

Deelvraag 3 (UGC)

- Veel informatie op social media is gebruiker-gegenereerd (user-generated content). Hoe wordt volgens u het beste de relevantie en betrouwbaarheid van deze informatie bepaald?
- Wat zou een visualisatie nodig hebben om inzicht te kunnen geven in de relevantie en betrouwbaardheit van gebruiker-gegenereerde informatie (user-generated content).
- Hiervoor de zelfde vraag als eerder: Hoe ziet u de rol van de computer (bijv. Google) die betrouwbaarheid kan aangeven versus de mens die dat zelf inschat? Wat zou voor u de optimale verhouding zijn?

Deelvraag 4 (usability/eindgebruikers)

- Wie zouden er volgens u allemaal baat hebben bij een visualisatie van social media informatie?
- Wie ziet u als mogelijke concrete eindgebruikers?
- Wat is er nodig om het mogelijk te maken dat een visualisatie hun werk goed zou kunnen ondersteunen en niet tegenwerken?
- Wat zijn volgens u aandachtspunten als het gaat om gebruiksvriendelijkheid van de visualisatie voor alle mensen uit de doelgroep?

Algemeen / afsluiting

- Wat wilt u zelf nog toevoegen?

E Second interview protocol

Algemeen

U krijgt nu 5-10 minuten de tijd om met het ontwerp te spelen. Let op: Sommige pagina's hebben even de tijd nodig om te laden.

- Wat vindt u van wat u ziet en tegenkomt?
- Wat zou u nu graag willen zien? Naar welke informatie zou u nu op zoek gaan? Waar zou u nu inzicht in willen krijgen?
- Wilt u hardop vertellen wat u denkt dat de tool zou doen als u ergens op klikt?

Opdrachten

De volgende vragen zijn kleine opdrachten om te kijken hoe het opzoeken van specifieke informatie zou gaan in dit ontwerp. Ik observeer de weg die de gebruiker verkiest om bij de gevraagde informatie te komen. De inhoud is hierbij niet van belang.

1. Wie verstuurde de eerste tweet waarin de naam 'Tarik' genoemd werd?
2. Vindt de zoektermen die gebruikt zijn om de tweets binnen te halen.
3. Op welke tijd tussen 20:00 en 21:00 zijn de meeste tweets verstuurd?
4. Welke niet-geannoteerde is toegevoegd aan het bord over 'Tarik Z'?
5. Vindt de persoon bovenaan de suggesties van Facebook-vrienden die Tarik in het echt zouden kunnen kennen.

Specifieke vragen over de zaak en generalisatie

- Welke informatie mist u nog? / Wat zou u verder nog willen opzoeken van deze zaak?
- Welke onderdelen zouden er anders moeten uitzien bij andere soorten delicten?
- Hoe kan dit ontwerp volgens u het beste gegeneraliseerd worden naar andere soorten zaken?

Over specifieke elementen van het ontwerp

- Wat vindt u van de kleuren van het ontwerp? Wat zou u anders doen?
- Wat vindt u van de indeling van het data-overzicht? Wat zou u anders doen?
- Wat vindt u van het SitRap-bord en hoe het er nu uitziet? Wat zou u anders doen?
- Wat vindt u van de mogelijkheden die u aan kunt geven bij het toevoegen van data aan een bord? Wat zou u anders doen?
- Wat vindt u van de plaatsing van de knoppen? Wat zou u anders doen?

- Wat vindt u van de horizontale en verticale tijdlijn? Waar zou u het voor gebruiken? En heeft u een voorkeur voor een van de twee?
- Ruimte voor eigen opmerkingen over specifieke elementen van het ontwerp

Over de inzicht-doelen

Vul in wat u van de volgende stellingen vindt (1=totaal mee oneens - 5=totaal mee eens).

- Deze tool zou me ondersteunen bij het rechercheproces.
- Deze tool zou me helpen om betrouwbaarheid van informatie te bepalen.
- Deze tool zou me helpen om relevantie van informatie te bepalen.
- Deze tool zou me helpen om objectief te redeneren over de informatie die ik zie.
- Ik heb het idee dat ik controle heb over de data.
- Ik heb het idee dat ik de data kan overzien.
- Ik heb het idee dat ik gemakkelijk inzicht krijg in de data.
- Deze tool kan me ondersteunen bij het komen van data tot intelligence.
- Deze tool kan me ondersteunen bij het beantwoorden van de gouden W's.
- Ik denk dat ik na het gebruiken van deze tool relevante data kan gebruiken als sturingsinformatie.
- Na het gebruiken van deze tool kan ik de gevonden informatie makkelijk delen met mijn collega's.
- Ik kan duidelijk zien wie van mijn collega's wat heeft toegevoegd.
- De mogelijkheid om data te annoteren helpt me om de juiste informatie te filteren.
- Ik kan me voorstellen dat deze tool bij kan dragen aan het eerder gebruik maken van social media informatie in politie-onderzoeken.
- Het zou handig zijn als deze tool ook op straat gebruikt kan worden (mobiele apparaten).
- Ik zou deze tool wel beschikbaar willen hebben op mijn mobiele apparaten
- Ik zou mijn baas aanraden dit te kopen/in gebruik te nemen.

SUS-vragenlijst

Vul in wat u van de volgende stellingen vindt (1=totaal mee oneens - 5=totaal mee eens).

- Ik denk dat ik dit systeem graag regelmatig wil gebruiken.
- Ik vond het systeem onnodig complex.
- Ik vond het systeem makkelijk te gebruiken.

