






Utilizing computational methods for analysing media framing of organizational crises: The 'Datalek' scandal during the COVID-19 pandemic in the Netherlands

Dennis Nguyen¹  | Sergül Nguyen²  | Phuong Hoan Le²  | Tessa Oomen²  | Yijing Wang² 

¹Department of Media and Culture studies, Utrecht University, Utrecht, The Netherlands

²Department of Media and Communication, Erasmus University Rotterdam, Rotterdam, The Netherlands

Correspondence

Yijing Wang, Department of Media and Communication, Erasmus University Rotterdam, Burgemeester Oudlaan 50, Rotterdam, South Holland 3062 PA, The Netherlands.

Email: y.wang@eshcc.eur.nl

Funding information

None

Abstract

Media framing of organizational crises is an important factor to consider in crisis communication since it can shape stakeholders' perceptions of organizations and discussions in the public sphere. This takes place in complex media ecologies where public communication happens at a large scale, both in the news and on social media. Here, computational methods offer new venues for analysing media framing in flux throughout the crisis life cycle. Especially methods for automated content analysis can quickly and efficiently reveal what media frames emerge in a crisis context and how they change over time across different channels and platforms. The present study showcases the benefits of such methodological approaches by critically exploring the example of the data breach at the national municipal health service in the Netherlands. Using computational methods for media frame analysis on news texts (N1 = 519) and social media postings (N2 = 2986), this article reconstructs how the incident was perceived throughout four crisis stages (build-up, outbreak, chronic stage, termination). The article critically discusses the relevance of researching media framing empirically with emphasis on the benefits but also limitations of computational approaches. It concludes with some general pointers for crisis researchers interested in such methods as well as their implications for practitioners in the field.

KEYWORDS

automated text analysis, computational methods, COVID-19, crisis communication, data breach, media framing

1 | INTRODUCTION

Media framing is an important factor in the emergence, development, and conclusion of organizational crises, with the potential to shape views on causalities, effects, and responsibilities among the public (Coombs, 2007; Cross et al., 2019). In contemporary media ecologies,

public discourses take place across a wide range of media channels and digital platforms in which a diversity of commentators share their perspectives with broad audiences (Nguyen, 2017). Arguably, the two most important sites for public discursivity are news media and social media. In both, organizational crises affecting external stakeholders are often portrayed as highly urgent and impactful negative incidents

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Author(s). *Journal of Contingencies and Crisis Management* published by John Wiley & Sons Ltd.

worthy of public attention. Importantly, media framing is not static but subject to dynamic change throughout the unfolding of a crisis and its different stages.

Various qualitative, quantitative, and mixed-methods approaches have been proposed to research media framing (D'Angelo et al., 2019). While valuable, these are often applied manually, demanding considerable time and resources. Computational methods offer quicker and potentially more cost-efficient solutions for analysing media texts from multiple angles in a relatively short amount of time (van der Meer, 2016). Especially methods in the field of natural language processing (NLP) focusing on automated text analysis can benefit media framing research that maps out changes over time (Nguyen & Hekman, 2022). This can yield important empirical insights for understanding the mutual affectivity between crisis development and media framing.

A particularly complex type of organizational crisis is data breaches. In short, a 'data breach is the intentional or inadvertent exposure of confidential information to unauthorized parties' (Cheng et al., 2017, p. 1). Data breaches affecting citizens have considerable news value due to their potentially profound consequences for all stakeholders involved. They pose serious crisis communication challenges for organizations: 1) they are complex issues with largely unforeseeable negative effects; 2) their impact is often delayed, possibly prolonging the necessity to engage with the issue over longer time spans; 3) they create various direct and indirect costs for organizations; and 4) data breaches are not always easy to explain to the public, given their inherent complexity and diverse causes (Kuipers & Schonheit, 2021).

The present paper aims to illustrate how computational methods can be utilized in mapping crisis discourses and media framing variations, focusing on the specific case of a data breach in the Netherlands during the COVID-19 pandemic as an empirical example. In January 2021, Dutch news media reported a significant data breach at the national municipal health service, known as the Gemeentelijke Gezondheidsdiensten (GGD), which collected data for COVID-19 contact tracing (Verlaan, 2021). GGD employees stole a large volume of citizen data from its databases with the intent to sell it on the black market. These data included names, contact information, residential addresses, patient numbers, phone numbers, COVID-19 test results, and citizens' service numbers. The incident triggered a public debate about governmental digitalization, data practices, and mismanagement across news media and social media, especially X/Twitter. It also prompted a Dutch parliamentary enquiry and demands from an NGO for the government to compensate victims, possibly costing billions of euros (Quekel, 2021). Different stakeholders—governmental, oppositional, media, and civic—contributed to the initiation and perpetuation of a crisis discourse that quickly escalated in the Dutch public sphere by connecting the data breach to fundamental questions about pandemic policies.

The GGD case underlines media framing as an important factor in the lifecycle of organizational crises that should be critically considered in crisis communication efforts. It presents a complex and insightful example for demonstrating how computational

methods can enhance understanding of data breach dynamics and quantify framing trends over the crisis timeline. The present article thus aims to address the research question: *How can methods for automated text analysis be utilized for the tracking of media framing of an organizational crisis across different media types and crisis stages?*

Analysing media framing of data breaches specifically has both academic and practical benefits. First, limited research is available on how data breaches are 'made sense of' from a media framing angle. More empirical research is needed to unearth the complexities and dynamics that shape perceptions and assessments of their impact, especially throughout the crisis life cycle. Second, insights into media framing help organizations assess how to craft effective communication strategies during different crisis phases. By understanding how data breaches are presented across media platforms, organizations can address public concerns more directly, mitigate panic, and provide clear, actionable information to prevent or respond to data breaches. The present article argues that computational methods can be effectively applied for both objectives.

In what follows, the article briefly discusses the specifics of data breaches as organizational crises before outlining the relevance of media framing in crisis communication. It then introduces different computational methods for analysing text data. The analyses and results offer insights into the impact of media framing on public perception and organizational reputation.

1.1 | Data breaches as organizational crises

Data breaches are recognized by researchers and practitioners alike as one of the most recurrent and damaging cyber incidents for organizations and their stakeholders (Kuipers & Schonheit, 2021). Illicit exposure of data constitutes a security incident in which sensitive and protected data are copied, revealed, or stolen, resulting in a breach of confidentiality, availability, or integrity and posing a risk to an individual's rights and freedoms (European Commission, 2016). Data breaches present ambiguous situations where it can be difficult to identify causes and offenders (Bentley et al., 2018). Arguably, their emergence often depends on factors endogenous to organizations, including inconsistent data retention and handling policies, system vulnerabilities, and human errors (Kuipers & Schonheit, 2021).

The effects of data breaches are not always straightforward, and the causes, scope, and consequences of risks are often difficult to identify (Kuipers & Schonheit, 2021). As a result, affected organizations experience high uncertainty and delayed urgency in their response (Lopes et al., 2019). The exposure of sensitive data can have detrimental effects on all stakeholders involved. Individuals may face various risks related to cybercrime, fraud, and even extortion, while organizations may suffer direct costs, such as forensic investigations, legal proceedings, and regulatory fines, as well as indirect costs, such as reputational damages and loss of stakeholder trust (Syed, 2019).

Data breaches constitute complex organizational incidents with potentially wide-reaching effects, and the lack of clarity can result in crisis escalation. The present study adopts Coombs' (2009) crisis

definition: '[a] crisis can be viewed as the perception of an event that threatens important expectancies of stakeholders and can impact the organization's performance' (p. 100). This emphasizes that crises are largely perceptual, and their emergence and impact are highly dependent on stakeholders' beliefs (Lopes et al., 2019). Data breaches only become crises when the necessity of immediate action is triggered by, for example, intense media attention or direct threats to stakeholders.

Organizations dealing with a data breach are challenged to find a suitable crisis communication strategy. Park (2017) advocates for more research on crisis responses, while Kuipers and Schonheit (2021) argue that investigating communication strategies in data breaches benefits both crisis communication and cybersecurity literature. However, there is a lack of research into media framing of data breaches in different crisis stages, or how framing characteristics pose risks to organizations in crisis and how to navigate these challenges.

