Big Data in Public Relations: A Conceptual Framework

Christian Wiencierz, Ph.D.
Research Assistant, Department of Communication,
University of Münster, Germany
christian.wiencierz@uni-muenster.de

Ulrike Röttger, Ph.D.
Professor, Department of Communication,
University of Münster, Germany
ulrike.roettger@uni-muenster.de

Abstract

The article reports the results of an interdisciplinary research project on the implementation of big data analysis in public relations. It presents a framework about the ideal-typical application of big data in strategic communications in which the phases of the communication planning process were linked with the data-generation and data-analysis approach Knowledge Discovery in Databases (KDD) used in (business) information systems. The results of two studies — in which guided in-depth interviews with 19 experts on big data and with 18 experts on social media analytics were conducted — show that the seven typical steps of big data applications are also applicable in public relations. The results confirm the theoretically derived framework and illustrate the enormous potential of big data for public relations, as well as the associated challenges. The description of an ideal-typical application process is an important step in the systematic exploration of the still-neglected research area of big data in public relations. This article is a contribution to a long-overdue scholarly discussion about the importance of big data for public relations and strategic communication in general.

Keywords: Corporate Communication, Strategic Communication, Big Data, Knowledge Discovery in Databases (KDD)

Introduction

Public relations is facing an irreversible revolution caused by big data. This is the assertion of Weiner and Kochhar (2016), claiming that big data is transforming the public relations profession. This means that communicators can use big data to better understand their stakeholders and competitors, and thus optimize their own performance. Whether a significant improvement in the planning, implementation, and evaluation of public relations measures by big data is a revolution or rather a significant step forward is debatable. The fact is: Despite the generally acknowledged importance of big data for organizations and their communications, scholarly research has thus far

largely ignored the potentials and challenges of big data for strategic communication in general and public relations in particular (Wiencierz & Röttger, 2017). We see strategic communication as a generic term for any kind of intentional-purposeful, persuasive, public use of communication by an organization to fulfill its mission (Hallahan, Holtzhausen, van Ruler, Verčič, & Sriramesh, 2007). Public relations as communicative management of the socio-political environment (political-administrative and sociocultural publics) with the aim to ensure scope for action and to legitimize concrete strategies through social integration is a form of strategic communication, alongside marketing communication and internal communication (Röttger, Preusse, & Schmitt, 2014, p. 190; Zerfass, 2008). Apart from a few use cases, the question how big data can be used in public relations is still largely unexplored. Due to the increased importance of big data, public relations must become more involved in this issue, especially given that disciplines such as marketing are ahead of public relations in this area (Wiencierz & Röttger, 2017).

There is also a lack of process models showing how big data can be ideal-typically applied in communication. Such an ideal-typical model of the application process can provide public relations practitioners with valuable guidance while they simultaneously advance systematic research in this area. By clarifying the individual steps of complex big data applications, possible approaches for public relations research can be identified. Insights into the individual steps of the application clarify the process as a whole. This article therefore starts with describing a seven-step, ideal-typical application process of big data in strategic communication — which can also be applied to public relations, as our study shows.

The ideal-typical application process was derived on the basis of previous scholarly models used in (business) information systems, which show how large information assets can be utilized efficiently and effectively, and how knowledge stored in databases can be discovered (Fayyad, Piatetsky-Shapiro, & Smyth, 1996a, 1996b). Big data is understood here as large information assets characterized by high volume, high velocity, high variety, and high veracity, generated by innovative computer and digital storage systems in such a way that organizations and individuals can use and analyze them. The term includes concepts, technologies and tools to generate, store, manage and analyze large heterogeneous amounts of structured and unstructured data, like customer data, user comments or audiovisual data with different formats, structures, and semantics systematically in high speed (e.g., Beyer & Laney, 2012; Scholz, 2017, pp. 12–20; Wiencierz, 2016). Big data is a relational term, the exact meaning of which depends on the size of the data-using organization, the industry in which it operates, and the ongoing evolution of data management and analysis systems that make it possible to generate and evaluate increasing volumes of data faster and faster (Gandomi & Haider, 2015).

Our process model provided the basis for the deductive formation of categories for the analysis of two studies, in which guided in-depth interviews with 19 experts on big data and 18 experts on social media analytics were conducted. The results show that generating knowledge from big data is a demanding, complex process that will change and optimize public relations. It leads to a new way of planning, implementing, and controlling communication processes. Transforming data into insights and knowledge for the communication planning process is not a linear process; it involves ongoing feedback processes. The framework we present provides valuable implications for public relations practice, which, along with the limitations of the studies, we will discuss in our conclusion.

Theory

Four Phases of Strategic Communication

Traditional conceptual models of both strategic communication in general and public relations in particular usually comprise four phases of problem solving: *situation analysis, strategy, implementation,* and *evaluation* — although the models vary in detail (Röttger et al., 2014, pp. 184–187). Other and new forms of planning and control are currently being discussed in research on strategic communication. Van Ruler (2015) questions whether an understanding of strategy, planning, implementation, and evaluation according to traditional conceptual theories is still appropriate in today's digitalized society. There is rather a need for new, iterative methods that enable flexible communication planning and in which change is a defining part during the communication planning process. She suggests using the agile method Scrum in public relations work. Despite this differentiated debate on strategy and planning, traditional systematization is helpful to demonstrate that different goals and methods for data application emerge from the strategic public relations planning process. That said, there is no question that the division into these four phases is artificial and agility is gaining in importance in public relations planning, especially in the increasingly fast digital communication environment.