- Ik denk dat ik ondersteuning nodig heb van een technisch persoon om dit systeem te kunnen gebruiken.
- Ik vond dat de verschillende functies in dit systeem erg goed geintegreerd zijn.
- Ik vond dat er teveel tegenstrijdigheden in het systeem zaten.
- Ik kan me voorstellen dat de meeste mensen zeer snel leren om dit systeem te gebruiken.
- Ik vond het systeem erg omslachtig in gebruik.
- Ik voelde me erg vertrouwd met het systeem.
- Ik moest erg veel leren voordat ik aan de gang kon gaan met dit systeem.

Afsluiting

- Hoe ziet u in de toekomst het gebruik van een SocialMedia QuickScan tool voor u?
- Hoe ziet u in de toekomst het gebruik van social media in de opsporing voor u?

F List of use cases

Login

1. The user loads the system on the web page
2. The system gives input fields
3. The user enters username and password and presses the 'Login' button
4. The system checks the username and password
5. The system loads the overview-page

4A1 Alternative: When the username and/or password are incorrect at step 4:

1. The system displays an error message
2. The system returns to step 2 of the login process.

Collect data

1. The user indicates he wants to add data
2. The system displays a pop-up with the input fields and options
3. The user enters search terms, and determines sources and date-range.
4. The system checks if the search terms are entered correctly
5. The system collects the requested data
6. The system loads the overview-page including the added data

2A1 alternative when the user prefers to add data manually instead of starting a new search:

1. The system displays a pop-up with options for uploading data
2. The system load the attached data
3. The system returns to step 6 of the collect-data process

4A1 alternative when the search terms are not entered correctly:

1. The system displays an error message
2. The system returns to step 2 of the collect-data process

Open case

1. The user indicates to open an existing case
2. The system displays a pop-up with available cases
3. The user selects a case
4. The system loads the overview-page of the case

Collect relevant information

1. The user explores the data by interacting with the system in one of the following ways:
   (a) View data
   (b) Filter data
   (c) Select data
   (d) Create overview
   (e) Visualize data

Export information

1. The user indicates he wants to export the selected data
2. The system displays the export options
3. The user selects the extension of preference
4. The system executes the selection way of export

