Big Data and CSR Communication

Ramón Reichert

Abstract This chapter discusses corporate social responsibility (CSR) in the area of digital media culture. Social media networks and online platforms are massive data collectors and have become the most important data source for collecting statistical data social large amount of data (i.e. big data) can be generated from all online communication. These data are, for example, used to identify moods and trends. Big data research has become very diversified in the past years by using machine-based processes for computer-based social media analysis. This article first summarizes current research on social networks, online communication and big data. Then three case studies are presented, focusing on (1) health monitoring and big data aggregated from Google search and social media monitoring, (2) Facebook data research and the analysis of data structures generated from this social network, and (3) big data research on Twitter. Finally, future developments, challenges and implications with regards to health communication, communication management and CSR are discussed.

Influential theoreticians investigating corporate social responsibility (CSR) agree that the rise and broad diffusion of social media has not only been accompanied by a transformation of everyday communication, but that it can also be regarded as an indication of the fundamental process of change that society is undergoing: "Corporate social responsibility (CSR) can be perceived as a reaction to this ongoing process of societal change."[1] Against this backdrop, recent studies of academic CSR have given rise to a debate about the relative importance of communicative strategies within the field of digital media culture.[2] Most authors are united in

---

[1] Bernd Lorenz Walter: Corporate Social Responsibility Communication: Towards a Phase Model of Strategic Planning, in Ralph Tench, William Sun, Brian Jones (ed.) Communicating Corporate Social Responsibility: Perspectives and Practice (Critical Studies on Corporate Responsibility, Governance and Sustainability, Volume 6) Emerald Group Publishing Limited 2014, pp. 59–79, here: 59.

[2] Oliver Meixner/Elisabeth Pollhammer and Rainer Haas: The communication of CSR activities via social media. A qualitative approach to identify opportunities and challenges for small and medium-sized enterprises in the agri-food sector, in: Jivka Deiters, Ursula Rickert, Gerhard Schiefer (eds.), Proceedings in Food System Dynamics and Innovation in Food Networks 2015, pp. 354–362.

R. Reichert (✉)
Department for Theatre, Film and Media Studies, Vienna University, Vienna, Austria
e-mail: ramon.reichert@univie.ac.at

believing that the communication with stakeholders is of great importance, particularly in the discursive sphere of social media. However, many actors remain convinced that they must convey communication in a top-down direction, in order to optimize communicative processes:

> Obviously, the possibilities and limitations of social media are misinterpreted and mainly connected to traditional marketing purposes. The experts assume that most of the social media tools like Facebook or Twitter are applied because the enterprises want to increase sales or improve the overall image of the company. The mistake here is that the responsible persons want to manage the processes like in the past as was with other media channels. But the professional application of social media demands a lot of authenticity and much higher efforts; enterprises must be willing to interact with their stakeholders.[3]

According to recent theories relating to Social Media Research, sustainable communication in the sphere of social media must engage with the digital communication culture, which deviates significantly from the diffusion of information and manipulation of opinion as evinced by traditional media channels and formats. Most authors represent the view that following key factors lend structure to the development of social media business communications geared towards sustainability: transparency, credibility, authenticity, and non-hierarchical communication can have a significant impact on the positive image of a corporation. However, this appraisal only applies to the user-generated interface of social media. Within the context of Big Data, on the other hand, a perceptible shift of all listed parameters has taken place, because the acquisition, modeling and analysis of large amounts of data, accelerated by servers and by entrepreneurial individuals, is conducted without the users' knowledge or perusal. Consequently, the socially acceptable communication of Big Data research seeks to integrate the methods, processes, and models used for data collection in publication strategies, in order to inform the users of online platforms (Facebook Data Team), or to invite them to contribute to the design and development of partially open interfaces (Twitter API).

Over the past few years, big data research has become highly diversified, and has yielded a number of published studies, which employ a form of computer-based social media analysis supported by machine-based processes such as text analysis (quantitative linguistics), sentiment analysis (mood recognition), social network analysis, or image analysis, or other processes of a machine-based nature. Given this background it would be ethically correct to regularly enlighten users of online platforms about the computer-based possibilities, processes, and results associated with the collection and analysis of large volumes of data.

The buzzword "big data" is on everyone's lips and not only describes scientific data practices but also stands for societal change and a media culture in transition. On the assumption that digital media and technologies do not merely convey neutral messages but establish cultural memory and develop social potency, they may be understood as discourses of societal self-reflection.

