
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Platanou, Kalliopi; Mäkelä, Kristiina; Beletskiy, Anton; Colicev, Anatoli
Using online data and network-based text analysis in HRM research

Published in:
JOURNAL OF ORGANIZATIONAL EFFECTIVENESS

DOI:
[10.1108/JOEPP-01-2017-0007](https://doi.org/10.1108/JOEPP-01-2017-0007)

Published: 01/01/2018

Document Version
Publisher's PDF, also known as Version of record

Please cite the original version:
Platanou, K., Mäkelä, K., Beletskiy, A., & Colicev, A. (2018). Using online data and network-based text analysis in HRM research. JOURNAL OF ORGANIZATIONAL EFFECTIVENESS, 5(1), 81-97.
<https://doi.org/10.1108/JOEPP-01-2017-0007>

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.



Journal of Organizational Effectiveness: People and Performance

Using online data and network-based text analysis in HRM research
Kalliopi Platanou, Kristiina Mäkelä, Anton Beletskiy, Anatoli Colicev,

Article information:

To cite this document:

Kalliopi Platanou, Kristiina Mäkelä, Anton Beletskiy, Anatoli Colicev, (2017) "Using online data and network-based text analysis in HRM research", Journal of Organizational Effectiveness: People and Performance, <https://doi.org/10.1108/JOEPP-01-2017-0007>

Permanent link to this document:

<https://doi.org/10.1108/JOEPP-01-2017-0007>

Downloaded on: 11 October 2017, At: 23:58 (PT)

References: this document contains references to 82 other documents.

The fulltext of this document has been downloaded 179 times since 2017*

Users who downloaded this article also downloaded:

(2017), "Human capital analytics: why aren't we there? Introduction to the special issue", Journal of Organizational Effectiveness: People and Performance, Vol. 4 Iss 2 pp. 110-118 https://doi.org/10.1108/JOEPP-04-2017-0035

Aalto-ympäristön kirjasto, Toimii
Aalto University Library, Toimii

Access to this document was granted through an Emerald subscription provided by All users group

For Authors

If you would like to write for this, or any other Emerald publication, then please use our Emerald for Authors service information about how to choose which publication to write for and submission guidelines are available for all. Please visit www.emeraldinsight.com/authors for more information.

About Emerald www.emeraldinsight.com

Emerald is a global publisher linking research and practice to the benefit of society. The company manages a portfolio of more than 290 journals and over 2,350 books and book series volumes, as well as providing an extensive range of online products and additional customer resources and services.

Emerald is both COUNTER 4 and TRANSFER compliant. The organization is a partner of the Committee on Publication Ethics (COPE) and also works with Portico and the LOCKSS initiative for digital archive preservation.

*Related content and download information correct at time of download.

Using online data and network-based text analysis in HRM research

Online data and network-based text analysis

Kalliopi Platanou, Kristiina Mäkelä and Anton Beletskiy
Aalto University School of Business, Helsinki, Finland, and

Anatoli Colicev

Graduate School of Business, Nazarbayev University, Astana, Kazakhstan

Abstract

Purpose – The purpose of this paper is to propose new directions for human resource management (HRM) research by drawing attention to online data as a complementary data source to traditional quantitative and qualitative data, and introducing network text analysis as a method for large quantities of textual material.

Design/methodology/approach – The paper first presents the added value and potential challenges of utilising online data in HRM research, and then proposes a four-step process for analysing online data with network text analysis.

Findings – Online data represent a naturally occurring source of real-time behavioural data that do not suffer from researcher intervention or hindsight bias. The authors argue that as such, this type of data provides a promising yet currently largely untapped empirical context for HRM research that is particularly suited for examining discourses and behavioural and social patterns over time.

Practical implications – While online data hold promise for many novel research questions, it is less appropriate for research questions that seek to establish causality between variables. When using online data, particular attention must be paid to ethical considerations, as well as the validity and representativeness of the sample.

Originality/value – The authors introduce online data and network text analysis as a new avenue for HRM research, with potential to address novel research questions at micro-, meso- and macro-levels of analysis.