Typically, the strategic public relations communication process begins with the situation analysis. Primary and secondary information is used to examine which public relations problems must be solved, which communicative tasks can be derived from the problem analysis, which organizational resources are available, and which external conditions for action exist. In strategy development, communication goals must be formulated on the basis of the communicative mission and the results of the situation analysis. Furthermore, target groups must be identified and tactics formulated, including the definition of concrete public relations measures and instruments for accomplishing goals, as well as time and responsibility schedules. In the implementation phase, the planned measures become specified enough to make them workable. Success monitoring finally takes place after completion of the public relations measures, but can also be carried out during the process. It includes the analysis of how effective which measures were, whether the communicative task has been accomplished, and whether KPIs, such as reputation, have been achieved (Macnamara & Likely, 2017; Röttger et al., 2014, pp. 184–187).

The Process of Big Data Analysis

In order to capture the complexity of data generation and processing in public relations theoretically, we considered the *Knowledge Discovery in Databases (KDD)* approach by Fayyad et al. (1996a, 1996b), which originated in (business) information systems. In information systems, there exist well-known, established technology- and analysis-centered approaches that describe data processing methods, such as Crisp DM, SEMMA, or CIA Intelligence Process (Alnoukari & El Sheikh, 2012). KDD "is the most authoritative description of the data mining process" (Li & Ruan, 2007, p. 170). Since its publication, KDD has been refined and modified by many authors and is "still the most cited to date" (Hendler, 2014, p. 206) when it comes to data application processes. It is often used as a basis for specific applications, for example in business administration (Assunção, Calheiros, Bianchi, Netto, & Buyya, 2015). We combined the process of strategic communication with the KDD process by Fayyad et al. (1996a) (see figure 1, see also Wiencierz, Berger, Röttger, & Wietholt, 2017, pp. 8-10). In doing so, it becomes clear how to extract information from big data and transform information into a comprehensible structure for further analysis in strategic communication in

general, but also specifically in public relations. The technical aspects regarding data management, data-processing, and post-processing of discovered structures, which are missing in strategic communication up to now, are brought into an interconnected relationship with the communication process.

Figure 1: The Big Data Application Process in Strategic Communication (own depiction).
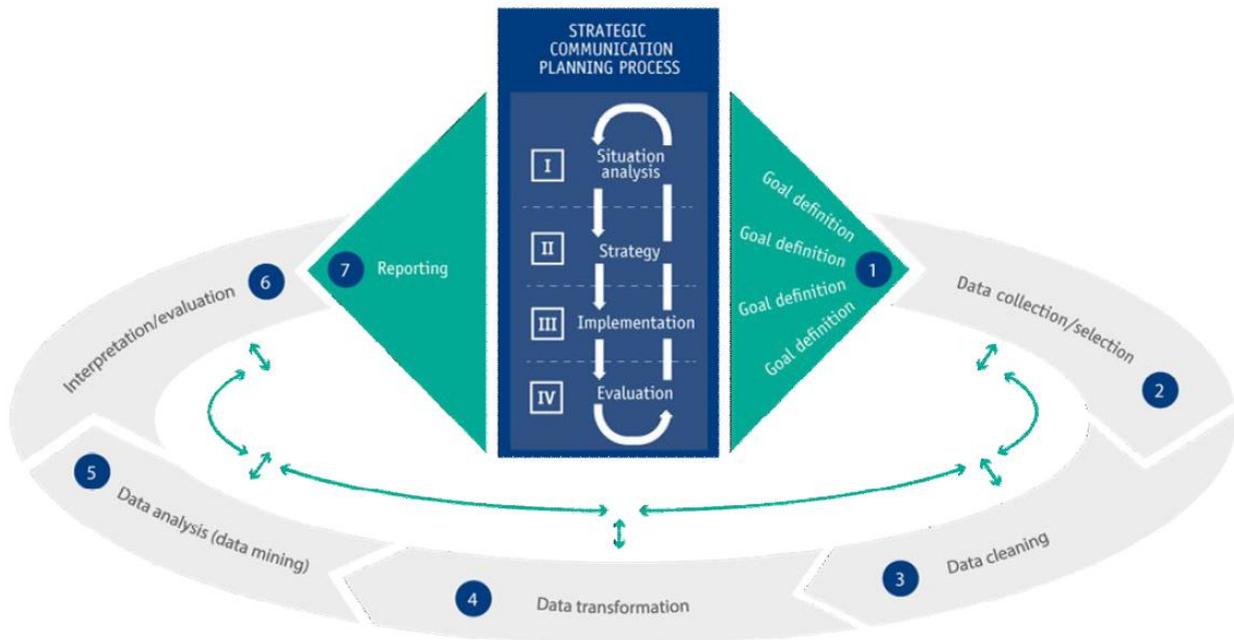
The KDD process is the "nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data" (Fayyad et al., 1996b, p. 29). With regard to our research context, the described overall process includes scaling of algorithms to massive data sets in the field of public relations, which run efficiently, the interpretation and visualization of results for the communication management, and the modeling and support of the overall human–machine interaction in public relations (Fayyad et al., 1996a). KDD requires a clear understanding of the area in which big data is to be used and a clear formulation of the goal in the first step (Sharafi, 2013, pp. 61–62). In public relations, the respective phase of communication planning determines the goals to be achieved using big data, e.g. to conduct detailed reputation analyses with big data for the situation analysis or to evaluate communication measures at the end of a campaign. In the second step, data selection, the necessary data to implement the determined communication goals is generated or acquired. This step is fundamental to the entire undertaking, because "if some important attributes are missing, then the entire study may fail" (Maimon & Rokach, 2010, p. 3). The third step, data cleaning, is an integral part of KDD. This step is about cleaning the data of outliers, inconsistencies, and errors, such as random measurement and transfer errors, or systematic errors due to incorrectly programmed algorithms for data generation (Maimon & Rokach, 2010). The next step, data transformation, describes the conversion of raw data into other data formats that can be processed by