---

[3]Ibid, p. 359.

The research methods of big data research, such as text, sediment, network and image analyses, are based on the insight that the social net has developed into the most important data source in the production and application of knowledge for government and control purposes. First, I would like to discuss the current state of research, as well as include some case studies and a discussion of future developments, paying particular attention to the challenges and implications for health communication, communication management and corporate social responsibility.

1   Media and Technologies

In the era of big data, the status of social networks has changed radically. Today, they increasingly act as gigantic data collectors for the observational requirements of social-statistical knowledge, and serve as a prime example of normalizing practices.[4] Where extremely large quantities of data are analyzed, it now usually entails the aggregation of moods and trends. Numerous studies exist, where the textual data of social media has been analyzed in order to predict political attitudes, financial and economic trends, psycho-pathologies, and revolutions and protest movements.[5]

Large quantities of data (big data) are being generated in all areas of Internet communication. Since the late twentieth century, digital big science with its big computer centers and server farms has been one of the central components of the production, processing and management of digital knowledge. Concomitantly, media technologies of data acquisition and processing as well as media that create knowledge in spaces of possibilities take center stage in knowledge production and social control. In their introduction to the "Routledge Handbook of Surveillance Studies", the editors Kirstie Ball, Kevin Haggerty and David Lyon draw a connection between technological and social control on the basis of the availability of large quantities of data: "Computers with the power to handle huge datasets, or 'big data', detailed satellite imaging and biometrics are just some of the technologies that now allow us to watch others in greater depth, breadth and immediacy."[6] In this sense one may speak of both *data-based* and *data-driven* sciences, as knowledge production has come to depend on the availability of computer infrastructures and the creation of digital applications and methods.

This went along with a significant change in the expectations towards twenty-first century science. In these debates, it is increasingly being demanded that the

---

[4]Berg, Kati Tusinski/Kim Bartel Sheehan: "Social Media as a CSR Communication Channel: The Current State of Practice", in *Ethical Practice of Social Media in Public Relations*. Eds. Marcia W. DiStaso and Denise Sevick Bortree, New York : Routledge, 2014: pp. 99–110, 103.

[5]Cf. Gerard George/Martine R. Haas/Alex Pentland: "Big data and management", in *Academy of Management Journal* Vol. 57, No. 2 (2014), pp. 321–326.

[6]Ball, Kirstie/Haggerty, Kevin D./Lyon, David, *Routledge Handbook of Surveillance Studies* (London: 2012), p. 2.

historically, culturally and socially influential aspects of digital data practices be systematically studied, with a view to embedding them in future science cultures and epistemologies of data production and analysis. A comparative analysis of data processing that takes into account the material culture of data practices from the nineteenth to the twenty-first century shows that mechanical data practices had a substantial influence on the taxonomic interests of scientists by the beginning of that period—long before there were computer-based methods of data acquisition.[7] Further investigations work out the social and political conditions and consequences of the transition from the mechanical data counting of the first censuses around 1890 through the electronic data processing of the 1950s to today's digital social monitoring. However, the collection of large quantities of data is also entangled in a history of power of the data-based epistemes.[8] At the interface between corporate business models and governmental action, biotechnology, health prognostics, labor and finance studies, risk and trend research experiments with predictive models of trends, opinions, moods or collective behavior in their social media and Web analyses.

As a buzzword, big data stands for the superposition of control knowledge grounded in statistics with a *macro-orientation* towards the economic exploitability of data and information grounded in media technology.[9] In most cases, exploring very large quantities of data aims to aggregate moods and trends. However, these data analyses and visualizations usually merely collect facts and disregard social contexts and motives.[10] Nevertheless, the big data approach has been able to establish itself in the human, social and cultural sciences.

Due to the Internet and the increasing popularity of social media services, research approaches for handling digital communication data are gaining in relevance. However, analog methods, e.g., surveys or interviews that were developed for studying interpersonal or mass communication, cannot simply be transferred to the communication practices in the Social Net. Richard Rogers, an influential researcher in the field of social media research, advocates no longer applying only digitalized methods (e.g., online surveys) in studying the network culture, but concentrating on digital methods to diagnose and predict cultural change and societal developments.

Thus, digital methods can be seen as approaches that do not adapt existing methods for Internet research but take up the genuine procedures of digital

---

[7]Cf. Driscoll, Kevin, "From Punched Cards to 'Big Data': A Social History of Database Populism." *Communication*. Vol. 1, No. 1, 2012. Online: http://kevindriscoll.info (accessed 20.06.2015).