Data breaches are a complex subject for media framing due to several factors. First, they involve diverse stakeholders (e.g., organizations, citizens/customers, regulators) each facing unique risks and vulnerabilities. Second, data breaches unfold over time, often not detected immediately, and their impacts, particularly on individuals, may only become evident later, adding complexity. Third, while data breaches represent considerable risks, they remain an abstract concept for the public. This underlines the crucial role of journalists as societal sense-makers (Robertson, 2010), who 'translate' complex issues for wider audiences. Coombs (2007) argues that media framing can substantially affect organizational reputation, and understanding what information media outlets cover can help organizations with making informed decisions regarding their crisis communication.

1.2 | Crisis communication, media framing, and the crisis life cycle

Organizational crises have high news value since they tend to signal potentially negative outcomes (Harcup & O'Neill, 2017) for their stakeholders and may raise public concern. During crises, news media play an important role as sources of information for public audiences and critical observers of societal actors. Additionally, social media platforms have become equally important, as affected organizations face a variety of reactions from political actors, representatives from civil society, media professionals, and a wider undefined public audience (Syed, 2019).

Media framing refers to the assumption that how an issue is portrayed in news reports or other media texts can affect how it is perceived by audiences (Scheufele & Tewksbury, 2007). Prior research shows that media framing can shape the perceived causes and severity of a crisis, as well as influence public emotions (Kepplinger et al., 2012). Coombs (2007) contends that emphasizing specific factors in media communication influences opinions about a crisis among recipients. Thus, the extent to which stakeholders blame an organization and assign responsibility for a crisis depends partly on the framing of the crisis within media discourses.

Media frames can be operationalized in various ways, but it is often useful to distinguish between emphasis frames (Chong & Druckman, 2007)—which address specific aspects, dimensions, and (sub-)topics of broader issues—and valence framing, which describes how an issue and its components are portrayed in terms of sentiment (e.g., positive, negative, supportive, oppositional). Media frames are not static entities but highly dynamic, changing in real-time corresponding to how events and issues unfold. New information, additional stakeholders, sudden shifts in the evaluation of relevance and other factors that Pentzold and Fraas (2023) call 'contextual configurations' all play into the emergence and transformation of media frames. Analysing the stakeholders referenced in media framing is also crucial as it reveals who is deemed significant in a crisis discourse, positioning societal actors in terms of responsibility, victimhood, and mediation (Nguyen, 2017).

Similar to the dynamism of media framing, crisis communication scholars argue that crises should be treated as a process, as opposed to a discrete event (Bundy et al., 2017). This implies that a crisis develops through a series of stages, and each stage involves distinct characteristics and dynamics (Sturges, 1994). Fink (1986) introduced the crisis stage theory which regards crisis as 'an unstable time or state of affairs in which a decisive change is impending' (p. 15). He proposes the notion of a lifecycle which comprises of four stages: *crisis build up*, *crisis breakout*, *chronic stage*, and *termination*. While newer models exist, they have been largely informed by Fink's original work (Stewart & Wilson, 2016).

The identification of crisis stages relies on retrospective analysis, as they cannot be discerned or predicted with certainty in advance of their unfolding. In the present study, Fink's crisis stage model is adopted; it has been empirically tested in the contexts of natural disasters (Spence et al., 2015) and corporate crises (Sung & Hwang, 2014). Employing the model to examine organizational crises allows to compare different crisis lifecycle stages in respect to how media framing changes over time. Mapping different (counter) frames that frequently emerge in news and on social media throughout the crisis lifecycle offers several benefits, both academically and practically. On the one hand, it contributes to a better understanding of how media framing, crisis perception, and organizational communication influence each other. On the other, it can support organizations in predicting what frames are picked up by journalists during specific lifecycles and to (re)frame their messages accordingly.

1.3 | Computational methods for researching media framing of organizational crises

A growing number of studies make use of computational methods to cluster media texts for frame detection (Nguyen & Hekman, 2022). These apply different forms of unsupervised and/or supervised machine learning in the NLP domain for identifying patterns in large text corpora, usually grouping documents based on shared features such as frequently co-occurring words. There are various approaches,

such as topic modelling, vectorization and clustering, and comparing word frequencies.

In the context of COVID-19, previous studies made use of, for example, topic modelling to explore polarization in political news reporting (Hart et al., 2020) or tracked changes over time in social media discourses on the pandemic (Bogdanowicz & Guan, 2022). In crisis communication research specifically, the potentials of automated text analysis have been repeatedly acknowledged (van der Meer, 2016; Vogler & Meissner, 2022). van der Meer (2016, p. 955) argues that especially 'the theory of framing in combination with automated content analysis can play a key role in crisis research to provide insights in how the situation is understood and framed and to what extent the framing of different actors is comparable or different'. If researchers have the data and programming skills, computational text analyses can quickly reveal dominant frames and framing practices across different media types.

However, so far, only a few crisis communication studies have made use of computational methods. Examples are automated text analyses of social media discourses and their implications for reputation management during disasters (Jin & Spence, 2021) or how crisis communication in digital networks is received by stakeholders (Dhar & Bose, 2022). Comparative studies considering different media types and charting out crisis stages over time have not been widely conducted yet. This is a missed opportunity since tracing media framing practices over time based on prevalent word features in a given period compared to another is relatively simple. It is here important, however, to provide a clear definition of what constitutes a media frame and to clearly demarcate how far the benefits of computational-quantitative methods go and where their limitations are.

In the present study, automated content analysis based on text clustering is deployed to identify emphasis frames, that is, the main thematic emphases of news media reporting and social media discussions. Indicators for different frames are frequently co-occurring words, especially nouns. For example, if a group of documents frequently shares the dominant words *police*, *investigation*, *forensics*, and *evidence*, then this could be considered the 'Police Investigation' emphasis frame, which is distinguishable from others (e.g., 'Political Debate' or 'Compensation Claim'). Manual validation by cross-checking samples per document cluster is essential for finalizing the frame labelling. A limitation is that tone of voice, use of metaphors, and other more complex, nuanced, and potentially ambiguous uses of language are not (easily) captured in this way. While advanced machine learning, especially involving supervised methods utilizing hand-annotated training data, can somewhat approach these more complex framing dimensions, automated text clustering has here clear limitations.

Based on the previous discussion, the present paper aims to comparatively analyse media framing across different media types and crisis stages. Furthermore, references to societal actors are considered as important hints for public allocation of responsibility in crisis contexts. Accordingly, to address the main research question, the analysis is structured around the following subquestions:

SQ1: *How did Dutch mainstream news and social media commentators on X/Twitter frame the GGD data breach?*

SQ2: *What societal actors were considered relevant for the GGD data breach and what are the implications for perceived crisis responsibility?*

To address these research questions, a representative corpus of news articles and social media postings from X/Twitter were analysed with computational methods for text analysis. The computational approach was chosen as it allows for quick comparative media frame analyses of larger text corpora. A fully manual approach would have been more time-consuming; coding hundreds of news articles and thousands of social media posts by hand can be costly and result in overload and/or coder fatigue. Furthermore, smaller samples would be delimiting in respect to different quantitative analyses for linking media framing to crisis stages.

Concerning crisis management practices, addressing these research questions helps in understanding how media framing of data breaches shapes crisis discourses across different stages. This, in turn, can provide valuable insights for devising effective communication strategies to address public concerns, provide clear information, identify risks to the organization's reputation, and support the development of desirable counter-frames.

2 | METHOD AND DATA

First, the researchers manually collected key news items for developing a timeline of the GGD data breach, beginning with the news item that broke the topic. Based on this initial search and developed timeline, Dutch news articles on the topic were retrieved using keyword searches in Nexis Uni with Boolean operators: '(datalek OR data-lek OR datadiefstal OR data-diefstal) AND GGD'. After removing duplicates, the final news corpus (N1 = 519) covered articles from April 2020 to January 2022. The data set was used for identifying emerging crisis stages using a temporal bracketing strategy (Langley, 1999). This approach draws on Giddens' (1984) structuration theory and has been used effectively in prior research on organizational crises (Nan & Lu, 2014). Langley's bracketing strategy involves temporarily isolating individual actions from their constraining structures to sequentially analyse processes. This strategy enables the structuring of a shapeless mass of process data into distinct, interconnected phases and the identification of stable patterns within those phases (Langley, 1999). Figure 1 shows the different crisis stages as identified for the GGD data breach case.

Additionally, tweets in Dutch containing 'datalek' and 'datadiefstal' from the same period were collected using X/Twitter's former Academic API, resulting in a corpus of N2 = 2986 unique tweets.