the respective analytical tools. The main goal of this phase is data reduction, for example by combining variables (Fayyad et al., 1996b; Sharafi, 2013, p. 64).

According to Fayyad et al. (1996a), data mining comprises three steps in KDD, which we have combined into one step in our model. In this fifth step, data mining, (1) the objective of data mining must be determined before (2) the algorithms can be created or the analytical methods can be selected that will (3) search for patterns in the data. Summarizations, classifications, regressions, or clustering can all be purposes of data mining. The methodological decisions also include all decisions as to how the algorithms used should and can become adaptive in order to be able to carry out the analyses. These developments relate to deep learning and artificial intelligence (Maimon & Rokach, 2010). In the sixth step, interpretation and evaluation, the identified statistical patterns are interpreted in relation to the goals defined by the communication management in step 1. Here it is also decided whether and to what extent additional parameters need to be changed in which step, in order to further improve the results in a new cycle, set other priorities, or incorporate the resulting patterns directly into the model. Finally, the analyses as well as the recommendations for action need to be visualized and presented in an understandable way to the communication management. Only then can the conclusions drawn ultimately be translated into concrete actions in order to complete the big data application process in public relations successfully. "The success of this step determines the effectiveness of the entire KDD process" (Maimon & Rokach, 2010, p. 5). Separating knowledge from laboratory conditions and translating it into communication practice is a major challenge. It should also be noted that the various steps are not normally carried out on a linear basis. New information or decisions in one step entail backtracking to previous steps and repeating them multiple times. This is the only way to approximate the optimum KDD result.

Research Question

The current state of research shows that the debate and research about the usage of big data in strategic communication is rather at the beginning (for a more detailed overview, see Wiencierz & Röttger, 2017). Many communication practitioners are not yet aware of the importance of such technologies (Wiesenberg, Zerfass, & Moreno, 2017). Mostly the potentials and challenges in the area of corporate communication are being discussed, particularly in terms of marketing communication. For instance, it is possible to identify core customers using visual data mining tools, and target-group-specific customer communication by means of advertising can be improved with the help of elaborate big data analyses (e.g., Ho, Chung, & Lau, 2010; Liu, Pancras, & Houtz, 2015). In addition to the possibilities of an individualized customer approach, big data also makes it possible to meet the challenges of globalization and to consider the idiosyncrasies and characteristics of relevant markets in advertising communication and media planning (O'Regan, Ashok, Maksimova, & Reshetin, 2011).

Compared to marketing communication, there is still a dramatic lack of research in the field of public relations with respect to big data applications (Wiencierz & Röttger, 2017). Weiner and Kochhar (2016) present an elaborate concept of how to implement big data in public relations, from situation analysis to evaluation. The case studies presented in their white paper illustrate the potential of big data for public relations. MasterCard, for example, uses the findings of its Conversation Suite, an algorithm-based computer program, to monitor, analyze, and visualize conversations around the world in real time, as a foundation for their communications decisions-making. The financial

services provider uses big data extracted from social media, in order to anticipate topics and discussions relevant to the company and to help define the respective agenda, for example on mobile payment.

A concrete knowledge generation process using big data in public relations and in strategic communication in general, including the technical complexity it entails, has not been described to date. In order to address this shortfall in research and to investigate whether the ideal-typical process described here is reflected in public relations practice, this article examines the following research question:

Which ideal-typical steps does a big-data application process involve in public relations?

Study

Method

To investigate the research question, the results of two guideline-based in-depth expert interviews were used. One expert interview was conducted to examine big data's potential for strategic communication in the context of globalization. Having an understanding of international communication processes and those factors which influence this kind of communication "has become imperative for communication practitioners" (Hallahan et al., 2007, p. 28). Moreover, data from social media is valuable for the communication for increasingly fragmented stakeholders. Thus, the other expert interview was used to analyze the potential of social media analytics (SMA) for individualized communication with stakeholders. SMA is a specific form of big data usage and includes concepts, technologies and tools for the systematical generation as well as the storage and analysis of large, heterogeneous amounts of social media data, i.e. big social data (Stieglitz & Duan-Xuan, 2013). Because the guidelines of both studies have the key question "Which phases or steps does the big data application process typically involve at (your) company(-ies) / agency(-ies) in order to use big data for corporate communication?", it is possible to combine the samples of both surveys together to one final sample with a total of 37 interviews for a secondary evaluation.

