DISSERTATION / DOCTORAL THESIS

Title der Dissertation / Title of the Doctoral Thesis

„A Framework for the Support of Cross-Media Communication During (Natural) Disasters“

verfasst von / submitted by

Dipl. Ing. Gerhard Backfried

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Doktor der Technischen Wissenschaften (Dr.techn.)

Wien, 2020 / Vienna 2020
Studienkennzahl lt. Studienblatt A 786 880 /
degree programme code as it appears on the student record sheet: A 786 880
Dissertationsgebiet lt. Studienblatt /
field of study as it appears on the student record sheet: Informatik
Betreut von / Supervisor: Univ.-Prof. Dipl.-Ing. Dr. Dr. Gerald Quirchmayr
Mitbetreut von / Co-Supervisor:

Contents

Abstract

Against the backdrop of an increase of the frequency and severity of natural disasters and a continuous extension and diversification of media used to communicate about them, the research performed within this thesis approaches the topic of disaster communication from a novel angle: its focus is the development of a framework – a conceptual structure - and model generic enough to represent the wide variety of data-sources and actors typically encountered during such incidents, taking into special consideration the connections between them. Their joint representation within a single model provides the advantage to detect commonalties and capitalize on them, to quickly be able to respond to newly emerging platforms, accommodate them within the model and to process data efficiently and comprehensively. The model is developed in an incremental manner, investigating what these generic elements and their attributes are, which connections exist between them and how the individual media and platforms can be represented. It allows to combine the different kinds of elements from relevant platforms and media, as determined by prior work, conversations with first responders and research on the state-of-the-art, in a cross-media approach and can be applied for analysis of communication during disasters. It furthermore provides a reference against which emerging technologies and tools can be compared and which can inform planning and management activities.

This thesis is based on findings of earlier work, in particular with respect to the mix of platforms and media and how these are employed and interact in disaster communication. However, where previous work tended to view and investigate individual platforms and media in isolation, the present thesis aims to abstract from this view and to accommodate all elements within a single model. In the process, the set of relevant platforms and media is refined and extended and the central topic of connections between elements is addressed explicitly. By focusing on commonalities and distilling generic properties, a model inspired by communication science is developed. This model is complemented by the identification of processes operating on the elements, resulting in a process model. The overall generic model allows a more systematic approach to and view of media-usage during disasters which does not only lend itself to technical implementation but may also be used for the design of guidelines, training, requirements analysis and procurement purposes. The work within this thesis is guided by principles of Design Science Research, ensuring relevance to real-world problems and contributing to a wider body of research. The model itself serves as an artifact and demonstrates that it is indeed possible to derive a generic model for cross-media disaster communication able to accommodate the full spectrum of elements involved. It comprises and describes the fundamental set of elements involved in communication, ways to link (combine) them,

attributes to describe and classify them and structures to process them and their content in an efficient manner. And it does so from a generic point of view applicable to the wide spectrum of platforms involved without requiring specific, repetitive work for each new platform. A graph-structure emerges as the natural way to represent the resulting network. This model is evaluated against a standard set of criteria employed to evaluate artifacts; a set of real-world examples taken from actual disasters serves to demonstrate the viability of the approach. A prototypical implementation and data collected during disasters provide further support for the motivation and practicality of the approach. The resulting structures allow for analyses previously not possible, the detection of connections which would have remained undetected and a possible deeper understanding of an evolving critical situation. At the same time, these additional possibilities also prompt new questions regarding cognitive load and appropriate manners of interaction which form alleys for further work.

Zusammenfassung

Vor dem Hintergrund einer Zunahme der Häufigkeit und des Ausmaßes von Naturkatastrophen und einer kontinuierlichen Erweiterung und Diversifizierung der Medien, in denen über sie kommuniziert wird, nähert sich die in dieser Arbeit durchgeführte Forschung dem Thema der Katastrophenkommunikation aus einem neuen Blickwinkel. Ihr Schwerpunkt liegt auf der Entwicklung eines (konzeptionellen) Frameworks und Modells, das generisch genug ist, um die große Vielfalt von Datenquellen und Akteuren darzustellen, die typischerweise bei solchen Vorfällen anzutreffen sind, wobei insbesondere die Verbindungen zwischen ihnen berücksichtigt werden. Die Darstellung all dieser Elemente innerhalb eines einzigen Modells bietet den Vorteil, Gemeinsamkeiten zu erkennen und zu nutzen, um schnell auf neu entstehende Plattformen reagieren und Daten effizient und umfassend verarbeiten zu können. Die Entwicklung des Modells erfolgt inkrementell; es wird dabei untersucht, welche generischen Elemente und Attribute existieren, welche Verbindungen zwischen ihnen bestehen und wie die einzelnen Medien und Plattformen dargestellt werden können. Das Modell ermöglicht es, die verschiedenen Arten von Elementen aus relevanten Plattformen und Medien, welche in früheren Arbeiten, durch Gespräche mit Hilfsorganisationen und durch Untersuchung des Stand-der-Technik bestimmt wurden, in einem medienübergreifenden Ansatz zu kombinieren und das daraus resultierende Framework für die Analyse der Kommunikation im Katastrophenfall anzuwenden. Darüber hinaus bietet es eine Referenz, anhand derer sich neue Technologien und Tools vergleichen lassen und die zu Planungs- und Managementaktivitäten herangezogen werden kann.

Diese Arbeit basiert auf den Erkenntnissen früherer Arbeiten, insbesondere in Bezug auf den Mix von Plattformen und Medien und wie diese in der Katastrophenkommunikation eingesetzt werden und interagieren. Doch wo frühere Arbeiten dazu neigten, einzelne Plattformen und Medien isoliert zu betrachten und individuell zu untersuchen, zielt die vorliegende Arbeit darauf ab, aus dieser Sicht zu abstrahieren und alle Elemente in einem einzigen Modell zu repräsentieren. Dabei wird die Menge der relevanten Plattformen und Medien schrittweise verfeinert und erweitert und dabei das zentrale Thema der Verbindungen zwischen den Elementen explizit untersucht. Durch die Konzentration auf Gemeinsamkeiten und das Herausarbeiten generischer Eigenschaften wird ein von der Kommunikationswissenschaft inspiriertes Modell entwickelt. Dieses Modell wird ergänzt durch die Identifizierung von Prozessen, welche auf den generischen Elementen aufsetzen. Diese führen zur Erstellung eines Prozessmodells. Das generische Gesamtmodell ermöglicht einen systematischen Ansatz und eine strukturierte Betrachtung der Mediennutzung bei Katastrophen, wodurch es sich nicht nur für die technische Umsetzung, sondern auch für die Gestaltung von Richtlinien, Schulungen,

Anforderungsanalysen und Beschaffungszwecke eignet. Die Arbeit wird von den Prinzipien des Design-Science-Research geleitet, um die Relevanz für reale Probleme sicherzustellen und zu einem breiten Forschungsgebiet beizutragen. Das Modell selbst dient als Artefakt und zeigt, dass es möglich ist, ein generisches Modell für die medienübergreifende Katastrophenkommunikation abzuleiten, welches das gesamte Spektrum der beteiligten Elemente berücksichtigt. Es umfasst und beschreibt den grundlegenden Satz von Elementen, die an dieser Kommunikation beteiligt sind, die Möglichkeiten ihrer Verknüpfung (Kombination), Attribute zur Beschreibung und Klassifizierung sowie Strukturen zur effizienten Verarbeitung. Dies alles aus einer allgemeinen Sicht, die für das breite Spektrum der beteiligten Plattformen anwendbar ist, ohne dass für jede neue Plattform spezifische, repetitive Arbeiten erforderlich wären. Eine Graphenstruktur erweist sich als der natürliche Weg, das resultierende Netzwerk darzustellen. Dieses Modell wird gegen ein Standard-Set von Kriterien zur Bewertung von Artefakten evaluiert. Eine Reihe von Beispielen aus realen Katastrophen dient dazu, die Motivation für diese Arbeit zu unterstreichen und die Anwendbarkeit des Ansatzes zu demonstrieren. Eine prototypische Implementierung und die bei Katastrophen gesammelten Daten bieten weitere Einsichten. Die daraus resultierenden Graphenstrukturen erlauben Analysen, die bisher in dieser Form nicht möglich waren, gestatten die Detektion von Verbindungen, welche ohne entsprechende Modellierung unentdeckt geblieben wären und führen zu einem tieferen Verständnis einer sich entwickelnden kritischen Situation. Gleichzeitig werfen diese zusätzlichen Möglichkeiten jedoch auch neue Fragen nach der kognitiven Belastung und den entsprechenden Interaktionsmöglichkeiten auf, die Gebiete für zukünftige Forschungsarbeiten darstellen.

Acknowledgements

I would like to express my sincere gratitude to my family and all friends and colleagues who have accompanied me during the work on this thesis. From its inception - following the experiences made during a research project in 2013 - to its completion, several years have passed. Throughout this period, I have received continuous support and motivation from many persons. I have also been able to draw upon the resources of the company I work for, which, especially regarding the environment to gather and process data, to implement and evaluate ideas and to visualize outcomes, has been invaluable. Many friends and colleagues have acted as sparring partners, provided me with feedback, and helped to proof-read and correct intermediate versions of this work. Without this support, it would not have been possible for me to complete this thesis. Thank you all very much!

1 Setting of this Thesis

Work on this thesis has been *framed* and *accompanied* by several natural disasters: from its inception, in 2013, the year of the worst floods in central Europe in over 500 years to its completion in 2019, when a sequence of cyclones never experienced in this intensity before, hit East Africa. Between these two dates, several of the largest Hurricanes ever to be recorded raged in the Caribbean and the US, Typhoons devastated large regions in South East Asia and an Earthquake with a subsequent Tsunami hit the Indonesian island of Sulawesi. These and numerous further disasters receive ample coverage on traditional and social media, on local, regional and global scale, allowing to glean valuable information about developments directly from the scene as well as from international perspectives. A large share of this information comes in the form of multimedia, is expressed in different languages and distributed over a wide and diverse array of platforms. Being able to use it efficiently and comprehensively allows to gain a better understanding of critical events and to increase situational awareness for citizens and first responders. This thesis aims to create a *generic framework*[1] able to accommodate all these ingredients involved in disaster communication and to demonstrate that it can be put into use for actual operation by practitioners. The framework itself as well as the methodology to develop it form the core of the work described below.

1.1 The QuOIMA Project

The KIRAS project QuOIMA (*Quelloffene Integrierte Multimedia Analyse*, KIRAS, 2015; Die Presse, 2015), addressing the use of social and traditional media in the context of natural disasters in Austria was launched in November of 2012 and completed in October 2014. Against a backdrop of increasing numbers and impacts of natural catastrophes, the project dealt with the use of open-source information and the combination of different types of media in the context of natural disasters based on a specific 5-phase disaster model (Backfried et al., 2013). Its goal was to allow decision makers to quickly make sense of a large, dynamic, multilingual and inhomogeneous body of data (Backfried et al., 2016). In collaboration with a first-responder (the Austrian Armed Forces)[2] the Central European Floods of 2013 were chosen as the specific use-case. Emphasis was placed on the creation of an infrastructure supporting the ingestion and analysis of data from several traditional and social media platforms with the aim of establishing a basis for decision-making. The inclusion and integration of social media had not been addressed by the end-user before and thus formed a central requirement

[1] The scope of *Framework* within this thesis is described in chapter 2.5 and further defined in chapter 5
[2] Also supported by a second first responder, the Austrian Red Cross

from their perspective[3]. This subject prompted several challenges and resulted in technical innovations and extensions to the previously existing infrastructure. During the project, the frequent phenomenon of parallel communication, within and across different channels and platforms became apparent. The potential kinds of cross-media communication and of communication-patterns occurring at different stages of a disaster were subsequently investigated using this infrastructure. Textual processing, audio-processing and visual-processing were combined, and an underlying multi-lingual ontology employed to also allow for cross-lingual processing[4]. A multi-modal corpus, spanning traditional as well as social media covering this event was created during the project (the first of its kind).

The architecture of QuOIMA was based on the Sail Labs Media Mining System[5] (MMS), a system designed for real-time processing of open-sources targeting the area of Open Source Intelligence (OSINT) (Backfried et al., 2012; Backfried et al., 2015). Even though its primary focus are applications in the realm of security and defense, the concepts, components and work-flows can readily be applied to other domains such as business intelligence (BI) and the management of disasters and crises. Within the MMS, individual components, such as the automatic speech recognition (ASR) engine used to transcribe TV or radio content or the web-crawler, used for websites and RSS-feeds, are built in a way to make them re-usable and extensible in a straight-forward manner. Components, technologies, models and the support for different languages can be extended step-by-step according to the requirements of the particular application domain. This flexibility resulted in an incremental approach leading to several iterations of component development and deployment within the project. As a result, a prototype was developed and deployed, demonstrating (some of) the capabilities of the MMS for disaster management.

QuOIMA's overall approach was based on the following assumptions:

- Both, traditional and social media play fundamental and complementary roles and will continue to do so in the foreseeable future

[3] The use of Social Media during disasters had not been investigated at all in Austria up to this point.
[4] Due to its geographic location, floods of the Danube in Austria also concern neighboring countries and communication in languages other than German, but also due to tourists and residents using languages other than German this aspect is of importance and should not be neglected.
[5] https://www.sail-labs.com/media-mining-system

- Their respective strengths can be increased by combination of sources and channels. Links detected between different media provide insights not available by approaches based on a single medium only
- Multimedia processing capabilities are required to ingest the complete spectrum of sources covering an event such as robust and extensible image-, text- and audio-processing technologies
- Crises and disasters often take place in multinational, cross-border, multilingual and multicultural settings. Consequently, media in different languages and dialects need to be processed.

Figure 1 below depicts a schematic overview of these assumptions concerning the cross-media nature of communication during disasters (Backfried et al., 2013a) at the outset of the project.

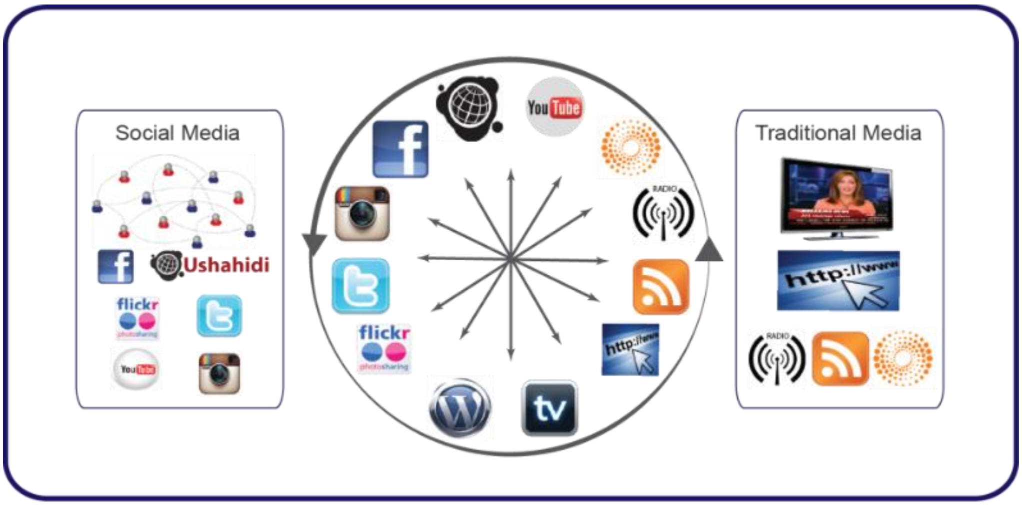


FIGURE 1 SCHEMATIC OVERVIEW OF TYPES OF CROSS-MEDIA LINKS

During the lifetime of the project, Central Europe experienced some of the worst flooding in 500 years, causing severe damage in several countries, among them in Austria. Figure 2 below provides a geographic view of the affected areas, showing that both the Danube/Inn rivers and their tributaries as well as the Elbe/Saale river-system presented areas of severe flooding.

FIGURE 2 MAP OF 2013 EUROPEAN FLOODS[6]

These floods were selected as the specific use-case for QuOIMA. Data from various media was collected during the time of the most severe flooding in Austria and a resulting corpus used for processing and analysis was created. This corpus, the first of its kind to span several media, is presented in the chapter on relevant inputs to this thesis from the QuOIMA project. In many ways, the work carried out in that project provided insights, prompted questions and exposed deficiencies and gaps of current approaches[7]. Findings and developments made within the project serve as a foundation for the current thesis and provide input to the models developed below. However, whereas QuOIMA prompted the development of individual technologies and components triggered by a singular use-case – and in particular the addition of the processing of social media in the context of disaster management in Austria - it did not go beyond these nor evolve into a generic framework. It lies in the nature of research-projects, that certain insights and results can only be obtained at the end of the project, leaving room for follow-up research as well as the possibility to take a *step back* and view matters from a more principled and general standpoint. The present thesis aims to refine the ideas developed in QuOIMA and take them one step further, equipped with the insights gained during the project. It aims to extend and systematically investigate these prior results, considering the state-of-the-art and results of further research, systems and architectures with the eventual goal of producing a generic framework and approach for cross-media and cross-platform communication during crises and disasters. Whereas the results of QuOIMA were based only on particular instances

[6] Copyright Alexrk2, licensed under CC BY-SA 3.0
[7] Detailed in chapter 2.2 below

of sources, media and data-sets, this thesis aims to take these insights *a step forward* by abstracting from the particular instances and establishing a framework on a generic level. In a way, this also entails taking a step back and creating a firm basis for such a model, defining the necessary terminology[8] and putting it into context with disaster communication and news consumption behavior. The resulting framework provides systematic insights into the mechanisms of communication during disasters and is capable to accommodate the different kinds of media and sources involved. Insights gained, data collected and the infrastructure which has been created as part of QuOIMA as well as the initial results from investigations of cross-media communication serve as the background on which the work of the present thesis is based.

1.2 Natural Disasters

Over the past decades, a rise in natural disasters[9,10,11] has been observed. Today more people are affected by disasters (and conflicts), more frequently, and for longer than in previous decades (UNOCHA, 2017). Extreme weather conditions – induced by global warming and climate change – have become more frequent and more severe and will likely continue to be so in the future. Their impact has become more profound both, from a humanitarian as well as from an economic point of view. A study undertaken by Munic Re[12] in 2018 (Figure 3) shows an increasing trend concerning the number of natural disasters on a global scale. Overall, the number of recorded natural disasters has approximately doubled over the past two decades alone.

[8] This is especially important in the light of the inter-disciplinary nature of the field.

[9] Whereas the proposed thesis focuses on natural disasters, many aspects equally apply to human-made or technological-disasters such as the large- scale migration effects currently witnessed or to industrial accidents. The mechanisms and structures devised may, to a large extent, also be applied to such kinds of disasters.

[10] The term *disaster* has been defined in many different ways. The current work adopts the definition by the United Nations Office for Disaster Risk Reduction (UNISDR), http://www.unisdr.org/we/inform/terminology on 2018/11/111

[11] See also the general classification proposed by Centre for Research on the Epidemiology of Disasters, CRED, https://www.emdat.be/classification

[12] Munic Re, https://www.munichre.com/topics-online/en/climate-change-and-natural-disasters/natural-disasters/2017-year-in-figures.html on 2018/11/06

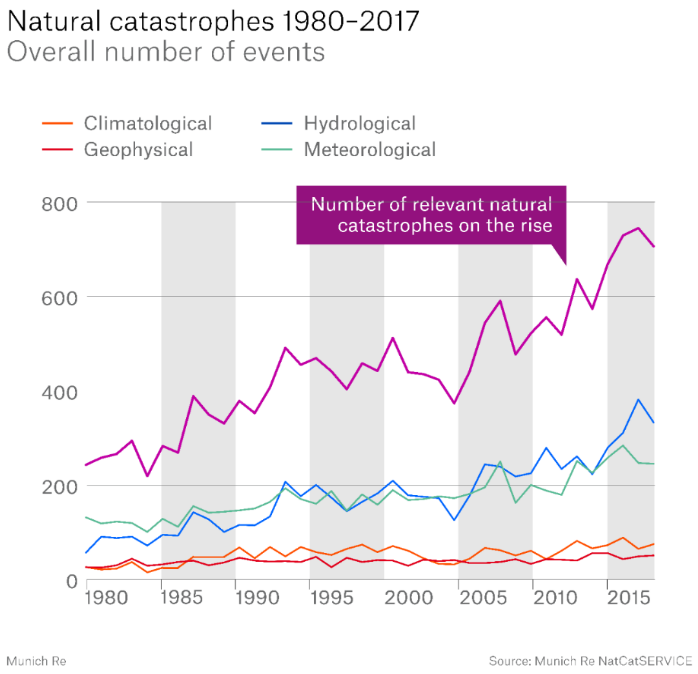


FIGURE 3 NATURAL CATASTROPHES OVERVIEW[13]

Regarding the US, the National Oceanic and Atmospheric Administration (NOAA) reports that 2018 has been a year with significant economic and humanitarian impact. The overall number of *Billion-Dollar-Disaster-Events* (large scale disasters whose cost surpasses USD 1 bn) displays an increasing trend as shown in Figure 4 below (NOAA, 2018).

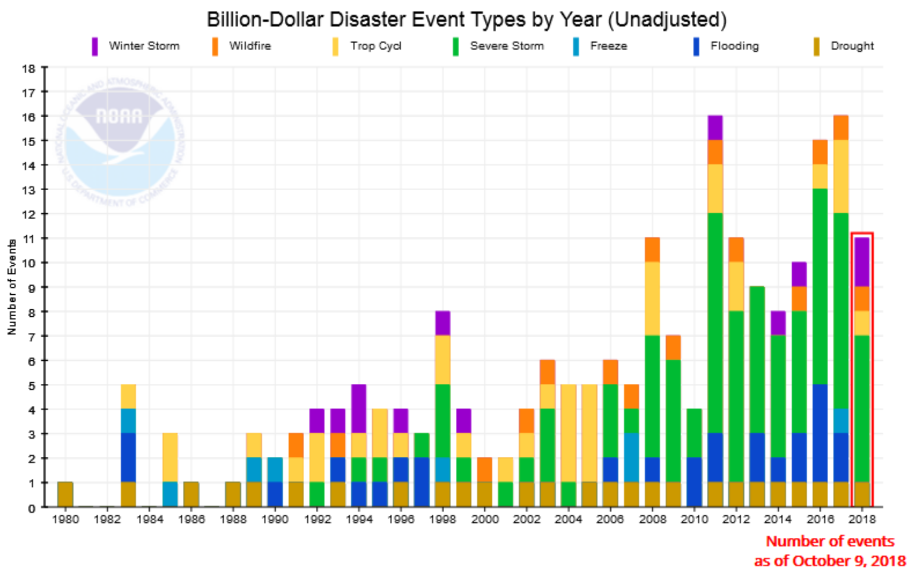


FIGURE 4 NOAA - BILLION DOLLAR DISASTERS IN THE US[14]

[13] Figure from Munic Re, https://www.munichre.com/topics-online/en/climate-change-and-natural-disasters/natural-disasters/2017-year-in-figures.html
[14] Figure generated on NOAA web-site, https://www.ncdc.noaa.gov/billions/time-series on 2018/11/10

In 2017, three major hurricanes – Harvey, Irma and Maria brought fear and widespread destruction to several countries in the Caribbean, Mexico and the United States. Irma, a storm the size of France and one of the most powerful hurricanes on record, made landfall in Florida and ravaged the gulf coast of the US. Only a short while before, another hurricane, Harvey had hit Texas and brought with it unprecedented amounts of rainfall in a very short period of time. It caused severe flooding in vast areas, among them the greater Houston metropolitan area, destroying homes and displacing thousands of people. Within a timespan of only four weeks, this trio of hurricanes had made 2017 the costliest North Atlantic hurricane season in history (Munic Re, 2018). The 2018 North Atlantic hurricane season boasted one of the most severe storms in history, Hurricanes Florence and Michael[15] making landfall in the US while the Pacific Typhoon season produced several super-typhoons, one of them Typhoon Mangkhut/Ompong[16] which made landfall in the Philippines and brought widespread destruction to large areas. By September of 2018, by far not the end of the Pacific Typhoon season, a total of 25 named storms had already hit countries of South East Asia[17].

In addition to storms, a host of further natural disasters hit North America in 2017 and 2018: several wildfires raged throughout the West of the US and Canada and even affected Greenland. In 2018, the Mendocino Complex fires in California became the largest fire ever recorded in that state.

Even though these specific disasters received widespread attention and media-coverage around the world – several of them were located in the US and concern the largest Western economy with a rich media landscape - they were by no means the only disasters with profound humanitarian and economic impacts. Figure 5 below (Munic Re, 2018) depicts global disasters from an insurance (loss) point of view for the year 2017. Differentiating between four categories of natural disasters (geophysical, meteorological, hydrological and climatological), the figure depicts the costliest events on a world-wide basis for that year. The disasters include slow-onset disasters, such as tropical storms and heat-waves as well as sudden-onset disasters such as earthquakes or landslides. They are distributed around the globe and concern developed regions as well as regions in development. Their impact is intricately linked to the scale and extent of the population and infrastructure affected in the areas hit by disaster. In the age of globalized commerce, far-reaching inter-dependencies and supply chains which stretch across the planet, the effects felt by economies and societies around the world may be stronger and more immediate than the perceived geographic distance of these disasters. Ultimately, many of these disasters have a global impact.

[15] https://en.wikipedia.org/wiki/Hurricane_Michael on 2018/11/10
[16] https://en.wikipedia.org/wiki/Typhoon_Mangkhut on 2018/11/10
[17] https://www3.nhk.or.jp/nhkworld/nhknewsline/backstories/disastroustyphoon/ on 2018/11/11

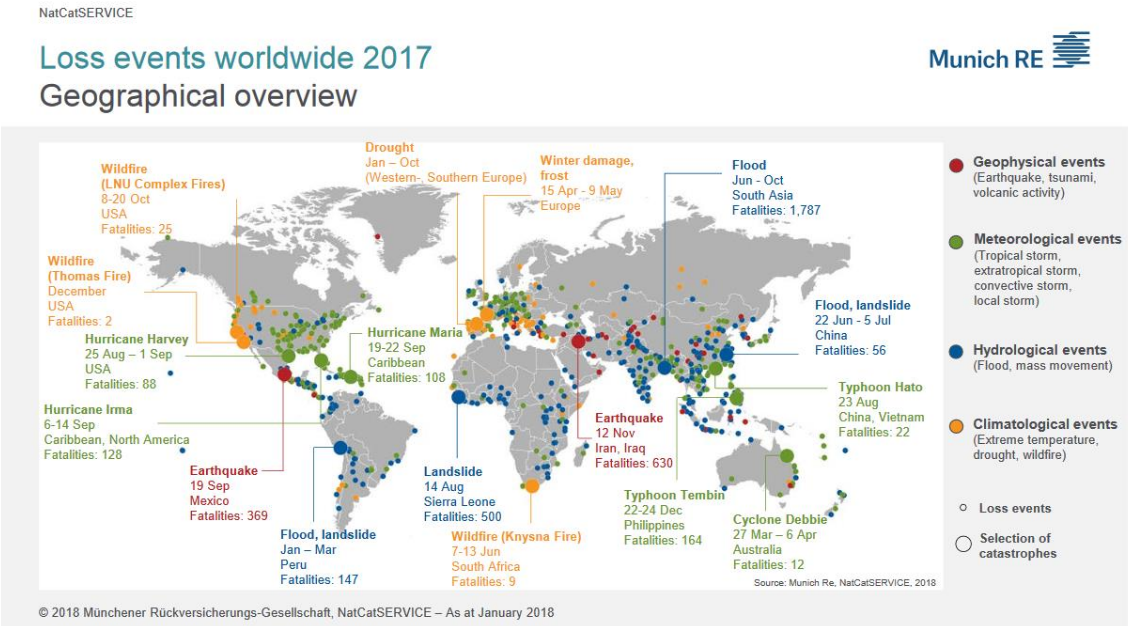


FIGURE 5 MUNIC RE, NATURAL CATASTROPHES 2017, GEOGRAPHICAL OVERVIEW[18]

The overall losses of 2017 even exceeded those of the extreme years of 2005 and 2008 and are significantly higher than the corresponding averages of the last ten years. The 730 events recorded as relevant on a world-wide scale meant that 2017 joined the years with the highest number of natural disasters. This rising trend in disasters and predictions that certain types of extreme weather events are likely to become even more frequent in the future may make 2017's disasters a mere foretaste of what to expect and to prepare for in the future (Munic Re, 2018). The economic and humanitarian impacts[19] on this scale will pose great challenges to societies and economies alike. Recent numbers for 2018 indicate that this year was yet another one with losses above average. In addition, it incurred a higher toll in fatalities with earthquakes and tsunamis in Indonesia claiming most lives[20].

Creating effective means to support organizations in dealing with such disasters and situations, allowing them to prepare, make sense and quickly respond to events has become an active area for research and commercial activities. All phases in the disaster cycle from preparation to recovery can be addressed and organizations such as first-responders be supported by such technologies. Comprehensive strategies and tools are essential and needed for all phases of disaster management.

[18] Figure from Munic Re, https://www.munichre.com/site/corporate/get/params_E976667823_Dattachment/1627370/MunichRe-NatCat-2017-World-Map.pdf
[19] The long-term (30 years) trend regarding fatalities shows a reduction in the number of victims.
[20] https://www.munichre.com/topics-online/en/climate-change-and-natural-disasters/natural-disasters/the-natural-disasters-of-2018-in-figures.html on 2019/05/30

The work in this thesis attempts to form part in this effort from the technical side of affairs and aims to provide insights and a framework which can be helpful for the design and implementation of an IT-infrastructure supporting disaster management activities.

1.3 Disasters and Media

Since their very beginnings, the media have played an important role in times of disaster and crisis[21]. They are crucial in communicating and understanding disasters and their implications and have gained prominence in the tasks of involving different actors, boosting actions of relief and support to the victims and generate trust among those affected (Reis et al., 2017). They form an integral part of disaster communication, rapidly reaching large portions of the population (Zemp, 2008). The pervasive nature of media in our lives and their choice of which events merit and attract attention shape and determine our experience of these events and influence how we come to understand them. In the spirit of Gerbner's Cultivation Theory (Moshrafa, 2015) the mass-media, especially TV may indeed shape our perception of disaster-reality to be aligned with the reality as portrayed on them. In a study focusing on the 2011 Great East Japan Earthquake, Cheng et al. (2015) investigate the role of the mass media and specifically find television to have been the most influential medium, particularly for those individuals not directly affected by the disaster. Radio likewise keeps playing a vital role during disasters, reaching wide and often remote audiences and is recognized as a powerful communication tool and a low-cost medium (UNISDR, 2016). But whether it's traditional media such as radio, TV or the press, or social media, such as Facebook, Instagram or Twitter, the contemporary media keep people in the affected areas and around the world informed about the state of an emergency as it unfolds, support response and resource mobilization, provide a way for authorities to listen to the needs of the affected population and foster a shared sense of identification (Greenberg & Scanlon, 2016).

During emergencies, it is essential for the affected population, first-responders and authorities to have access to broad, reliable and up-to-date information. For the affected population, warnings, alerts and timely information regarding safety and protection are essential. Frequently, local members of the affected community will be able to share first-hand information and provide detailed accounts of what is happening. For first responders, such irregular events require a profound and comprehensive situational understanding of the relevant circumstances and actors involved in order to put them into

[21] The terms *disaster* and *crisis* are often used interchangeably, as outlined by Coombs and Holladay, 2012. They propose to interpret *crisis* within an organizational context whereas disaster to be used in contexts external to an organization. For the scope of the current work and its focus on natural disasters, the term disaster is preferred.

context, form a proper and timely response policy and to ensure efficient allocation of the required resources (Backfried et al., 2012).

Information is produced across a spectrum of media around the globe and on a multitude of sources varying widely regarding their nature, language, reliability, quality and quantity. The amount is ever increasing and its content changing continuously. Content is put on the Internet by professional news agencies or individuals, broadcast on TV- or on video-sharing platforms, announced on radio stations or posted on social media platforms. Some of the most immediate information about an event might be produced by citizens with little experience, using mobile devices being unaware of official terminology[22] and using local slang and jargon. In particular, the multitude of languages involved forms a critical factor, as often one of the key-problems of disasters with international or even global relevance is that information is only available from local sources and in local languages or dialects. Frequently, organizational structures and technologies to support the processing of such information are not up to the task and do not support these languages. Hence valuable information may be ignored and lost.

Due to the local nature of many disasters, the public is often only informed about them in the mass-media rather than experiencing them first-hand. In this process, the media form an intermediary between authorities in charge of risk- and crisis-communication and the general public, framing the meaning, scope and significance of disasters. Journalists, however, are also often located far away from the scene of disasters, unaccustomed to the discomfort and anxieties caused by them. They select and filter events and news, cooperate with, complement or even act against official crisis communication mechanisms and can be either an *unwelcome irritant* or a *useful ally* in such circumstances (Scanlon, 2007). Whereas local media will typically be more concerned with the information needs of the affected population, more distant media will cater to the expectations of an imaginary, unaffected and distant audience (DKKV, 2000, p6). Tabloids and the so-called *yellow-press*[23] will typically exaggerate events in a sensationalist manner; broadsheets and quality-journalism will tend to present them in a more balanced and comprehensive way (DKKV, 2000, pp 19). Fueled by the desire to capture audience attention, disasters have become prime events to report about (Zemp & Bonfadelli, 2008). In some regions of the world (e.g. in the US) reports are delivered in apocalyptic

[22] Remark during a keynote given at ISCRAM 2014 conference: during hurricane Sandy, messages about flooding were received by first responders via Social Media. A team was quickly dispatched to the location only to find out an insignificant amount of water had accumulated – a situation which would be referred to as *ponding* by experts, not warranting an immediate dispatch of responders, particularly in times when these resources were scarce and in high demand.

[23] Yellow Press or Yellow Journalism, https://en.wikipedia.org/wiki/Yellow_journalism on 2017/09/30

tones so frequently that dire warnings may lose their impact and ensure that nobody takes warnings too seriously because they have all heard it too many times before (Cockburn, 1011).

The relationship between authorities involved in official crisis-communication and the media plays an important role in the selection, acceptance, trust (or mis-trust), support or contradiction of official disaster communication. Especially when assigning responsibility for the causes or the extent of a disaster or pointing out deficiencies in crisis-management, this relationship may experience disruptions and conflicts. Ultimately, however, the media will tend to orientate themselves on the actual events which they choose to report about and reduce complexity, simplifying and condensing issues along the way (DDKV 1998, p22).

What will be communicated by the mass-media also depends strongly on the willingness or ability of authorities to communicate about a disaster. Freedom of the press, censure, outright oppression or the motivation to deny, hide or blur matters for political or economic reasons or merely to disguise a lack of knowledge may further complicate matters. Figure 6 below shows an image of the "Freedom of Press" map of the Newseum[24] in Washington, D.C, USA. It divides countries on a global scale into three categories: *Free* (green), *Partly Free* (yellow) and *Not Free* (red). Whereas this map is certainly biased from a US point of view and the situation in the depicted countries changes over time, the map still provides a striking image to the fact that in large regions of the world, mass-media are not free, controlled or influenced by the state and authorities and thus also in times of disasters cannot be expected to provide credible, reliable or comprehensive information.

FIGURE 6 FREEDOM OF PRESS MAP, NEWSEUM, WASHINGTON, D.C., 2013[25]

[24] http://www.newseum.org/ on 2018/11/11
[25] Personal photograph of author, 2013

Under such circumstances, social media may play an even more prominent role (e.g. during the Arab Spring (Khondker, 2011) or the protests in Iran[26]), filling the information vacuum created by the absence of mass-media coverage. However, even in countries with a free-press, such situations may occur: a prominent case of disaster communication problems – occurring in Japan, a country with a free press - is the Fukushima-Daiichi nuclear disaster caused by the tsunami following the Tohoku earthquake on 11 March 2011 (Acar & Muraki, 2012; Peary et al., 2012). Not only is this event a prime example of a cascade of disasters – an earthquake leading to a tsunami leading to a nuclear meltdown – and the immediate connections between natural disasters and human-made disasters, it also demonstrates the rapid transition of a local disaster into one of global relevance (Lipscy et al., 2013). The Fukushima nuclear disaster is also a striking example of poor communication (resp. the *lack* of communication) and delays causing utmost confusion, an erosion in public trust and a loss of valuable time for response and rescue activities (Pacchioli, 2013). Omissions and failures on part of TEPCO and the Japanese government created an information vacuum which left ample room for speculations and rumors and which was consequently filled by other actors including many users on social media.

[26] https://www.bbc.com/news/world-middle-east-42566083 on 2018/11/10

2 Thesis Approach

2.1 Motivation for the Cross-Media Approach

Increases in the number and intensity of natural disasters paired with a growth in the diversity and variety of traditional and social media reporting about them – often inter-linked and used in combination - in an increasingly international and sometimes even global setting and typically involving multiple platforms, languages and modalities, provide a strong case for an approach accommodating this wide variety of platforms, media and processing capabilities for and across the different types of media – a ***cross-media approach***. Insights and data from a research project in the domain of disaster communication, the experiences made with first-responders, their dependence on comprehensive information from a broad set of sources as well as current trends in research further support this view.

Such a cross-media approach should account for the following principles:

- Multilinguality
  - Data about disasters is often available in a variety of languages and dialects. Public media as well as individuals use different languages and language-registers when communicating about disasters, providing information or expressing their concerns, needs or intentions.
- Multimedia
  - In addition to the press, TV and radio, the ubiquitous use of mobile-devices and abundance of cameras is leading to an increased production of multimedia documents (images and videos) producing information spanning different modalities.
- Multi-platform
  - A wide array of platforms exist which are used for communication during disasters. Users prefer different platforms to communicate in their daily lives and typically use the same (familiar) platforms for communication during disasters. However, the landscape of platforms is diverse (technically, geographically, socio-economically) and dynamic with new contenders entering the stage over time.
- Linking between Media and Platforms
  - Media and platforms are frequently used in combination, linking actors and contents across several platforms. The complete picture can only be obtained by the combined view rather than by individual views from solitary platforms. These are by necessity only a small window into a larger communication landscape.

Figure 7 below provides a schematic overview of these elements: multilingual, multimedia content from multiple platforms, spanning traditional as well as social media. All of these need to be taken into account for the creation of a generic framework, defining and structuring elements identified as being relevant on a high level of abstraction in order to provide an overview and classification of these elements and their interrelationships and spanning all relevant types of media and platforms for communication during disasters.

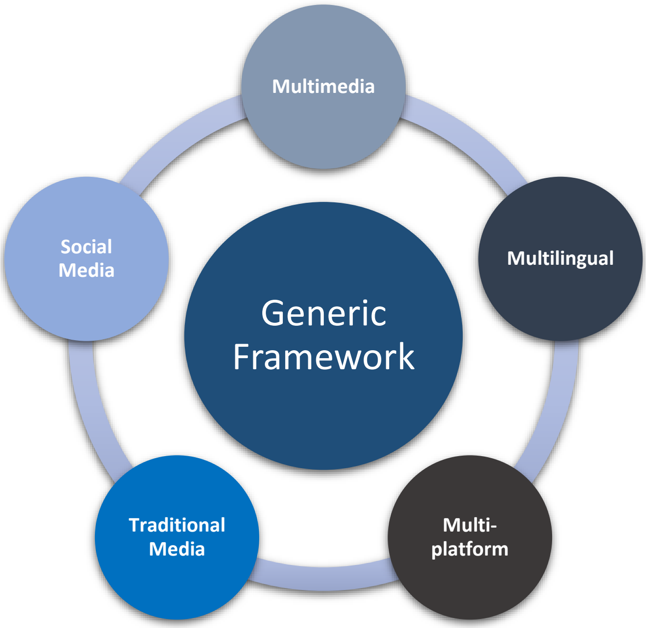


FIGURE 7 ELEMENTS OF DISASTER COMMUNICATION

Boundaries between traditional media and social media are becoming increasingly blurred resulting in *compounded documents:* documents and links between them producing a mash across platforms. The links between documents corresponding to different kinds of media and residing on different platforms provide the fabric for such *compound-documents* and play a central role in a generic framework.

Against the backdrop of an increase in the frequency and severity of natural disasters and a continuous extension and diversification of media and technologies, the goal of this thesis is to develop a generic framework accommodating the wide variety of data-sources and documents typically encountered in the communication during natural disasters and crises. The framework places emphasis on the commonalities across the different kinds of media and platforms, the possible links between the different types of sources, documents and actors and the resulting networks created by them. Its development is built upon insights gained in the KIRAS project QuOIMA, in particular concerning the use of actual platforms as well as findings from the data sets established within that project. By abstraction and generalization from these, as well as the extension to further types of documents, media and platforms, a more systematic approach is developed resulting in a generic framework applicable to natural and human-created disasters alike[27]. Focusing on the generic elements and methods this framework provides a reference against which emerging technologies and tools can be weighed and which can inform planning and management activities.

The development is expected to have an impact by scientifically investigating, completing and extending work performed within QuOIMA as well as in terms of practical implications for first responders (e.g. for use with guidelines, training, procurement, context of emerging technologies or development of communication strategies) and commercialization of the capabilities identified and implemented within the framework. The focus on the framework and methods rather than on individual platforms and tools (which may become outdated or discontinued) is expected to provide relevant input also for the foreseeable future beyond the completion of this thesis.

2.2 Research Gaps

In many ways, previous work - in particular in the field of open source information gathering and the QuOIMA project - prompted questions and exposed deficiencies and gaps of current approaches: existing approaches typically only target individual platforms, operate in a single language, on textual data only and do not take into account cross-referencing at all. These findings are supported by a recent study by Reuter et al. (2018a[28]) who examined the complete set of papers presented at the social media tracks of the past 14 ISCRAM conferences[29] spanning the period from 2004 – 2017. Whereas ISCRAM certainly does not encompass all efforts by communities on a global scale, it can certainly serve as an indicator and witness of the limited nature of many (if not most) efforts. Castillo (2016) likewise notes this *overreliance on a single data source* (Twitter) as a current research problem. From these investigations and further research on the state-of-the-art, it is notable that most research connecting social media with natural disasters has been carried out or at least been focusing on a limited set of aspects:

- a single language (mostly English)
- a single medium (text) and
- a single source (Twitter)

In addition, a large proportion of the investigated disasters concerned the US.

[27] In addition, a common terminology is developed for further work
[28] The author of this thesis is a co-author of the paper
[29] ISCRAM: Information Systems for Crisis Response and Management, http://www.iscram.org

The limited scope of social media processing in connection with natural disasters may be attributed to a number of factors (Backfried, 2013a):

- the ease of Twitter in accessing data and establishing data-sets
- the affected population may have preferred to use Twitter for communication
- internationally visible target organizations of such systems operate primarily in English speaking countries
- researchers are limited in the languages they focus on, sometimes ruling out portions of a collected data set (due to them not speaking the language and/or the absence of adequate technology and models)
- research programs may prefer to fund particular settings
- natural language processing (NLP) tools may support a limited set of languages only, in particular English
- automatic speech recognition may not be available for particular domains (and a wider range of languages)
- visual processing technology may not be available for a particular domain or may require resources on a scale not available to a project
- data sets may have been collected by entities in the English-speaking world
- data sets are available for a single medium only (no parallel corpora across different media)
- data may simply be accessible more easily for certain platforms than for others, particularly for Twitter (privacy and legal concerns naturally play a strong role)

The combination of social media processing with processing of traditional media content comprising the complete spectrum of media from textual, auditory and visual media-types, has not been addressed thoroughly in the scope of disasters and disaster management. The rich set of connections and links possible between traditional and social media has only been investigated to a very limited extent.

Twitter, although only used by relatively few citizens compared to other platforms like Facebook, has been the center of attention of the majority of studies (Kaufhold & Reuter, 2016; Reuter et al., 2018a, Reuter et al., 2018b). One reason might be that Twitter's open API, terms of use and the simplicity of 140/280-character messages allow for a convenient way of gathering data and establishing substantial data samples (Imran et al., 2015). Further reasons for the focus on particular platforms might be related to the location of the researchers themselves, of funding bodies and research programs and to a host of socio-economic factors influencing the platform use and distribution within affected populations (Backfried et al., 2013b).

A further limitation concerns the location of events and language of social media content. While the analysis of social media during crises has become commonplace in the US since about 2006 (Palen & Hughes, 2018), there is still relatively little research on this topic in Europe (Backfried et al., 2015) and elsewhere as well as on non-English speaking communities (Reuter & Kaufhold, 2018). Related studies, mainly on automatic information extraction from social media that can help disaster prevention and management, have been conducted in Japan as a country prone to earthquakes (Sakaki et al., 2010), and similarly in Qatar[30] (Imran et al., 2014).

Natural Language Processing (NLP) tools used to analyze data are generally (much) more advanced for English language content. They have frequently been developed with an *English mindset* and often provide the best performance for input data in English. As in the case of the focus on particular platforms, the location of research programs and the researchers themselves have doubtlessly had an influence on the targeted languages (Reuter et al., 2018a). Even though claims may be made as to the coverage of dozens of languages, these are often only supported to a lesser extent. The corpora, required to train statistical models for Machine Learning, are distributed very unevenly between different languages, again with English boasting most resources in this respect. This resembles the situation in several related areas such as Automatic Speech Recognition (ASR), Named Entity Recognition (NER) or Topic Detection (TD) which have all received substantial attention in the US, driven by DARPA sponsored evaluations such as the HUB, TDT or MUC evaluations (Fiscus & Doddington, 2002; Grishman & Sundheim, 1996; Przybocki et al., 1999). Extensions to further languages and groups of languages - each one exhibiting particularities in structure and use on several linguistic levels - often require substantial amounts of effort and time and up to this date remain a significant challenge (Backfried et al., 2013b). Dialects and colloquial registers typically encountered on social media pose further challenges[31]. Languages with scarce resources, frequently referred to as *low-resource languages*, have more recently become the focus of attention of projects and at conferences such as LREC[32] and Interspeech[33] , leading to gradual improvements on corpora and language specific technologies. Advances in deep-learning and transfer-learning (Pan & Yang, 2010)

[30] Qatar Computing Research Institute, QCRI, https://www.qcri.org
[31] Dialects, colloquial or vernacular versions of languages may not possess standard spellings, making processing of texts even more complicated and sometimes close to impossible.
[32] The International Conference on Language Resources and Evaluation, http://www.lrec-conf.org
[33] http://www.isca-speech.org/iscaweb/index.php/conferences/interspeech

paired with the availability of numerous corpora can be expected to further aid in the extension of NLP technologies to such languages.

The majority of research in the field of disaster communication has been carried out on textual data - rather than audio, images or video - leading to a narrow focus on text-processing capabilities and mechanisms and limiting the possibilities to this media type. However, in the age of ubiquitous cameras on smart-phones and mobile devices, the amount of visual content is growing tremendously and may indeed provide valuable information: during Hurricane Sandy (2012), more than 1.3 million images were posted (Taylor, 2012)[34]. Users commenting and referencing of images and videos, their linking of images and connecting visual and textual content, provide additional valuable information for analysis and situational awareness. This rich pool of associated visual and textual data so far has only been tapped marginally for the disaster domain.

Whereas the concept of *cross-media communication* has been examined from an information-dissemination point of view (Davidson, 2010) and been promoted primarily for marketing purposes (Winer, 2009), the reverse side of the coin – the collection, processing and analysis of information in a cross-media manner – has only received little attention in the past. To date, in the domain of disaster communication, this phenomenon has not been systematically examined at all. The connections and interplay between traditional media and social media form a rich information landscape which has been acknowledged to provide valuable information during disasters. To this end, the QuOIMA project prompted the development of individual technologies and components (Backfried et al., 2013a, Backfried et al., 2014, Backfried et al. 2017). However, its goal was not to investigate the generic mechanisms underlying those components nor to produce a systematic model for the elements involved in disaster communication. Whereas the project did accomplish to combine traditional and social media in the disaster context, it did not evolve into a generic framework or approach.

Such a generic framework – a model of the elements involved, their interplay and the operations possible on them and a method to derive them – provides a generalization of previous approaches and efforts having taken place independently. Especially in the dynamic and fluid media-landscape comprising traditional and social media, the absence of a generic model means that commonalities between existing platforms cannot be capitalized upon, new platforms cannot easily be accommodated and the linking of information between the platforms remains largely unused. The lack of such a generic model encompassing the various elements of disaster communication and

[34] Hurricane Sandy is often regarded as the first large disaster where visual social media played a central role.

including the links and combinations of these elements thus forms a gap in the current research landscape. Furthermore, no principled approach to develop such a framework has been developed to date.

2.3 Expected Contribution and Impact

Several of the questions, findings and developments made within QuOIMA serve as a basis for the current thesis. However, whereas this prior work prompted the development of individual technologies and components triggered by a singular use-case, it did not go beyond these individual cases and did not evolve into a general framework. This thesis thus aims to extend those results, build upon the insights gained and take efforts one step further. It systematically investigates, extends and unifies prior results, taking into account the developing state-of-the-art and current research, systems and architectures and establishes a generic framework able to accommodate all elements involved. In the process, previous insights and results are revisited, updated and put into the context of the framework developed here. Work proceeds in line with the strategy of open source information gathering and –processing, focusing on openly available data[35].

Presthus and Munkvold (2016) discuss how research may contribute to knowledge in the field of information science (IS) and how potential contributions may be framed. As presented in Figure 8 below, they postulate that (knowledge) contributions should involve theoretical contributions (*the Area of concern*, *Academia's call for research on a subject*) as well as practical contributions (*to a real-world problem)*. Research is motivated by theoretical and practical factors, and consequently its contributions also should have theoretical as well as practical implications and contribute to the underlying knowledge base as well as to the target environment (Hevner et al., 2004). The framing of the research questions itself is influenced by the contributions as may the method by which research is conducted.

[35] Sources which are publicly available to anyone though possibly not free of charge.

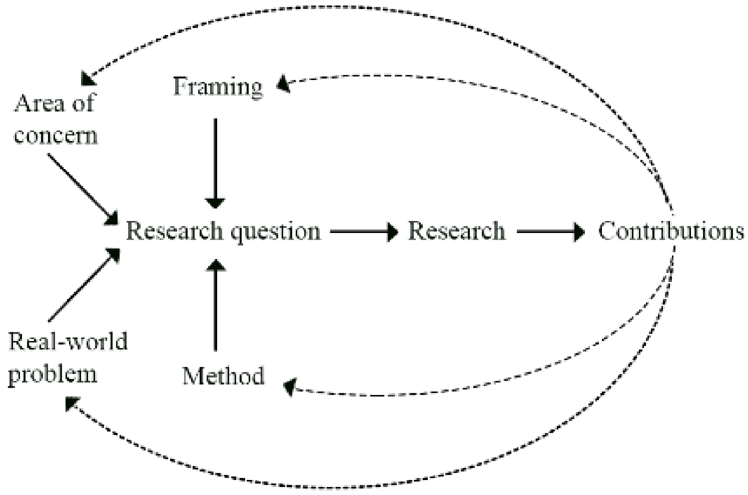


FIGURE 8 STYLE COMPOSITION FOR RESEARCH PROJECTS[36]

Presthus and Munkvold (2016) furthermore suggest a taxonomy of contributions[37] of which the following kinds of contributions can be identified to apply to this thesis:

Contributions to practice: Lessons Learned and Guidelines
Contributions to theory/science: Frameworks and Models

This thesis is expected to provide a theoretical/scientific contribution by investigating how a model can be built which overcomes the shortfalls of the above identified gaps and satisfies the identified requirements. It proceeds by systematically extending prior work, producing a framework contributing to a better understanding of the kinds of elements involved and their connections in disaster communication. Following a common perspective from computer science, it divides the overall problem of framework creation into the development of a data model, a process model and a method to derive these models (the term *framework* thus denotes the models themselves as well as the method to derive them). The data model describes the relevant elements, their properties and connections whereas the process model describes how these elements are created and processed. The thesis aims to produce such a combined model, and doing so, to produce a new and innovative *artifact* (Hevner et al. 2004).

Furthermore, this thesis is expected to provide a practical contribution in terms of lessons learned from past work and work building upon it, insights and implications for first responders, for use with

[36] Figure from Presthus & Munkvold (2016)
[37] Chapter 4 of Presthus & Munkvold (2016)

guidelines, procurement, training and strategic planning. This is complemented by insights for IS practitioners on how to accommodate new functionalities when devising systems for disaster management.

In the sense of Design Science Research (DSR), an *innovative technological artifact* (the framework/model) is developed as well as an *instantiation* representing an improvement to existing approaches and solving a real-world problem of relevance to an application domain and its practitioners (Hevner & Chatterjee, 2010). Real-world cases provide the background for exemplary application of the models developed, connecting them to the application domain of disaster communication and demonstrating their practical applicability and relevance.

The overall contribution of this thesis aims to answer the question of how to construct a model overcoming prior shortcomings and satisfying the requirements collected from end-users and demonstrates how such a model can be constructed and applied to real-world data. Its contributions are thus expected to be of a theoretical as well as practical nature.

2.4 Research Questions

The above analysis of current gaps emphasizes the noticeable limitations of current approaches, e.g. the types of media or the contents which can be represented and processed within the scope of disaster communication or the *isolated manner,* in which these have been investigated and modelled in this highly relevant domain. Particularly in a time of increasingly rich media representations and an extensive and diverse landscape of platforms, these limitations seem remarkable and hint at a multitude of untapped possibilities to capitalize on commonalities and combinations. These gaps and the motivation discussed above prompt the question whether a (generic) model can be devised which does not suffer from the same constraints and how to construct such a model in a principled manner. The nature and scope of the model and the manner of how it can be constructed thus form the central topics of this thesis and are reflected in the research questions below.

The over-arching **research question** (**RQ**) of this thesis can be formulated as follows:

How can a model be built which represents the elements and processes involved in cross-media disaster communication, such that all relevant components and technologies found within the set of relevant platforms and media and their interplay can be represented in a generic and extensible manner, in a way which overcomes current limitations and satisfies the identified needs of end-users of the target domain?

This research question can be divided into a set of **sub-research-questions** concerning the structure of the model itself as well as the methodology to create it:

- **RQ1:** What should a framework for cross-media disaster communication consist of and which elements should it contain?
- **RQ2**: What are the qualities and features of such a generic framework and how should it be constructed?
- **RQ3**: Which platforms and media should be accommodated in the framework?
- **RQ4**: Which components and functionalities are required and how are they to be combined?
- **RQ5**: Which types of cross-media communication exist and how can they be characterized?

These sub-research-questions will be addressed in the work below and their combined analysis be employed to answer the primary research question.

In addition to the above, two secondary research questions - which are of particular practical relevance - will be examined:

- SRQ1: Which phases in the disaster-cycle can be covered by the capabilities offered by such a framework?
- SRQ2: Are the patterns static or do they change dynamically during the different phases of a disaster?

2.5 Methodology

As outlined above, the project QuOIMA provides the backdrop for the current work. This project resulted in a specific model and a set of insights which the current thesis can build upon. These entail knowledge about the field of disasters, the state-of-the-art about social media platforms, technical and social aspects of their use and the contact to practitioners providing crucial feedback. The author participated in that project, working on the creation of this specific model and its implementation in a deployed system. This work, the experiences made and feedback from using the implemented elements in practice as well as the author's professional experience led to several publications and inform the present thesis. However, they only provide a *starting point* for the present research, which aims to go beyond this prior work. Its goal is to develop a generic framework defining and structuring the elements and interactions identified as being relevant on a high level of abstraction. The framework encompasses a set of models, accommodating the relevant elements and operations, as well as a method to create these models, following a principled approach and using a consistent terminology for all elements involved. Figure 9 below provides a schematic view of the constituents of the framework (its development is described in chapter 5).

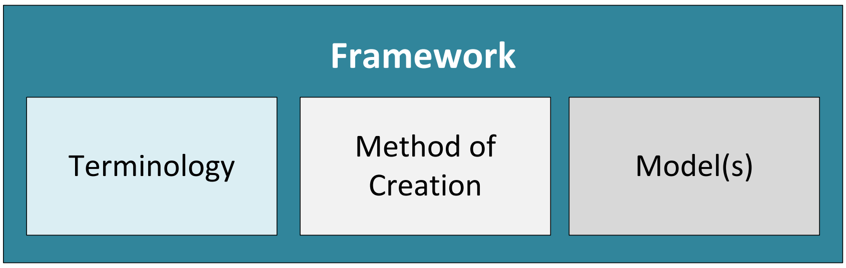


FIGURE 9 FRAMEWORK - CONSTITUENTS

In establishing this framework, top-down and bottom-up elements are combined for the current work:

- Top-down: starting from the insights and requirements of first responders as well as guidance of literature research, the state-of-the-art and developments on the topic of natural disasters and disaster management
- Bottom-up: driven by the existing platforms employed during natural disasters and citizens' use of these platforms, data collected in a running system (the Media Mining System) as well as technical capabilities, in particular in the area of natural language processing

Figure 10 below provides an overview of the wider context and the general approach taken within the current thesis. The Generic Framework depicted at the center and the method to develop it form the focus of the current research.

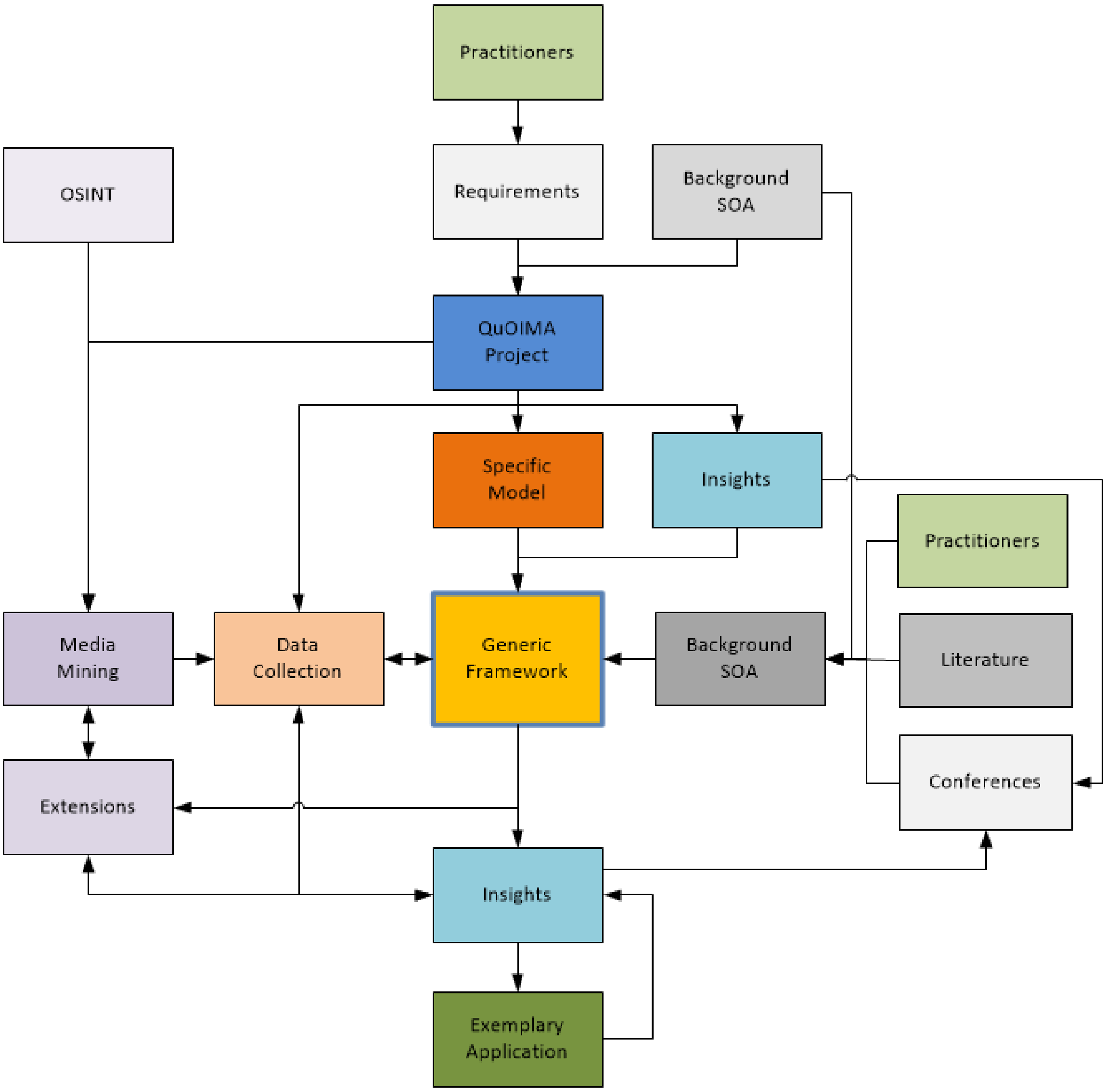


FIGURE 10 METHODOLOGY AND APPROACH

In the spirit of Design Science Research (Hevner et al., 2004; Hevner & Chatterjee, 2010; Vaishnavi et al., 2017), research takes place in an iterative manner, is related to a relevant application domain, and both, draws from and contributes to knowledge. Hevner outlines these research cycles as depicted in Figure 11 below.