[8]Cf. Leistert, Oliver/Röhle, Theo (eds.), *Generation Facebook. Über das Leben im Social Net*. (Bielefeld: 2011).

[9]Oliver Meixner/Elisabeth Pollhammer/Rainer Haas: "The communication of CSR activities via social media A qualitative approach to identify opportunities and challenges for small and medium-sized enterprises in the agri-food sector", in *Proceedings in Food System Dynamics* (2015): pp. 354–362, 357.

[10]Cf. Rick Edgeman, "Sustainable Enterprise Excellence: towards a framework for holistic data-analytics", in *Corporate Governance* Vol. 13, No. 5 (2013), pp. 527–540.

media.[11] According to Richard Rogers, digital methods are research approaches that, on the one hand, use large quantities of digital communication data produced by millions of users in the Social Net every day, and, on the other hand, filter, analyze, process and present this data using computer-based techniques. In the tradition of the actor-network theory, numerous representatives of Internet research postulate digital actors such as hyperlinks, threads, tags, PageRanks, log files and cookies, which interact with each other and with dataset subjects. The actor networks can only be observed, recorded and assessed using digital methods, though they often turn out to be unstable and ephemeral events. This gives rise to a novel methodology, which combines aspects of computer science, statistics and information visualization with research approaches of the social sciences and humanities. However, the vision of such a native digital research methodology, whether in the form of a "computational social science"[12] or of "cultural analytics"[13] is still incomplete and demands an epistemic inquiry into the digital methods. Against this background, the objectivist and positivist postulates of big data research will be data-critically queried with respect to technological infrastructure, the pragmatics of research and the politics of knowledge using three practical case studies.

2   Case Study: Health Monitoring

In the research field of "social media data," health diagnostics has emerged as an evidence-based prevention practice that influences the institutional development of state-run preventive health care and the cultural techniques for making life choices. Preventive health care observes with great interest how millions of users worldwide use the Internet search engine Google daily to search for information on health. When influenza is rampant, search queries on flu become more frequent, so that conversely the frequency of certain search terms can yield indications on the frequency of flu cases. Studies on search volume patterns have found a significant connection between the number of flu-related search queries and the number of persons with actual flu symptoms.[14] This epidemiological structure of relationships can be expanded to cities, regions, countries and continents and represented in detail to allow for early warning of epidemics. With the epidemiological evaluation of textual clusters and semantic fields, the Social Net acquires the status of a big

---

[11]Cf. Rogers, Richard, *Digital Methods* (Cambridge/MA: 2013), p. 13f.

[12]Lazer, David et al., "Computational Social Science." *Science*, Vol. 323, No. 5915, 2009, pp. 721–723.

[13]Manovich, Lev, "How to Follow Global Digital Cultures: Cultural Analytics for Beginners," in *Deep Search: The Politics of Search Beyond Google* Becker, Konrad/Stalder, Felix (eds.) (Edison/NJ: 2009), pp. 198–212.

[14]Cf. Freyer-Dugas, Andrea et al., "Google Flu Trends: Correlation With Emergency Department Influenza Rates and Crowding Metrics," *Clinical Infectious Diseases*, Vol. 54, No. 7, 2012, pp. 463–469.

database that reflects social life in its entirety and thus affords a representative data source for preventive health policymaking.

Most monitoring projects that study large quantities of data in the Social Net are being carried out by computer linguists and computer scientists. They generally interpret the communication as collectively shared and culturally specific knowledge structures with which individuals try to interpret their experiences. The investigation of these knowledge structures aspires to gain socially differentiated insights into public debates and socially shared discourse networks. The research begins by creating a digital corpus consisting of terminological entities that are usually categorized as "canonical." Thus, some hypotheses only result from the empirical resistance of the big data and then develop in the process of description. In this way, the category catalogues at first sight suggest scientific objectivity, but in view of the huge quantities of data, a precise validation of the terminological choices, i.e., of the interpretative selection of the big data, often remains unclear and vague. This uncertainty in hypothesis formation is due to the fact that the extensive dataset no longer allows for an overview and thus can no longer be linguistically coded. Often the quantity of data acquired is so extensive that further weighting and restriction are required to reduce the complexity after an initial exploration of the material. This methodological restriction of big data monitoring has been criticized on the basis of the fact that that the resulting findings merely provide an atomistic view of the data and thus necessarily largely forgo a contextualization of the text material and hence also a context-sensitive interpretation of the use of signs.[15]