Since the state of research shows that communication practitioners are often unaware of big data (Wiencierz & Röttger, 2017; Wiesenberg et al., 2017), an interdisciplinary sample was compiled in both interview studies with experts with extraordinary knowledge of big data or SMA and skills in this field. The experts had to have a certain proximity to corporate communications by working in related fields such as marketing, or by offering big data tools that can also be used in corporate communications. The publication *Germany – Excellence in Big Data* (Bitkom, 2016) was taken into account to identify German expertise in this area. Other German and also international experts, especially from the United States, were found by searching the professional networks Xing and LinkedIn, using keywords such as *corporate communication*, *big data (analytics)*, or *data scientist*. A few experts were recommended by other interviewees. For the sample of big data 19 experts (seven representatives of companies, 10 service providers, and two scientists) were interviewed between May and November 2016. For the other sample 18 experts from the field of SMA (seven representatives of companies, 10 service providers, and one scientist) were interviewed between July 2016 and February 2017. This is consistent with Kvale (1996) who suggested a number between 10 and 15 for exploratory qualitative interview studies. Data collection was finished when data saturation was reached and nothing new emerge in relation to the research objectives (Saunders & Rojon, 2014). Altogether, in the sample are 18 executives from communication departments or

software service providers and 16 employees as well as three scientists. With 20 experts, most of them are from the field of IT or data science, 11 are from the field of corporate communications and six from the field of marketing. The telephone interviews lasted on average 45 minutes. In our analysis of the audio-recorded interviews, we used the big data applications model in the strategic communication planning process, with the individual phases of strategic communication and the steps of the KDD (see figure 1), as the basis for formulating the category system, and identified and analyzed all public relations-specific expert statements with MAXQDA.

Findings

Potentials of Big Data for Public Relations

First of all, it is striking that the experts were significantly more reticent regarding the potential of big data in public relations compared to marketing communications. Some experts could not name any potential. The greatest opportunities, relatively speaking, were seen in issues management and crisis communication. Issues, or potentially image-damaging information, can be identified and anticipated more precisely and much faster with data from social media, for example. Accordingly, company representatives and service providers carry out real-time crisis monitoring, particularly by means of SMA, which they also use as an early-warning system for potential crisis issues.

Big data also provides communicators with added value for reputation management. Big data analyses establish a broader knowledge base about how one's brand or company is perceived in different regions compared to the competition, and who are the main players there. For such a situation analysis, communicators use tools such as big data to analyze the share of voice regarding their own reputation, i.e. their organization's online visibility and share of the discussion. Big data also helps communicators to optimize analytical methods for strategy formulation. Some of the surveyed experts conduct extensive network analyses of user-generated content in any form to identify, for example, influential bloggers and influencers in the different regions who represent topics and opinions of relevance to the company and have a wide reach. The analysis of the content can also provide information on which suitable topics should be addressed with which wording via which channels. Online touchpoint analyses based on big data, in turn, help to formulate strategies, because they examine in greater detail how, when, and above all where — at which online touchpoints — the target group seeks information about the company or a topic, and in what tonality it expresses its opinion. According to the experts, it is possible to conduct detailed target group analyses on aspects such as region and demography, but also interests and occupation, in order to learn to understand the user behavior of age- and gender-segmented target groups in relation to a topic. The experts also describe how large datasets are used to inform public relations activities. For instance, two versions of a communication activity are subjected to A/B testing by examining in real time whether version A or version B of a message, website, or app is better accepted.

Finally, the experts' statements show that big data applications significantly optimize performance measurement as they allow it to be carried out on a broader data basis and even in real time. Since the surveyed company representatives and service providers discuss using big data for the purpose of optimizing existing communication processes, they rely on established KPIs for performance measurement, such as individual reach, interaction rates, share of voice, share of buzz, or conversion rates.

Big Data Generation for Public Relations

In line with the ideal-typical application process outlined here, the experts emphasize the necessity of formulating clear goals and content-related questions at the outset, which can then be translated into technical questions. Situation analysis regarding the share of voice, for example, requires a different big-data knowledge generation process than issues management. This ideal approach is at odds with the reality that these goals are often not clearly defined, as some of the experts criticize: "I think everything comes back to the question, 'Why?' That is not asked enough by people" (interviewee 31). The vague definition of the goal then leads to deficits in the final evaluation as well as an unclean measurement of the status quo, according to the service providers we interviewed. "I think the big mistake is to start working and looking at this big data and say, 'How do I measure success?', without having defined success in the first place" (interviewee 36).