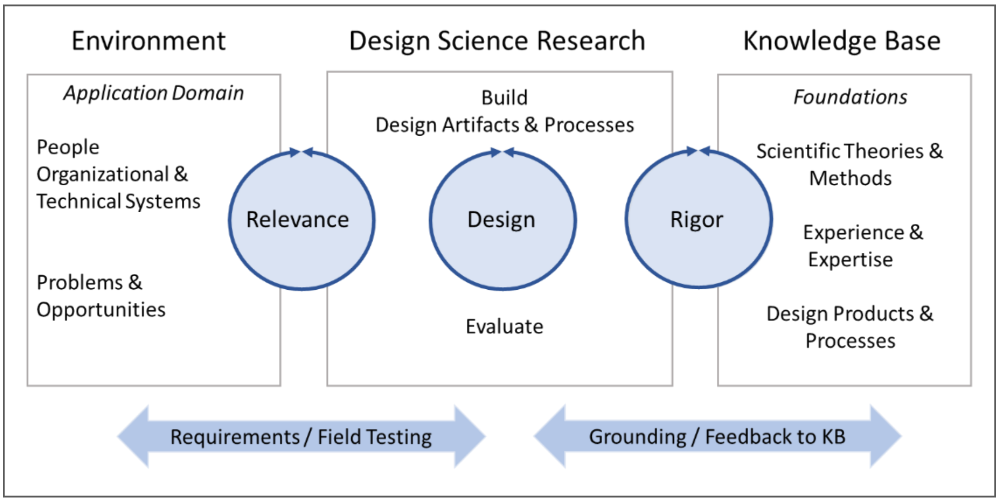


FIGURE 11 DESIGN SCIENCE RESEARCH CYCLES[38]

Comparing the chosen approach for this thesis from Figure 10 with the above figure, several common activities and key-elements can be identified:

- Incremental approach of framework and model creation, extending a specific model to a generic one in several iterations, refining and extending the model at each step (design)
- Information from a prior project involving first responders, connecting the work to a relevant problem in an application domain of current interest to practitioners (relevance)
- Drawing upon existing research, an implemented system, practice and prior projects, communicating and publishing results, providing feedback to the community (rigor)

Within this thesis, an *artifact* (in this case a *model*) relevant to a real-world application domain is developed. The process is carried out in an iterative manner with feedback-cycles regarding the relevance, design and grounding, with the ultimate goal to construct a framework and generic model. The model itself as well as the process to develop it represent improvements to existing research and solve a *significant* problem. The environment, consisting of the end-users (practitioners) and existing technology inform the research process, providing relevant input and ensuring practical applicability of the current work. Prior work of the author, the state-of-the-art, literature review, conference participation and work with peers provide a sound knowledge-base from which the author can draw for the development of the models. All of these steps are undertaken in several iterations, considering the above listed top-down and bottom-up elements, and in a continuous manner during the work on this thesis. Finally, the developed models are examined and evaluated against a set of standard

[38] Figure adapted from Hevner & Chatterjee (2010)

criteria. To a large extent, the methodology thus conforms to the Information Systems Research Framework laid out by Hevner et al. (2004).

The design and development of the generic framework and its grounding in the relevant scientific context is the focus of this thesis. Establishing such a model has scientific implications as it goes beyond prior work, contributing to knowledge in the field, as well as practical implications as it addresses a problem identified as relevant by several stakeholders and furthermore lends itself to actual implementation.

The state-of-the-art (SOA) as well as end-user requirements collected during previous work provide input to the respective work within the current thesis: this is extended by a literature search, conference participation, personal communication with experts as well as an overview paper investigating the use of social media during disasters of the past 15 years presented at the ISCRAM 2018 conference (Reuter et al., 2018a[39]). In addition, the end-user requirements were refined through discussions at various conferences, by presentations to the Association of the Austrian Fire Brigades[40] and through work with the Austrian Red Cross as well as the National Defence Academy of the Austrian Armed Forces[41].

Whereas the manner of traditional media coverage of natural disasters can be assumed to be rather constant, social media use is highly fluid and thus needs to be monitored constantly. To this end, the state-of-the-art is kept up-to-date in a continuous manner. This is accomplished by repeated literature searches as well as by the regular participation in conferences including in program committees and performing reviewing tasks. Regarding traditional media, the author was able to participate in an inter-disciplinary[42] research project with the Department of Communication of the University of Vienna, being able to exchange ideas and verify elements of the chosen approach[43]. Furthermore, due to the previously established contact to the Austrian Red Cross as well as to the Austrian Armed Forces, continuous relevance to practical issues could be ensured.

Due to the inter-disciplinary nature of the field and the variety of platforms and proprietary naming schemes, a common communication model and terminology needs to be devised. This communication

---

[39] The author of this thesis is a co-author of the paper

[40] Österreichischer Bundesfeuerwehrverband (Austrian Fire Brigade Association)

[41] The Austrian Armed Forces render assistance in case of natural catastrophes and disasters in Austria

[42] *Inter-disciplinarity* as defined by http://www.arj.no/2012/03/12/disciplinarities-2 on 2019/03/24

[43] Project COMRAD, „Kommunikationsmuster der Radikalisierung“, https://empcom.univie.ac.at/forschungsprojekt-radikalisierung-comrad/ on 2019/02/27

model is based on an established model of communication science and allows to map all relevant items involved onto a common set of elements and attributes[44]. Based on the communication model, the generic model is developed in an iterative manner refining features with each additional step (integration of a further platform and medium). As such, the resulting terminology by itself may also be useful for future inter-disciplinary work.

Adopting an established approach from computer science, model development is divided into the development of a data model, describing the data and its organization, and of a process model, describing operations such as ingestion and manipulation (Simsion & Witt, 2004). Within this approach, all elements involved in communication are represented by the data model whereas the processing steps applicable to them are represented by the process model. The two models describe static and dynamic aspects of the domain on a conceptual[45] level and are developed *hand-in-hand*.

In order to be able to demonstrate the relevance of the chosen approach, selected parts are implemented and applied to *real-world events*. This leads to (further) data being collected and processed which serve as exemplary evidence for the practical applicability of the generic model and allow to gain further insights. These processes and the resulting structures are examined from a qualitative and quantitative point of view. The proposed approach and developed models are evaluated against a set of established criteria as defined by Moody (2003) and Prat et al. (2014).

2.6 Author's Own Publications Related to the Current Thesis

Prior to this thesis, the author worked on the research project QuOIMA. As pointed out above, insights and experiences from this project inform the current thesis as well as providing the basis for the original work developed within this thesis. Several publications were made by the author (as primary author and as co-author) during the period of QuOIMA as well during the work on this thesis. These publications investigate ideas related to the current thesis. Some of them proved to be dead-ends, others side-tracks and yet others paved the way and developed ideas for the current thesis. Selected parts of these publications were incorporated in the current thesis where appropriate and relevant. They mainly concern chapters 1-4. Output and insights from these papers also influence the research questions and findings of this thesis. All of the sections where prior work was re-used or integrated are marked with references and footnotes.

[44] The terms *attribute* and *property* are used in a synonymous manner within this thesis.
[45] The term c*onceptual* refers to a technology-independent, high-level, semantic specification.

The following publications of the author are of relevance to the current thesis:

Backfried G., Aniola, D., Mak, K., Pilles, H.C., Quirchmayr, G., Winiwarter, W., Roth, P. (2010), Multimedia Documentation Lab, 10th International Conference on Knowledge Management and Knowledge Technologies, Graz, Austria, September 2010

Backfried G, Schmidt, Ch., Pfeiffer, M., Quirchmayr, G., Glanzer, M., Rainer, K., (2012) Open Source Intelligence in Disaster Management, EISIC 2012, European Intelligence and Security Informatics Conference, Odense, Denmark, August 2012

Backfried, G., Göllner, J., Quirchmayr, G., Rainer, K., Kienast, G., Thallinger, G., Schmidt, Ch., Pfeiffer, M., Meurers, Ch., Peer, A., (2013a) Cross-Media Analysis for Communication during Natural Disasters, IAIT 2013, International Conference on Advances in Information Technology, December 2013, Bangkok, Thailand

Backfried, G., Göllner, J., Quirchmayr, G., Rainer, K., Kienast, G., Thallinger, G., Schmidt, Ch., Peer, A., (2013b) Integration of Media Sources for Situation Analysis in the Different Phases of Disaster Management, EISIC 2013, Uppsala, Sweden

Backfried, G, Prinz, K., Göllner, J., Quirchmayr, G., Czech, G., (2014) Cross-media, Multimedia, Multilingual Communication in Open Sources During Crises and Disasters, ISCRAM 2014, University Park, PA, USA

Backfried, G., Schmidt, Ch. and Quirchmayr, G., (2015) Cross-Media Linking in Times of Disaster. In Proceedings of the Information Systems for Crisis Response and Management (ISCRAM). Kristiansand, Norway

Backfried, G., Kais, I., Quirchmayr, G., (2016) Towards a Generic Data-Model for Cross-Media Communication During Disasters and Crises , The 3rd International Conference on Information and Communication Technologies for Disaster Management, ICT-DM 2016, 13-15 December 2016, Vienna, Austria[46]

Backfried, G., Schmidt, Ch., Aniola, D., Meurers, Ch., Mak, K., Göllner J., Peer, A., Quirchmayr, G., Czech, G. and Glanzer, M., (2016) A General Framework for Using Social and Traditional Media During Natural Disasters: QuOIMA and the Central European Floods of 2013, in Rogova, G. and Scott, P (Eds), Fusion Methodologies in Crisis Management, Springer

The following publications have been co-authored by the author of this thesis:

Bailer, W, Kienast, G., Thallinger, G and Backfried, G., (2016) Data Fusion Across Traditional and Social Media in Fusion Methodologies in Crisis Management, Rogova G and Scott, P, Eds, Springer

Bara, G., Backfried, G., Aniola, D., (2019) Fake or Fact? Theoretical and Practical Aspects of Fake News, in Information Quality in Information Fusion and Decision Making, Bossé, Éloi, Rogova, Galina (Eds.), Springer

Reuter, Ch., Backfried, G., Kaufhold, M-A and Spahr, F., (2018a) ISCRAM turns 15: A Trend Analysis of Social Media Papers 2004-2017, ISCRAM 2018, Rochester, NY, USA

[46] Received Best-Poster Award at ICT-DM 2016 conference

2.7 Overview of Chapters

The overall structure of this thesis is divided into two sections: chapters 1-4 provide the background on natural disasters, the role that the media play in their coverage, the results of prior work of the author and the motivation and strategy for the writing of the current thesis. The second section, comprising chapters 5-7, deals with the (method of) construction of a generic data and process model, its evaluation against a set of standard criteria and application to selected cases collected during actual disasters as well the examination of the current thesis' contributions and possible future alleys of research.

Chapter 1 outlines the state of disasters in recent years: due to climate change, natural disasters have become more frequent and more severe, having far-reaching and profound impacts on the affected populations as well as widespread economic ramifications. An active media landscape - dominated by different mixes of traditional and social media - creates the shape and perceptions of disasters, forming an essential part of communication for affected populations, official institutions and first responders. Both, disasters and communication about them frequently span diverse geographic regions and multiple languages or dialects. Communication takes place in real-time; media are typically employed in a mixed, mashed-up way by eye-witnesses and officials alike. Consequently, any model aiming to accommodate this type of communication in a generic manner needs to take these factors into account.

Chapter 2 describes the motivation and approach chosen for this thesis. The initial motivation for the current work is provided by the developments covered by Chapter 1, by experiences and insights gained during a research-project the author participated in and by the absence of a generic model encompassing the different elements of disaster communication. The lack of such a model became obvious during and after that project; limitations of activities and research in the field became apparent at many conferences. To respond to these shortcomings, an overall approach guided by Design Science Research is adopted, reflecting the iterative nature of the approach for model developed and evaluation.

Chapter 3 discusses the interplay of media and disasters in more detail, discussing changes in the media-landscape, their role during disasters and in media-consumption in recent years. Social media have become a central element in news consumption in general as well as during disasters. This concerns the affected populations as well as first responders and officials. Anyone communicating during disasters will need to be present on such platforms for gathering and dissemination of information. At the same time, traditional media continue to play a strong role, especially in remote areas and need to be taken into consideration.

Chapter 4 presents relevant insights from the QuOIMA project. A corpus spanning the communication across several media serves to emphasize the importance of cross-media communication and provides support for the motivation of the current thesis. Data from this corpus is used for quantitative insights in chapter 6. Furthermore, relevant results regarding practical aspects of the project are discussed.

Chapter 5 deals with the development of the generic data- and process model. Due to the multi-disciplinary nature of the domain, some terminology is established first and a communication model from communication science is adopted to serve as the basis for model development. The construction of the data model takes into account the key elements of this communication model and investigates all relevant platforms with regard to these elements and their properties. The linking of elements is considered explicitly and separately as it provides an extension to the selected communication model and is of utmost importance for disaster communication. Starting from a high-level perspective of processing, the development of the process model considers the elements of communication and how they and their contents are to be processed and inter-linked. A set of technologies required to perform this processing is aligned with the individual elements and their attributes. A property-graph model is selected as the appropriate fit for implementation of the model.

Chapter 6 examines the model in a qualitative and quantitative manner and provides several real-world cases demonstrating its practical applicability. Quantitative insights provide support for the chosen cross-media approach and for the motivation of model-development. A set of standard evaluation criteria is identified and applied to evaluate the model in a qualitative manner, determining to what extent it achieves its ultimate goal: to be a viable generic model for the elements and operations involved in cross-media communication during natural disasters. Finally, a set of increasingly complex examples is provided demonstrating practical applicability.

Chapter 7 provides an overview of the research carried out, summarizing the insights gained through the present thesis. Results are discussed by revisiting the research-questions in light of the work performed, examining the contributions made on a practical as well as a scientific/theoretical level and presenting the conclusions of the current work. Furthermore, possible alleys for future research are described.

3 News, Media and Disasters

3.1 A Changing Media Landscape

News – its production as well as its consumption – is not the same today as it was in pre-satellite TV times when people waited for their morning papers or sat down in the living-room at a fixed time for the evening news on TV (Alejandro, 2010). Nor is it the same as in pre-Internet times, as nowadays an *always-on attitude* and a proliferation of mobile devices paired with high-speed networks allow people to tune in to events happening anywhere on the globe in a continuous 7x24 cycle. Users have become accustomed to *being online* – for social interaction as much as to obtain news (and to *stay connected*) (Alejandro, 2010) - and with it have adopted new habits, expectations and requirements when searching for news and information. The advent of Web2.0 (O'Reilly, 2005) and subsequently of SM in the early 2000s announced a new era of news consumption and production – for daily life as much as during disasters. The changes in technical and organizational possibilities suggest a distinction between *traditional media* (TM) – the ones dominant before this period, such as newspapers, TV, radio and news agencies – and *social media* (SM) on the other hand. Traditional media are available in physical as well as digital formats and include purely digital distribution mechanisms such as RSS-feeds. In contrast, SM are typically available exclusively in digital form with social interaction and communication aspects forming key elements. Certain features, such as a 'letter to the editor' or its digital re-incarnation, the comments section to a newspaper article, form hybrid elements which indicate that the distinction between traditional and social media is not a clear-cut and unique one but rather gradual and fluid. The modern media-landscape is characterized by a blurring of the boundaries between print, broadcast, and digital media, and increasingly between and among legacy news organizations, social media platforms that aggregate both professional and citizen reporting, blogs, and other non-traditional media outlets (Greenberg & Scanlon, 2016). Technical developments were accompanied by the rapid adoption of new channels for distribution of existing media entities and the introduction of new breeds of distributors and news-aggregators, such as The Huffington Post[47] or Buzzfeed[48] who entered the market during the same time-frame. These institutions add to the mix of traditional professional news producers and a myriad of individuals active on SM.

Such changes alter the terrain in which sources shape coverage, the possibilities for how journalists and citizens tell stories, and also how the public learns about an incident or disaster. While media have

[47] https://www.huffingtonpost.com/
[48] https://www.buzzfeed.com/

arguably never been as fractured and fragmented as they are today, they have also never been more omnipresent (Greenberg & Scanlon, 2016). Bearing in mind that the digital revolution is full of contradictions and exceptions, that countries and technologies started in different places and are moving at different paces, hindered or fostered by political and technical conditions, this is to be expected (Newman et al., 2017). Even though the distinction between traditional media and social media may not be clear-cut, the two terms serve to denote the extreme ends of the spectrum and allow to separate relevant aspects which are of importance from an information-gathering as well as a technical-processing perspective.

Although the tendency to consume news via SM and mobile devices has been growing tremendously and one might be tempted to regard SM as *the* primary source of information, this does not hold true in general and depends on a series of factors. On a global scale, access to SM may be limited to only a subset of a population by socio-economic, religious or political factors. Following a disaster, the infrastructure and electricity supply might be damaged and inhibit people from connecting to SM. Differences in geography and age are likewise major factors of influence for news consumption and choice of source. Increasingly, SM are not regarded as something new or separate, but rather have become just another part of people's every-day mix of media in a complex media landscape, in which the boundaries between traditional media and social media are becoming less and less pronounced. For example, 8 out of the 10 most active Twitter users reporting on an earthquake in Chile in 2010 were found to be directly related to the mass-media (Mendoza et al., 2010). Likewise, TV-stations receive input about storm damage from SM which is then broadcast immediately on TV (Holmes, 2011). In other cases, such as the Queensland floods (Bruns et al., 2011) the SM themselves became the primary source and a driver for the mainstream media, as affected individuals and authorities used them extensively for communication.

According to the Reuter's Institute for Digital Journalism's Report for 2017 (Newman et al., 2017), a consistent pattern of growth for news consumption via social media has been observed for many countries. However, TV and online news still play a significant role (see Figure 12 below for the US) and are expected to do so also in the foreseeable future.

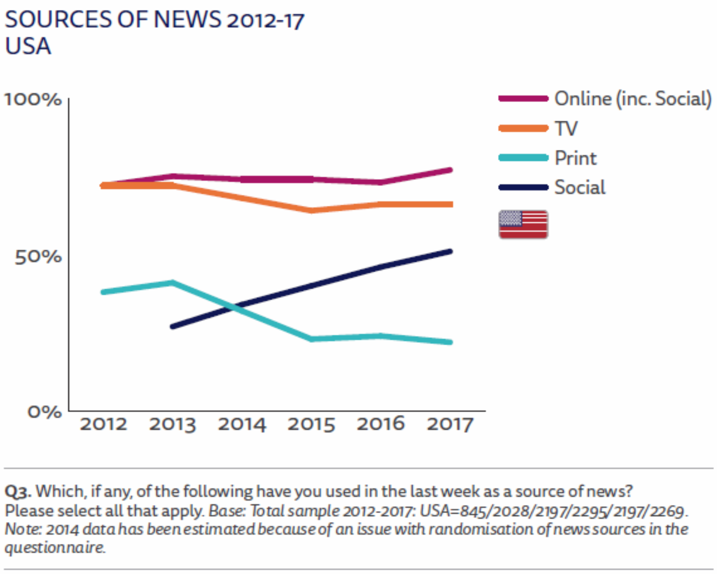


FIGURE 12 SOURCES OF NEWS - USA[49]

Radio likewise still plays a central role in many regions of the world, particularly in those where TV and SM are less present. A prominent example of the importance of radio is the Ebola outbreak in West Africa in 2014, where vital information was broadcast via radio networks (Greenberg & Scanlon 2016). But even in the US, more listeners still turned to radio (93%) for news in 2017 than watched TV or used smart-phones (Nielsen, 2018). Figure 13 below depicts the situation and the reach of various types of media in the US in Q2 of 2018.

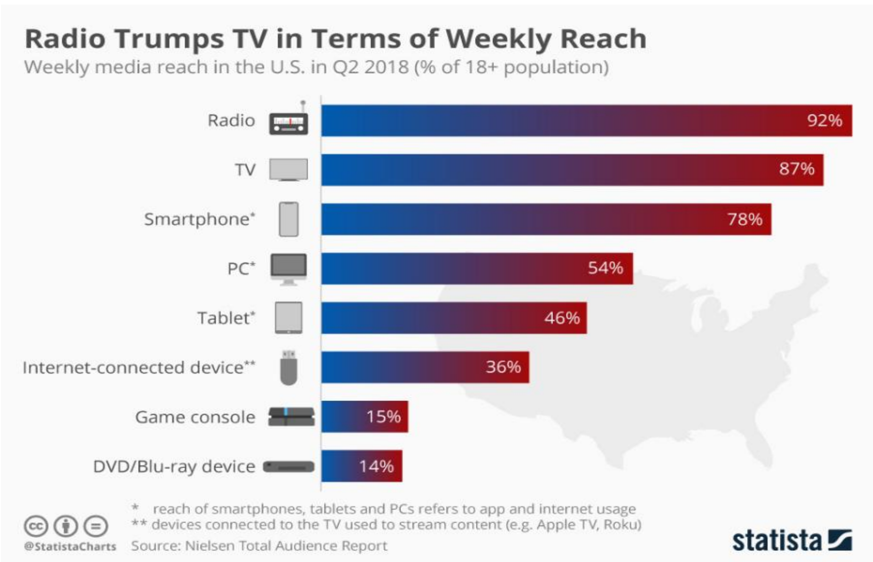


FIGURE 13 IMPORTANCE OF RADIO AND TV[50]

[49] Figure from Reuters Institute Digital News Report 2017, Newman et al 2017
[50] Figure from Statista, https://www.statista.com/chart/3589/weekly-reach-of-media-platforms/ on 2019/05/10

A significant split in the main source for news can be observed across age-groups, with younger segments being much more likely to rely on social and digital media, and older groups tending to rely more on TV, radio and printed media as depicted in Figure 14 below (a phenomenon often referred-to as the *age-gap* or *generation-gap* of media).

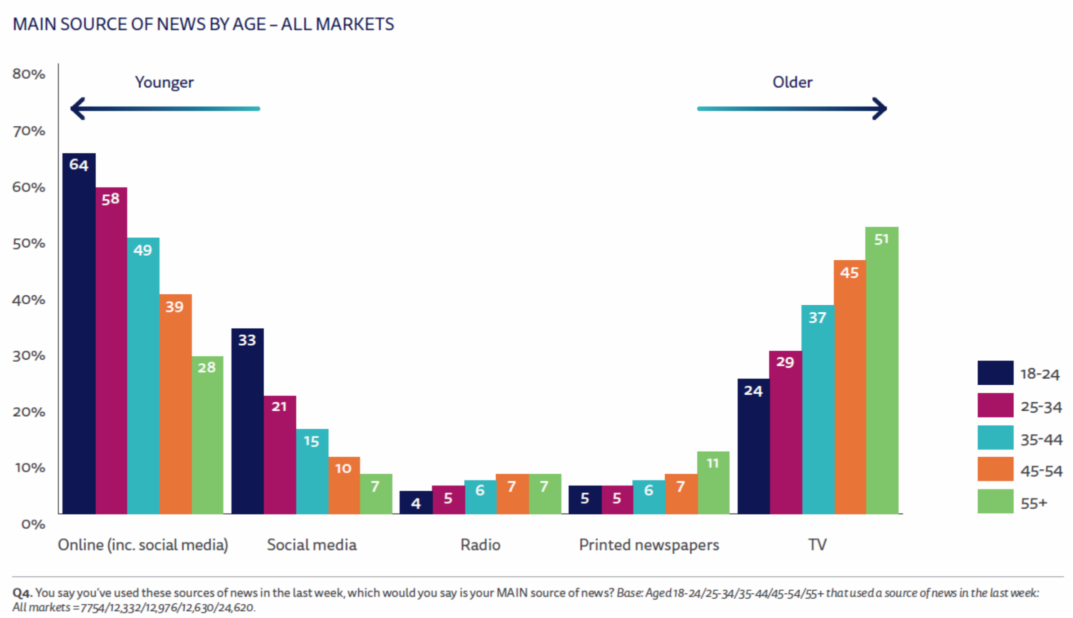


FIGURE 14 MAIN NEWS SOURCES ACROSS AGE-GROUPS[51]

Based on these findings, it can be assumed that neither traditional media nor social media alone provide the exclusive source for the dissemination of information in general nor during disasters. Both types of media are likely to be used as information sources by individuals as well as by organizations. Both types have their advantages and merits and need to be considered when trying to establish a clear picture of events and devising strategies to deal with an uncertain and complex situation. Consequently, any comprehensive strategy or generic framework needs to consider both, traditional as well as social media and take into account that mixes of media will be used depending on the particular qualities of an event.

[51] Figure from Reuters Institute Digital News Report 2017 (Newman et al., 2017)

3.2 Media Coverage of Disasters[52]

In describing news-coverage of disasters, the terms *news* and *information* are to be interpreted in the context of mass-media communication and are used largely interchangeable within this thesis. They refer to the communication of relevant facts (resp. to relevant *information*) regarding events of interest affecting a population. In case of disasters[53], they specifically refer to relevant information - novel or already known - regarding an incident of immediate interest to individuals, organizations or authorities. From the standpoint of the disaster cycle, the phases immediately before, during and following an impact are most likely to be covered by the media (as opposed to quiet phases of diligent preparation or long-lasting reconstruction, where media attention is typically limited, decreases over time or is simply even non-existent). The event under consideration is thus typically one about to take place, still taking place or one which has only recently taken place - a *new event*, a current emergency, rather than a past event being investigated in retrospect. Information in this context is hence often associated with *breaking news* and recent developments but equally includes updates of known information. This temporal proximity affects real-time communication behavior as well as influencing information gathering strategies and possibilities.[54]

On the one hand, the affected population, first responders and authorities dealing with the disaster are consumers of these news, trying to extract relevant information and facts with the aim of increasing situational awareness (Endsley, 2003) or simply trying to *make sense* of the situation as events unfold (Muhren & Van de Walle, 2009). On the other hand, they are also active producers of information - on-the-ground, as eyewitnesses or from a distance, institutionalized via press-releases and press-conferences in an interplay with the media – and in turn contribute to the news about the incident.

The fundamental mechanisms to generate, spread and receive news in such situations can be assumed to be largely identical as for news during *normal times* as individuals and organizations can be expected to use identical sources, platforms, distribution mechanisms and formats during regular operation as well as during disasters. The emphasis, however may shift to more resilient mechanisms,

[52] This chapter contains material previously published by the author in chapters 10 and 22, “Data Fusion Across Traditional and Social Media” (author of present thesis is author and co-author respectively of these book-chapters) see Backfried et al., 2016 and Bailer et al., 2016, also see chapter 2.6 above.

[53] In contrast to disaster, the term crisis will be interpreted to be relevant primarily to an organization (Coombs and Holladay, 2012).

[54] Likewise, it imposes limitations on the disaster management and -communication strategies. However, these will not be of concern in the present thesis and are addressed extensively in the literature.

such as radio broadcasts, e.g. in case of infrastructural damage. Specific actors and strategies may differ according to the kind of event and the media attention generated by that event. Overall, people can be expected to turn to the platforms they already know from their daily lives also in times of disaster – they will already be knowledgeable and familiar with the use of these platforms rather than wanting to learn and adapt to new platforms in times of distress and emergency (Haddow & Haddow, 2014)[55]. This does not only apply to SM, but also applies to traditional media, in particular radio, which remains one of the most popular forms of communication during disasters.

As disasters unfold[56], communication about them will spread rapidly. Information and content will be shared and referenced, (re-)packaged and assembled, in the course quickly encompass different media and span multiple platforms as individuals and news organizations report first hand and from afar.

Disasters have been covered by mass-media since the beginning of their existence. The geographical location of disasters and the identities of victims, the increasingly visual nature of disaster events, the ready accessibility of eye-witnesses and first-hand accounts and the pervasive nature of media in our lives, all shape and influence which kinds of emergencies attract global media and public attention. There is nothing inherent about these events that give them meaning - they occur in a real, material world; but for most of us, our experience of these events is shaped and determined in large part by our interactions with media industries, institutions, and technologies (Greenberg & Scanlon, 2016).

When disaster strikes in distant places, established and start-up news outlets quickly become the primary sources of information about events as they unfold (Franks, 2009). Social media platforms, such as Facebook, Twitter or Instagram, abound with reports and posts as locals publish images and messages from the ground. Even with the rise of social media and start-up news outlets like Buzzfeed, mainstream media continue to provide a pivotal role of validating and providing a coherent, reliable gatekeeper to the information about such events (Greenberg & Scanlon, 2016; Franks, 2013).

Traditional mainstream media (i.e. the press, corporate and state broadcasters, news agencies) have played and continue to play an important role in disasters, with respect to early signaling, problem framing, and definition (Haddow & Haddow, 2009). They play a critical role before, during and after an incident, are essential for warnings to be effective and may be the single most important source of

[55] The actual contents and topics, however, will be dramatically different in times of disaster.
[56] Sudden onset disasters or slow onset disasters alike

public information in the wake of a disaster (Scanlon, 2009). In doing so, they have assumed myriad and complex roles in supporting emergency response and notification, and in helping constitute a shared experience (Greenberg & Scanlon, 2016). They also function as gate-keepers on all levels of the media-universe – influenced by journalistic values and practices and economic considerations - deciding which stories merit coverage and which information is conveyed. Even in a complex media-universe, TM often still provide an important and highly additive source of information. In certain regions of the world, they remain the primary source for large parts of the population. However, under certain circumstances, e.g. cross-border or international settings of crises, official and traditional channels may not be available to provide comprehensive information or only be able to – or willing to – provide partial information.

Social media allow professionals as well as affected individuals to provide and retrieve insights, different angles, and unfiltered (and to some extent unbiased) news on events and are able to add valuable dimensions like detailed-micro perspectives and (near) real-time images of dynamically changing situations to an already multilingual, multi-source and multimedia environment. They enable targeted two-way communication in disaster-scenarios, allowing information to be collected and distributed through these channels. This type of communication is of great value not only because it allows first responders to have access to critical updates from individuals who eyewitness the situation on site but also enables them to send important disaster-related information to the persons who need it most (Backfried et al., 2013b). Their value in the scope of disaster and crisis communication is widely accepted and has led to an increasing number of systems and approaches integrating their strengths (Holmes, 2011; Nilsson et al., 2012; American Red Cross, 2010; Lindsay, 2011; DHS 2018). Especially multimedia data in the form of images and video is becoming more common-place with the ubiquity of portable devices and cameras. Individuals carrying such devices will often be on-site, delivering visual content and associated meta-data accompanied by short comments in real-time. Even though they may sometimes not be the actual drivers of news (Lee Hughes & Palen, 2009), they can be viewed as amplifiers and in many cases may be ahead of traditional media in reporting about a certain incident (a study comparing trending hash-tags on Twitter and headlines on CNN found that in approximately 50% of the cases CNN was ahead of Twitter (Kwak et al., 2010)). As technology is still evolving, people will continue to have rising expectations of first-responders and vice-versa. Affected persons may indeed already expect authorities to quickly respond to information shared on SM (Johansson et al., 2012). However, SM weren't created with first-responders and their requirements in mind, but rather for the general public to connect and share their thoughts and interests with family and friends. Hence certain constraints and limitations remain when considering the scope of SM for crisis communication.

To offer citizens faster and more effective responses, national and international risk and communication experts and projects are starting to focus on new ways to systematically incorporate and integrate SM and the related technologies into risk and crisis communication. Technological advances are accompanied by the development of best practices which enhance the goals of communication in crisis and disaster situations. From the point of view of first responders, this concerns communication in both directions – gathering information from citizens and the media as well as communication of messages from first responders out to the world. The public has thus turned into an important stakeholder in the community of crisis management. Today, everyone is a potential source of information, a potential *sensor* in the field and a target for messages. People share their observations, emotions and opinions and communicate via SM and thus form a crowd-sourced means for gathering and dissemination of information. They have used and will continue to use SM to describe what they see and experience (Earle et al., 2011). In contrast to TM, which generally aim for neutral and objective reporting, SM is frequently dominated by personal goals and attitudes. It is therefore especially important, that this information be combined, reconciled, vetted and contrasted with information provided by traditional sources and traditional media to create a more complete and balanced picture of the situation at hand (in a cross-media approach).

Based on these observations and experiences of previous projects, a strategy of fusing the diverse, mixed, and complementary sources of information, of combining contents from traditional and social media presents a promising approach. The combination of SM and TM allows to obtain information in a variety, at a pace and depth which would otherwise not be available. On the one hand, information from TM – which can often be assumed to possess a certain quality - can be used to assess the trustworthiness and veracity of SM messages. For example, it might be possible to verify the location of a SM post by identifying that the image included in the post and an image published in TM contain overlapping visual content (e.g. the same background depicting the same bridge). Reports on traditional media may also be expected to provide more comprehensive and verified background material. On the other hand, SM messages may be available faster, be more up-to-date and also be available when professional news teams are not (yet) present at the scene of a disaster (or are not allowed to be present). They may thus provide valuable and unfiltered information to first responders, e.g. identifying the exact location where help is needed, the state of local infrastructure or background information about the security-situation or the state of victims. But clearly, SM are not only useful to gather information. Outbound information activities include the dissemination and targeted communication with persons immediately affected by a disaster or corrective measures such as clarifying rumors or correcting misinformation (Bailer et al., 2016).

3.3 Media Requirements of First Responders[57]

First responders have been facing the challenge to find new ways to systematically incorporate and integrate social media[58] and the related technology into disaster communication and combine them with the information gained from the ground and through traditional media. Their specific needs may change as a disaster evolves: at the onset of a disaster, the priority might be to understand the context and scope. At later stages, the focus may shift to assessing the impact on the affected population, the kinds and severity of damage incurred, the types of goods and services needed and eventually to gathering information about donations and external aid. Two prototypical information needs can be distinguished: strategic information, capturing the *big picture* and tactical information, capturing *actionable insights* (Castillo, 2016).

Current approaches typically focus on the gathering and processing of textual data. Technologies, which are not purely based on textual input, operating on audio or visual data form natural extensions of such systems, augmenting, extending and complementing existing sources with information which up to this date remains largely untapped.

Fusing the diverse, mixed and complementary sources of information and allowing decision makers to make sense of a large and inhomogeneous body of data in a short period of time is of utmost importance to first responders and their coordinating organizations. Access to data from different modalities just minutes after they have been published would allow practitioners to perform various types of visualization and analysis activities by connecting and inter-linking information from different sources and media. The early detection of structures and patterns of communication as well as of perceived incidents are required to allow authorities to react earlier, swifter and more adequately in situations of crisis or disaster. It enables them to position vital assets adequately in order to minimize impact, speed up recovery and rebuilding efforts and even anticipate critical events before escalation.

Figure 15 below depicts a generic 5-phase model developed at the National Defence Academy of the Austrian MoD (KIRAS, MUSE, 2013). First responders can benefit in all phases by the combined processing of data across media and languages and their underlying patterns of communication in a unified, fused manner.

---

[57] This chapter contains material previously published by the author in Backfried et al. (2013a and 2013b).
[58] The increased presence of TM sources on SM and their influence on SM only adds to the complexity of this problem. Processing data from SM, a distinction between institutional actors and individuals should be made.

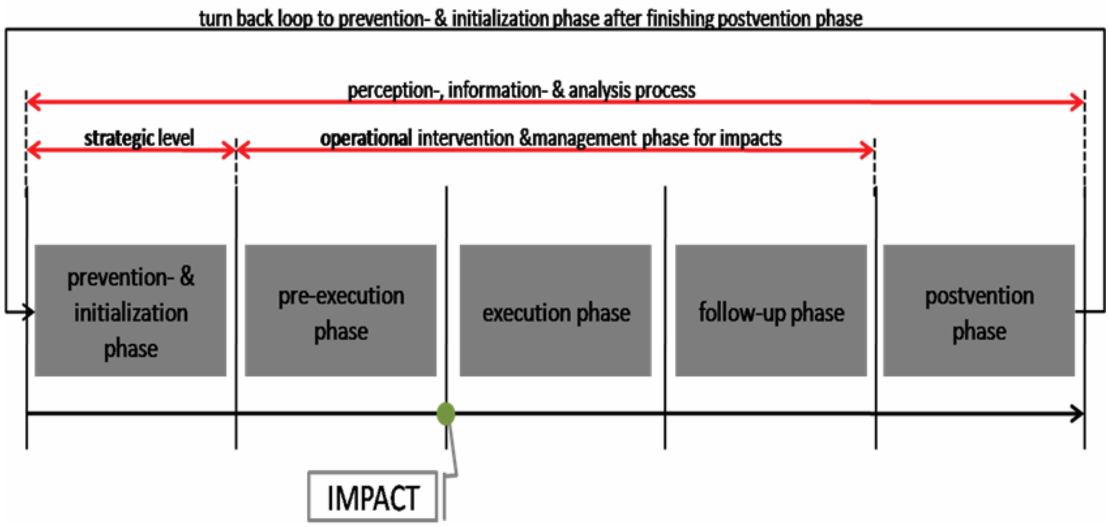


FIGURE 15 FIVE PHASE DISASTER MODEL[59]

According to the above model, first responders face different challenges and requirements in crisis and disaster communication depending on the management / leader-ship level (strategic, operational etc.) as well as the phases themselves. On the strategic level, first responders should have the ability to use the results of media fusion for analytical tasks to identify communication patterns and to link them to particular disaster phases. Emerging risks can be identified more quickly and reliably, and the behavior of the affected population interpreted, monitored and influenced more effectively. On the operational level first responders require a large amount of information to understand the local requirements and tasks and to gain insights about emerging risks and possible threats to the population. The combination of media sources is expected to help to improve shared situational awareness beyond organizational limitations which may exist between the different first responder organizations.

Additionally, first responders require complex (semi-) automated analysis and visualization tools and techniques to complement and support their decision-making processes. The technologies involved have to be multi-lingual and provide cross- and multi-media analysis facilities to ensure access to a large variety of diverse media and information sources. Accessibility and persistence of information has to be guaranteed for later justification of actions or decisions taken or not taken. Visual analysis is required to tap into the growing number of photographic and video-content collected and uploaded to social media sites. These documents do not exist in a vacuum but rather form a complex network with textual and audio-data available across different media. The combinations of these modalities can serve as the base for a holistic approach of disaster communication and management. A combined

[59] Figure copyright J.Göllner & A.Peer, figure taken from Backfried et al. (2013b)

system may also serve as a platform for first responders to share awareness, coordinate the operation, capitalize on synergy effects and synchronize their forces in a theater of operation. It cannot be stressed enough, that all methods and technologies should not be designed to replace people but to help decision makers handle the increasing flood of information and fulfill their tasks in a more efficient way. Technological advances have to be accompanied by the establishment of best practices that enhance the goals of communication in crisis and disaster situations (Holmes, 2011). Their relevance to an application domain, comprising people, organizational and technical systems has to be ensured (Hevner & Chatterjee, 2010).

3.4 Combination of Traditional Media and Social Media

Whereas considerable attention has been paid to the use of social media during disasters (Reuter et al., 2018a; Reuter et al., 2018b), only little work has been devoted to the combination of different types of social media with each other and/or their relation and inter-linking with traditional media. Incorporating the different modalities, angles and aspects allows to obtain an even more complete, timely and diverse picture as events unravel (much in line with the so-called *All-Source-Intelligence* approach advocated for the Intelligence domain (US DoD, 2005)). Clearly, the different media and channels possess different qualities and strengths, operate on different time-scales and granularities and their combination should yield advantages over each individual one. Technically, this process has to be accompanied by the fusion of processing capabilities for different modalities (textual, audio and visual media) as well as multi-lingual and multi-platform processing. Fusion allows to combine modalities during processing (early fusion) or to combine the results of individual processing chains (late fusion). Late fusion is deemed to be more powerful and flexible allowing end-users to emphasize different aspects without having to re-perform processing steps (Snoeck et al., 2005). Multi-lingual processing is indispensable, especially in view of multi-cultural, cross-border or international disaster settings and multiple platforms being accessed by users from around the globe. Foreigners and tourists present at the site of an incident may add further to the mix of languages. Varying time-scales in the publication of content and different granularities of information need to be accounted for in this process as well as for later analysis and visualization.

The products and effects of disaster communication – originating from affected individuals, by-standers, first-responders, authorities or media outlets - form the diverse types of content to be collected, processed and provided for analysis with the eventual goal of improving situational awareness. Typically, this content is provided by a wide variety of actors, in a multitude of formats, on a series of platforms and in multiple languages.

The focus of the current thesis thus lies on the utilization of multiple media and the patterns and linking of media as well as the gathering and processing of content with the aim of uncovering and documenting structures which can only be obtained from the observation of this combined media landscape.

3.5 Existing Work[60]

The first decades of the 2000s have witnessed the birth of a number of social media platforms. These have been adopted on a wide scale and are being used by increasingly large portions of the population for personal and professional purposes. They have changed the digital landscape profoundly and in a lasting way and produced new online social- and communication behaviors in our daily lives as well as during emergencies and disasters. Clearly, there are regional, demographic, cultural and socio-economic differences in the popularity and accessibility of individual platforms and the way users embrace them. However, some large networks - above all Facebook - have managed to dominate increasingly large regions of the globe and continue to extend their reach, even to the point of almost establishing a de facto monopoly across many countries.

Social media platforms may roughly be classified into the following categories:

- Social Networks (e.g. Facebook, LinkedIn, Google+, VKontakte,…)
- Micro-Blogging (e.g. Twitter, Sina Weibo,…)
- Photo-Sharing (e.g. Flickr, Instagram,…)
- Video-Sharing (e.g. YouTube, Vimeo,…)
- Bookmarking (e.g. Delicious, StumbleUpon,..)
- Social News (e.g. Digg, Reddit,…)
- Discussion Fora
- Incident-related platforms (e.g. Ushahidi, Crisis-Tracker,…)

These platforms target different user-groups, environments and objectives and offer different sets of features and functionalities (sometimes referred to as *affordances* (Bucher & Helmod, 2017)). Many of them involve processes of sharing textual and/or multi-media content and the possibility to establish links and communication threads between them and/or the involved persons and entities.

[60] Chapter includes work from papers published by the author in Backfried (2013a) and in Reuter et al. (2018)

Some of them, like Ushahidi involve crowd-sourcing or activist-mapping. The nature of these platforms ranges from purely commercial ventures to open-source projects.

Since their respective time of introduction, social media have also been used to collect and disseminate information about natural disasters. Platforms like Twitter or Facebook, allowing to disseminate snippets of information in real-time to a selected group or to the general public and allowing users to comment and enrich information have found wide adoption during crises and disasters. Photo-sharing services, being fed by an increasing number of mobile devices have been used for publishing and linking of information across media. Images and videos are frequently referenced from platforms other than the one they actually reside on. Web sites and RSS-feeds are used for similar purposes in addition to providing a separate source in their own right. Linking between the different media, e.g. including links to videos in tweets or mentioning Facebook accounts or hashtags on TV has become a common way of communicating and referencing information.

Since 2004, several disasters have been investigated in the context of social media and crises. Figure 16 below depicts some of the relevant platforms involved and their respective years of launch.

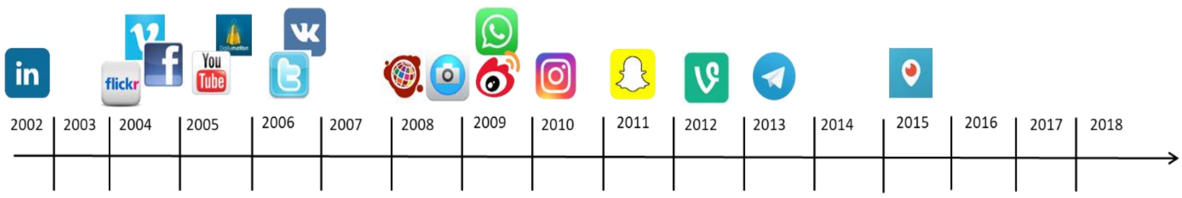


FIGURE 16 LAUNCH OF SOCIAL MEDIA PLATFORMS

Over the same time-span, crisis scenarios have become an important field of application for social media. Under the term of C*risis Informatics* (Palen & Anderson, 2016), this phenomenon has been an extensive area of research since 2006 (Palen & Hughes, 2018; Reuter & Kaufhold, 2018; Reuter et al., 2018a). Ever since their inception, there has been almost no large emergency where the use of social media has not been studied (Reuter & Kaufhold, 2018).

Even as early as 2005 during the London bombing attacks[61] different types of media were already being mashed-up and connected by numerous actors (Peary et al., 2012). The same year witnessed widely observable changes in media mixes and information flows during Hurricane Katrina (Haddow & Haddow, 2014). Liu et al. (2008) report on the role of the photo-sharing platform Flickr for six notable disasters between 2004 and 2007. More recently, the use of multiple platforms during an

[61] A *man-made disaster*, see chapter 5.1

event has been investigated by Kaufhold and Reuter (2014) and St.Denis et al. (2014). Backfried et al. (2014) stress that traditional and social media provide complementary and different angles and should be used in combination rather than viewed in an isolated manner and provide an overview and findings on how different natural disasters have been covered and investigated with regard to media utilization. Bruns et al. (2012) provide an in-depth and comprehensive overview of how social media played a central role during the 2011 Queensland floods where Twitter was used extensively by the Queensland Police Service (@QPSMedia) to communicate with affected population.

Social media have likewise played important roles during and following the earthquakes in New Zealand (Bruns & Burgess, 2012), Haiti (Dugdale et al., 2012), Chile (Mendoza et al., 2010), China (Eearle et al., 2010) and the US (Starbird & Palen, 2010), for grassfires in the US (Starbird & Palen 2010), hurricanes Katrina (Faustino et al., 2012) Ike and Gustav (Lee Hughes & Palen, 2009) and Sandy (Faustino et al., 2012; DHS, 2013), typhoons and floods in the Philippines (Parilla-Ferrer et al., 2014) and floods in Brazil (Nagar et al., 2012) and Pakistan (Murthy & Longwell, 2013). Their role in the floods in central Europe in May and June 2013 was the subject of the project QuOIMA (Backfried 2013b; Backfried et al., 2013). The rapid use of newly emerging platforms for the case of Periscope during Hurricane Joaquin is documented in Fichet et al. (2016). Reuter et al. (2018a) provide further examples and general background on the history of social media in crisis management.

In the overview of social media in the context of crisis information provided by Nilsson et al. (2012), it is found that media – traditional as well as social – form an important source for sending out and retrieving information about different aspects of a crisis or disaster. Mendoza et al. (2010), Lee Hughes and Palen (2009) and Kwak et al. (2010) elaborate on the structure and types of communication which emerge on Twitter, Holmes (2011) does so particularly for the crisis-case. Faustino et al. (2012) provide an overview of social media in the disaster context using a set of case spotlights covering specific events. Harrington et al. (2012) focus on the connection between Twitter and TV from the point of view of how to combine social network interaction with programming. Kwak et al. (2010) provide a brief overview of existing and potential future uses of social media in disasters. Olteanu et al. (2015) investigate a series of natural and human-induced disasters and identify different types of content and information sharing across crises. They provide an excellent overview of previously examined disasters and the respective findings. Relatively little attention seems to have been paid to the actual use of language in the context of social media and natural disasters: Medoza et al. (2010) and Faustino et al. (2012) deal with this topic to some extent.

Numerous research projects aim at improving the effectiveness of communication and alerting during crises. For example, in the EU FP7 Alert4All (https://cordis.europa.eu/project/rcn/98427/factsheet/en), the screening of new media (SNM) tool (Nilsson et al., 2012) serves to enhance situational awareness using opinions and emotions found in online web content and social media. MASSCRISCOM (http://www.masscriscom.eu) takes social media into account as an emerging means of mass communication during crisis situations. The H2020 project E2mC (https://www.e2mc-project.eu) aims at assessing the quality and credibility of social media gathered by crowd-sourcing. At QCRI in Qatar, the AIDR project[62] aims at developing semi-automated mechanisms, combined with crowd-sourcing, to form the foundations of a Twitter-based dashboard for analytical purposes during crises. The system has been used during several large-scale disasters such as the 2015 Nepal Earthquake, and typhoons Hayan and Hagupit in 2013 and 2014 respectively (Imran et al., 2017). In the Austrian Security Research Program KIRAS[63], the project QuOIMA Backfried et al. (2013a) aimed at combining cross-media, multi-lingual analysis with visual analytics methods.

It is notable that the above works, with few exceptions, focus on social media and on individual platforms only. The combination of social media processing with processing of traditional media content – comprising the complete spectrum of media from textual, auditory and visual media-types - has not been addressed thoroughly in the scope of disasters and disaster management. Furthermore, the connections between different types of media have only been investigated to a limited extent.

3.6 Summary

The advent of the Web2.0 and social media have had a profound impact on the media landscape. What used to be predominantly a one-way street of communicating has been extended with two-way, interactive and participatory modes of engagement. Technological changes, such as the ubiquity of mobile and portable devices have accompanied these changes and lead to dramatically new ways of how news is produced and consumed. Members of the public have been turned into *prosumers* of news, during normal times as much as during times of disasters. Changes are still on-going, and the dynamics of the media landscape will continue to evolve, spawning new platforms and formats along the way. Disasters have always been covered by the traditional media; new and social media now complement this coverage and bring in novel and different angles along the way. Coverage is

[62] http://aidr.qcri.org/ on 2018/11/12
[63] https://www.ffg.at/programme/kiras on 2018/11/12

becoming more diverse, but with the diversity new challenges are appearing, such as the veracity of news, the mixes and mashes of multimedia content and use of (non-standard) language. In the midst of this all, first responders and officials need to be aware of developments across the complete spectrum of media. In particular with cross-border and international settings, multilinguality plays a crucial role. Not to participate is not an option if one does not want to leave the resulting communication vacuum to individuals and outside actors. In order to comprehensively monitor disaster coverage and produce the best situational awareness possible, a multi-platform, multimedia and multilingual approach is required. Due to the dynamics of the landscape, such an approach should be generic and open – a framework flexible enough to accommodate newly emerging platforms – rather than focused and constrained to individual platforms and media only.

4 Relevant Insights and Results from QuOIMA[64]

The KIRAS project QuOIMA was launched in November of 2012 and completed in October 2014. It addressed the use of social and traditional media in the context of natural disasters in Austria (Backfried et al., 2013). In the summer of 2013 Austria experienced the worst flooding in 500 years, an event which was chosen to serve as a use-case for the project. The data collection and processing efforts, as well as the insights gained, which are relevant to this thesis, are described below.

4.1 Corpus

To allow the analysis of - postulated - cross-media communication phenomena, a parallel corpus spanning various types of media, sources and languages all pertaining to a single event is required. To this end, a corpus covering the floods on various types of media was created within the QuOIMA project. Data-sources and the respective amounts of data collected are listed below in Table 1.

Medium	Amount	Comment
TV	218h	13 TV programs, 9 TV stations, 4 countries, German and English
Web	3500 pages	102 German and English sources
Tweets	470k tweets	German, English, Dutch
Facebook	9800 posts/comments	16 public pages, German and English posts and comments from more than 1000 users
Press Agencies	750 press releases	German

TABLE 1 QUOIMA CORPUS AND DATA SOURCES

Traditional sources (TV, web, press agencies) were selected according to geographical and content-related coverage. TV sources include public and private stations from Germany and Austria (e.g. ZDF, ORF, NTV) as well as TV-stations broadcasting on a global scale (e.g. CNN, BBC) but covering events in Central Europe and Austria. Web-sources included the news feeds of international, national (Germany and Austria) as well as local newspapers and magazines. Facebook data included posts and comments from public-pages of organizations and individuals involved in disaster-relief. Twitter data was generated in an incremental approach by starting from generic disaster-related terminology and finally producing an active set of hash-tags and user-accounts which were used with Twitter's streaming-API to retrieve tweets. Images referenced from documents were downloaded from Twitter, Facebook, Flickr and Instagram and resulting in an additional set of almost 50.000 images and

[64] This chapter contains material previously published by the author (including figures), also see chapter 2.6

associated meta-data. All data collected was processed (transcribed, extracted, cleaned, tokenized,…) annotated and enriched with meta-data for subsequent analysis purposes.

Data collection was carried out throughout the flooding with the period of May 20 to June 23, 2013 representing the most severe period corresponding to the execution phase of the disaster management cycle in Austria. The data corresponding to this core period was subsequently used as the base for analysis.

4.2 Coverage Patterns of Media

A first analysis was performed on the frequency and volume of documents from the above corpus. Over the observed period from May 20 to June 23 of 2013, all data sources, traditional as well as social, generally follow the same trend. Activity is low at first and then rises slowly towards the final days of May, as the high-water level on the Danube was approaching Austria. Within the first days of June, however, the picture is quite different with a sudden surge of activity across all sources. TV coverage displays a first peak on May 31 during the ramp-up leading to the height of the flood. As the situation was gradually getting worse with forecasts indicating extreme conditions, these forecasts received substantial attention across all TV stations monitored[65]. Social media coverage yields most documents relating to the event on June 3 and 4 - following the impact of flooding by a couple of days - the activity on Twitter leading Facebook posts and comments by one day. Web-sources peak one day later on June 5.

All sources display multiple activity peaks corresponding to the relative peaks of the Danube- and Elbe[66]-related floods within a few days only. TV-coverage quickly dropped after the initial peaks (potentially due to other, international events replacing the flood-topic swiftly) whereas social media and web-sources continued with constant high coverage of the events. Figure 17 below displays the activity over all observed media by plotting the percentage of documents relative to the total volume within a medium, over the observed period (May 20 to June 23 2013).

[65] Radio was not monitored separately but it can be assumed that the pattern and volume of radio coverage resembles that of TV in Austria.

[66] The floods on the river Elbe followed the ones on the Danube with a delay of a few days.

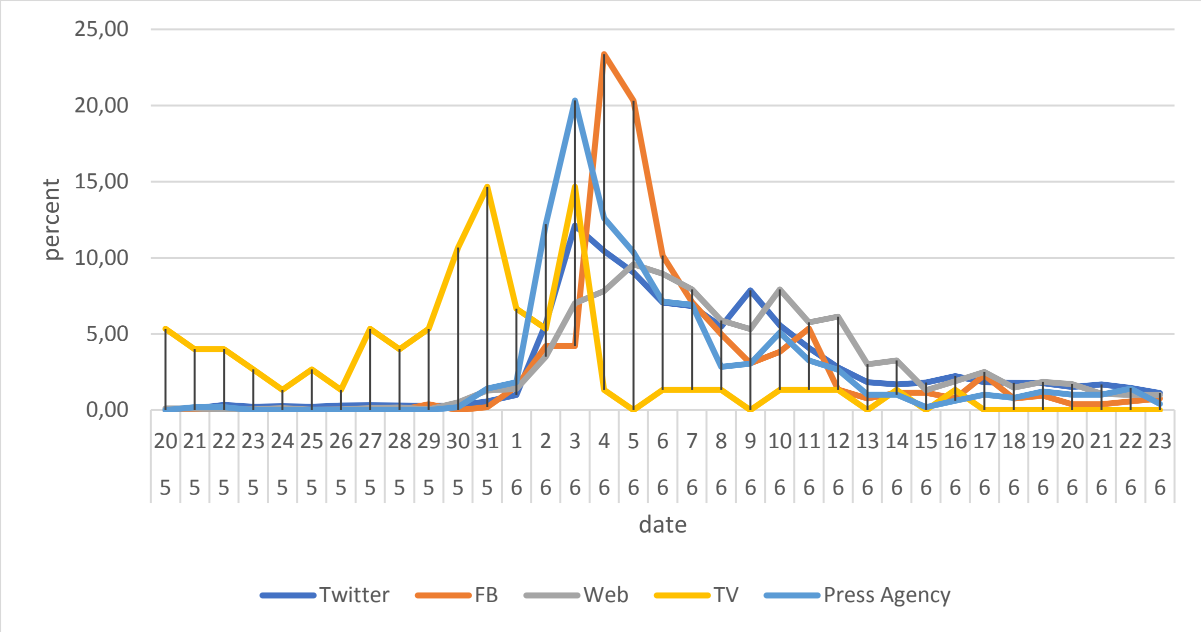


FIGURE 17 PERCENTAGE OF DOCUMENTS PER MEDIUM COMPARED TO TOTAL VOLUME OF DOCUMENTS

The observed behavior seems to suggest that neither traditional nor social media have a strong lead in their coverage of events – the relative peaks of documents all lie within a couple of days. This may have been due to the nature of the natural disaster, having been correctly predicted by meteorologists some time ahead of the actual impact. The *attention span* of the different kinds of media, however, seems to differ, with interest ebbing off quite quickly on TV.

The overall analysis demonstrates that activity indeed does span the complete spectrum of observed media and patterns are comparable over time, meaning that all media should be observed with the same intensity regardless of the disaster-phase and for the duration of all phases in the disaster cycle.

4.3 Cross-Media Communication

Various forms of cross-media communication can be expected to take place during a major disaster. TV programs may list hash-tags and accounts to be used for communication on social media. Posts on social media platforms may link to newspaper feeds or image-sharing platforms, newspapers in turn may link to video-sharing sites etc. These interconnections form just a few examples of the many possible combinations and links between and within media types and sources. Not all connections may occur with equal intensity during the different phases of a disaster - e.g. links to photo-sharing sites might occur more frequently during the initial phases (e.g. showing images of the impact of a disaster) whereas links to web-sites might prevail during later stage (e.g. donations). All of these links provide ample opportunities to extend mechanisms for data collection - by adding the appropriate sources and by downloading and processing the actual documents and associated meta-data – as well as for analysis of actors, their messages and agendas. Furthermore, feedback on relief measures and the perception of first responders can readily be made visible through this data and supporting or contradicting information be highlighted.

Table 2 below provides an overview of the most frequently linked-to sites from tweets referring to the flood. A large proportion of links emanating from tweets points to other social networking sites (*social*). Links to Facebook feature among the most frequent ones. Links to public pages – as private ones were not considered - can be followed and the target pages added to data-collection and analysis tasks. Video- and image-sharing web-sites such as YouTube, Instagram, Twitpic and Flickr likewise feature prominently. A portion of the targeted videos as well as images can be downloaded and processed using audio- or visual-processing components such as OCR, scene detection or the detection of near duplicates. Visual objects receiving most links (and/or rated on their respective platforms according to meta-information available via APIs) can be identified and presented to analysts. Using this approach, the most-frequently shared image on a popular Facebook page could be identified as having been taken during a flood several years ago (also along the Danube, so location and flood condition might be genuine, however, the time is not) – a case of *fake news*. A further group of link-destinations are news-outlets associated with newspapers, magazines and TV-stations - this supports the statements made above regarding the increasingly blurred boundaries of the media landscape (*news*). Finally, weather stations posting river-levels and specific weather information (*weather*) as well as political entities such as parties and local governments are linked to (*org*). Conspicuously missing from the table are first responders, such as the Austrian Red Cross[67] or the "Technisches Hilfswerk" (the Federal Agency for Technical Relief, THW) in Germany[68] or government agencies. This could be an indicator that the presence of such organizations on Twitter is still weak or that information posted by them is not perceived as relevant or interesting and is consequently not referenced or passed-on frequently. Both cases could motivate a more active stance of such organizations on social media platforms and the improvement of corresponding communication strategies.