To analyse media framing in Dutch news and social media reactions on X, various computational methods for a multilevel content analysis were applied, using Python 3 (Burscher et al., 2016). The respective corpora underwent different text pre-processing steps, including tokenization, stop word removal, lemmatization, lower-casing, and the removal of symbols and numbers. Additionally,

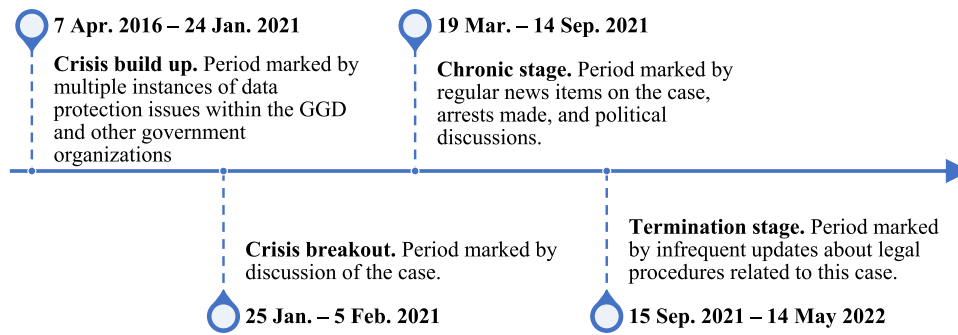


FIGURE 1 Crisis stages of the GGD data breach case.

Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4
vaccin	politie	gemeente	kabinet	minister
locatie	aanhouding	technologie	versoepeling	kamer
prik	verband	overheid	avondklok	kamerlid
afspraak	onderzoek	partij	team	debat
uitnodiging	callocenter	burger	routekaart	taak
levering	politie_onderzoek	piratenpartij	school	kabinet
regio	privégegeven	politiek	variant	overheid
kwartaal	verkoop	token	lockdown	partij
crisis	zaak	onderwerp	besmetting	controle
priklocaties	verdacht	tijd	premier	probleem
miljoen	arrestatie	digitalisering	aantal	ministerie
huisarts	adresgegevens	privacy	peiling	vaccinatie
zorgmedewerker	verdenken	probleem	restaurant	d66kamerlid
plan	bovenkanspel	risico	persconferentie	minister_volkgezondheid
ggds	datadiefstal	samenleving	maatregel	signa
ziekenhuis	coronasystemen	beleid	rapport	volksgezondheid
nacht	politie_verdenken	politicus	cijfer	cyberweerbaarheid
vaccinatielocatie	burgerservicenummer	kennis	virus	vragenuurtje
eind	systeem_coronatest	zetel	ouder	opposite
personeel	koop	kant	basisschol	kruid
Vaccinations	Police Investigation & Corona Security Issues	Political Debate over Privacy and Data Security	Corona Measures	Governmental Responsibility

FIGURE 2 Screenshot of example clusters from term frequency-inverse document frequency news texts.

for the identification of emphasis frames only nouns were retained via Spacy's Dutch language model (*nl_core_news_lg*).

The automated content analysis aimed to 1) identify dominant emphasis frames over time in Dutch news and on X comparatively, and 2) pinpoint dominant stakeholders during specific crisis stages. For the framing analysis, *term frequency-inverse document frequency* (TF-IDF) and K-Means clustering were deployed on both corpora. TF-IDF highlights important words in a text by weighing their frequency against their commonness across all texts, thus emphasizing distinctive terms (Figure 2). K-means clustering then groups texts into clusters based on the similarity of their TF-IDF profiles. This process can reveal emphasis frames by clustering texts with similar thematic focus based on the use of similar indicative words.

Determining the optimal number of clusters, *k*, was achieved with the help of silhouette scores. Silhouette scores measured how similar each text is to its own cluster compared to other clusters. The optimal number of clusters is identified at the point where the silhouette score, ranging from -1 to 1, is the highest, signalling the best balance of intracluster similarity and intercluster dissimilarity. However, these quantitative measures only provide orientation and

human validation and experimentation with different *k* is essential for determining how a range of clusters best represent the content of a given corpus. For both news texts and tweets/X postings, 18 clusters yielded interpretable results, with silhouette scores of 0.059 and 0.037, respectively. It is important to note that the media discourse on the GGD data breach scandal focused on a very specific issue with considerable thematic homogeneity, which explains these relatively low scores. Yet the clustering approach still allowed to distinguish between different topical emphases. To provide a more general overview, the 18 emphasis frames were further grouped based on their thematic similarities into four meta-frames for each corpus (Tables 1 and 2, and Appendix S1). Meta-frames further collapse and summarize specific emphasis frames based on their similarities. Meta-frames are higher level frames that propose more general yet distinguishable interpretative frameworks, like generic frames (e.g., economic consequences, conflict, morality; Semetko & Valkenburg, 2000).

Emphasis frames were labelled by reviewing and interpreting the top 15 most frequent words in each cluster as thematic indicators. Frame labels' validity was confirmed through human review of representative topic sub-samples (at least 10% per cluster). Intercoder

agreement between a human annotator and the algorithm was assessed using Krippendorff's α (Krippendorff, 2011). The objective was to verify if a human annotator would categorize texts into the same (meta-)frame as accomplished with the algorithm. For example, the human annotator had to decide which frame a given news article belongs to, and this decision was then compared with the algorithm's classification. For news texts, an agreement score of 0.88 was achieved, while tweets showed a lower yet acceptable agreement of 0.75. Valence framing was mostly qualitatively analysed by reviewing sample texts from each identified emphasis frame across the corpora. Part of the analysis then investigated whether there are hints for

statistically significant relationships between specific (meta-)frames and crisis stages.

Finally, named entity recognition (NER) via FLAIR (Akbik et al., 2019) served for retrieving stakeholders mentioned in the texts (persons and organizations). FLAIR/ner-dutch is an NLP model hosted on Hugging Face, designed for NER specifically in the Dutch language, utilizing the FLAIR framework for enhanced context-aware embeddings. This allowed to probe what stakeholders were considered dominant during different crisis stages and in specific media frames across news coverage and on X.

TABLE 1 Meta-frames news.

Meta frames	Topics
Corona policies, vaccinations, stats & use of data	0, 3, 13
Police investigation, legal consequences, victims & cybercrime	1, 6, 7, 10, 11, 15
Corona app, data systems, cybersecurity & corona governance	8, 14, 16, 17
Political debate, parliament, governmental responses & responsibility	2, 4, 5, 9, 12

TABLE 2 Meta-frames tweets.

Meta frames	Topics
Governmental response, responsibility, corona policies & criticism	0, 1, 7, 11, 14
GGD, insecure data systems and data theft	5, 6, 9
Extent of data leak, impact/effects & victims and public debate	2, 4, 10, 13, 12, 15
Compensation claim	3, 5, 8

3 | RESULTS

3.1 | Media framing of the GGD data breach in Dutch mainstream news and on X

Concerning the first set of subquestions 'How did Dutch mainstream news and social media commentators on X/Twitter frame the GGD data breach?', the results show considerable similarities but also important distinctions across mainstream news and among social media commentators.

News media primarily framed the data breach as a societal issue with relevance beyond the immediate organizational context. Focus was placed on different political, social, legal, and technical aspects of the case (Figure 3). Most articles varyingly emphasized breaking the news about the GGD data breach and the ensuing police investigation, the role of data-driven systems in governing the corona crisis, the impact of the data breach on citizens, and questions of governmental responsibility.

The 18 distinct topics can be grouped into four dominant themes or meta-frames (Figure 4). Generally, news coverage of the data breach peaked around January to February 2021, though some of the relevant

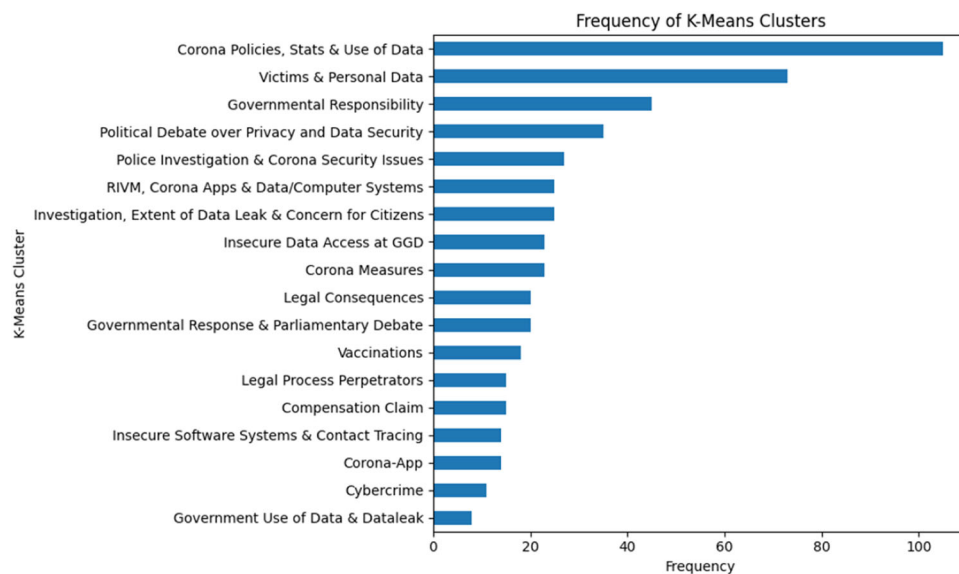


FIGURE 3 Topics in news (N = 516).

