# 2. Foundations of Content Analysis

The drastically changing nature of content in the move from print and elsewhere (e.g.,television) to the web has challenged the techniques and tools of content analysis, which, upon its inception, concerned itself mostly with large but static groupings of texts. Unlike modern print media such as newspapers or books, web content is often unstable and dynamic. It is also networked, which poses more problems for the researcher regarding the demarcation of the ‘text’ under study. Before further exploring this difference that technicity makes when aiming to do content analysis across the web, it is necessary to review the foundational status, methodologies, and tools of content analysis that existed as developed for (pre-web) mass media content. This chapter offers a historical perspective on the foundations of content analysis, discussing its scholarly roots and exploring how it has been modified as a field of research along with the changing technicities of content that it engages with. My historical reappraisal of the concepts and methods of content analysis considers first the work of Klaus Krippendorff, a groundbreaking content analysis scholar and, not coincidentally, a co-organizer of the first content analysis conference at Annenberg in 1969. After a brief examination of the foundational work by Berelson and Gerbner, I will come to describe Krippendorff’s seminal work *Content Analysis: An Introduction to its Methodology,* in which he lays out the requirements of a sound content analysis research framework.[[1]](#footnote-1)

Secondly, I will address the challenges this approach faces since the computer has become more of a content producer and site of production and publication, rather than merely a research aid for large-scale analyses of ‘texts’, broadly defined. Here, I will build on responses to the work of Krippendorff by communications and advertising scholar Sally McMillan and linguist and information scientist Susan Herring, who further developed Krippendorff’s techniques to grapple with the technical specificities of web content, which I refer to as its technicity. The term technicity, as described in the introduction, refers to the technologically composed nature of web content—the fact that content can hardly be separated from its carrier (a specific web platform for instance), and that technical agents such as hyperlinks and shares are not mere features, but *part of* the content under study.[[2]](#footnote-2)[[3]](#footnote-3) Accordingly, when looking at previous applications of content analysis to web content, I ask how the *technicity* may be made part of the definition, collection, and analysis of content being studied, which is the central question of this book.

Thirdly, I will ask whether content analysis should be enhanced to suit the analysis of networked and dynamic information online. Looking at the traditions in content analysis, a return to its roots may prove more productive. I would argue that conventional content analysis still holds valuable insights for current (online) approaches of web content. However, what needs to be explored are the necessary steps towards networked content analysis that takes the technicity of web content and the variety thereof as its point of departure. Lastly, I will describe how I will apply networked content analysis to study the issue of climate change, in the case studies in this book. I underline the importance of the issue for our day and age, but also describe strong preceding research in the content analysis of climate change content.

## Emergence of a Research Field

The field of content analysis considers its first seminal work to be that of Berend Berelson of 1952, titled *Content Analysis in Communication Research*, which describes content analysis as an important research technique for social scientists and media scholars for reading social and cultural change from (the analysis of) mediated messages.[[4]](#footnote-4) For example, in a study from 1948, Berelson and Salter study prejudice against minority groups through the analysis of popular magazine fiction.[[5]](#footnote-5) In the same tradition, as mentioned in the Introduction, George Gerbner has studied violence on television and the representation of for instance women and children during primetime programming, to derive *cultural indicators,* the indicators of their position in society at a given time.[[6]](#footnote-6)

Scholars often refer to the inclusion of the definition of ‘content analysis’ in Webster’s Dictionary of the English Language in 1961 as a milestone in the establishment and public recognition of the field. Here, content analysis was defined as the ‘analysis of the manifest and latent content of a body of communicated material (as a book or film) through classification, tabulation, and evaluation of its key symbols and themes in order to ascertain its meaning and probable effect’.[[7]](#footnote-7) In November of 1969, another milestone took place with the content analysis conference of the Annenberg School of Communications, where over 400 scholars gathered from approximately 85 educational and scientific institutions in the United States and Canada to discuss the application of content analysis to and from a wide range of academic disciplines.[[8]](#footnote-8)[[9]](#footnote-9) The conference also featured a panel dedicated to ‘Computer Techniques in Content Analysis and Computational Linguistics’, focusing solely on different ways in which content could be analyzed by the computer and by computer-aided content analysts. The scholars who presented computational analyses, in particular, at this inaugural event also came from a diverse set of fields, including ‘political science, psychiatry, sociology, English, and social psychology’.[[10]](#footnote-10)[[11]](#footnote-11) It is worth keeping these early, partially interdisciplinary beginnings in mind when negotiating contemporary applications of content analysis by different academic fields. With the more recent infusion of culture with information technology, content analysis’ early trajectory, as well as its focus on text and image analysis, merges with the interests of information science and allied fields in data-driven contemporary cultural analysis; this situation and convergence of practices and methods continues to create confusion about the possibilities of techniques for studying culture through content.

The most significant disciplinary figure of early content analysis,Klaus Krippendorff, defines content analysis as a ‘scientific tool’ and ‘a research technique for making replicable and valid inferences from text to the contexts of their use’.[[12]](#footnote-12) He deployed terms and concepts from outside the qualitative humanities normally concerned with content, like for example ‘scientific’, ‘replicable’ and ‘valid’, to emphasize the need for formalization of techniques and tools of analysis. At the same time, however, his use of the word *text* does not refer only to written materials but expansively may include ‘works of art, images, maps, sounds, signs, symbols, and even numerical records’ and other data.[[13]](#footnote-13) Krippendorff makes the significant conceptual point that it is precisely one’s definition of what content is, and how that is delimited, that leads to specific kinds of analytical results. As we will see with the analysis of networked content, it is indeed this refinement of definitions and approaches to the time and materiality of ‘content’ that needs to be amended. This is important for the recognition of the technicity as an active material agent and *part of* the content, rather than as a challenge that disturbs or supposedly renders difficult the demarcation and study of content online.