Keywords Network text analysis, Online data, HRM research

Paper type Conceptual paper

Introduction

Recent advances in digital technology have resulted in the production and availability of large amounts of online information from social media posts to digitised libraries (Evans and Aceves, 2016; Light, 2014). This vast body of material provides an increasingly important source of data that complements traditional quantitative and qualitative data and allows researchers to ask novel research questions as well as unravel robust patterns about organisational and social phenomena. Online data represent a source of real-time yet longitudinal behavioural data that do not suffer from researcher intervention or hindsight bias (Hookway, 2008; Thelwall, 2006), but the relationships can be examined as they naturally occur (Scandura and Williams, 2000). The availability of large online data sets – online “Big Data” – with detailed information at the macro-, meso- and micro-level has been growing exponentially in the last couple of years due to the rapid development of data

© Kalliopi Platanou, Kristiina Mäkelä, Anton Beletskiy and Anatoli Colicev. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

The authors are grateful to the Finnish Funding Agency for Technology and Innovation (Tekes) (No. 40325/14) and the Academy of Finland (No. 298225) for their generous support for this research.



storage and digital monitoring. Such online “Big Data” have the potential to tap into collective patterns of perceptions, attitudes, behaviours and social structures in ways that other sources cannot (George *et al.*, 2014; Hannigan, 2015). Consequently, online data have been used to gauge various topics ranging from stock prices and political opinion to consumer reactions and supply chain analytics (Bollen *et al.*, 2011; Chae, 2015; Lipizzi *et al.*, 2015; Sobkowicz *et al.*, 2012; Tetlock, 2007), and as an emerging decision-making tool both in the business community and for policy makers. Overall, online “Big Data” are evolving rapidly and at a speed that leaves scholars and practitioners alike attempting to make sense of its potential opportunities and risks (George *et al.*, 2016).

We argue that human resource management (HRM) scholars would benefit from embracing this new source of data that comes with the potential to open up novel research opportunities. Research in the HRM field has for a long time employed quantitative survey-based data, complemented by an increasing focus on qualitative interview and case data (Sanders *et al.*, 2014). These data sources continue to be of major relevance. However, in today’s digital age, HRM-related professional discourse is fast moving to online domains, where HR practitioners read about the latest news and trends in the field, share their experiences, and discuss current people management issues. Such interactions occur mostly in textual format embedded in online articles, blogs, tweets and discussions in online forums. Although unstructured, they provide evidence of “what people do, know, think, and feel” (Evans and Aceves, 2016, p. 22) in a far more naturally occurring form than questionnaires and interviews do.

In this paper, we first identify and discuss key sources of online data for HRM. We then turn to the methodological implications of utilising online data in HRM research – how online data can be analysed both through more traditional methods such as content and discourse analysis, and through more novel approaches, in particular network-based text analysis (referred to as network text analysis in the remainder of this paper; see e.g. Popping, 2000; Roberts, 2000; Smith and Humphreys, 2006). We focus on the latter, suggesting a four-step process for network text analysing online data. Finally, we conclude by setting out a future HRM research agenda, in which online data and network text analysis can complement traditional quantitative (survey-based) and qualitative (interview-based) methods at different levels of analysis.

Online data in HRM research: value added, sources and challenges

A number of different terms and definitions have been used to describe what we, in this paper, call online data. Online data have been referred to as “digital data” (Venturini *et al.*, 2014) on the one hand, and “community data” on the other (George *et al.*, 2014), with subtle differences. Furthermore, recent business and management literature has referred to extensive, various and complex online data as “Big Data” (George *et al.*, 2016), a term borrowed from computer science (Mashey, 1997). We adopt the term online data and define it as the textual data that have been produced in one or more online domains, such as websites, online news, social media platforms, blogs and discussion forums, and, which, upon their production, were not intended to be used in organisational research (Venturini *et al.*, 2014). Other potential internet-based data sources, such as internet surveys and online focus groups (Granello and Wheaton, 2004; Hewson and Laurent, 2008), as well as numerical big data (e.g. from company-internal systems), are beyond the scope of this paper. In this section, we discuss the value added of online data, together with potential online data sources for HRM research, and challenges related to working with online data.

Potential value added

New digital technologies have allowed individuals to share their experiences, opinions and thoughts about various topics related to people management online and make them accessible

to practitioner communities, which, in turn, provide the opportunity for others to read and react through comments (Barros, 2014; Dellarocas, 2003). This user-generated content and the dynamic and “fluid” nature of interactions are distinctive aspects of online data (Barros, 2014; Scott and Orlikowski, 2009, p. 9). Compared to other more traditional qualitative data sets such as company material, media texts or interviews, online data represent “new forms of distributed, collective knowledge-sharing” (Scott and Orlikowski, 2009, p. 2), which may capture and represent organisational, professional and public discourse at different levels. Another benefit of using online data lies within its potential of providing us with longitudinal data in a much less tedious and time-consuming way than longitudinal interview or survey designs do (Granello and Wheaton, 2004). For HRM researchers, this means that we could – relatively easily and inexpensively – collect large volumes of longitudinal behavioural data produced by specific groups (e.g. business managers and professionals, HR professionals) in the form of real-time, naturally occurring communication. As useful analogical examples in the broader field of organisation studies, Sillince and Brown (2009) analysed police websites in order to identify how multiple organisational identities are constructed, and Barros (2014) explored how organisations use corporate blogs as tools of legitimacy.