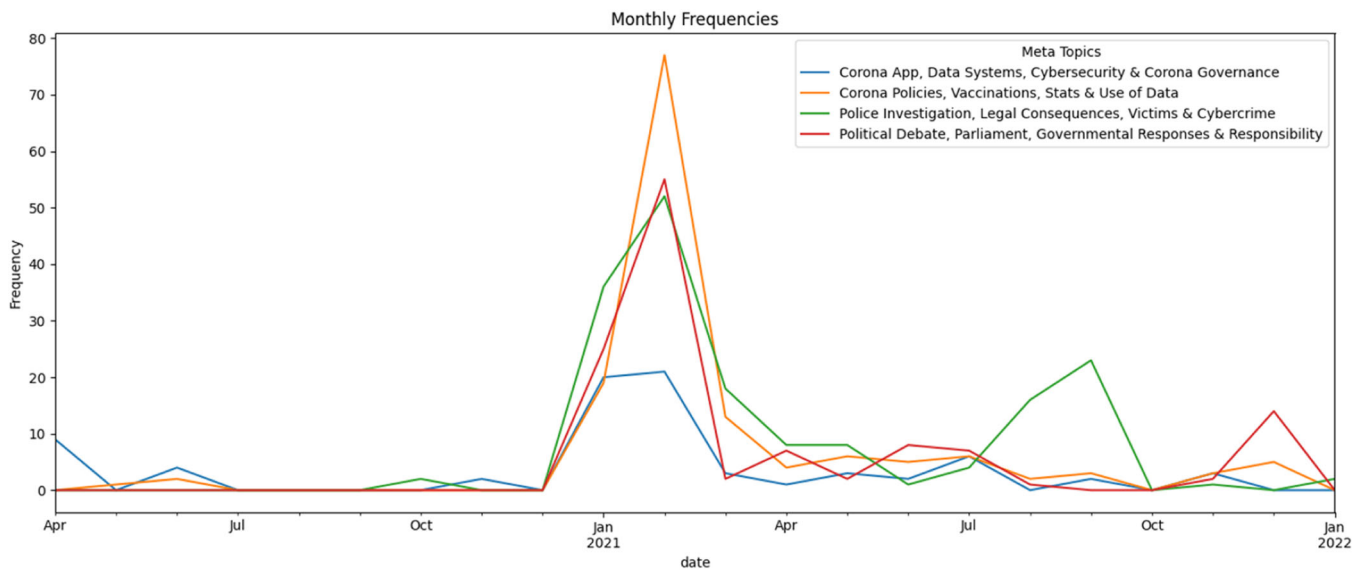


FIGURE 4 Meta-frames news ($N = 519$).

political and legal developments reattracted media attention at later stages. This implies that different dimensions of the crisis follow different temporal trajectories and may maintain prolonged news value.

The largest meta-frame is *Police Investigation, Legal Consequences, Victims & Cybercrime* ($N = 176$). Articles in this cluster cover the data breach primarily as a crime story with insecure data systems, suspects and perpetrators, victims (unforeseeable) risks of the data theft, police investigations, and the legal repercussions for different stakeholders involved.

The second largest meta-frame is *Corona Policies, Vaccinations, Stats & Use of Data* ($N = 146$). Here, news reporting frames the data breach as part of broader trends in governing the pandemic, the efficiency of different corona measures—especially vaccinations—and the use of data in tackling the pandemic.

The third meta-frame, *Political Debate, Parliament, Governmental Responses & Responsibility* ($N = 123$), portrays the data breach as a political issue. It focuses on questions of responsibility, scrutinizes governmental data practices and responses to the incident, and highlights critical debates within the political landscape, particularly in the Dutch parliament. This includes a collective compensation claim by a foundation demanding the government to pay billions for the loss of personal data of affected citizens.

The smallest meta-frame, *Corona App, Data Systems, Cybersecurity & Corona Governance* ($N = 76$), focuses on technologies and systems for managing pandemic data. These articles often highlight the weaknesses, vulnerabilities, and limitations of techno-centric pandemic strategies, along with cybersecurity concerns.

Figure 5 illustrates the meta-frame distribution by crisis stage, highlighting a tendency for news reporting during the crisis build-up to focus on *Corona App, Data Systems, Cybersecurity & Corona Governance* ($\chi^2[3] = 52.99, p < .001$). This is particularly true for the emphasis on Corona-Apps as technological solutions for pandemic management. It includes narratives on their intended use, benefits,

and the risks associated with apps, algorithms, and data usage in the COVID-19 context. Risks are latent and hypothetical at this stage.

During the crisis breakout, news reporting first 'broke the news' about the incident and then shifted focus to *Police Investigation, Legal Consequences, Victims & Cybercrime* ($\chi^2[3] = 13.12, p < .001$), with *Victims & Personal Data* as the most frequent particular emphasis frame for this period. When the crisis reached its chronic stage, news media frequently covered *Corona Policies, Vaccinations, Stats & Use of Data* ($\chi^2[3] = 9.26, p = .03$). News coverage tended to focus on the specific emphasis frame *Political Debate over Privacy and Data Security*. Finally, during the termination stage, media attention shifted to *Police Investigation, Legal Consequences, Victims & Cybercrime* ($\chi^2[3] = 13.12, p < .001$). Most articles evoke the specific emphasis frame *Legal Consequences* in this period.

In comparison, X/Twitter discussions tend to frame the data breach primarily as a controversy. The social media discussion shares some similarities with news media narratives on the data breach but shows diverging emphases (Figure 6). Social media commentary is notably more critical of various stakeholders, especially the GGD. Users often connect the data breach and cybercrime risks to other criticized pandemic policies. For instance, the government's handling of the data breach is frequently compared to the vaccination programme, critiqued as a flawed effort symbolic of broader official mismanagement and political incompetence in addressing pandemic-related issues.

The 18 different emphasis frames on X/Twitter can also be grouped into four broader meta-frames (Figure 7). The largest by far is *Extent of Data Leak, Impact/Effects & Victims* ($N = 1907$). Tweets included here tend to highlight the scope of the incident, which is considered massive, as well as various risks and negative impacts for citizens. For example, numerous tweets refer to different types of personal information that were lost to criminals, such as citizen service numbers or contact details. It is then alleged that such