In other words, how one chooses to define content paves the way for specific research questions, methodological choices, and analytical consequences to play out over others. Content analysis, in this sense is not an entirely standardized or standardizable practice but is applied *across* scholarly disciplines that have used many different strategies of coping with the challenges posed by content on the web. Krippendorff dates this broadening as coinciding with some of the earliest applications of content analysis to the (further) growth of mass media after WWII. This rise of the field of content analysis to deal with expanded media formats, he argues, meant a loss of focus already then, as ‘everything seemed to be content analyzable and every analysis of symbolic phenomena became Content Analysis’.[[14]](#footnote-14) Krippendorff describes how various disciplines began to apply the research techniques of content analysis differently: ethnographers were interacting with their informants (something content analysts usually do not do, as they prioritize ‘unobtrusive' analyses) and also analyzing their own personal field notes as ‘content’, while social scientists studied educational materials to identify societal trends. At this point, Krippendorff develops a conceptual framework for content analysis that serves not only as a tool with which to (re-)establish a focus for this research methodology but also as a practical, analytical and methodological guide for researchers to *apply* the methods to diverse types of content. In the next section, I will describe this framework as introduced by Krippendorff and briefly reflect on how its components may hold in networked.

Krippendorff's framework lays out six components necessary for a content analysis research project, all of which are to be included though not necessarily in this sequential order:

- A body of text, the data that a content analyst has available to begin an analytical effort;

- A research question that the analyst seeks to answer by examining the body of text;

- A context of the analyst's choice within which to make sense of the body of text;

- An analytical construct that operationalizes what the analyst knows about the context;

- Inferences that are intended to answer the research question, which constitute the basic accomplishment of the content analysis;

- Validating evidence, which the ultimate justification of the content analysis.[[15]](#footnote-15)

Importantly, from the beginning point of his procedural outline, Krippendorff does not describe how content should be collected for well-formed content analysis to take place. The content to be analyzed seems not in question, in the sense that the text is already assumed to be accessible to the scholar (as, for example, a set of recent newspaper articles might be), demarcated, and readily available for study. The formulation of the research question and context narrows the broad scope of content analysis’ disciplinarity slightly more. (Again, the term ‘text' also refers to images, websites, music, etc.) In the next outlined step, Krippendorff emphasizes the importance of tailoring appropriate research questions, when stating that — in contrast to a deliberately open-ended interpretive approach to texts — strong research questions enable the researcher to read a text with more analytical distance. This allows the analyst not to just follow the author (in the Latourian sense described in the introduction) in what the actor says is in the text but instead, read off content with a specific question in mind. In this sense, the research question could also be described as a methodological tool in itself, with which to create a selection or sample of the data appropriate for answering the question.

As Krippendorff asserts, ‘[data] become[s] text to the analyst within the context that the analyst has chosen to read [it], that is, from within the analysis’.[[16]](#footnote-16) The analyst's background and scope and the research questions in combination provide the texts with a novel interpretive mechanism, within which they can be analyzed. A political scientist and an anthropologist might analyze the same piece of text differently, for instance. With regards to the analytical construct, Krippendorff stresses the importance of the *research* context in which a given text ‘would arguably make sense’.[[17]](#footnote-17) The analysis of text should be conducted in line with what is known about its uses. Krippendorf's fifth point constitutes the core of content analysis, in so far as the analysis enables the researcher to make inferences that scale appropriately. Krippendorff emphasizes the strength of abductive inferences — meaning those findings that are made across ‘logically distinct domains’ where multiple variables are taken into account — and compares this approach to the logic of reasoning employed by Sherlock Holmes, who uses clues to solve or sort through a larger reality and situation.[[18]](#footnote-18) For example, one can date a text by analyzing the vocabulary it uses, or infer the poignant issues of a city by studying letters sent to the municipality or local newspaper.[[19]](#footnote-19)

Lastly, and clearly in the interest of *not* letting abductive inferences over-reach, or otherwise become scientifically suspect, Krippendorff argues that all content analyses should be ‘validatable in principle’.[[20]](#footnote-20) Importantly for Krippendorff, this means there is a necessity to enable correlation of the research results with other data or information that stands *outside* the scope of the original analysis. The question of when the data requires a baseline outside of the content under study is one that resonates in the study of web content.[[21]](#footnote-21) In the realm of content analysis, this discussion has also taken place, including the suggestion of validating mass media content analysis (of culture) with audience interviews.[[22]](#footnote-22) For example, Gerbner on multiple occasions tried to correlate his Cultural Indicators research on violence in prime-time television with a survey on whether people also concurrently perceived the world as a violent place.[[23]](#footnote-23)

While the definition and demarcation of content were never that straightforward in the case of offline mass media materials, the rise of digital media has further complicated these matters. Digitization of content changed the nature of the materials already, raising new questions (e.g., Should column-width still be considered?). With hyperlinks, content became networked and thus harder to demarcate (Where does this content end?). Search engines brought about new ways of presenting and ranking data (What is the most important source?), and platformization gives shape to the far-stretching entanglement of social media with other web content.[[24]](#footnote-24)

As I will discuss in this chapter, the defining characteristics of web content pose new challenges to the above outlines, conditions, and expected consequences of what once fell under the purview of content analysis. To make a move into what I name networked content analysis, namely the application of content analysis on the web and the challenges thereof, it is important to engage with the challenges of this transition as these have been pre-conceived and processed by scholars identifying with the foundations of the field. This includes the work of Sally McMillan, who describes the study of web content as like looking at ‘a moving target through a microscope’.[[25]](#footnote-25) Web content in the late 1990s was in many respects different from web content in 2009 or 2014; this is a fact that should never be lost hold of. In the late 1990s, which is the period in which the studies McMillan reviews in her paper were conducted, the web did not yet have ‘platforms’ and was still in its early days of search engines and web browsers. Content was, however, already networked by hyperlinks and website *features*, which thus were the focus of many analyses of this period.