[67] www.roteskreuz.at
[68] www.thw.de

domain	count	social	news	weather	org
www.facebook.com	10030	x			
www.seismoblog.de	5703			(x)	
YouTube	4946	x			
instagram.com	3524	x			
www.spiegel.de	3187		x		
www.welt.de	2552		x		
www.focus.de	2263		x		
www.mz-web.de	2170		x		
www.tagesschau.de	1778		x		
www.n24.de	1717		x		
www.n-tv.de	1642		x		
www.wetterblogger.de	1598			x	
www.sueddeutsche.de	1476		x		
www.br.de	1458		x		
twitter.com	1420	x			
www.mdr.de	1331		x		
twitpic.com	1152	x			
stuke2.piratenpartei-bayern.de	1031				x
fluddhilfe.de	1022			x	
www.epochtimes.de	1009		x		
www.wetter.com	931			x	
www.dnn-online.de	890		x		
www.faz.net	848		x		
www.morgenpost.de	813		x		
www.zeit.de	812		x		
www.ndr.de	781		x		
www.deutscher-warndienst.de	748			x	x
www.bild.de	741		x		

TABLE 2 DESTINATION SITES OF LINKS

4.4 Linking Behavior over Time

The linking behavior of social media posts might vary depending on the phases of a disaster and the amount of news-coverage an incident receives. The ubiquity of smart-phones and hand-held devices allows individuals witnessing events to post images in real-time (sometimes even to the point where people have to be reminded to seek shelter before or instead of tweeting). These can subsequently be passed-on and linked-to by the recipients or the general public. Later phases of the disaster cycle might be characterized by news agencies or organizations (such as first responders) providing in-depth and broader coverage which might then be disseminated along the same channels. Figure 18 below shows a graph of links to social media sites, news-outlets and sites for donations over the core-period of the flooding in Austria (May 20 to June 23, 2013). The graph depicts the percentage of links to the particular kind of medium as compared to the overall volume per day of links to the same kind of medium.

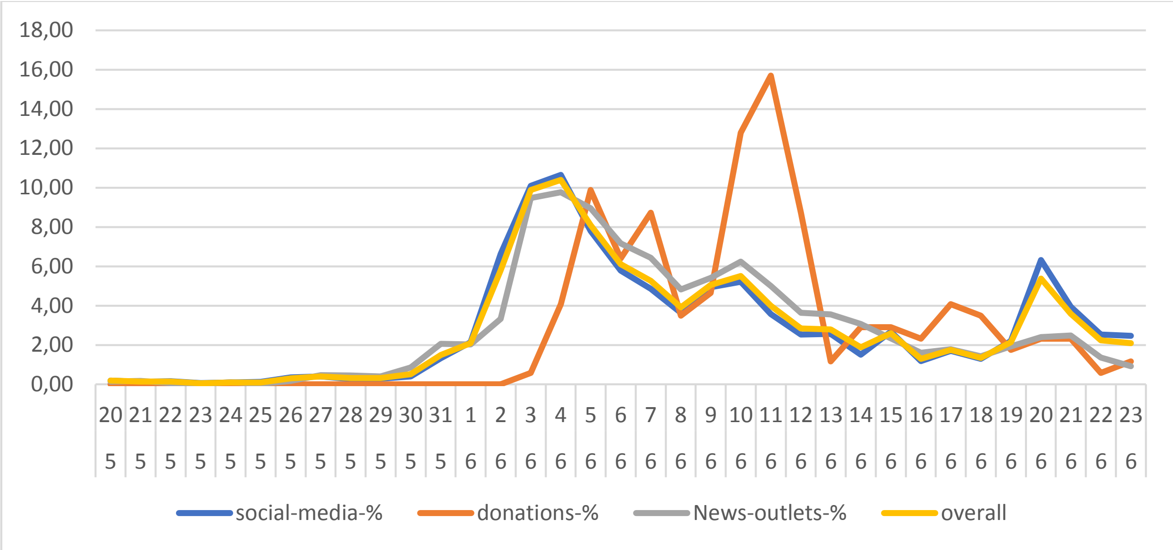


FIGURE 18 DESTINATIONS OF LINKS OVER TIME FOR THE CORE-PERIOD OF FLOODS IN AUSTRIA

No clearly discernible difference in the pattern of linking behavior can be observed over time. That is, links to social media as well as to news outlets follow a similar trend with neither one seemingly being preferred at any stage of the observed period. Links to sites dealing with donations, however, do display this time-related behavior with links increasing dramatically in the days following the highest flood levels (peaking around June 12[69]), several days into the incident, once the population had been exposed to information and images regarding the flood's impact and consequences. This observation indicates that information gathering should not assume any staged time-distribution over the different types of media but rather should address all types of media – traditional and social – simultaneously and in parallel, i.e. in a phase-agnostic manner. Changes in the types of destinations, however, may occur allowing to focus efforts on particular topics over time.

4.5 Inter-platform Linking

On the one hand, due to the concise nature of tweets, Twitter is frequently used as a communication platform during disaster. On the other hand, this very feature – a limit to 140 (resp 280) characters frequently prompts the inclusion of links to other platforms and media to complement and extend the tweet itself. In this respect, Twitter may be regarded as an engine for cross-platform and cross-media linking. An analysis of the tweets collected during the flood reveals that several social media platforms are linked to from within the tweets. They comprise social networking sites (Facebook), video sharing (YouTube, Vine) and image-sharing platforms (Instagram, Flickr, Twitpic) as well as the location-based platform Foursquare. Figure 19 below depicts the linking behavior to those social media platforms

[69] The 2nd peak around June 20 was likely caused by activity due to related flooding in Germany.

over time (absolute number of links to the respective platform). The overall frequency of linking follows the activity on Twitter, with the various platforms exhibiting similar trends over time (as above in Figure 18 the smaller peak around June 20 may have been caused by further flooding in Germany).

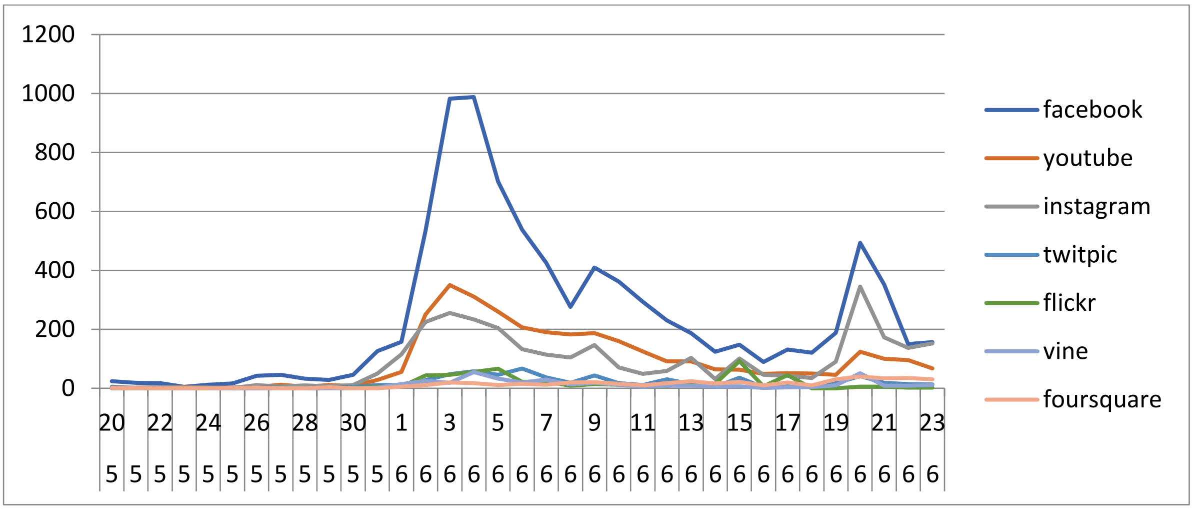


FIGURE 19 SOCIAL MEDIA PLATFORMS AS DESTINATION FROM TWEETS

As for the general trend of linking over time, the parallel development of linking to social media destinations over time suggests a collection strategy which is active continuously. Due to potentially large surges in volume and frequency, processing must be dynamically adjustable to sudden spikes of activity.

4.6 Summary

QuOIMA represented the first systematic endeavor studying the phenomenon of social media in the context of natural disasters in Austria. As such, it provided a series of valuable insights from a technical as well as an organizational point of view. The ones relevant for this thesis are listed below.

- Social media (Twitter, Facebook) do play a role in the communication about natural disasters in Austria. Even though Twitter is sometimes viewed as a niche-phenomenon in Austria (and other countries), its nature of concise messages, real-time communication and the possibility to interlink content from various platforms and media make it an adequate and practical means of disaster-communication also for such conditions.
- A lack of use on part of first responders and officials may cause an information vacuum filled by individual players such as in the case of alleged coli bacteria in drinking water, where an individual account managed to rapidly concentrate a large portion of all social media communication, distributing incorrect information.

- Automated tweeting (bots) occurred with some automated accounts providing valuable information (e.g. water levels at specific locations) and others providing irrelevant information (e.g. travel). Detection of such accounts and distinction of their nature would be beneficial.
- Traditional media coverage of the disaster was substantial. Peak coverage on TV preceded peaks on social media. However, this might be explained due to the specific setting of the flood. Social media may lead in coverage; the quality and content may differ however. Neither traditional nor social media should be preferred but rather both types should be regarded as a continuous source of information and should be considered when collecting information about the state of the incident.
- Linking of information and across media types occurs in a variety of ways. Images and web-content are frequently linked-to by social media posts. Inter social-media linking can be observed in a variety of ways; web-pages are likewise linked-to from social media. The destination of links – and the absence of certain destinations, such as web-pages of first responders or governmental agencies – provide valuable clues as to the information needs of the affected population as well as whether certain official measures and information is even perceived. The various types of links provide strong support for the argument of cross-media communication.
- Even though the flood was examined from an Austrian perspective, it became clear that content is not only provided in German. This may be due to foreigners and tourists[70] present at the time of disaster. Languages used internationally, such as English as well as local languages of surrounding areas should be taken into account.
- Information was provided and shared during the complete period of the flooding. The amounts and quality of information shifted during this time. However, the activities provide evidence that all disaster phases should be monitored.

The results and insights of the QuOIMA project outlined above prompted a more thorough investigation of the mechanisms and platforms involved. They provide support for the proposed approach, in particular for the phenomenon of cross-media links between the different types of documents and actors in disaster communication, which forms the work undertaken within this thesis.

[70] More recently, *Dark Tourism* – tourism involving travelling to disaster-sites – has added another facet to this phenomenon.

5 Framework

Chapter 2.1 above outlines the motivation for the current work and the development of a generic framework for cross-media communication during disasters. The term *framework* is heavily overloaded with a wide variety of definitions, as much in computer science as in a more general sense. To delineate and guide the work performed within this thesis, a definition similar to the one provided by Meise (2001) is adopted. Within the current thesis, the (generic) framework defines and structures the elements identified as relevant in the domain of cross-media disaster communication on a high level of abstraction in order to provide an overview and to allow for the classification of these elements and their interrelationships. It aims to clarify the terminology and to represent the involved elements on a generic und unified level. The framework consist of a model describing the elements of the target domain and a model for operations for their creation and processing as well as a method to create these models. Both models are developed on a conceptual level and *hand-in-hand*, following the suggested method, in the chapters below. This approach aligns well with a common perspective from computer science, where model-development is divided into the development of a data model and of a process model (Simsion & Witt, 2004). The ***framework*** thus comprises the structure of these models as well as the method to develop them.

An increase in natural disasters paired with a changing media-landscape, fluid and more diverse than ever before with a multitude of actors producing and consuming information during such events pose substantial challenges to anyone wanting to obtain a comprehensive and broad picture of a situation. This concerns individual citizens, the press and media, NGOs and official institutions alike. At the forefront, it concerns first responders requiring and relying on adequate and timely information for the coordination of their efforts. In order to comprehensively monitor disaster coverage and produce the best situational awareness possible, a multi-platform, multimedia and multilingual approach is required. Several aspects regarding these statements were confirmed by the findings of the QuOIMA project (outlined in chapter 4) preceding the work on this thesis. Due to the dynamics of the landscape and the heterogeneity of sources and platforms, a generic model is developed within this section, going beyond the individual elements investigated by previous research. The model is envisioned to be generic and open, and flexible enough to accommodate newly emerging platforms, rather than focused and constrained to individual platforms and media only.

In order to be able to develop this framework, a terminology is defined first, allowing to refer to the involved elements in a consistent manner. Due to the inter-disciplinary nature of the domain, no single (pre-defined) terminology exists which would satisfy and be in complete agreement with all the fields involved. The terminology is thus defined primarily from an information processing/computer science background and to allow to refer to the elements involved in this thesis in a consistent manner. An established communication model is chosen, providing the basic elements of the generic model and adding to the terminology. The technologies involved, methods to access and process data of various kinds are combined with the former elements in the actual data- and process model development. Finally, the resulting structure and mechanisms are evaluated against real-world data. Figure 20 below provides a schematic overview of these steps which are detailed in the following chapters.

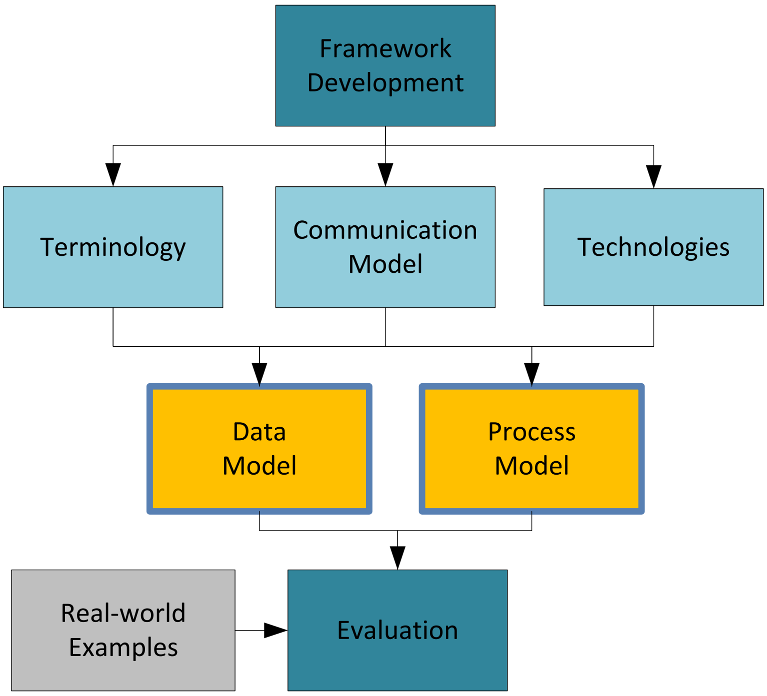


FIGURE 20 FRAMEWORK DEVELOPMENT OVERVIEW

The goal of establishing a generic framework is to map all involved elements and terms to the same concepts, where possible, and thus allow to classify and compare them and to determine commonalities and differences. This in turn allows a more principled approach to information collection and processing and to the matching of requirements with technical capabilities and platform affordances.

5.1 Terminology

Cross-media communication during crises and disasters is an inter-disciplinary endeavor. With origins in different disciplines, terms such as *medium*, *channel* or *platform* are heavily overloaded and defined in a multitude of ways[71]. In practice, their exact meaning differs substantially depending on the context: the same term may refer to fundamentally different *concepts* or different terms may refer to one and the same *concept*. The resulting ambiguities became apparent and tangible during earlier work of the author within the QuOIMA project. The ensuing difficulty in comparing different sources of information and in determining their particular differences and commonalities - and being able to build upon them - formed a strong motivation for the current thesis. The highly heterogenous media-landscape, the data-structures (documents, actors, …) involved as well as the operations on and between them provide the starting point for the creation of a generic data- and process model. However, to arrive at such a model, a common terminology – a common ground to be able to compare apples with apples and allow to specify the operations on them - first needs to be established.

The author advocates the view that only the combination of different types of media (a *cross-media approach*) rather than focusing on one specific kind of medium can provide the adequate and broad information required to efficiently manage disaster- and crisis response. Particularly in a multimedia, multi-source, multi-actor and multi-platform news environment, links and possibilities to mash-up information between individual sources become apparent rapidly and add an important dimension to communication. The current thesis reflects this with a *cross-media approach* in mind. An over-arching data gathering, analysis[72] and information-retrieval perspective rather than a dissemination viewpoint is assumed. This stands in contrast to the traditional notion of *cross-media communication* and traditional crisis- and disaster-communication as described in Haddow and Haddow (2014) focusing on dissemination aspects.

The following definitions describe the central elements and dimensions mentioned in the title of the current thesis:

A Framework for the Support of Cross-Media Communication During (Natural) Disasters

<u>Framework</u>

Framework here refers to a supporting, general structure, a *frame* providing the context for the developments within this thesis and is used as an overall term encompassing the models developed, the method to do so and the terminology established in the process within this thesis. It does not refer to any particular framework as frequently used in the field of computer science. Focusing on a

[71] Chapter 5.2 below provides their definitions for the scope of this thesis.
[72] Analysis: process of collection, enrichment, processing, visualization

framework aims to abstract from individual, existing platforms and to find the generic and common structures and mechanisms behind them, allowing to also accommodate emerging and new platforms.

Cross-Media Communication

In the context of the current approach, cross-media communication comprises the communication of individuals and organizations on a variety of platforms and media. The present thesis specifically targets the collection[73] of information on open platforms, the analyses possible by such mechanisms and does not address the aspects of story-telling or (outbound) communication strategies. The approach thus complements the traditional information-dissemination focused concept of cross-media communication (Davidson, 2010) with an information-gathering and -retrieval one.

Crisis Communication and Disaster Communication

The discipline of crisis-communication is dedicated to communicational aspects during crises. It is typically concerned with public relations, the management of reputation and the partnership with the media and press, officials and the public in general (Coombs & Holladay, 2012). In contrast, disaster communication is framed slightly differently with a focus on inter-agency communication (Haddow & Haddow, 2014). In the context of this thesis, it entails the communication of individuals and organizations about natural disasters and other large-scale incidents. Only this latter type of communication is addressed by this thesis.

Disaster

Due to the many disciplines using the term *disaster* there is no single definition of the term which is universally accepted. On the contrary, a multitude of definitions exists, focusing on different aspects and related concepts such as *crisis*, *incident*, *hazard, emergency* or *issue* even to the point of sometimes using these as synonyms (Shaluf et al., 2003). Al-Dahash et al. (2016) provide a review of the terms *disaster, crisis* and *emergency* and their manifold definitions. Olteanu et al. (2015) consider different taxonomies of disasters used in Europe and the US and stress the point that the type of disaster impacts the information and volume of data generated and shared. Particularly regarding the dimensions of time and space, a distinction into instantaneous (sudden-onset) and progressive (slow-onset) disasters as well as into focalized (small-area) or diffused (large area) disasters can be made as these factors all have a strong influence on the dissemination and flow of information. Castillo (2016) discusses disasters as social phenomena characterized by a disruption of routine and of social

[73] Collection is used as a synonym for ingestion, gathering, harvesting etc.

structure and investigates the role of social media during disasters. In the present work, a pragmatic approach is followed using the definition of *disaster* as provided by the UNHCR (UN, 2009):

A serious disruption of the functioning of a community or a society, involving widespread human, material, economic or environmental losses and impacts, which exceeds the ability of the affected community or society to cope using its own resources.

Regarding the relationship between *disaster* and *crisis*, the latter term is often viewed as a pre-cursor to a disaster, an abnormal and difficult situation which, if not handled timely and properly, may lead to a disaster. *Crisis* is also typically used in an organizational context (Coombs & Holladay, 2012; Shaluf et al., 2003).

Figure 21 below provides an overview of these terms. For the scope of the current work, the term *disaster* is viewed and discussed primarily with n*atural disasters* in mind. Many of the concepts found in the framework may equally be applied to man-made disasters and, to some extent also to (organizational) crises as many common communication mechanisms can be encountered across such types of events. However, for the current work, examples are drawn primarily from natural disasters (unless explicitly specified otherwise) and results and statements thus apply foremost to these.

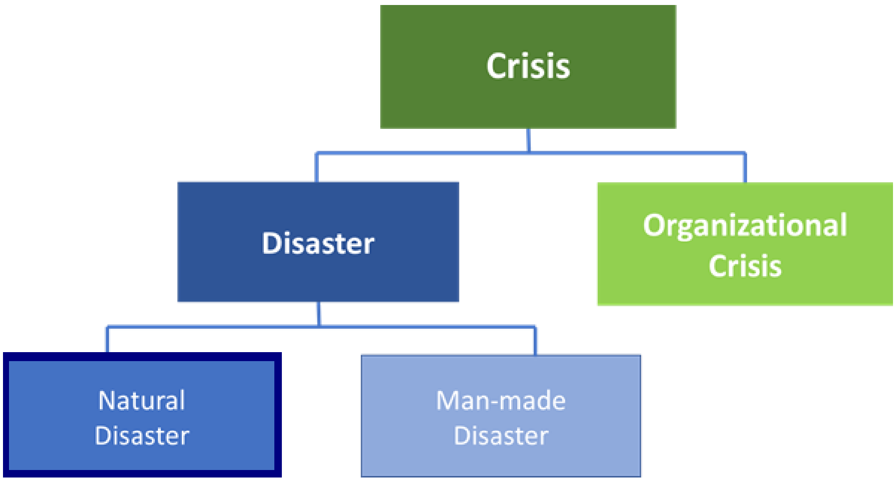


FIGURE 21 CRISES AND DISASTERS

Further terminology will be defined in regarding the elements of the communication model below.

5.2 Communication Model

Communication models are conceptual models used to describe and explain processes of human or machine communication[74]. Triggered by the technical developments in the mid-20th century and with a mathematical and information-theoretic background, the seminal work of Claude Shannon and Warren Weaver was published in 1949 (Shannon & Weaver, 1949). The *Shannon-Weaver-Model*[75] subsequently became the basis of communication modeling, above all in technical disciplines, sometimes even being referred to as the "mother of all models" (Hollnagel & Woods, 2005). Since then, models of communication have been investigated by numerous other fields of science. A considerable number of conceptual models have been developed and refined within the area communication theory including models addressing the phenomena of mass communication (Kübler, 2003).

As the focus of the current work lies on data-collection, -processing and -analysis and the means by which these can be achieved – rather than on the effects generated by communication – a simple, but yet comprehensive model, allowing to describe the elements within the scope of these processes is provided by Harold Lasswell's well-known model of communication (Lasswell, 1960) depicted in Figure 22 below. The model allows to accommodate all relevant elements of disaster-communication while not incurring additional unneeded complexity, making it an adequate fit for the current work. The individual elements, their properties, roles and how they are related to Lasswell's model are detailed below.

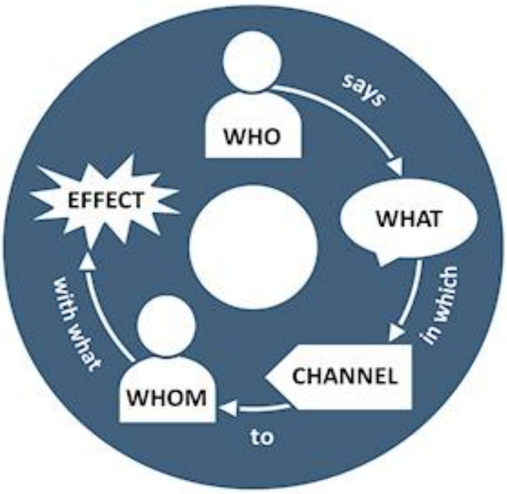


FIGURE 22 LASSWELL'S MODEL OF COMMUNICATION[76]

[74] For an overview of different communication models see chapters 1 and 2 in Cobley & Schulz (2013) and also https://en.wikipedia.org/wiki/Models_of_communication (on 2018/12/31).
[75] https://en.wikipedia.org/wiki/Shannon%E2%80%93Weaver_model (on 2018/12/31)
[76] http://www.osa.fu-berlin.de/puk/beispielaufgaben/lasswell-formel/index.html (on 2018/12/31)

The following elements regarding the creation and transfer of information can be identified in Lasswell's model:

- <u>Who and Whom</u>: the communicating party/parties, generically referred to as *Communicator* (C) below
- <u>What</u>: the *message*, the actual *content* and how this content is conveyed. This will be represented by the possible kinds of *Document* (D) below. Salient elements of the content, so-called *Entities* are represented separately as they form key elements for connecting information across *Documents.*
- <u>Channel</u>: the *Medium* on which a message is transmitted and on which communication takes place. *Channel* is interpreted as a specific kind of Medium - determined by a set of properties - and in this manner is related to Medium as outlined below.

Due to the focus on collection rather than dissemination strategies, the effect generated at the receiver is not addressed below.

Communicator: the originator of a message may generically be referred to as a *Communicator*. Communicators possess several properties, they can be persons (a primary source, an eye-witness or a secondary source, relaying information), organizations or digital identities (e.g. bots or even IoT[77]-devices). A Communicator may be connected to several accounts on a single platform[78] or on different platforms, i.e. a set of accounts may be associated with one or several Communicators. For the scope of this thesis, the *Recipient* of the message will be treated in analogy to the *Communicator*, i.e. their properties within the model to be developed here are assumed to be largely identical (in particular on SM, Communicators and Recipients also frequently do change their respective nature and role). The concept of *Communicator* corresponds to the *Who(m)* in Lasswell's model[79].

Document[80]: content is regarded as the abstract idea or message to be communicated and packaged into Documents, thus making a *Document* the *embodiment* of content. As for the Communicator, a list of attributes can be associated with Documents, such as the publication date, language or its type

[77] IoT: Internet of Things, in disaster settings e.g. sensors measuring water-levels, reporting wind-speeds etc.
[78] The term *platform* is related to *Medium* via a set of attributes associated with *Medium* as described below.
[79] From a communication science perspective, a distinction should indeed be made. For the purposes of this thesis, however, and as particularly on SM these roles intermix, this distinction will be omitted for the current work: the Recipient of a message will be modelled like a Communicator.
[80] The use of *document* (lower-case) denotes documents in a more general sense, e.g. originating from social media platforms whereas *Document* specifically refers to the element of the current model.

(e.g. audio, image or video, similar to the well-known MIME types (MIME, 2017)). Within a Document, specific ***Entities*** - some of which may also be referred to as ***Named Entities*** (Grishman & Sundheim, 1996) - and ***Topical Markers*** may occur. Entities can be found in Documents' content or meta-data associated with Documents or Communicators and play a central role in Information Extraction (IE) (Okurowski, 1993; Balke, 2012). They may be present in textual, auditory or visual form.

Documents, their content and the *Entities* contained in them correspond to the notion of *What* in Lasswell's model. The explicit modelling of Entities in the current work is motivated by their pivotal function in the linking of information across Media, Platforms and Documents and forms an extension not present in Lasswell's communication model.

Medium/Channel/Platform: the term Medium in our context is to be understood as the means of delivering and receiving Documents. The concept of Channel is directly related to Medium and can be regarded to be a particular instance of Medium. Media can be characterized by a set of properties such as whether they are digital or analogue by nature, by the type of Documents that can be exchanged on them or whether they are uni- or bi-directional. A Channel may be viewed as a specific combination of properties and their associated set of values, possibly have a specific name and be organized temporally and/or by topic. In addition, Channels may contain Sub-Channels. The term *Platform* corresponds directly to Channel and can be regarded as a synonym of Channel within our approach. Both Medium and Channel correspond to the notion of C*hannel* in Lasswell's model.

Connections/Links[81]: Links, Connections between the elements represented by the communication model, are not addressed by Lasswell's model. However, they do form an integral element of (disaster) communication in today's highly interconnected and multi-platform world. Their detection and processing are essential as they often provide crucial context and background. Thus, they are also represented and handled explicitly by the current model, forming a further extension to Lasswell's original communication model.

A number of Links may exist between the above elements: Documents (D) Communicators (C), Entities (E) and Medium (M).

Examples of Links between the different elements include:

- C -> C: one Communicator following another Communicator

[81] The terms *Link* and *Connection* are used as synonyms

- C -> D: author or commentator of a Document
- D -> C: a Document mentioning the name of a Communicator
- D -> D: one Document referencing another Document via a hyper-link
- C -> E: Communicator referring to a topic or hashtag
- E -> C: an Entity being the name of a Communicator
- M ->C: a particular Communicator owning an account on a particular Platform
- E -> M: an Entity corresponding to the name of a Platform

As the above elements, Links can be characterized by a set of properties. They can be explicit, visible in a Document, such as a Link to an image on a photo-sharing site from within a message (D->D), or implicit, such as between a user (Communicator) and a message sent by them (C->D). They can be created by users ad-hoc – deciding on the spot whether to add a URL to an image (D->D) or may be created automatically (and always) by a mechanism implemented by the Platform – e.g. attaching of the author of a Document (C->D). They can be unidirectional, such as one user following another user (C->C) or bi-directional, e.g. two Documents on the web referring to each other (D<->D). Links can be direct or indirect, as in the case of a message mentioning a user (D->C) where the message's author will also be connected to the mentioned user indirectly (C->C) or in a sequence of interactions (C -> C -> C). In principle, all indirect links can be converted to become direct links. This may be beneficial for processing and efficiency purposes when operating on a graph-based model. Direct links are represented explicitly in the graph-model whereas indirect links can only be inferred by operations on the graph.

The framing of messages, the context in which a message is presented and received as well as information about the sender and Medium, is not addressed explicitly by the model developed here. Information quality - or perceived information quality - in particular factors of information relevance and reliability, are known to play a crucial role in users' propensity to share and re-share information (Koohikamali & Sidorova, 2017). Especially in times of *fake news*, these factors become of immediate importance and can be incorporated as an extension of the current model regarding the linking between Communicators and Documents (Bara et al., 2019).

5.2.1 Characterization and Classification of Elements

The main elements involved in communication as described above - Communicators, Documents, Entities, Media and the various Links between them - can all be characterized by a set of attributes. Associating such attributes to the elements of the communication model permits classifying them in a manner similar to *facets* (*facets* can simply be viewed as corresponding to these *attributes*[82]). Classifying the elements accordingly in turn allows to determine commonalities for modeling of data and process structures. These commonalities are subsequently used for alignment of data and operations.

Classification schemes of media are frequently organized in a hierarchical and enumerative manner (e.g. the Dewey Decimal Classification System, DCC (Dewey, 2011)), the Universal Decimal Classification (UDC) (UDC Consortium[83]) or the Library of Congress Classification, LCC (LCC, 2017). Alternatively, they may be based on facets or be of a hybrid nature (Hemerly, 2011; Spiteri, 1998). The disciplines of library science, Knowledge Management and the Semantic Web have witnessed ample activities in structuring and presenting information using combinations of hierarchical and facet-based approaches (Slavic, 2011). These are deemed more flexible and allow several advantages over strictly enumerative systems. Following this vein, the different types of Communicators, Documents, Media and Links are characterized and classified by a combination of attributes and not positioned in disjoint branches of a rigid classification scheme. This approach also removes the need for the strict classification of media into traditional and social ones altogether, rather allowing for a gradual classification focusing on discriminative elements where needed and common qualities where possible[84].

Based on the above definitions of the central communication-elements, attributes can be defined and used to:

1) characterize the elements, qualify them regarding the kinds of information they provide, where they appear, how they are used in disaster communication, how they interact and *act* in linking, etc.

as well as to

2) specify and classify the technologies involved in their processing, depending on the types of documents and data produced and exchanged, the kinds of contents involved, the nature of publishing and distribution offered, the types of linking which may occur etc.

[82] The terms *facet* and *attribute* are used in a synonymous way; facets only being referenced due to their established role in the classification of media. The mode of classification is inspired by faceted classification, however, here it is based on attributes and their associated values.
[83] http://www.udcc.org
[84] This quality also agrees with the tendency noted above of different kinds of media and their boundaries becoming increasingly blurred.

Item 1) above mainly concerns the construction of an object model, modelling the elements of the communication model and their relations whereas item 2) concerns the construction of a process model, specifying the possible operations and interactions between the entities of 1) and the population[85] of the object model with instances at runtime. The application of operations to elements of the object model is guided by common attributes.

5.3 Technologies

In addition to the elements themselves involved in the communication process, a suite of technologies is required for their collection and processing, taking into account their different qualities and the qualities of the platforms they originate from. The term *technologies* is used here to refer to the processing steps of a processing pipeline, each one transforming and enriching its input further (e.g. the Processing Resources in GATE (Bontcheva et al., 2002; Cunningham, 2002) or Text Analysis Engines (TAEs) in UIMA (Ferruci & Lally, 2004). These technologies are mainly concerned with the ingestion of data, the extraction of content, information and links between different elements as well as the processing and enrichment of different kinds of data. The former technologies (ingestion) focus on the connection to content sources (typically via APIs), segmentation and processing of content and meta-data. The latter concentrate on the content itself and largely qualify as technologies from the area of information extraction (IE), especially natural language processing (NLP) technologies (Manning & Schütze, 1999) and visual processing technologies (Bailer et al., 2005). All processing is performed with the ultimate aim of extracting the relevant information and making it available for analysis, search and retrieval purposes. Documents and Communicators, characterized by their attributes, form the connection between the different Platforms and the processing technologies implemented within the framework. These technologies[86], such as cleaning and tokenization tools for textual data, optical character recognition (OCR) for still-images or automatic speech recognition (ASR) for auditory data, can be characterized by the same sets of attributes as the Documents and Communicators. This provides a direct connection between Documents and Communicators, their attributes and the required and possible processing steps to handle and enrich them. This framework can also serve as a blueprint for requirements analysis and procurement and to position newly developing platforms and technologies. It may be used by first-responders when designing systems and devising new strategies for the collection and processing of information or to assess existing systems.

[85] *Population* in the sense of *instantiation*, adding instances to the model
[86] The specific technologies will be examined in detail below, see chapter 5.6

The diverse nature of the platforms typically involved in disaster-communication provides a wide spectrum of multimedia, multilingual and interlinked content. The set of technologies involved and required for the processing of such heterogeneous content has been investigated in the project QuOIMA (Backfried et al., 2016b).

Technologies to process and enrich content can broadly be classified into the following categories[87]:

- <u>Textual</u>: technologies operating on textual content, either from Documents or textual content associated with Communicators, such as account descriptions or textual meta-data
- <u>Auditory</u>: technologies operating on audio-content, non-speech or speech such as the audio-track of a video or a radio program
- <u>Visual</u>: technologies operating on still-images or videos, such as images from photo-sharing platforms or videos from TV or social media video-sharing platforms
- <u>Link-detection and classification</u>: technologies operating on textual, audio or visual content as well as operating on meta-data associated with the linking of the various elements of communication. Links may appear within one platform or across different platforms.

Table 3 below provides some examples of how the different kinds of technologies are related to the elements of the communication model.

	Textual	Auditory	Visual
Communicator	Detection of location in the description of an account	Identification of a speaker by voice	Identification of a person using a profile image associated with an account
Document / Entities	Detection of organizations in documents with textual-content or in descriptions and titles of auditory or visual documents	Keyword detection on sound files or audio-tracks of multimedia documents, audio segmentation and -classification	Logo detection on images or in videos, scene detection, object recognition
Link	Detection of URLs or hashtags embedded in text documents	Detection of names of users or persons mentioned by speaker	Detection of inserts of hashtags, URLs or account-handles within a video

TABLE 3 TECHNOLOGIES AND COMMUNICATION MODEL

[87] All of these technologies may also be employed to verify the authenticity of content, e.g. the detection of stock-images known to be used frequently for fake social media accounts.

All of the above technologies stem from research fields in their own right and often concern several sub-fields, such as text-mining, machine-learning, information retrieval and information extraction. Within this thesis they can only be treated insofar as their relevance to the subject of cross-media communication is concerned.

Textual Processing: technologies for textual content comprise a wide set of NLP and semantic technologies operating on levels of the individual characters (e.g. character-set detection or cleaning), to morphemes, sub-words, words (e.g. tokenization), phrases and sentences (e.g. sentence boundary detection), paragraphs, complete Documents or even to entire collections of Documents. Segmentation may create a sequence of text Documents from a single input Document, each characterized by individual homogeneous qualities and requiring different kinds of processing, such as segments of text in one language followed by text in another language. Processing of other data-types, such as auditory-data, may result in text (the transcript) which can subsequently be processed by text-processing technologies. However, these texts may qualitatively differ from actual texts input by humans, e.g. in granularity, style, punctuation, structure or the kinds of errors present (*typos* vs *recognition errors*). Text-processing with the aim of information extraction - turning unstructured and semi-structured information into structured information - comprises a wide-range of algorithms and technologies operating on all levels described above (Okurowski, 1993; Balke, 2012). Frameworks like GATE (Cunningham, 2002; Bontcheva et al., 2002; Maynard et al., 2017), the Stanford CoreNLP (Manning et al., 2014) or UIMA (Ferruci & Lally, 2004) offer technologies, components and models which can be combined into workflows (pipelines) on many levels for processing and semantic enrichment of content. Components within these frameworks typically focus on Information Extraction and NLP tasks (but are limited to content in textual format).

Auditory Processing: technologies operating on auditory content comprise technologies operating on the raw audio, improving quality, segmenting and classifying the audio for further processing (e.g. classification as speech or non-speech and *diarization*[88]), technologies operating on non-speech auditory content (e.g. audio-event detection of explosions) as well as technologies operating on speech content (ASR or keyword-spotting). As in textual processing, segmentation may partition a Document into several Sub-Documents, each with their own specific character (e.g. sequences of audio containing music/background noise and segments containing speech or segments with audio in different languages when Documents from multi-lingual environments are processed)

[88] Partitioning of audio into homogeneous stretches taking into account acoustic conditions and speaker characteristics

Visual Processing: processing of visual content includes technologies for individual (still) images as well as for videos. This includes the improvement of image quality, the creation of visual descriptors, spatio/temporal segmentation, the detection of key-frames, the detection and identification of faces and objects, the detection and recognition of text (OCR) or scene-classification. As for the other types of processing, segmentation into smaller, homogeneous units may be carried out during initial processing steps, such as the segmentation of a broadcast video into scenes of disaster and studio settings or the detection of images containing maps (Backfried et al., 2010).

Meta-Data Processing: Platforms typically provide a set of meta-data together with the Documents and accounts present on them. This meta-data may be of a textual nature - in which case the application of the above text technologies is possible. It may be visual (e.g. an image associated with an account) or may also contain particular kinds of semi-structured and structured-data requiring specific processing, such as geo-coordinates or time-stamps.

The (geographic) localization and temporal alignment of Documents, Communicators and interactions form central requirements for situational awareness and consequently crucial dimensions for the analysis and visualization of data:

- Positioning, GPS-coordinates: localization data may be provided in a set of different formats. These formats can be converted to a single standard format for further processing, analysis and visualization.
- Timestamps: various timestamps may be associated with a Document, such as the Document's creation, publication or last-modification date and time. All of these may be provided in different formats and often contain additional information such as the time-zone or daylight savings time settings. These formats can be converted into a standard format and unit, e.g. into UTC (coordinated universal time) in order to be able to make them comparable and accessible for further processing, analysis and visualization.

Link Processing: the detection and processing of links concerns all of the elements listed above, as links occur in a variety of manners. Processing emphasizes the detection of links between the Communicators, Documents and Entities. It comprises technologies from the above list and operates on textual, visual and auditory content as well as on meta-data. Links may occur within a Document or across Documents and within a Platform or across multiple Platforms. Processing will reflect this as certain kinds of operations can only be performed once data from elements from across Platforms becomes available.

Examples of the detection of links:

Textual: e.g. detecting the name of (another) Communicator in a text-Document

Auditory: e.g. the detection of a hashtag when spoken by a reporter

Visual: e.g. the detection of an account name while inserted on TV

5.4 Data and Process Model

The above chapters describe the motivation and approach to characterize and classify[89] the relevant elements of disaster communication[90]. These elements and their attributes provide the ingredients for a generic data model representing all elements involved as well as for a process model representing possible operations which can be applied to them in a generic manner. Content on the relevant platforms is provided in visual, auditory and textual form or as a combination (multi-media) and is in many instances accompanied by meta-data (mostly in textual form). The different media-types are reflected in the kinds of operations possible on them, the *affordances* offered by them (Bucher & Helmond, 2017) and in the ways that links between individual elements can be detected and processed.

In the current thesis, both models are dealt with on a conceptual level and aim at a logical description of the modelled entities and their interrelations. The primary aim of the models is not to provide a detailed specification or to directly allow for an automatic implementation. They also do not address organizational or business-related aspects. Rather, they are meant to provide a general and structured way to describe and understand the specificities of the entities involved (data model) and to capture the main characteristics of operations and activities which can be performed on them (process model)[91]. Some aspects of the resulting model – a graph-based model as will be outlined below - as well as the processes employed to create, modify and query it have been implemented to demonstrate the viability of the approach and evaluate some key aspects.

5.4.1 From Diverse Elements to Generic Ones

Whereas an individual platform may refer to a Communicator as a user or as an account or refer to a document as a tweet, post, message, event or status update – each with its own, proprietary legacy set of attributes - the generic model tries to map these heterogeneous entries to uniform ones. This

[89] In the current work, the term *classification* is preferred over *categorization* (Jacob, 2004).
[90] From an information processing perspective
[91] As such, the process model does not concern the construction of an abstract description of the activities by which software is developed (Finkelstein et al., 1992).

allows to refer to all entities and their attributes in a generic way and to define possible/required operations on them. Likewise, actions performed which may be referred to as commenting, replying, sharing, re-tweeting or re-blogging can be mapped to generic activities creating Links between Communicators, Documents and their content. To arrive at this mapping, the individual attributes encountered on different platforms are organized semantically in a first step and are then converted into a set of common attributes. These attributes form the link between Platforms, Communicators, Documents, Entities and processing-technologies and provide a *common denominator* by allowing to align them with each other. The elements themselves as well as the attributes characterizing them are derived in an iterative and incremental manner, taking into account the set of platforms and technologies involved in disaster communication. Together, they form an integral part of the data and process model. The initial set of platforms was determined by the requirements of first-responders as identified in the QuOIMA project[92] and extended during work on this thesis.

The choices and requirements of how to combine different platforms and media could be confirmed by the author in various conversations with first responders (Austrian Army, Austrian Red Cross) during national events (e.g. visit of representatives of the Austrian Fire-Brigades) as well as through discussions and presentations at international conferences (ISCRAM[93], ICT-DM[94]) and is further supported by projects and activities such as described in chapter 3.5, in Rogers (2017) and Reuter et al. (2018a).

It should be stressed that the resulting model is not limited to these exact platforms, but rather is envisioned to evolve over time as new platforms are identified to play an important role during disasters[95]. The ability to accommodate additional platforms as well as changes regarding existing platforms[96] in a dynamic way also in the future is indeed expected to be one of the key advantages of such a generic model.

All types of Media concerned - across the spectrum from traditional ones to social ones - can be characterized using the resulting common set of attributes. Individual Platforms can be modeled as instances of more general class of Media. Technologies, required to process these Media, Documents and Communicators can be classified along the same dimensions, allowing for a direct alignment of Platforms and technologies. In addition, within-media and cross-media links can be modelled by the

[92] During that project, a relational approach was chosen and the implementation carried out using MySQL.
[93] http://www.iscram.org
[94] http://ict-dm2018.net
[95] For example, during the work on this thesis, the platform *Periscope* was introduced and integrated into the model without difficulties.
[96] E.g. changes in the availability and scope of data available on Facebook following the Cambridge Analytica Scandal, (https://en.wikipedia.org/wiki/Facebook%E2%80%93Cambridge_Analytica_data_scandal, accessed on 08/10/2018)

same attributes. Their presence on certain Platforms and technologies to detect and process them can be represented within the framework. Chapter 5.6.5 provides an overview and table with the alignment. Figure 23 below provides a schematic overview of the approach and framework which will be discussed in detail below.

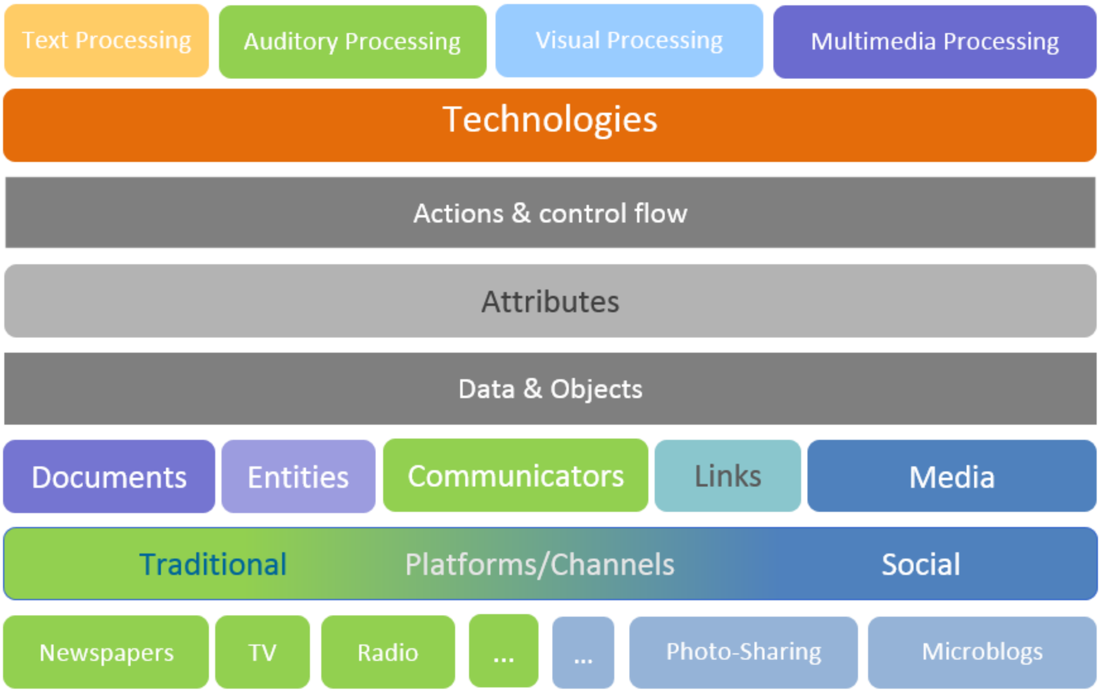


FIGURE 23 SCHEMATIC OVERVIEW OF FRAMEWORK

5.5 Data Model

The **data model** is developed by aggregating all elements corresponding to the communication model from the set of platforms and media which were found to be employed during disaster communication and abstracting from these the common, generic elements. Its aim is to describe and model the specificities of the individual elements involved, each one of them being characterized by a set of attributes which are aggregated in an iterative manner. Elements can be connected (*linked*) in a variety of ways; these links can be described by a further set of attributes. The resulting structure - elements (nodes or vertices[97]) with attributes, being connected by links (arcs or edges[98]) with attributes - provides a conceptual model of the domain and suggests a graph structure. The *Property Graph Model (PGM)*, a directed, labeled, attributed multi-graph model (Rodriguez & Neubauer, 2009; Robinson et

[97] The term *node* will be used in this thesis rather than *vertex.*
[98] The terms *link* or *edge* will be used (as synonyms) in this thesis rather than *arc.*

al., 2015) is a natural fit to this structure and is thus chosen to model the elements and relations involved.

The construction of the data model is carried out in two steps:

- Creation of a generic model (abstracting from individual platforms)
- Creation of a Property Graph Model (creating a graph representation from the generic model)

These two steps are outlined in Figure 24 below.

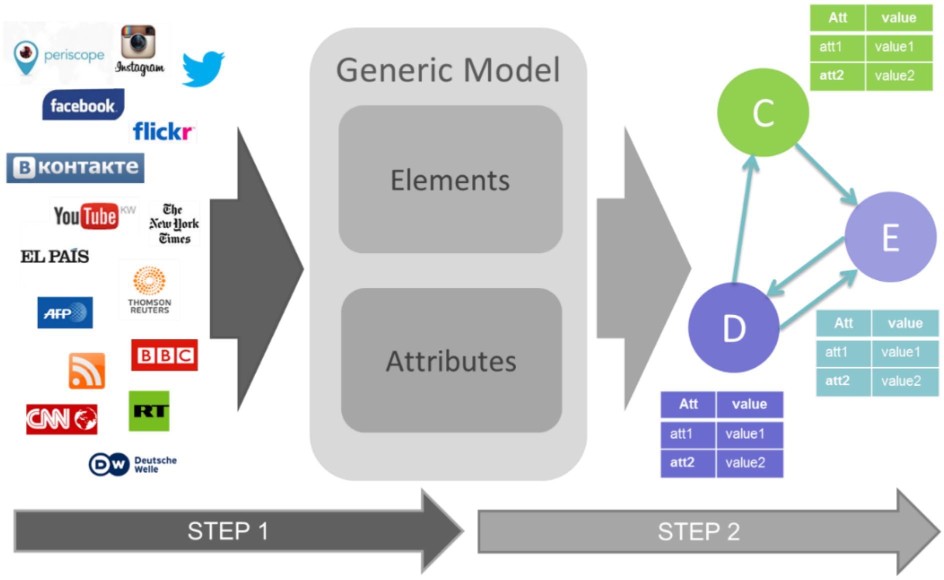


FIGURE 24 TWO-STEP MODEL CREATION

The result of the first step generically represents Platforms and technologies and can be used to align them with each other to determine the requirements and possibilities of processing. It also presents the basis for the construction of the PGM in the second step.

5.5.1 Method to Derive Elements and Attributes

In the first step, depicted schematically in Figure 25 below, all elements and their attributes are identified, aggregated and grouped in a semantic way by iterating over all investigated platforms (and their respective implementation, naming scheme and structure). This step leads to a common view, a uniform and generic representation of the various elements involved.

In a second step, this representation is used to construct a PGM reflecting these entities and their relationships, converting the uniform, semantic representation created in step 1 into a graph-based representation on which the process model operates.

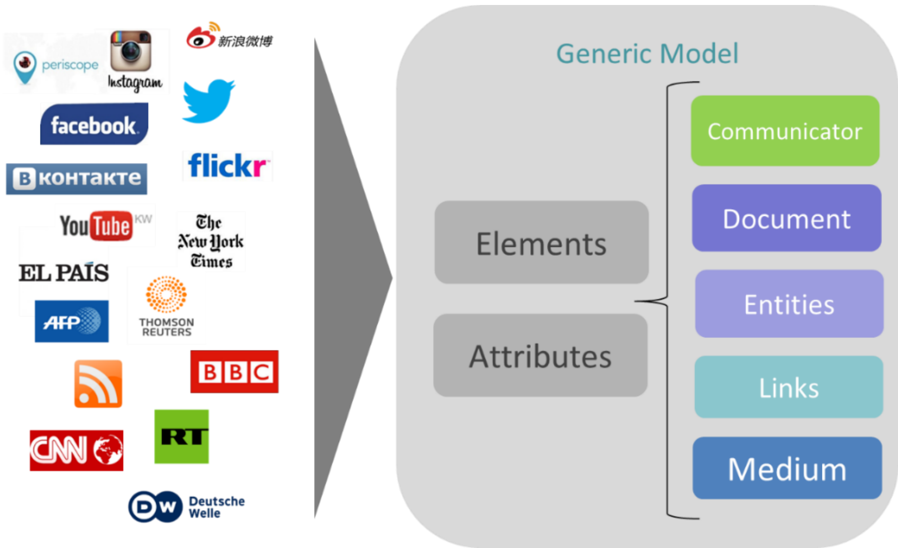


FIGURE 25 FIRST STEP OF MODEL CREATION

The **procedure**[99] followed for the first step consists of the following activities.

For each platform:

- Identify the elements corresponding to the communication model (*Medium*, *Communicator*, *Document and Entities)*
  - If these platform-specific elements have not been encountered, add them to the respective element
- Determine their attributes
  - If the attribute present on the platform has not been seen before, add to set of attributes
  - Group attributes according to their semantic and syntactic properties
- Identify types of links and how these are realized
  - If a type of linking offered by the platform has not been encountered before, add to the set of links
  - Group and refine links for the elements of the communication model (*Communicator*, *Document* and *Entities*)
  - Determine and group attributes of these links

[99] This iterative procedure was carried out over all platforms identified by QuOIMA and by literature research on further relevant projects.

- Once all platforms have been iterated over
    - Aggregate all collected attributes and distill into generic ones
    - Aggregate all collected links and distill into generic ones

The resulting model incorporates the elements of the communication model and their attributes in a systematic manner. Further emerging platforms can be accounted for by following the same procedure.

Once step 1 has been carried out and elements and attributes have been aggregated for all platforms, the attributes can be used to relate Communicators, Documents, Entities, Platforms and technologies operating on them. The generic set of elements and attributes allows to represent them and their properties on a common level and is used for the creation of the process model below.

In a second step, a PGM is constructed from this first, conceptual representation as depicted in Figure 26 below. For the creation of the PGM, the individual attributes are further refined and syntactically specified by assigning data types and formats. For example, all time-related attributes are converted to the format *YYYY-MM-DDThh:mm:ss* and to UTC. Such standardization is of crucial importance as different platforms use different formats and may set their attributes according to different time-zones, making cross-platform comparison and relation difficult or impossible if not accounted for correctly[100].

Within this thesis, the actual construction of the graph is only carried out to the point of being able to demonstrate its viability and for selected cases for evaluation. However, it should be apparent that the remainder of the graph could be constructed in analogy to the exemplary work carried out here.

[100] For search, visualization and analysis, the local time of an organization or of an analyst can then be uniformly applied to provide a consistent localized point of view: "How did things evolve viewed from *my timescale*?" This is of importance especially in international and cross-border settings of disasters.

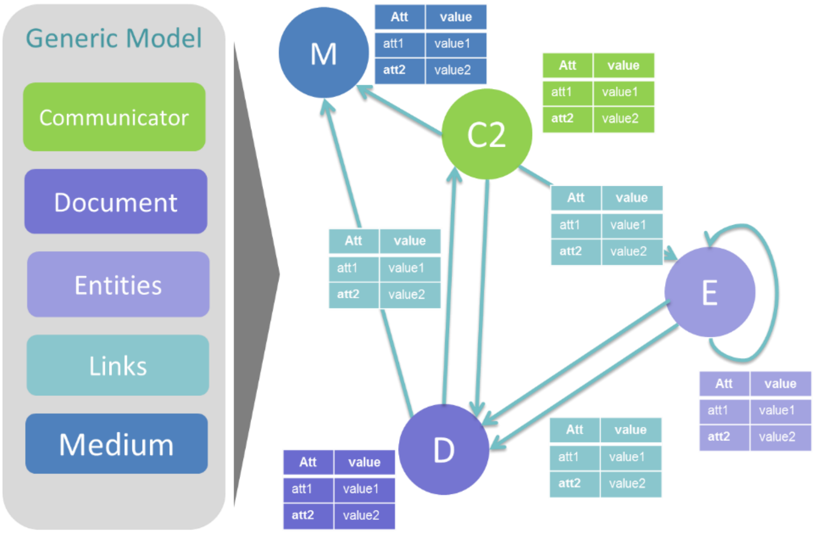


FIGURE 26 SECOND STEP OF MODEL CREATION

The different attributes, their characteristics and how and where they can be detected and processed within media will be discussed below. Furthermore, the creation process of the property graph based on this model will be outlined.

5.5.2 Media/Platforms

As described above in chapters 3 and 4, a series of media and platforms are typically employed for communication during disasters by individuals and organizations. The following types of media and platforms have been identified as being of relevance for disaster and crisis-communication (Backfried et al., 2013b; Backfried et al., 2014; Backfried et al., 2015; Backfried et al., 2016; Koohikamali & Sidorova, 2017; Reuter & Kaufhold, 2018; Reuter et al., 2018a) and have been integrated into the data model:

- TV (digital, analogue, IP, sub-titles, teletext)
- Radio (digital, analogue, IP)
- Press-Agencies (e.g. Reuters, AFP)
- Print-Media (newspapers, journals, magazines)
- Web-Pages (government, NGOs, media-corporations)
- RSS-feeds, ATOM-feeds
- Fora
- Blogs
- Digital-Distribution (newsletters distributed via email)
- Microblogs (Twitter)

- Social Networks (Facebook, VKontakte)
- Podcasts (NGOs, individuals)
- Photo-Sharing (Flickr, Instagram)
- Video-Sharing (YouTube, Periscope)
- Instant-Messaging (Telegram, WhatsApp[101])

The elements of the communication model encountered in the above platforms are discussed below regarding their relevant attributes. For each element, the associated links are discussed in more detail as linking between the elements and within and across platforms represents a key factor for cross-media communication. The remaining attributes are discussed from the viewpoint of the respective element.

Every instance of an element of the communication model is assumed to possess a unique ID (possibly derived from the ID of the originating platform) allowing to determine its uniqueness and to avoid duplication[102]. If the ID is not determined by the originating platform, it is created and assigned by the interface accessing the platform at the time of first ingestion[103].

5.5.3 Communicator

Figure 27 below depicts the model for the element *Communicator* along with its attributes which have been determined by following the iterative procedure outlined above. It comprises attributes from across the complete spectrum of platforms from traditional media as well as from social media.

[101] Closed (non-open) platforms such as Telegram, WhatsApp or Linked-In can likewise be integrated into this scheme. As the current thesis focuses on open source information only, these are not considered here in detail. However, *Telegram* and *WhatsApp* have indeed already been integrated as of 2019/01 which provides further support for the generic and flexible nature of the proposed approach.
[102] The detection of duplicates can be based on this ID or on the content. In the latter case, it may be possible to also detect *near-duplicates*, documents which are not identical but resemble each other to a large extent (see e.g. Broder, 2000)
[103] Certain types of elements may be ingested several times, in order to retrieve updated or changed states, e.g. the number of followers of an account typically changes over time requiring re-ingestion to maintain an up-to-date view.

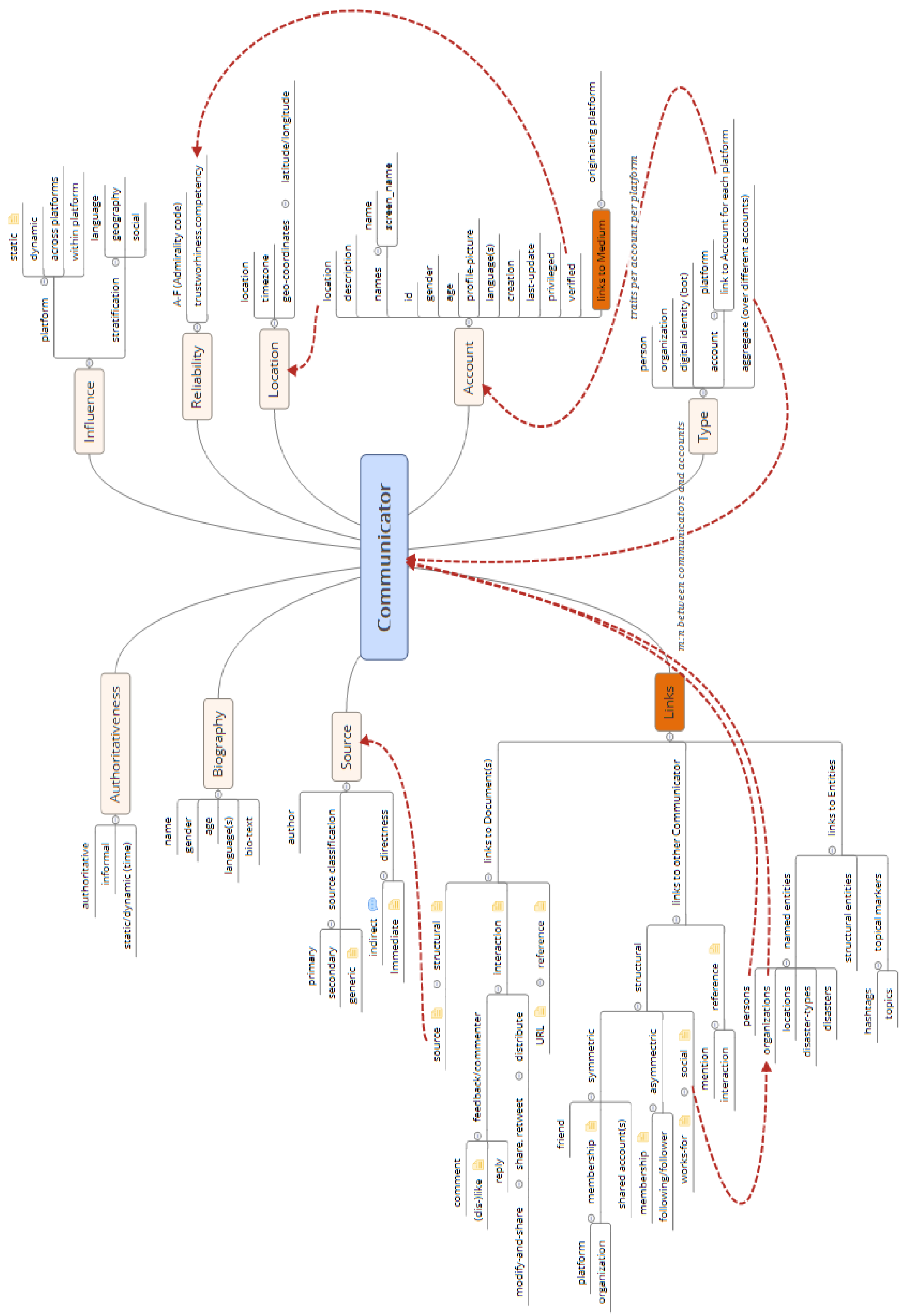


FIGURE 27 MODEL OF COMMUNICATOR

Following the proposed method, attributes have been grouped semantically and are discussed according to this grouping below.