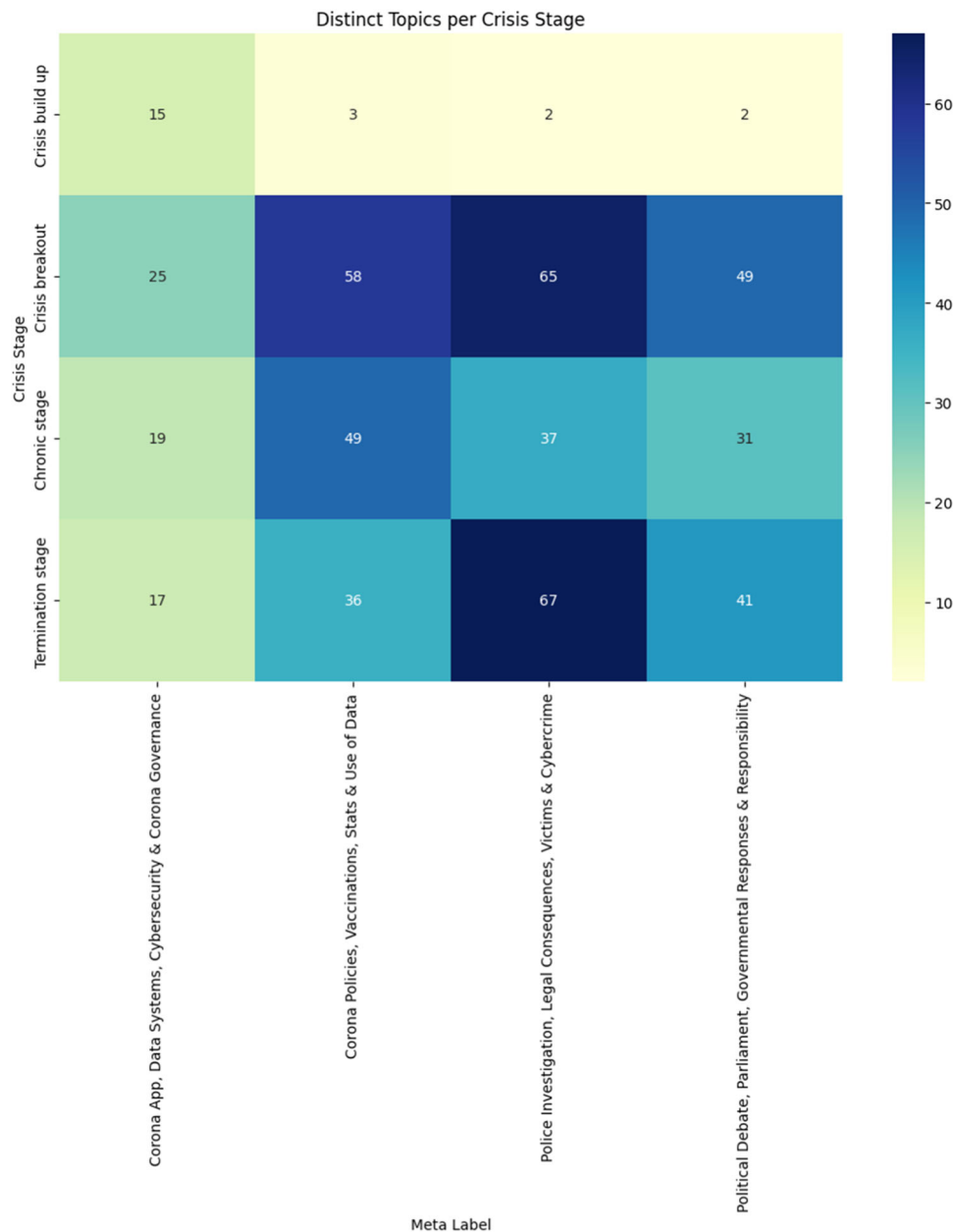


FIGURE 5 Meta-frames per crisis stage news.

information could be easily abused for identity fraud and other harmful activities. This is often directly tied to criticism on key actors, namely the GGD and governmental representatives.

The second-largest meta-frame focuses on the *Compensation Claim* ($N = 424$), also dubbed 'megaclaim' for its significant financial implications. The third largest is *Governmental Response, Responsibility, Corona Policies & Criticism* ($N = 375$), where tweets critically address the government's data breach management, question its responsibility, and link the incident to other perceived corona policy failures. The fourth meta-frame, *GGD, Insecure Data Systems & Data Theft* ($N = 280$), highlights the GGD's inadequate data management as a facilitator of the large-scale data leak.

Figure 8 highlights the most frequent hashtags on X/Twitter related to the data leak, showing that many directly link to specific

stakeholders, indicating responsibility allocation and explicit evaluations via social media. Notable examples include #ggd, #ggddatalek, #klaarmetrotte ('#donewithrotte'), and #hugodejongekanniks ('Hugo de Jonge is "useless"). This illustrates the association of the data breach with controversial COVID-19 policy topics, like vaccinations (#vaccindebacl, #vaccinatiestrategie), and other scandals not related to the pandemic, such as the #toeslagenaffaire (welfare scandal), highlighting another case of data misuse.

There is less variation in the X/Twitter discourses on dominant meta-frames per crisis stage than in the news reporting based on sheer frequency counts (Figure 9). Throughout the crisis stages, most tweets fall into the meta-frame *Extent of Data Leak, Impact/Effects & Victims*.

When zooming-in on specific emphasis frames, two of them dominated the different crisis stages: during the crisis build up, *Data*

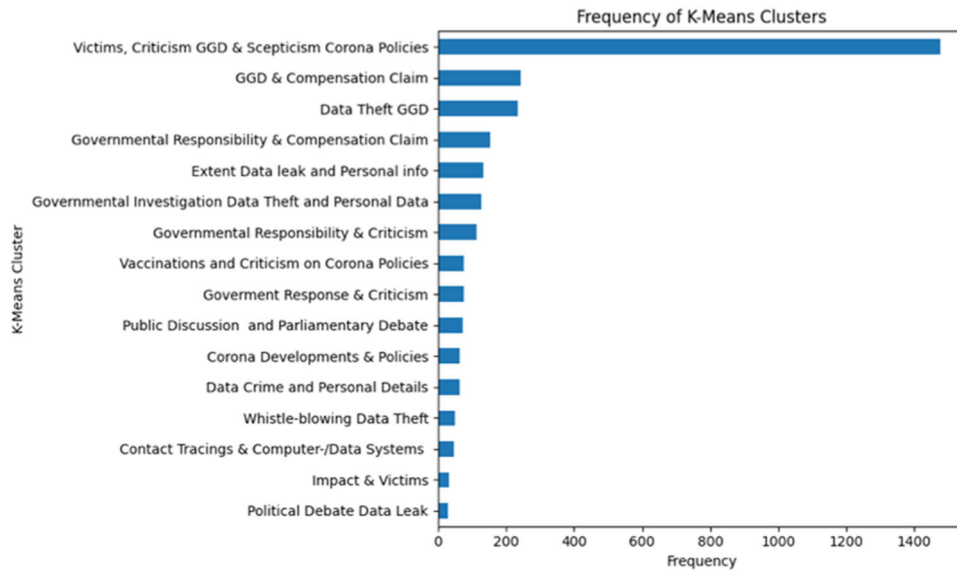


FIGURE 6 Topics in tweets (N2 = 2986).

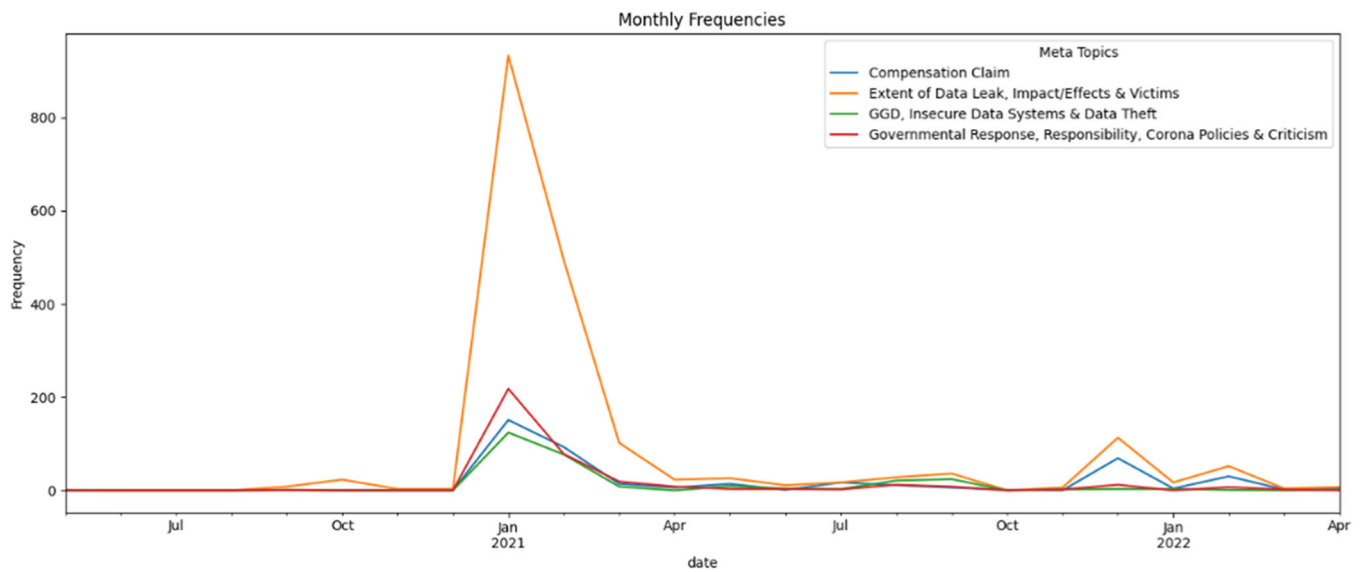


FIGURE 7 Meta frames tweets (N2 = 2986).