## Web Content Analysis: A Moving Target Seen Through a Microscope (McMillan)

In her article *Web Content Analysis: A Moving Target Seen Through a Microscope*, notably included in Krippendorff’s 2004 edited volume *Content Analysis: An Introduction to its Methodology*, McMillan takes stock of the challenges researchers face when applying traditional content analysis techniques to the study of web content. Interestingly, McMillan takes up certain directives from Krippendorff’s original content analysis framework to systematically track the present theoretical varieties of contemporary content analysis methods and theories in this paper. Firstly, McMillan compiles a collection of papers by searching the Social Sciences Citation Index (SSCI) for the keyword combinations ‘web’ and ‘content analysis’ as well as ‘internet’ and ‘content analysis’. Secondly, McMillan seeks papers from selected communication journals as well as communication conferences not indexed in the SSCI. Finally, she expands the list by checking the bibliographies of all the found sources, and adding relevant cited studies to the list. In all, she finds a total of nineteen studies dedicated to content analysis on the web and another eleven studies that are dedicated to the analysis of other digital content, such as email and ListServs, both of which were important online media at the time but which are not included in her final study.[[26]](#footnote-26)

Having collected her sources, McMillan relies on a research protocol close to Krippendorff's, checking each study for the resemblance of its components and methods to the original content analysis framework.[[27]](#footnote-27)[[28]](#footnote-28) She then compares the 19 articles to identify how the challenges of applying content analysis to web content research were being dealt with by each of the authors. From this, McMillan induced five steps that in her view, should be part of all web content analysis studies, and which should be compared to Krippendorff’s original list above:

1. Formulate the research questions and/or hypotheses;

2. Create a sample;

3. Further define categories:

(a) Establish the time period of the study, as web research calls for rapid collection of data,

(b) Identify context units;

(c) Develop coding units;

4. Train the coders and check the reliability of their coding skills;

5. Analyze and interpret data.

I want to discuss this paper in more detail because McMillan does try to address the issue of content collection that goes unstated in Krippendorff. Firstly, aiming to summarize how scholars collected their content, McMillan carefully lists the sampling strategies she has found in her list of 19 studies. She notes a wide variety of ways in which the researchers compiled their collections of websites to be analyzed. Most of the studies identify existing lists of reputable sources. In a footnote, McMillan issues a warning for web researchers using search engines, a novel tool at the time, stressing the importance of knowing as much as possible about how a search engine chooses and prioritizes its results before deciding how to use it for sampling. (I will more fully elaborate on this issue through the case study and argumentation of chapter 4 that deals with search engine results for the analysis of the position and resonance of climate skepticism on the web.)

Comparing McMillan's assembled lists of steps to the original provided by Krippendorff, one important component is now missing, which is validatability. This omission, I argue, very directly points to one of the key problems in using traditional content analysis methods without alteration in the analysis of web content: the fact that validation, which presumes an offline reference as a baseline, is not always possible in the analysis of digital and networked content. I would propose that an offline validation of online research in many cases is impossible. Thought-provokingly, scholars have asked in which cases the online *is* the only relevant baseline.[[29]](#footnote-29)[[30]](#footnote-30) Linguist and information scientist Susan Herring recognizes that web content is indeed a different kind of object compared to the pre-web content of Krippendorff’s time. In her 2010 paper, *Web Content Analysis: Expanding the Paradigm*, she calls for a widening of the research paradigms and methods attendant to web-oriented content analysis.

## Web Content Analysis: Expanding the Paradigm (Herring)

Rather than proposing wholly novel means of analysis, Herring proposes a combination of methods from various disciplines that can help the analyst to research the new kinds of content that occur on the web. Herring begins her contribution by noting the semantic differentiation between ‘web [content analysis]’, a narrower kind of research where traditional content analysis methods are applied to the web, and ‘[web content] analysis’, or what she calls WebCA, which is the analysis of web content in a broader sense, where various ‘traditional and non-traditional techniques’ can be applied.[[31]](#footnote-31) Herring promotes the latter by showing how traditional content analysis can be combined with methodologies from disciplines such as linguistics and sociology to offer a more workable response to the challenges offered by ‘new online media’.[[32]](#footnote-32) She illustrates this with examples from blog analyses and conversations online.

Herring's ‘more general’ definition of web content covers ‘various types of information “contained” in new media documents […], all of which can communicate meaning’.[[33]](#footnote-33) This definition is very similar to the earlier definitions of content analysis that were critiqued by Krippendorff, for the presumption that content is ‘contained in messages, waiting to be separated from its form and described’, as the true nature of content ‘resides *inside* a text’.[[34]](#footnote-34) The broadening that Herring proposes is, in fact, a return to another specific idea of content, where various content types (all of which Krippendorff would refer to as *text*) can each communicate meaning. The broadening of the paradigm in her paper's title refers on the one hand to the inclusion of the analysis of these various types of online content. In other words, besides the more traditional content *elements* that might be considered by content analysts, such as images, themes, and features, she includes a range of newer online-only (or: natively digital) elements, such as the hyperlink.

Furthermore, Herring argues that the research practice she denotes as ‘[web content] analysis’ would benefit from a broadening of its methodology, by including methods from other disciplines (see Figure 1). From sociology (and social network analysis), it is possible to attend to link analyses, from communication science (and content analysis) one can do feature analyses, and from linguistics (and discourse analysis) the contributing methodologies make it possible to produce computer-mediated discourse analysis. Rather than proposing medium-specific approaches to ‘web content analysis’, she proposes to broaden the methodological apparatus, by including other non-web-specific methods from different disciplines.