5.5.3.1 Type

Communicators may be individuals (persons), organizations or digital identities (such as bots). They may appear in person for media such as TV, radio or videos on social media, author and comment articles on press-feeds, web-pages, blogs or printed papers or appear on social media platforms as authors, commenters or distributors. They may own accounts on one or several platforms which may lead to further information about them via associated meta-data for these accounts[104]. Actions and interactions then actually take place via these accounts. However, from the account, we assume that the Communicator can be determined[105] and in any case generally refer to the links from the standpoint of the Communicator rather than only from one of their individual accounts.

5.5.3.2 Source, Reliability, Authoritativeness

These attributes concern and define the Communicator's relationship regarding the creation or transfer (forwarding or passing-on) of a Document. Authorship in this context does not address copyright and legal issues but is meant only to state who created/authored a certain Document. For textual Documents, this information may be part of the Document and visible in the Document's content. For audio Documents, the voice of the speaker may provide information about authorship. For visual Documents, the author may be visible him- or herself. In all cases, the author may appear in the associated meta-data. Groups of Communicators and organizations may be regarded as joint authors of a Document. For cases like news-presenters working for a TV network or journalists writing for a newspaper, both, the individual as well as the associated organization may be regarded as authors, with the organization being linked only indirectly as depicted in Figure 28 below.

[104] https://blog.globalwebindex.com/chart-of-the-day/internet-users-have-average-of-7-social-accounts
[105] For cases where the Communicator's identity cannot be determined, an anonymous Communicator is assumed. Anonymity does not affect the potential types of links in principle and may be able to be *resolved* upon the processing of further data, i.e. an identity assigned through detection of links between accounts.

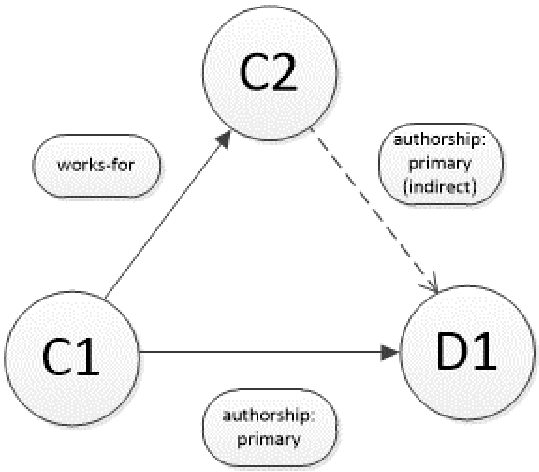


FIGURE 28 DIRECT AND INDIRECT AUTHORSHIP

The model does not account for script-writers but rather attributes authorship to the *messenger* – which for practical purposes provides sufficient information and context. A spokesperson of a first-responder organization may thus also be regarded as the author of the conveyed message. On occasion, the individual may not be known and only the organization which a Document originates from is available. In such cases, the organization in its role as Communicator may assume the role of author.

How closely a Communicator is related to a Document's content, whether they are reporting live from the location of a disaster, reading a pre-fabricated script in a studio or simply relating reports provided by others, determines whether they can be classified as a primary, secondary or generic source (*generic* here in the sense that the exact relationship is unknown). During disasters, especially primary sources, providing eyewitness reports from the ground, are of great value when assessing the situation at the disaster site. The relaying of information (re-tweeting) during disasters has been found to be an indicator of recommendation of the original source and to provide further support for the reliability of information (Parilla-Ferrer et al., 2014). Communicators and Documents may be related by different levels of authorship: a Communicator may be a primary-source for one Document (which is a primary source itself) and a secondary source for another Document (which is a secondary source itself).

Communicators may be an immediate source, when Documents are received directly from them (e.g. an anchorman presenting news on TV, a spokesperson of an NGO reporting on events at a press-conference, a user tweeting or commenting information about the developing situation). For platforms allowing for referral of information (*passing-on* of information), involving several Communicators in sequence, the most recent Communicator in such a chain may be regarded as the immediate source with all previous ones being regarded as indirect sources. These Communicators may not be known or only be known to some limited extent (e.g. with a distance of two for the re-tweeting of a tweet). The reliability and authoritativeness ascribed to a Communicator – also through verification mechanisms on platforms where a Communicator owns an account - may also be combined with the source classification on certain Documents when estimating trustworthiness, veracity and authoritativeness of contents.

5.5.3.3 Influence

The influence of a Communicator is a multi-dimensional, soft factor which is difficult to define or determine. Within the domain of social media marketing, a variety of measures and KPIs (key-performance-indicators) have been developed to estimate the *influence*, *impact*, or *reach* of a particular message, user or account. These measures commonly combine static factors such as the number of followers or friends with dynamic factors such as the number of replies, forwards or mentions and are of special importance in the field of influencer-driven marketing[106]. For traditional media, influence is commonly expressed in terms of audience measurement (readers, viewers or listeners) which are largely based on geographic and socio-economic factors. Overall estimates for a Communicator may be generated from scores of associated accounts and information from public sources or companies specializing in media observation and analysis (e.g. Nielsen[107]). Attempts to combine several diverse factors from multiple platforms and media, such as the Klout-Score[108] have failed in establishing a standard for condensing a multidimensional factor into a single score. Furthermore, the generality and sensibility of such approaches must be viewed with caution.

Within the context of disaster communication, it may be assumed that influence can be viewed from a social or geographic dimension with language playing an important role for the distribution and acceptance of information. Measures developed for social media marketing (within a single platform or across multiple platforms), based on static and dynamic features of traditional and social media may be considered. For TM, the reputation and credibility of an organization or individual may provide important clues. However, any method of combination and approximation has to consider the variety of factors and different scales and granularities of measurement and should not aim to compress such a diverse measure into a single number. Rather, depending on the particular platforms and types of Communicators involved, a portfolio of different indicators should be provided.

[106] https://en.wikipedia.org/wiki/Influencer_marketing
[107] www.nielsen.com
[108] https://en.wikipedia.org/wiki/Klout

Even though influence can be considered primarily as an outbound, communicative measure, it is represented in the current model[109] as the messages received can also be assumed to be read by a portion of the affected population and consequently influence the behavior of these readers. Opinion leaders and central distributors of information may be identified by such measures as centers of exchange and gateways of information flow. For practical purposes, though, it may suffice to classify influence into general categories in order to be able to assess possible effects of statements on a broad scale. In combination with other factors such as authoritativeness or reliability (discussed above) it may be possible to estimate the influence of messages in a more principled manner.

5.5.3.4 Accounts, Location, Biography

One or several accounts may be associated with a Communicator, in particular for social media platforms. These accounts may be hosted on one platform (e.g. one person with several accounts on Twitter) or on several platforms (one person with one account on Twitter and another account on YouTube). Several Communicators may share one account (as is allegedly the case with the account @therealdonaldtrump on Twitter[110]). Accounts may provide a variety of information on the account-holder, such as a description, name, age, profile-image, time-zone, location, account creation time and the time of the last update. Several attributes have been used when determining the authenticity of an account, such as the use of stock-images for the profile-image, the use of automatically generated screen-names or the frequency with which this name is updated (Subrahmanian, 2016). Stylometric measures and measures of authorship may be applied in combination with other measures to assist in the determination or verification of the identity. It is known that accounts created shortly before an event, e.g. during the onset of a disaster, can oftentimes not be trusted. Account activity and alleged time-zone or location may provide further clues as to the veracity of information. Since the localization of information is one of the highest priorities during disasters, the different ways of how to retrieve and combine such information is of utmost importance. If a Communicator is linked to several accounts, the combination of location information across all accounts and related Documents can be used in determining the likely location and the reliability of the location information for the Communicator and their Documents. Frequently, location information also includes latitude/longitude information in addition to a verbal description of the location. Such more precise location information in the form of geo-coordinates provides valuable information to pin down events. If information about the languages spoken by a Communicator is available, this might be used to triangulate information about Documents authored or forwarded by

[109] Which focuses on inbound information

[110] https://www.wired.com/story/tell-when-someone-else-tweets-from-realdonaldtrump on 2018/04/22

them and allow to assess trustworthiness. Age and gender information may be useful for crowd-sourcing activities, e.g. when asking individuals to provide further information on a crisis or perform tasks (crowd-sourcing) and likewise allow to frame trustworthiness and reliability. Accounts may be verified Accounts on certain platforms (e.g. Twitter or Facebook) indicating a certain level of authenticity and authoritativeness. Likewise, certain Accounts may be privileged, e.g. in administering a forum or Facebook-page. All Accounts within the model are linked back to their originating platforms in order to allow for direct traceback[111].

5.5.3.5 Links

Figure 29 below depicts the different kinds of Links encountered which involve Communicators. Communicators may be linked to Documents (C -> D), to other Communicators (C -> C) as well as to Entities (C -> E) by a variety of mechanisms.

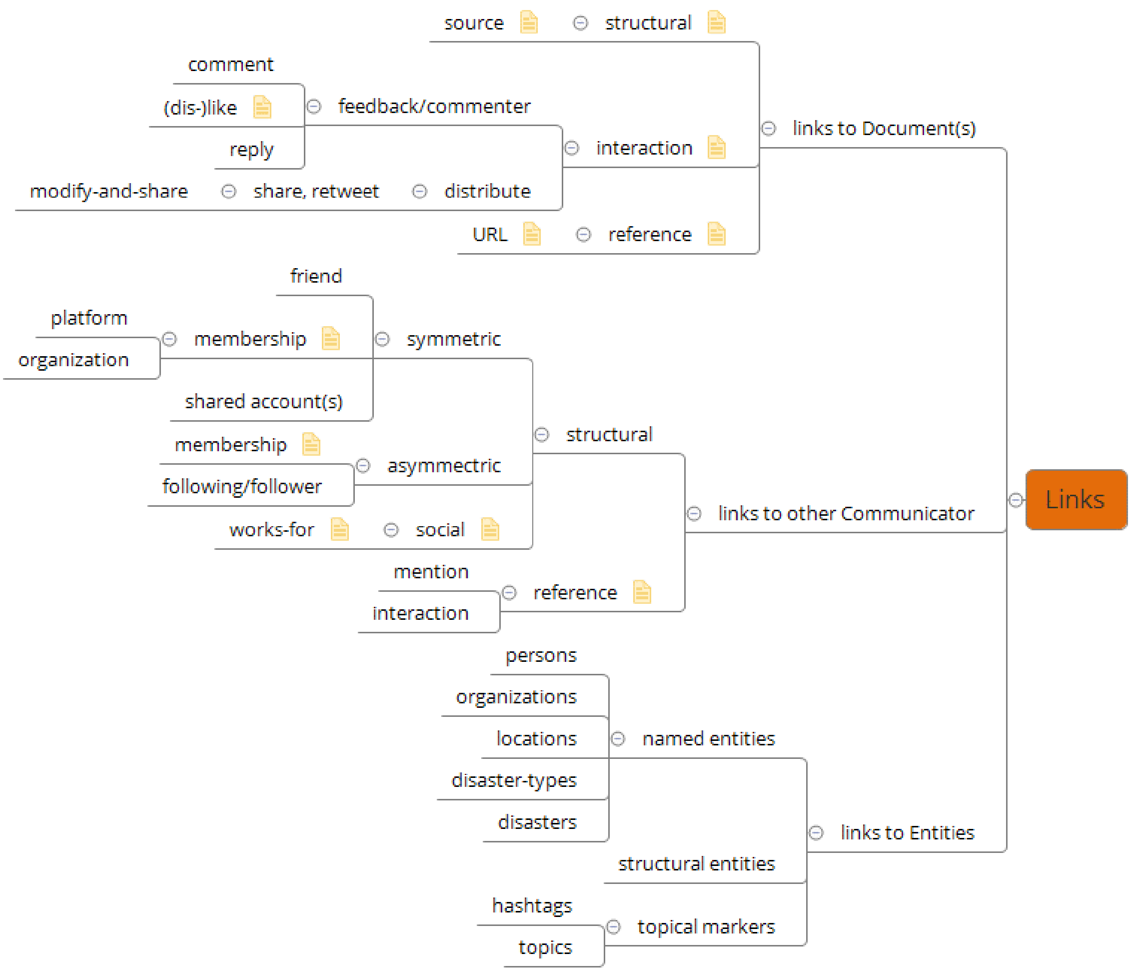


FIGURE 29 LINKS INVOLVING COMMUNICATOR

[111] Also to check whether an account has been closed, removed or is still active.

5.5.3.5.1 Links to Document (C -> D)

Communicators can be the authors of Documents or interact with Documents in a variety of ways (for platforms allowing for interactivity). A Communicator can be a primary source[112] of a Document - in case of firsthand experience or eyewitness reports - or a secondary source in case of summarizing or discussing events, e.g. broadcasting from the TV studio or writing an article for a newspaper. Being a primary source and detecting that someone is indeed a primary source may be crucial during disasters. In case of interactions, the Communicator may be distributing or commenting on someone else's Document as a secondary source. For cases where this relation cannot be determined, the Communicator can be considered an unknown or generic source. Further attributes, such as the Communicator's authoritativeness and reliability, their location and information about languages spoken provide context for the interpretation and assessment of related Documents.

For mechanisms of *referral* or *passing-on* of a Document, such as re-tweeting or forwarding, a whole cascade of Communicators may be involved. At the time of processing, it may not be clear or possible to state who the original author of a Document was or even whether the author of the Document was a primary, secondary or unknown source. Especially, if the Document has been modified (even only slightly), these questions may not be answerable or traceable at all. The most recent one of such a sequence of Communicators, however, will be accessible through the collection mechanism and can be marked as being the *immediate source*. All other Communicators known to have taken part in the communication chain may be regarded as indirect sources. In case a Document references another Document (e.g. via inclusion of a URL within a tweet), the Communicator also becomes an indirect source for the referenced Document (e.g. for the web-page referenced by the URL embedded in that tweet). Indirectness between a Communicator and a Document can thus be established via sharing/commenting or referencing mechanisms. Links between Communicators and Documents are either permanent, as in the case of authorship, or temporally limited, as in the case of interaction with a Document (where the time to interact is usually short and the end or beginning time may be utilized). The link itself remains in both cases, however the attributes regarding time differ.

5.5.3.5.2 Links to Communicator (C -> C)

Links to other Communicators may be of a structural nature, by means of interaction with another Communicator's Documents or by reference of other Communicators (resp. reference of their accounts). Structural relationships comprise asymmetrical ones such as follower/following or

[112] Definition, e.g. http://infoguides.pepperdine.edu/c.php?g=287355&p=1915752 on 2019/04/04

symmetrical ones as having befriended each other on a platform or belonging to the same group in a forum. In addition, it may be known through *real-world information* that some relationship exists between Communicators *outside of* the observed Platforms (e.g. they are known to work together for the same organization). Structural links provide direct links between Communicators. In addition, Communicators may share *accounts* on a Platform, i.e. have access to and use the same account. Referencing other Communicators' Documents or referencing Communicators by mentioning their name or an account associated with them provides further links between Communicators. All relations between Communicators may in principle also be realized through respective accounts associated with them. Links between Communicators may exist for a limited timespan only (which might be open-ended) or may be instantaneous and temporally limited in case of reference.

5.5.3.5.3 Links to Entities (C -> E)

Communicators may be linked to Entities via Documents authored by them, by association in the meta-data of their associated accounts or by direct association to specific Named Entities. In the latter case, Communicators may correspond to specific persons or organizations, which in turn correspond to a single concept or referent. As such, they can be represented by Named Entities in the current model. Named Entities corresponding to a Communicator may be mentioned in a Document, creating a (direct) D->E link, a further link to the referenced Communicator (E->C-ref) and indirect links from the author of the Document C to C-ref (C->C-ref) as well as from the Document to the referenced Communicator, D->C-ref. Further Entities (Named, Structural or Topical Markers) may be referenced by a Communicator in a Document or in meta-data associated with the Communicator or associated accounts. A direct (C->E) link from the Communicator to the referenced Entity is created in the former case, an indirect one in the latter case.

5.5.3.5.4 Examples of Links Involving Communicators

The following examples of links between Communicators and other elements illustrate the above points.

Indirect links between two Communicators (C1 and C2) created due to their sharing of a common account (A1); a structural, symmetric link:

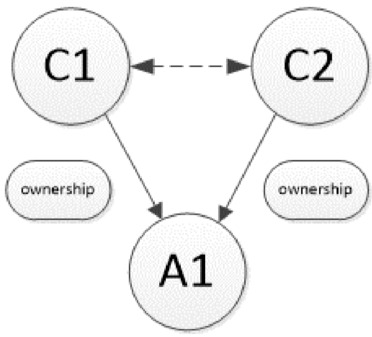


FIGURE 30 COMMUNICATORS SHARING AN ACCOUNT[113]

An identical situation of linking arises if Communicators (C1 and C2) are members of the same group on a SM platform or the same forum or working for the same organization (groups and organization being represented as Entities, E).

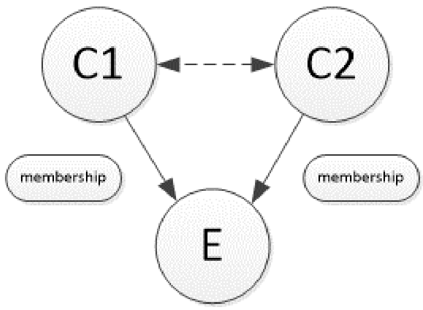


FIGURE 31 COMMUNICATORS LINKED THROUGH MEMBERSHIP

Indirect link between Communicators (C1 and C2) due to one of them *following* the other one (resp. their Accounts (A11, A12 and A21) following one another on a particular Platform). The structural relationship between Communicators is inferred from the structural relationship between (their) associated Accounts.

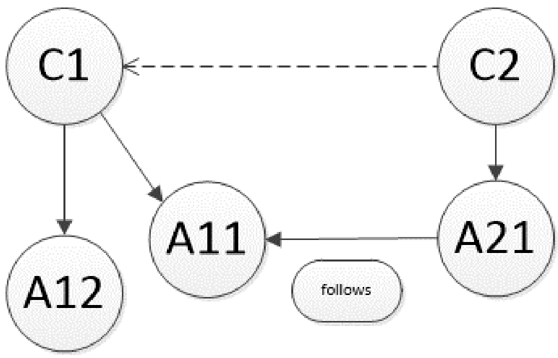


FIGURE 32 COMMUNICATORS LINKED THROUGH STRUCTURAL RELATIONSHIP (FOLLOWING)

[113] Directed links are assumed throughout this work – if bi-directional links are shown - as in Figure 30 - they represent two individual directed links.

Indirect link between two Accounts (A11 and A21) and two Communicators (C1 and C2) due to one of the Accounts interacting with Documents authored by the other one (e.g. *like, wow, hate, favorite*, …). Both, Accounts and Communicators are linked indirectly by this interaction.

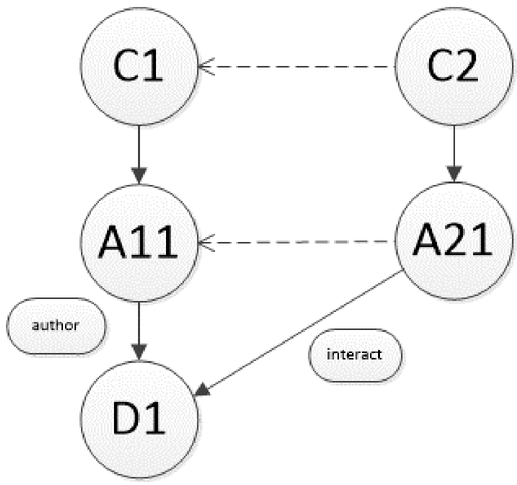


FIGURE 33 COMMUNICATORS LINKED THROUGH INTERACTION

Indirect link between Communicators (C1 and C2) resulting from one *mentioning* or *citing* the other one in a Document (D1), e.g. textually in a tweet or verbally in a broadcast.

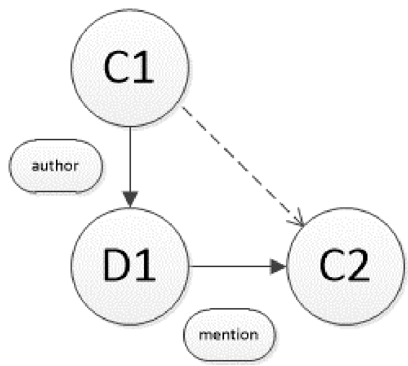


FIGURE 34 COMMUNICATORS LINKED THROUGH MENTION

Indirect link between Communicators (C1 and C2) due to one Communicator authoring a Document (D2) which comments on a Document (D1) authored by the other, e.g. by C2 commenting on a SM post of C1.

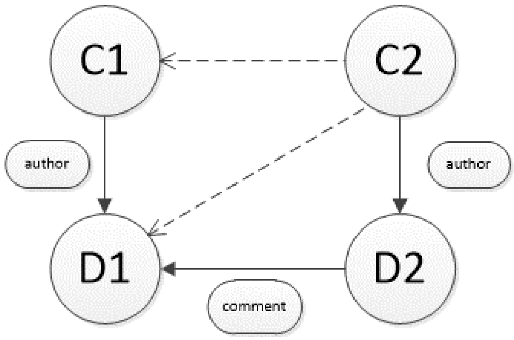


FIGURE 35 LINKS BETWEEN COMMUNICATORS CREATED BY COMMENTING

Links between Communicators (C1 and C2) or between associated Accounts (A11, A12 and A21) can be assumed always to *propagate*. In case two Communicators are related via a link, all associated Accounts are likewise related and vice-versa (resulting in direct or indirect links, bidirectional links in figure Figure 36 symbolize the various kinds of possible links in this situation).

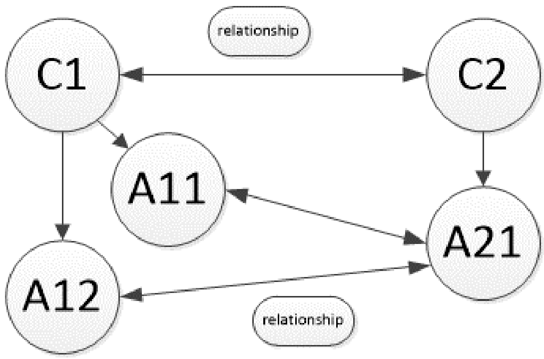


FIGURE 36 LINKS BETWEEN COMMUNICATORS AND ACCOUNTS

5.5.3.6 Conceptual Model for Communicator

The following attributes, depicted below along with their (basic-) types, can be identified for the *Communicator*. Types are in agreement with technologies developed in the process model below.

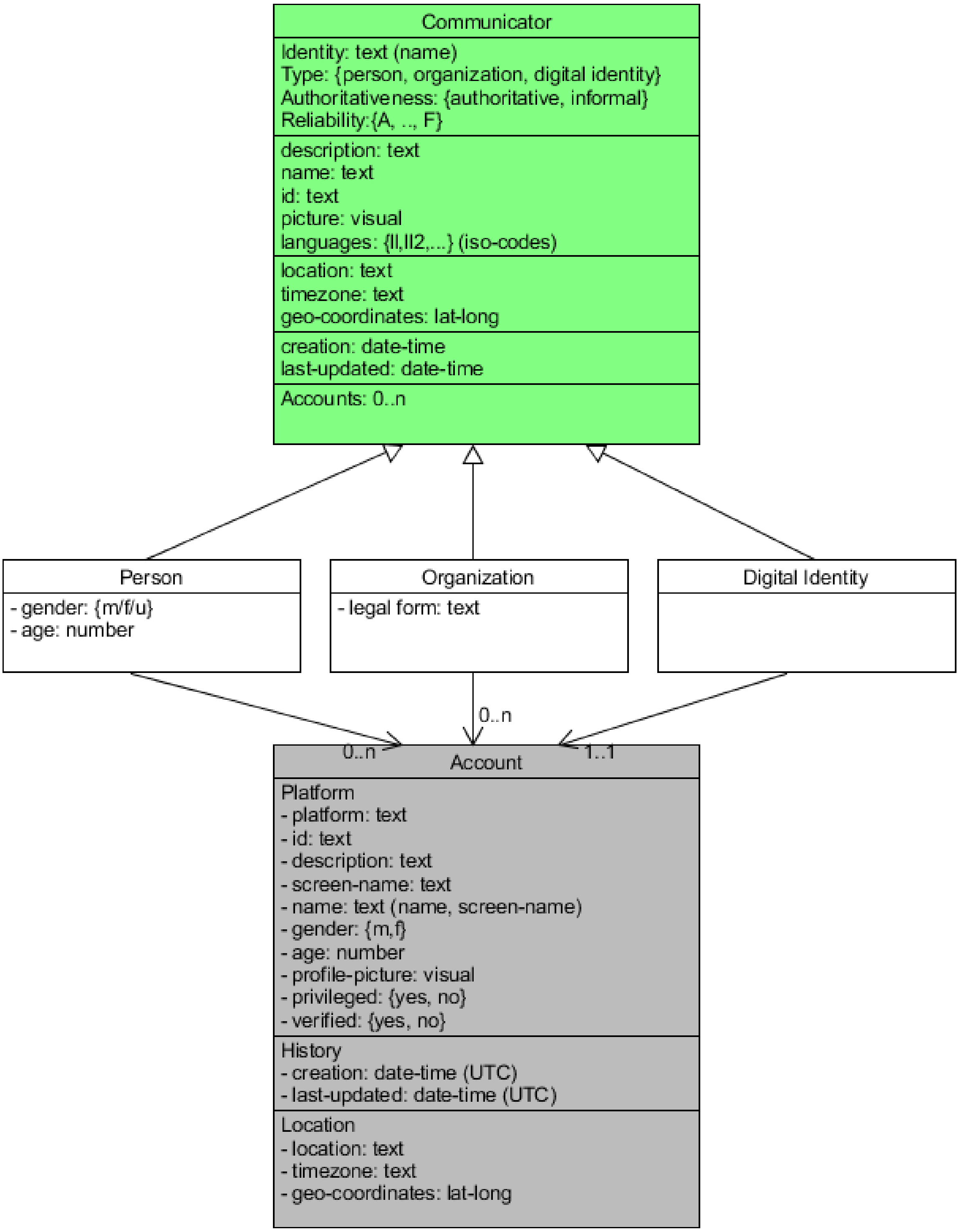


FIGURE 37 PROPERTIES OF COMMUNICATOR AND ACCOUNT

5.5.3.7 *Attributes of Communicator*

Table 4 and Table 5 below list the set of attributes together with a specification of their type corresponding to the elements of Communicator and Account described above.

Communicator	***attribute***	***type***
	identity	text
	type	{person, organization, digital identity}
	authoritativeness	{authoritative, informal}
	reliability	{A,..,F}
	description	text
	name	text
	id	text
	picture	visual
	languages	{iso-639-1}
	location	{text, timezone, geo-coordinates}
	timezone	{iso-8601}
	geo-coordinates	latitude,longitude {iso-6709}
	creation	{iso-8601},UTC
	last-updated	{iso-8601},UTC
	accounts	{0..n} (Account)
Person	gender	{m,f}
	age	text
	voice-print	audio
Organization	legal-form	text
Digital Identity		

TABLE 4 ATTRIBUTES OF COMMUNICATOR

Account	***attribute***	***type***
	platform	text
	id	text
	description	text
	name	text
	screen-name	text
	gender	{m,f}
	age	text
	profile-image	visual
	creation	{iso-8601},UTC
	last-updated	{iso-8601},UTC
	location	text
	timezone	{iso-8601}
	geo-coordinates	latitude,longitude {iso-6709}

TABLE 5 ATTRIBUTES OF ACCOUNT

5.5.4 Document

Figure 38 below depicts the model for the element *Document* along with its attributes which have been identified by following the iterative procedure outlined above. As for the Communicator, it comprises the set of all attributes found to be of relevance across the spectrum of platforms and attributes pertaining to traditional media as well as to social media can be found.

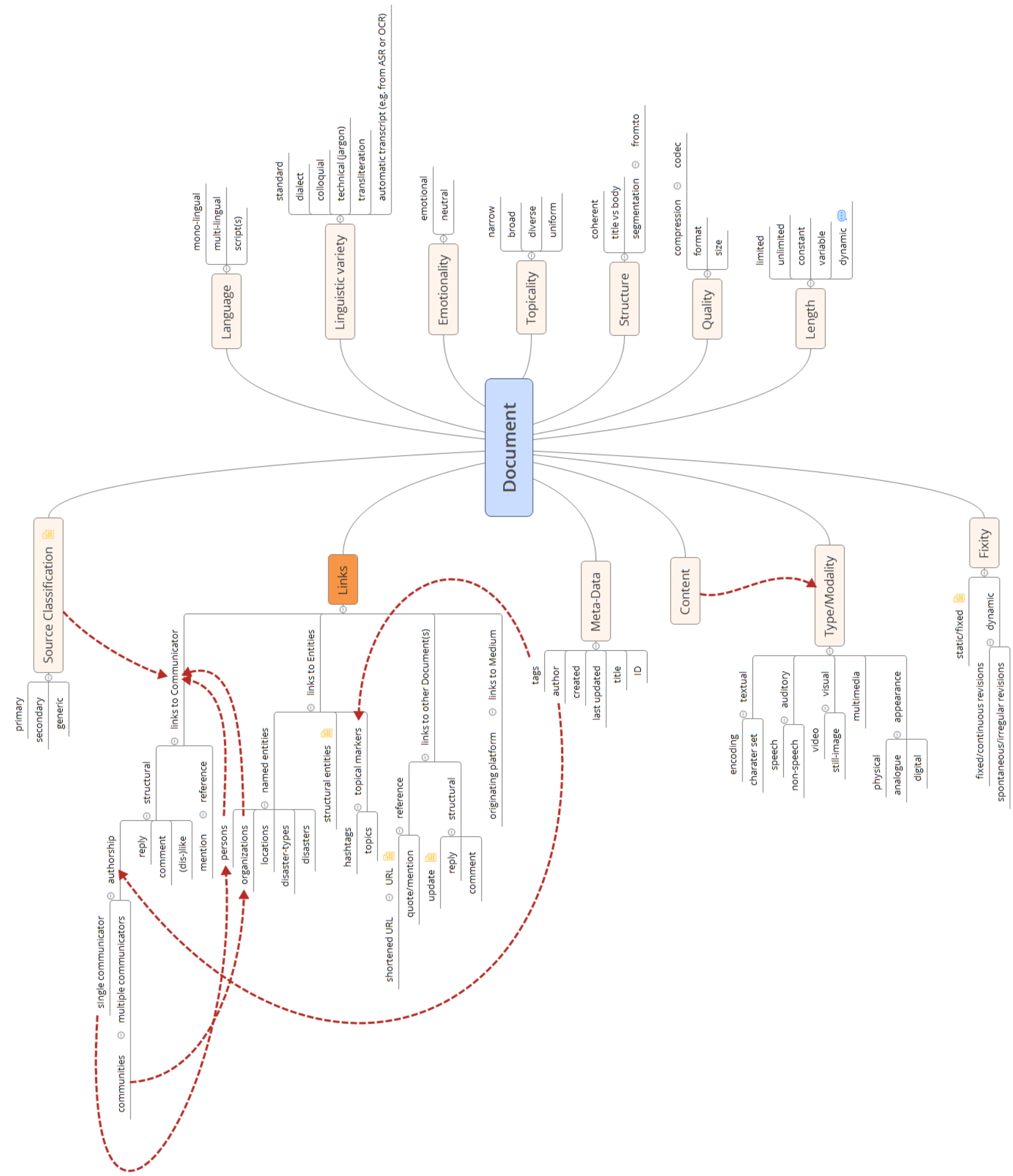


FIGURE 38 MODEL OF DOCUMENT

5.5.4.1 Type, Structure, Content

Documents can be distinguished according to the type of content they contain: textual, auditory, visual (still-images and videos) or multimedia content, combining several different types within a single Document. For textual content, different languages, scripts and encodings may be used to represent content. These issues are important when processing and enriching textual content and may play a crucial role in cross-border scenarios and international disaster-settings where one language (e.g. English) associated with one character-set (Western Characters) might be used to report on events mentioning locations and organizations which are originally spelled with a different character set (e.g. Arabic) leading to transliterations which may be non-standard and ambiguous. In case of locations, this may lead to additional steps having to be taken to reconcile several different spellings of a location. For auditory Documents a broad distinction into Documents containing speech and Documents containing *non-speech* such as music or (background) noise must be undertaken in order to assign the appropriate processing schemes. Documents carrying visual content can be distinguished into Documents with still-images and Documents of type video. Videos may also provide audio-tracks which can extracted and processed like audio-Documents. For certain kinds of sources, Documents might be provided in a physical rather than digital appearance. In many parts of the world – including those regularly hit by disasters – printed media, such as newspapers, may provide important and oftentimes the only source of information. Physical media may be converted to digital form by scanning and OCR'ing them, thus making automatic processing and further enrichment possible.

The structure of a Document determines and describes whether a Document may be segmented with each segment representing a coherent and consistent part of the overall Document. Segmentation may comprise stretches of text, pieces of audio or regions of images and videos. Segments may have different levels of importance, e.g. the headline of a newspaper article versus the body of the same article. The assignment of attributes and associated processing steps (described by the process model) may then take place on the segment-level. Meta-Data may be provided with each kind of Document on the Document-level as well as on the level of the Segment. The nature of meta-data (mostly textual) can be distinguished as for the nature of the content of Document and processed accordingly.

5.5.4.2 Source Classification

This attribute denotes a measure for the directness of relation between Communicators and Documents as well as between Documents and events. It states whether a Document forms a primary, secondary or generic (unknown) source. Primary sources are defined as Documents created by authors who were in direct contact with events as they unfolded, such as eye-witness news or journalists reporting from the scene of a disaster (*reporters*). Both, the Document as well as the related Communicator are labelled as primary sources in such cases. The concept of secondary sources is used for Documents which were forwarded or referenced and possibly authored by another Communicator. In case the directness of relationship is unknown, a generic source can be assumed. The attribute is not to be interpreted as in historical or literature research but rather meant to provide a measure of relationship only. In many cases, the order of relationship may not be obvious, and it may not be possible to assign it in an exact manner.

5.5.4.3 Length, Fixity

Documents may be of a fixed length, determined at the time of their creation, or may be dynamic in length for Platforms and Media where updates and additions are supported. Updates may be made in regular intervals or spontaneously, such as an update of the content of a web-page as information becomes more precise or needs to be corrected. The length of Documents on a Platform may be limited (such as Twitter's famous 140-character limit (now 280 characters)) or in principle be unlimited in length or size. Documents may all be of the same size (constant) or differ (substantially) in size. These attributes influence the type of collection, processing and storage required.

5.5.4.4 Quality

Documents may be stored multiple times on Platforms offering different levels of quality, such as different formats and resolutions of videos on YouTube[114]. Certain visual operations, like the detection of objects, persons or logos may depend on a certain minimum quality being available. For audio Documents, automatic processing typically requires a minimum quality level and models are trained to work best with a particular setup. For the case of printed textual Documents, the quality and resolution may play a role in digitizing and OCR'ing the content. Differences in quality of content influences the performance and accuracy of processing, causing degradation or resulting in inapplicability of processing in extreme cases when quality requirements are not met.

5.5.4.5 Topicality

Documents on a Platform may exhibit a wide variety of topics, each one being associated with a host of technical jargon and expressions. Typical cases of such Documents can be found on platforms like

[114] https://www.quora.com/Does-YouTube-store-multiple-resolution-copies-of-a-video-or-transcode-on-the-fly on 2019/03/17

Twitter or Facebook. Particular blogs, public pages or TV or radio channels may focus on specific topics only and thus present much more homogeneous content in the Documents present. Documents found on such platforms are more uniform in content and will use similar language and terminology. The types of Entities and hashtags mentioned will be more homogeneous and narrower for such platforms, whereas they might be differing widely on the former type of platforms.

5.5.4.6 Language, Linguistic Variety, Emotionality

The kind of language found in Documents may range from formal, official and standard[115] language to colloquial and jargon, depending on the Platform and Communicators present on them. Official newscasts on TV programs, press-conferences by spokespersons of government agencies or NGOs may exhibit strongly different language than interviews on the street (especially when emotionally charged after or during a disaster) or Documents on social media platforms. Depending on socio-linguistic factors, dialects may be spoken or even written (especially on social media) and for languages like Arabic[116] indeed be the only variety of a language used by communities. For several languages (e.g. Arabic or Greek) a diglossia results where the spoken and the written language of people differ considerably. For Arabic dialects, no one universally accepted written form exists resulting in a myriad of different spellings and the phenomenon of transliteration which is neither standardized nor consistent. Such situations impose particular challenges to automatic processing methods. Whereas standard typed text can be assumed to be largely free of mistakes, text on social media platforms is commonly full of typos. Text output by automatic processing such as OCR or automatic speech recognition (ASR) will include different kinds and qualities of errors, e.g. with ASR output the spellings of individual words will be correct whereas the words themselves might be misrecognized and phrase- and sentence structure might be incomplete and incorrect. Such situations require special consideration and robustness in further processing steps.

5.5.4.7 Links

Documents can be involved in a series of links, which are depicted in Figure 39 below. These are described in detail below.

[115] Not all languages have a single *standard* variety. For some languages, multiple standard-varieties exist (*pluricentricity*) whereas for others no standard variety may exist. In addition to purely linguistic aspects, historical and political motives may dominate the particular situation and official position.

[116] Modern Standard Arabic (MSA), is considered a standard and literary variety of Arabic but not spoken in people's everyday lives where typically local dialects are used. This is a form of diglossia which needs to be taken into account when processing data from speakers of 'Arabic'. Arabic dialects are usually not written but only spoken.

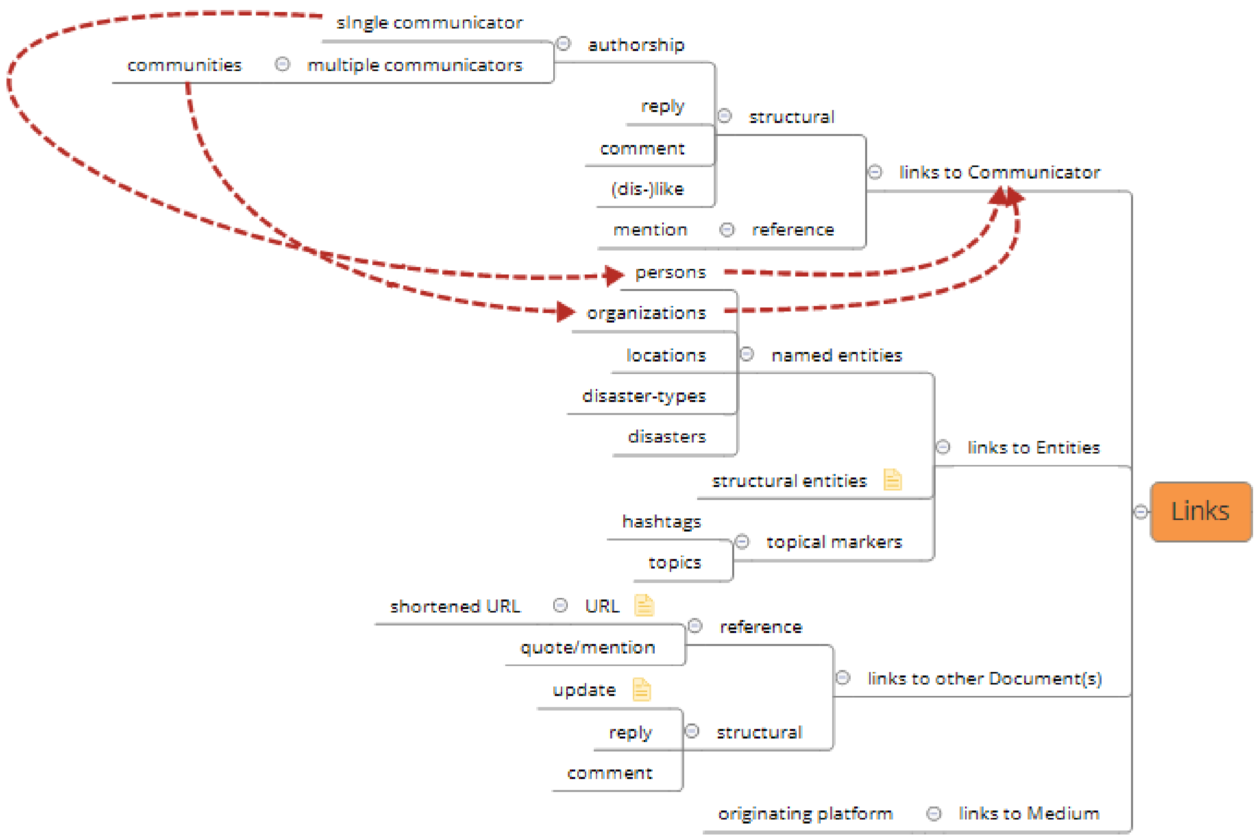


FIGURE 39 LINKS INVOLVING DOCUMENTS

5.5.4.7.1 Links to Communicator (D -> C)

Links from Documents to Communicators can be based on authorship or reference. Authorship is permanent and may refer to a single Communicator or to a group/community of Communicators. These may correspond to persons or organizations which in turn can be Communicators. Source classification of Documents and Communicators are related, a Document being a primary source also making the authoring Communicator a primary source for this Document. Links to Communicators can be explicit, mentioning the Communicator in the content of the Document, e.g. when an entity corresponding to a Communicator is mentioned within a textual Document. Links can also be implicit, e.g. when replying to or commenting on another Communicator's Document without explicitly mentioning them.

Figure 40 below provides an example of a Document (D1) being linked to a Communicator (C2) by citing (quoting, reference) this Communicator. The author or D1 (C1) is linked indirectly to C2 by this reference whereas both Communicators and the Document are linked directly. The explicit mention in the citation of C2 in D1 makes this link an explicit one.

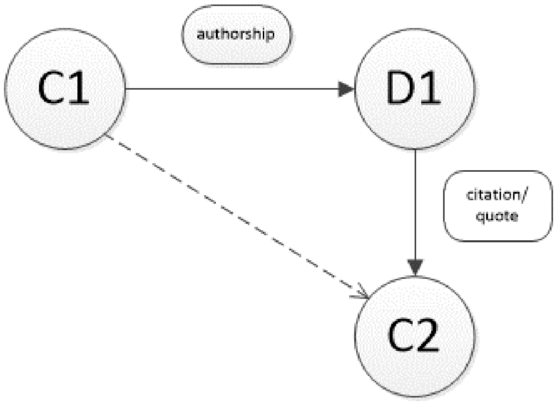


FIGURE 40 COMMUNICATORS LINKED THROUGH CITATION

5.5.4.7.2 Links to Entities which are not Communicators (D -> E)

Entities not corresponding to Communicators, such as the names of locations affected by a disaster, the name and type of a disaster, (names of) diseases spreading after a disaster, or hashtags, which may be used by populations in a dynamic way to tag Documents with informal mark-up, may likewise be linked-to from within a Document. Such links provide explicit and direct topical-references and may be used to cluster related Documents. Such clusters might be beneficial for summarization or the localization of a disaster and allow processing and filtering according to a wide variety of first-responders requirements. Entities are typically detected by technologies such as *Named Entity Recognition* (NER). Linking different entities to an underlying single unique concept is commonly referred-to as *Named Entity Linking*. Both technologies will be covered in the technologies section below (process model).

Links to entities are explicit and direct. They may occur within or across Platforms, e.g. the same hashtag being used on different social media platforms, thereby providing a link indicating the relatedness of content across Platforms.

Figure 41 provides an example of a Document (D1) mentioning a person (an Entity, E1) and containing a URL (leading to a referenced Document, D2). Communicator 1 (C1) as the author of the Document (D1) is also linked indirectly to the contained elements as well as to the further elements referred to by them. This results, for example, in indirect links between Communicators C1 and C2 and Document D2. Not all indirect links are shown in Figure 41. The creation and use of indirect links and whether they make sense for practical purposes (to what extent they can and should be expanded and/or made direct-links) is also determined to a large extent by the cognitive complexity and mental load imposed on the end-user (analyst) by such links and inferencing.

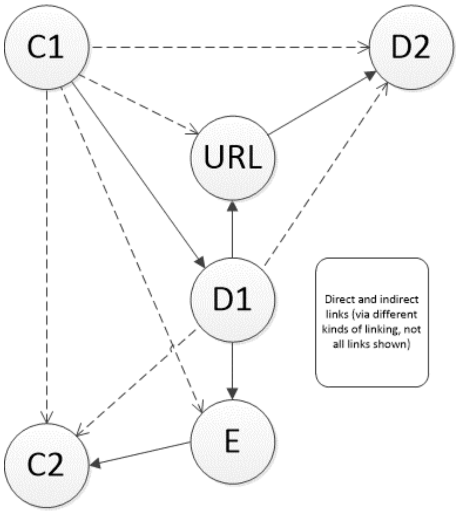


FIGURE 41 DIRECT AND INDIRECT LINKS

5.5.4.7.3 Links to Document (D -> D)

Documents may reference (other) Documents directly or indirectly. Direct references include URLs, which may occur in a shortened-version which need to be extended, or verbal references of another Document on audio-visual media (e.g. reference to an upcoming program covering a disaster). Links between Documents may also be of an indirect nature, e.g. if one Document is a comment or reply to another Document. Updates of Documents over time, e.g. web-pages which regularly get updated as a disaster evolves, and more information becomes available, replies or comments to other Documents are further possible direct links between Documents. Embedded URLs are often cases of cross-platform links with a textual Document (e.g. a tweet) residing on one platform and the referenced document (e.g. an image on Flickr) residing on another platform. Web-pages with URL-links to live-streams likewise form direct, cross-platform links.

Figure 42 provides an example of two Documents (D1 and D2) connected via a direct link due to commenting (reference). D2, authored by Communicator C2, is a comment to D1, authored by Communicator C1. The two Documents are thus linked directly with the involved Communicators and associated Accounts being linked indirectly to each other’s Document.

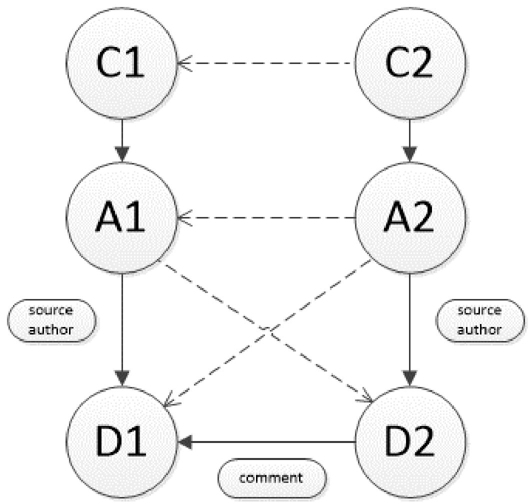


FIGURE 42 DOCUMENTS LINKED THROUGH COMMENTING

Figure 43 depicts an example of re-tweeting of a tweet. Communicator C2 retweets Communicator C1's tweet (D1) thus creating an updated version of Document1 (D1*). D1*'s content has been altered to contain an additional "RT" and updated meta-data and is thus not identical to Document D1. The two Documents are related via a reference (update) link, C1 is indirectly related to D1* and Accounts and Communicators are in turn linked indirectly by this interaction.

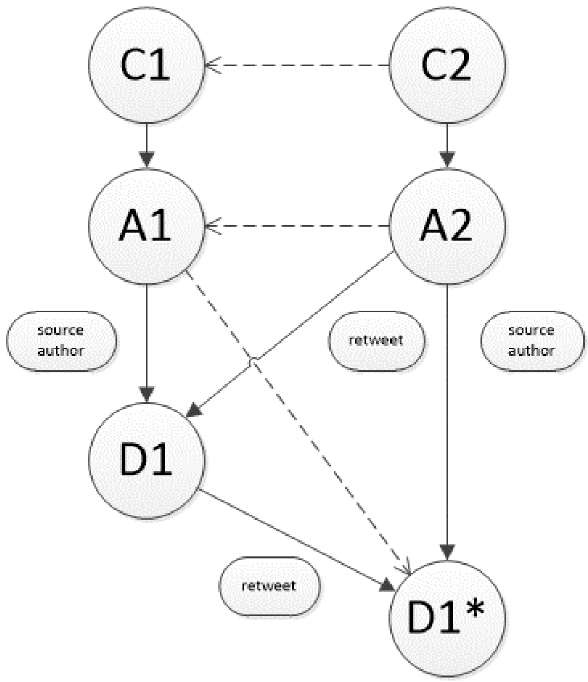


FIGURE 43 LINKING THROUGH RE-TWEETING

5.5.4.8 Conceptual Model for Document

Figure 44 below depicts a conceptual model of Document, Segment, and their sub-types. A Document may consist of one or several Segments, each possessing their own attributes. Segments may be of a textual, auditory or visual nature with further distinctions. Segments are created by initial processing, yielding homogeneous segments which can be passed to further, segment-specific processing. The respective attributes are discussed below.

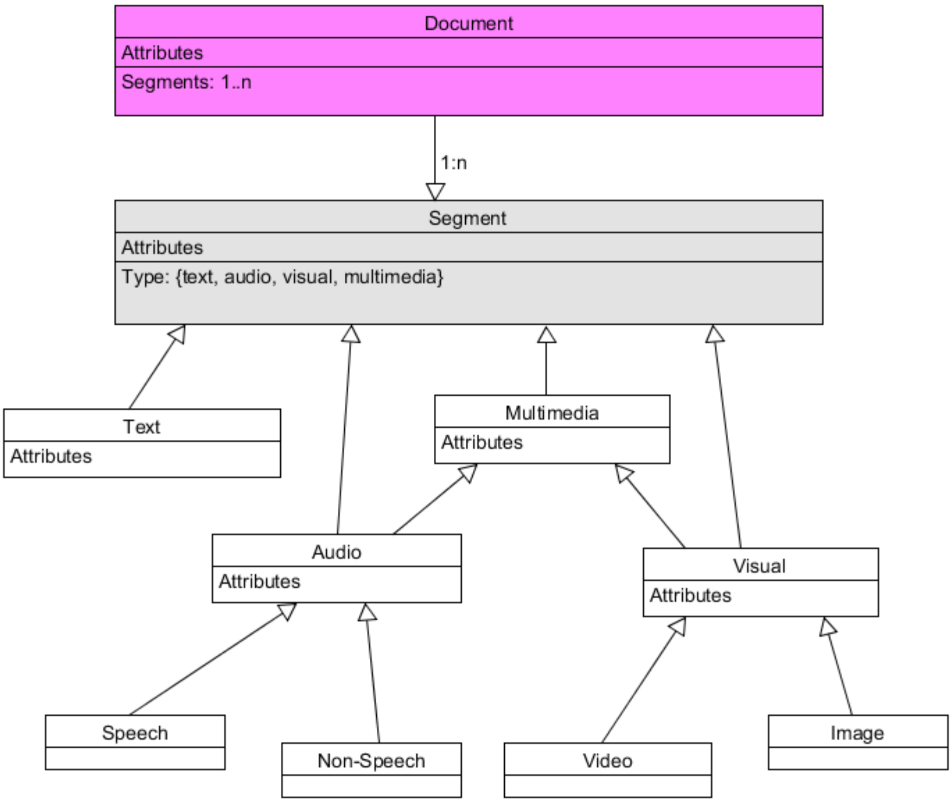


FIGURE 44 CONCEPTUAL MODEL OF DOCUMENT

5.5.4.9 Attributes of Document

Table 6 and Table 7 below list the set of attributes of Document together with a specification of their type corresponding to the elements of Document and Segment described above.

Document	*attribute*	*type*
	type	{text, audio, visual, multimedia}
	appearance	{digital, physical}
	fixity	{static, dynamic}
	length	{limited, unlimited, variable, constant, dynamic}
	quality	text
	language	{monolingual, multilingual}
	linguistic-variety	{standard, dialect, colloquial, technical, transliteration, automatic}
	topicality	{narrow, broad, uniform, diverse}
	emotionality	{neutral, emotional}
	segments	{1..n} (Segment)
	source	{primary, secondary, generic}

TABLE 6 ATTRIBUTES OF DOCUMENT

Segment	*attribute*	*type*
	type	{text, audio, visual, multimedia}
	duration	{from:to} {iso-8601}
	location	{from:to} (from-char:to-char)
	language	{iso-639-1}
Text	type	{standard, transliteration, automatic}
	character-se	text
	encoding	text (UTF-8)
Audio	format	text
	quality	text
	codec	text
Video	format	text
	quality	text
Image	format	text
	quality	text

TABLE 7 ATTRIBUTES OF SEGMENT

5.5.5 Entities

The term *Named Entity* was coined by the DARPA MUC evaluations (Grishman & Sundheim, 1996) where it comprised a set of designators for persons, locations and organizations, referred to ENAMEX, as well as numerical and temporal designators, referred to as NUMEX and TIMEX respectively. The former types include so-called proper-names, where several surface forms (*designators*) may reference the same underlying concept (*referent*). Nadeau and Serine (2007) provide a survey of the field of Named Entity recognition, its development over time and the different types of named entities developed and applied in various domains and scenarios. Entities can be found in Documents' content or in the meta-data associated with Documents or Communicators. They may be present in textual, auditory or visual form. In the context of the current thesis they extend more traditional versions of Entities insofar as they are to be viewed in a multimedia, multi-lingual and multi-script environment, allowing them to serve as links between Documents in different languages[117] and on different media.

For the present work, three basic different kinds of Entities as depicted in Figure 45 below are considered: Named Entities, Structural Entities and Topical Markers.

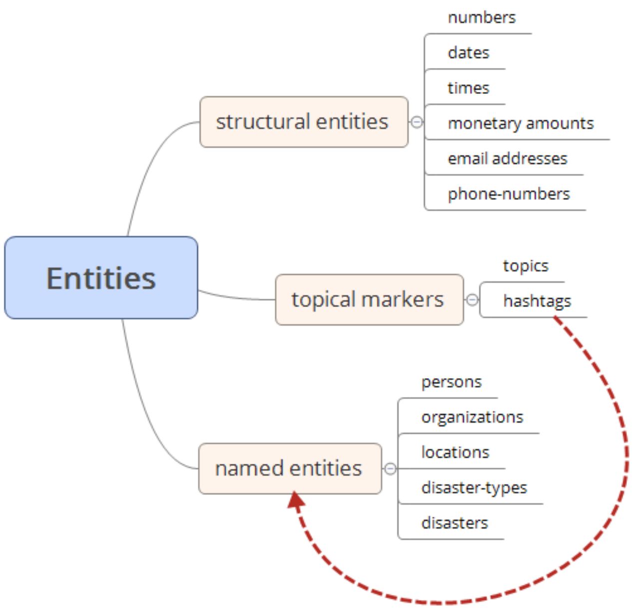


FIGURE 45 ENTITIES

Named Entities may refer to individuals or organizations – in which case they may refer to Communicators. They may also refer to locations, disaster-types (e.g. *Hurricane*), names of disasters (e.g. *Mangkhut*) and a set of further concepts which can be employed to classify and organize

[117] In addition to being cross-lingual, Entities may be viewed as nodes in an ontology (concepts) allowing for a diversity of connections between them through a multitude of possible links.

information pertaining to disasters. Like in MUC, they are based on the idea of associating multiple designators, possibly in different languages, to a single concept. They include the ENAMEX types from MUC but extend these with disaster-relevant types[118]. The addition of disasters and disaster-types provides a link to work performed in the domain of (new) event detection where entities provide fundamental components in determining the details of an event. Tanev et al. (2008) describe an approach and a system developed at the Joint Research Centre (JRC) of the European Union aiming at the monitoring of violent and disaster events.

Structural Entities are defined by their prototypical structure and are based on patterns or regular-expressions. They are a superset of the types used in MUC (NUMEX and TIMEX classes) and may refer to dates, times, currency amounts, account- or telephone-numbers or email-addresses. Their detection can be based on the structural specifics of such information and be performed via list- and pattern-matching (also taking into account morphological aspects depending on the language). Whereas it is assumed that the Named Entities are represented by a set of designators – in principle a closed set but with the possibility to adjust and extend it according to the particular domain or for a particular disaster - Structural Entities are an open, possibly infinite set of elements only defined by their common structure.

Topical Markers comprise topic-labels as produced by statistics-based topic-models (TDT (Allen et al., 1998; Wayne, 1998), LDA (Blei et al., 2003)) or, more recently by word-embedding models (Liu et al., 2015) and also comprise hashtags - ***#hashtag*** or in Chinese characters also ***#hashtag#***, ***#北京#*** (Murzintcev & Cheng, 2017). Topical markers may originate from a pre-defined, fixed set of elements (as the Named Entities) or originate from unsupervised, clustering processes whereby labels are selected or created automatically from documents (Popescul & Ungar, 2000). Hashtags are user-generated, folksomic, Topical Markers which originated on Twitter but are now popular on a variety of platforms[119]. Rogers (2017) even ventures that as hashtags "*are often embedded social issues (and events), the hashtags themselves could be considered the content*". Hashtags may correspond to Named Entities for cases where a designator of a Named Entity is used as the hashtag itself (e.g. #DonaldTrump for the Named Entity of type Person "Donald Trump" where the correspondence between hashtag and associated Named Entity may not be 100% as in the example).

[118] Using the classes and concepts from the disaster-ontology developed within QuOIMA.
[119] For an overview of the history of hashtags, their use, strengths and weaknesses refer to Rogers (2017).

Several *surface forms/designators* may refer to one *underlying Entity or Concept*[120]. However, designators may in certain cases not allow to uniquely identify the corresponding Named Entity directly, requiring further context and disambiguation mechanisms (Chang et al., 2016). Furthermore, entities may not be mentioned explicitly within a document but rather only be referred to by reference (e.g. by pronouns), requiring mechanisms for co-reference resolution (Ng, 2017). Disambiguation and co-reference resolution are active fields in the NLP domain. They are not dealt with further in this thesis as their application is not essential for the model developed within this thesis. However, their application is certainly beneficial and will result in better quality of information when the model is applied.

5.5.6 Medium

Figure 46 below depicts the model for *Medium* along with attributes which have been identified by the iterative procedure outlined above. As in the previous chapters, it comprises the set of all attributes found to be of relevance across the complete spectrum of platforms from traditional media as well as from social media. As outlined above, the *Medium* is viewed as the *means of delivering and receiving of Documents*, and the concepts of *Channel* and *Platform* may be viewed as specific sets of attributes and their associated values within the model of Medium. *Medium* thus represents the superset of attributes and possible values of which any particular channel/platform only uses a subset - a particular channel/platform thus forms an *instance* of the more general concept of *Medium*. Documents and Communicators are related to Medium by their respective attributes and by which Documents and Communicators exist on a Medium. Beyond the attributes shared between Medium and these elements, the Medium's attributes influence processing strategies and technologies which can be applied as reflected in the process model.