The evaluation of the data from Google trends can be extended to other trend developments. There are now numerous studies that investigate the textual data of social media to make early predictions of *political attitudes*,[16] *financial trends* and *economic crises*,[17] *psychopathologies*,[18] *uprisings* and *protest movements*.[19] From the systematic evaluation of big data, prognosticians expect more efficient company management with statistical assessments of demand and sales markets, customized service offers and improved societal control. The algorithmic prognosis of collective processes is particularly significant politically. In this context, the Social Net

---

[15]Cf. Boyd, Danah/Crawford, Kate, "Six Provocations for Big Data. Conference Paper, A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society" [presented in September 2011, Oxford]. Online: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1926431 (accessed 27.12.2013).

[16]Conover, Michael D. et al., "Predicting the Political Alignment of Twitter Users" [presented at 3rd IEEE Conference on Social Computing 2011, forthcoming]. Online: http://cnets.indiana.edu/wpcontent/uploads/conover_prediction_socialcom_pdfexpress_ok_version.pdf (accessed 27.12.2013).

[17]Gilbert, Eric/Karahalios, Karrie, "Widespread Worry and the Stock Market" [presented at 4th International AAAI Conference on Weblogs and Social Media (ICWSM), Washington, DC 2010].

[18]Wald, Randall/Khoshgoftaar, Taghi M./Sumner, Chris, "Machine Prediction of Personality from Facebook Profiles" [presented at 13th IEEE International Conference on Information Reuse and Integration, Washington 2012], pp. 109–115.

[19]Yogatama, Dani, "Predicting the Future: Text as Societal Measurement," 2012. Online: http://www.cs.cmu.edu/~dyogatam/Home_files/statement.pdf (accessed 27.12.2013).

has become the most important data source for producing government and control knowledge. The political control of social movements thus shifts to the Net when sociologists and computer scientists jointly take part e.g., in "riot forecasting", using the collected text data of Twitter streams: "Due to the availability of the dataset, we focused on riots in Brazil. Our datasets consist of two news streams, five blog streams, two Twitter streams (one for politicians in Brazil and one for the general public in Brazil), and one stream of 34 macroeconomic variables related to Brazil and Latin America."[20]

Big Data offers a specific method and technology for statistical data evaluation, which arises at the epistemic interface of business informatics and commercial data management and combines the fields of "Business Intelligence," "Data Warehousing"[21] and "Data Mining".[22] The discussion about the significance of big data for technology, infrastructure and power indicates that the numerical representation of collectivities is one of the fundamental operations of digital media and constitutes a computer-based knowledge technique that allows collective practices to be mathematically described and thus quantified. The determination of the multiplicities using numerically structured quantifications serves mainly for orientation and can be interpreted as a strategy that translates collective data streams into readable data collectives. In this sense, social network media like Facebook, Twitter and Google + act in the public media sphere as a mirror of the general state of the economy[23] or as a prognostic indicator of national mood swings.[24] To this end, they themselves form arenas of popular attention and popularizing discourses that ascribe certain exterior roles to them—e.g., as indicators of economic cycles and social welfare.

Social Media Monitoring represents a new paradigm of *Public Health Governance*. While traditional approaches to health prognosis operated with data collected in the clinical diagnosis, Internet biosurveillance studies use the methods and infrastructures of *Health Informatics*. That means, more precisely, that they use unstructured data from different Web-based sources and targets using the collected and processed data and information about changes in health-related behavior. The two main tasks of Internet biosurveillance are (1) the early detection of epidemic

---

[20]Yogatama, Dani, "Predicting the Future: Text as Societal Measurement," 2012, p. 3, Online: http://www.cs.cmu.edu/~dyogatam/Home_files/statement.pdf (accessed 15.04.2015).

[21]Data warehousing is an infrastructural technology that serves to evaluate data inventories.

[22]In the commercial sector, the term "data mining" has established itself for the entire process of "knowledge discovery in databases." "Data mining" refers to the application of exploratory methods to a data inventory with the aim of pattern recognition. Beyond representing the data, the goal of exploratory data analysis is to search for structures and peculiarities. It is thus typically employed when the problem is not well-defined or the choice of a suitable statistical model is unclear. With data selection as its point of departure, its search comprises all activities required for communicating patterns recognized in data inventories: problem definition, selection and extraction, preparation and transformation, pattern recognition, evaluation and presentation.