G Screenshots of the wireframes

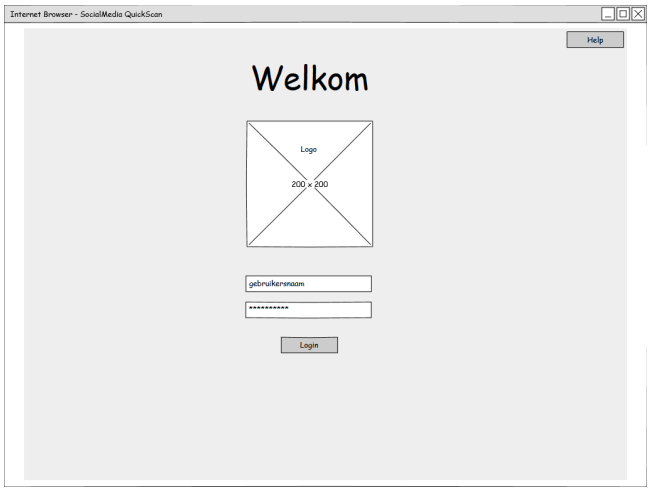


Figure 37: Wireframe of the Login page

Figure 38: Wireframe of the data overview page

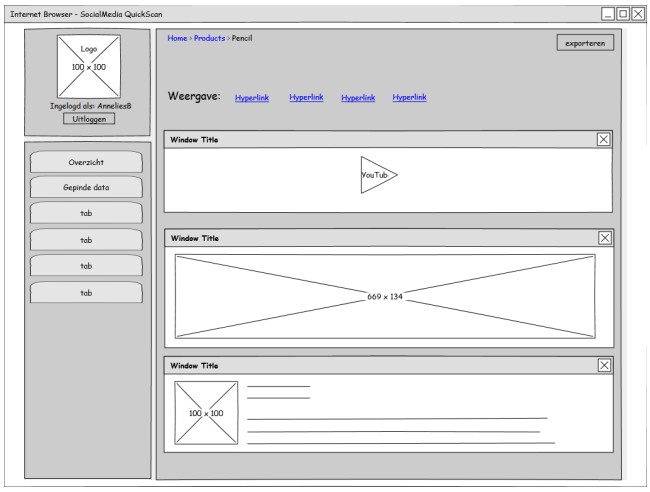


Figure 39: Wireframe of the 'pinned data' page

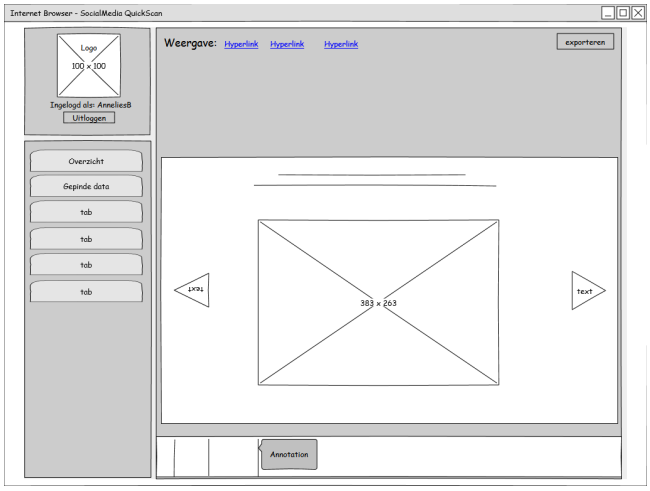


Figure 40: Wireframe of the timeline page

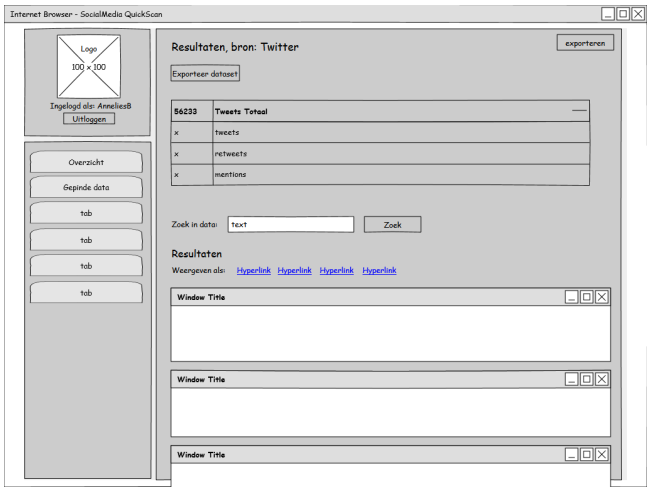


Figure 41: Wireframe of the page with Twitter results and panels with space for visualizations.

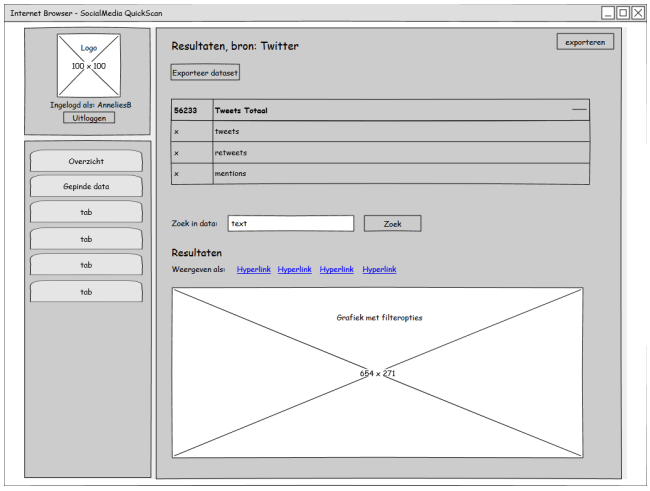


Figure 42: Wireframe of the page with Twitter results with space for a large visualization.

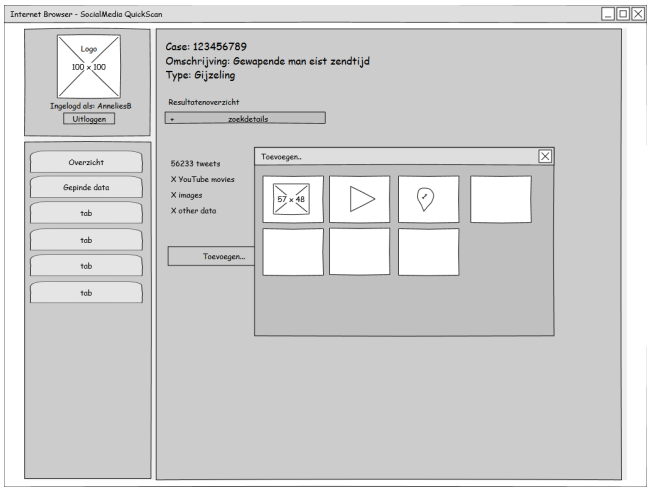


Figure 43: Wireframe with an example pop-up with options to add new data

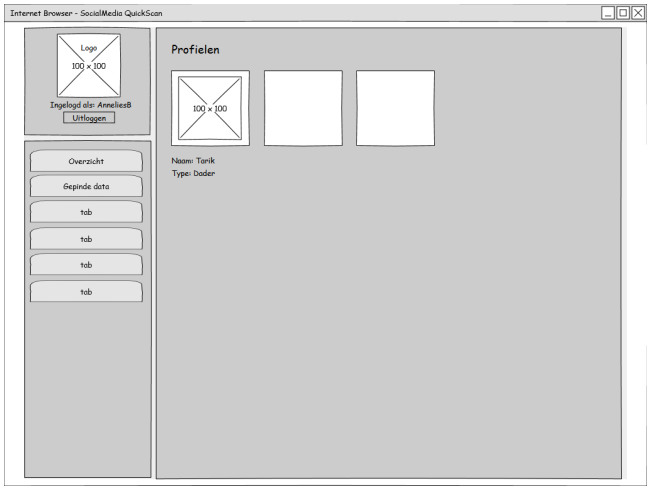


Figure 44: Wireframe of the page with an overview of the available profiles of persons