The analysis of the experts' statements shows that building a data pool with the three core characteristics of big data volume, variety, and velocity can be described as a second step: Which data must be available, in what volume, and at what speed or over what period of time, in order to achieve the set goals? In this context, many experts note that, counterintuitively, the added value of big data is generated not primarily by the characteristic of volume, but rather by variety and velocity. Even for public relations, innovative software and database infrastructures enable to use "all kinds of data formats from any data sources [...], no matter how large, no matter how fast, no matter from which source and in which format" (interviewee 16). In other words, these technologies allow public relations practitioners to collect and process unstructured data — text files, videos, Internet-of-things data, etc. — at an exponentially increasing speed. The service providers generate such data, or make the tools to do so available to businesses. The development of proprietary tools by companies is the exception.

To build up a data pool for the analyses, the experts advise to consider which data is available and which external data sources can be used, depending on the specific application. To better understand data types, it is helpful to distinguish between first, second, and third party data, as proposed by two of the experts. First party data refers to internal data owned by a company. This includes all data from a data warehouse or a customer relationship management (CRM) system, as well as market and competition data and data from the company's communication channels. Second party data is provided by another company, for example as part of a sponsorship or a loyalty-and-reward program. Finally, the data pool can also be enhanced with third party data. The company can generate such external data on its own or purchase it from a service provider. The experts consider social media data to be the most important for public relations.

As an ideal for data infrastructure, some experts describe a data lake into which all data streams from different sources in different formats and structures can flow, and which employees can access according to their needs. However, such an infrastructure requires complex technical solutions and is difficult to reconcile with existing privacy laws. According to the service providers, businesses are still a long way from such an ideal because, due to confusing data infrastructures involving different storage systems, they often have no overview of the data they already possess.

The third step, data cleansing and preparation, is described by most experts as complex but crucial. "The biggest time guzzler is cleaning up the data. This editing process is very time-consuming — and if you're not asking a good question, the whole thing is basically pointless" (interviewee 19). Tasks such as value-adding topic analyses or crisis monitoring can only be carried out if the collected data is cleaned of spam and other irrelevant data, such as content created by bots. This aspect mentioned by the experts corresponds to our fourth big-data characteristic, veracity.

According to the experts, cleaning is mostly algorithm-based and therefore automated. At the same time, company representatives and service providers emphasize that they also perform manual spot checks of data such as user comments, and, if necessary, adjust the search algorithm accordingly.

None of the experts surveyed explicitly described the data transformation step. However, the step is implied when the experts describe, for example, the necessity to convert raw data into actionable data. As mentioned, data must be transformed such that it can be processed by the appropriate analytical tools when, for example, text or image files are available in large quantities but different data formats.

Big Data Analysis for Public Relations

When it comes to data analysis, the experts hold that initially the analytical method should be determined based the goal, in order to search for patterns, correlations, or trends using computer-assisted statistical methods. "The first thing you do is a descriptive analysis. You look at the data and analyze how it's structured, what are typical values. That's how you learn to understand your data" (interviewee 13). In this descriptive analysis, company representatives and service providers examine frequencies such as the number of likes or page views.

After the descriptive analysis, the experts recommend that the reasons, interrelationships, and patterns should be examined in greater depth — a process that can be termed diagnostic analysis. "In other words, we're not just collecting information and putting out vanity metrics, but we're really trying to go deeper and establish some meaning in what the information is telling us" (interviewee 1). This makes it possible to show, for example, the virality of topics or the career of issues alongside the reasons for their evolution.

Some experts also experiment with predictive analysis, which involves calculating the probabilities of future events. The rule of thumb is: The more data and patterns from past analyses are available and can be applied to future developments, the more reliable are the forecasts. However, the reliability of predictive analyses is controversial among the experts surveyed. However, most experts do not yet have a clear idea about the areas where predictive analysis can be used within the field of public relations. One potential public relations use of such analyses that was mentioned is the prediction of the careers of issues with the purpose of formulating a communication strategy based on these calculations. The skeptics doubt that predictive analyses, both in public relations and in strategic communication generally, can be sensibly carried out and produce reliable results, given the large number of intervening variables. Especially unexpected trends are often the most interesting ones, and crises are all but impossible to predict within an issues management.

According to the experts, big data technologies make it possible to link heterogeneous data and combine analyses, and thus to unlock in-depth knowledge in an innovative way, for example regarding issues and reputation. Case in point: An interviewed expert of an insurance company uses sophisticated listening tools and combines the resulting data with market data to conduct market analyses. Interviewee 27 recommends close cooperation with classical market and opinion research. The combination with monitoring-data yields more detailed results, for example on stakeholders and their discussion topics, than would be available without combining both methods.

According to the ideal-typical process described, the analyzed patterns, interrelationships, or forecasts must be interpreted in the next step, and recommendations for action must be derived from them. The experts note, however, that the entire data application process should still be evaluated before the results are presented in a report to the decision makers. It should be checked whether the correct analyses were carried out reliably with the appropriate data. As a general rule, "If you're not

going to evaluate them, you didn't need to do them in the first place" (interviewee 9). There is always the possibility that false correlations were produced or that a result is a fluke, and that correlations or relations can only be found in the dataset used and are not systematic.