[23]Cf. Bollen, Johan/Mao, Huina/Zeng, Xiaojun Zeng, "Twitter Mood Predicts the Stock Market," *Journal of Computational Science*, Vol. 2, No. 1, 2011, pp. 1–8.

[24]Cf. Bollen, Johan, "Happiness Is Assortative in Online Social Networks," *Artificial Life*, Vol. 17, No. 3, 2011, pp. 237–251.

diseases, biochemical, radiological and nuclear threats and (2) the implementation of strategies and measures of sustainable governance in the target areas of health promotion and health education.

Internet biosurveillance uses the accessibility to data and analytic tools provided by digital infrastructures of social media, *participatory sources* and *non-text-based sources*. The structural change generated by digital technologies, as main driver for Big Data, offers a multitude of applications for sensor technology and biometrics as key technologies. Biometric analysis technologies and methods are finding their way into all areas of life, changing people's daily lives. In particular the areas of sensor technology, biometric recognition processes and the general tendency towards convergence of information and communication technologies are stimulating Big Data research. The conquest of mass markets through sensor and biometric recognition processes can sometimes be explained by the fact that mobile, web-based terminals are equipped with a large variety of different sensors. More and more users are thus coming into contact with the sensor technology or with the measurement of individual body characteristics. Due to the more stable and faster mobile networks, many people are permanently connected to the Internet using their mobile devices, providing connectivity an extra boost. With the development of apps, application software for mobile devices such as smartphones (iPhone, Android, BlackBerry, Windows Phone) and tablet computers, the application culture of biosurveillance changed significantly, since these apps are strongly influenced by the dynamics of *bottom-up* participation. Andreas Albrechtslund speaks in this context of the "Participatory Surveillance" (2008) on the social networking sites, in which biosurveillance increasingly assumes itself as a place for open production of meaning and permanent negotiation, by providing comment functions, hypertext systems, ranking and voting procedures through collective framing processes.[25] Therefore, the algorithmic prognosis of collective processes enjoys particularly high political status, with the social web becoming the most important data-source for knowledge on governance and control.

3 Case Study: Facebook Data Research

While there has been much public speculation about how the social network Facebook transforms the online behavior of its members into metadata, relatively little is known about the actual methods used by Facebook to generate this data knowledge. The data-structures modeled by Facebook therefore need to be analyzed methodologically. Two areas in particular require attention: first, the positivism of large-scale data analyses; second, the status of theory in the online research

---

[25]Anders Albrechtslund: "Online Social Networking as Participatory Surveillance", in: *First Monday* Vol. 13/3 (2008), Online: http://firstmonday.org/ojs/index.php/fm/article/viewArticle/2142/1949

carried out by the Facebook data team. How is Facebook to be understood in the context of the "digital turn" in the social and cultural sciences? And how does the methodology underlying big data research bear out in concrete data practices?

Which music will one billion people hear in the future when they've just fallen in love, and which music will they hear when they've just broken up? These questions prompted the "Facebook Data Team" to evaluate the data of more than one billion user profiles (more than 10 % of the world's population) and 6 billion songs of the online music service Spotify in 2012, using a correlative data analysis that determines the degree of positive correlation between the variable "relationship status" and the variable "music taste".[26] This prognosis for collective consumption behavior is based on feature predictions expressed in a simple causal relationship using data mining. Led by sociologist Cameron Marlow, the group of computer scientists, statisticians and sociologists investigated the statistical relationship behavior of Facebook users and on February 10 of that year published two hit lists of songs that users heard when they changed their relationship status, succinctly calling them "Facebook Love Mix" and "Facebook Breakup Mix".[27] The back end research group not only distilled a global behavioral diagnosis from the statistical investigation of "big data"[28] but also transformed it into a suggestive statement about the future. It asserted: We researchers in the back end of Facebook know which music one billion Facebook users will want to listen to when they fall in love or break up.[29] Under the guise of merely collecting and passing on information, the "Facebook Data Team" research group establishes a power of interpretation with respect to users, by prompting users via the automatically generated update mode "What's going on?" to regularly post data and information.

However, the Facebook Data Team's statements about the future are only superficially mathematically motivated; they point towards the performative origin of future knowledge. Despite the advanced mathematization, calculization and operationalization of the future, the performative power of future knowledge is always drawn from the speech acts and propositional orders that can differentiate into literary, narrative and fictional forms of expression.[30]

The format of the hit list with its ten most popular songs tries to allow complex facts to be represented at a glance. It is a popularizing future narrative that is meant

---

[26]Facebook Data Science, https://www.facebook.com/data (accessed 28.12.2013).