[120] The designators *US, United States* and *США* all refer to the same underlying referent/concept of *USA,* e.g. represented by a unique URI allowing to associate several surface forms with the single concept.

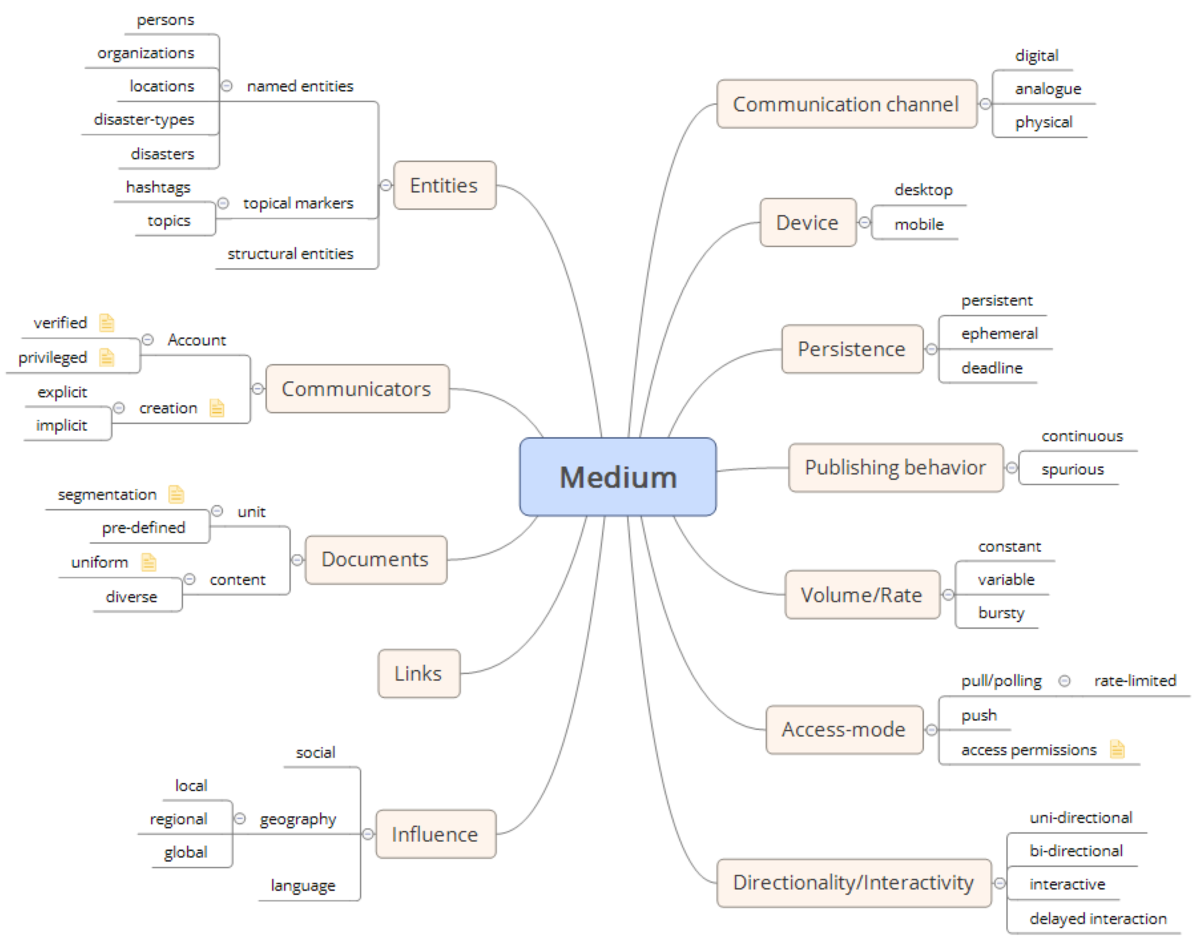


FIGURE 46 MODEL OF MEDIUM

5.5.6.1 Type

The type of a Medium is not described by any single attribute. Rather, a list of attributes and their associated values characterize a Medium. This approach was chosen as it allows for a maximum of flexibility in the classification of Media as well as in the assignment of technologies which may be applied to the Documents, Communicators and Links present on a specific Medium[121]. The type of Documents on a Medium allows to correlate processing technologies regardless of the other attributes of the Medium itself. E.g. for Documents with Type=textual, text-based language identification may be applied without any further knowledge of the Medium, whether it is a microblogging service such as Twitter, a page of an RSS-feed or an automatically generated transcript of an interview on TV. The common denominator is the Document-Type which is sufficient to determine the requirements (text requires language identification, if not provided via meta-data of

[121] Following the *facet-oriented approach* described above in chapter 5.2.1

the Document) and possibilities – if Document-Type=text then certain operations can be carried out and insights gained.

In addition to the attributes of the Communicators and Documents present on a Medium, the following attributes characterize a Medium:

5.5.6.2 Communication Channel

Media channels can be of a digital nature (social media, digital TV, Web-pages), of an analogue nature (FM radio) or physical nature (printed newspapers). The Communication Channel determines the means to ingest/collect the Documents on the Medium. Physical Media may first have to be digitized, e.g. via scanning and OCR of newspaper pages. Analogue Media, such as FM-radio may have to be digitized before further processing, e.g. subsequent automatic speech recognition (ASR).

5.5.6.3 Directionality/Interactivity

Whereas traditional media tend to be uni-directional, Media of the Web2.0 or social media age typically allow for interactivity and feedback. This feedback may be possible in real-time (e.g. users interacting live with the author of a movie on Periscope) or in a delayed manner (e.g. commenting on a post sometime after the post has been published). Directionality determines the kinds of interactions (links) possible between Communicators and Documents and influences strategies to collect and update these interactions.

5.5.6.4 Access Mode

The methods to retrieve Documents from Media may be of a polling/pull nature, in which a platform is queried through its API (e.g. REST APIs) for (new) Documents. Keeping track of which Documents have already been processed, for instance with an *iterator*, is required in such cases to avoid the introduction of duplicates. Push interfaces allow to retrieve new Documents via registration and subsequent reception (e.g. via a streaming API such as Twitter's *real-time streaming API*[122]). In the former case, a scheduling methodology is required to retrieve new Documents at appropriate moments in time. Both cases need to account for completeness/recall issues, i.e. they have to ensure that all relevant Documents are indeed obtained. Certain rate-limitations, e.g. regarding the maximum allowed number of calls to an API within a specified period of time, or certain conditions limiting the

[122] https://developer.twitter.com/en/docs.html

complexity of the query or amount of data retrieved may apply. In the push-case, the platform itself decides when and how to transmit (further) Documents. Duplicates are usually not present in such scenarios but the reception, queuing and processing all have to be adequately designed to progress with the platform's pace. Technical or platform-induced limitations or restrictions may require a distributed setup to access information and call for coordination activities to accommodate the resulting different input streams. Pull-interfaces are typically more common than push-interfaces with social media APIs (Castillo, 2016).

5.5.6.5 Volume/Rate

The volume (or rate) of Documents produced by a Medium may be constant, such as for TV stations producing 24 hours of newscasts per day or it may be variable, such as for Web-pages, microblogging, video- or photo-sharing platforms. In case of large-scale events (including disasters), certain platforms may exhibit a highly *bursty* behavior of the publishing of Documents. Ingestion and processing mechanisms need to reflect this and have to be designed appropriately to deal with such behavior for the respective Media. Volumes of data, especially on social media platforms change rapidly with dramatic increases in volume and speed over short periods of time. Selective processing strategies and various types of limitations require focused processing of selected Documents or topically determined sub-sets of Documents. Automatic load-balancing and parallel processing may be required to accommodate such dynamic setups[123].

5.5.6.6 Publishing Behavior

New Documents may be published in a continuous manner, such as a radio-program transmitting news 24 hours per day. Microblogging platforms with constant high volume likewise may be considered to be producing Documents continuously (yielding a quasi-real-time stream of Documents). On the other hand, web-publishing of news articles may take place only at certain times of the day, resulting in a few or even only one (possibly very large) batch of new Documents per day. Other platforms may proceed in a more continuous and constant manner with short intervals between the publishing of (possibly) small sets of Documents. Bursts of Documents may ensue from large-scale events, predominantly on platforms allowing for user-generated content (UGC). Depending on search criteria and access mode, Documents may only be provided by a platform at infrequent and irregular intervals. In certain cases, this may trigger the question whether a specific platform is still operational at all (in case of disasters, this question may be of high importance concerning local sources).

[123] Containerization of individual services, e.g. with Docker and orchestration and scaling with Kubernetes or Docker-Swarm provide state-of-the-art approaches for such environments.

5.5.6.7 Persistence

Documents may be stored and be available for extended periods of time on certain platforms, whereas they may exist for a short period of time only on other platforms before being removed[124]. The latter type of persistence is frequently referred to as *ephemeral*. In case Documents remain available on a platform, they may be removed after a given period of time, for example content on a Web-page of newspaper might be removed due to legal and copyright restrictions or because other Documents require the web-space. Certain Documents may persist for longer periods of time, such as Documents in archives, and may be regarded as persistent. The type of persistence dictates whether Documents can be linked-to or need to be downloaded and stored for future analysis or to justify and document actions taken during a disaster (if a crucial Document is no longer available, an action taken due to the Document's content may become a complicated if not impossible task).

5.5.6.8 Influence

In the fields of advertising and media analysis, Media are often associated with measures of influence such as *reach, impressions* or *impact*. These aim to reflect the potential number of recipients typically reached by a platform. Whereas reach can be defined as a continuous figure, in cases of disasters it may be viewed predominantly in terms of local – e.g. a local newspaper – or wider reach, such as nationwide or global-scale. The latter type typically only applies to cases of cross-border or international incidents and disasters with an impact large enough to be covered also by the international press (see chapter 3). Local reach may be interpreted as corresponding to local sources and as such may provide valuable information from the scene of a disaster. Global reach on the other hand may translate to international aid and donations being generated. The influence of a medium or platform may be combined with the influence of Communicators (see chapter 5.5.3.3 above) and Documents on these platforms.

5.5.6.9 Device

Social media platforms tend to be accessed from mobile devices (but also from desktop-devices). The layout and scope of information may differ according to the type of access device. Likewise, stability and robustness of access may be influenced by the style of device with mobile-device interfaces typically displaying simpler layouts. The kinds of devices used may also allow to gain information about

---

[124] Removal may also be due to ethical or legal reasons, e.g. hate-speech or *fake-news.*

the location of a Communicator. Using a mobile-device paired with geo-location information may provide hints of whether a user may be an eyewitness and a primary source for an incident.

5.5.6.10 Entities

Entities – in all of the forms represented in the model – provide a fundamental way to tag Documents and Communicators and connect them *on the surface* as well as semantically. Several platforms meanwhile support the use of hashtags to allow Communicators to tag Documents topically. Whereas this phenomenon originated on social media platforms like Twitter, Facebook, or Instagram, hashtags are meanwhile also used on traditional media such as TV via inserts or in audio Documents by pronouncing the hashtag's name (or spelling it). Hashtags thus provide an important cross-Document and cross-media way of linking topically related information (Documents). Named Entities may occur on any platform and be detected in textual, visual or auditory content. Like hashtags, these entities also provide cross-Document and cross-Platform links for related Documents.

5.5.6.11 Attributes of Documents

Documents on a platform may be uniform or diverse in nature or content. All attributes of Document play a role for collection and processing, such as the language or register used, the structure or quality of Documents, whether textual, auditory, visual or multimedia Documents are supported by the respective platform. Only a subset of the possible attributes of Document may apply to any given platform.

5.5.6.12 Properties of Communicators

As for Document above, Communicators may be homogeneous or exhibit a certain variety on a platform. The various attributes of Communicator may be employed to characterize a Medium. As for Document, not all possible attributes of Communicator may apply to any given platform.

5.5.6.13 Links

Media allow for different kinds of links depending on the nature of the Documents and Communicators present on them as well as on the meta-data associated with them. Information about Links may be present in a structural or referential form in the Documents and structures provided by a platform. Links may exist between the elements within an individual Medium or across Media (Cross-

Media Links). The nature of persistence of elements on a platform limits the kinds and duration of the validity of possible links. In case a video is removed after a certain amount of time, all links to the video will be void (unless the video has been downloaded and stored separately from the original platform, which again depends on the platform's technical and legal framework). In case of newspaper articles, papers are sometimes obliged to remove articles from news agencies after a certain period of time. The directionality of platforms has a strong influence on the kinds of links where interactivity plays an important role. The different kinds of Documents supported likewise influence the resulting kinds of links with different kinds of media allowing for more diversity of linking Documents of different media types.

5.5.6.14 Conceptual Model of Medium

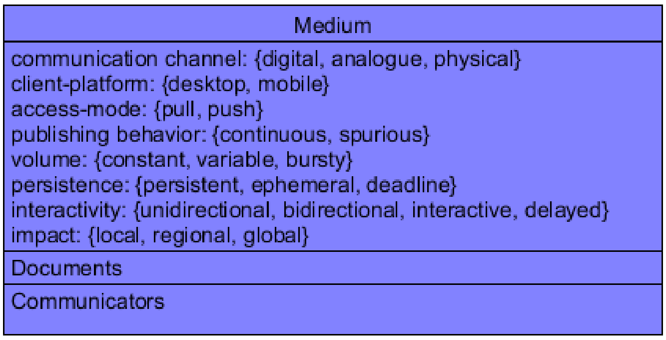


FIGURE 47 CONCEPTUAL MODEL OF MEDIUM

5.5.6.15 Attributes of Medium

Medium	***attribute***	***type***
	communication-channel	{digital, analogue, physical}
	device	{desktop, mobile}
	access-mode	{pull, push}
	publishing-behavior	{continuous, spurious}
	volume	{constant, variable, bursty}
	persistence	{persistent, ephemeral, deadline}
	interactivity	{unidirectional, bidirectional, interactive, delayed}
	impact	{local, regional, global}
	entities	{hashtags, named-entities}
	communicators	{standard, verified, privileged}
	documents	{uniform, diverse}

TABLE 8 ATTRIBUTES OF MEDIUM

5.5.7 Links

Links have been covered from the perspective of the two elements Document and Communicator in the chapters above. These elements correspond to the *Who* and *What* in Lasswell's model of communication whereas the Medium corresponds to the notion of *Channel*. Links, as previously stated, go beyond this original model and are not represented by it. The diversity and richness of such links, the possibility of linking elements actively in multiple ways and the resulting indirect links, which are not directly detectable *on the surface* form a rich source of information which can be used effectively when processing communication during disasters. The connected structure created by the elements and links forms the basis to infer further information which may be crucial and otherwise not be detected. In the model developed within this thesis, they form the central mechanism of connection between the individual elements. In this chapter the Links themselves and their attributes will be discussed.

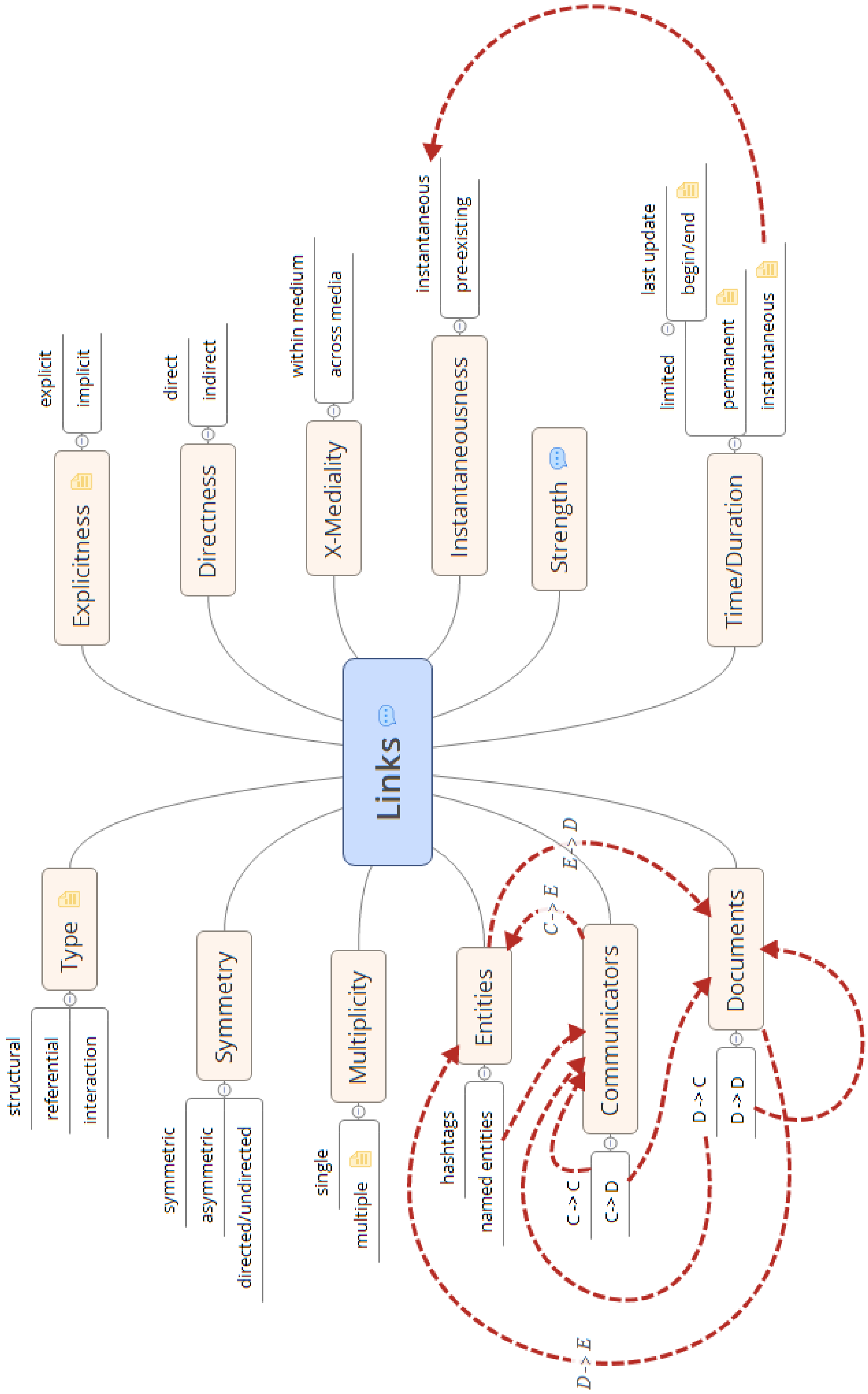


FIGURE 48 MODEL OF LINKS

5.5.7.1 Type

Links can be of a structural, referential (referencing) or interactive nature. Structural links encompass friend- (both on a virtual level as well as real-world level) or follower relationships as well as authorship as a special kind of structural relationship of high importance in determining the immediacy, source, relevance or credibility of information. Interactive links encompass the mentioning of Entities as well as various kinds of mentioning Communicators. Interactive Links encompass various kinds of interactions for feedback, commenting, sharing or liking of Documents - for platforms allowing for interactivity within the platform, for platforms not allowing such interactivity in a cross-platform manner, e.g. by tweeting about a (concurrent) discussion on TV.

One kind of Link – the *non-existing Link* – is not part of this classification but may, for practical and research purposes prove to be a crucial and particularly valuable object of interest. Why are certain elements in the communication model <u>not</u> linked within a certain dataset, e.g. among the Communicators and Documents involved in communication during a disaster. And how could such gaps in the Links be detected and modelled? Might this be a sign of an anomaly? And could they possibly be hypothesized, predicted or completed in order to provide a more complete picture of developments and foresight of things to come?

5.5.7.2 Directness

Links between elements (Communicators, Documents, Entities) can be of a direct nature or indirect nature[125]. Direct Links are introduced by a single interaction between elements of the communication model. Multiple indirect links may result from such an interaction by following direct links and establishing further links between nodes. Links may also be eliminated by another single interaction (e.g. unfollowing a former friend) or be marked as having been removed. Structural links which tend to be of a permanent nature are also direct links and indeed result from a singular action like establishing a follower/following relationship.

Direct links in the form of Communicator -> Communicator can be of a structural nature or be established via reference of Communicators. Structural links include social links, e.g. two (or several) Communicators working for the same organization, where the organization itself may also be regarded as a Communicator such as for TV- or radio-networks, newspapers, media-companies, first-responders

[125] *Direct* and *indirect* have different meanings from *directed* and *undirected*: the former denote the kinds of links established by processing whereas the latter denotes the kind of arcs which in the current work generally is assumed to be of a directed nature (see below).

or NGOs. Furthermore, they include links provided by platforms such as *follower/following* or *friendship* of accounts and as a consequence of their associated Communicators. Creation of Documents creates a direct C->D link between a Communicator and a Document, referencing of other Documents forms a direct D->D and an indirect C->D link.

By following connections (direct links) between nodes - bearing in mind graph-structure, symmetry and a reasonable distance between nodes (as otherwise relations cannot be grasped semantically or processed efficiently anymore) - further indirect links may be created temporally or added statically to the structure. Links established this way (by *inference*) become direct links in the graph.

Whether to do this and where to draw the line between pre-computation and dynamic inference at runtime is something which needs to be determined on a case-by-case basis. The complexity introduced by increasingly elaborate chains of connections for indirect links is a limiting factor for computation as well as for practical use by end-users. Cognitive overload and the possibility to formulate queries and interpret results are limiting factors, in particular when under pressure in times of disaster.

5.5.7.3 Explicit/Implicit

The term *explicit link* refers to a link which is visible on a Document's surface, e.g. the name of a person spelled-out in a Document of text-type, an anchor pronouncing the name of a disaster on the radio or an insert on TV displaying the name of a first-responder (an organization). Implicit links are established by being present in the meta-data associated with a Communicator or Document, such as the mention of the location where a user on Twitter resides in the account's meta-data or an author's name in the meta-data of a web-page. The distinction between explicit and implicit may not always be a clear-cut one. However, the emphasis is on the way that a links is conveyed in the content and whether to a regular user (e.g. an analyst of volunteer) the information is visible directly and explicitly. The distinction between explicit and implicit is relevant when it comes to deciding which data needs to be processed and whether the two kinds are available on a platform. A radio-program typically provides information of type audio and thus only explicit links will be present (for IP-radio where additional meta-data are available there will also be implicit links).

5.5.7.4 Symmetry and Multiplicity

Elements of the communication model can be connected with each other in multiple ways. The individual connections may be of a symmetric or asymmetric nature and be represented by directed or undirected links. In case of a symmetric relationship, this relationship can be modelled with one undirected or two directed links. Within this thesis, connections are generally regarded to be directed, allowing to associate different meanings to connections between elements depending on their direction. Thus, symmetric relationships are represented by two links of identical nature, whereas asymmetric relationships are typically represented by a single link. All links may account for the different type of link (structural, referential or via interaction). Furthermore, multiple links may exist between any two elements due to multiple relations or multiple interactions between them over time, i.e. due to several instantaneous actions (e.g. favorite or mention) or actions with limited time-spans (one of them ending before another one starts).

5.5.7.5 Within Medium / Across Media

Links between the elements of the communication model may occur within the same medium (within the same platform) or across different media. Whereas structural links tend to be present only within media - e.g. one Twitter user following another Twitter user but not an account on Instagram or a journalist publishing on the Web-Site of the New York Times - referential Links and Links stemming from interaction occur both within one platform as well as across different platforms. Referential linking between Documents via embedded URLs or based on shared hashtags (referred to as *co-linked and cross-hashtagged content* by Rogers, 2017) allow to cross the borders between different platforms and thus put communicative patterns on a broader basis, drawing upon the different affordances provided by the individual platforms and reaching different communities present on them. From an information gathering standpoint, this cross-platform activity and linking can be expected to provide a wider array of angles and a greater variety of content. Different platforms may be the *leading* platform of information dissemination during a particular disaster. Detecting and processing cross-media links provides a broader basis for data gathering activities as many different ramifications of information flow can be taken into account simultaneously due to the multi-platform approach (Backfried et al., 2016).

5.5.7.6 Strength

The strength of a Link between two elements may be used to assign a level of confidence in the Link, e.g. when detecting links automatically, this may be associated with a score of the detection mechanism. It may be regarded as a measure of confidence in the validity of the link.

5.5.7.7 Communicator / Document / Entity

The fundamental elements of the communication model: Communicators, Documents and Entities can be connected (or linked) in a variety of manners. The respective kinds of possibilities for such connections have been described in the chapters above. Each kind of element can be linked to the all other elements of the communication model as well as to themselves as depicted in Figure 48 above.

5.5.8 Property Graph Model for Cross-Media Communication

The above analysis regarding the central elements of the communication model – Communicator, Document, Entity and Medium/Platform – shows that each of the elements can be characterized by a set of attributes. A variety of manners to connect (link) these elements exist, describing various kinds of relationships between them. Furthermore, these diverse kinds of connections can likewise be described by a set of attributes. The resulting image is one of a network of nodes and edges, each with associated properties. All of the elements and their connections lend themselves to being modeled by a graph, a set of nodes and edges with associated attributes – as a Property Graph Model. Not only does this model allow for a direct semantic relationship between the elements and the model, it is indeed a semantic representation of the elements and their interactions. The focus on elements and relations mirrors the overall approach of this thesis of identifying key elements and their relations and modeling these on a generic level. The focus on interconnections and many-to-many relationships provides a strong argument for the choice of this data model (Kleppmann, 2017). Choosing a graph-based data model allows for flexible and direct extension in case of new platforms needing to be accommodated (as they appear and start playing a role in disaster communication). Commonalities can be modeled easily by common attributes, focusing on generic attributes and not requiring idiosyncratic elements when these can be omitted. Regarding the actual implementation and processing, choosing a graph-based approach means that a host of proven graph-DB technologies - both regarding storage as well as graph-DB processing methods and technologies - can readily be applied. The close affinity of logical and physical model allows to bridge the semantic gap between the conceptual model and the instantiation of that model (Robinson et al., 2015). Weaknesses in terms of integrity constraints require a careful approach to the modelling of attributes and connections (Šestak et al., 2016). These aspects are reflected in the design of the elements of the data model. Detecting temporal and indirect relationships, involving sequences of nodes and edges can be done efficiently with state-of-the-art graph-DB frameworks such as Apache TinkerPop[126]. Technologies to create and extend the graph, to detect nodes, connections and Links as well as their attributes, however, still need to be applied during the construction of the network. As outlined in chapter 5.3 above they rely largely on a set of NLP-technologies which are discussed with the process model below.

5.5.9 Summary

Chapter 5.2 dealt with the choice of a communication model and its central elements: *Communicator*, *Document*, *Entity* and *Medium*. The multitude and richness of connections between these elements – and the additional information which can be gained by explicit modelling of these links - motivated the addition of *Link* as a separate entity to the model. Based on these basic ingredients, chapter 5.5 describes a method to construct models for these elements in a two-step approach, taking into account all relevant platforms and media encountered during disaster communication. Elements are aggregated, their attributes are compiled by an iterative procedure and the possible links treated in separate sub-chapters. Characterization of elements via attributes provides a similarity to facet-based classification. Mind-maps and class diagrams were chosen to present a clear overview of the resulting models and their associated attributes. A view centered on attributes allows to align the elements of the communication model with the technologies required to process them (their content). The network created by elements with attributes connected by links (with attributes) suggests a graph representation leading to the choice of the property-graph-model as an adequate representation and manner to implement the resulting network.

5.6 Process Model

The **process model** aims to model the operations and activities which can be performed on the elements of the data model outlined above. Their characteristics are based on the attributes of the respective elements developed in chapter 5.5. The process models deals with the issues of how to construct these elements from the various kinds of media and platforms and how to process them, their content and meta-data. It comprises the steps from accessing data on different media, ingesting, processing and enriching it and finally creating a structure representing the elements and their

[126] http://tinkerpop.apache.org/

relationships. The result of this process is a graph consisting of enriched nodes and connections between them representing a variety of relations. This graph can then be employed for further inference, online analytical processing (OLAP[127]), querying or be used within a visual-analytics[128] framework. In addition, the *indexing* of content – for efficient and rapid access during information retrieval (IR) - as well as the *persistence* of information – for consistent and long-term access[129] – form part of the process model. The application of the resulting model to actual data and several real-world settings are discussed in chapter 6 below.

5.6.1 Method to Derive the Process Model

The process model development is carried out by the following steps:

- Construction of the overall workflow
    - Based on elements of the data model: Medium, Document and Communicator and their associated properties, design an initial generic workflow for the access and processing of elements
- For each platform/medium involved
    - Determine the characteristics of medium-specific access according to the attributes of Medium and design access accordingly
    - Identify the elements on the platform corresponding to Communicator and Document. Determine how the platform specific elements and properties correspond to the generic ones and how these are to be set (mapped)
    - Identify (structural) links provided by the platform and determine how these links can be accommodated in the generic model and how their attributes can be set.
    - Depending on their nature, identify technologies (types of processing) required to process Communicators and Documents (their content and meta-data), distinguishing between the different modalities (text, audio, visual, multimedia) where needed
- Categorize different technologies and identify dependencies between them and align technologies, platforms and the elements of the data model
- Process elements for graph construction and graph *population* (node and arc creation and attribute assignment)

[127] https://en.wikipedia.org/wiki/Online_analytical_processing
[128] https://en.wikipedia.org/wiki/Visual_analytics
[129] Persistence of data plays a crucial role as decisions taken during a disaster may have to be documented or even justified after the incident – the original documents may have disappeared or have been altered on the original platforms. Details of indexing and persistence are not dealt with in the current thesis.

- Enrich the initial graph-structure and to infer further insights (infer further connections and their attributes)
- Determine elements for indexing guided by issues of efficiency of retrieval and address issues of persistence

The above steps are based on the generic data model, in particular on the elements representing Medium/Platform, Communicator, Document and Entity. They deal with the issues of how these elements of the communication model can be retrieved from the various platforms (*Medium Specific Access*), how their content and meta-data can be processed, and respective attributes be assigned (*Medium Specific Processing*) and a how a graph-structure can be created and instantiated from this information. The elements of the data model and their attributes thereby guide and influence the processing model in terms of design choices and tasks. For the framework, a general streaming point of view is assumed for ingest and processing, although actual access and processing may be undertaken in incremental steps or via batch-processing depending on platforms and interfaces. The detection of links is addressed explicitly as they form a particularly rich base for insights and inference not possible by examining individual platforms or media only and thus go beyond the tools in use today (which are typically limited to a single platform only). Based on an initial graph, the subsequent step focuses on further processing and the creation of additional structure and links between elements of this graph.

The possibilities and limitations of the Medium (its *affordances* (Bucher & Helmond, 2017)) reflected in its attributes determine which processing is possible or required on the respective elements (this information can also be employed when determining processing requirements for a given set of Media/platforms for requirements analysis). Attributes of Document and Communicator further determine possible processing steps and sequences.

Within the process model, dependencies between technologies can be identified, determining the possible sequences of operations and influencing the overall design of processing flows. Furthermore, the potential for parallelization of activities as well as gaps and alternatives can be identified. Sequences and combinations of technologies can be grouped into larger units, allowing for a mapping to broader activities.

Figure 49 below provides a high-level description of the process model. Data is ingested from a specific platform using the access-methods provided by the platform (*Medium Specific Access*). In case of physical or analogue platforms, digitization is carried out as a prior step. Incoming data may need to be segmented and results in a stream of documents (*documents* here in the sense of the platform, not yet in the sense of the data model which are referred to as *Document*). The elements of the communication model of Communicator and Document are extracted by platform specific processing along with their respective meta-data (*Medium Specific Processing*). During this step, platform-specific attributes can already be assigned to a set of generic attributes. For example, proprietary attributes like *profile*, *peer* or *user* and their respective ID's are assigned (and their format is normalized) to the generic *id-field* for the Communicator. Processing of Documents, their content and of Communicators is then performed. Entities are extracted as part of this processing. The resulting elements provide the input for graph-construction, indexing for full-text searching (e.g. via Apache Lucene[130]) as well as for persistence (for storage and later retrieval, e.g. via MongoDB[131]). The latter issues are not addressed by this thesis, they do however form an integral part of the suggested architecture

[130] https://lucene.apache.org
[131] https://de.wikipedia.org/wiki/MongoDB

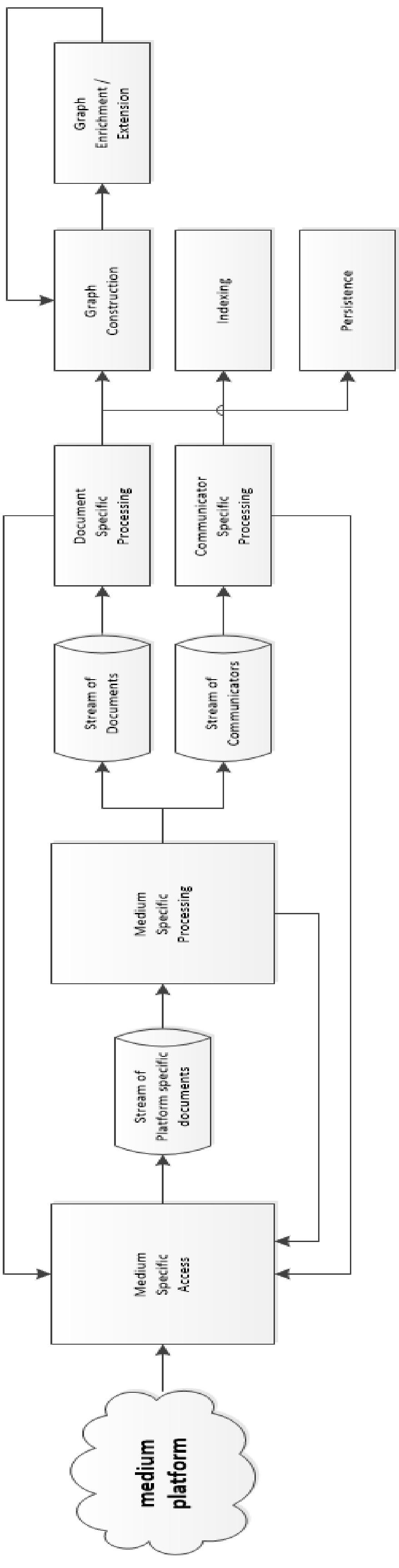


FIGURE 49 PROCESS MODEL – OVERVIEW

Several feedback loops can be identified: *Medium Specific Processing* may trigger further access within the same Medium or across Media, e.g. when one Document references another Document from a different platform which should be retrieved. The platform Twitter has been used in this respect to retrieve images from Flickr for triangulation of information during wildfires (Panteras et al., 2014), for location information from Foursquare and videos from YouTube for information diffusion (Jain et al., 2013). Document and Communicator specific processing may yield further links to important related information which may be fed back to the *Medium Specific Access* as well (e.g. after Document Specific Processing detected a particularly frequent, new hashtag in Facebook comments, querying of that hashtag on Twitter's streaming API may be triggered). During these phases, irrelevant or duplicate items may be detected and filtered out to increase the efficiency of downstream processing (Castillo, 2016). A further feedback loop exists between graph construction and extension where inferencing may be employed to extend the graph, adding new nodes and edges representing further insights. Such mechanisms may be used to infer and mark *indirect connections* which may only be detectable by network analysis of the graph structure.

The process model is guided by the attributes collected for the Medium describing the different aspects which determine access to and processing of data. Based on the presence and nature of Communicators, Documents, Entities and their attributes, further technologies and processing steps can be identified. The elements, their attributes and their role in this process are described below.

5.6.2 Medium/Platform

The characteristics and attributes of the Medium (platform) determine to a large extent the strategy to be employed for accessing and processing Documents and Communicators. They are specific to the platform and can be assumed to be constant only for a limited period of time. Changes due to technological advances, changes in business models, new legislation (e.g. the introduction of the GDPR[132]) or general trends, e.g. attitude changes of platform users towards privacy make platforms and (API-) access to their documents a dynamic and continuous challenge[133]. The access to the Medium itself determines which kinds of Documents and Communicators can be retrieved and which attributes will be available for them. These in turn determine the content, related Entities and the possible processing steps including the creation of elements and the detection and use of Links between them. The attributes of Medium relevant to the process model are further detailed below.

[132] https://en.wikipedia.org/wiki/General_Data_Protection_Regulation
[133] Frequently, changes are announced only on short notice or triggered by an incident or scandal and then need to be reacted upon quickly and previous functionalities are abandoned or restricted.

5.6.2.1 Communication Channel

Many Media employed for communication during disasters are of a digital nature. However, on a global scale various analogue media are still of importance in large parts of the world such as terrestrial AM and FM radio[134]. The content of such sources may first need to be digitized in order to be processable by audio-processing technologies. Printed newspapers and press-releases form another common source of analogue, physical documents. These documents may be scanned and OCR'd and thus turned into digital versions of the original document. The availability and quality of OCR however, still varies widely. Tools such as Tesseract[135] provide support for a wide variety of languages and scripts. Performance has been improved continuously, however, many scripts such as Arabic or various Indian scripts are still only supported to a limited extent.

5.6.2.2 Device

Social media platforms are frequently accessed from mobile devices, such as cell-phones or tablets, equipped with (high-quality) cameras and geo-positioning. Taking into account which *Device* a Document was created on may help to localize the Communicator and indirectly determine the availability of networks and infrastructure. If a Document stems from a mobile device, the possibility of the Communicator being an eyewitness, reporting from the scene is possible or even likely. This information can be triangulated with geo-positioning (exact positioning) or the detection of locations (approximate positioning) from the textual content of the Document itself, the Account or Communicator authoring the Document and be used to obtain further information. Communicators may operate more than one account and device (such as is allegedly the case with Donald Trump, using one iOS and one Android device) or several persons may be using the same account[136]. In such cases, there might be a relevant difference of the role the Communicator is acting in, using one device in an official (authoritative) role, whereas another device may be used only privately. This may also be reflected in time-patterns of device-use.

5.6.2.3 Access Mode

Platforms may offer pull, push or mixed types of access modes to obtain data or subsets of data (Castillo, 2016). The APIs themselves are commonly REST-based or proprietary APIs. Pull modes are

[134] During the 2014/2015 Ebola outbreak in West Africa, radio was a key source of information, https://www.bbc.co.uk/mediaaction/publications-and-resources/research/summaries/africa/sierra-leone/local-radio-stations-ebola

[135] https://en.wikipedia.org/wiki/Tesseract_(software)

[136] http://varianceexplained.org/r/trump-tweets

often regulated according to specific terms of service and platform rate-limitations, e.g. based on the number of calls per unit of time, on volume, the number of calls per application or IP-address or the number of documents returned. Push-access, such as Twitter's streaming-API continuously return a certain percentage of all items generating a quasi-continuous stream of documents. It is not clear, however, how exactly this sample is actually produced (Morstatter et al., 2013) prompting questions about bias and recall. For pull-scenarios, a batch-approach can be employed, fetching documents in batches from the platform at specified, regular intervals or depending on a model adapted to publication frequency. In combination with rate-limitations, sophisticated strategies may be employed for maximum access under the given constraints (e.g. accumulating and partitioning requests in order to get a maximum of data per unit of access or access via distributed nodes). From the perspective of the process model, a stream of Documents and Communicators is produced and provided to further processing steps.
In order to avoid delivering documents multiple times, some platforms allow to pass-in parameters stating where the last call has left off (a *cursor* or *iterator*, e.g. the ID or time-stamp of the last document retrieved). For platforms and methods not supporting such parameters (e.g. RSS-feeds) explicit de-duplication of Documents is required. The detection of near-duplicates may also be performed on the contents of Documents. Push-setups involving continuous streams of documents require special handling (buffering) for periods of processing or maintenance in order not to skip or miss-out on documents. If these are used on platforms with *bursty* publication behavior, bandwidth, sampling and buffering may become limiting factors.

Access to a particular SM platform may require a review-procedure, such as recently introduced by Facebook following the Cambridge Analytica Scandal[137], before being able to connect and use an API.

Audiovisual and audio content from TV and radio can be accessed and processed live (e.g. via SAT-receivers or cable), in which case the amount of data can be controlled and stays constant given a fixed number of channels to be monitored. In case of terrestrial reception, as in the case of local radio stations, analogue signals need to be digitized first for further processing. Processing of data is typically carried out in a 24x7 manner, allowing for the immediate ingestion of breaking news about emergencies as they happen. Alternatively, processing may take place in a pre-programmed manner, e.g. by only processing major newscasts, the evening news, or specialized programs (e.g. dealing with environmental issues). Alternative access to content of TV and radio-stations may be gained through

[137] https://en.wikipedia.org/wiki/Facebook%E2%80%93Cambridge_Analytica_data_scandal

specific libraries of multimedia content and livestreams (e.g. ORF's TVThek[138], ARD's Mediathek[139] or CNN's livestream[140]). The dominance of radio in certain parts of the world makes radio an invaluable and locally connected source in times of disaster. IP radio provides an Internet-based streaming alternative for the reception of radio content.

5.6.2.4 Publishing Behavior

The publishing behavior of a platform determines the rate at which new documents need to be fetched and processed. As near-real-time processing is a crucial factor for disaster management, the periods must be tailored to the cycle of the platforms, possibly with mechanisms adapting to the activity patterns of a platform. Such patterns may depend on the locality of platforms and authors and fluctuate dramatically in times of disasters. Even under regular operating conditions, it is not uncommon for RSS-feeds to publish a large number of new documents in a single instant whereas other feeds may publish new Documents almost continuously, spread out over the whole day. The rhythm of publishing and the resulting set of documents determine to what extent live (streaming) data analysis or incremental data analysis is possible or feasible. Publishers may employ certain tricks, like pre-dating of documents to influence search engine behavior. Such mechanisms need to be countered and corrected to arrive at a homogeneous set of true time-stamps. This is of utmost importance as especially during times of disaster obtaining a clear timeline of events and reports is essential.

5.6.2.5 Volume/Rate

The absolute number of documents published by a platform is one of the major factors determining the required processing power and storage. The rhythm of this volume, however, might be quite different from platform to platform. Whereas some platforms exhibit a fairly constant rate and volume of documents over the course of a day, allowing to tailor processing power to this constant rate, others exhibit a variable and – especially during disasters and crises – highly bursty nature. Such bursts may result in volumes of orders of magnitudes higher than on average and – as these situations are typically the crucial ones when processing and analysis of communication is needed most – require careful planning ahead and the allowance of contingencies to accommodate such extreme periods. Volume is also strongly influenced by the number of multimedia documents, their formats and resolutions. The general pattern of a strong increase in multimedia content paired with the trend towards high-

[138] http://tvthek.orf.at
[139] http://www.ardmediathek.de/tv/live
[140] http://www.livenewson.com/american/cnn-news-usa.html

definition (HD) and ultra-high-definition (UHD/4K) formats is expected to have a multiplier effect on internet-traffic (in particular on mobile networks) and consequently on processing and storage requirements (Cisco, 2017). Keeping the best possible quality for analysis or re-processing purposes, resulting in higher volumes of data, needs to be balanced with space constraints, sometimes resulting in a mixed approach regarding persistence of keeping high-quality content only for a limited period of time and lower-quality content for longer periods (see persistence below).

5.6.2.6 Persistence

Platforms may allow to access documents for an extended period of time (persistent), for only a limited amount of time (deadline) or for only an instant (ephemeral). This strategy may be due to copyright and legal reasons or for efficiency and determined by the available resources. As the content of documents may play a crucial role in the decision-making-process during disasters, it is imperative that the documents leading to such decisions be kept in the system for possible later analysis or transparency and justification of actions. Tendencies and a higher awareness towards increased privacy may have an impact on future persistence strategies of platforms. Legislation like the European GDPR[141] have an impact not only on persistence, but on all levels of access to data of persons via APIs. Storing local copies of documents also allows for the re-processing of content with improved or refined models but likewise needs to consider technical (storage space) and legal constraints.

5.6.2.7 Interactivity/Directionality

Platforms allowing for interactivity offer a variety of interaction patterns as laid out above (in chapter 5.5.7). Interactivity may be uni-directional, only allowing a subset of privileged Communicators to take action (e.g. post messages). When it is bi-directional, interaction may take place on an immediate level, such as seen on platforms like Periscope which allow users to provide feed-back to the creator of a video even while the video is being created, or it may be on a delayed level, such as providing comments to a post sometime after the post has been published. Whereas the mere presence of bi-directionality means that connections can and should be accounted for, the immediacy of such interaction determines the rhythm with which such interactions need to be scanned and (re-) processed, e.g. when, for how long and how often a certain document should be re-fetched for updates.

[141] https://www.eugdpr.org

5.6.2.8 Influence

Whereas the previous attributes were of a technical nature, the attribute *influence* of a Medium plays a role in determining the strength and radius of effect on an audience, including social, linguistic and geographic reach regarding an affected population during a disaster. Sources can be marked for their geographical influence (geographical impact, reach) in general categories such as local, regional or global. This measure may subsequently be used to allow filtering of Documents to gain insights about the locality or spread of information. In combination with attributes of the Communicator (e.g. *influence*, see chapter 5.5.3.3) this attribute may provide insights into what regions or groups of an affected population might have been reached and which regions form part of the communication – or have been cut off from communication (e.g. by devastation of infrastructure). Locality may be an important magnifier during disasters for local sentiment or be an indicator of the possible level of detail of information gathered. Linguistic factors (the language typically used on a Medium) and social factors (the audience typically present on a Medium) likewise provide important factors to estimate the influence a Medium may have on an affected population, even though social factors may be much harder to categorize.

5.6.2.9 Summary of Attributes of Medium Relevant to the Process Model

Table 9 below summarizes the attributes identified as being of relevance to the process model for the element Medium. These are of particular importance to the initial stages in the process model (Medium Specific Access and Medium Specific Processing in Figure 49 above) and influence the design and processing in these stages. From a requirements and design perspective, they provide a list of key issues to be considered and reflected in the design.

Medium	***attribute***	***type***
	communication-channel	{digital, analogue, physical}
	device	{desktop, mobile}
	access-mode	{pull, push}
	publishing-behavior	{continuous, spurious}
	volume/rate	{constant, variable, bursty}
	persistence	{persistent, ephemeral, deadline}
	directionality/interactivity	{unidirectional, bidirectional, interactive, delayed}
	geographical influence	{local, regional, global}
	entities	{hashtags, named-entities}
	communicators	{standard, verified, privileged}
	documents	{uniform, diverse}

TABLE 9 ATTRIBUTES OF MEDIUM

5.6.3 Communicator

The communication channel determines to a large extent the manner of creation of Communicators. They may be emitted explicitly from a platform via its API, be present in the meta-data of Documents or be implicitly determined by the particular platform (e.g. by knowledge of the platform's owner or operator). Via *Medium Specific Access* and *-Processing*, a stream of Communicators is produced. These may be linked to the stream of Documents from the same platform via information such as platform-specific IDs or references (see Document <-> Communicator links for possible settings).

Figure 50 below presents an overview for the processing of Communicators.

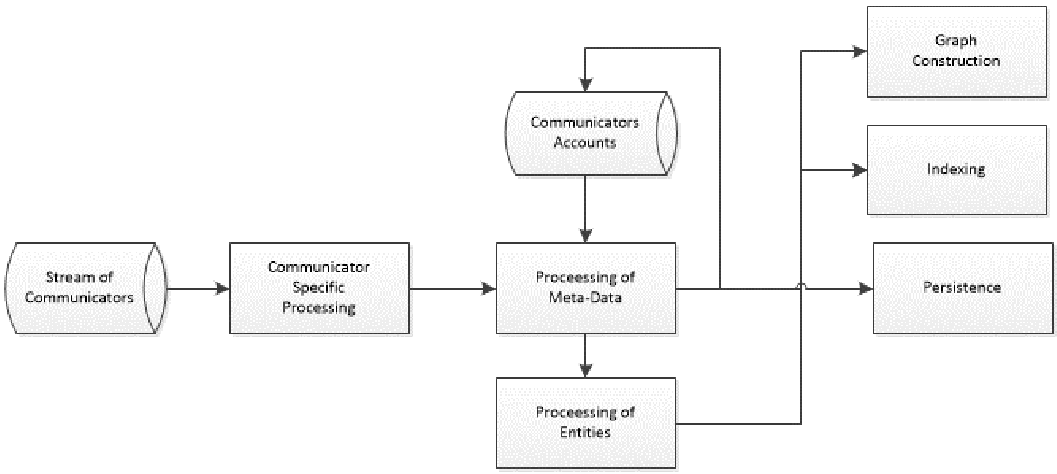


FIGURE 50 PROCESSING OF COMMUNICATORS

Processing of a Communicator's data focuses on the processing of the attached meta-data with the aim of associating further information with the attributes of an already existing Communicator or creating a new Communicator node in the graph. Processing may trigger further processing of other Communicators or Accounts (e.g. the followers of a Communicator) and may result in the detection and association with Entities as well as the creation of different kinds of links between Communicators and Documents. Incorrect information, e.g. a creation-date which is in the future, may be detected at this stage and corrected or marked. The detection of duplicates takes place at this stage as well in order to avoid unnecessary processing and inconsistencies. The result of this process is a unique node to be used for all connections with Documents or other Communicators. Attributes of the node may be enriched in the process.

Communicators may be known for traditional media platforms, such as anchors in TV programs, journalists in newspapers, or spokespersons of organizations or governments etc. in which case it is possible to associate them to existing nodes. For social media, only an account may be known or, as recently put into effect by Facebook's graph-API, only an *unknown user* may be obtainable[142]. In the latter case, a place-holder *unknown Communicator* can be assumed which might be resolved by later actions, e.g. by inferences from the graph or by human interaction, e.g. an analyst connecting this entry manually based on their personal knowledge of the situation and actors. In case where Account and Communicator can be related directly, e.g. in the case of a public Facebook-Page of an organization, this connection can be established immediately and remains constant.

The set of Communicators and associated accounts and attributes is kept as part of the graph and extended continuously. The creation and extension of Communicators may occur in a variety of manners:

- by manual interaction with a component designed for the explicit management of Communicators, e.g. via a graph-editing interface. This approach does not scale well but may be required due to the specificity of the application domain and/or the lack of existing datasets for particular domains, languages or geographies. The merging of existing Communicators found to be one and the same Communicator can also be carried out here.
- by automatic extraction and transfer from a knowledge-graph such as DBPedia[143]. This method lends itself to a kind pre-population of Communicators in the graph, depending only on external sources and pre-dating any addition of Documents from processing.
- by interaction with a particular platform (e.g. via its API) if sufficient information can be obtained from the platform's API.
- by the structure of the platform itself, e.g. when an account is directly associated with the owner, e.g. the account and owner of a public Facebook page which can be linked directly-.

Some of the above may require a mechanism for the fusion of information with pre-existing data, e.g. by semi-automatic matching of attributes to an existing set of Communicators. The possible use of knowledge-graphs and linked-open-data in such a construction process links the current work on a generic model to the wider landscape of the Semantic Web and the topic of fusion of ontologies (Allemang & Hendler, 2011). Information about organizations and individuals in a knowledge graph can serve as a source for the creation of Communicators (e.g. an entry for the *International Red Cross and Red Crescent Movement* on DBPedia[144] can be used to provide attributes for the Communicator

[142] Such a lack of more specific information may make certain types of analyses impossible. However, others such as ones concentrating on general behavior and trends may only be affected marginally.
[143] https://wiki.dbpedia.org
[144] http://dbpedia.org/page/Category:International_Red_Cross_and_Red_Crescent_Movement

model of this organization within the data- and process model developed here. Communicators may thus be linked to entries in (open) knowledge-graphs[145].

For social media, the unit of interaction is typically an Account. As discussed above, accounts may be shared by Communicators and Communicators may own/use multiple accounts. In both cases, the Account may provide further information on the attributes of the underlying Communicator as well as regarding connections to other Accounts and Communicators. Documents from social media platforms may contain information about Accounts or Communicators implicitly, e.g. by containing the IDs of other accounts (e.g. accounts who re-tweeted a tweet). Information may also be available explicitly by API calls allowing to retrieve multiple attributes of an Account. This may result in a combined strategy when retrieving information about a Communicator or Account: IDs may be retrieved from Documents and subsequently be used in different interactions to gain further information about the Account or Communicator.

5.6.3.1 Summary of Attributes of Communicator Relevant to the Process Model

Table 10 below summarizes the set of attributes which can be associated with a Communicator for the Medium Specific Processing and Communicator Specific Processing steps. It summarizes the kinds of processing to be applied and the context of the resulting information. Attributes can be assigned during processing of a Communicator; respective models are required for some processing, e.g. the identification of a Communicator’s voice or image by auditory or visual technologies.

[145] Changes in the underlying structures of organizations and individuals pose a problem and will have to be accounted for over time.

Communicator	*attribute*	*type of processing*
	identity	text, link to existing Communicator(s)
	description	text, link to existing Communicator(s)
	name	text, link to existing Communicator(s)
	id	text
	picture	visual, link to existing Communicator
	languages	text conversion to {iso-639-1}
	location	text conversion to geo-coordinates and/or entity representing location
	timezone	text conversion to iso-8601
	geo-coordinates	text conversion to latitude,longitude iso-6709
	creation	text conversion of timestamp to iso-8601,UTC
	last-updated	text conversion of timestamp to iso-8601,UTC
Person	gender	text, conversion to {m,f}
	age	text, conversion to number
	voice-print	audio, creation of or link to voice-model
Organization	legal form	text, conversion to existing entity

TABLE 10 ATTRIBUTES OF COMMUNICATOR AND TYPES OF PROCESSING

Table 11 below provides the respective attributes for the Account(s) linked to a Communicator. As for Communicator above, the types of processing and the context of the resulting information is provided.

Account	*attribute*	*type of processing*
	platform	text, link to existing Medium/platform
	id	text, link to existing Communicator(s)
	description	text, link to existing Communicator(s)
	name	text, link to existing Communicator(s)
	screen-name	text, link to existing Communicator(s)
	gender	text, conversion to {m,f}
	age	text, conversion to number
	profile-image	visual, link to existing Communicator
	creation	text conversion of timestamp to iso-8601,UTC
	last-updated	text conversion of timestamp to iso-8601,UTC
	location	text conversion to geo-coordinates and/or entity representing location
	timezone	text conversion to iso-8601
	geo-coordinates	text conversion to latitude,longitude iso-6709

TABLE 11 ATTRIBUTES OF ACCOUNT AND TYPES OF PROCESSING

Further attributes may be computed after the initial creation of a Communicator. These may depend on inference from the initial network but also on mechanisms requiring more context than is available at the stage of Communicator creation, e.g. the computation of bot-factor and subsequent assignment of an Account and its associated Communicator to the class of *Digital Identity* – an activity which requires broader information about activity patterns and communication behavior.

Text, audio and visual processing is described below and can be applied to all attributes of the meta-data of a Communicator marked as such as in the tables above. These tables can be used for direct alignment of which kind of processing should be applied to each of the attributes depending on their particular values.

5.6.4 Document

The properties of the Medium, its communication channel, publishing behavior and access-mode, influence the creation of Documents. In case of non-digital communication channels, some pre-processing of (external) data may be required before being able to ingest data and create Documents. For example, a printed newspaper may first need to be scanned and segmented before being able to process it as a Document. The publishing behavior may require initial segmentation of an incoming stream (in case of continuous publishing) in order to obtain useful units which can serve as Documents. External information such as an Electronic Program Guide (EPG) for TV or radio may be employed for segmentation. In case such information is not available, different kinds of technologies may be applied to the content itself to generate coherent and reasonable segments in an automatic manner. Segments are considered to be homogeneous sub-parts of the original Document; their processing is in analogy to the Document itself. Volume and the rate of publication of new information on part of the platform needs to be considered for timely and correct construction of Documents. Persistence of information likewise has to be accounted for as Documents might otherwise vanish before being accessed and processed.

Figure 51 below presents an overview for the processing of Documents.

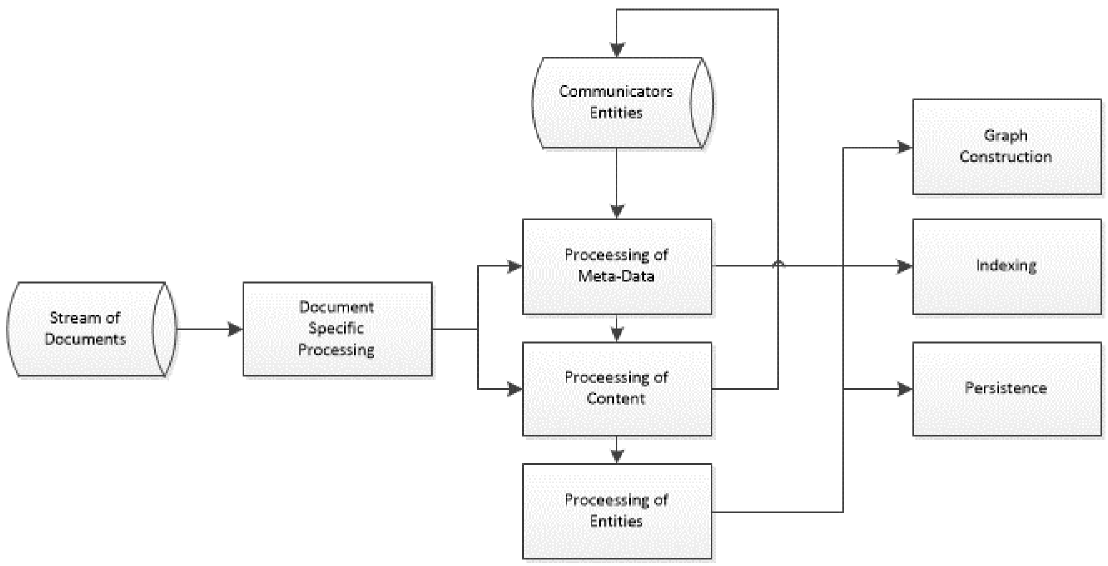


FIGURE 51 PROCESSING OF DOCUMENTS

The processing of a Document focuses on the processing of the Document's content as well as on the processing of any associated meta-data. Both kinds of processing may yield the detection of Entities and Communicators and result in various kinds of links.

5.6.4.1 Summary of Attributes of Document Relevant to the Process Model

Table 12 and

Table 13 below summarize the attributes identified as being of relevance to the process model for the elements Document and Segment. They are of particular importance to the enrichment steps in the process model (Medium Specific Processing and subsequent Document Specific Processing in Figure 51 above). For Documents consisting of several Segments, the respective attributes determine the processing of the individual Segments.

Document	***attribute***	***type of processing***
	type	content is {text, audio, visual, multimedia}
	appearance	determines whether digitization is required prior to further processing
	fixity	processing of different versions of the 'same document' required
	length	effect on processing resources
	quality	may have an effect on the selection of models to be applied
	language	application of models of different languages or of a single one
	linguistic-variety	may have an effect on the selection of models to be applied
	topicality	may require further segmentation, effect on the selection of models to be applied
	emotionality	may have an effect on the selection of models to be applied
	structure	determines strategy for segmentation (or processing of coherent content)
	segments	segmentation and (parallel) processing, final fusion

TABLE 12 ATTRIBUTES OF DOCUMENT

Segment	attribute	type of processing
	type	content is {text, audio, visual, multimedia} invoke appropriate processing
	language	determines models to be applied
Text	type	may have an effect on models to be applied
	character-set	effect on models depending on language
	encoding	all content converted to UTF-8 encoded Unicode
Audio/Image/Video	format	effect on models to be applied
	quality	effect on models to be applied

TABLE 13 ATTRIBUTES OF SEGMENT

After the initial creation of a Document, its content provides further information which needs to be processed in order to extend (and complete) this initial information.

Text, audio and visual processing is described below and can be applied to all attributes of the content and meta-data of a Document marked as such as in the tables above. These tables can be used for direct alignment of which kind of processing should be applied to each of the attributes depending on their particular values.

5.6.5 Relating Elements and Processing via Attributes

The attributes of Medium, Communicator and Document allow to relate these elements to the kinds of platforms employed during disaster communication and the kinds of processing which can be performed on them. Specific values of attributes of these elements – e.g. the type of content of a Document being *text* and the presence of such Documents on a specific platform allow to align the platform with a set of text-processing technologies. Any platform hosting this particular kind of Document can thus be related directly to a set of processing technologies (which in turn are required to be able to process the Documents of this platform).