The last step is the preparation and communication of the core results in an aggregated and visualized form in order to report them to decision makers. The experts note that it must be determined beforehand in which form and when or how often reports will be produced. It is crucial that the core results that support the communication goal be prepared in an understandable way, using the right KPIs — for example in the form of network graphs or diagrams, or with the aid of real-time dashboards. Company representatives and service providers tend to use automated procedures, such as alerts, to keep track of any increase in critical mentions of the company. The result is a report that serves as basis for decisions in the individual phases of the strategic communication process.

Data-based Public Relations as an Iterative Process

In line with our theoretical model, most experts generally describe the data application process as an iterative process. Iterative means that knowledge generation from big data is a process with multiple repetitions in which the same or similar actions are carried out step by step and decisions are reassessed until the result of the application meets the requirements for achieving the goal. For example, if the goal in public relations is automated, permanent issues management with big data tools, the big challenge is to optimize the algorithms to produce permanently valid results. Search terms must first be identified and tested until the required communications are captured with these search term combinations, the so-called queries. Likewise, in the further course of the knowledge generation process, new developments and decisions in one step lead to changes affecting previous steps. Such cases require backtracking to previous steps, which is represented in our model by the inner arrows. For example, some communicators and service providers describe the ongoing requirement to identify new relevant search terms or previously unconsidered channels and target groups, resulting in the need to repeat and adapt previous steps under changed conditions. "We are continually receiving new data, and new stakeholders are being added. There are new topics that we address. So we see this as a process in which we constantly have to reorientate ourselves" (interviewee 21). Only with ongoing feedback loops, which can also lead to an adjustment of the set goal, can communicators meet the constantly changing requirements and approximate the optimum result. Revisions are common and necessary.

The data applications should be carried out according to the principle, "Try, measure, learn, improve" (interviewee 25). 'Try' also means to dare something that could fail, because this is the only way to optimize applications and develop new processes in an explorative way. "Learn quickly and try a bunch of them. Discard whatever doesn't work. Fail fast" (interviewee 8). With such an iterative way of thinking, new analytical methods can be developed and implemented to give communicators a competitive edge.

Some experts even emphasize the exploratory character of big data, meaning that the focus should not be on answering a straightforward question, but on the existence of data that can be gathered in large quantities and with which communicators can experiment. In this way, relations and patterns can be found that would not become apparent with a targeted application geared at a clear goal, as described above. Critics of this purely explorative approach see the risk of getting "lost in minutiae" (interviewee 3).

Especially the described explorative, iterative way of thinking and acting seems to be at odds with the highly targeted application process with clear goals that most experts emphasize. The synthesis of the seemingly contradictory expert positions results in a "hybrid approach" (interviewee 28) as an ideal. Be it in terms of experimentation with heterogeneous datasets and the resulting innovative analyses, such as the integration of Google search data into market analyses, or of the more liberal implementation of conventional analyses — analysts and communicators should be given leeway to take exploratory approaches, even as the actual data application should be done in a targeted manner. Accordingly, interviewee 23 grants his employees the freedom also to analyze new topics that are not on the company's agenda. Crises could also be triggered by such topics.

Discussion

Increasing digitization is dramatically reducing the transaction costs for generating information, while at the same time, the volume of data is growing rapidly. The statements of the interviewed experts show that communicators obtain a broader decision-making basis for their communications planning if they use their big data application tools to capture and process many more messages from various stakeholders in different regions of the world in a much faster and more efficient manner. Depending on the available data and goals, experts are able to carry out very broad analyses at a global level or more focused, in-depth, local analyses. They use these findings to address regional or cultural specificities and to communicate in a much more targeted manner than in the past.

An ideal-typical method of data application in public relations was described in this article and checked against statements by experts on big data and SMA. No expert describes the application process of big data in corporate communication in the aforementioned theoretical form. Experts also emphasize that the use of big data is always case-dependent. Taken together, however, the statements elucidate the individual steps of the described process. The experts' statements show that it is possible to optimize the planning, execution, and evaluation of public relations using big data, even though the data application process is complex and demanding. Due to the complexity of the big data application, the experts' experience to date has been that the great vision of big data is followed by disillusionment and the insight of only being capable of a slow pace on the road towards data-driven corporate communications.

With the model presented here, research gaps and guardrails for future public relations research can be identified, making a valuable contribution to the discussion on data-based public relations. For example, it should be investigated which data can be generated and which analysis is appropriate for which phase of public relations. The other steps raise research questions such as which methods can be used to clean data, for example of bot-generated information, or how heterogeneous data can be prepared most effectively for data analysis. More research should be carried out to identify efficient structures that organizations have to implement in order to centrally consolidate and prepare heterogeneous data — which in international enterprises exists often in multiple languages — and to optimize it for fast access, in keeping with applicable privacy law. There are still shortcomings regarding anonymization and pseudonymization procedures, which need to be remedied to be able to implement applications in accordance with privacy regulations.

Conclusion

Our study is one of the first attempts to make big data in public relations tangible, although it comes with limitations. Our article does not provide a clear-cut definition of big data. This would

also hardly be possible, because no fixed standard exists for what exactly constitutes big data — such as how big is big — as the experts interviewed also note. Furthermore, the qualitative method of expert interviews used here results in a small data basis. Therefore, the plausibility and applicability of the described ideal-typical process should be tested with further studies — also in areas of strategic communication other than public relations.