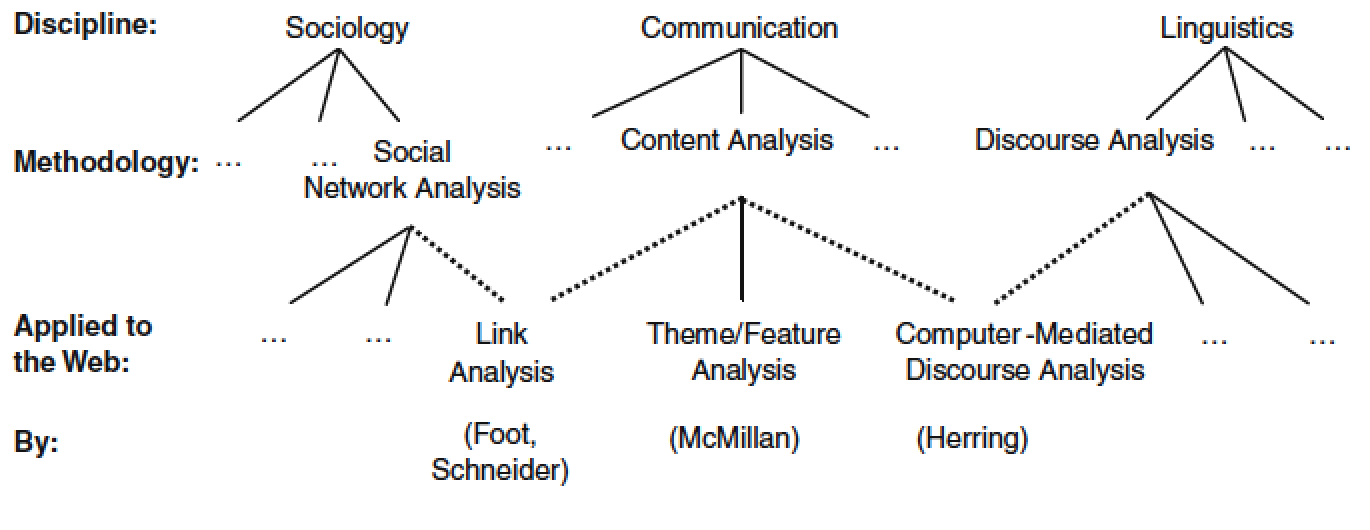


Figure 1: Widening of the content analysis paradigm. Herring's brief overview of approaches to web content analysis. Herring, ‘Web Content Analysis’, 240.

In her critique of McMillan's five-step research protocol, Herring argues that web content analysis follows ‘somewhat different norms from those traditionally prescribed for the analysis of communication content by researchers such as Krippendorff and McMillan’ and may even be developing new norms.[[35]](#footnote-35) She stresses that Krippendorff's framework also has been used rather liberally in content analysis practices. Furthermore, she notes, ‘a growing number of web studies analyze types of content that differ from those usually studied in CA — such as textual conversations and hyperlinks — using methodological paradigms other than traditional CA’.[[36]](#footnote-36) Herring offers a new list of five steps for web content analysis, or more specifically that of ‘computer-mediated discourse analysis’ (CMDA), which she initially developed in 2004.[[37]](#footnote-37) CMDA is described as ‘language-focused content analysis supplemented by a toolkit of spoken conversation and written text analysis’.[[38]](#footnote-38)

Herring's checklist for web content analysis is similar to that of McMillan but offers in her view a more ‘pragmatic’ point of departure:[[39]](#footnote-39)

a) Articulate research question(s);

b) Select computer-mediated data sample;

c) Operationalize key concept(s) in terms of discourse features;

d) Apply method(s) of analysis to data sample;

e) Interpret results.

Like Krippendorff and McMillan, Herring does not begin her procedures for content analysis with any specific mentioning of the exact means of collecting data but instead takes the data set to be something already given. Although the checklist may suggest that the research question would lead the analysis, at the same time she urges researchers to ‘choose a research question that is “empirically answerable from the available data”’.[[40]](#footnote-40) Herring also promotes flexibility in determining the sample types and coding categories based on the available data set. She builds a plea for a widening of the paradigms of content analysis, including the objects of such analyses, based on the assertion that most preceding approaches to content analysis focus on features and themes alone. She finds that in her own research practice of computer-mediated discourse analysis as applied to blogs, the research techniques of content analysis are indeed ‘well suited for analyzing structural features of blog interfaces’ and ‘analyzing themes represented in blog entries and comments’.[[41]](#footnote-41) Furthermore, although Herring rightly points out that the field of web content analysis nowadays extends beyond the use of conventional pre-web content analysis methods being merely applied to the web, it is clear that even this multi-disciplinary approach still attempts to separate content from its carrier. In this book, working beyond these concerted but insufficient attempts to update content analysis for the (changing) present media age, I want to show how and why this (persistent) separation of content and carrier can no longer hold with online networked content.

## Technicity of Content

As outlined in the introduction to this chapter, my emphasis on the technicity of content stems from the observation that web content is networked. The networked character of online content means that content now includes technical agents that network it, such as in-text hyperlinks, tags, and social buttons.[[42]](#footnote-42) Re-considering the early disciplined approaches of content analysis, we can see how networked content raises numerous methodological questions, many of which have been pointed out already, for instance in the above work of McMillan and Herring. When demarcating and collecting the relevant content at stake in analysis, one may wonder, for instance, where exactly the content of an article in an online newspaper ends. Should the hyperlinked pages be included in the study? How should social buttons be treated? Are all these links and buttons mere features to be counted and quantified, or should they be analyzed otherwise?[[43]](#footnote-43) My propositions for networked content analysis urge the analyst to move beyond the analysis of web page features to treat the particular technicities of content—exactly this complexity—as part of the *text* under study, as Krippendorff would phrase this. Only when we include these technical specificities in the analysis of content rather than attempting to separate content from its carrier, can we meaningfully apply still-key foundational content analysis techniques to natively digital content.