Table 14 below provides an example, relating different kinds of Documents, Media (Platforms) and Technologies to process the particular type of content[146]. In this table, the attribute Type states which type of content a Document contains. Relating the types to the various kinds of platforms as well as relating them to particular kinds of processing allows to determine which processing steps are

[146] Markers in the table (*x*) indicate that a particular feature is relevant/present; markers in more general classes also apply to the more specific ones, e.g. if *radio* is marked it applies to *digital*, *analogue* as well as *IP-radio.* Elements of the communication model. Platforms and technologies can be related via these markers.

required in order to process data from a particular kind of platform and which factors need to be taken into consideration for processing.

For example, Documents with content in the form of (still-) images can be processed with visual processing technologies (such as OCR and Logo-Detection). This kind of Document is found on Photo-Sharing platforms. Hence, to process Documents from a newly emerging and relevant photo-sharing platform, these technologies and corresponding models need to be provided to process content from the platform. From the viewpoint of the platform, this information can be used as requirements when designing new systems and solutions (if a certain platform is to be supported, these technologies will be required to process its contents). For the end-user, the table can serve as guidance when designing systems and deciding on requirements and possibilities regarding the processing of data from a given set of platforms.

			Traditional Media																	Social Media									Technologies			
			TV				Radio				Press-Agencies	Print-Media					Web-Pages	Feeds (RSS, Atom)	Digital Distribution	Microblogs	Social Networks	Podcasts	Forums	Blogs	Photo-sharing	Video-sharing	Social bookmarking	Instant messaging	Auditory	Textual	Visual	Multimedia
				digital	analogue	IP		digital	analogue	IP	digital		Newspapers	Journals	Magazines	Books																
Document	***attribute***	***type***																														
	type	text											x							x	x		x	x	x	x	x	x		x		
		audio					x																				x	x	x			x
		visual-still-image											x							x	x		x		x			x			x	
		visual-video	x																	x	x	x				x		x			x	
		multimedia	x																			x						x	x	x	x	x
	appearance	digital	x				x													x	x	x	x	x	x	x	x	x				
		physical											x																			
	fixity	static	x				x						x							x	x	x	x	x	x	x	x	x				
		dynamic																		x			x	x								
	length	limited											x							x	x	x			x	x	x	x				
		unlimited	x				x														x		x	x	x	x	x					
		variable	x				x													x	x	x	x	x	x	x	x	x				
		constant										x	x	x	x					x												
		dynamic	x				x				x	x				x	x	x	x	x	x	x	x	x	x	x	x	x				
	quality	(text)																														
	language	monologual	x				x					x	x	x	x	x	x	x	x		x	x	x	x			x	x	x	x		x
		multilingual																		x	x				x	x	x	x	x	x		x

TABLE 14 DOCUMENTS, TECHNOLOGIES AND MEDIA RELATED VIA COMMON ATTRIBUTES

The set of attributes of all elements, platforms and technologies (as developed in the data model) allows to relate them to the different kinds of processing of traditional and social media platforms. The absence or presence of a certain attribute and its type (the kind of data it represents) can be used to quickly categorize a newly emerging platform and determine the kinds of processing required/possible.

5.6.6 Medium Specific Access

The different kinds and factors influencing this processing step have been discussed above in the chapter describing the properties of Medium within the data model. They comprise a set of proprietary and standard mechanisms to access data from a platform (the component *encapsulates* this access). In practice, this access is typically implemented via code communicating with the platform's specific API(s) and tailored to the particular platform. As such, it needs to be specifically implemented for every further medium to be supported. Changes in versioning, updates, error-handling or format changes (syntactic and semantic) require continuous attention. Rate-limits need to be taken into account, prompting potential parallelism and distribution of access mechanisms. Regardless of the actual platform, methods of access or types of documents, it is assumed that this stage provides a stream of documents (*documents* in the sense of the platform) in digital form with associated meta-data and ready to be processed by a component with specific information about the exact format and content.

5.6.7 Medium Specific Processing

Each of the involved platforms and media may require special processing to handle the specifics of the platform. The Medium Specific Processing component receives its input as a stream of digital documents in the format of the respective platform (obtained via the *Medium Specific Access*). It segments these, processes and transforms them and outputs streams of (generic) Documents and Communicators in the sense of the data model, which lend themselves to specific types of processing depending on their attributes and content. Many attributes can be set via a mapping process during the Medium Specific Processing step with the values obtained from the originating Medium; e.g. mapping JSON structures (first converted to hashes, such as `$tweet->{user}->{id}`[147]) to a Communicator's ID and associated hashtags (`$tweet->{entities}->{hashtags}`) to the respective hashtags attribute of a Document. Platform proprietary attributes and formats are mapped to generic and homogeneous ones. Attributes containing information which cannot be mapped, are processed by the appropriate means during subsequent specific processing. The Documents and Communicators resulting from this process maintain information linking them to each other where possible (e.g. a Document will contain information about the Communicator who is the author of the Document). Again, this is performed through appropriate mapping of attributes of the documents obtained by the

[147] PERL syntax as most of the code implemented for this component is in the programming language PERL

Medium Specific Access. Furthermore, these elements are linked to the platform of origin, allowing for later filtering and platform-specific querying.

5.6.8 Document Specific Processing

Document Specific Processing is centered on the Document's type and content. These determine the applicable processing steps and technologies. For the possible kinds see chapter 5.6.10 below. In addition to the content, a Document's meta-data may be processed using the same technologies, e.g. a (textual) title be processed by the same text-based enrichment as textual content. In case of segmentation, the individual segments will be processed separately. Several meta-data fields are normalized to conform to a common standard and unit. Duplicates and near-duplicates can be determined at this stage to avoid duplication of information and potential bias of resulting statistics and analyses. The detection and processing of links to other Documents, Communicators and Entities is discussed below.

5.6.9 Communicator Specific Processing

In contrast to the element of Document above, Communicator specific processing is focused on the meta-data associated with a Communicator and on the links involving the Communicator (chapter 5.6.3). These can relate the Communicator to other Communicators, Documents or Entities as described above in the data model. Associated Accounts may require further processing steps. Several meta-data fields can be normalized to conform to a common standard and unit. The Detection and processing of links to other Documents, Communicators and Entities will be discussed below.

5.6.10 Processing of Content

Even in a highly digitized world, input is sometimes still provided in analogue form which needs to be digitized (and meta-data added) before being processed by the proposed framework. Corresponding to the different classes of content, audio-, text- and visual processing steps may be applied. For multimedia documents, a combination of these technologies may be required. Determining what constitutes and delimits a document in the first place, especially from a streaming source, may require pre-processing to segment a continuous stream into individual documents. This may be based on an underlying logical structure of the documents or, in case this is not available, determined by time (e.g. segmenting a broadcast into chunks of an equal, predefined length).

Meta-data associated with documents corresponds to the same categories and can be processed in analogy to the types of content (albeit resulting in assignment to different attributes). For example, the 'location' meta-information of an account may be specified in textual form and as such be submitted to the same text-processing technologies as textual content but lead to the assignment of attributes of the Account and connected Communicator rather than being assigned or linked to a Document.

The processing of content and meta-data may lead to the detection of further elements, such as Communicators (e.g. via structural follower-information or via a reference in the content) which in turn may trigger further processing (e.g. issuing another set of queries to a platform's API in order to obtain more detailed information about an account referenced in the original document). Entities and Topical Markers such as hashtags as well as references to further documents (e.g. via embedded URLs) likewise may yield a series of Links. This setup leads to a feedback-loop of data ingestion and further queries based on the data obtained as outlined in Figure 49 above.

5.6.10.1 Processing of Textual Content

Textual content – as the content of Documents or present in meta-data – can be associated with a corresponding set of processing steps. A typical set of steps and sequence of actions is depicted in Figure 52 below (these technologies have been found to be of relevance for the processing of content during natural disasters in the incremental manner described above in chapter 5.6.1).

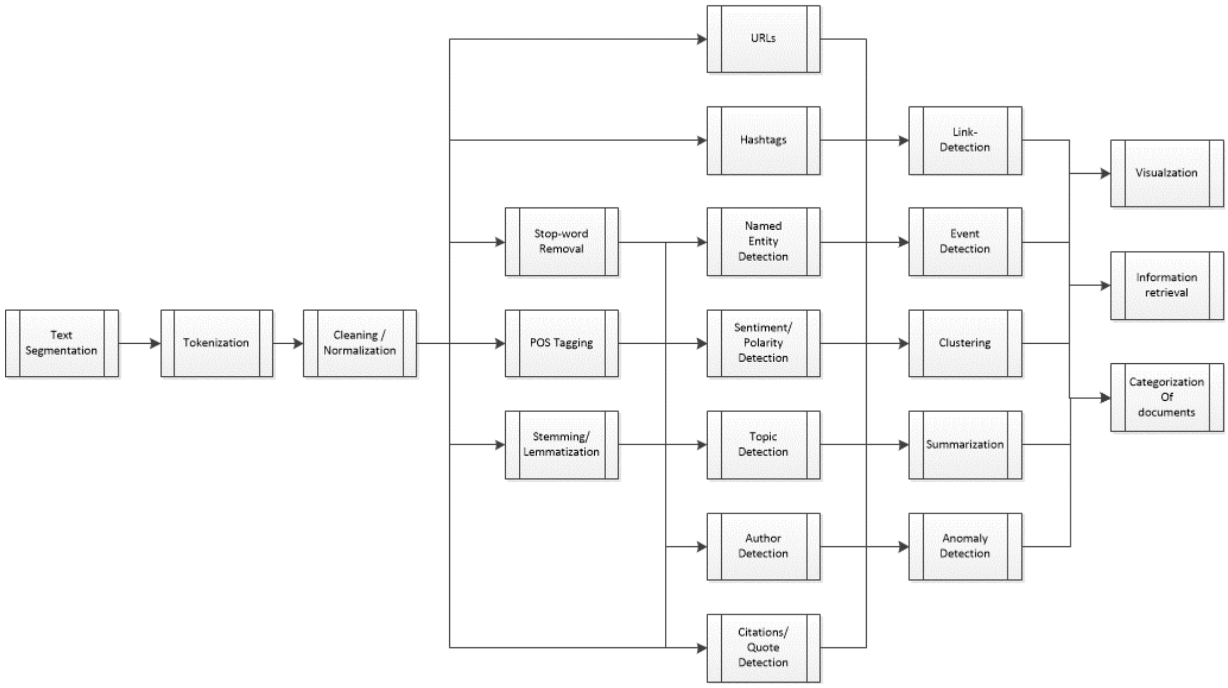


FIGURE 52 TEXT PROCESSING STEPS

The steps comprise sequences of pre-processing, normalization and cleaning, transforming the content into an adequate state for the remainder of the processing steps. Some technologies, such as Anomaly Detection, may need more context and thus may be applied only partially during this process, e.g. generating indicators for the actual process which can then be applied on the constructed graph during later graph extension.

The detection of Links plays an important role in the text-processing sequence: text Documents may contain different kinds of explicitly mentioned Entities, such as the names and titles of persons, organizations, locations, disasters, diseases etc. as well as a number of structural entities such as timepoints, dates, email-addresses or (monetary) amounts. Persons and organizations may lead to Communicators, allowing to directly relate the Document to them. Entities may have different ways of being referred to, e.g. by abbreviations, acronyms, nicknames and also exhibit misspellings (e.g. typically on social media or in case of little-known locations). In addition, transliteration of foreign names may vary depending on different conventions and on the native language, language skills and knowledge of the author of a Document. Hashtags, Topical Markers created by communities of users, can be found on several platforms. Whereas they originated on the short-messaging platform Twitter, they are also being used widely on Facebook as well as on Instagram and displayed on TV. During disasters, governments and first-responders have been trying to streamline the use of hashtags by asking people to use a pre-defined and standardized set of hashtags rather than ad-hoc ones to allow for better monitoring and easier retrieval of relevant documents (UNOCHA, 2014) – an approach which has had mixed effects on communication. A further set of elements providing linking information are URLs which relate the current Document to other Documents and possibly to Communicators. Especially on SM URLs may appear in a variety of shortened versions (provided by shortening-services such as bitly[148] or TinyURL[149]) which need to be expanded to retrieve the actual domain and URL. This in turn can provide important information as to the domains which are referred to and which may contain additional and popular information. On the other hand, a lack of referrals to domains of authorities and first-responders may prompt corrections to communication strategies.

Table 15 lists technologies which have been found to be of relevance for the processing of textual data in the scope of disasters and crises.

Textual	
Cleaning	Character encodings (all to UTF-8), elimination of ctrl-characters, mark-up,...
Tokenization	segmentation into words (units), separation of punctuation
Normalization	Acronyms, abbreviations, spelling-reform
Stop-Word-Removal	Removal of language and/or domain-dependent stopwords
Stemming/Lemmatization	replace words with canonical form (root-form)/strip-off-endings
POS-tagging	Tagging words with their resp. part-of-speech (word-class)
Dependency/Constituency Parsing	Syntax-analysis, yielding dependencies between constituents
Coreference Resolution	Relate expressions referring to the same entity
Language-ID	Determination of language/dialect from text
Segmentation	Segmentation of (longer) documents into sections, chapters, paragraphs,... also for OCR'd text segmentation into text-blocks or paragraphs
NED	Detection and classification of Named Entities (persons, organizations, locations, diseases,..), disambiguation, linking and resolution
NED-temporal	Detection and resolution of temporal expressions, "last-night", "next Saturday"
Disambiguation	Identyfying a particular meaning for a word or entity depending on context
Topic Detection	Detection and classification of topicality of text (acc to pre-defined taxonomy)
Clustering	Clustering/grouping of documents according to topicality
Summarization	Extractive or abstrative summarization of document
Sentiment/Polarity detection	Detection and classification of sentiment or polarity of document
Event-Detection	Detection of events (and associated properties: type, actor(s), location(s), time(s),...)
Localization	Geo-tagging (explicit coordinates) or geo-coding (from documents or user properties etc.)
Author detection	detection of authoring information, meta-data or authorship detection via statistical models
Citations	detection of quotes
Collocations	detection of sequences of words which co-occur with unusual frequency
Anomalies	detection of deviation of context and/or frequency of documents, communicators or key elements such as entities

TABLE 15 TECHNOLOGIES FOR TEXT-PROCESSING

It should be stated that not all of the above technologies need to be applied all the time, but rather that the list provides an overview of potentially relevant and meaningful technologies. Models and technologies may be supported at different levels of quality and performance depending on factors such as a Document's language. Further text-based technologies may be added to this table and can easily be accommodated in the overall processing scheme.

[148] bitly.com
[149] tinyurl.com

Processing examples:

- Language-ID to determine a Document's language in order to be able to apply further technologies. For social media content this may sometimes not be possible, but even an approximation may yield helpful results.
- NER: to detect occurrences of the names of first responders or kinds of diseases spreading in the aftermath of a disaster
- Topic-Detection: to classify a Document as covering economic aspects (e.g. regarding financial aid or donations) concerning a disaster
- Clustering: grouping of Documents about the same event in order to allow more rapid and focused access for first-responders and analysts, also to allow the tracing back of the flow of information
- Summarization: producing compressed versions of a Document or a set of Documents in order to allow faster scanning of events and be able to view them on portable devices
- Sentiment/Polarity Detection: to detect trends in Documents, shifts in public perspective, opinion or regarding attitudes towards donations after an event
- Event-Detection: to be able to detect coherent events in a stream of Documents, quickly alerting a crisis team to a developing situation, assignment of properties to particular types of events, e.g. the number of casualties caused by a disaster
- Localization: localizing information, actors and documents rapidly and being able to put them onto a map for deployment and response as well as to determine the geographic scope of damage
- Citations: to be able to attribute statements to actors and monitor the spread of information and influence of actors
- Collocations: to allow to discover terms used in connection which are specific to a certain event and allow to gain insights due to their combined meaning
- Anomalies: to detect changes in trends and situations rapidly and be able to react swiftly. This may include unusual occurrences of disaster-related terms but also unusual absence of such terms

5.6.10.2 Processing of Visual Content

As for textual content, the processing of visual content is typically carried out by a sequence of steps. As an initial step, image quality may be enhanced before further processing (Katsaggelos, 2012). Videos may be segmented into scenes or sequences, individual (key) images extracted and subsequently be processed as a set of still-images. Some technologies can be applied to videos directly and yield better performance than when applied to individual images as they can also take into account context over a sequence of images (e.g. applying tracking for the detection of faces or objects). The overall task of detection, analysis and interpretation of visual information is also referred to as *image understanding*, whose goal is to produce a semantic interpretation of visual content (Zhang, 2017). The technologies considered here form part of this larger field of image understanding.

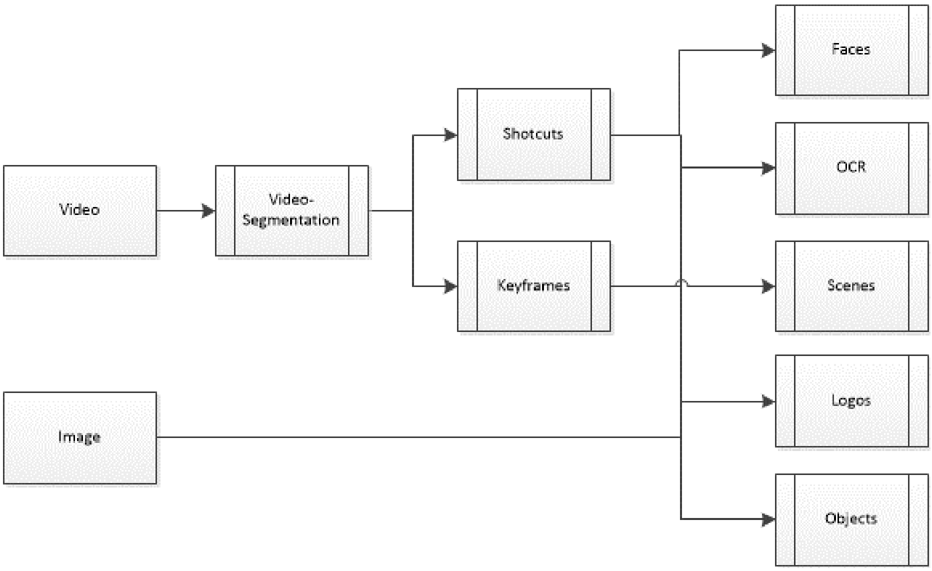


FIGURE 53 VISUAL PROCESSING STEPS

Images may contain Entities in textual form, such as inserts about the type and location of a disaster, requiring visual technologies for their detection and OCR for their subsequent conversion into processable text. Entities may also be of a visual nature, such as the faces of known persons (Communicators), the logos of first-responders and emergency services, flags or emergency symbols (Akella, 2009). For example, detecting the logo of the Red Cross/Red Crescent Movement may be used to help detect the presence of a vehicle, of supplies, containers or personnel at specific critical sites. This information paired with time and location information may in turn be employed for planning purposes. In general, technologies for detection and actual recognition/identification are required.

Object detection and recognition (Zhao et al., 2017; Liu et al., 2018) as well as scene segmentation and -detection/-classification (Del Fabro & Böszörmenyi, 2013; Abdulhussain et al., 2018) can be used to detect the presence of equipment, machinery or to detect hazards, e.g. fire or smoke and can also be used to cluster similar scenes, such as flooding scenarios (Bailer, 2016). Advances in deep learning (deep neural networks, DNNs) have emerged over the past decade as powerful accelerators for the detection and classification of a large number of objects yielding approaches which can also readily be applied to the domain of natural disasters. The detection and identification of faces (Wang & Deng,

2018) can be applied to visual content in order to link persons (Communicators) to a Document. The detection of manipulated material – the typical shark photoshopped into the image of a flood struck area – provides another valuable field of application for visual processing technologies. All of these fields remain highly active areas of research with a rapidly advancing state-of-the-art.

In addition to their visual content, images may contain meta-data embedded within them in the form of EXIF[150] tags and attributes. These entries may include location information in verbal or latitude/longitude format, information about the image itself, the author or camera, the circumstances of the image etc. all of which may contain Entities providing further background information. Meta-data in text form accompany images and videos and can be processed as regular text. The resulting information be used to enrich the information about the image or video itself as well as to link the image to further Entities, Communicators or Documents. Furthermore, this processing may be used for triangulation of information when trying to determine the veracity of (visual) content. In recent years a trend can be observed for platforms to actively remove such meta-data before providing images via their APIs. It can be assumed that this takes place primarily as this meta-data provides commercially valuable information for the platforms themselves and only less so because of privacy concerns. A lack of meta-data may thus put the focus of visual processing on the visual content itself even more.

Over the last years, a strong increase in the amount of visual data has been observed on social media. This general increase does not only concern still images, but to an even greater extent also concerns content in forms of videos (YouTube and Instagram are notable players in this platform-landscape, (Pew, 2018)). The accelerated growth of this type of content (and the increased absence of meta-data, see above) puts visual processing technologies at the forefront when aiming to extract critical information in general and especially during disasters. The omnipresence of mobile devices equipped with high-quality cameras paired with the rapid dissemination allow to gain immediate insights into developing disasters as they unfold – if the appropriate processing of visual content is available.

Table 16 lists several technologies which have been found to be of relevance for the processing of visual data in the scope of disasters.

[150] https://en.wikipedia.org/wiki/Exif

Visual	
OCR	Detection of text and recognition of text
Object Recognition	Detection and recognitoin of objects, e.g. acc to pre-defined taxonomy like "bridge"
Logo Detection/Recognition	Detection and recognition of logos (e.g. of NGOs)
Face-Detection/Identificaiton	Detection of faces and identification of detected faces
Scene Detection	Classification of images, e.g. acc to pre-defined taxonomy like "flooded-area"
Shotcut Detection	Detection of shot-cut boundaries
Keyframe Detection	Detection of keyframes in videos

TABLE 16 TECHNOLOGIES FOR VISUAL PROCESSING

As for text, it should be stated that not all the above technologies need to be applied all the time, but rather that the list provides an overview of potentially relevant and meaningful technologies. Models and technologies may also be supported at different levels of quality and performance depending on factors such as image quality or format or the underlying models for ML approaches.

Processing examples:

- Face-Detection to detect and identify spokespersons of emergency-organization or of politicians or officials
- OCR to detect text, segment it and convert it into text for further processing, e.g. to detect written inserts about a location or hashtags inserted to streamline communication. Information about the language and script may be required for this step.
- Logo-detection, e.g. to detect the presence of equipment, personnel or volunteers at the scene of a disaster
- Scene-classification to detect hazards and the level of destruction or flooding
- Object-recognition to detect infrastructural damage or important equipment
- Extraction of meta-data such as EXIF resulting textual information, processing to learn more about location, timing or authors

5.6.10.3 Processing of Audio Content

As other types of content, the processing of audio content can be carried out by a sequence of steps. Similar to the processing of visual data, steps of enhancing and normalizing content may be applied initially. Audio Documents may need to be segmented and the ensuing Segments be processed individually using a different set of technologies. Whereas certain technologies can operate on the audio in a language-independent way, other technologies require information about the actual language used (e.g. ASR). For the latter technologies, determining the language via language-identification may be required (and, as an even earlier step, determining whether and where an audio document contains speech at all).

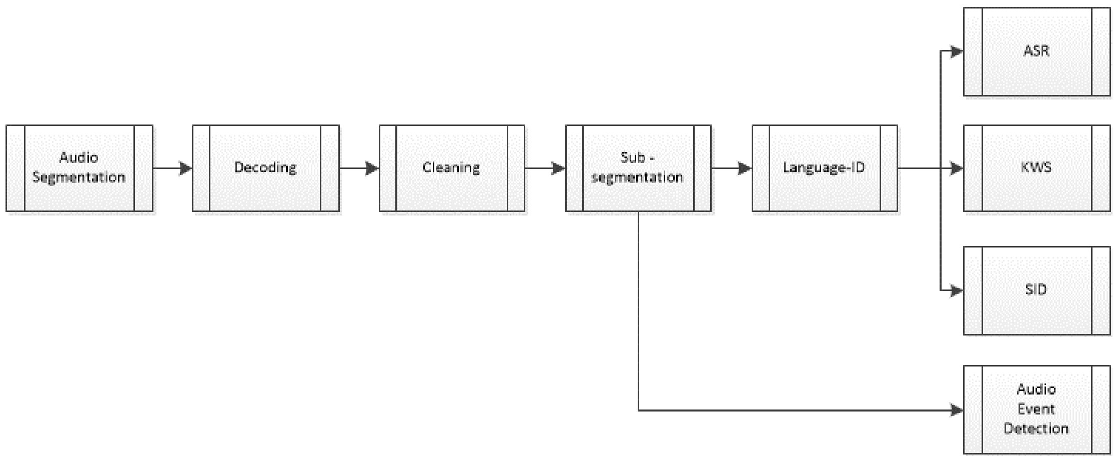


FIGURE 54 AUDIO PROCESSING STEPS

Table 17 provides a list of technologies which have been found to be of relevance for the processing of auditory data in the scope of disasters and crises.

Auditory	
Cleaning	SNR-ratio, noise-reduction
Segmentation	Detection of speech/non-speech, music, noise (explosion, screaming,...)
Language-ID	Determination of language/dialect from audio
ASR	Automatic Speech Recognition, transcription of speech/non-speech
KWS	Keyword Spotting, detection of content words/keywords
SID	identification of speaker, e.g. to determine the author of an audio document
Audio-Event Detection	Detection of events such as explosions, screaming,..

TABLE 17 TECHNOLOGIES FOR AUDIO PROCESSING

<u>Processing examples:</u>

- Detection of sections in audio containing speech in videos captured at the scene of disaster, eliminating sections of noise, silence or music and allowing to focus processing on relevant portions only
- Transcription of radio broadcasts covering the impact of a natural disaster, frequently involving local FM radio of inferior quality
- Detection of keywords for monitoring of TV broadcasts, interviews with locals and first-responders
- Detection and identification of officials and government spokespersons responsible for disaster communication based on their voice-print

5.6.11 Processing of Meta-Data

Meta-data might be present for Documents and Communicators and be available in textual or visual form (audio is theoretically also possible but has not been encountered in the media investigated for the current work). It may be processed in analogy with the processing of content described above. However, the use of the resulting information might be different, e.g. being able to locate a Communicator (or an Account) based on the description which may contain the name of locations and/or time-zone information. Special processing might be required for formats such as geo-locations and time-information which need to be normalized and converted to a standard format and unit to be comparable to other information.

5.6.12 Processing of Entities

The Entities (Named Entities, Structural Entities and Topical Markers) encountered on a platform are determined largely by the content of Documents but to a lesser extent also by the Communicators and their associated meta-data. Topical Markers such as hashtags are often used within one platform and medium but over time have also come to be used across different platforms and media and as such form an important source for cross-media linking of information (as well as playing a role in the dissemination of patterns and memes (Jain, 2013)). Entities and Topical Markers may be detected using textual, visual and auditory processing. They occur in the content of Documents, in written form, graphically (e.g. as logos or inserts on TV) or acoustically (e.g. by being mentioned on a radio program or audio-clip). Textual and graphical meta-data associated with Communicators and Documents provide further sources for the detection of Entities and Topical Markers. Topics as produced by a *Topic Detection component*[151] can be regarded as an automatically assigned Topical Marker and be processed in analogy to hashtags. Together these elements create topical links between Documents and Communicators on different platforms and in different media-types and thus serve to relate them within and across platforms and media borders. The detection of Entities can be performed depending on their nature: structural entities, such as dates, monetary amounts or email-addresses can be detected via pattern-matching (e.g. using regular-expressions). Topical Markers such as hashtags may likewise be detected via matching mechanisms. Topics typically are associated with meta-data of the Document or Medium or with phrases within the content. Named Entities can be detected via matching mechanisms which need to take variations and the morphology of the language into account. They can be linked to lists, ontologies, gazetteers or based on statistical mechanisms for tagging, typically operating on the context of an entity, the surrounding words or features of the entity itself (Nadeau & Sekine, 2007). In addition, Entities may already be marked as such in meta-data of documents in which case they can simply be mapped (e.g. with the separate elements for hash-tags as provided by the JSON structures available via Twitter's API).

[151] Either from a pre-defined set using an existing model and classifier or from (unsupervised) clustering with automatically generated topic-labels.

5.6.13 Processing of Within Media Links

The detection and processing of links between the elements of the communication model is the basis for the creation of connections between the entities themselves. As outlined above in the chapter on Links within the data model (chapter 5.5.7), a variety of connections exist and need to be accounted for. Moreover, whereas several links may be detectable from the elements themselves, others can only be established by inference over an initial network (*indirect Links*) in a second phase of graph extension. For example, by authoring one Document (D1) containing an entity (E) also appearing in another Document (D2), the Communicator (C1) also becomes indirectly related the other Document's author (who is also a Communicator, C2) and the two Documents likewise become related indirectly. These indirect connections are depicted by the dotted arrow[152] in Figure 55.

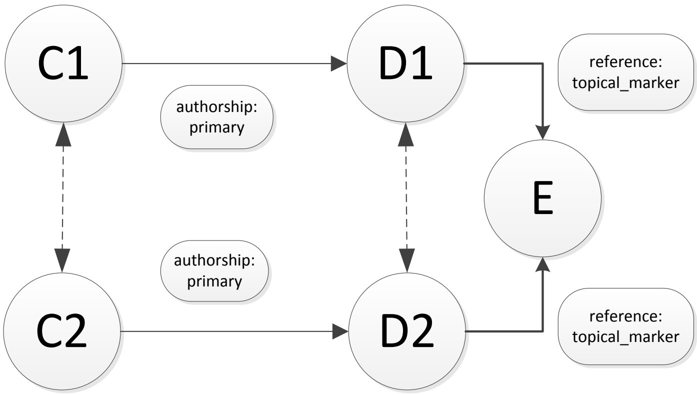


FIGURE 55 INDIRECT LINK BETWEEN COMMUNICATORS AND DOCUMENTS VIA A SHARED ENTITY

Several processing steps regarding the detection of links are carried out already in the phase of Medium Specific Processing as the precise content of documents from a platform also determines the possibilities and allows mapping to the generic attributes and types of links as defined in the data model. Further processing, especially concerning the content and meta-data, takes place in the Document and Communicator Specific Processing phase, where the detection of Topical Markers and further entities is carried out.

5.6.13.1 Communicator -> Medium

A link between the Communicator and a particular Medium can be established during Medium Specific Processing if the identity of the Communicator is known. An Account may indeed be the *driver* of the processing - e.g. when querying for Documents from a specific *user* (Account) - and thus allow to link

[152] Bi-directional arrows representing two separate directed connections.

the Communicator with the Account's Medium. Linking back to the Medium allows to view and manipulate elements in their *original environment*, which may provide additional context (e.g. information not available via API access).

5.6.13.2 Communicator -> Document

Communicators and Documents may be related in a structural, interactive or referential manner. Authorship links a Communicator structurally with a Document. The level of authorship (primary, secondary) and directness are assigned to attributes of the link. Authorship may be detected from meta-data or by detection within the content of the Document when stated there explicitly. If the publisher of a Document is known, the respective Communicator can be directly linked to the Document as author.

Interaction between a Communicator and a Document may take place in two manners:

- Providing feedback on a Document, possibly by creation of another Document or by altering the state of the original Document (e.g. the number of *likes* or *hates*)
- (Re-)Distribution of a Document which may be passed-on in its original form or in an altered form (e.g. re-tweeting)
- The former type of interaction may directly involve the author of the Document – as the action is directed at the Communicator (the author) - whereas the latter may involve only further structurally connected Communicators of the interacting Communicator (e.g. their followers) and not concern the original Communicator (author) him- or herself.

Links of Structural or Interactive nature have to be detected and processed at the Medium Specific Processing stage due to the platform dependent specificities.

Regarding the referential connection, a Communicator's meta-data may include links to further information, e.g. a URL of a person's or organization's home-page. If Communicators refer to another Document within a Document authored by them, this connection is modelled by a corresponding direct Document -> Document connection.

5.6.13.3 Communicator -> Communicator

Structural links between Communicators may be inferred from platform specific information available with the Communicator (typically in meta-data or by access to specific, platform dependent functionality via specific API calls). Depending on the platform, it may be of a symmetric (e.g. C1 is-friend-of C2) or asymmetric (e.g. C1 is-follower-of C2) nature. Due to this dependence, the type and its attributes are set in the Medium Specific Processing stage.

In case structural links are known from external sources, such as knowledge about membership in organizations, or working relationships (e.g. cooperation during disasters), these links may be added manually or via specific processing of external structured information, e.g. FOAF[153] or an organigram of an organization.

Referential links between Communicators are created via a reference to a Communicator in another Communicator's Document or via interaction with such a Document. These links are of an indirect nature. For example: C1 authors a Document D1 citing the name of C2. An indirect link between C1 and C2 is created by this reference as depicted in Figure 56 below.

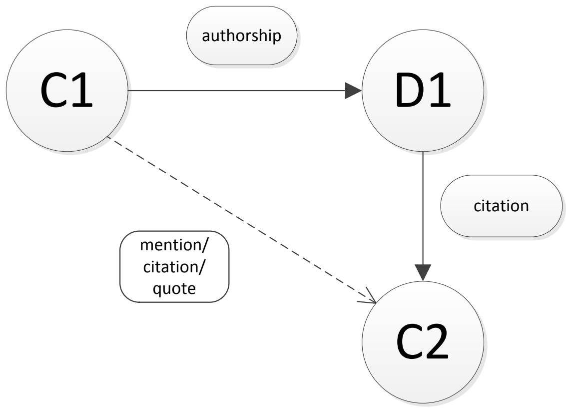


FIGURE 56 INDIRECT LINK BETWEEN COMMUNICATORS VIA CITATION

This type of link can be detected in the Medium Specific Processing stage as it may entail knowledge about the exact mechanism (and syntax) of reference for the specific platform. For example, the use of the character '@' with an Account or Communicator's name may be a mechanism to refer to the Communicator (resp. their Account) on certain platforms. Naming the Communicator in plain text or audio is another way of creating such a link, requiring the appropriate processing of the Document's content (Document Specific Processing) and detecting the Communicator as an entity mentioned in the content.

153 http://www.foaf-project.org/

5.6.13.4 Communicator -> Entities

In case a Communicator corresponds to a known person or organization (Entity), the Communicator can be associated with this Entity (i.e. linked to it). This link may be of a temporal nature only, e.g. when the Entity represents an office held temporarily by an individual (Communicator). In case of unknown Communicators or digital identities, a *proxy* Communicator can be assumed. The identity may be known as part of the data provided by a particular platform and be associated in the Medium Specific Processing stage. It may also be inferred from a Document's content. Further links to other types of Entities and Topical Markers may likewise be established during Document Specific Processing (from the content) or Communicator Specific Processing (from the Communicator's meta-data).

5.6.13.5 Document -> Medium

A link between the Document and a particular Medium is established by the Medium Specific Processing as the Document is extracted in a platform specific manner and converted to a generic Document with attributes.

5.6.13.6 Document -> Document

Documents may be related in a structural way or via reference. Medium Specific Processing may include information as to whether Document-1 is an updated, newer version of Document-2, in which case these two Documents are linked. For platforms allowing for interactive feedback and commenting, Document-1 may be a reply or comment to a Document-2, in which case this specific information may likewise be used to link the two Documents. Documents may also reference each other in their content, e.g. by quoting or mentioning explicitly another Document or by inclusion of URL linking to the referenced Document. The latter type can be detected in the Document Specific Processing stage.

5.6.13.7 Document -> Communicator

As outlined above, a structural relationship between a Document and a Communicator exists if the Communicator is the author of the Document. In case of platforms allowing for feedback or commenting mechanisms, the Communicator performing this action may be linked directly to a Document (e.g. in case of like, hate etc.) or indirectly via another Document which is a reply or comment to the original Document (also see above). This information may be gathered during Medium Specific Processing. A Document's content may refer to a Communicator by mentioning their name or an associated Account which allows to create a referential link between the Document and the mentioned Communicator (and possibly an Entity linked to the Communicator). Such links can be detected during Document Specific Processing.

5.6.13.8 Document-> Entities

A Document's content may refer to a set of Entities and Topical Markers. In case such Entities correspond to Communicators, the Document and the respective Communicator can be linked (see also above). Further links to other types of Entities and Topical Markers may likewise be established during Document Specific Processing (of the content). The different kinds of processing of the contents of Documents has been described in chapter 5.6.10 above. If several Documents refer to the same Entity/Entities, these Documents are linked indirectly.

5.6.14 Processing of Cross-Media Links

Most of the different kinds of links discussed above do not cross boundaries between the different Media they originate from. Many structural and interactive connections result from activities within one and the same platform. URLs, however provide one kind of link which may cross different kinds of media, e.g. in the case of a textual Document containing a URL whose domain leads to an image on a different platform. Entities and Topical Markers provide an even wider field of potential links across and between platforms as they may originate in Documents from different Media, of different media types as well as in the meta-data associated with Communicators. Detecting them in Documents originating from different media allows to link previously unrelated information and infer connections which are not visible *on the surface* going beyond the individual possibilities of working with data from a single platform only. Detection and extension of such linking is performed on the initial graph constructed during the Graph Construction phase and capitalizes on the nodes and edges created in that phase. The extended context provided by the original graph allows to relate elements on a wider scale and to infer further relationships.

Examples of links which can be detected and used for graph extension are:

- Hashtags detected during the Document Specific Processing of textual Documents and also detected during Document Specific Processing of images where they appear as text-inserts.
- Entities, such as names of organizations mentioned in textual Documents and also mentioned by speakers on TV
- Entities mentioned on radio and depicted in an image on a social media platform

- URLs depicted on images or appearing in video, leading to a web-page

5.6.15 Graph Construction and Extension

The processing of Documents and Communicators, of their associated meta-data and the content of Documents as described above results in a series of nodes and relations for initial graph creation as depicted in Figure 57 below. Properties of all nodes as well as edges are assigned according to the data- and process model described above.

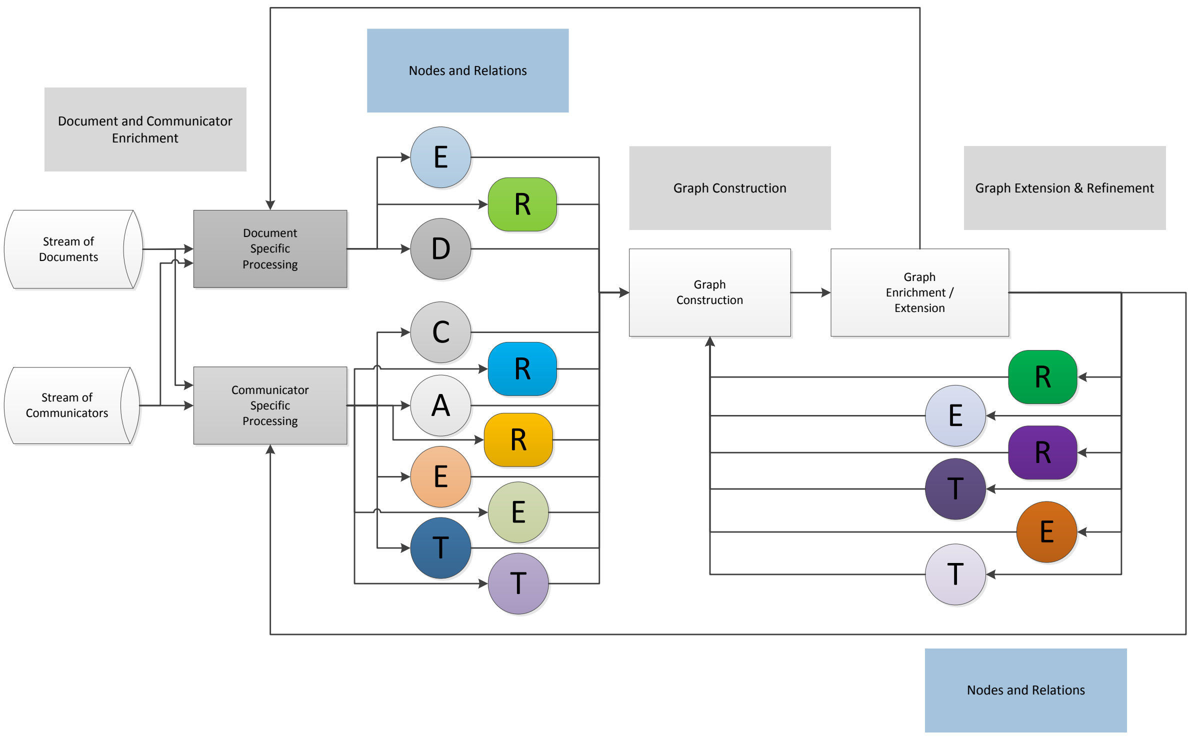


FIGURE 57 GRAPH CONSTRUCTION AND EXTENSION

Similar to the way that integrity constraints have to be considered in relational databases, the dependencies of node and edge-creation have to be taken into account (Šestak et al., 2016). A suggested general sequence of creation is for Communicators, Accounts, derived Entities and Topical Markers to be added to the graph first followed by the addition of Documents and their respective derived Entities and Topical Markers. Information about the connectivity of Documents and Communicators is passed to the Document and Communicator Specific Processing stages in order to be able to provide the respective edges and add them to the graph structure.

As for the creation of the individual nodes, edges and their respective properties, the same steps can be applied to extend the graph and further enrich its elements in a second phase of graph extension and refinement. Cross-media links can be detected and added at this stage, resulting in indirect links between elements. Whether and to what extent to create direct links from these indirect links needs to be investigated and has to take into account processing and usability factors. In case updated and improved models for the enrichment technologies are available (e.g. an updated version of Named Entity Detection), these can be applied to the content of Documents and further Entities and Topical Markers can be added. Subsequently, inference of further relations can be carried out, resulting in a cyclic model of processing as depicted by the feedback arrows in Figure 57 above.

The overall processing thus combines incoming streams of Communicators and Documents, their processing and enrichment, resulting in a series of nodes and edges describing these elements and their relationships with the construction and extension of a graph structure representing them. The same structure can be used for querying and OLAP. From a practical perspective, the resulting graph will need to be pruned back at some stage as continuous addition of elements will cause an otherwise infinite growth. The use of date and time information on both nodes as well as on edges can be employed for such reduction. Another method could be to keep track of nodes and edges, their use in references and OLAP activity and to prune the least important (i.e. least frequently used) ones. Human intervention by an administrator with explicit knowledge of the contents and demands on the graph provides a further layer of intervention to keep structures manageable.

5.6.16 Summary

Chapter 5.6 dealt with how to construct the elements of the data model identified previously from the various kinds of media and platforms (described in chapter 5.5) and how to process them, their content and meta-data. An overall processing model was introduced, comprising the steps from accessing data on different media, ingesting, processing and enriching it and finally creating a structure representing the elements of the communication model and their respective relationships (see Figure 49). Processing of connections between the elements provides additional information and was detailed in a separate sub-chapter. The result of this process is a property-graph consisting of enriched nodes and connections between them representing a variety of relations. This graph is constructed in a first iteration and then extended (by further links) in a second iteration, inferring non-trivial connections from the structure constructed by the first iteration.

All processing is envisioned to take place in a continuous manner, so as to provide frequent updates and near-real-time situational awareness – a crucial aspect when dealing with natural disasters. The graph can be employed for querying, to create visualizations, and for online processing and is extended and enriched over time. Interactivity and drill-down functionality on these graphs is envisioned to allow increasing levels of detail with the appropriate level of granularity available for analysis during all steps of disaster management.

6 Exemplary Application to Disaster Communication and Evaluation

Having devised a data- and process model in the chapters above, the question arises how these models are to be evaluated. The artifact's *intended purpose* - the model's scope and functionality and the method to produce the model - has been laid out in the research questions formulated in chapter 2.4. These research questions form the context for the discussions below. The research questions themselves are revisited and individually addressed in chapter 7.3.

Venable et al. (2012) provide a set of guidelines for the evaluation of DSR artifacts. Three criteria to be considered are:

- what will be evaluated?
- when will evaluation take place?, and
- how will evaluation be performed?

Regarding the first criterion, the model and its partial implementation (*instantiation*) is considered to represent an *artifact* as defined by Hevner et al. (2004). Consequently, the model itself and the method followed to create it will form the focus of evaluation.
Regarding the second criterion, an ex-post, summative perspective is adopted. Evaluation takes place after construction and partial implementation of the artifact[154].
Regarding the third criterion, the artifact will be evaluated from the perspective of a set of factors contributing to *utility* and *efficacy* and the extent to which the artifact achieves its intended purpose (Hevner et al., 2004). More specifically, the question as to whether the developed model represents a viable generic model for the elements and operations involved in cross-media communication during natural disasters will be examined. In addition, exemplary cases of the application are presented to

[154] Several iterations remain to a full implementation of the model. Hence this evaluation may also be viewed from an ex-ante/formative perspective.

illustrate the viability and scalability of the approach and to emphasize that the artifact does indeed meet its goal.

Parts of the models and processes have been implemented - attesting to Hevner et al.'s (2004) criterion of *implementability of an artifact* - allowing to gain some quantitative insights and to assess their practical applicability. These quantitative insights support the motivation and choice of the selected approach and model. Following the general methodology laid out in chapter 2.5, the construction of the model did not proceed linearly, but rather consisted of several iterations, with insights of partial evaluation feeding back to repeated steps of design and implementation. In this respect, it follows the general iterative *generate/test* pattern propagated by Hevner et al. (2004). In a larger context, prior work and the results and insights gained from it provided a first large iteration of design, implementation and feedback leading to an improved design of the models within this thesis.

The resulting implementation serves to provide evidence and support regarding the assumptions underlying the construction of the model and will be considered when revisiting the research questions below. However, as it is limited only to individual aspects of processing, it cannot be employed to quantitatively assess the complete overall model or set of processing steps. To assess this overall model, a qualitative view, considering a set of established evaluation-criteria, will thus be adopted below.

6.1 Quantitative Insights

In the sections below, several issues will be discussed which quantitatively characterize parts of the model or which can be viewed as supporting evidence for the practicality and utility of such models. Moreover, they serve to underline the motivation of the chosen approach. In the book-chapter "A General Framework for Using Social and Traditional Media During Natural Disasters: QuOIMA and the Central European Floods of 2013" Backfried et al (2016) describe some of the phenomena in a quantitative manner. These provide insights from practical, real-world events and are discussed below.

6.1.1 Use of Social Media in Austria during Disasters

Whereas the models developed are expected to be of a generic and general nature, some of the foundations were initially motivated by work on disasters in the context of Austria. At the time, less than 1.5% of all Austrians were active on Twitter, with a highly a skewed distribution of activities focused on a handful of personalities and accounts (Ausserhofer et al., 2012[155]). This marginal use of Twitter is typical for many parts of the world and thus may also apply to other regions where disaster strikes.

Nevertheless, a remarkable number[156] of Tweets could be found during the floods of 2013 (the period of May 20th to June 23rd, 2013 was studied in detail as this formed the period of most severe flooding in Austria). Not only is this an indicator that Twitter is indeed a source of information to be considered during disasters in Austria, it also demonstrated the typical links featured so prominently in the proposed model. More than 17% of all links encountered in Tweets linked to Facebook, 9% linked to videos on YouTube and 6% linked to images on Instagram – the latter providing support for the proposed approach and postulated importance of cross-media analysis. Many Tweets linking to newspapers and TV stations provide further evidence of a highly-connected media-landscape during disasters. Social media and traditional media indeed form an inter-connected network. Consequently, they should also be considered in this way rather than be viewed in isolation – a shortcoming of current approaches which was noted in the chapter on gaps above (chapter 2.2).

Automation of accounts (*bots*[157]) played a considerable role during the event. A bot-detection component (Chu et al., 2012) using a *conservative detection setup* yielded approximately 1% of the accounts involved to be bots, creating approximately 22% of all tweets encountered. Manual inspection showed that several bots provided regular information about water-levels whereas others were used purely for commercial means (e.g. insurance for water damage). Information from such accounts should neither be discarded nor included blindly or completely for analysis. Rather, a distinction between the different kinds of bots should be made. Marking accounts as automated or partially-automated was perceived as beneficial by end-users and should be included in an overall system.

In contrast to Twitter, Facebook is used on a much broader scale in Austria (as of April 2019 an estimated 68% of the population; in May and June of 2013 approximately 80%[158]). The investigation of the communication on Facebook – limited to public pages only - showed that during the floods of 2013, a single page operated by an individual managed to absorb a large share of communication

---

[155] Presentation available at: https://www.univie.ac.at/twitterpolitik/studie/TwitterPolitik_Studie.pdf

[156] A total of more than 470.000 tweets was collected during this period; however, these also include tweets from Germany and the Netherlands as they were experiencing a flood at the same time.

[157] The phenomenon of mis- and disinformation was not investigated during the project. However, this issue has clearly gained in importance and provides a strong argument for detection of (partially) automated accounts. Cross-media behavior may provide additional, valuable clues for such identification.

[158] http://gs.statcounter.com/social-media-stats/all/austria/#monthly-201301-201804 on 2019/04/23

within only a few days after the flood hit Austria (approximately 66% of all posts and comments collected over all FB-pages during that period (Backfried et al., 2016)). Facebook pages of official institutions (government, authorities), the Red Cross or NGOs attracted much less attention (if any) and were rarely linked-to by other platforms. Especially in the light of rumors and incorrect information which was spread on the individual's account, this resulted in a negative situation for first responders, who were forced to intervene and correct later. Clearly, there is a need for first responders to continuously monitor the flow of information across a wider spectrum of media in order to avoid such potentially dangerous situations and to be present on traditional as well as on popular social media platforms - if they are not active, an *information vacuum* may result which will quickly be filled by other actors.

6.1.2 Locations and Disaster Phase

Triggered by interest from a first responder, the Austrian Red Cross, several locations along the affected rivers in Austria (Salzach, Inn, Danube) were investigated regarding the level of tweeting behavior in connection with links to images. Locating information and gaining insights through associated images may lead to a better understanding of a local situation and lead an increase in situational awareness for first responders. Table 18 below lists the selected locations and relates their approximate population with a number of factors regarding communication on social media. Among these factors is the number of tweets containing links to photo-sharing sites which were found to exist even for medium-sized communities. It can thus be assumed that even for regions with limited population, information including images can be found on social networking sites. Given the recent trends towards increased proportions of multimedia content, it is safe to assume that this situation is even more pronounced today.

	Geißau	*Braunau*	*Salzburg*	*Wels*	*Linz*	*Enns*	*Melk*	*Aggsbach*	*Krems*	*Korneuburg*	*Klosterneuburg*
Inhabitants	1600	16000	147000	58000	191000	12000	5000	682	24000	12000	26000
Tweets with like %{loc}%	0	27	991	512	912	194	201	0	318	48	97
tweets/inhabitant	0,000	0,002	0,007	0,009	0,005	0,016	0,040	0,000	0,013	0,004	0,004
Users with Location like %{loc}%	0	32	503	46	306	28	2	0	25	19	24
Tweets with photo-links	0	7	37	1	130	6	14	0	23	1	2
Tweets with photo-links (%)	0	25,93	3,73	0,20	14,25	3,09	6,97	0,00	7,23	2,08	2,06
Tweets with lat/long	0	0	4	0	7	0	1	0	1	0	0
Tweets with lat/long (%)	0	0	0,40	0,00	0,77	0,00	0,50	0,00	0,31	0,00	0,00
FB posts like %{loc}%	0	4	18	5	60	7	14	2	9	11	16
FB comments like %{loc}%	0	6	53	20	182	74	37	4	34	25	17
FB sum	0	10	71	25	242	81	51	6	43	36	33

TABLE 18 LOCATIONS AND SOCIAL MEDIA IN AUSTRIA

Table 19 below depicts the activity on Twitter over time and for three phases of the disaster cycle: pre-execution (blue), execution (orange) and follow-up (purple) demonstrating activity in all three phases (for the phases see Figure 15).

tweets / time / location	05		06								
	30	31	01	02	03	04	05	06	07	08	09
Braunau	0	0	0	7	9	6					
Salzburg	7	4	6	554	166	62	37	21	10	9	24
Wels	0	0	0	6	22	28	27	24	25	26	24
Linz	0	21	3	119	391	161	80	46	12	14	3
Enns	1	12	2	27	25	26	5	4	8	8	6
Melk	0	0	0	27	55	26	6	6	4	9	5
Krems	0	0	10	16	44	40	43	18	18	6	8
Korneuburg	0	0	0	2	15	10	13	4	1	2	1
Klosterneuburg	0	0	0	22	8	23	11	7	5	5	0

TABLE 19 ACTIVITY ON TWITTER ACROSS DISASTER PHASES

The lag between the cities of Melk and Korneuburg can be explained by their relative geographic position with Korneuburg being located further downstream on the river Danube from Melk as depicted in Figure 58 (locations mentioned in table above marked in red, map created with Google Maps).

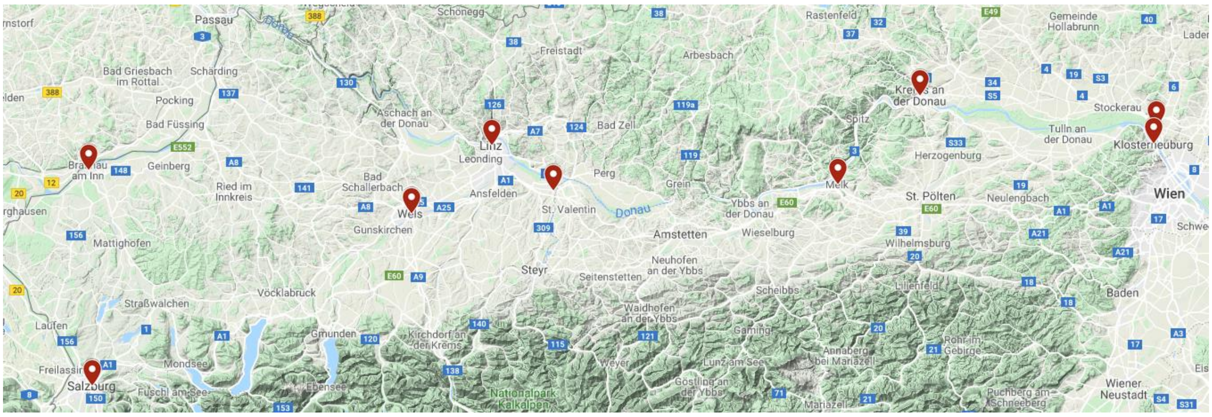


FIGURE 58 LOCATIONS OF SM-ACTIVITY ACROSS AUSTRIA

The destination of links on Twitter – and the kind of content communicated - changed over the core period of the flood, with linking to donations-related sites becoming more prominent during later stages of the disaster as depicted Figure 59 below (percentage of tweets over the core period of the flood, May 20 - June 23 2013. This is the same as Figure 18 above, repeated here for convenience).

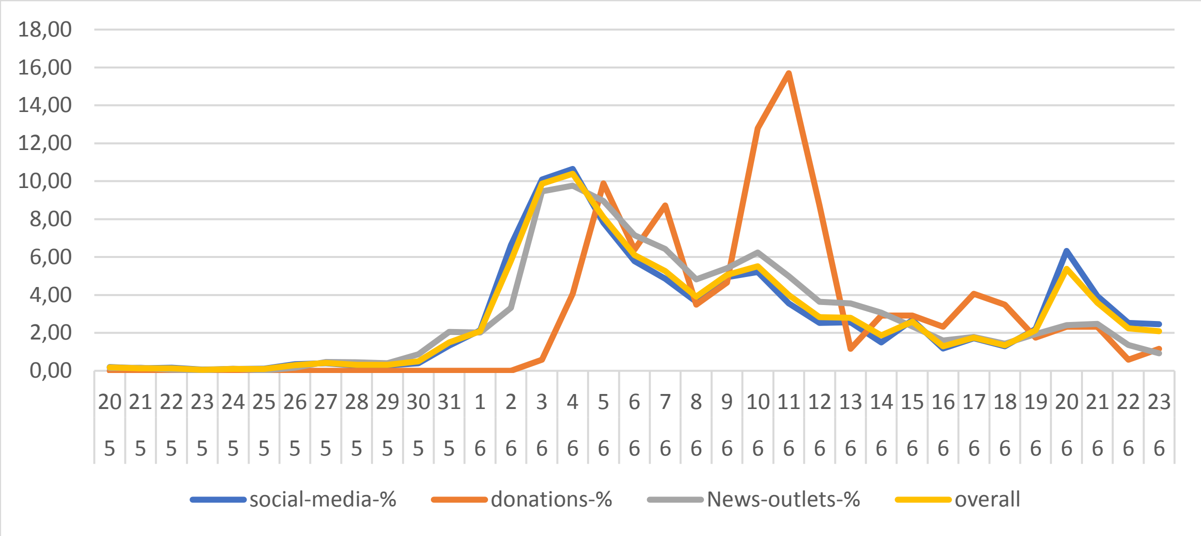


Figure 59 Destination of links over time

6.1.3 Traditional and Social Media Coverage of Crises

One of the underlying assumptions justifying the general approach selected in this thesis is an increased use of social media - including a variety of platforms and media – in combination with a continuing use of Traditional Media in the scope of natural disasters. Whereas it can be assumed that traditional mainstream media indeed still do cover natural disasters (Greenberg & Scanlon, 2016) and will also continue to do so in the future, a strong increase in social media coverage could be observed over the last decade. In a recent study, Reuter et al. (2018a[159]) investigated the presence of social media in papers presented at ISCRAM (the International Conference on Information Systems for Crisis Response and Management, one of the prime conferences in the field) conferences over the last 15 years. Figure 60 below displays the total number of papers presented (red line) in comparison to the percentage of papers discussing topics of social media (blue bars). The increased academic and professional attention over time reflects the higher significance and more widespread use of social media during disasters on a global scale.

[159] The author of the current thesis is a co-author of the referenced paper; figures 60 and 61 from that paper.

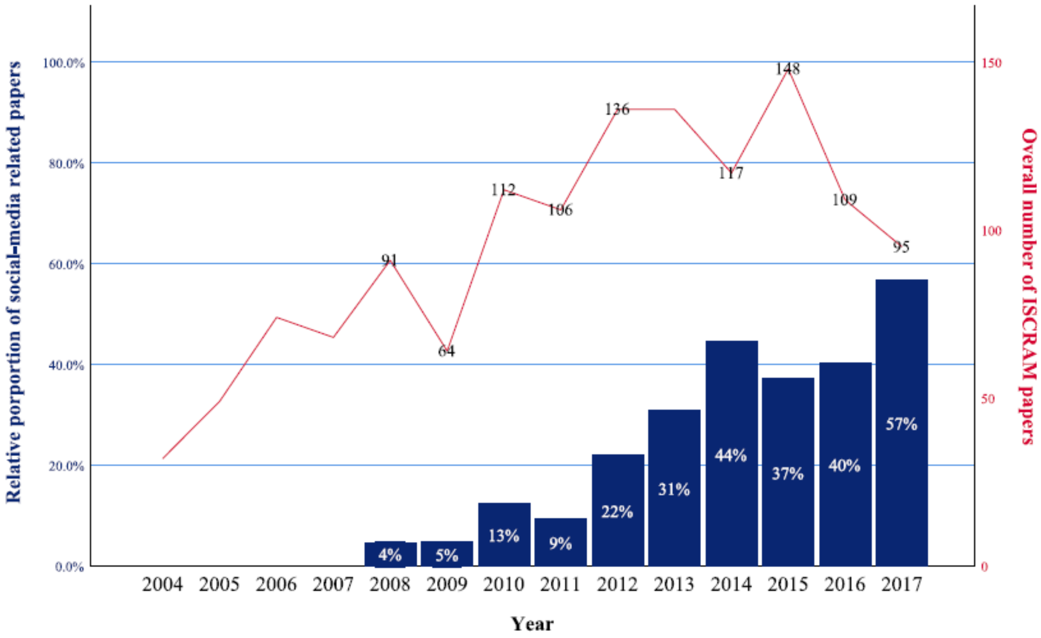


FIGURE 60 TOTAL NUMBER OF PAPERS AND SM-PAPERS AT ISCRAM CONFERENCES

The different kinds of platforms investigated are depicted in Figure 61 below. Whereas Twitter has been the focus of the majority of investigations (for possible reasons see Reuter et al., 2018a), other platforms also play a significant role with multimedia platforms on the rise. The increase in the portfolio of platforms and media supports the chosen model of this thesis and the linking of information across platforms.