[27]Under the title "Facebook Reveals Most Popular Songs for New Loves and Breakups," "Wired" raved about the new possibilities of data mining; see www.wired.com/underwire/2012/02/facebook-love-songs/ (accessed 28.12.2013).

[28]Wolf, Fredric et al., "Education and Data-Intensive Science in the Beginning of the 21st Century," *OMICS: A Journal of Integrative Biology*, Vol. 15, No. 4, 2011, pp. 217–219.

[29]The collective figure "We" in this case refers to the researchers in the back end and fueled futurological conspiracy theories that imagine the world's knowledge to be in the hands of a few researchers.

[30]Cf. Lummerding, Susanne, Facebooking. "What You Book is What You Get—What Else?" in *Generation Facebook. Über das Leben im Social Net*, Leistert, Oliver/Röhle, Theo (eds.) (Bielefeld 2011), pp. 199–216.

to fulfill a behavior-moderating, representative and rhetorical function and to highlight research on the future as entertaining and harmless. To be credible in this sense, futural epistemology must always be staged persuasively in some way; it must be exaggerated theatrically, enacted promotionally and narrated in order to generate attention. Thus the futural propositional forms inherently include a moment of prophetic enactment of self and knowledge, with which the scientific representatives aim to prove the added value of the social networks for purposes of societal diagnosis.[31] Social network media like Facebook nowadays act as global players of opinion research and trend analysis and play a decisive role in modeling statements about the future and futurological enactment of knowledge.

Happiness research is also increasingly using friendship networks like Facebook for evaluating its mass data. Within big data prognostics, happiness research is a central research direction. But the socioeconomic study of happiness occurs mostly in seclusion from the academic public. In this connection, influential theoreticians such as Lev Manovich[32] and Danah Boyd[33] warn of a "Digital Divide" that distributes future knowledge lopsidedly and could lead to power asymmetries between researchers *inside* and *outside* the networks.

This unequal relationship consolidates the position of social networks as computer-based control media that appropriate future knowledge along a vertical and one-dimensional network communication: (1) They enable a continuous flow of data (digital footprints), (2) they collect and order these data and (3) they establish closed knowledge and communication spaces for experts and their expertise, which condense the collective data into information and interpret them. Thus the future knowledge passes through different layers of media, technology and infrastructure that are arranged in a hierarchical pyramid:

> The current ecosystem around Big Data creates a new kind of digital divide: the Big Data rich and the Big Data poor. Some company researchers have even gone so far as to suggest that academics shouldn't bother studying social media data sets—Jimmy Lin, a professor on industrial sabbatical at Twitter argued that academics should not engage in research that industry 'can do better'.[34]

---

[31]Cf. Doorn, Niels Van, "The Ties that Bind: The Networked Performance of Gender, Sexuality and Friendship on MySpace," *New Media & Society*, Vol. 12, No. 4, 2010, pp. 583–602.

[32]Cf. Manovich, Lev, "The Promises and the Challenges of Big Social Data," in *Debates in the digital humanities*, Matthew K. Gold (ed.) (Minneapolis: University of Minnesota Press, 2012) pp. 460–475.

[33]Cf. Boyd, Danah/Crawford, Jane, "Six Provocations for Big Data. Conference Paper, A Decade in Internet Time" [presented at Symposium on the Dynamics of the Internet and Society, September 2011, Oxford]. Online: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1926431 (accessed 27.12.2013).

[34]Boyd, Danah/Crawford, Jane, "Six Provocations for Big Data. Conference Paper, A Decade in Internet Time" [presented at Symposium on the Dynamics of the Internet and Society, September 2011, Oxford]. Online: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1926431 (accessed 27.12.2013).

These statements illustrate—in addition to the factual seclusion of future knowledge with respect to technological infrastructure—that the strategic decisions are made in the back end and not in peer-to-peer communication. While peers, with their restricted agency, may distort results, create fake profiles and communicate nonsense, they cannot actively shape the future beyond tactical activities.