In line with Krippendorff, who states that the meaning of content emerges *through its analysis*, we could say here that the technicity of the content, and further, the algorithmic logic behind platforms (such as Twitter and Facebook) and search engines (like Google) that rank and organize content, both serve and give shape to this technicity while forming the unique *context* of web content. The fact that online content is networked and dynamic shapes the context, and in turn, the means of the analysis. In the last part of this chapter, I will give an example of technicities of content from the platforms I study through Digital Methods, methods in which the quantitative measures that are built into the medium are deployed for networked content analysis. Krippendorff’s sensitivity to the context of the text and the materiality thereof, which I observe to have receded in later content analysis methods formulations by scholars like McMillan and Herring, can from this point regain prominence for a networked content analysis.

There is no single common *type* of online content, as we have seen from McMillan and Herring’s attempt at an overview, alongside many other attempts, and as is evident from the examples of different types of web content I provide in the case studies that follow.[[44]](#footnote-44) Rather than emphasizing the pluriformity of the web’s content ‘types’, I would like to *productively* distinguish between various platforms with which content analysis must come to terms.[[45]](#footnote-45) Platforms are ‘portals or applications that offer specific Internet services, frameworks for social interaction, or interfaces to access other networked communications and information distribution systems’.[[46]](#footnote-46) Many researchers have described how the Internet can easily be observed to be changing into a constellation of platforms, which are fast becoming our main means of accessing online information.[[47]](#footnote-47) This tendency adds to the urgency for content analysis approaches to be able to deal with platform-specific aspects of content.

The approach of networked content analysis that I put forward, given these above considerations, is based on two overarching principles. The first is that web content is increasingly accessed and organized through search engines and platforms. The second principle is that the technicity of content should be part of the analysis of such networked content proposed. In this way, I consider techniques of content analysis that are inclusive of the specificity of the *platform* in networked content analysis, and that enable the researcher to study content, with an enhanced literacy for its dimensionality and movement, within and through the technical specificities and cultures of online content in context. This entails analytical sensitivity that recognizes that each platform networks, handles and serves content differently, for instance, search engines serving search results in a ranked list, Wikipedia cleaning and organizing its content with robots, and Twitter linking content through hashtags.

## Networked Content Analysis With (or as) Digital Methods

Perhaps the most significant difference of emphasis, also from Krippendorf, that I am making in the proposition for networked content analysis is for the research question to lead to the collection of data or to a specific query within an existing data set, rather than the other way around. To emphasize this research need on the level of methodology and protocol is clearly quite contrary to the pre- and early-digital methods of content analysis (as shown in the research protocols earlier in this chapter). Networked content analysis can start with a question involving a set of actors in a specific issue, as I engage climate change skeptics (detailed in Chapter 3), and in a Latourian way follow these actors across platforms and sources, looking at their resonance, their language, and their networks. Such an approach to online content is partly drawing on the techniques and strengths of issue mapping, the multidisciplinary research practice described in the Introduction, where the objects of study are ‘issues’ themselves, and where analysis may include how these issues manifest online, within specific platforms. Issue mapping can follow a topic as it traverses sources, for example, or capture multiple online spaces in a comparative analysis. An example of this is offered by Climaps, where a mapping of the issue of climate change across sources and platforms resulted in an online atlas of climate change adaptation.[[48]](#footnote-48)

Given that this demarcation of content is such an essential part of networked content analysis research, much attention needs to be paid to the design and fine-tuning of search strings when using engines and related tools. Clarifying refined queries for specific source sets enables the researcher to answer the research questions with their gathered data. Rather than using predefined categories or translating jargon into more familiar terms, such inquiries also aspire to ‘follow the actors’ in their own (issue).[[49]](#footnote-49) Thus, research queries respect the terms employed by the actors. Source sets may be conventional, such as from leading environmental or human rights organizations' public data, or they may be derived more directly from web engines or platforms, e.g., the leading organizations according to a search engine query, or the sub-issues resonating in a set of tweets.

Critical views on issue mapping with digital methods highlight the problem of the methods’ and tools’ dependency on already problematic proprietary walled gardens, and otherwise volatile ever-innovating commercial web platforms, such as Facebook and Twitter.[[50]](#footnote-50) Scholars particularly warn of the sheer impossibility of distinguishing between the working logic of web platforms and the exemplarity of ‘platform artifacts’.[[51]](#footnote-51)[[52]](#footnote-52) For example, the most ‘retweeted’ content on Twitter might be the most Twitter-friendly content; therefore, we may only be finding out more about the logic of the platform itself, rather than the issue under study or the eventfulness of a particular tweet.[[53]](#footnote-53) Consequently, when dealing with online content, we need to take into account the socio-technical logic of the platform itself as part of any analysis.[[54]](#footnote-54) In fact, with the explosive rise of (big) data, attention to socio-technical logics of platforms must be further prioritized as social research increasingly makes use of what is called *Live Research*, where masses of content (with specific forms and technicities) are aggregated in real-time, copied onto other networks, and archived across the (social) web.[[55]](#footnote-55) [[56]](#footnote-56) Furthermore, data analysis and the tools that enable this are built on highly dynamic web services. In a critique of the famous Google Flu Trends project, David Lazer et al. write how Twitter, Facebook, Google, and the Internet more generally are continually changing because of the actions of millions of engineers and consumers.[[57]](#footnote-57) Understanding and studying these platforms as socio-technical systems for what they are, is of utmost importance, as they are ‘increasingly embedded in our societies’.[[58]](#footnote-58) In this book, I develop such a socio-technological perspective on the controversy surrounding climate change as presented and debated on the web.