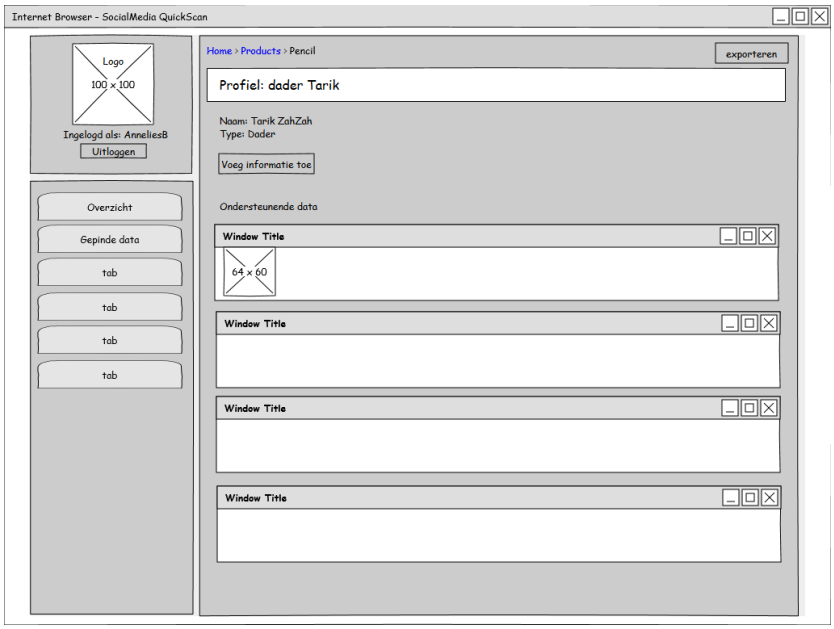


Figure 45: Wireframe of the page with detailed data belonging to a profile

H Screenshots of the design

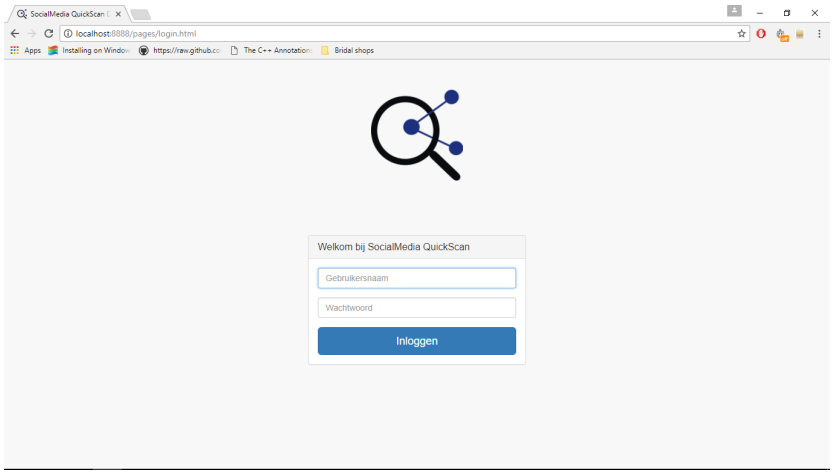


Figure 46: Login page

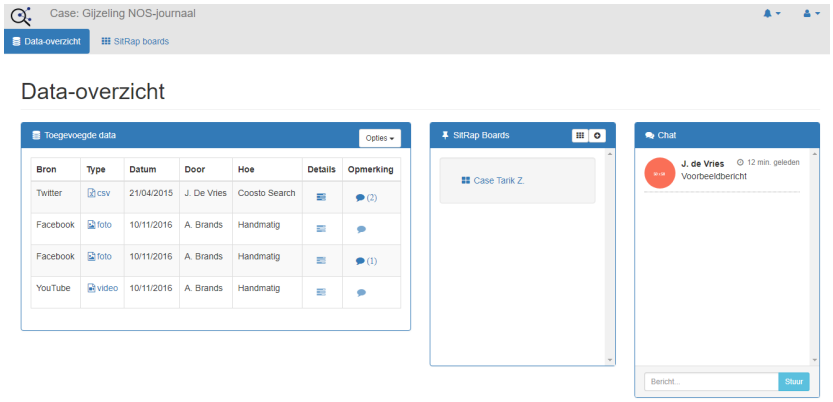


Figure 47: Upper part of the dashboard

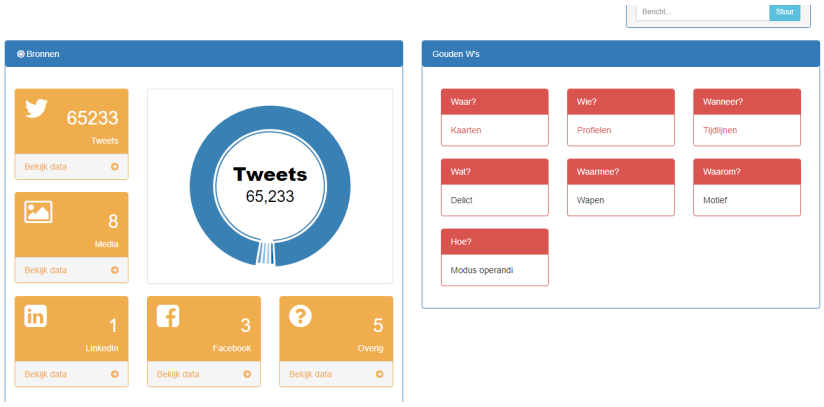


Figure 48: Lower part of the dashboard

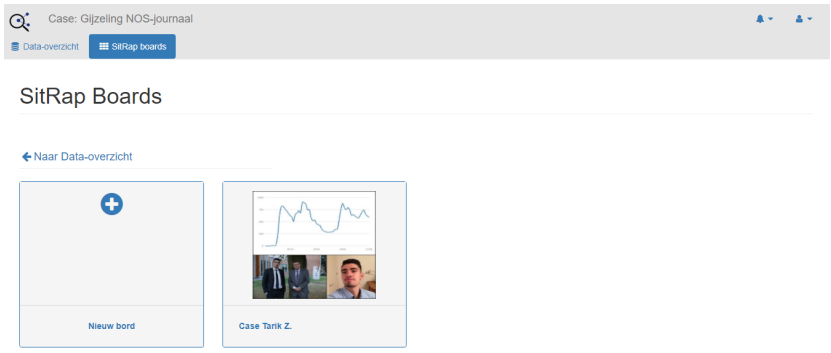


Figure 49: Overview of SitRap boards

Figure 50: Overview of SitRap board

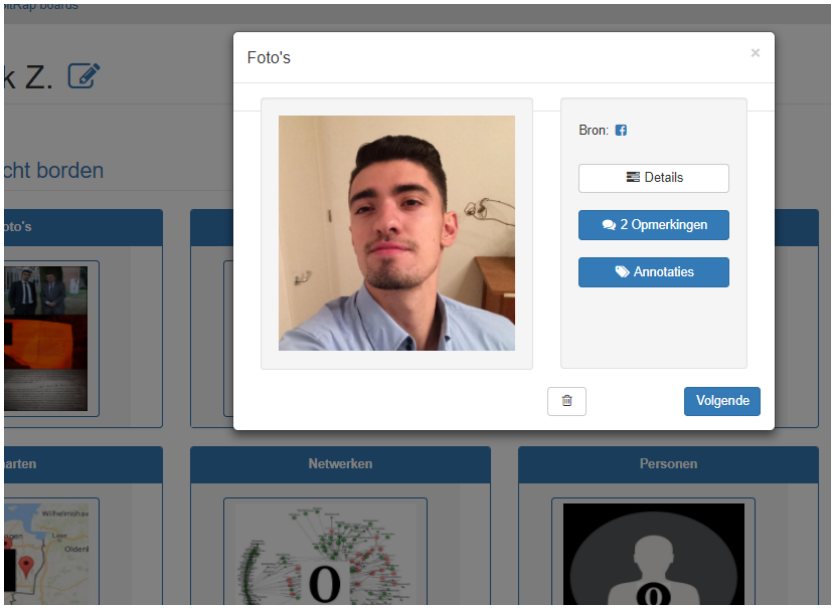


Figure 51: Pop-up data SitRap board

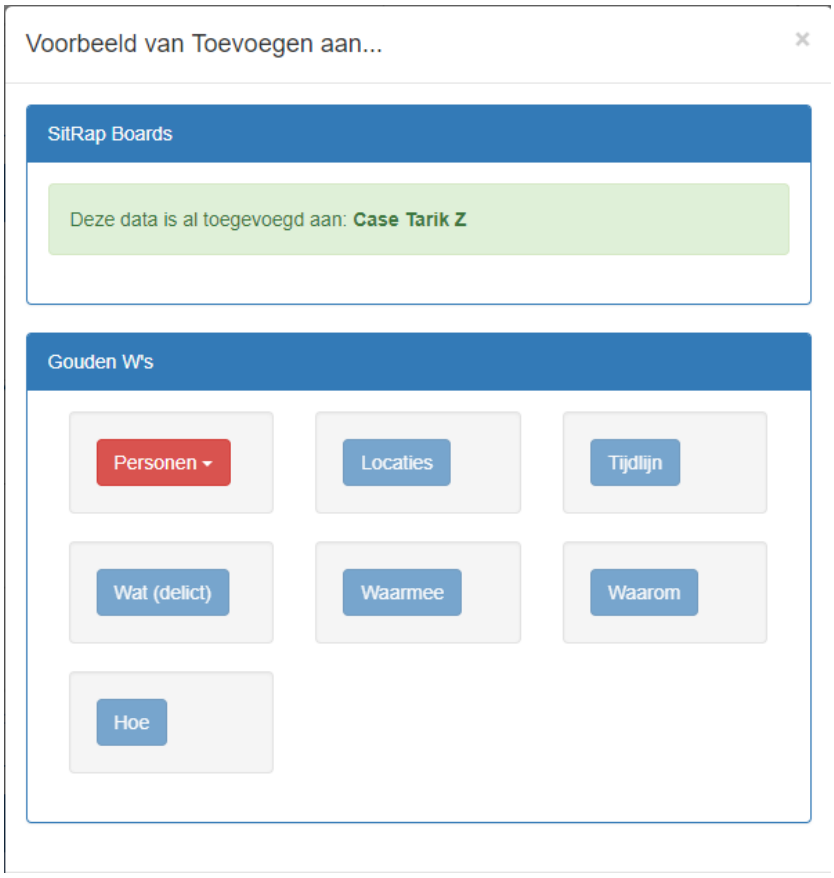


Figure 52: Add data to board or attach data to W-question

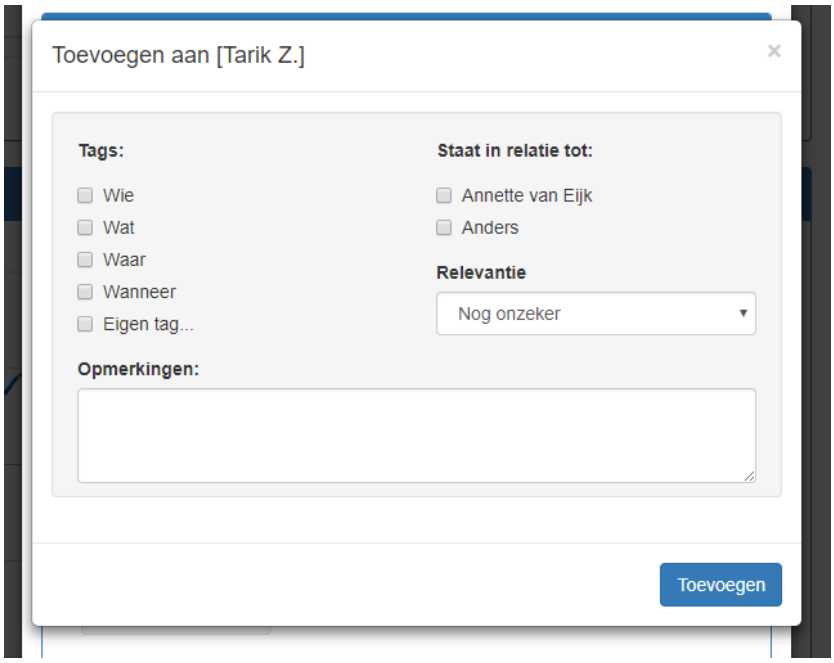


Figure 53: Add data to board or attach data to W-question

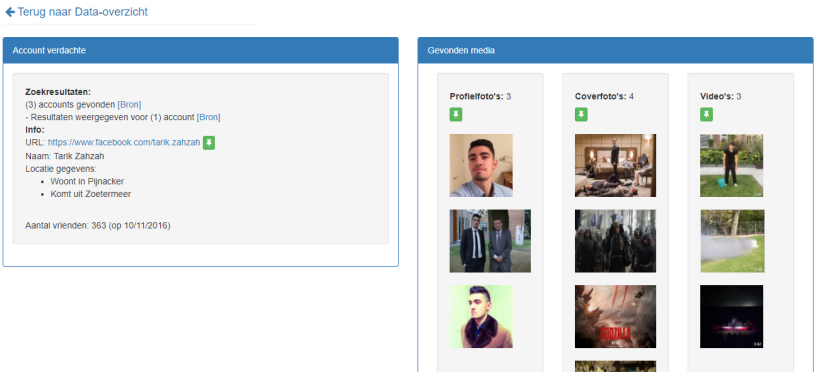


Figure 54: Upper part of the Facebook page

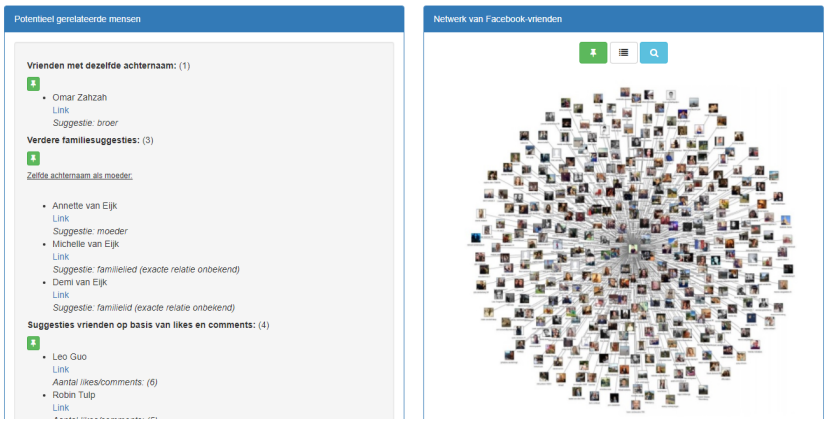


Figure 55: Lower part of the Facebook page

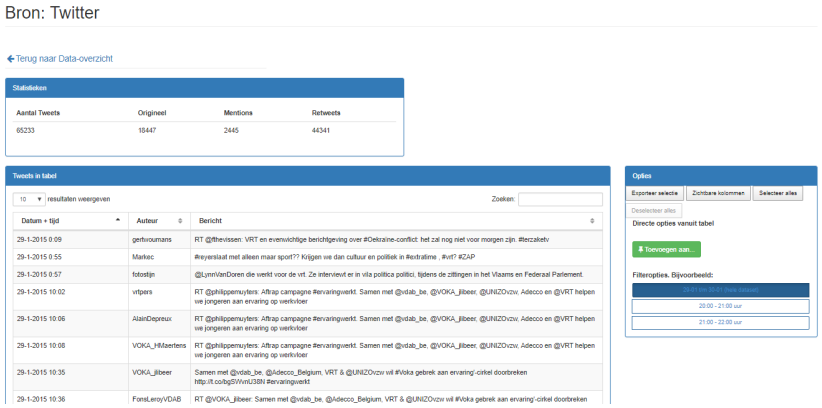


Figure 56: Upper part of the Twitter page

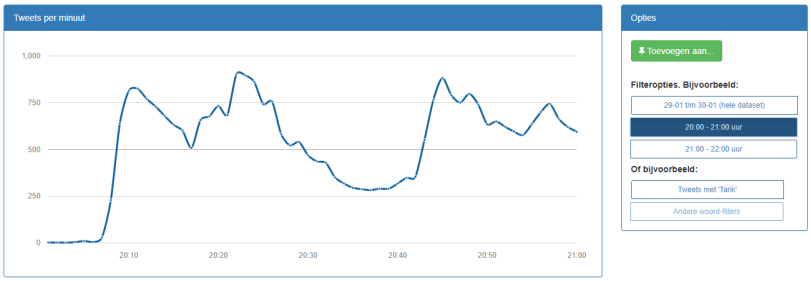


Figure 57: Tweets per minute visualization on Twitter page

Figure 58: Tweets with geodata on Twitter page

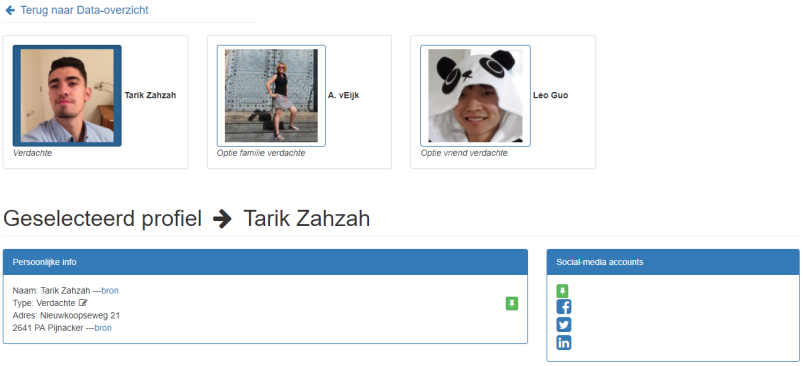