FIGURE 8 Hashtags in data leak tweets.

Theft & Personal Data was the most frequently evoked emphasis frame ($\chi^2[3, N = 22] = 254.85, p < .001$). In all other stages, the dominant emphasis frame was *Victims, Criticism of GGD & Skepticism of Corona Policies* (breakout $N = 947$, chronic $N = 186$, termination $N = 330, \chi^2(3)$ of $9.78, p = .02$).

3.2 | Visibility of societal actors in the GGD data breach media discourse

Regarding the second subquestion, 'What societal actors were considered relevant for the GGD data breach and what are the

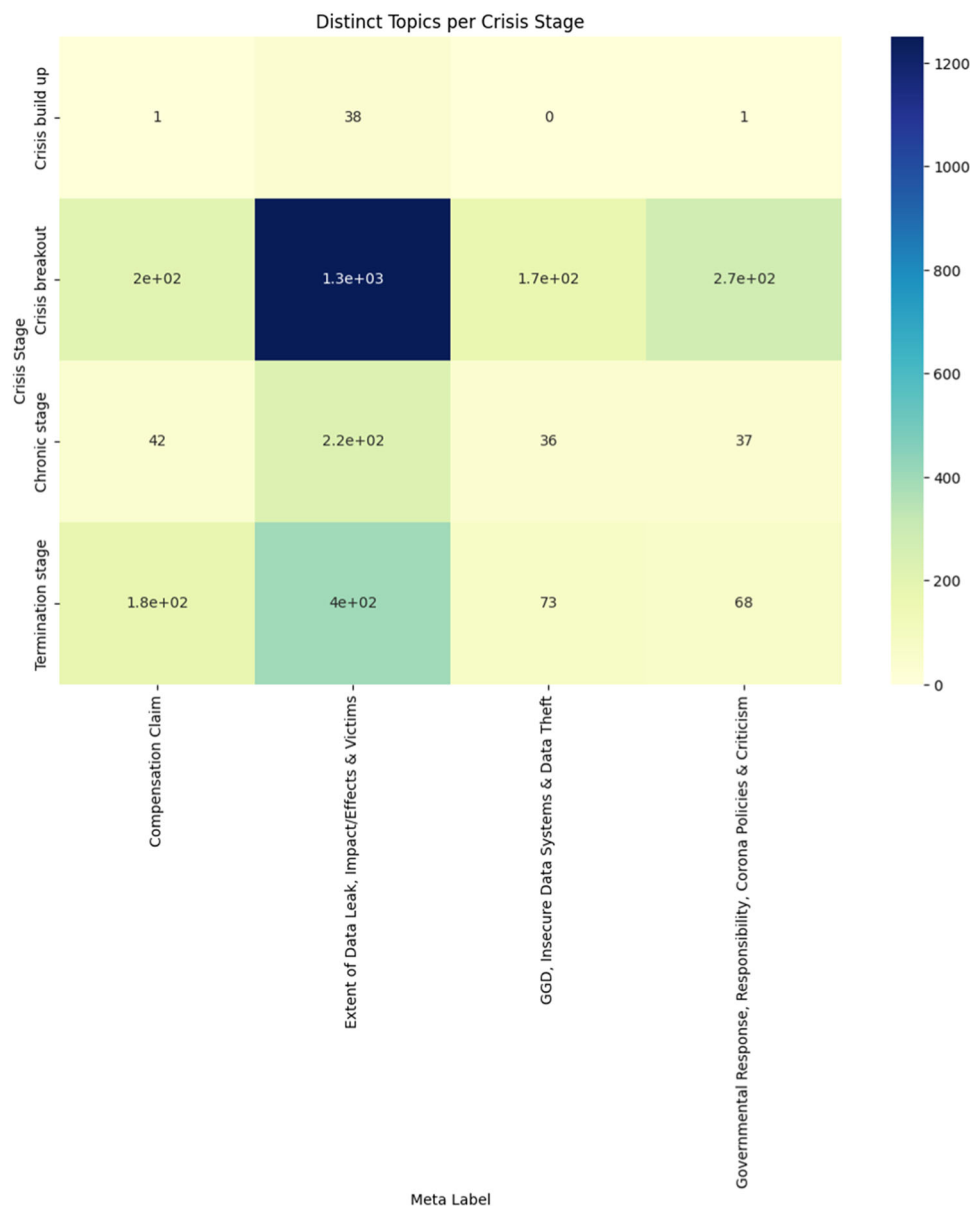


FIGURE 9 Meta-frames per crisis stage tweets ($N = 2984$).

implications for perceived crisis responsibility?', it appears that a limited number of recurring, high-ranking Dutch governmental representatives are primarily mentioned across news articles and social media posts. This suggests that perceived responsibility for the data breach crisis transcends the organizational context in the public discussion.

News reports mostly mentioned governmental stakeholders, but also suspects involved in the data breach, journalists, and other public personalities in Dutch society (Figure 10). The relevance of each actor varies per crisis stage, however (Table 3). Hugo de Jonge, Dutch Health Minister during the pandemic, was the most mentioned political figure in data breach news, indicating his perceived responsibility across all crisis stages. Prime Minister Mark Rutte was the second most cited. Although GGD representatives such as regional director Ellis Jeurissen were mentioned, more focus was placed on top government officials.

Regarding frequently mentioned organizations, it is political parties, state institutions, and news media that dominate the data breach news reporting (Figure 11).

X/Twitter discussions largely overlap with the findings for the news reporting with some notable differences (Figure 12). Social media postings tended to mostly mention specific government representatives, such as Health Minister Hugo de Jonge, his successor Ernst Kuipers and Prime Minister Mark Rutte but also journalists such as Daniël Verlaan or Eva Jinek.

The situation looks similar for the different crisis stages, where social media posts tend to focus on high-ranking government officials (Table 4).

Tweets mention mostly the same organizations as news articles (Figure 13). However, there is one noticeable difference: Initiatives Collective Actions Mass Damage (ICAM), which is the driving force behind the compensation claim and sued the Dutch state, is one of the

FIGURE 10 Top 15 actors in news.

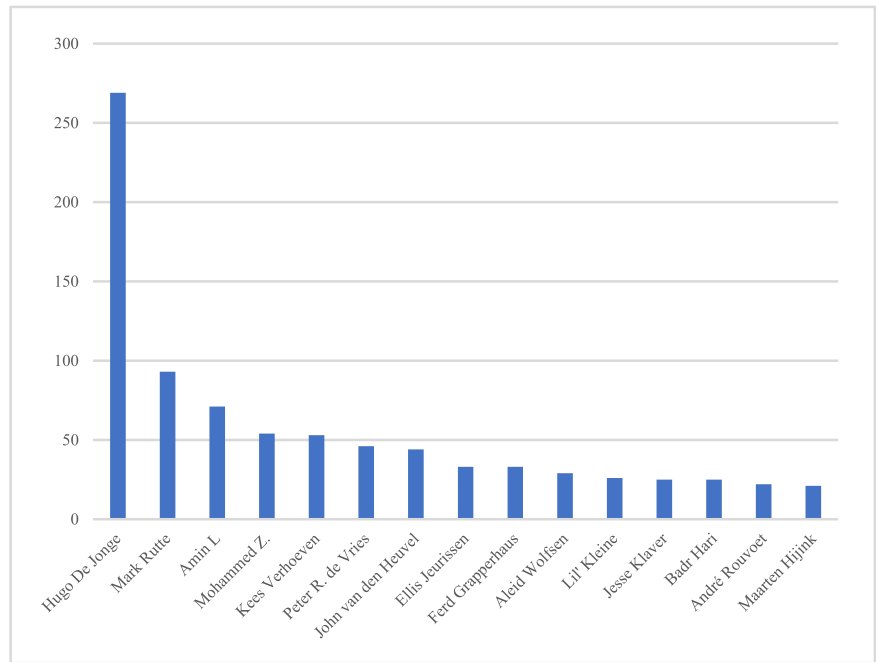


TABLE 3 Top 5 persons in news per crisis stage.