Consequent to the process of data collection, and then the querying of that data through refined search queries, the decision to visualize the data arises. Visualization here is not a mandatory step in the analysis. However, it can be considered an applied tool for the purpose of visual and descriptive analysis. While the ‘descriptive turn’ has been ‘embraced’ in contemporary sociology, it does come with its own ethical questions, if you will.[[59]](#footnote-59) Each time a map is made, the researcher has to consider the appropriate output of the analysis ‘in ranked lists, in cluster graphs, in line graphs, in clouds, on maps’ and on a more abstract level, the visual, critical and even political aspects of map-making in their work.[[60]](#footnote-60)[[61]](#footnote-61) Sociologist Tommaso Venturini, when discussing controversy maps, has described social maps as a visual interface to complex issues: ‘To be of any use, social maps have to be less confused and convoluted than collective disputes. They cannot just mirror the complexity of controversies: they have to make such complexity legible.’[[62]](#footnote-62) Similarly, visualization of data layered onto a geographic map of an area should render legible the complexity of the area, as well as the ways in which the social media platforms from which the data is taken from actually deal with that geo-location. It must be constantly borne in mind that map-based visualizations have been criticized for their oversimplification and reductionist approach to vast and multifarious data, highlighting some information and obfuscating other data for the sake of creating a ‘display [of] what we already know’.[[63]](#footnote-63)

I would, therefore, like to stress that in this book and in related research, the practice and objects of mapping are not efforts to ignore the distributed nature of today’s technologies or data, or the richness of public debate, but in fact to gain a better understanding of the complex patterns and intersections of competing technologies as they intertwine form and content.[[64]](#footnote-64) These maps then function as a navigational tool through a complex debate, rather than aiming at a reductionist narrative.[[65]](#footnote-65) The map is neither the end product nor an aesthetic inquiry into the data. Here, the visualizations function as an analytical tool.[[66]](#footnote-66) The maps then enable researchers to essentially zoom out, navigate the issue, and decide the directions for further analysis. I endeavor to accomplish this here by studying a single issue across multiple platforms from different viewpoints and by creating not one all-encompassing ‘mother map’, but a series of different maps and descriptions, of variegated utility, which underline both the complexity of studying issues through online content and the entanglement of content with its technicity. Of course, offline mass media content will also be present in these maps and analyses, as news and other media (which have traditionally been subject to content analysis) are referred to in and thus form part of online networked content.

When applied to the study of controversy, as in this case the climate change debate, the key contribution of networked content analysis lies in the development of adaptive research techniques that are rooted in content analysis while suited to networked digital media content. These methods allow researchers to follow debates and actors across diverse sources and online platforms. In the next chapters, I will operationalize such an approach, in which I discuss first how content is networked (on the web and accessed through Google, Wikipedia and, lastly, Twitter) and then address a question considering a specific aspect of the climate change debate. In the case of the web, I assess the place and status of climate change skeptics, within climate science and on the web. Are they professional climate experts, or professional skeptics? I operationalize this question by taking a set of key actors and profiling them, if you will, by assessing their prominence within climate science, their networking behavior, their resonance in search engine results for the issue of climate change and, lastly, by appraising and discussing their ‘related content’. In the case of Wikipedia, a network of interlinked articles on climate change and global warming allows for a reconstruction of the debate over time. Lastly, through the platform of Twitter, I provide a comparative view of the different stages of climate change (skepticism, mitigation, adaptation, and vulnerability/conflict), and explore the sub-issue of climate vulnerability in detail.

Conclusions  
Content analysis has made longstanding contributions to the broadest definitions of mediated ‘textual’ analysis, but when applied to networked content evokes a myriad of analytical issues: demarcating the object of study (where does a website end?), dealing with the dynamic character of the web (how can you redo the research, when the object of study constantly changes?), dealing with the unknown algorithms of search engines (how does one rely on Google without knowing its exact algorithm?) and so on. Where some content analysts, such as McMillan, prefer to stay close to the foundations of content analysis, others, such as Herring, make a plea for the widening of this research paradigm and its object of study, through the inclusion of methods from adjoining scientific fields. However, while Herring regards content as contained in media documents, I argue that the separation between content and its carrier no longer holds with networked content.

As Krippendorff pointed out, it is the specificity of the definition of ‘content’ one chooses that leads to specific kinds or varieties of content analysis.[[67]](#footnote-67) The inclusion of web content’s technicity into the idea of content itself leads to analyses that make use of and deal analytically with these technical agents. As I have demonstrated in this chapter, the collection and analysis of web content that follow the specificities of each platform and operationalizes the specific technicities at play will lead to more precise analysis, one that is sensitive to the networked nature and dynamical movement of online content. I have here realigned my work with Krippendorff’s inceptive call to keep the content together with its carrier (or context), and accordingly propose that in networked content analysis researchers include not only the carrier (e.g., the Wikipedia article, the search engine result, the Tweet) but also the technicity (e.g., the editing history and content robots of the Wikipedia article, the ranking of the search results, the hashtags and retweets networking the collection of tweets) as part of their analytical approaches.

For the collection of content in early instituted content analysis methods, the data available was always shaping or assumed to be setting up the point of departure for research. In other words, research questions enabled the researchers to sample more specific queries from that available data only. In the methods proposed here, I assert the value of working the other way around. The collection of data occurs *after* the research questions are formulated, and starts with the careful composition of source lists that are to be queried. After the sources are collected, and the spreadsheets are in place, the queries for the content sphere’s dominant engine are designed, tested and, if necessary, tweaked. Subsequent analysis of the content under study often comes with a map or visualization of the data.

The way forward presented here is a first step in the description of the contribution of medium-specific digital methods to the field of content analysis.Of course, it will need and welcome further elaboration, and, to stay in line with traditions of content analysis, should offer both a description of the approach on a theoretical, conceptual and historical level and eventually also hands-on guidelines that lay out the recipe for a solid project of content analysis. Clearly, I am valuing and making progress towards tools and methods for networked content analysis that stay tied to the inceptive work of Krippendorff. In line with his thinking, a contemporary web-literate approach titled networked content analysis remains open to all kinds of content and includes contents’ technical specificities in the value of such. The case studies in the next chapters offer such medium-specific approaches to climate change content on the web (and Google), Wikipedia, and Twitter to ask which methods might be further tailored towards platform-specific ends, and which can be scaled from or between platforms.