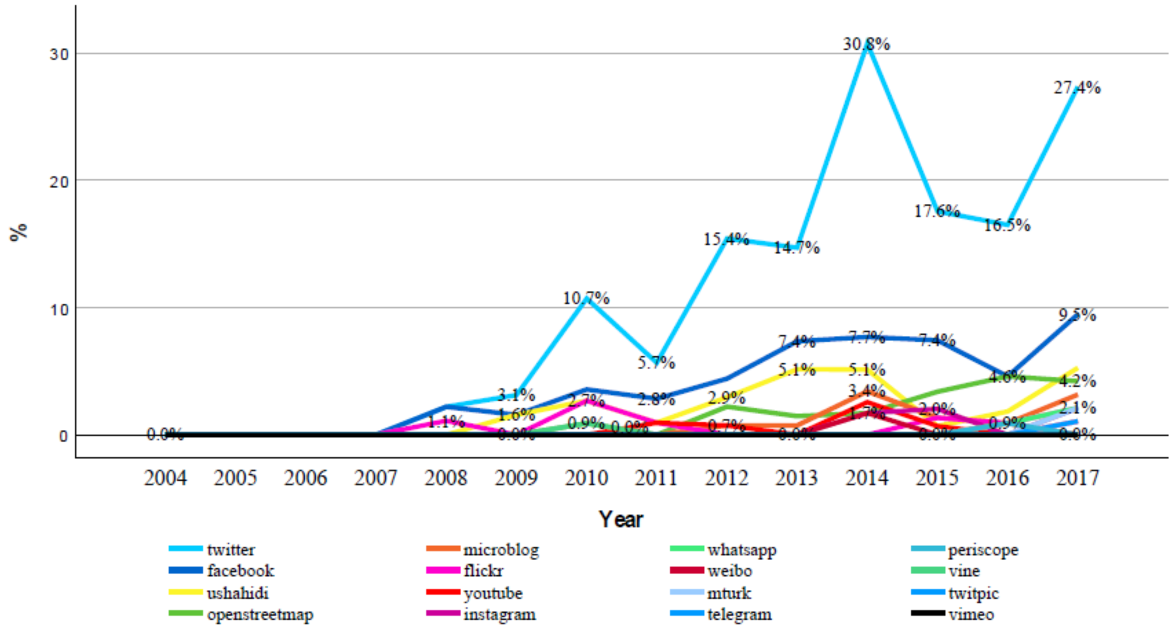


FIGURE 61 PLATFORMS INVESTIGATED

6.2 Qualitative Insights

The overall model developed in this thesis consists of a data model and process model. The data model focuses on the ingredients, the *What*, whereas the process model is dedicated to the *How* in the context of processing multiple media across multiple platforms. Naturally, the two models are intricately connected and have been developed hand-in-hand. When referring simply to *the model* without further qualification, their combination is meant below. Their development followed the incremental, iterative approach outlined in chapter 2.5. This approach combines a top-down and bottom-up manner of model construction and aligns well with the recommended cycles of DSR. Incremental steps of model extension were followed by (informal) evaluations on whether the extensions were indeed able to accommodate the newly added elements – adjustments and corrections were applied before entering another cycle.

In the sense of DSR, the model forms an artifact, which partially also lead to an implementation, an *instantiation* (Hevner et al., 2004). For the evaluation of an artifact, Venable et al. (2012) suggest examining the artifact's *utility* and *efficacy* for achieving its intended purpose and further elaborate that these complex concepts can be described in terms of functionality, completeness, consistency, usability and a set of related criteria. Building on the above, Prat et al. (2014) define a set of criteria for the evaluation of an artifact of which the ones listed in Table 20 are deemed relevant to the current work.

Efficacy	the degree to which the artifact produces the desired result
Validity	the degree to which the artifact correctly achieves its goal
Utility	the quality of the artifact in practical use
Understandability, Clarity	the ease with which concepts and structures in the data model can be understood[160]*
Ease of Use	the ease of practical use of the artifact
Completeness	whether the data model contains all information required to support the required functionality of the system*
Simplicity	whether the data model contains the minimum possible entities and relationships*
Level of detail	granularity and detail of the model with regard to reproducibility of the artifact**[161]
Consistency	the internal consistency of the artifact

TABLE 20 ARTIFACT EVALUATION CRITERIA

[160] *: Neither Prat el al. (2014) nor their reference provide an exact definition of this term. However, as the terms are identical to the ones defined by Moody (2003), these definitions are replicated here.

[161] **: Interpretation of the author, as the term is not explicitly defined in the referenced papers (nor in their included references).

These evaluation criteria provide the *dimensions* along which the developed model and its exemplary application should be examined. In addition, the *instantiation*, according to Prat et al. (2014) can also be considered an evaluation of the (abstract) artifact per se.

From an information science (IS) centric perspective, an empirical evaluation of the use of quality metrics for data models undertaken by Moody (2003) provides a series of factors to measure the quality of such models. In a later study carried out by Gosain and Heena (2015) who reviewed 22 studies in the field of measuring data model quality for data warehousing, these quality factors were confirmed to feature among the most important metrics.

The data model quality factors identified by Moody (2003) and further investigated by Gosain and Heena (2015) comprise two basic dimensions:

- Business dimensions: completeness, integrity, flexibility, understandability,
- Technical dimensions: correctness, simplicity, integration, implementability

Table 21 below lists Moody's definitions of these terms[162].

Correctness	whether the model conforms to the rules of the data modelling technique, i.e. whether it is a valid data model
Completeness	whether the model represents all information required to support the targeted functionality and scope of the system
Integrity	whether the model defines all business rules which apply to the data
Simplicity	means that the model contains the minimum possible entities and relationships
Flexibility	the ease with which the model can cope with business and/or regulatory change
Integration	the consistency of the model with the rest of the organization's data
Understandability	the ease with which concepts and structures in the model can be understood
Implementability	the ease with which the model can be implemented within the time, budget and technology constrains of the project and organization

TABLE 21 MODEL QUALITY FACTORS

162 Whereas Moody (2003) refers to *data models*, the definition can be interpreted in a more general sense and the more general term *model* is consequently used here.

Both studies, Moody (2003) and Gosain and Heena (2015) stress the importance of *Understandability* as a key element. In the same vein, Simsion and Witt (2004)[163] emphasize the related factor of *Communication*, referring to the effectiveness of the model for communication among stakeholders. They also conclude that terminology issues become especially important when communicating about models with a high level of complexity and unfamiliar concepts (a point of view thoroughly embodied in the present work). This factor of *Communication* is viewed in terms of understandability for communication with stakeholders below. None of the studies addresses quality factors for the evaluation of process models. However, to some extent the identified dimensions may be applied within the context of a qualitative assessment of a process model as well; this view is adopted below.

The criteria for evaluation of an artifact by Prat et al. (2014) and the quality factors identified by Moody (2003) overlap in their aim and can be employed to assess the quality of the developed model (in a qualitative[164] and descriptive manner). Several of the listed criteria and factors are of relevance in the context of the current work. Due to their interrelatedness, these are combined for further discussion as shown in

Table 22 below.

Understandability, Simplicity, Clarity	the ease with which concepts and structures in the model can be understood; whether the model contains the minimum possible entities and relationships; the consistency of the model
<u>Efficacy</u>, Completeness, Flexibility	the degree to which the artifact produces the desired result; whether the model contains all information required to support the required and targeted functionality of the system; how easily the model can cope with change
<u>Utility</u>, Ease of Use, Implementability, Integration	the quality of the artifact in practical use; the ease with which the model can be implemented and integrated into existing infrastructure

TABLE 22 COMBINED EVALUATION CRITERIA

[163] Their factors differ only slightly from the two cited studies; where appropriate, a special comment is made in the discussion of the individual factors below.
[164] This, however, is not a *qualitative evaluation* as described e.g. by Kaplan and Maxwell (2005) whose aim is to induce knowledge by examining individuals and learning from them in interviews and by observation, from their interactions, feedback, perspectives or behavior typically exhibited in their natural environment.

As the models developed in this thesis are mainly on a conceptual level, not all of these criteria are equally applicable or of equal importance. The relevant criteria and to what extent these have been met by the proposed model are discussed below[165].

6.2.1 Understandability, Simplicity, Clarity

Especially in an inter-disciplinary endeavor such as the analysis of communication during crises and disasters by means of algorithms and models from computer science, it is of utmost importance to clarify the inventory and scope of the elements in play. The disciplines involved bring with themselves a variety of jargon and legacy vocabulary, referring to identical elements and operations in a different way or using the same term for different elements. A clarification of concepts and terminology is thus especially important in this setting and increases the understandability of the model. This is also echoed in Simsion and Witt's (2004) factor of *Communication* stressing the crucial importance of terminology and clarity when communicating with stakeholders. These might have different backgrounds and hence be accustomed to different terminology. Making terminology and concepts explicit in a clear and understandable way not only allows to establish common ground for discussions and inter-disciplinary development, but also increases stakeholders' options when talking about requirements of future systems and procurement.

An elementary model of communication, Lasswell's model (Lasswell, 1960) was chosen as the basis for the modelling of entities, relations and actions. Building upon this model, attributes and links between these entities were identified and a generic set of elements and interactions devised. This methodology effectively reduces the involved elements to the minimum needed to model the target domain of this thesis. The number of resulting elements and attributes can thus be kept to a minimum, reducing model complexity and ensuring logical consistency. Elements which appear to be different in nature at first sight – and which were given different names by their originating platforms – emerge to be identical elements with identical attributes. Based on them, operations for their processing can be identified and limited based on the set of elements. In a similar manner, the operations offered by the individual originating platforms can be reduced to a small set of generic ones.

One of the fundamental propositions underlying the current thesis is that many elements which appear to be different on the surface can be reduced to a few common elementary units. This is true of the entities involved as much as for the relationships between them. These relationships and links

[165] It should be noted that any model is by necessity a compromise of potentially competing factors and can only aim to provide the best balance between these factors (Simsion & Witt, 2004).

across the boundaries of individual platforms provide the basis for insights which could otherwise not be obtained. Elements and relations with properties, connected in a many-to-many-manner, with a particular focus on the connections lend themselves to a graph model (Kleppmann, 2017, p.49ff). Such an approach yields a simple, expressive and intuitive model, eliminating intermediate steps required by other approaches which are often the basis of semantic dissonance (Robinson et al., 2015).

The choice of method and creation of models may be viewed as aiming at a high level of understandability, simplicity and clarity. The explicit and clear conceptualization of the elements involved, of their structure and interactions, the connection to a standard model of communication science, the iterative approach combining top-down as well bottom-up elements follow a principled approach and yield a model which is superior to existing models, concise enough to model all facets of all elements involved and which can be reproduced and serve as the basis of further work.

6.2.2 Efficacy, Completeness, Flexibility

The models developed span all elements – objects as well as operations - as far as they have been identified to be of importance and of practical relevance in the scope of cross-media disaster communication. They comprise traditional as well as social media and map a set of diverse legacy elements to a small set of common, fundamental elements. All platforms involved have been examined and mapped to a common, generic model. Direct association of platforms with technologies allows to identify possibilities for processing as well as requirements of processing (when new platforms emerge) and provides for flexibility in a dynamically changing media landscape. Ease of addition of new platforms and straight-forward accommodation of new processing steps are a consequence of the chosen approach. The ability to scale with an increasing amount and complexity of communication elements provides a further benefit of the selected model, pointing out the factor of *scalability* – a criterion not explicitly mentioned by Moody (2003) or Prat et al. (2014) but aligning well with other criteria. The examples presented in chapter 6.3 below cover a variety of settings, indicating that the model and approach is indeed able to scale. During the work on this thesis, Periscope, a new type of platform allowing for interactivity in videos was used within the context of disasters for the first time (Fichet et al., 2016). This new platform could easily be accommodated within the proposed framework.

The inclusion of all relevant platforms, the ability and ease to accommodate a newly emerging platform as well as the degree to which all targeted platforms could be accommodated indicate that the generated artifact does indeed meet the goal of this thesis and produce the intended (desired) result. In particular, the integration of cross-media links within the developed model extends previous approaches and represents a novel and essential aspect. The overall model thus exhibits a strong level regarding the factors of completeness, flexibility and efficacy.

6.2.3 Utility, Ease of Use, Implementability , Integration

Several parts of the suggested model (data- as well as process model) have been implemented. This concerns foremost the phase of data ingestion (Medium Specific Access and Medium Specific Processing) as well as the enrichment phase of the workflow (Document Specific Processing and Communicator Specific Processing). These steps are outlined below and marked in green in Figure 62 below. The elements of the data model are constructed and processed according to the mechanisms described by the process model. Graph-construction has only been implemented partially; Graph-Extension has not been implemented yet and forms part of ongoing and future work.

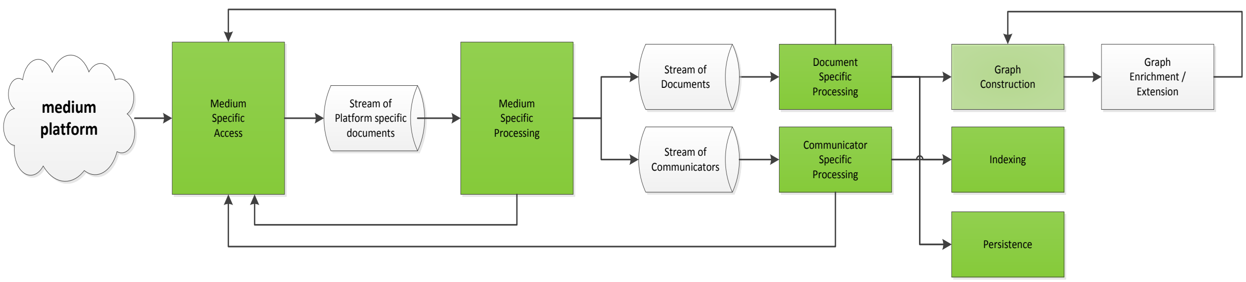


FIGURE 62 IMPLEMENTATION OF PROCESS MODEL

Following the scheme and models developed above, the following platforms have been accommodated in the proposed workflow:

- Satellite-TV
- Radio
- IP-radio
- News agencies
- RSS-Feeds
- Atom-Feeds
- Newsletters via Email
- Scanned documents
- Twitter
- Facebook
- VKontakte

- Flickr
- Instagram
- YouTube
- Telegram*
- WhatsApp*

These currently already integrated platforms serve as support for the implementability of the suggested model.

The two platforms marked with asterisks (*) above require access to specific groups or channels in order to be processed and thus do not conform to the definition of an *open-source*. However, they already play an important role in a variety of domains (e.g. for the role of WhatsApp during elections see Moreno and Best (2018)) and may also be of importance during disasters. The integration of such *newly-relevant* platforms into the framework is possible in an efficient manner and can be carried out in a short period of time. Both of the above platforms have indeed already been integrated; this provides a further argument for the implementability, generality and flexibility of the proposed models as well as the practicality of the chosen approach. Furthermore, the generic model allows to better estimate efforts and adjust planning for further work. Integration within existing infrastructure has taken place but existing resources (relational databases) still exist side-by-side with the newly developed model. Whether to completely integrate (replace) existing structures and to what extent this will be possible is an active area of development and has not been conclusively answered.

Regarding the overall functionality of the model, all elements and processing steps have been identified and accommodated in a flexible manner. The model allows for a number of steps which can be performed in parallel, which lend themselves to modular processing and containerization and which can take place in (near) real-time. In the face of massive amounts of data as typically witnessed during disasters, such flexibility is key for obtaining actionable insights in time for proper planning and response measures. Furthermore, elements may be duplicated to provide redundancy on a number of levels. These factors and the comprehensive set of elements and processing-steps result in a model providing greater utility than existing approaches. The criteria regarding the utility, ease of use, implementability and integration can thus be said to have been adequately met. The set of increasingly complex examples in chapter 6.3 below further demonstrates the approach's *scalability* using data from *real-world disasters* and supports this claim.

6.3 Examples of Application

To illustrate and support the above statements, the following chapter provides several actual examples of data encountered during natural disasters and outlines the steps from the Medium Specific Access to the Graph Construction phase. Following the proposition by Prat et al. (2014), the instantiation of the artifact developed within this thesis (the model) and its application to real-world examples may be considered an evaluation of the artifact: "*demonstrating the use of the IS artifact in one or several examples is a very common way of verifying if the artifact meets its goal*" (Prat el al., 2014). The examples themselves exhibit increasing levels of complexity; with this increase, the importance of the above discussed criteria of *Understandability*, *Simplicity* and *Clarity* also come to play a more important role. Regarding the criteria of *Completeness, Flexibility and Efficacy*, the examples demonstrate that is it indeed possible to accommodate a variety of settings within the developed model and that the model also scales with the complexity of the setup. Concerning the criteria of *Implementability, Utility, Efficacy, Ease of Use and Integration*, the examples provide evidence that the developed model can indeed be implements and put to work in real-world settings.

6.3.1 The Use of Topical Markers Across Platforms

As described above, hashtags are considered to be a Topical Marker and can act as a link between different Documents. They may occur in textual, visual or even in auditory form, e.g. when displayed on screen on TV or uttered by a speaker on radio.

In 2014, the UN Office for the Coordination of Humanitarian Affairs (UNOCHA[166]) issued a report on the suggested use of hashtags during disasters (UNOCHA, 2014). During the Ebola outbreak in West-Africa of the same year, the following hashtags (Figure 63) were suggested to be used[167]:

[166] https://www.unocha.org/

[167] https://irevolutions.org/2014/11/05/social-media-hashtag-standards-disaster-response/

FIGURE 63 UNOCHA – HASHTAGS FOR EBOLA[168]

This information was published by UNOCHA on a web-page in the form of an image as depicted in Figure 63 above, listing the suggested hashtags and their respective proposed use. To effectively detect and process this information, the content of Web-Pages needs to be ingested, segmented and processed, using text- and image-processing technologies and the hashtags within images be converted into text via OCR. These hashtags were subsequently used (and are still being used as of 2018) on Twitter, Flickr (see Figure 64 and Figure 65 below) and other platforms. The spread of such a Topical Marker across media-types and platforms provides an excellent example of how information is connected and can be used to propagate information across the boundaries of individual platforms.

FIGURE 64 EBOLA HASHTAGS ON TWITTER

[168] This image and the following images taken from the respective platforms (Twitter, Flickr & Facebook)

FIGURE 65 EBOLA HASHTAGS ON FLICKR

The detection of the hashtag *#EbolaResponse* on UNOCHA's web-site allows to create a Topical Marker, link this to the web-page (Document) of UNOCHA (Communicator) and, using the above examples, to further link it to the tweet (Document) of Doctors without Borders[169] (Communicator) and to the image of doctors in a lab above (Document) by user UNMEER[170] (Communicator).

The resulting network of Communicators, Documents and the Topical Marker is depicted in Figure 66 below (not all links and elements are shown for simplicity). Even though the original elements have different names, different attributes and originate on different platforms, the conversion into the generic elements of the model allows to relate them in direct manner, represent them as a graph and furthermore allows to infer further knowledge from this graph. For example, a link of (the Communicator) Doctor's-without-Borders and (the Communicator) UMEER can be established via their respective authorship of Documents referring (i.e. linking-to) an identical Topical Marker (in this case a hashtag), a connection which may prove crucial during operations and which may otherwise not be detected and go unnoticed.

[169] Doctors Without Borders, https://www.doctorswithoutborders.org
[170] UNMEER: the UN Mission for Ebola Emergency Response, https://ebolaresponse.un.org

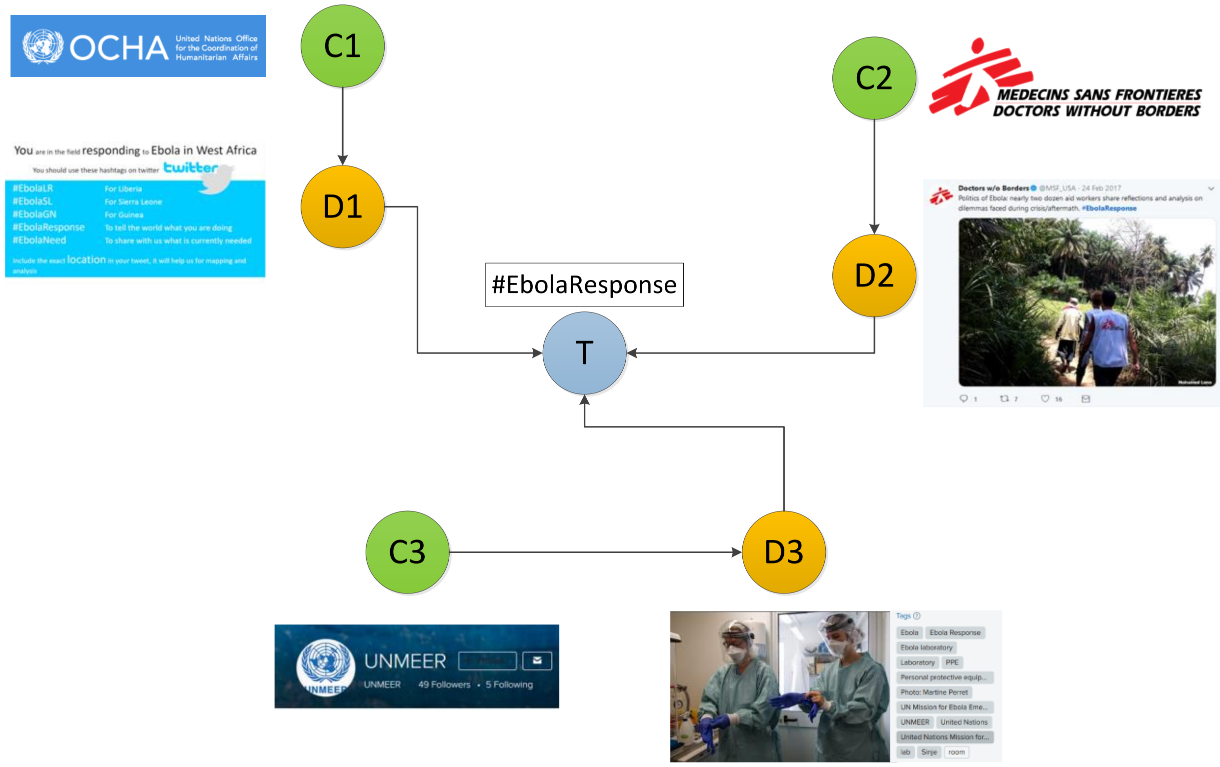


FIGURE 66 COMMUNICATORS, DOCUMENTS AND TOPICAL MARKERS FOR EBOLA

Detecting such links may not only be helpful in the orchestration of efforts but equally to further streamline and focus communication. Hashtags and Topical Markers in general are already used on a variety of platforms and thus form an important source for cross-media communication. Occasionally, the transition from one platform to another will be automated by bots such as in the case depicted in Figure 67 below. Systematic use and coordination of hashtags in such situations and across platforms provides an even richer environment for cross-media analysis.

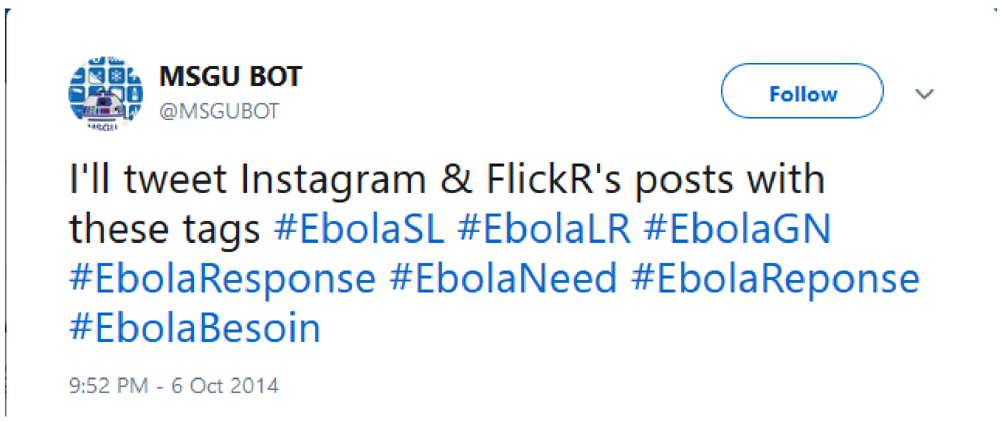


FIGURE 67 CROSS-PLATFORM COMMUNICATION BY BOTS

6.3.2 Links between TV and Social Media

Traditional media, such as TV stations and news-networks are typically also active on a variety of social media platforms, using these as additional and complementary distribution channels. In this setup, they increasingly also mix formats and add pointers to alternative channels. This is also the case during disasters as shown in Figure 68 below. News coverage of Storm Doris which hit the UK in February of 2017 also included a live report by Juliet Dunlop from Blackpool on ITV's Good Morning Britain[171], a screenshot of which is depicted below. In addition to transcribing what the reporter is saying on the ground, further information can be gleaned from the inserts such as the location of reporting or the kinds of damages reported or to be expected (this is visible by watching the video following the link in the footnote). In addition, the social media accounts are displayed, allowing to link them with the content of the particular day's newscast (and the program in general). Frequently, these might also be said by the reporter and corresponding hashtags might be displayed on screen, allowing to establish links between the TV station, the program, the reporter, the social media accounts as well as the content of the newscast. The link between the TV station and the respective account is typically constant, allowing to associate them for longer periods of time.

FIGURE 68 ON-THE-SCENE REPORTING DURING STORM DORIS

[171] https://www.thescottishsun.co.uk/news/629228/storm-doris-batters-good-morning-britain-reporter-juliet-dunlop-live-on-tv-as-raging-80mph-smash-into-uk

6.3.3 Automatic Speech Recognition and Entities

The following text provides snippets from an automatically generated transcript (applying ASR to the audio track of the newscast)[172] of the introduction to the above report and the ensuing live report:

***Britain** is bracing itself for storm Doris … widespread travel disruption … could be power cuts and damage to buildings…. issued several amber warnings … Juliet Dunlop is in Blackpool this morning where it looks very windy … … gusts up to 80 mph … we are certainly in the path of Doris … up to 30cm of snow in Scotland … meaning that travel could be especially difficult this morning … flights have been cancelled … delays on the railways and on the roads …*

In the above text, various entities have been highlighted which can be detected within the transcript by a Named Entity component. They comprise locations (Britain, Scotland, Blackpool), persons (Juliet Dunlop), types of natural disasters (storm), of infrastructural-damage (power cuts, damage to buildings, delays, flight and railway delays). Additional facts such that *flights* have been *cancelled* can also be elicited.

Due to the scheduled recording, it is known that the program is called *Good Morning Britain* and was aired on ITV on February 23, 2017 at 7:01AM local time. Segmentation of the newscast allows to create a topically narrow and coherent segment dealing with the approaching storm Doris.

The resulting network of Communicators (ITV, Juliet Dunlop), Documents (Good Morning Britain, Feb 23, 2017, Segment 07:01-07:05), Entities (Britain, Blackpool, Power Cut, Flight Cancellation, Infrastructure Damage, Storm Doris) is depicted below in Figure 69 (not all connections are displayed for better visibility). The Entity “Infrastructure Damage” is not mentioned explicitly in the newscast but may be inferred from static background information relating the specific kinds of disruption to the more general concept via an ontology (an example of how further linking may be applied).

[172] Image and transcript taken from https://www.youtube.com/watch?v=nqUhFhN86iU

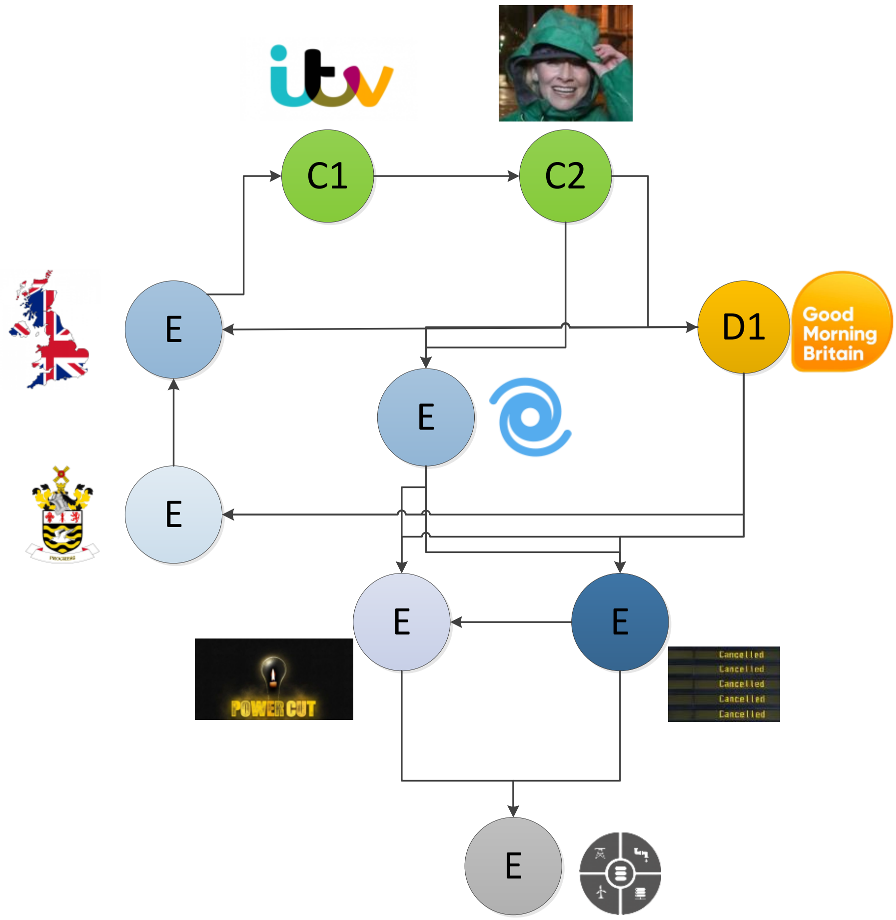


FIGURE 69 NETWORK OF COMMUNICATORS, DOCUMENTS AND ENTITIES

This resulting structure uses the elements of the generic data model. The created graph-structure allows for queries such as "Are any natural disasters causing infrastructural damage in the UK?" and yield the precise footage from ITV. Documents from other platforms, such as Tweets related to Storm Doris will link to the same node and can be used for further context (e.g. to obtain images from the ground). As processing is envisioned to be continuous and in real-time, this allows to provide further insights and to improve situational awareness even while the event is still taking place.

6.3.4 Location and Maps from Multimedia

In the context of (natural) disasters, the detection and recognition of maps on images and within videos is of high interest. Frequently, maps are shown in connection with further details mostly in textual form. The examples below were published after an earthquake in Alaska in 2016[173].

[173] https://www.accuweather.com/en/weather-news/breaking-magnitude71-earthquak/54982626

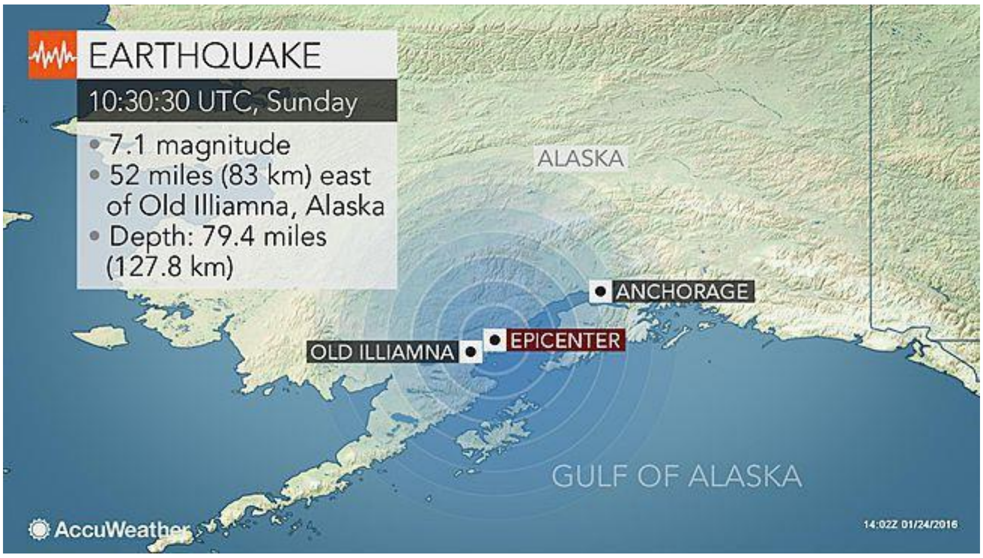


FIGURE 70 EARTHQUAKE IN ALASKA

In the above image, the type of natural disaster (Earthquake), its location (Old Illiamna, Anchorage, Alaska, Gulf of Alaska) as well as further attributes, such as the magnitude and depth can be detected and processed. The disaster-type as well as the various locations are represented as Entities by the framework. Properties, such as the magnitude or exact location can be extracted from images as the one above by text-detection, segmentation and OCR technologies.

6.3.5 Multilingual, multi-platform and multimedia Communication Following an Earthquake and Tsunami

On the 25th of September 2018, at approximately 18:30h local time (10:30h UTC) a tsunami struck the city of Palu in Sulawesi's Donggala province following an earlier earthquake. Waves of several meters' height entered the city, destroying large parts of Palu. Soil liquefaction resulting from the quake caused further destruction inland (a cascade of disasters).

The following documents and actors are taken from data collected after the event. The example is merely meant to demonstrate the possibility to model the different elements – it is by no means an attempt to thoroughly reconstruct the sequence or complexity of events which unfolded following the earthquake and tsunami! One Tweet, two posts on Facebook, a report on a web-page and a TV program all reporting on the event are combined to demonstrate the representation of Communicators and Documents in a common graph.

Natural disasters, especially earthquakes are observed by a variety of systems and organizations world-wide. The USGS is one of the most prominent actors, detecting earthquakes not only in the US but also on a global scale. The tweet below was issued by user @everyearthquake, an account *Tweeting every earthquake occurrence reported by USGS* according to its description. It references the USGS, the event of an earthquake and the location and time of the reported quake.

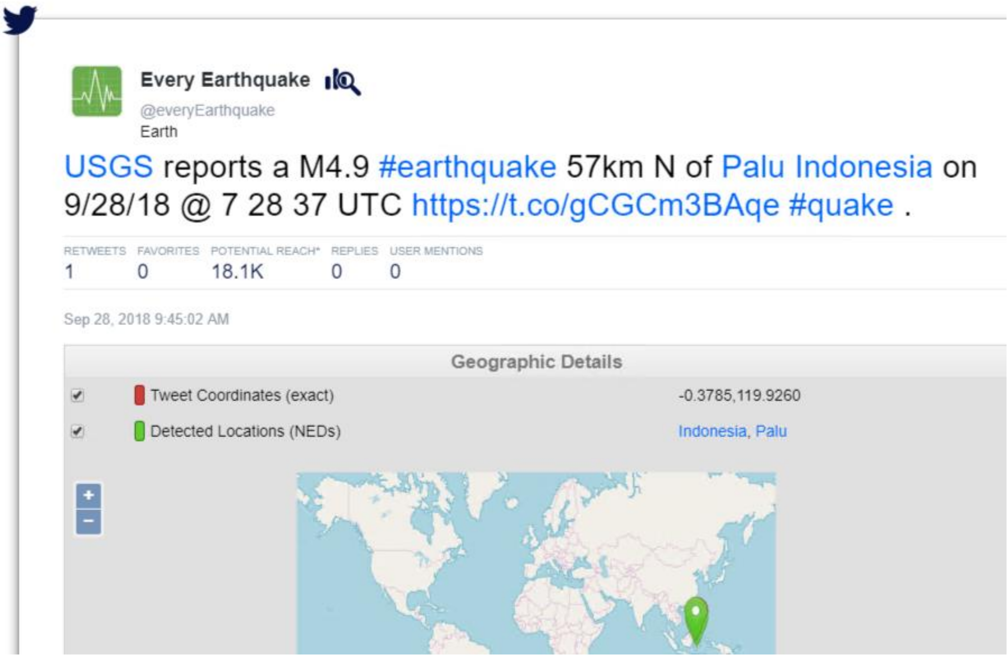
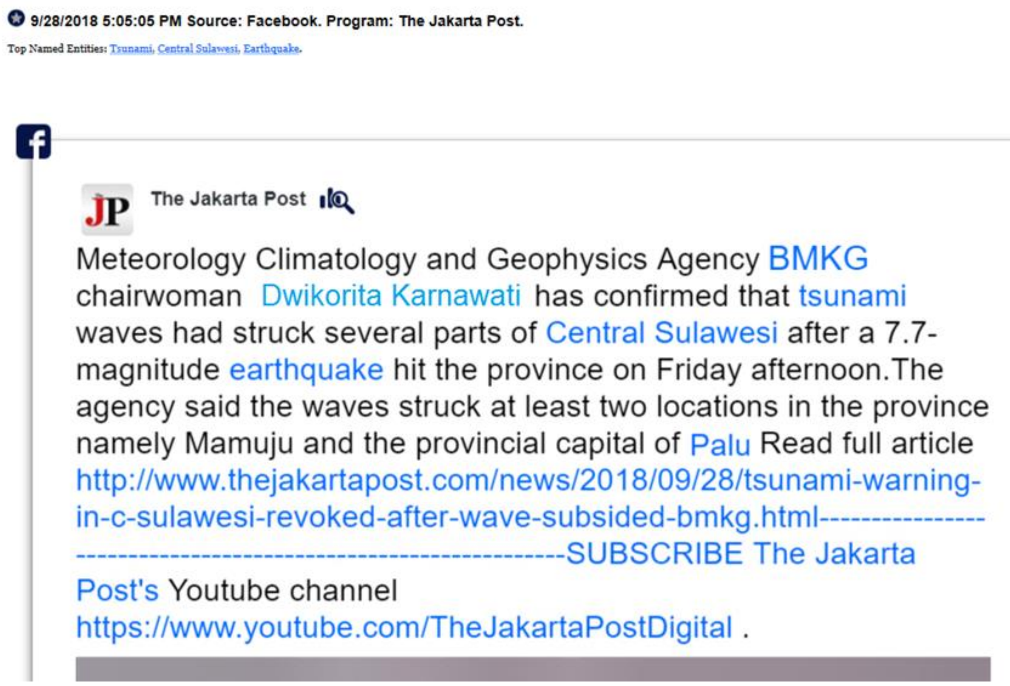


FIGURE 71 TWEET ABOUT EARTHQUAKE IN INDONESIA

The Jakarta Post published (posted) on its public Facebook page a confirmation of the tsunami hitting Palu (Figure 72). It mentions the Indonesian Meteorology, Climatology and Geophysics Agency, BMKG, its chairwoman Dwikorita Karnawati, the region of Central Sulawesi and several locations, among them Palu. Furthermore, it includes a link to the Jakarta Posts YouTube channel.


9/28/2018 5:05:05 PM Source: Facebook. Program: The Jakarta Post.

Top Named Entities: Tsunami, Central Sulawesi, Earthquake.

The Jakarta Post

Meteorology Climatology and Geophysics Agency BMKG chairwoman Dwikorita Karnawati has confirmed that tsunami waves had struck several parts of Central Sulawesi after a 7.7-magnitude earthquake hit the province on Friday afternoon.The agency said the waves struck at least two locations in the province namely Mamuju and the provincial capital of Palu Read full article http://www.thejakartapost.com/news/2018/09/28/tsunami-warning-in-c-sulawesi-revoked-after-wave-subsided-bmkg.html--SUBSCRIBE The Jakarta Post's Youtube channel https://www.youtube.com/TheJakartaPostDigital .

FIGURE 72 FACEBOOK POST ABOUT EARTHQUAKE AND TSUNAMI IN INDONESIA

The German Red Cross (Deutsches Rotes Kreuz) also confirmed the incident, posting on its Facebook page that at least 400 casualties had been reported and aid was underway (Figure 73). This document serves as a *placeholder* of the many different languages used in covering the incident. The event of Tsunami is mentioned with the affected island of Sulawesi and the country of Indonesia.


Deutsches Rotes Kreuz

Mindestens 400 Menschen sind nach dem Tsunami auf Sulawesi in Indonesien ums Leben gekommen Es wird mit noch viel mehr Opfern gerechnet Der Einsatz läuft.Hier mehr Infos https://www.drk.de/presse/pressemitteilungen/meldung/tsunami-trifft-indonesien-nach-schwerem-erdbeben/PALANG MERAH INDONESIA | International Federation of Red Cross and Red Crescent Societies | #Sulawesi | #Tsunami .

FIGURE 73 FACEBOOK POST BY GERMAN RED CROSS

Figure 74 below displays the resulting graph after processing the three Documents above. Medium Specific Access has retrieved the original documents and their meta-data, Medium Specific Processing has turned them into a sequence of Documents and Communicators. Document- and Communicator

Specific Processing have enriched the contents and meta-data and created nodes and edges yielding the graph-structure below. Entities of type organization are colored in blue, disaster-types in purple, locations in light blue and persons in light-green. Nodes in the graph have been connected with statically linked nodes, such as the German Red Cross (DRK) being located in Germany or the USGS being located in the US. The resulting graph-structure allows a variety of complex queries such as "Which organizations have been reporting about the tsunami in Palu?" or "Which organizations from Germany have been communicating about the earthquake and tsunami?"

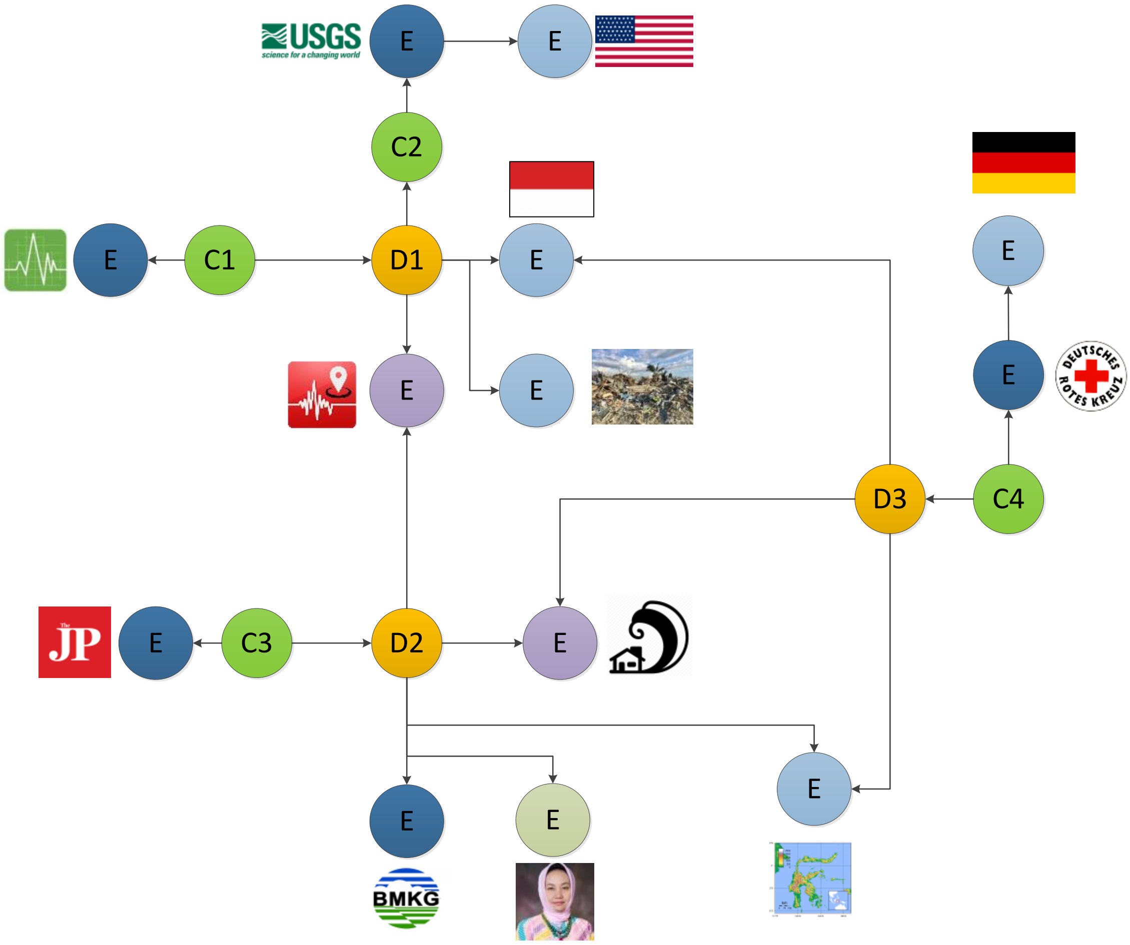


FIGURE 74 NETWORK FOR INDONESIA EARTHQUAKE AND TSUNAMI

The following document is a web-page from the Chinese Global Times[174] (in English), reporting about the reconstruction after the efforts following the earthquake (Figure 75).

HOME >> CHINA,WORLD

Chinese efforts underway in reconstruction of quake-hit Indonesian city

By Global Times Source:Global Times - Xinhua
Published: 2018/10/11 20:23:41

The earthquake and tsunami disaster zone is littered with debris in Palu, Indonesia, on October 9, 2018. Photo: VCG

Under a tall signal tower, dozens of tents have been set up. They are "offices" of Chinese telecom company ZTE in Palu, the capital of Indonesia's Central Sulawesi province, hit by devastating quakes and tsunami.

Rindhu, one of ZTE's local staff in Indonesia, told Xinhua on Wednesday that ZTE's emergency team arrived in Palu immediately after the strong earthquakes struck last month to assist local authorities in the recovery efforts.

FIGURE 75 REPORT ON RECONSTRUCTION EFFORTS

Figure 76 below displays the extracted relevant text after the Document Specific Processing (for contents of a Web-Page, image taken from the SAIL LABS Media Mining System[175] into which part of the work of this thesis has been integrated).

[174] http://www.globaltimes.cn
[175] https://www.sail-labs.com/media-mining-system

Mixed

Chinese efforts underway in reconstruction of quake-hit Indonesian city

Links | PDF | Web |

Properties

10/13/2018 12:19:48 PM Source: Global Times China. Program: News.
Link: http://www.globaltimes.cn/content/1122599.shtml
Publishing date: 10/13/2018 12:19:48 PM
Topics: General Language.
Top Named Entities: Indonesia, Palu, Earthquake, People's Republic of China, Tsunami.

The earthquake and tsunami disaster zone is littered with debris in Palu Indonesia on October 92018.

Under a tall signal tower dozens of tents have been set up.

They are offices of Chinese telecom company ZTE in Palu the capital of Indonesia's Central Sulawesi province hit by devastating quakes and tsunami.

Rindhu one of ZTE's local staff in Indonesia told Xinhua on Wednesday that ZTE's emergency team arrived in Palu immediately after the strong earthquakes struck last month to assist local authorities in the recovery efforts.

Devastating earthquakes and an ensuing tsunami ravaged the central Indonesian province on September 28leaving more than 2000 dead and thousands of people homeless so far with Palu worst hit.

After the disaster the Chinese government social organizations and enterprises offered help to the affected people in Indonesia and participated in the reconstruction efforts in the disaster-hit areas.

Thanks to the effective cooperation between the Indonesian government its people and the international communities including China the daily life in Palu is returning to normalcy with debris being cleaned up more supplies delivered to the needy and banks and supermarkets reopened.

Rindhu 34is the only woman in ZTE's temporary working site in Palu as well as the head of the network maintenance team.

Her colleagues see her as a brave woman.

Rindhu was asked to travel to Palu to repair the damaged telecom equipment network after the disasters.

Flying from Jakarta to Massakar the capital of South Sulawesi province before taking a bus for two days from Massakar to Palu Rindhu finally arrived in Palu on October 3.

Figure 76 Processed and enriched text from web-page

The final two documents below show an Indonesian Web-Page from the newspaper *Republika Online*[176] (Figure 77 and Figure 78) and a screen-shot of an Indonesian newscast of *Astro Awani*[177] taken from a YouTube channel (Figure 79). Both demonstrate the importance of supporting local languages as both are in Bahasa Indonesia and refer to the earthquake (*gempa bumi*) and Central Sulawesi (*Sulawesi Tengah*). Due to the multi-lingual processing, these can be associated with the same Entities as the elements in other languages (English and German) in the documents above. This in turn allows for immediate cross-lingual analysis, linking elements which originally appear in different languages. Especially in large-scale disasters of global importance involving organizations from around the globe, such multilingual processing and linking of information is of utmost relevance.

[176] https://www.republika.co.id
[177] http://www.astroawani.com

FIGURE 77 REPUBLIKA ONLINE WEB-PAGE

Properties

9/28/2018 12:20:02 PM Source: Republika. Program: News.
Link: https://nasional.republika.co.id/berita/nasional/daerah/18/09/28/pfrgpe384-donggala-diguncang-gempa-6-sr
Publishing date: 9/28/2018 12:20:02 PM
Topics: Daerah.
Top Named Entities: Earthquake, Central Sulawesi, Palu.

REPUBLIKA.CO.ID JAKARTA Badan Meteorologi Klimatologi dan Geofisika (BMKG) mencatat wilayah Kabupaten Donggala Sulawesi Tengah diguncang gempa bumi tektonik berkekuatan 6 SR Jumat (28/9).

Beruntung gempa tidak berpotensi tsunami.

Kepala Pusat Gempabumi dan Tsunami BMKG Rahmat Triyono mengatakan hasil analisis BMKG menunjukkan informasi awal gempabumi ini disebut 5,9 SR yang selanjutnya dilakukan pemutakhiran menjadi 6,0 SR.

Episenter gempabumi terletak pada koordinat 0,40 LS dan 119,81 BT atau tepatnya berlokasi di darat pada jarak dua kilometer arah utara Kota Donggala.

Pusat gempa berada pada kedalaman 10 kilometer.

Dengan memperhatikan lokasi episenter dan kedalaman hiposenter gempa bumi ini tampak bahwa gempa bumi dangkal ini terjadi akibat aktivitas di zona sesar Palu Koro.

Hasil analisis mekanisme sumber menunjukkan bahwa gempa bumi ini dibangkitkan oleh deformasi batuan dengan mekanisme jenis sesar.

FIGURE 78 PROCESSED AND ENRICHED TEXT FROM REPUBLIKA ONLINE

FIGURE 79 YOUTUBE VIDEO FROM ASTRO AWANI CHANNEL

Adding these documents – two web-pages as well as (the transcript of) one video on YouTube results in the following extended graph (Figure 80, not all edges are shown in the graph).

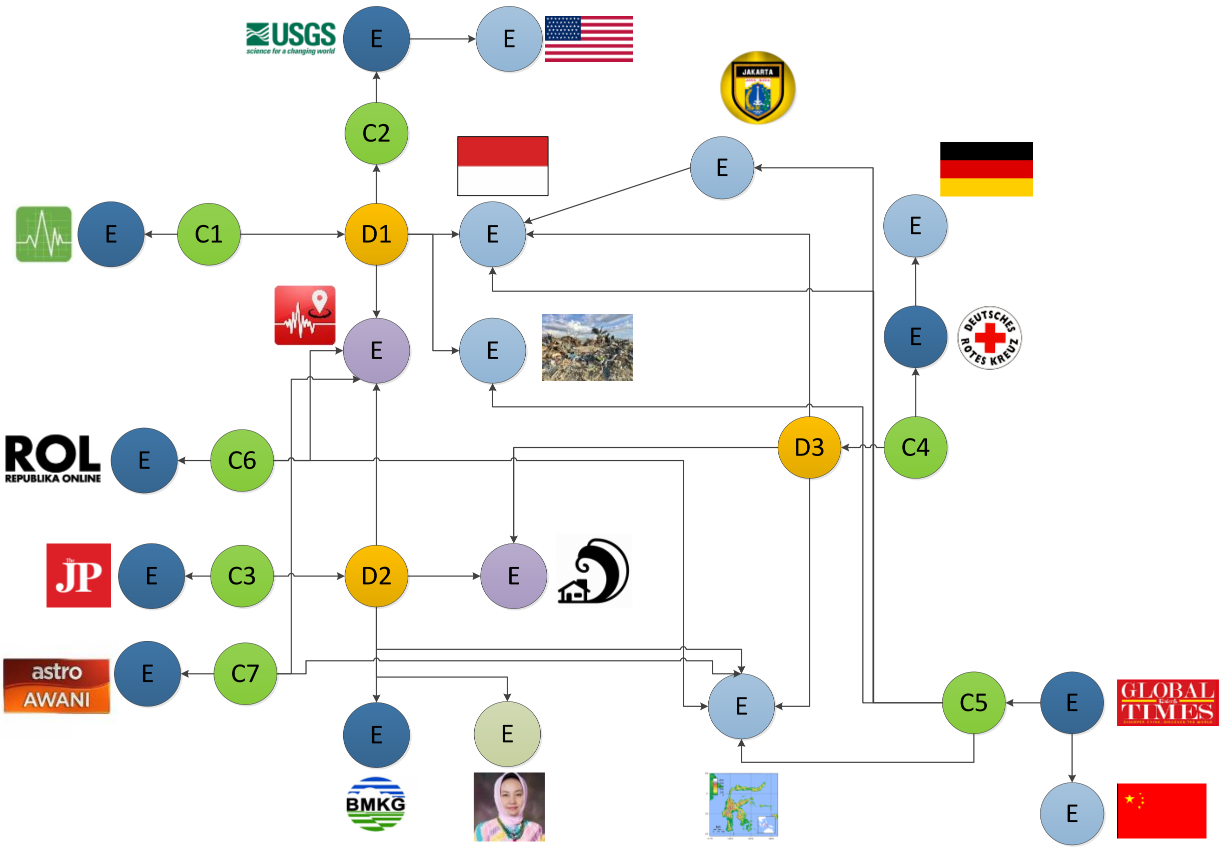


FIGURE 80 COMMUNICATORS, DOCUMENTS AND ENTITIES FOR INDONESIA EARTHQUAKE AND TSUNAMI

A multitude of queries is possible on this small section of a theoretically much bigger graph. In order to measure the propagation of information by state-run outlets, e.g. the Indonesian News Sources (Communicators) talking about the disaster but not mentioning the National Agency for Meteorology could be determined. Using such information, adjustments may be applied to the communication strategy. Furthermore, different ways of contrasting information – e.g. Chinese sources talking about reconstruction when national sources are still focusing on search and rescue – are possible. The ways to issue queries for search and the resulting complexity of the graph structure need to be considered carefully in order to avoid cognitive overload of the persons interpreting their results. But clearly, a wide variety and powerful queries can be answered based on the structure created following the proposed approach of this thesis.

7 Conclusions and Outlook

7.1 Introduction - Discussion of Results

The aim of the research performed in this thesis has been to develop a framework - a generic model and a method to create it - for cross-media communication during disasters in a multitude of media, languages and platforms. No such model existed prior to the current work, as confirmed during earlier work of the author and initial research for this thesis[178]. Against the backdrop of an increase in the frequency and severity of natural disasters and a continuous extension and diversification of media and technologies used to communicate about them, the development of a model generic enough to represent the wide variety of data-sources and documents typically encountered during such incidents has been the focus of this thesis. Investigating what these generic elements and their attributes are and how the individual media and platforms can be represented by them, a method to create such a generic model was devised and the actual model developed accordingly. This model allows to combine the different kinds of elements in a *cross-media approach* and to apply the resulting framework for analysis of communication during disasters, leading to improved situational awareness. It furthermore provides a reference against which emerging technologies and tools can be compared and which can inform planning and management activities.

The model is based on a standard model of communication theory and extends it by adding the concept of *links*, different types of connections between the individual elements. These links, some of them crossing the boundaries of individual media (*cross-media links*), provide additional insights not obtainable without this extension. A process model, defining operations to create the elements and operating on the elements of the data model themselves has been developed together with the data model. The work performed in this thesis could build upon earlier work of the author performed during a research-project dealing with disaster communication in Austria during the floods of 2013. Insights gained within that project, findings in the corpus of collected data and several publications[179] provided input to the current thesis, which was thus not developed in a *vacuum*, but rather connected to real-world events and requirements from actual end-users. The insights gained can further serve as lessons-learned and guidance for first-responders and underscore the practical relevance of the current work.

[178] See chapter 2.2 (Research Gaps)

[179] Authored and co-authored by the author of this thesis, see chapter 2.6 (Own Publications)

An overall approach guided by Design Science Research (DSR) was adopted for this thesis, which is reflected in the iterative nature of the approach for model-development and exemplary evaluation. In the sense of DSR, an *innovative technological artifact* (the framework/model) has been created as well as an *instantiation* representing an improvement to existing approaches and solving a real-world problem of relevance to an application domain and its practitioners (Hevner & Chatterjee, 2010). Chapter 6 above provides a set of insights evaluating different aspects of model development and the model itself. This chapter concludes this research effort by revisiting the aims and steps taken to achieve them. The research questions are revisited in the light of the results of the current work. The different kinds of contributions and their implications are discussed. Finally, limitations of the current work and alleys for future research are presented.

7.2 Overview of the Research

The first section of this thesis comprises chapters 1-4. The motivation and approach are presented, followed by chapters examining the background of disaster communication. This comprises a literature review, determining the state-of-the-art and reviewing insights of projects and earlier work regarding the various aspects and roles that media play in the context of natural disasters. These form a challenge which is highly relevant and of increasing importance to a broad spectrum of stake-holders, from affected citizens to first-responders, authorities and society as a whole. Even in a *media world* where social media make the creation and sharing of information a ubiquitous and non-stop activity, traditional media still maintain their firm place as valuable, (often) credible and sometimes the only available sources of information during disasters. The conclusion is drawn, that the development of any generic model aiming to cover the full spectrum of media communication during disasters thus needs to consider a wide range of traditional and social media. Current research generally does not provide such a broad approach, focusing largely on individual platforms, languages or technologies. These and further factors are discussed in the section on gaps in current research as detailed in chapter 2. The current thesis aims to fill these gaps, by creation of a framework – a data and process model paired with a method on how to create these models – reflecting the complete spectrum of media employed.

The second section of this thesis comprises chapters 5 and 6, dealing with the actual development of the data- and process model, their exemplary application to real-world data and the assessment of the results and contributions. First, a method to develop these models is introduced. Subsequently, the elements and their properties are derived in an iterative scheme, accommodating all relevant platforms and actors. Generic aspects are extracted from the particular instances, refined and aggregated into the common models (elements and their attributes). Technologies required to process the resulting structures are identified in an iterative manner, allowing to align the elements and their processing. Special attention is paid to the various kinds of linking between these elements – reflecting the cross-media approach propagated throughout this thesis. A set of examples are provided demonstrating the practical viability of the proposed approach and the resulting insights are discussed from a quantitative as well as qualitative point of view. Chapter 6 recapitulates and concludes the work on this thesis.

7.3 Research Questions Revisited – Answers and Insights

The **primary research-question** of this thesis was formulated in chapter 2.4:

How can a model be built which represents the elements and processes involved in cross-media disaster communication, such that all relevant components and technologies found within the set of relevant platforms and media and their interplay can be represented in a generic and extensible manner, in a way which overcomes current limitations and satisfies the identified needs of end-users of the target domain?

To address this research question, several **sub-research questions** were formulated at the outset of the current thesis (in chapter 2.4). These sub-questions cover the different aspects of the primary RQ and will now be revisited in the light of the work carried out within this thesis. The primary research question is answered by the combination of the answers to these sub-research-questions.

RQ1: What should a framework for cross-media disaster communication consist of and which elements should it contain?

To delineate and guide the work within this thesis, the notion of *framework* has been interpreted from a point of view similar to the one defined by Meise (2001) there referred to as “*Ordnungsrahmen*” (German for framework). This approach resulted in a generic framework defining and structuring the elements identified as relevant in the domain of cross-media disaster communication on a high level of abstraction, providing an overview and allowing for the classification of these elements and their interrelationships. The framework has been shown to be general and flexible enough to accommodate existing and even newly emerging elements. As an essential component, a standard model from communication science was selected (Lasswell, 1960), providing the basis for a common terminology and allowing to identify and map the building blocks for model development (i.e. the *elements*). Motivated by a data-centered perspective from computer science (Simsion & Witt, 2004), the development of the overall model was divided into the development of a data model, representing all relevant elements of cross-media communication, and into the development of a process model, representing the operations possible and required[180] on these elements. An iterative procedure, following the principles of Design Science Research (DSR) (Hevner et al., 2004) was devised and applied to derive a uniform representation and a common set of attributes. Both models were developed hand-in-hand and following this common methodology. The resulting model forms an artifact in the sense of DSR.