We see our article as a first contribution to an urgently needed discussion about the application of big data in public relations and strategic communication in general. The experts' statements make it clear that big data is going to help communication practitioners work smarter. The experts pointed out that this new technology could be used to improve communications controlling – – from the planning of communications to performance measurement. Our study shows that big data makes it easier to identify correlations between practitioner communication actions and the behavioral actions of stakeholders or organizational outcomes. With the big data application tools, existing analyses can be optimized by conducting them with a much broader database. For example, the optimized analysis of the share of voice was mentioned in order to gain detailed insights in real time about the public attention to the brand, the products or the company compared to the competitors. Issues management can also be extended to automated, permanent real-time crisis monitoring. As our model illustrates, it is crucial that big data measures such as real-time crisis monitoring are integrated into the communication process. In this case, an internal crisis plan must contain information on who receives which automatic messages from the system in a crisis situation and how to proceed in this situation.

All big data measures, e.g. the programming of chat bots for one's own public relations, should be carried out according to the iterative principle described here. A procedure according to this principle is also compatible with the necessity of agile communication in companies as described by van Ruler (2015) in her pioneering work. Thus, for a succuessful public relations in a digitalized society, a more flexible planning method is needed in which change is a defining part of the communication process. Big data applications based on the principle presented here can provide detailed insights into what changes are necessary.

However, the experts' statements confirm that many communicators and companies are still in their infancy when it comes to data-based public relations, and that there is still great uncertainty surrounding big data. Moreover, not even big data will be able to make the communication process completely transparent. It will never be possible to read the minds of the corporation's stakeholders. Potentially, the fact that communicators are able to operate strategically on a broad data basis reduces the probability of misunderstandings. However, it is a fallacy that an algorithm must be and should be perfect: "The crucial thing is that it is better than what I had before […], knowing that still there happen mistakes" (interviewee 11).

The results provide indicators for the future job market and for how educators should prepare students for the occupational field. Strategic and creative skills will remain important, but IT and data analytics will become a part of strategic communication in the future. When it comes to the use of big data for strategic communication, "communicators can't just be a person of words and language. You need to be a person of numbers as well" (interviewee 35). Consequently, public relations will need skilled personnel who are able to explore, digest, synthesize and explain incoming insights gained from big data. However, communicators do not need to become data scientists, and there is no need to train them as such. Knowledge must be imparted to students to enable them to form their own opinions on the use of big data and to enable them to communicate and cooperate in interdisciplinary teams, e.g. with data scientists. The use of big data should always be accompanied

by decision-making and action competence in order to allow these applications to be questioned.

Finally, two conflicting trends should be kept in mind: privacy and personalization. If public relations wants to address stakeholders in a targeted and individualized way based on data, communication practitioners must always take into account the skepticism and privacy concerns of stakeholders. Public relations must manage the balancing act between exploiting the potential of big data and establishing trust among stakeholders and the public (Wiencierz, 2018; 2016, p. 12).

References

Alnoukari, M., & El Sheikh, A. (2012). Knowledge Discovery Process Models: From Traditional to Agile Modeling. In A. A. R. El Sheikh & M. Alnoukari (Eds.), *Business Intelligence and Agile Methodologies for Knowledge-Based Organizations: Cross-Disciplinary Applications* (pp. 72–100). Hershey PA: Business Science Reference.

Assunção, M. D., Calheiros, R. N., Bianchi, S., Netto, M. A. S., & Buyya, R. (2015). Big Data Computing and Clouds: Trends and Future Directions. *Journal of Parallel and Distributed Computing*, *79–80*, 3–15. doi: 10.1016/j.jpdc.2014.08.003

Beyer, M. A., & Laney, D. (2012). The Importance of "Big Data": A Definition. *Gartner Research Report*.

Bitkom. (2016). *Germany - Excellence in Big Data*. Berlin. Retrieved from https://www.bitkom.org/noindex/Publikationen/2016/Leitfaden/Germany-Excellence-in-Big-Data/Germany-Excellence-in-Big-Data.pdf

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996a). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, *17*(3), 37–54.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996b). The KDD Process for Extracting Useful Knowledge from Volumes of Data. *Communications of the ACM*, *39*(11), 27–34.

Gandomi, A., & Haider, M. (2015). Beyond the Hype: Big Data Concepts, Methods, and Analytics. *International Journal of Information Management*, *35*(2), 137–144. doi: 10.1016/j.ijinfomgt.2014.10.007

Hallahan, K., Holtzhausen, D., van Ruler, B., Verčič, D., & Sriramesh, K. (2007). Defining Strategic Communication. *International Journal of Strategic Communication*, *1*(1), 3–35. doi: 10.1080/15531180701285244

Hendler, J. (2014). Data Integration for Heterogenous Datasets. *Big Data*, *2*(4), 205–215. doi: 10.1089/big.2014.0068

Ho, Y., Chung, Y., & Lau, K. (2010). Unfolding Large-Scale Marketing Data. *International Journal of Research in Marketing*, *27*(2), 119–132. doi: 10.1016/j.ijresmar.2009.12.009

Kvale, S. (1996). *Interviews: An Introduction to Qualitative Research Interviewing*. Thousand Oaks, CA: SAGE.