1. In this chapter, I will refer mostly to the second edition published in 2004, as this was thoroughly revised to describe the analysis of ‘computer readable’ content and presents a more mature method and technique of content analysis since the first edition of 1980. K. Krippendorff, *Content Analysis: An Introduction to its Methodology*, second edition, Thousand Oaks: Sage Publications, 2004, xiv. I will occasionally refer to the third edition of 2013, e.g., when addressing recent discussions or techniques not included in the previous editions. K. Krippendorff, *Content Analysis: An Introduction to its Methodology*, third edition, Thousand Oaks, CA: Sage Publications, 2013. [↑](#footnote-ref-1)
2. S. Niederer and J. van Dijck, 'Wisdom of the Crowd or Technicity of Content? Wikipedia as a Sociotechnical System', *New Media & Society* 12.8 (2010): 1368–1387. [↑](#footnote-ref-2)
3. S. Niederer and J. van Dijck, 'Wisdom of the Crowd or Technicity of Content? Wikipedia as a Sociotechnical System', in M. David and P. Milward (eds) *Researching Society Online*, London: Sage, 2014. [↑](#footnote-ref-3)
4. B. Berelson, 'Content Analysis in Communication Research', 1952, http://psycnet.apa.org/psycinfo/1953-07730-000. [↑](#footnote-ref-4)
5. B. Berelson and P.J. Salter, 'Majority and Minority Americans: An Analysis of Magazine Fiction', *The Public Opinion Quarterly*, 10 (1948): 168–190. [↑](#footnote-ref-5)
6. Annenberg School for Communication, *George Gerbner Archive*, University of Pennsylvania, 2006. [↑](#footnote-ref-6)
7. A. Merriam-Webster, *Webster’s New Collegiate Dictionary*, G.&C. Merriam Company, Publishers, 1961. [↑](#footnote-ref-7)
8. Presently called Annenberg School for Communication. ‘Annenberg School for Communication’, https://www.asc.upenn.edu/. [↑](#footnote-ref-8)
9. G. Gerbner, O. Holsti, K. Krippendorff, W.J. Paisley, and P.J. Stone, eds. *The Analysis of Communication Contents: Development in Scientific Theories and Computer Techniques*, Wiley, 1969, xiii. [↑](#footnote-ref-9)
10. Stone in Gerbner et al. *The Analysis of Communication Contents,* 335. [↑](#footnote-ref-10)
11. It is worth mentioning here that at this historical moment, the computer being brought to work on content analysis was, specifically, a machine reading text from punch cards or microfilm, or otherwise dealing with content ‘typed in from a computer console’. Accordingly, the approaches to content analysis presented were often captured in pieces of software and developed in different ways that directly reflected the specific state of the technology. Some approaches were programmed by the scholars themselves or programmed by others, including technicians, under close supervision from scholars, while yet other scholars completely outsourced programming responsibilities in full. Stone in Gerbner et al. *The Analysis of Communication Contents,* 336. [↑](#footnote-ref-11)
12. Krippendorff, *Content Analysis,* 2004, 24. [↑](#footnote-ref-12)
13. Krippendorff, *Content Analysis*, 2013, 25. [↑](#footnote-ref-13)
14. Krippendorff, *Content Analysis,* 2004, 12. [↑](#footnote-ref-14)
15. Krippendorff, *Content Analysis,* 2004, 29-30. [↑](#footnote-ref-15)
16. Krippendorff, *Content Analysis,* 2004, 33. [↑](#footnote-ref-16)
17. Krippendorff, *Content Analysis,* 2004, 35. [↑](#footnote-ref-17)
18. Krippendorff, *Content Analysis,* 2004, 38. [↑](#footnote-ref-18)
19. Krippendorff, *Content Analysis,* 2004, 42. [↑](#footnote-ref-19)
20. Krippendorff, *Content Analysis,* 2004, 39. [↑](#footnote-ref-20)
21. See for instance R. Rogers, F. Janssen, M. Stevenson, and E. Weltevrede, 'Mapping Democracy', in *Global Informaton Society Watch*, The Hague: Hivos, 2009, pp. 47-57. [↑](#footnote-ref-21)
22. Krippendorff, *Content Analysis,* 2013, 44. [↑](#footnote-ref-22)
23. G. Gerbner, L. Gross, N. Signorielli, M. Morgan, and M. Jackson-Beeck, 'The Demonstration of Power: Violence Profile No. 10', *Journal of Communication* 29.3 (1979): 177–196. [↑](#footnote-ref-23)
24. See also Helmond, *The Web as Platform.* [↑](#footnote-ref-24)
25. McMillan, ‘The Microscope and the Moving Target’, 80. [↑](#footnote-ref-25)
26. McMillan, ‘The Microscope and the Moving Target’, 88. [↑](#footnote-ref-26)
27. McMillan, ‘The Microscope and the Moving Target’. [↑](#footnote-ref-27)
28. Krippendorff, *Content Analysis,* 2004. [↑](#footnote-ref-28)
29. D. Moats, 'From Digital Methods to Digital Ontologies: Bruno Latour and Richard Rogers at CSISP', 2012, http://www.csisponline.net/2012/03/12/from-digital-methods-to-digital-ontologies-bruno-latour-and-richard-rogers-at-csisp/. [↑](#footnote-ref-29)
30. R. Rogers, *The End of the Virtual: Digital Methods*, Amsterdam: Vossiuspers UvA, 2009. [↑](#footnote-ref-30)
31. Herring, ‘Web Content Analysis’, 235. [↑](#footnote-ref-31)
32. Herring, ‘Web Content Analysis’, 246. [↑](#footnote-ref-32)
33. Herring, ‘Web Content Analysis’, 245 [↑](#footnote-ref-33)
34. Krippendorff, *Content Analysis,* 2004, 20. [↑](#footnote-ref-34)
35. Herring, ‘Web Content Analysis’, 237. [↑](#footnote-ref-35)
36. Herring, ‘Web Content Analysis’, 238. [↑](#footnote-ref-36)
37. Herring, ‘Web Content Analysis’, 238. [↑](#footnote-ref-37)