The "Facebook Happiness Index" introduced in 2007, which empirically evaluates users' moods in the status messages using a word index analysis, constitutes an important variety of futurological prophecy. Based on the status update data, the network researchers calculate the so-called "Gross National Happiness Index" (GNH) of societies. The sociologist Adam Kramer worked for Facebook from 2008 to 2009 and created the Happiness Index together with the Facebook Data Team, social psychologist Moira Burke, computer scientist Danny Ferrante and the director of data science research, Cameron Marlow. Adam Kramer was able to use the internally available data volume for this purpose. He evaluated the frequency of positive and negative words in the self-documenting format of the status messages and contextualized these self-recordings with the individual life satisfaction of the users ("convergent validity") and with significant data curves on days on which different events occupied the public media ("face validity"): "'Gross national happiness' is operationalized as a standardized difference between the use of positive and negative words, aggregated across days, and present[s] a graph of this metric."[35] The individual practices of self-care analyzed by the sociologists are ultimately reduced to the opposites "happiness"/"unhappiness" and "satisfaction"/"dissatisfaction." In the end, a mood state with binary structure is employed as an indicator of a collective mentality that has recourse to certain collectively shared experiences and expresses specific moods. The sociological mass survey of the self-documentations ("self reports") in social networks has so far determined the mood state of 22 nations. With the scientific correlation of subjective mental states and population statistics, the "Happiness Index" can be assessed not only as an indicator of "good" or "bad" governance, but as a criterion of a possible adaptive response of the political sphere to the perceptual processing of social networks. In this sense, the "Happiness Index" represents an extended tool set of economic expansion and administrative preparation of decisions.

4 Case Study: Twitter Research

For big data research, application programming represents a central tool of data-driven knowledge production. The term "application programming interface" is abbreviated here as API.

---

[35]Kramer, Adam D. I., "An Unobtrusive Behavioral Model of 'Gross National Happiness,'" in *Conference on Human Factors in Computing Systems*, Association for Computing Machinery (ed.), Vol. 28, No. 3, New York 2010, pp. 287–290, here p. 287.

It is an irony of history that the usability of the commercially motivated Twitter API has given a boost to big social data research in the field of media and communication sciences. The online research depends on the versions that the Twitter company makes freely available. The API prescribes, among other things, the quantity and mode of selection options and access restrictions. A dispositive analysis of the media could draw on the data critique of the research infrastructures articulated by Danah Boyd, Jane Crawford und Lev Manovich and could enhance the analysis of media arrangements and their discourses with the analysis of the power effects that arise at this interface.

Richard Rogers developed the concept of "online groundedness" "to conceptualize research that follows the medium, captures its dynamics, and makes grounded claims about cultural and societal change."[36] Can online research independently decide to follow a medium, or does it not rather affix itself to predetermined tendencies as a tactical effect? So far, the Twitter API has exercised constitutive power over the upsurge in applied Twitter research. The application programming made available by Twitter can be critically discussed in two respects as a media-dispositive infrastructure. As a programming paradigm for Web applications, it upholds the logic of the back and front end and hence does not act as a window to the world of social data but rather leads to automated preselection by creating software-based filters of selective knowledge generation that cannot be scrutinized by ordinary research. The filtering function of the application interface therefore systematically establishes a lack of transparency, regulatory gaps and epistemic ambiguities. In this regard, the methods of acquisition and processing that operate in the framework of application programming can be regarded as *foundational fiction*. The pertinent research literature[37] took a close look at the reliability and the validity of the Twitter data and came to the conclusion that the Twitter data interfaces can more or less be regarded as dispositive arrangements in the sense of a gatekeeper. This can be demonstrated through the different access points of the Twitter API: "Statuses/filter" returns all public tweets that correspond to a set of filter parameters passed. The standard access privileges allow up to 400 keywords, 5000 user IDs and 25 locations. "Statuses/samples" returns a small random subset of all public tweets. "Statuses/firehouse" is a limited access point that requires special privileges to connect to it. Thus, as filter interfaces the APIs always also produce economically motivated exclusion effects for network research, which the latter cannot control independently.

Digital methods and analyses situate themselves in the interplay between media innovations and the limitations of technical infrastructures. Thus, in order to precisely work out the power asymmetries of socio-technical interactions, one would have to avoid building on the sheer existence of individual communication technologies and their digital methods, and instead engage with the concrete data practices and technical infrastructural restrictions associated with them.

---

[36]Rogers, Richard, *Digital Methods* (Cambridge/MA: 2013), p. 64.

[37]Cf. Burgess, Jean/Puschmann, "Cornelius: The Politics of Twitter Data." Online: www.papers.ssrn.com/sol3/papers.cfm?abstract_id=2206225 (accessed 20.06.2015).