The following elements can consequently be seen as essential ingredients of a framework for cross-media disaster communication:

- a common terminology, to be able to reference all elements involved and link the present work to communication science
- a data model, representing the common set of elements and generic enough to be able to represent all platforms, media and actors
- a process model, representing a common set of operations, accounting for the creation and processing steps necessary for all elements involved
- a method to derive the above models, employing the common terminology and allowing to create the models in an iterative manner

In a simplistic manner, the following statement may be used to describe the necessary components of the overall framework:

> Framework = terminology + model (data+process) + method to create models
>
> (all ingredients need to be generic and flexible enough to extend to and to cover the wide range of media, platforms and actors).

RQ2: What are the qualities and features of such a generic framework and how should it be constructed?

It was found that communication during disasters takes place on traditional media as well as on social media, depending on a variety of geographic and socio-economic factors. Even though social media play an important role, traditional media must not be ignored or excluded when aiming to create a

[180] *Required* in the sense of what operations are required to be able to process the element and construct the respective generic element and relations from it.

comprehensive overview of events - both kinds of media merit attention and both contribute to situational awareness. Communication about news and disasters has turned into a two-way affair, with citizen actively taking part in the production of material. This content includes first hand, eye-witness accounts from the scene of disaster allowing to gain rapid insights into a yet developing situation. The variety of platforms made use of and the ubiquity of portable devices has created a media-landscape containing information in a multitude of languages and with an increasing amount of multimedia data. Consequently, technologies going beyond the mere processing of text are required to tap into the growing amount of audio and visual content. The sharing and referencing of information effectively create a network of documents and actors, related in a variety of ways. These phenomena do not only exist on social media but have likewise been adopted on traditional media. Representing these networks allows to gain additional insights into the flow and distribution of information, allows to identify commonalities and central elements in the production and distribution of information. The dynamic landscape of media, in particular of social media, requires that a generic framework be flexible enough to account for newly emerging platforms in a short amount of time. The set of platforms encountered in disaster communication has been examined within this thesis and an iterative procedure applied to create a data and process model, covering the complete set of platforms. A standard model from communication science was chosen to provide the basis upon which to create the generic model for disaster-communication. Using this model as a guideline, the generic elements and their associated properties were distilled and aggregated. A graph structure was determined to provide the most adequate fit to further model these elements and their interconnections. Parts of this model have been implemented; several examples demonstrate the practical viability of the chosen approach.

RQ3: Which platforms and media should be accommodated in the framework and how?

Even though a lot of weight has been given to social media, traditional media are a long way from being *dead* or useless in times of disaster. On the contrary, frequently the public only learns about a disaster from TM, especially when the disaster is taking place in a remote area not of immediate interest to this public. Press-agencies act as intermediaries and information brokers, researching and verifying information[181] before feeding their reporting to TM as well as providing information on SM channels. Depending on the location, the local level of freedom of press, socio-economic factors or the media landscape of a region, one or the other type of media may be able to supply more relevant and timely information. Social media may spark coverage and provide first insights at the time of

[181] When working according to professional journalistic standards - which may not be the case for all agencies.

impact of a disaster. Traditional media, with sound and professional research may shed further insights and provide credible facts and a deeper understanding of a situation. The different media-types complement each other and should be used in combination. Consequently, any generic framework needs to account for this spectrum of media and include traditional as well as social media. In a recent paper, the relevant social media were investigated by Reuter et al. (2018a[182]). Insights from this paper were used in the selection of platforms within the current thesis. The iterative approach applied to accommodate the different platforms resulted in a generic set of elements, properties and relations representing all relevant platforms and actors as well as a set of technologies allowing to perform the processing required to construct and instantiate the model. The approach is deemed flexible enough to also allow to accommodate further platforms.

RQ4: Which components and functionalities are required and how are they to be combined?

The advent of social media, portable devices and cell-phones with high resolution cameras has brought with it a deluge of audio-visual material. This is also true in times of disaster where images may provide more detailed and deeper insights than a purely textual description made by a non-expert at the scene of disaster. As a consequence, in addition to a series of text-processing technologies, audio and visual technologies thus also need to be accommodated in a generic framework. The set of relevant textual, audio and visual technologies has been determined by the study of collected data-sets, literature review about disasters and interaction with practitioners. Certain technologies require the application of prior processing, creating a logical sequence of processing steps and dependencies between components. The availability of models and technologies can be aligned with the types of data available to determine the possibilities and options for disaster management. Furthermore, the set of technologies can be used for procurement purposes and to detect gaps in the technology mix when designing disaster management systems.

RQ5: Which types of cross-media communication exist and how can they be characterized?

Following the iterative procedure outlined in chapters 5.5.1 and 5.6.1 the set of possible connections (links) between the various elements of the communication model have been determined and a set of technologies identified which is required to process data and detect those links. The links themselves can be classified as *structural*, *interactive* and *referential* and can be characterized by a set of attributes as detailed in chapter 5.5.7. Links can be extracted from platform-provided data (within

[182] The author of this thesis is a co-author

the *Medium Specific Processing*) as well as from the Documents and Communicators created by subsequent processing. These links form the connections between the elements extracted and play a central role in creating a graph-structure from these elements. Further inference can be applied to the graph created this way, resulting in further (often indirect) links providing additional insights.

The two **secondary research questions** - which are of particular practical relevance – formulated above in chapter 2.4 are addressed below:

SRQ1: Which phases in the disaster-cycle can be covered by the capabilities offered by such a framework?

Findings of earlier projects as well as by investigation of several international crises indicate that communication about disasters takes place during all phases of a disaster. The intensity of communication depends on a variety of factors and may fluctuate depending on media coverage and perceived relevance as discussed in chapter 3.2. The quality and types of information and communication may change according to the disaster phase, e.g. warnings and advice for caution may dominate during the onset of a disaster whereas calls for donations may surface and prevail at later stages (Olteanu et al, 2015). The types and quantity of links (e.g. from social media to traditional news media) may likewise change during the course of a disaster. However, during all phases of the disaster cycle, relevant and new information may appear leading to improved situational awareness for first responders. Information gathering cannot assume any staged time-distribution over the different types of media but rather should address all types of media – traditional and social – simultaneously and in parallel, i.e. in a phase-agnostic manner (Backfried et al., 2016).

SRQ2: Are the patterns static or do they change dynamically during the different phases of a disaster?

As stated above, the intensity, patterns and contents of communication may change during the different phases of a disaster. Links between media can also be expected to shift over time (Backfried et al., 2016). A prudent approach has to assume that many different media are active and should be monitored throughout an incident. Observing certain shifts may indicate a change in the situation, the sentiment reflected by the affected population or yield newly emerging facts. These may require further assessment as they could be valid and relevant but also simply spark rumors and spread incorrect information going viral if not attended to and countered immediately (as experienced during QuOIMA, see chapter 4).

7.4 Contributions

According to Hevner et al. (2004), *The ultimate assessment for any research is, "What are the new and interesting contributions?"*. This chapter summarizes the contributions made by the current thesis from a theoretical/scientific as well as a practical point of view. The following types of contributions were laid out in chapter 2.3 above and are repeated here for reference.

- Contributions to theory/science: Framework and Model
- Contributions to practice: Lessons Learned and Guidelines

These are discussed in the following chapters below.

7.4.1 Theoretical and Scientific Contributions

The theoretical contributions of this thesis concern the creation of a framework consisting of a terminology, a data- and process model and a method to create these models. These models are able to accommodate the wide spectrum of platforms employed during disaster communication - the complete set of platforms and technologies identified to be of relevance can indeed also be represented within the model. The framework was developed in a principled way and linked to a standard communication model from communication theory, which it extends to account for a prevalent phenomenon in media coverage of disasters: the linking and mashing-up of documents and actors. Due to the inherent inter-disciplinarity, the terminology was considered first and a common scheme to refer to all elements proposed. This may also contribute to improve communication across the borders of the disciplines involved. The model is motivated by real-world data and end-user requirements and has been validated by partial implementation. Although a substantial number of research studies has been carried out on how social media are used within the scope of disasters (Olteanu et al., 2015; Castillo, 2016; Reuter & Kaufhold, 2018; Reuter et al., 2018a), very little work has been dedicated to how different kinds of social media are combined and no work had previously been devoted to how traditional and social media can be combined under such circumstances. Furthermore, no generic model had been devised trying to capture all elements and platforms involved. These very aspects have been investigated in this thesis and a model developed which is able to accommodate all of the different elements involved. The construction of this model shows that it is indeed possible to unify all relevant elements into one model which allows a more comprehensive understanding of the nature of communication during disasters on traditional and social media. Moreover, the focus on the various types of links which exist between these elements allows to gain insights which could not be obtained by examining the individual elements only. This approach thus extends our understanding of how media are employed during disasters and outlines how this understanding can be used to increase situational awareness. This claim is confirmed by the experiences made in the QuOIMA project and the use of the implemented components in the setting of intelligence gathering (Backfried et al., 2016).

The following list summarizes the theoretical contributions:

- Systematic investigation of social and traditional media involved in disaster communication.
- Framework (i.e. terminology, data and process model and method to create them) comprising both traditional and social media and method to construct these models. This forms a generalization of prior work, from a constrained, individual platform-based view to a holistic and generic view. Furthermore, a connection to a standard model from communication science was established.
- Systematic investigation on linking between multiple platforms for disaster communication. Links within platforms and especially between platforms had not been examined thoroughly before this thesis but are known to provide valuable insights during disasters. These links have been integrated into the framework.
- Graph-based approach as the natural and appropriate fit for the variety of elements involved.

7.4.2 Practical Contributions

The practical contributions of this thesis concern the insights gained by systematically investigating and extending the results of a previous research project. Abstracting from the insights gained in that project, the generic framework developed within this thesis allows to understand, based on theoretical assumptions, how media interact and can be linked during disasters, which types of linking occur, whether linking takes place in a disaster-phase dependent manner and how links can be detected and processed. The thesis thus contributes to an enhanced understanding of the mechanisms of how different types of media interact during disasters which can in turn influence guidelines for future system design, training, procurement and deployment. Implementation of the proposed model should ultimately help to obtain an extended view and to improve situational awareness during disasters.

The following list summarizes the practical contributions:

- Practical insights for first responders and practitioners, who will benefit directly and indirectly from the development of the framework. Implementation and use of the features allow to gain improved situational-awareness and support sense-making. The model and properties of the elements can be employed to inform requirements analysis and procurement and serve for educational purposes or as lessons-learned. Furthermore, it may be used to gauge expectation management in terms of what is possible, what could be relevant or which constraints exist.
- Exemplary implementation to demonstrate the practical viability of the approach by implementing parts of it and applying this to disasters. This demonstrates the viability of the approach which is expected to be beneficial for commercial efforts.
- Demonstration of the actual use of social media in Austria during disasters in terms of platform use and linking of information.
- Awareness for crisis managers/first responders as to the presence and importance of cross media links and that these form a crucial source of information during disasters.
- Insights for first responders (the Austrian Red Cross and the Austrian Ministry of Defence) through partial implementation.
- Ability of the framework to quickly place and categorize newly emerging platforms. This is especially important for an area as dynamic as SM.
- Awareness of terminological issues due to its inter-disciplinary approach, combining a model from communication theory with methods from computer-science. This may help in establishing a common terminology or at least help raise awareness about terminology and foster inter-disciplinary cooperation.
- Broader view how different platforms interact, instead of focusing even more on a single platform, thus broadening the basis for further research for this typical interaction of platforms and media encountered in today's communication patterns.

The practical relevance of results has furthermore resulted in the writing of several research proposals (H2020) providing support for the accepted applicability of the generated ideas and models.

7.4.3 Putting the Contributions into Perspective

Apart from the theoretical and practical contributions above, some more general questions and issues arise which are discussed within this chapter. Whetten (1989) proposes several important aspects to be taken into account to assess theoretical (research) contributions. A set of questions derived from these will be used below to put the current contributions into a more general and broader context.

What is new? Why has this not been done before?

No single model to cover the different media employed during disaster communication had been developed before this thesis. Furthermore, the connections between the elements of communication in such a model have not been dealt with. This may be due to a host of technical (access to data, technologies, language processing models, ...) and organizational reasons (location of researchers, funding bodies and research schemes). The theoretical contributions of the current thesis thus form novel contributions to the field (of information systems for crisis management).

Why now?

The current thesis has its roots in the QuOIMA project, dating back to 2012 and the Central European Floods of that year. The relevance of social media in disasters had been acknowledged before. However, this issue had not been investigated in an Austrian context, nor had the more general question of how social and traditional media interact in such cases been studied. This project – the infrastructure to collect data, the coincidence of a large-scale natural disaster during the runtime of the project and the many insights gained from interaction with first responders thus provided an excellent base for the present thesis. The topic of media communication during disasters – which are becoming more severe and more frequent, having profound impacts on society and the economy – is of utmost importance, now more than ever.

Who cares? Who is interested in this topic? Is the topic of contemporary interest to scholars and practitioners in this field?

The topic of communication during disasters, and how to obtain relevant information for the affected population, first responders and official institutions presents an established field of research (ISCRAM – Information Systems for Crisis Response and Management). Many first responders have devised strategies for the inclusion of social media into their disaster management activities (American Red Cross, 2010; UNOCHA, 2017) and rely on established practices for traditional media communication. Clearly, a combination for the analysis of social media and traditional media would be advantageous. Creating the framework of this thesis not only provides a model for how to include the complete spectrum of sources and actors, but also allows first responders to benefit from the practical aspects, forming part of the contributions. Furthermore, for large-scale disasters and disasters of wide-geographic scope, the accommodation of platforms and languages used in geographies and media-landscapes unlike those of typical (US or European) research projects might be beneficial to quickly be able to deploy systems.

7.5 Limitations and Suggestions for Further Research

The limitations of this this thesis are discussed from the viewpoint of their potential impact on the research questions (objectives). Some of them form alleys for future work and will be presented as such below. Furthermore, several directions for possible future research are outlined.

Although the present thesis provides a framework encompassing a wide spectrum of media, it is certainly biased from the research and work environment of the author. For example, specifically Chinese platforms (used only/predominantly in China) have not been taken into account which may exhibit certain characteristics not fully covered by the framework. Furthermore, the processing of content in Chinese may require additional technologies not presented here in detail. Some language dependencies and blind spots regarding platforms and technologies may result from this bias. Regarding the integrated technologies, the author's background in NLP and speech recognition (rather than visual processing) may have led to an emphasis on these technologies. The visual processing technologies considered represent the ones encountered during literature review and practical work; however, additional specialized technologies may not yet be fully represented and need to be added by future work. Regarding the spectrum of sources, all relevant ones have been included into the framework and the set of platforms has remained stable over the course of this thesis. This is viewed as an indicator of the generality of the chosen approach. However, *change* is an ongoing process and further sources can be expected to emerge sooner or later. An extension of the framework for such sources might be required.

The focus on open sources may pose a limitation, as closed sources may yield further important information increasing situational awareness for first responders. In particular in regions with limited freedom of press and controlled SM networks, these may provide deeper and otherwise unaccessible information. The extent to which such platforms are integrated, however, is not only determined by technical factors, but even more so by privacy concerns, ethical considerations and legal requirements. Certain sources (like WhatsApp and Telegram) have indeed been accommodated into the framework, showing that this is possible in principle. Proprietary data-sets and sources (which might be available in first-responders' IT infrastructures), however, may need to be pre-processed and only then fit into the model.

Two opposing trends - on the one hand platform providers increasingly *motivating* users to use proprietary apps instead of general web-interfaces and trying to gain access to ever increasing amounts of personal data along the way - on the other hand increased awareness of privacy and activities to regain ownership of one's data like the Solid Project[183] headed by Tim Berners-Lee at MIT - will have profound implications as to how the access to data will be managed and possible in the coming years. This is expected to have strong effects also on the way that data can be gathered and processed during disasters and on the way that links between documents and actors can be utilized. In combination with changes in the legal realm like the GDPR and (European) copyright laws[184], the impact is going to be fundamental.

The quantitative aspects of evaluation of the model are mainly limited to data from one disaster, which is mostly due to no further corpus being available which would contain the richness of data needed to perform such an investigation. Other disaster corpora and literature have been consulted and taken into account for qualitative evaluation. However, a different data set would be a beneficial cross-check to confirm the quantitative findings. The author hopes that this weakness is somewhat countered by the insights from the partial implementation and an ongoing use of the resulting model in an active system.

Even though an exemplary implementation has been performed, only a portion of the suggested complete framework has been implemented as a property-graph. This leaves issues regarding the eventual scalability of the proposed mechanism unanswered to some extent. A measurement performed recently for the different kinds of data entering into a live system (accessed by the Austrian Red Cross and Austrian Military) yielded the following numbers (per second):

- 20 tweets
- 1.96 Internet pages
- 1 FB posts/comments
- 0.86 YouTube comments
- 0.4 VK posts/comments
- 0.06 TV transcripts of radio and TV broadcasts
- In addition, approximately 20 hours of social media videos and almost 300 hours of TV and radio content are processed continuously per day

[183] https://solid.mit.edu
[184] WIRED, https://www.wired.com/story/europes-copyright-law-could-change-the-web on 2019/01/11

This dataset provides a realistic example of the amount of data ingested and processed by a system corresponding to the suggested framework. Continuous processing and network creation under these conditions can still be performed in real-time. Access to the resulting structures for retrieval results in response times of typically no more than a few seconds. However, no formal evaluation has been carried out. The performance, however, is also determined by the underlying technology stack and needs to be investigated in detail to arrive at a sound conclusion.

A graph-structure has been identified and devised as the appropriate model to represent the manifold elements and links within the proposed framework. Whereas it is apparent that this structure allows for a rich variety of queries, it is unclear as to what point this power can and will be employed by end-users when using the system. A balance between the power of querying and the cognitive load incurred by such queries will need to be found and ways to transparently manage queries and visualize query-results will need to be devised in the future. The developed framework and the corresponding infrastructure can serve as an environment to investigate these effects and determine the optimal operating conditions and parameters, in particular for stressful conditions such as during disasters.

The resulting corpus, spanning different media and sources is linked with a multitude of intellectual property and copyright issues. Thus, it is not possible to make this corpus available for further research (outside of the team of the involved partners).

Suggestions for further Research

The resulting model accommodates a set of generic elements which is assumed to be constant for the foreseeable future. However, the actual platforms used during disasters are likely to change over time and depending on preferences of the affected populations and first responders. It would thus be beneficial to use the model as a starting point for the integration of such further platforms and – should this be required – to extend the model to accommodate such platforms.

Concerning the set of technologies, the availability of very-large datasets paired with progress on machine learning (ML) algorithms (deep learning, transfer learning, convolutional neural networks, long short-term memory (LSTM) networks, …) and massively parallel hardware (in form of high-performance GPUs) has generated a boost in applicable, ML-based technologies for classification and prediction. In particular, classification – e.g. related to the relevance or veracity of information – could be a beneficial addition to the model as well as create valuable information for end-users in a running system. However, such models should not be regarded as static but rather will need to change and adapt continuously over time, suggesting an integration into the workflow with regular updates based on recent data[185].

Due to the availability of additional data sets and progress in transfer learning, it may also be possible to add processing for languages and dialects not considered during the work on the current model. These languages and dialects may require additional processing steps which will then need to be accommodated in the processing scheme. However, support for such languages may be crucial to deploy a resulting system in diverse areas of this planet hit by disaster.

The model assumes streaming, continuous processing of data and subsequent creation of structures. These structures, however, must be assumed to shift and change over time. Data as well as relations may become obsolete - at some point (and ultimately this might even be motivated by limits of scalability) shrinking/pruning of the graph will become a topic of interest. How to best apply such mechanisms, both in logical as well as in technical terms, provides ample opportunity for further research.

The incremental and continuous construction of the suggested graph-model, strategies to combine growth with 'forgetting' (i.e. pruning of sub structures) in combination with the underlying technology stack (especially the graph-database) poses several challenges and provides a wide area for further research. To this end, the author is actively participating in the further development of this structure within a commercial system.

The trend towards service-oriented-architectures and with it the topics of containerization and micro-services can be envisaged to have implications on the flexibility, extensibility and scalability of the proposed model. Whereas a certain (logical) linearity is inherent in the proposed process model (Figure 49), it is possible to perform several of the steps by loose-coupling and parallelization of processing. This concerns the processing of Documents and Communicators but likewise the enrichment activities on the resulting Property Graph. The extent and granularity of such extensions remains to be explored and well-balanced against an increase of coordination.

From the point of view of the end-user, the possibilities to query this graph and use it for OLAP and visualization purposes presents a field for further research. In particular, the balance between technical possibilities and potential cognitive overload of users provide ample opportunities for further work.

[185] The author believes that such continuously learning ML components provide a promising addition to the suggested framework. A scheme developed to semi-automatically update the vocabulary and language model for ASR has been developed by the author and is already applied in practice on a daily basis.

Whereas the model is focused on communication during natural disasters, many aspects generalize to other types of disasters – e.g. human-induced disasters – as well as to areas beyond disasters, where the same assumptions about mixing and mashing-up of information may hold, such as certain kinds of business intelligence.

7.6 Summary and Outlook

The research presented in this thesis has approached the topic of disaster communication from a novel angle. A framework consisting of a generic model and a method to create this model has been developed and its practical viability been assessed using real-world data. The framework draws on knowledge from a variety of fields and comprises a wide spectrum of media, all of which have been shown to be of importance during disasters and all of which can be accommodated within the proposed model. It has been developed in a principled manner, guided by ideas from design science research and extends prior approaches by the combination of traditional and social media and a special focus on cross-media communication. As such, it extends prior knowledge and contributes to our understanding of how media are used in connection during disasters. The research was motivated by an increase of natural disasters, with tremendous effects on the affected populations and society and thus addresses a highly relevant topic. This immediate relevance and practical insights also provide input for guidelines, training and procurement on the part of the intended users, first-responders and authorities. Even though this thesis has been written from a computer-science perspective, the research-domain is an inter-disciplinary field situated at the cross-roads of several disciplines such as computer-science, linguistics or communication science. A clear terminology and clarity of the developed concepts thus plays an important role and is hoped to foster future cooperation. Several possible directions for future research have been proposed. They include work on the integration of additional platforms and technologies as well as on the proposed graph-structure and ensuing topics of application such as investigation of the cognitive load on users when working with such structures. During the work on this thesis, technical advances as well as ethical and legal changes have taken place, some of which influence the assumptions and statements made, without invalidating them (see *Changes During the Writing of This Thesis* below). Some of the work investigated by this thesis has already been implemented and deployed in an actual system, triggering yet new thoughts and ideas. Some ideas have found their way into research projects and proposals for further research. Thus, the work performed here is hoped not to be an end by itself, but rather also the beginning and basis of further work!

Changes During the Writing of This Thesis

The current thesis has been written in the course of several years. Whereas some of the areas concerned remained relatively constant during this time, others exhibited a rapid growth and dynamic change. Even though the media landscape itself and the way how media are consumed have changed - the Digital News Report[186] of the Reuters Institute for the Study of Journalism of the University of Oxford provides an excellent overview of these changes over time – the types of data and contents which can be retrieved remained relatively constant. Notable changes concern a trend towards pay-walls and premium content having an impact on open access to information from such platforms. Tailored formats taking into account personalization and the increased use of mobile platforms for news consumption provide further changes. In the realm of social media, a trend towards more multi-media content could be observed paired with raised concerns about data-reuse and privacy concerns, not least caused by scandals such as Cambridge Analytica in connection with Facebook. This also led to a trend towards closed platforms or communication within closed groups on messaging platforms. The current thesis only considers such closed communication to a limited amount and does not aim to generally integrate communication which is meant to be private. Some of these changes have caused initial ideas for this thesis to be revised or extended as work on it proceeded. New approaches or platforms (such as Periscope) were considered and integrated into this thesis. The following list provides a short summery on two of the most salient areas of change touching on the research questions of this thesis.

Technical Advances

- An increase of multimedia data, images and videos with higher resolution: this results in largely increased amounts of visual information also being available during disasters and a wider application of visual processing technologies for detection and classification of the content of these images and videos. The quality of cameras and the resulting images and videos increases continuously and so do the requirement of space to store and processing power to process them. This is expected to place higher emphasis on the visual processing technologies and components in the future.

[186] http://www.digitalnewsreport.org

- Altered expectations on natural language processing caused by the emergence of personal assistants such as Amazon's Alexa, Microsoft's Cortana, Google's Assistant, or Apple's Siri: not only will these assistants make interaction with computers more natural, they will also bring with them increased expectations as to what is possible, e.g. communicating via voice also under distress during an emergency. Regarding the lessons-learned and insights for first responders, advances in these technologies need to be monitored closely and a clear understanding of the possibilities and risks obtained. Maturity and robustness paired with a profound understanding will be required to separate *technological gimmicks* from practical tools, ready for deployment in real-world settings.

- Advances in graph-databases, in particular in the area of scalability and integration with streaming data: these changes had a strong influence on this thesis and the models developed due to their natural fit as well as their enabling some of the processing structures which were previously not considered to be practical in the particular setting.

- Advances in ML and data-sets for training them have had an impact on the technologies available to process the spectrum of media targeted by this thesis: this is true for text, audio and visual types of content with audio and visual content benefitting more than textual content due to large advances in the processing of speech and images. Transfer of models and the possibility to retrain models and adapt them dynamically over time in connection with more powerful and cheaper hardware (GPUs) allows practical deployment of powerful technologies. ML and data-science are expected to have an impact also on models used within the proposed framework and within the scope of analysis of disaster communication. The presence of adequate training data, sound evaluation methodologies and novel skills on part of technical personnel present challenges to be met and considered for strategic planning.

Ethical and Legal Changes

In an era of increased polarization of society and particularly during the period following the US election campaign of 2016, propaganda and manipulation (including by outside powers) have led to increased concerns and awareness about the veracity of news, the automation and manipulation of communication via bots and the *fake news* phenomenon. The Cambridge Analytica scandal has led to deepened concerns about data re-use and privacy while the introduction of the GDPR in Europe changed the legal parameters for the use of and access to data. Platforms are reacting, introducing lengthy review procedures and narrowing platform access through a set of evolving mechanisms. These frameworks change irregularly, causing confusion and puzzlement among the developers using the corresponding APIs and introducing uncertainty on the future availability of content. In addition, platforms proceed in removing more and more (meta-) data via API access, limiting the use of the resulting content profoundly. On the one hand, much of these effects underscore the multimedia, multi-platform and multilingual emphasis of this work – in a polarized and slanted environment, in an environment where any platform may suddenly change, limit or even cancel access to content, it may become even more important to draw upon a wide array of sources and media and not be locked-in by a singular platform. On the other hand, the issue of veracity and trust and how these should be incorporated in the developed model and processes poses new questions beyond the current work.

Abbreviations

ASR	Automatic Speech Recognition also often referred to as TTS, Text-to-Speech
DSR	Design Science Research
IE	Information Extraction
IR	Information Retrieval
IS	Information Science
MMS	Media Mining System
NED	Named Entity Detection
NER	Named Entity Recognition
NLP	Natural Language Processing
OCR	Optical Character Recognition
OSINT	Open Source Intelligence
PGM	Property Graph Model
SID	Speaker Identification
SM	Social Media
TD	Topic Detection
TM	Traditional Media
UGC	User-Generated Content

References

Abdulhussain S.H., Ramli A.R., Saripan M.I., Mahmmod B.M., Al-Haddad S.A.R., Jassim W.A., (2018) Methods and Challenges in Shot Boundary Detection: A Review, Entropy, 20 (4) , art. no. 214

Acar, A, and Muraki, Y., (2011) Twitter for crisis communication: lessons learned from Japan's tsunami disaster, Int. Journal Web Based Communities, Vol. 7, No. 3, 2011

Akella, M.K., (2009) First responders and crisis map symbols: Clarifying communication. Cartographic and Geographic Information Science 36(1): 19-28, 2009

Al-Dahash, H F, Thayaparan, M and Kulatunga, U, (2016) Understanding the Terminologies: Disaster, Crisis and Emergency. In: P W Chan and CJ Neilson (Eds) Proceedings of the 32nd Annual ARCOM Conference, 5-7 September 2016, Manchester, UK

Alejandro, J., (2010) Journalism in the Age of Social Media, Reuters Institute Fellowship Paper, University of Oxford, available at https://reutersinstitute.politics.ox.ac.uk/sites/default/files/research/files/Journalism%2520in%2520the%2520Age%2520of%2520Social%2520Media.pdf on 2018/10/21

Allan, J., Carbonell, J., Doddington, G., Yamron, J., and Yang, Y., (1998) Topic detection and tracking pilot study: Final report. In *Proceedings of the DARPA Broadcast News Transcription and Understanding Workshop*, pages 194–218

Allemang, D., Hendler, J. (2011) Semantic Web for the Working Ontologist, 2nd edition, Elsevier

American Red Cross, (2010), The Case for Integrating Crisis Response with Social Media, White Paper

Ausserhoder, J., Kittenberger, A., Maireder, A., (2012) Twitterpolitik – Netzwerke und Themen der politischen Twittersphäre in Österreich, Studie an der Universität Wien

Backfried G., Aniola, D., Mak, K., Pilles, H.C., Quirchmayr, G., Winiwarter, W., Roth, P. (2010), Multimedia Documentation Lab, 10th International Conference on Knowledge Management and Knowledge Technologies, Graz, Austria, September 2010

Backfried G, Schmidt, Ch., Pfeiffer, M., Quirchmayr, G., Glanzer, M., Rainer, K., Open Source Intelligence in Disaster Management , EISIC 2012, European Intelligence and Security Informatics Conference, Odense, Denmark, August 2012

Backfried, G., Göllner, J., Quirchmayr, G., Rainer, K., Kienast, G., Thallinger, G., Schmidt, Ch., Pfeiffer, M., Meurers, Ch., Peer, A., (2013a) Cross-Media Analysis for Communication during Natural Disasters, IAIT 2013, International Conference on Advances in Information Technology, December 2013, Bangkok, Thailand

Backfried, G., Göllner, J., Quirchmayr, G., Rainer, K., Kienast, G., Thallinger, G., Schmidt, Ch., Peer, A., (2013b) Integration of Media Sources for Situation Analysis in the Different Phases of Disaster Management, EISIC 2013, Uppsala, Sweden

Backfried, G, Prinz, K., Göllner, J., Quirchmayr, G., Czech, G., (2014) Cross-media, Multimedia, Multilingual Communication in Open Sources During Crises and Disasters, ISCRAM 2014, University Park , PA, USA

Backfried, G., Schmidt, Ch. and Quirchmayr, G., (2015) Cross-Media Linking in Times of Disaster. In Proceedings of the Information Systems for Crisis Response and Management (ISCRAM). Kristiansand, Norway

Backfried, G., Schmidt, Ch., Aniola, D., Meurers, Ch., Mak, K., Göllner J., Peer, A., Quirchmayr, G., Czech, G. and Glanzer, M., (2016) A General Framework for Using Social and Traditional Media During Natural Disasters: QuOIMA and the Central European Floods of 2013, in Rogova, G. and Scott, P (Eds), Fusion Methodologies in Crisis Management, Springer

Bailer, W, Kienast, G., Thallinger, G and Backfried, G, (2016) Data Fusion Across Traditional and Social Media in Fusion Methodologies in Crisis Management, Rogova G and Scott, P, Eds, Springer

Balke, W-T., (2012) Introduction to Information Extraction: Basic Notions and Current Trends, Datenbank Spektrum (2012) 12:81-88, Springer

Bara, G., Backfried, G., Aniola, D., (2019) Fake or Fact? Theoretical and Practical Aspects of Fake News, in Information Quality in Information Fusion and Decision Making, Bossé, Éloi, Rogova, Galina (Eds.), Springer

Blei, D.M., Ng, A.Y., Jordan, M.I., (2003) Latent Dirichlet Allocation, Journal of Machine Learning Research 3, p993-1022

Bontcheva, K., Cunningham, H., Tablan, V., Maynard, D., and Saggion, H., (2002) Developing Reusable and Robust Language Processing Components for Information Systems using GATE. In Proceedings of the 3rd International Workshop on Natural Language and Information Systems NLIS'2002), IEEE Computer Society Press

Broder A.Z. (2000) Identifying and Filtering Near-Duplicate Documents. In: Giancarlo R., Sankoff D. (eds) Combinatorial Pattern Matching. CPM 2000. Lecture Notes in Computer Science, vol 1848. Springer, Berlin, Heidelberg

Bruns, A., Burgess, J., (2012) Local and Global Responses to Disaster: #eqnz and the Christ-church Earthquake, 2012 Disaster and Emergency Management Conference, Brisbane, Australia

Bruns, A., Burgess, J., Crawford, K. and Shaw, F., (2012), #qldfloods and @QPSMedia: Crisis Communication on Twitter in the 2011 South East Queensland Floods, Brisbane: ARC Centre of Excellence for Creative Industries and Innovation

Bucher, T., Helmond, A., (2017) The Affordances of Social Media Platforms, in The SAGE Handbook of Social Media, edited by Jean Burgess, Thomas Poell, and Alice Marwick. London and New York: SAGE Publications Ltd.

Castillo, C., (2016) Big Crisis Data, Social Media in Disasters and Time-Critical Situations, Cambridge University Press

Cheng JW, Mitomo H, Otsuka T, Jeon SY., (2016) Cultivation effects of mass and social media on perceptions and behavioural intentions in post-disaster recovery–The case of the 2011 Great

East Japan Earthquake, Telematics Informatics. 2016;33(3):753-72

Chu, Z., Gianvecchio, S., Wang, H., Jajoda, S., (2012) Detecting automation of Twitter accounts: are you a human, Bot or Cyborg?, IEEE Transactions of Dependable and Secure Computing 9(6):811-824

Cisco, White-Paper "The Zettabyte Era: Trend and Analysis", (2017) https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/vni-hyperconnectivity-wp.html on 2018/05/05

Cobley, P, Schulz, P.J. (Ed. and Introduction), (2013) Theories and Models of Communication, de Gruyter Mouton, Berlin, Germany

Cockburn, P., (2011) Catastrophe on camera: Why media coverage of natural disasters is flawed, The Independent, available at https://www.independent.co.uk/news/media/tv-radio/catastrophe-on-camera-why-media-coverage-of-natural-disasters-is-flawed-2189032.html on 2018/11/10

Coombs, T.W., Holladay, S., (2012) The Handbook of Crisis Communication, Wiley

Cunningham, H., (2002) GATE, a General Architecture for Text Engineering, Computers and Humanities, 36: 223-254, Kluwer Academic Publishers

Davidson, D., (2010) Cross-Media Communications: An Introduction to the Art of Creating Integrated Media Experiences, available at http://bit.ly/1X356eG on 2016/05/28

Del Fabro, M., Böszörmenyi, L. (2013) State-of-the-art and future challenges in video scene detection: a survey, Multimedia Systems, 2013, 19:427-454, Springer

Dewey, M., (2011), Dewey classification and relative index, 23rd ed. Joan S.Mitchell, Julianne Beall, Rebecca Green, Giles Martin and Michael Panzer Eds, Dublin, OH

DHS Report, (2013) Lessons Learned: Social Media and Hurricane Sandy, available at https://communities.firstresponder.gov/DHS_VSMWG_Lessons_Learned_Social_Media_and_Hurricane_Sandy_Formatted_June_2013_FINAL.pdf, 03/26/2018

Die Presse, (2015) Krisenmanagement mit Facebook, https://diepresse.com/home/science/4634937/Krisenmanagement-mit-Facebook on 2018/11/06

DKKV, Deutsches Komitee für Katastrophenvorsorg e.V. (DKKV), (2000) Naturkatastrophen und die Medien, Herausforderungen an die öffentliche Risiko- und Krisenkommunikation, Dokumentation des IDNDR-Expertenworkshops vom 3.-4. Dezember 1998 in Königswinter

Dugdale, J., Van de Walle, B., Koeppinghoff, C., (2012) Social Media and SMS in the Haiti Earthquake, SWDM 2012 Workshop, Lyon, France

Earle, P., Guy, M., Buckmaster, R., Ostrum, Ch., Horvath, S., Vaughan, A., (2010) OMG Earthquake! Can Twitter Improve Earthquake Response?, Seismological Research Letters (2010) 81 (2): 246-251

Earle, P., Bowden, D.C., Guy, M., (2011) Twitter earthquake detection: earthquake monitoring in a social world, Annals of Geophysics, 54, 6, 2011; doi: 10.4401/ag-5364

Endsley, M., Bolte, B., Jones, D., (2003) Designing for Situational Awareness: an Approach to User-centered Design. Taylor Francis Group, Boca Raton

Faustino, J.D., Liu, B., Jin, Y., (2012) Social Media During Disasters: A Review of the Knowledge Base and Gaps, Final Report to Human Factors/Behavioral Sciences Division, U.S. Department of Homeland Security, College Park, USA

Ferrucci, D., Lally, A., (2004) UIMA: an architectural approach to unstructured information processing in the corporate research environment. Natural Language Engineering , 10:327–348

Fichet, E., Robinson, J., Dailey, D, Starbird, K., (2016) Eyes on the Ground: Emerging Practices in Periscope Use during Crisis Events, ISCRAM 2016

Fiscus, Jg; & Doddington, Gr., (2002) Topic Detection and Tracking Evaluation Overview. Topic Detection and Tracking, pp. 17–31., https://doi.org/10.1007/978-1-4615-0933-2_2

Finkelstein, A., Kramer J., Hales M., (1992) Process modelling: A critical analysis in Integrated Software Engineering with Reuse: Management and Techniques, UK, London: Chapman & Hall, and UNICOM, pp. 137-148

Gosain, A., Heena, M., (2015) Literature Review of Data Model Quality metrics of Data Warehouse, Procedia Computer Science 48 (2015) 236-243, Elsevier

Greenberg, J., Scanlon, T.J., (2016) Old Media, New Media, and the Complex Story of Disasters, Oxford Research Encyclopedia of Natural Hazard Science, July 2016

Grishman, R.; Sundheim, B., (1996) Message Understanding Conference-6. Proceedings of the 16th Conference on Computational Linguistics , vol. 1, pp. 466, https://doi.org/10.3115/992628.992709

Haddow, G., and Haddow, K., (2014) Disaster Communications in a Changing World, 2nd edition, Elsevier

Harrington, S., Highfield, T. Bruns, A., (2012) More Than a Backchannel: Twitter and Television, 2012 COST Action ISO906 Transforming Audiences, Transforming Societies

Hemerly, J, (2011) Chapter 8, Classification in Intellectual Foundations for Information Organization and Information Retrieval, http://people.ischool.berkeley.edu/~glushko/IFIOIR/ on 2018/03/28

Hollnagel, E., Woods, D.D., (2005) Joint Cognitive Systems: Foundations of Cognitive Systems Engineering. Boca Raton, FL: Taylor & Francis. ISBN 0-8493-2821-7

Hevner, A., March, S., Park, J., Ram, S., (2004) Design Science in Information Systems Research, IS Quarterly Vol. 28 No. 1, pp. 75-105/March 2004

Hevner, A., Chatterjee, S., (2010) Design Research in Information Science, Theory and Practice, Springer

Holmes, W., (2011) Crisis Communication and Social Media: Advantages, Disadvantages and Best Practices, Univ. of Tennessee, CCISymposium 2011, Knoxville, USA

Imran, M.; Castillo, C.; Lucas, J.; Meier, P.; & Rogstadius, J. (2014) Coordinating Human and Machine Intelligence to Classify Microblog Communications in Crises. In S. R. Hiltz, M. S. Pfaff, L. Plotnick, & A. C. Robinson (Eds.), Proceedings of the 11th International ISCRAM Conference (pp. 712–721). University Park, Pennsylvania, USA

Imran, M., Castillo, C., Diaz, F. and Vieweg, S., (2015) Processing Social Media Messages in Mass Emergency: A Survey. ACM Computing Surveys, vol. 47, no. 4, pp. 1–38., https://doi.org/10.1145/2771588

Imran, M., Alam, F., Ofli, F., & Aupetit, M., (2017) Enabling rapid disaster response using artificial intelligence and social media, International Roundtable on the Impact of Extreme Natural Events: Science and Technology for Mitigation, pp. 80-81, 13th-15th December, Colombo, Sri Lanka

Jain, P., Rodrigues, T., Magno, G., Kumaraguru, P., & Almeida, V.A., (2011) Cross-Pollination of Information in Online Social Media: A Case Study on Popular Social Networks. 2011 IEEE Third Int'l Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third Int'l Conference on Social Computing, 477-482.

Jacob, E.K., (2004) Classification and Categorization: A Difference that Makes a Difference, Library Trends 52(3) Winter 2004: 515-540

Johansson F., Brynielsson, J., Narganes Quijano M, (2012) Estimating Citizen Alertness in crises using SM monitoring and analysis, EISIC 2012 Odense, Denmark, 2012

Kaplan, B., Maxwell, J.A., (2005) Qualitative Research Methods for Evaluating Computer Information Systems, in Evaluating the Organizational Impact of Healthcare Information Systems, pp30-55, Springer

Katsaggelos, A.K., (2012) Digital Image Restauration, Springer

Kaufhold, A, Reuter, Ch., (2014) Vernetzte Selbsthilfe in Sozialen Medien am Beispiel des Hochwassers 2013, i-com Zeitschrift für interaktive und kooperative Medien, 13(1)

Kaufhold, M-A, Reuter, Ch., (2016) The Self-Organization of Digital Volunteers across Social Media: The Case of the 2013 European Floods in Germany. Journal of Homeland Security and Emergency Management (JHSEM), vol. 13, no. 1, pp. 137–166., https://doi.org/10.1515/jhsem-2015-0063

Khondker, H.H., (2011) Role of the New Media in the Arab Spring, Globalizations, 8:5, 675-679, DOI: 10.1080/14747731.2011.621287

Kleppmann, M., (2017) Designing Data-Intensive Applications, The Big Ideas Behind Reliable, Scalable and Maintainable Systems, O'Reilly

Koohikamali, M., Sidorova, A., (2017) Information re-sharing on social network sites in the age of fake news. Informing Science: the International Journal of an Emerging Transdiscipline, 20, 215-235., Retrieved from http://www.informingscience.org/Publications/3871 on 2019/09/01

Kübler, H-D., (2003) Kommunikation und Medien. Eine Einführung (S.91-129), LIT Verlag

Kwak, H, Lee, C., Park H., Moon S, (2010) What is Twitter, a Social Network or a News Media, WWW 2010, Raleigh, USA

Lasswell, H., (1960) The Structure and Function of Communication in Society, in Schramm, W. (Ed), Mass Communications (2nd ed), University of Illinois Press

LCC, (2017) Library of Congress Classification, https://www.loc.gov/catdir/cpso/lcco/ on 2018/03/28

Lee Hughes, A., Palen, L., (2009) Twitter Adoption and Use in Mass Convergence and Emergency Events, ISCRAM 2009, Gothenburg, Sweden

Lindsay, B., (2011) Social media and disasters: current uses, future options and policy considerations, Congressional Research Service

Lipscy, P., Kushida, K., and Incerti, T., (2013) The Fukushima Disaster and Japan's Nuclear Plant Vulnerability in Comparative Perspective." Environmental Science and Technology 47 (May), 6082–6088.,https://en.wikipedia.org/wiki/Fukushima_Daiichi_nuclear_disaster#Poor_communication_and_delays on 2017/09/30

Liu, Sophia B., Palen, L., Sutton, J., Hughes, A, Vieweg, S., (2008) In Search of the Bigger Picture: The Emergent Role of On-Line Photo Sharing in Times of Disaster, ISCRAM 2008, Washington, D.C., USA

Liu Fisher, B., Jin, Y, Briones, R. and Kuch, B., (2012) Managing Turbulence in the Blogosphere: Evaluating the Blog-Mediated Crisis Communication Model with the American Red Cross, Journal of Public Relations Research, 24, 353-370

Liu, L., Ouyang, W., Wang, X., Fieguth, P., Liu, X., Pietikäinen, M., (2018) Deep Learning for Generic Object Detection: A Survey, https://arxiv.org/abs/1809.02165v4

Liu, Y., Liu, Z., Chua, T., Sun M., (2015) Topical Word Embeddings, Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence

Manning, Ch., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S.J., McClosky, D., (2014) The Stanford CoreNLP Natural Language Processing Toolkit, Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pages 55–60, Baltimore, Maryland USA, June 23-24

Maynard, D., Roberts, I., Greenwood, M.A., Rout, D, Bontcheva, K., (2017) A framework for real-time semantic social media analysis, Web Semantics: Science, Services and Agents on the World Wide Web 44 (2017) 75-88

Meier, P., (2012) Behind the Scenes. The digital operations center of the American Red Cross, iRevolutions, Blog Post, April 17 2012

Meise, V. (2001) Ordnungsrahmen zur prozessorientierten Organisationsgestaltung.Kovač, Hamburg, ISBN 3-8300-0354-4 (Studien zur Wirtschaftsinformatik. Band 10), p.62

Mendoza, M., Poblete, B., and Castillo, C., (2010) Twitter Under Crisis: Can we trust what we RT?, SOMA 2010, Washington D.C., USA

MIME, (2018) Multipurpose Internet Mail Extensions, https://de.wikipedia.org/wiki/Multipurpose_Internet_Mail_Extensions on 2018/03/28

Moody, D. L., (2003) Measuring the Quality of Data Models: An Empirical Evaluation of the Use of Quality Metrics in Practice, ECIS 2003 Proceedings. 78

Moreno, A., Best, M.L, (2018) Crossmedia Election Monitoring, Report March 2018, United Nations University, Institute on Computing and Society (UNU-CS), available at http://collections.unu.edu/eserv/UNU:6473/Crossmedia_Election_Monitoring.pdf on 2018/12/25

Morstatter, F., Pfeffer, J., Liu, H., and Carley, K. M., (2013) Is the Sample Good Enough? Comparing Data from Twitter's Streaming Api with Twitter's Firehose, In Proceedings of ICWSM. Cambridge, MA: AAAI Press

Moshrafa, E., (2015) All you Need to Know About: The Cultivation Theory, 215, Global Journal of Human-Social Science: (A) Arts & Humanities- Psychology, Volume 15, Issue 8, Version 1.0, Global Journals Inc.

Muhren, W.J., Van de Walle, B., (2009) Sensemaking and Information Management in Humanitarian Disaster Response: Observations from the TRIPLEX Exercise, Proceedings of the 6th International ISCRAM Conference, Gothenburg, Sweden, May 2009

Murthy, D., Longwell, S.A., (2013) TWITTER AND DISASTERS, Information, Communication & Society, 16:6, 837-855, DOI: 10.1080/1369118X.2012.696123

Murzintcev, N., Cheng, Ch., (2017) Disaster Hashtags in Social Media, International Journal of Geo-Information, 2017, 6, 204, doi:10.3390/ijgi6070204

Munic Re, (2017) TOPICS Geo, Natural Catastrophes 2017, Analyses, assessments, positions, 2018 https://www.munichre.com/content/dam/websites/munichre/mrwebsites/topics-online/2018/2017-was-a-wake-up-call/302-09092_en%20(6).pdf on 2018/11/06

Nagar, S., Aaditeshwar, S., Joshi, A., (2012) Characterization of Social Media Response to Natural Disasters, SWDM 2012 Workshop, Lyon, France

Nadeau, D. and Sekine, S., (2007) A survey of named entity recognition and classification, Lingvisticae Investigationes, 30(1):3–26

Nielsen, 02-14-2018, (2018) AS THE AUDIO LANDSCAPE EVOLVES, BROADCAST RADIO REMAINS THE KING, https://www.nielsen.com/us/en/insights/news/2018/as-the-audio-landscape-evolves-broadcast-radio-remains-the-comparable-king.html available on 2018/11/11

Nilsson, S., Brynielsson, J., Granasen, M., Hellgren, Ch., Lindquist, S., Lundin, M, Narganes Quijano, M., Trnka, J., (2012) Making use of New Media for Pan-European Crisis Communication, ISCRAM 2012, Vancouver, Canada

NOAA (2018), National Centers for Environmental Information (NCEI) U.S. Billion-Dollar Weather and Climate Disasters (2018). https://www.ncdc.noaa.gov/billions/ on 2018/11/11

O'Reilly, T., (2005) What Is Web 2.0 - Design Patterns and Business Models for the Next Generation of Software, Retrieved from http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html on 2017/09/26

Okurowski, M.E, (1993) Information Extraction Overview, TIPSTER '93 Proceedings of a workshop on held at Fredericksburg, Virginia: September 19-23, 1993, Pages 117-121

Olteanu, A.; Vieweg, S.; & Castillo, C., (2015) What to Expect When the Unexpected Happens: Social Media Communications Across Crises. In Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (pp. 994–1009). New York, NY, USA: ACM., https://doi.org/10.1145/2675133.2675242

Pacchioli, D., (2013) Communication in the Fukushima Crisis, How did officials, scientists, and the media perform?, Oceanus Magazine, http://www.whoi.edu/oceanus/feature/communicating-science on 2017/09/30

Pan, S.J, Yang, Q., (2010) A Survey on Transfer Learning, IEEE Transactions on Knowledge and Data Engineering (IEEE TKDE). Volume 22, No. 10, Pages 1345-1359, October 2010

Palen, L., Anderson, K. M., (2016) Crisis informatics: New data for extraordinary times. Science, vol. 353, no. 6296, pp. 224–225., https://doi.org/10.1126/science.aag2579

Palen, L., Hughes, A.L., (2018) Social Media in Disaster Communication. In H. Rodríguez, W. Donner, & J. E. Trainor (Eds.), Handbook of Disaster Research (pp. 497–518). Cham: Springer International Publishing, https://doi.org/10.1007/978-3-319-63254-4_24

Panteras, G., Wise, S., Lu, X., Croitor, A., Crooks, A. and Stefanidis, A., (2014) Triangulating Social Multimedia Content for Event Localization using Flickr and Twitter, Transactions in GIS

Parilla-Ferrer, B.E, Fernandez, P.L., and Ballena, J.T., (2014) Automatic Classification of Disaster-Related Tweets, International conference on Innovative Engineering Technologies (ICIET'2014) Dec. 28-29, Bangkok (Thailand)

Peary, B., Shaw, R. and Takeuchi, Y., (2012) Utilization of Social Media in the East Japan Earthquake and Tsunami and its Effectiveness, Journal of Natural Disaster Science, Volume 34, Number 1, pp3-18

Prat, N., Wattiau, I., Akoka, J, (2014) Artifact Evaluation in Information Systems Design Science Research? A Holistic View, Proceedings - Pacific Asia Conference on Information Systems, PACIS 2014

Przybocki, M.A., Fiscus, J.G.; Garofolo, J., Pallett, D.S., (1999) 1998 Hub-4 information extraction evaluation. Proceedings of DARPA Broadcast News Workshop, vol. 225, pp. 13–18

Popescul, A., Ungar, L.H., (2000) Automatic Labeling of Document Clusters, available at https://pdfs.semanticscholar.org/9cf5/a7c658926f8c7898cae1bd4687f67a040617.pdf on 2018/11/24

Presthus, W., Munkvold, B.E., (2016) How to frame your contribution to knowledge? A guide for junior researchers in information systems. Paper presented at NOKOBIT 2016, Bergen, 28-30 Nov. 2016, NOKOBIT, vol. 24, no. 1, Bibsys Open Journal Systems, ISSN 1894-7719.

Reis, C., Mattedi, M., Reimondo Barrios, Y., (2017) Media and disasters: overview of international scientific production from 1996 to 2016, Intercom, Rev. Bras. Ciênc. Comun. vol.40 no.2 São Paulo May/Aug. 2017

Reuter, Ch. and Kaufhold, M-A., (2018) Fifteen Years of Social Media in Emergencies: A Retrospective Review and Future Directions for Crisis Informatics. Journal of Contingencies and Crisis Management (JCCM), pp. 1–17., https://doi.org/10.1111/1468-5973.12196

Reuter, Ch., Backfried, G., Kaufhold, M-A and Spahr, F., (2018a) ISCRAM turns 15: A Trend Analysis of Social Media Papers 2004-2017, ISCRAM 2018, Rochester, NY, USA

Reuter, Ch.; Hughes, A. L., Kaufhold, M-A., (2018b) Social Media in Crisis Management: An Evaluation and Analysis of Crisis Informatics Research. International Journal on Human-Computer Interaction (IJHCI), vol. 34, no. 4, pp. 280–294., 2018, https://doi.org/10.1080/10447318.2018.1427832

Newman, N., Fletcher, R., Kalogeropoulos, A, Levy, D.A., Nielsen, R.K., (2017) Reuters Institute for the Study of Journalism Digital News Report 2017, http://www.digitalnewsreport.org/survey/2017 on 2019/03/10

Robinson, I., Webber, J. Eifrem, E., (2015) Graph Databases, New Opportunities for Connected Data, 2nd edition, O'Reilly

Roddriguez, M.A., Neubauer, P., (2010) Constructions from Dots and Lines, Bulletin of the American Society for Information Science and Technology, American Society for Information Science and Technology, 36,(6), pp. 35-41, ISSN:1550-8366, August 2010

Rogers, R., (2017) Digital Methods for Cross-platform Analysis in the SAGE Handbook of Social Media, Chapter: 5, Publisher: Sage, Editors: Jean Burgess et al., pp.91-110, 2017

Sakaki, T., Okazaki, M., Matsuo, Y., (2010) Earthquake shakes Twitter users: real-time event detection by social sensors. WWW '10: Proceedings of the 19th International Conference on World Wide Web, pp. 851., https://doi.org/10.1145/1772690.1772777

St.Denis, L.A, Palen, L., Anderson, K.M, (2014) Mastering Social Media: An Analysis of Jefferson County's Communications during the 2013 Colorado Floods, ISCRAM 2014, , University Park , PA, USA

Scanlon, J., (2007) Unwelcome irritant or useful ally? The mass media in emergencies. In H. Rodriguez, E.L. Quarantelli & RR Dynes (Eds), Handbook of Disaster Research, New York, Springer

Scanlon , J. (2009) Research about the Mass Media and Disaster: Never (Well Hardly Ever) The Twain Shall Meet Disciplines, Disasters and Emergency Management 308, Textbook Retrieved from http://www.training.fema.gov/emiweb/downloads/ScanlonJournalism.pdf

Šestak, M., Rabuzin, K., & Novak, M., (2016) Integrity constraints in graph databases – implementation challenges, Central European Conference on Information and Intelligent Systems

Shaluf, I.M., Ahmadun, F., Said A.M., (2003) A review of disaster and crisis, Disaster Prevention and Management: An International Journal, Vol. 12 Iss 1 pp. 24 – 32, http://dx.doi.org/10.1108/09653560310463829

Shannon, C., Weaver, W., (1949) The Mathematical Theory of Communication, The University of Illinois Press, Urbana

Simsion, G, Witt, G., (2004) Data Modelling Essentials, 3rd edition, Morgan Kaufmann Publishers

Slavic, A., (2011) Classification revisited a web of knowledge. In: Innovations in information retrieval: perspectives for theory and practice. Eds. Allen Foster and Pauline Rafferty. London: Facet, pp. 23-48, arXiv:1705.07058v1

Snoeck, CGM, Worring M, Smeulders, AWM, (2005), Early vs late fusion in semantic video analysis, Proceedings of the 13th annual ACM International Conference on Multimedia, Singapore

Spiteri, L., (1998) A Simplified Model for Facet Analysis, Canadian Journal of Information and Library Science

Starbird, K., Palen, L., (2010) Pass It On?: Retweeting in Mass Emergency , ISCRAM 2010, Seattle, USA

Subrahmanian, V.S, Azaria, A., Durst S., Kagan V., Galstyan A., Lerman K., Zhu, L., Ferrara E., Flammini, A., and Menczer, F., (2016) The DARPA Twitter bot challenge, Computer, volume 49, number 6, pp. 38–46., doi: http://dx.doi.org/10.1109/MC.2016.183, accessed 2018/11/25

Tanev, H., Piskorski, J., Atkinson, M., (2008) Real-time news event extraction for global crisis monitoring, in Lecture Notes in Computer Science, Vol.5039, pp.207-218

Taylor, Ch., (2012) Sandy Really Was Instagram's Moment: 1.3 Million Pics Posted on Mashable, Nov 05 2012, available at https://mashable.com/2012/11/05/sandy-instagram-record/?europe=true#P2q5PfgpO8qM on 2018/11/11

Tierney, K., Bev C. and Kuligowski, E., (2006) Metaphors matter: Disaster myths, media frames, and their consequences in Hurricane Katrina, Annals of the American Academy, 604(1), 57-81

UN, (2009) United Nations: UN International Strategy for Disaster Reduction, 2009, UNISDR Terminology on Disaster Risk Reduction (Geneva: UNISDR, 2009) at 9; The Nansen Initiative, “Towards a Protection Agenda,” supra note 32

UNISDR, (2016) United Nations Office for Disaster Reduction, “Radio in times of disasters”, available at https://www.unisdr.org/archive/47813 on 2018/11/10

UNOCHA, (2014) United Nations Office for the Coordination of Humanitarian Affairs, Think Brief, Hashtag Standards for Emergencies, OCHA Policy and Studies Series, October 2014, 012

UNOCHA, (2017) United Nations Office for the Coordination of Humanitarian Affairs, World Humanitarian Data and Trends 2017, available at http://interactive.unocha.org/publication/datatrends2017/resources/WHDT2017_Final_Singles.pdf on 2018/11/11

US Department of Defense (DoD), (2005) Dictionary of Military and Associated Terms, Retrieved October 21 2018 from https://www.thefreedictionary.com/all-source+intelligence

Vaishnavi, V., Kuechler, W., and Petter, S. (Eds.) (2004/17). “Design Science Research in Information Systems” January 20, 2004, created in 2004 and updated until 2015 by Vaishnavi, V. and Kuechler, W.; last updated (by Vaishnavi, V. and Petter, S.), December 20, 2017. URL: http://www.desrist.org/design-research-in-information-systems on 2018/12/31

Venable, J.R., (2010) Design Science Research Post Hevner et al.: Criteria, Standards, Guidelines , and Expectations, in Global Perspective on Design Science Research, Winter, R., Zhao, J.L., Aier, S. (Eds.), 5th International Conference, DESRIST 2010, St.Gallen, Switzerland, June4-5. 2010, Proceedings

Venable, J., Pries-Heje, J., Baskerville, R., (2012) A Comprehensive Framework for Evaluation in Design Science Research, in: Peffers, K., Rothenberger, M, Kuechler, B (Eds) Design Science Research in Information Systems. Advances in Theory and Practice, Proceedings of the 7th International Conference, DESRIST 2012, Las Vegas, NV, USA

Wang, M, Deng, W., (2019) Deep Face Recognition: A Survey, https://arxiv.org/abs/1804.06655, on 03/30/2019

Wayne, Ch., (1998) Topic Detection and Tracking (TDT): Overview and Perspective, Proceedings of the Broadcast News Transcription and Understanding Workshop, Landsdowne, Virginia

Whetten, D.A. (1989) What constitutes a theoretical contribution, The Academy of Management Review 14(4):490-495, DOI: 10.5465/AMR

Winer, R.S., (2009) New Communication Approaches in Marketing: Issues and Research Directions, 2009, Journal of Interactive Marketing 23 (2009) 108-117, Elsevier

Zemp, H., Bonfadelli, H. (2008). Hochwasser im Fokus der Medien und in der Wahrnehmung der Bevölkerung; Hochwasserereignisse im Spiegel der Presse. 10.5167/uzh-8844.

Zhang, Y-J, (2017) Image Understanding in Image Engineering, Volume III, De Gruyter, Berlin

Zhao, Z-Q., Zheng, P., Xu, S-T, Wu, X., (2018) Object Detection with Deep Learning: A Review, arXiv:1807.05511

Table of Figures

List of Tables