38. Herring, ‘Web Content Analysis’, 238. [↑](#footnote-ref-38)
39. Herring, ‘Web Content Analysis’, 238. [↑](#footnote-ref-39)
40. Herring, ‘Web Content Analysis’, 238. [↑](#footnote-ref-40)
41. Herring, ‘Web Content Analysis’, 241. [↑](#footnote-ref-41)
42. This term ‘agents’ implies that these pieces of content have agency, which I argue is indeed the case. These technical specificities not only present or structure text differently, they are also co-authoring the text. The chapter on Wikipedia will provide examples of this, when I zoom in on the activity of software robots authoring and editing articles. [↑](#footnote-ref-42)
43. Similar questions arise in the research (and practice) of web archiving, where national libraries and other organizations aim to demarcate and archive a ‘national web’. Similarly, internet censorship tries to demarcate ‘forbidden content’, and grapples with similar questions (see our study on the Iranian web: R. Rogers, E. Weltevrede, S. Niederer, and E. Borra, 'National Web Studies: The case of Iran', in J. Hartley, J. Burgess and A. Bruns (eds) *Blackwell Companion to New Media Dynamics*, Oxford: Blackwell, 2013, pp. 142-166.). [↑](#footnote-ref-43)
44. For an example of another attempt, see e.g., Weare and Lin, ‘Content Analysis of the World Wide Web’. [↑](#footnote-ref-44)
45. The idea of content segmentation has been popular in Internet marketing since the early 2000s, where it refers to the segmentation of content within one website, to attract various audiences. See for instance 'Content Segmentation: Differentiate Your Brand Online', 5 April 2012, http://contentmarketinginstitute.com/2012/04/use-content-segmentation-to-differentiate-your-brand/. [↑](#footnote-ref-45)
46. Platform Politics, '*Platform Politics: Call for Papers: A Multidisciplinary Conference*', Cambridge, UK, 2011, http://www.networkpolitics.org/content/platform-politics-call-papers. [↑](#footnote-ref-46)
47. Helmond, *The Web as Platform.* [↑](#footnote-ref-47)
48. EMAPS, 'Climaps: A Global Issue Atlas of Climate Change Adaptation', 2014, http://climaps.eu/. [↑](#footnote-ref-48)
49. Latour, *Reassembling the Social.* [↑](#footnote-ref-49)
50. J. van Dijck, *The Culture of Connectivity: A Critical History of Social Media*, New York, NY: Oxford University Press, 2013. [↑](#footnote-ref-50)
51. Marres, ‘Why Map Issues?’. [↑](#footnote-ref-51)
52. N. Marres and E. Weltevrede, 'Scraping the Social? Issues in Live Social research', *Journal of Cultural Economy* 6.3 (2013): 313–335, Rogers, *Digital Methods.* [↑](#footnote-ref-52)
53. Marres, ‘Why Map Issues?’ [↑](#footnote-ref-53)
54. Niederer and van Dijck, 'Wisdom of the Crowd or Technicity of Content?’ [↑](#footnote-ref-54)
55. Back et al. ‘Doing Real Time Research’. [↑](#footnote-ref-55)
56. Marres and Weltevrede, ‘Scraping the Social?’ [↑](#footnote-ref-56)
57. D. Lazer, R. Kennedy, G. King, and A. Vespignani, 'The Parable of Google Flu: Traps in Big Data Analysis', *Science* 343 (2014): 1205. [↑](#footnote-ref-57)
58. Lazer et al. ‘The Parable of Google Flu’, 1205. [↑](#footnote-ref-58)
59. M. Savage, 'Contemporary Sociology and the Challenge of Descriptive Assemblage', *European Journal of Social Theory* 12.1 (2009): 158. In this article, Savage makes a strong case for visualization research, stating that ‘there needs to be more sociological interest in visualization as process, social artifact, and research tool’. [↑](#footnote-ref-59)
60. Digital Methods Initiative, *DMIR Unit #5: Cross-Platform Analysis*, Amsterdam: University of Amsterdam, 2015. [↑](#footnote-ref-60)
61. D. Wood and J. Fels, *The Power of Maps*, New York: The Guilford Press, 1992. [↑](#footnote-ref-61)
62. T. Venturini, 'Building on Faults: How to Represent Controversies with Digital Methods', *Public Understanding of Science* 21.7 (2010): 797. [↑](#footnote-ref-62)
63. G. Lovink, *Social Media Abyss: Critical Internet Cultures and the Force of Negation*, Cambridge, UK: Polity Press, 2016, 152. [↑](#footnote-ref-63)
64. See also S. Niederer, G. Colombo, M. Mauri, and M. Azzi, 'Street-Level City Analytics: Mapping the Amsterdam Knowledge Mile' in *Hybrid City 2015: Data to the People*, Athens: University of Athens, 2015, www.media.uoa.gr/hybridcity. [↑](#footnote-ref-64)
65. As some of my research was part of the EMAPS EU-project (2014), I’d like to refer here the way in which the analysts of the program articulated their practice of mapping while showing full awareness of the value of these emerging critiques, which can be read in *Climaps*, the collaborative issue atlas of climate change adaptation produced as the result of EU FP7 project EMAPS (with principal investigator Bruno Latour, Sciences Po) and Digital Methods at the University of Amsterdam collaborated with international parties (Barcelona Media, Politecnico di Milano, the Young Foundation, and the Dortmund Technische Universität) in mapping the issue of climate adaptation. EMAPS, ‘Climaps.’ [↑](#footnote-ref-65)
66. S.K. Card, J.D. Mackinlay, and B. Shneiderman, *Readings in Information Visualization: Using Vision to Think*, San Francisco, CA: Morgan Kaufmann, 1999. [↑](#footnote-ref-66)
67. Krippendorff, *Content Analysis,* 2004. [↑](#footnote-ref-67)