5   Development and Outlook

In recent years, social networks and online platforms have become important sources for mass collection of statistical data and have given rise to new forms of socio-empirical knowledge using data-based digital methods. Their gigantic databases serve to systematically acquire information and are employed for collecting, evaluating and interpreting statistical social data and information. Under the perspective of corporate social responsibility in their role as a medium for storing, processing and distributing mass data, social networks have brought forth comprehensive data aggregates that are used for predicting societal developments.[38]

However, the future knowledge contained in social media is not equally available to all involved. This asymmetric relationship between ordinary users and exclusive experts has been described in the pertinent literature as a "participatory gap."[39] Though it suggests a new form of government and administration, the future knowledge established by the social networks is excluded from public discussion.

Social networks have opened up new possibilities of tapping sources for empirical social research. The future knowledge of social networks overlaps two fields of knowledge. Empirical social science and media informatics are responsible for evaluating the communication in mediated interactive network media. Social research sees a decisive force for societal development in the communication media of the social networks. Its research perspective onto socialization via information technology in multimedially networked media has developed a coordinate frame of different knowledge sources and knowledge techniques to produce prognostic knowledge. Thus, for instance, the acquisition of knowledge is delegated to search bots that can access the public information. However, the future knowledge can also be used for enacting expected constellations of the statistical data aggregates, e.g., when the Facebook Data Team popularizes certain segments of its activities on its web page. In this sense, statistical data and information are built into the outward representation of the social networks and acquire an additional performative component.

In the course of its modeling, the future knowledge passes through different fields of production, acquisition and mediation and can be employed as a procedure, argumentation or integration. Against this background, the future knowledge can be viewed as a heterogeneous field of knowledge that absorbs empirical, formal-mathematical, semantic, psychological and visual knowledge. Accordingly, a futural episteme has affixed itself to the social media and brought forth a plurality of planning and consulting practices that act as multipliers of a computer-based power differential and a time-based system of rule. The heightened interest of market and opinion research in the trend analyses and prognoses of social networks

---

[38]W. Lance Bennett, "The Personalization of Politics: Political Identity, Social Media, and Changing Patterns of Participation", *The Annals of the American Academy of Political and Social Science*, November 2012 Vol. 6, No. 44: pp. 20–39.

[39]Taewoo, Nam/Stromer-Galley, Jennifer, "The Democratic Divide in the 2008 US Presidential Election," *Journal of Information Technology & Politics*, Vol. 9, No. 2, 2012, pp. 133–149.

illustrates that social, political and economic decision-making processes are being rendered highly dependent on the availability of prognostic knowledge. In this respect, this essay has drawn attention to the fact that prognostic techniques can always also be viewed as techniques of power manifested in medial arrangements and infrastructural precepts.

In view of the increased public participation in scientific research and development, enterprise-internal actors (researchers, public relations staff, CEOs), have a significant role to play in communicating data-driven insights to the wider public. The opportunities provided by social media, in particular, illustrate the development toward collaborative processes in the co-determination of the cultural formation of meaning. As such, open platforms and networks are well suited to the integration of critically reflecting communication processes in the field of business-oriented big data research. Establishing an ethical foundation for the CSR approach recognizes the challenges and requirements of online-based research in terms of data protection, and aims at creating binding directives for the development of digital methods and for data work involving electronic sources and resources. The planned preparation of guiding principles for ethical research identifies "best practice" examples for specific research processes. The ethical catalogue of guidelines for research relates to shaping the relationship between the researcher and research subject, it is intended to stimulate self-reflection during the research process, and it comprises the full spectrum of empirical data acquisition, data analysis and data preparation. The guiding principles for ethical research are developed in collaboration with legal experts in the field of data protection and primarily serve to protect personal rights and the cultural sustainability of research results.

To ensure the long-lasting availability and ongoing development of the research results, the participatory processes of the specialized and collective roles are continuously evaluated and processed in feedback procedures for the innovative development of socio-cultural sustainability. In addition to the technical organization of crowdsourcing, the institutional anchoring of cultural sustainability also seeks to answer the question about how to design the bridge across to new forms of medial and cultural expression and use, in order to ensure the sustainable integration of the users within the agenda of big data research on social networks sites.

6 Exercise and Reflective Questions

1. Describe the relationship between social media and corporate social responsibility (CSR).
2. Please describe the role social media networks play for generating big data.
3. How can social media data be applied to health monitoring and how can the data be used? Please illustrate with an example.
4. What role does the "Facebook data team" play in terms of predicting future knowledge about Facebook user's music choices?
5. Please discuss chances and problems when using social network data for future knowledge.