Figure 59: Upper part of the Profiles page

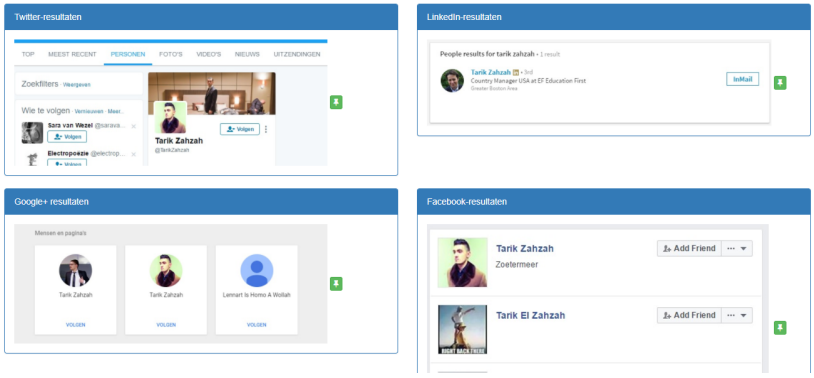


Figure 60: Middle part of the Profiles page

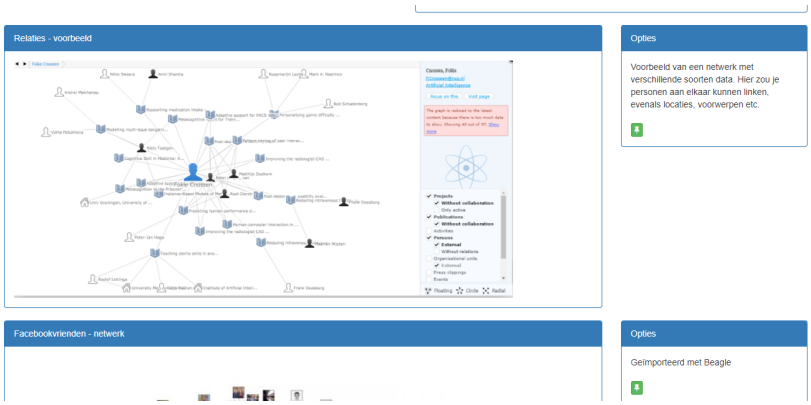


Figure 61: Lower part of the Profiles page

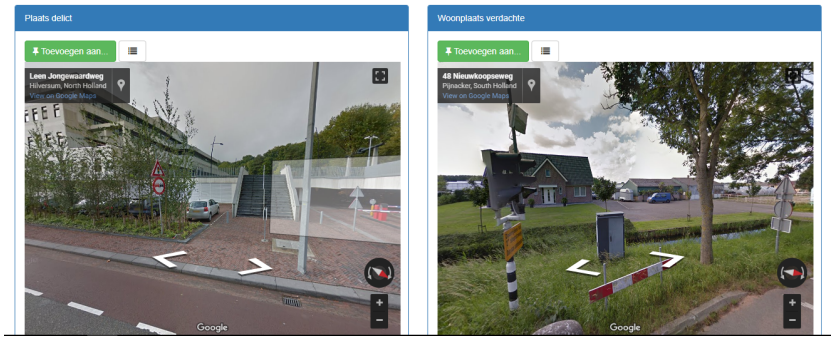


Figure 62: Screenshot of the Maps page

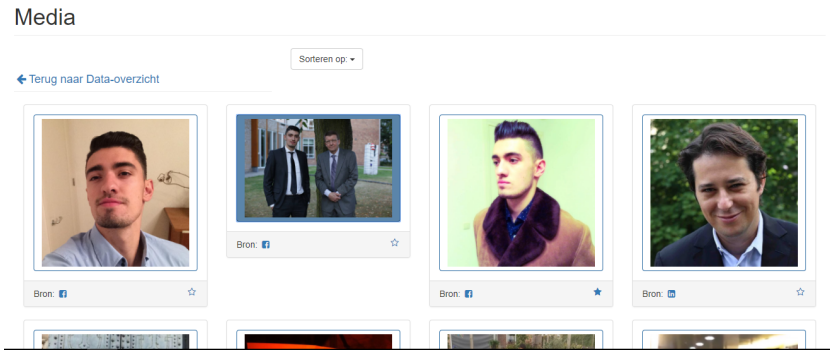


Figure 63: Screenshot of the Media page

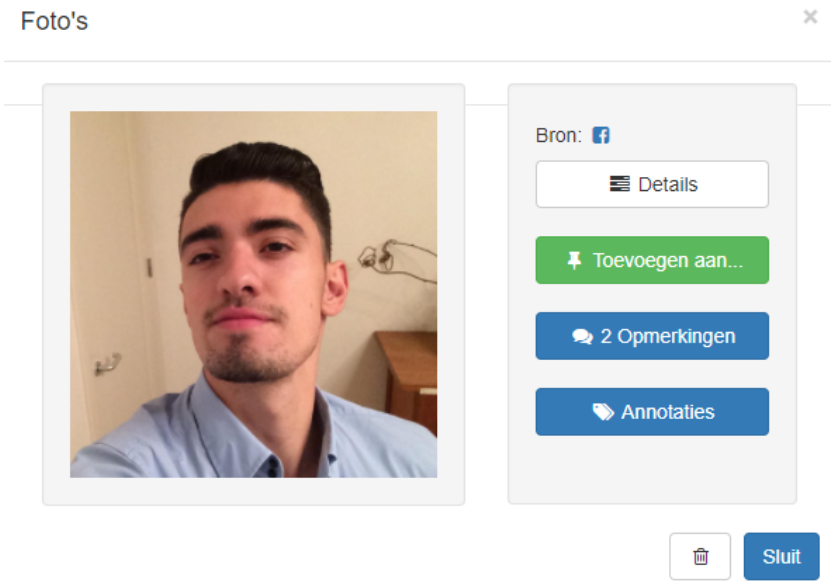


Figure 64: Pop-up of data at the Media page

Figure 65: Screenshot of the horizontal timeline

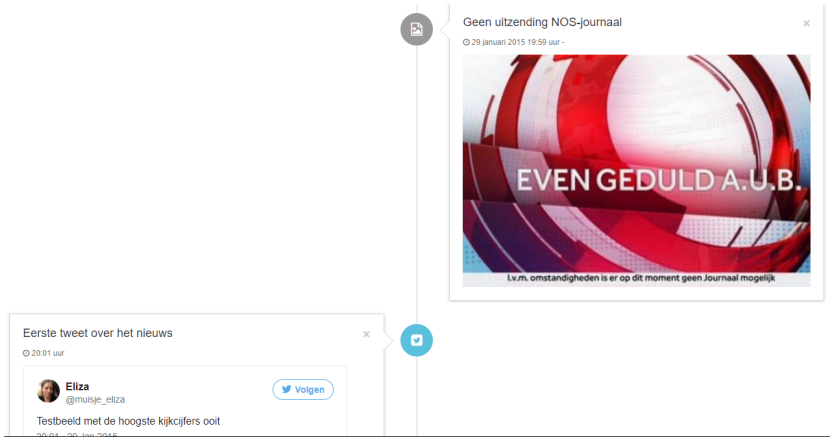


Figure 66: Screenshot of the vertical timeline

I List of individual comments and suggestions from the evaluation

All individual suggestions are listed in this appendix, grouped by subject.

Suggestions on the layout

- Exchange the panels 'SitRap boards' and 'Added data'
- Put the 'Sources' panel at the top of the page
- Change the chat panel to a pop-up at the bottom of the page
- Make the location in the tool clearer to the user by adding more 'bread-crumbs' on the pages.
- Adapt the title of tab to the page the user is at.
- Change the name of the 'Overview of data' to 'Homepage', because it is the page you return to everytime.

Suggestions on the overview of the data