	Hugo de Jonge	Mark Rutte	Kees Verhoeven	Peter R. de Vries	John van den Heuvel
Crisis build-up	21	1	9	2	0
Crisis outbreak	125	38	28	3	2
Chronic stage	52	19	6	41	42
Termination stage	71	34	10	0	0



FIGURE 11 Organizations mentioned in the news.

more frequently mentioned organizations on X/Twitter. In the news reporting, it did not emerge as a particularly visible organization.

4 | DISCUSSION

The present paper set out to address the research question 'How can methods for automated text analysis be utilized for the tracking of media framing of an organizational crisis across different media types

and crisis stages?' To this end, it analysed the empirical case of the GGD data breach in the Netherland during the COVID-19 pandemic. The results show how computational approaches allow for charting media frames over time and how they shape different crisis stages. Concerning the role of media framing in crisis perception and crisis communication, the findings imply that it is important to consider the wider social-political context in which data breaches happen. That includes but also expands beyond the specific organizational context and the most immediate stakeholders in its sector of operations (or ecosystem). Gaining specific empirical insights on relevant media frames and stakeholders with automated methods can be here an important asset for crisis communication practitioners, who need to devise and adjust their strategies in extremely dynamic situations that involve a wide range of distinct yet interconnected media types and communicators. Many of the methods used in this study are readily available in form of media monitoring tools that often serve for, for example, social media marketing and reputation management.

The GGD data breach specifically was perceived and assessed in connection to governmental strategies for the wider COVID-19 pandemic. Arguably, journalists and commentators contributed to a 'crisis escalation' from an organizational context to a much broader societal level, where the GGD data breach fed into an already critical discussion of, for example, vaccination programmes, lock-down measures, and the manifold ways to which data were used for pandemic governance. Organizations need to be keenly aware of the wider context in which they are affected by a crisis. How are they connected to other organizations? Who is likely to be held responsible? What are other lingering issues that may be related to a new incident? It seems prudent for crisis communication professionals to have awareness for the interconnectedness and dynamism of crisis discourses. They should also consider the

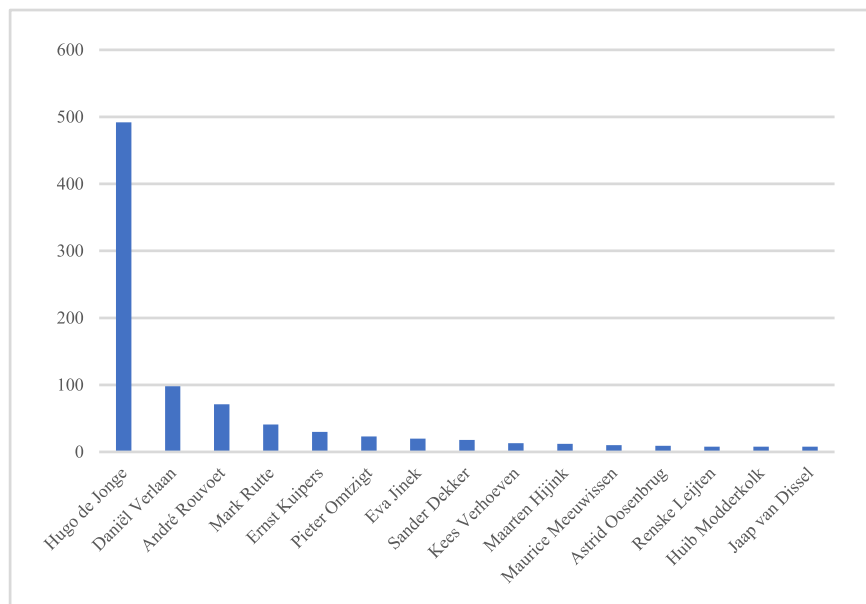


FIGURE 12 Top 15 actors X/Twitter.

TABLE 4 Top 5 persons in tweets per crisis stage.

	Hugo de Jonge	Daniël Verlaan	André Rouvoet	Mark Rutte	Ernst Kuipers
Crisis build up	1	0	0	0	0
Crisis breakout	449	79	60	27	6
Chronic stage	11	7	7	4	0
Termination stage	31	12	4	10	24



FIGURE 13 Organizations mentioned on X/Twitter.

relationship and differences between media outlets and online platforms. Each part of the wider digital public sphere asks for a different communication approach that caters to the needs and sensitivities of different media publics. Importantly, understanding dominant media commentators' agendas and motivations is recommendable in devising responses to their framing practices that are potentially harmful for an organization.

It is not unlikely that different data breaches go through similar stages in media discourses, at least in the context of public organizations: 1) 'Concerns about data use and possible data risks'

(*crisis build-up*), 2) 'breaking the news and immediate effects' (*crisis outbreak*), 3) 'political and public debate' (*chronic stage*), and 4) 'legal and political consequences' (*termination stage*). The proposed methodology revealed these stages of media framing through a comparative analysis of different media types that can be further expanded, if data sources are accessible (e.g., press releases, parliamentary debates, broadcasting subtitles, other social media). Commonalities and differences between outlets, platforms, and crisis stages can be researched on scale with relative ease using text analytical methods. However, data access is a major challenge. Especially social media data are increasingly difficult to retrieve but also news articles and other text sources need a viable strategy that may include both manual downloading and automated data collection, dependent on where relevant documents are located.

Furthermore, future studies should explore supervised machine learning and transformer-based models for text clustering (e.g., BERTTopics) for more fine-grained analysis into the nuanced use of language that factor into media frames. Examples are ways to measure 'sensationalism' in news headlines, the expression of emotions in textual form, or gauging not only *who* is cited but also to *what extent* (e.g., who has the most verbatim quotes?). It is also important to consider mixed-methods research designs that logically combine automated-quantitative approaches with qualitative steps. One important limitation of computational approaches is that they tend to remain descriptive. Here, methods from crisis communication research that place emphasis on inferential statistical analyses and hypothesis-testing may find useful application in follow-up research. Nevertheless, computational methods offer researchers the potential to analyse public discourses through media communication on a large scale within feasible timeframes, while a reducing costs and circumventing risks of cognitive overload. This can then provide the basis to develop hypotheses and models for how crises emerge, evolve, and are being perceived in the digital public sphere.

5 | CONCLUSION

This study critically examined media framing of the GGD data breach in news and social media throughout the crisis life cycle using computational text analysis methods. It aimed to explore how such methods can be utilized to research media framing and its role in public crisis perception, with implications for how organizations benefit from empirical insights to refine their crisis communication strategies.

The study has limitations: first, it examines a single data breach incident during the COVID-19 pandemic. Future research should broaden the scope to include multiple incidents across various contexts to identify commonalities and differences, aiding in the theorization of data breaches as organizational crises. Second, the focus was on media framing of the breach across crisis stages, without analysing the responses of the GGD and government to public reactions.

The methodological approach could extend to analysing organizational communication (e.g., press releases and public statements), to examine their impact on media framing and vice versa. Additionally, the study did not account for the political leanings of different news outlets; a more detailed analysis is needed to assess how these backgrounds influence media framing. Lastly, while this research began charting media framing of the data breach, future work should further explore the audience perspective, examining how citizens perceive and evaluate the breach to discern differences and similarities with media and social media discussions.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author.