Li, T., & Ruan, D. (2007). An Extended Process Model of Knowledge Discovery in Database. *Journal of Enterprise Information Management*, *20*(2), 169–177. doi: 10.1108/17410390710725751

Liu, H., Pancras, J., & Houtz, M. (2015). Managing Customer Acquisition Risk Using Co-Operative Databases. *Journal of Interactive Marketing*, *29*, 39–56. doi: 10.1016/j.intmar.2014.09.002

Macnamara, J., & Likely, F. (2017). Revisiting the Disciplinary Home of Evaluation: New Perspectives to Inform PR Evaluation Standards. *Research Journal of the Institute for Public Relations, 3*(2), 1–21.

Maimon, O., & Rokach, L. (2010). Introduction to Knowledge Discovery and Data Mining. In O. Maimon & L. Rokach (Eds.), *Data Mining and Knowledge Discovery Handbook. Second Edition* (pp. 1–15). Boston, MA: Springer. doi: 10.1007/978-0-387-09823-4_1

O'Regan, M., Ashok, K., Maksimova, O., & Reshetin, O. (2011). Optimizing Market Segmentation for a Global Mobile Phone Provider for Both Targeting and Insight. *Journal of Advertising Research*, *51*(4), 571–577. doi: 10.2501/JAR -51-4-571-577

Röttger, U., Preusse, J., & Schmitt, J. (2014). *Grundlagen der Public Relations [Basics of Public Relations]*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Saunders, M., & Rojon, C. (2014). There's no Madness in my Method: Explaining how your Coaching Research Findings are built on Firm Foundations. *Coaching: An International Journal of Theory, Research and Practice, 7*(1)*, 74–83. doi: 10.1080/17521882.2014.889185

Scholz, T. M. (2017). *Big Data in Organizations and the Role of Human Resource Management. A Complex Systems Theory-Based Conceptualization*. Frankfurt a. M.: Peter Lang GmbH.

Sharafi, A. (2013). *Knowledge Discovery in Databases. Eine Analyse des Änderungsmanagements in der Produktentwicklung*. Wiesbaden: Gabler Verlag. doi: 10.1007/978-3-658-02002-6

Stieglitz, S., & Dang-Xuan, L. (2013). Social Media and Political Communication: A Social Media Analytics Framework. *Social Network Analysis and Mining, 3*(4), 1277-1291. doi: 10.1007/s13278-012-0079-3

Van Ruler, B. (2015). Agile Public Relations Planning: The Reflective Communication Scrum. *Public Relations Review*, *41*(2), 187–194. doi: 10.1016/j.pubrev.2014.11.008

Weiner, M., & Kochhar, S. (2016). *Irreversible: The Public Relations Big Data Revolution*. Retrieved from http://www.instituteforpr.org/wp-content/uploads/IPR_PR-Big-Data-Revolution_3-29.pdf

Wiencierz, C. (2016). Der Big-Data-Anwendungsprozess in der Unternehmenskommunikation. Potenziale und Herausforderungen [The Big Data Application Process in Corporate Communications. Potentials and Challenges]. In B. Günter, M. Piwinger, & G. Schönborn (Eds.), *Kommunikationsmanagement. Strategien, Wissen, Lösungen [Communication Management. Strategies, Knowledge, Solutions]*. Köln: Luchterhand, Beitrag 4.49.

Wiencierz, C. (2018). Vertrauen in gemeinwohlorientierte Big-Data-Anwendungen. Ethische Leitlinien für eine datenbasierte Organisationskommunikation [Trust in Big Data Applications that are Oriented towards the Common Good. Ethical Guidelines for Data-based Organizational Communication]. K. Liesem & L. Rademacher (Eds.), *Die Macht der Strategischen Kommunikation. Medienethische Perspektiven der Digitalisierung [The Power of Strategic Communication. Media Ethical Perspectives on Digitalization]* (pp. 109-126). Baden-Baden: Nomos.

Wiencierz, C., & Röttger, U. (2017). The Use of Big Data in Corporate Communication. *Corporate Communications: An International Journal*, *22*(3), 258-272. doi: 10.1108/CCIJ-02-2016-0015

Wiencierz, C., Berger, K., Röttger, U., & Wietholt, C. (2017). *Startklar für Big Data. Chancen, Voraussetzungen und Anwendungen für die Unternehmenskommunikation [Ready for Big Data. Opportunities, Prerequisites and Applications for Corporate Communications]. Communication Insights 4*. Leipzig: The Academic Society for Corporate Management & Communication.

Wiesenberg, M., Zerfass, A., & Moreno, A. (2017). Big Data and Automation in Strategic Communication. *International Journal of Strategic Communication*, *11*(2), 95–114. doi: 10.1080/1553118X.2017.1285770

Zerfass, A. (2008). Corporate Communication Revisited: Integrating Business Strategy and Strategic Communication. In A. Zerfass, B. van Ruler, & K. Sriramesh (Eds.), *Public Relations Research* (pp. 65–96). Wiesbaden: VS Verlag für Sozialwissenschaften. doi: 10.1007/978-3-531-90918-9_5