- Link the parts of the donut visualization to the individual source pages.
- in the 'Added data' panel, the date and person can be put in a pop-up and be shown only on demand
- Show the officer's number next to its name, because multiple people can have the same name at the department.
- Add a button to directly export the chat data to a pdf for in a report

Twitter page

- Add a 'search' button next to the input field of the table
- Show the search query on top of the Twitter page.
- Show the number of results on top of the table instead of below the table
- Make separate filter buttons to quickly filter on witness statements, like 'I saw...' or 'I heard...'
- Add a filter button to filter on terrorism
- Show locations on a map based on locations mentioned in the content of a tweet
- Show hints to click on the markers on the map to find the hidden information
- add a 'loading' screen when the user has to wait for the data to load
- Add the option to click on points in the 'tweets-per-minute' figure that pops-up the tweets that belong to that point

Suggestions on functionalities

- Show notifications when new data is added
- Add the option to link work assignments to specific data and send it to a colleague.
- Add the option to sort data on data were colleagues have added comments on.
- Set out monitors that add data real time
- automatically search for relevant data
- Add news sources and fora
- connect the data with data from police sources
- Add layers that are or aren't visible to specific users.

Suggestions on the SitRap board

- Show the number of added photos on each page of the pop-up. And a suggestion to show it between brackets on the board, like this: Photos **(5)**
- Add the option to archive boards
- When clicking on the photos for example, show an overview of all selected photos that the user can scroll through
- Add the option to show the pictures or videos in full size when clicking on them

Suggestions on the timelines

- Enable the user to stack timelines from multiple events to look for connections
- On the vertical timeline: show the date and time in the middle right next to the timeline. That will give immediate insight in some context of the data.

Suggestions on the maps

- Load the maps with Globetrotter instead of Google Maps. It can be more accurate
- Add the option to add photos of the crime scene.

Bugs found

- The embedded YouTube videos continue running, even when the pop-up is clicked away.
- the 'tweets-per-minute' doesn's show the correct figure when the whole dataset is displayed.

Individual comments on items of the questionnaires

- On item 4, if the tool would support me in objective reasoning: objectivity is a relative term and it heavily depends on the search queries. Therefore I cannot say that this tool would support me on that.
- On item 9, regarding the tool to support the user to come from data to intelligence: I rated this with 2 (disagree), because this tool only is not enough to transform the data into intelligence. Social media data is always insecure when it comes to reliability. Therefore it has to be cross-checked with other systems, including police systems, before we can speak of intelligence.
- On item 16, regarding the contribution of the tool earlier in police investigations: the earlier use of social media in police investigations also depends on the capabilities of the person who is designated to this investigation.
- The single participant whose SUS results scored below average commented: I see great potential in the tool, but it is not very usable the way it is right now, and several functionalities are unclear to me. I have to see how the real-time parts get implemented and I want to work with the tool, before I can really comment on the usability.