ORCID

Dennis Nguyen  <http://orcid.org/0000-0001-6982-775X>

Sergül Nguyen  <http://orcid.org/0000-0002-1238-5206>

Phuong Hoan Le  <http://orcid.org/0000-0002-5406-1373>

Tessa Oomen  <http://orcid.org/0000-0002-2872-4006>

Yijing Wang  <http://orcid.org/0000-0002-7886-8366>

REFERENCES

- Akbik, A., Bergmann, T., Blythe, D., Rasul, K., Schweter, S., & Vollgraf, R. (2019). FLAIR: An easy-to-use framework for state-of-the-art NLP. In W. Ammar, A. Louis, & N. Mostafazadeh (Eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations)* (pp. 54–59). Association for Computational Linguistics. <https://www.aclweb.org/anthology/N19-4010>
- Bentley, J. M., Oostman, K. R., & Shah, S. F. A. (2018). We're sorry but it's not our fault: Organizational apologies in ambiguous crisis situations. *Journal of Contingencies and Crisis Management*, 26(1), 138–149. <https://doi.org/10.1111/1468-5973.12169>
- Bogdanowicz, A., & Guan, C. (2022). Dynamic topic modeling of Twitter data during the COVID-19 pandemic. *PLoS One*, 17(5), e0268669. <https://doi.org/10.1371/journal.pone.0268669>
- Bundy, J., Pfarrer, M. D., Short, C. E., & Coombs, W. T. (2017). Crises and crisis management: Integration, interpretation, and research development. *Journal of Management*, 43(6), 1661–1692. <https://doi.org/10.1177/0149206316680030>
- Burscher, B., Vliegthart, R., & Vreese, C. H. (2016). Frames beyond words: Applying cluster and sentiment analysis to news coverage of the nuclear power issue. *Social Science Computer Review*, 34(5), 530–545.
- Cheng, L., Liu, F., & Yao, D. (2017). Enterprise data breach: Causes, challenges, prevention, and future directions. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 7(5), 1–14. <https://doi.org/10.1002/widm.1211>
- Chong, D., & Druckman, J. N. (2007). Framing theory. *Annual Review of Political Science*, 10, 103–126.
- Coombs, W. T. (2007). Protecting organization reputations during a crisis: The development and application of situational crisis communication theory. *Corporate Reputation Review*, 10(3), 163–176. <https://doi.org/10.1057/palgrave.crr.1550049>
- Coombs, W. T. (2009). Crisis, crisis communication, reputation, and rhetoric. In R. L. Heath, E. L. Toth, & D. Waymer (Eds.), *Rhetorical and critical approaches to public relations II* (pp. 249–264). Routledge.
- Cross, C., Parker, M., & Sansom, D. (2019). Media discourses surrounding 'non-ideal' victims: The case of the Ashley Madison data breach. *International Review of Victimology*, 25(1), 53–69. <https://doi.org/10.1177/0269758017752410>
- D'Angelo, P., Lule, J., Neuman, W. R., Rodriguez, L., Dimitrova, D. V., & Carragee, K. M. (2019). Beyond framing: A forum for framing researchers. *Journalism & Mass Communication Quarterly*, 96(1), 12–30. <https://doi.org/10.1177/1077699018825004>
- Dhar, S., & Bose, I. (2022). Victim crisis communication strategy on digital media: A study of the COVID-19 pandemic. *Decision Support Systems*, 161, 113830. <https://doi.org/10.1016/j.dss.2022.113830>
- European Commission. (2016). *What is a data breach and what do we have to do in case of a data breach*. https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/obligations/what-data-breach-and-what-do-we-have-do-case-data-breach_en
- Fink, S. (1986). *Crisis management: Planning for the inevitable*. American Association of Management.
- Giddens, A. (1984). *The constitution of society: Outline of the theory of structuration*. University of California Press.
- Harcup, T., & O'Neill, D. (2017). What is news?: News values revisited (again). *Journalism Studies*, 18(12), 1470–1488. <https://doi.org/10.1080/1461670X.2016.1150193>
- Hart, P. S., Chinn, S., & Soroka, S. (2020). Politicization and polarization in COVID-19 news coverage. *Science Communication*, 42(5), 679–697.
- Jin, X., & Spence, P. R. (2021). Understanding crisis communication on social media with CERC: Topic model analysis of tweets about Hurricane Maria. *Journal of Risk Research*, 24(10), 1266–1287. <https://doi.org/10.1080/13669877.2020.1848901>
- Kepplinger, H. M., Geiss, S., & Siebert, S. (2012). Framing scandals: Cognitive and emotional media effects. *Journal of Communication*, 62(4), 659–681. <https://doi.org/10.1111/j.1460-2466.2012.01653.x>
- Krippendorff, K. (2011). Computing Krippendorff's alpha-reliability. <https://repository.upenn.edu/handle/20.500.14332/2089>
- Kuipers, S., & Schonheit, M. (2021). Data breaches and effective crisis communication: A comparative analysis of corporate reputational crises. *Corporate Reputation Review*, 25(3), 176–197. <https://doi.org/10.1057/s41299-021-00121-9>
- Langley, A. (1999). Strategies for theorizing from process data. *The Academy of Management Review*, 24(4), 691–710. <https://doi.org/10.2307/259349>
- Lopes, I. M., Guarda, T., & Oliveira, P. (2019). Implementation of ISO 27001 standards as GDPR compliance facilitator. *Journal of Information Systems Engineering & Management*, 4(2), 1–8. <https://doi.org/10.29333/jisem/5888>

- van der Meer, T. G. L. A. (2016). Automated content analysis and crisis communication research. *Public Relations Review*, 42(5), 952–961. <https://doi.org/10.1016/j.pubrev.2016.09.001>
- Nan, N., & Lu, Y. (2014). Harnessing the power of self-organization in an online community during organizational crisis. *Mis Quarterly*, 38(4), 1135–1157. <https://doi.org/10.25300/MISQ/2014/38.4.09>
- Nguyen, D. (2017). *Europe, the crisis, and the internet: A web sphere analysis*. Palgrave. <https://doi.org/10.1007/978-3-319-60843-3>
- Nguyen, D., & Hekman, E. (2022). The news framing of artificial intelligence: A critical exploration of how media discourses make sense of automation. *AI & Society*, 39, 437–451. <https://doi.org/10.1007/s00146-022-01511-1>
- Park, H. (2017). Exploring effective crisis response strategies. *Public Relations Review*, 43(1), 190–192. <https://doi.org/10.1016/j.pubrev.2016.12.001>
- Pentzold, C., & Fraas, C. (2023). Media frames as adaptive networks of meaning: A conceptual proposition. *Language & Communication*, 93, 95–106. <https://doi.org/10.1016/j.langcom.2023.09.001>
- Quekel, S. (2021, December 5). Megaclaim in de maak: Stichting eist miljarden van ministerie om datalek GGD. *Algemeen Dagblad*. <https://www.ad.nl/binnenland/megaclaim-in-de-maak-stichting-eist-miljarden-van-ministerie-om-datalek-ggd~a4fb05bf/>
- Robertson, A. (2010). *Mediated cosmopolitanism: The world of television news*. Polity Press.
- Scheufele, D. A., & Tewksbury, D. (2007). Framing, agenda setting, and priming: The evolution of three media effects models. *Journal of Communication*, 57(1), 9–20. <https://doi.org/10.1111/j.0021-9916.2007.00326.x>
- Semetko, H., & Valkenburg, P. (2000). Content analysis of press and television news. *Framing European Politics: A Journal of Communication*, 50(2), 93–109. <https://doi.org/10.1111/j.1460-2466.2000.tb02843.x>
- Spence, P. R., Lachlan, K. A., Lin, X., & del Greco, M. (2015). Variability in Twitter content across the stages of a natural disaster: Implications for crisis communication. *Communication Quarterly*, 63(2), 171–186. <https://doi.org/10.1080/01463373.2015.1012219>
- Stewart, M. C., & Wilson, B. G. (2016). The dynamic role of social media during hurricane# sandy: An introduction of the STREMI model to weather the storm of the crisis lifecycle. *Computers in Human Behavior*, 54, 639–646. <https://doi.org/10.1016/j.chb.2015.07.009>
- Sturges, D. L. (1994). Communicating through crisis: A strategy for organizational survival. *Management Communication Quarterly*, 7(3), 297–316. <https://doi.org/10.1177/0893318994007003004>
- Sung, M., & Hwang, J. S. (2014). Who drives a crisis? The diffusion of an issue through social networks. *Computers in Human Behavior*, 36, 246–257. <https://doi.org/10.1016/j.chb.2014.03.063>
- Syed, R. (2019). Enterprise reputation threats on social media: A case of data breach framing. *The Journal of Strategic Information Systems*, 28(3), 257–274. <https://doi.org/10.1016/j.jsis.2018.12.001>
- Verlaan, D. (2021, January 25). Illegale handel in privégegevens miljoenen Nederlanders uit coronasystemen GGD. *RTL Nieuws*. <https://www.rtlnieuws.nl/nieuws/nederland/artikel/5210644/handel-gegevens-nederlanders-ggd-systemen-database-coronit-hpzone>
- Vogler, D., & Meissner, F. (2022). Tackling the information overload. In W. T. Coombs, & S. J. Holladay (Eds.), *The handbook of crisis communication*. Wiley. <https://doi.org/10.1002/9781119678953.ch4>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Nguyen, D., Nguyen, S., Le, P. H., Oomen, T., & Wang, Y. (2024). Utilizing computational methods for analyzing media framing of organizational crises: The 'Datalek' scandal during the COVID-19 pandemic in the Netherlands. *Journal of Contingencies and Crisis Management*, 32, e12595. <https://doi.org/10.1111/1468-5973.12595>