Data sources

The amount of both public and private online data is constantly growing, and key current sources include popular platforms such as LinkedIn and Twitter, in particular. LinkedIn, the largest professional online social network with over 200 million members, has the potential to become a particularly important online data source for HRM research (Fawley, 2013), as it enables the collection of detailed data on professional discussions, together with public user profiles and connections between users. Tools for LinkedIn data extraction and analysis already exist in the field of computer science (Barroso *et al.*, 2013). At the same time, private conversations (i.e. closed groups and messages) are protected and cannot be accessed by automatic data extraction tools (i.e. “crawlers”), unless users allow it.

Twitter’s benefit, in turn, is that it has a very large user base with over 300 million active users per month, producing around 500 million tweets a day (Statista, 2015). Researchers can access Twitter’s own Automated Programming Interface (API) and use keywords to collect textual information in the form of tweets. Topic-specific tweets can be collected through hashtags (#), and retweets contain information on the user network structure and content’s virality. In terms of limitations, Twitter’s 140 character counts is a significant downside, as classic tools for analysing text content tend to perform better with more text input. Another important limitation of Twitter is that researchers can only go back in time for no more than three months; beyond this point, data must be bought from Twitter or secondhand parties.

Although large with over 1.7 billion users (Statista, 2015) and the most accessible API of all social media sites and full history (via the Open Graph Interface), Facebook’s content has, at least, thus far been more relevant for fields such as marketing than for HRM. Typical data include public discussions in the form of user posts and comments on companies’ walls, and as companies keep their walls public, such information can be freely collected (although the content may be heavily moderated). In addition, volumes of historical likes, shares and comments have been used, for instance, to investigate the public perception about a certain topic, company or brand, and Facebook’s availability in different languages has created opportunities for regional analyses. Again, and importantly, private profiles and conversations are protected and cannot be used without user consent.

Other potentially interesting online data sources include YouTube, the most popular platform for video content consumption with four billion minutes of videos watched per month (Statista, 2015). Users and companies increasingly spread professional content on YouTube in the form of lectures, tutorials on best practices, which can also be commented on.

Lastly, several professional associations and commercial platforms in the HRM field publish extensively through their websites and host blog and discussion forums. Although these provide perhaps the most HRM-focussed online material, their texts are often published under copyright and/or in semi-private forums (e.g. under membership), and require consent.

Challenges

While the value of using online data in opening up new research opportunities seems undisputed, the inherent features of online data also introduce a number of challenges. Most importantly, we do not know the “conditions of their production” (Venturini *et al.*, 2014, p. 2). We need to acknowledge that this type of data was not created by online users for the purpose of research, which although providing the benefit of being unfiltered and naturally occurring, has some crucial limitations with regards to its use for research purposes, ethical issues, representativeness, validity and applicability.

Despite the increasing use of online data in research, ethical guidelines have not yet received the required attention. In an attempt of reviewing the current frameworks and guidelines for conducting online research, Sugiura *et al.* (2016) conclude that the present ethical guidance remains somewhat inconsistent and does not resolve the challenges regarding the issues of the informed consent, privacy and anonymity of informants. Hence, the first major challenge in using online data for research relates to ethical considerations. The lack of clarity with regard to ownership and control over online texts has important ramifications for anyone carrying out online research. At heart, this is a question of what exactly constitutes private and public in an online environment, which is often difficult to determine (Markham and Buchanan, 2012; Orton-Johnson, 2010; Waskul and Douglass, 1996). Although online textual data are public information (with the exception of discussions that take place in a protected online environment such as private social media or forums that require a membership), they do not necessarily fall into public domain, meaning they would be free from any form of copyright protection. In other words, although a body of online material can be publicly viewed, the terms and conditions of the website may not allow their use for other than personal purposes.

To resolve these issues, it is important for researchers to familiarise themselves with potential intellectual property rights and copyright issues before starting data collection (Hookway, 2008; Orton-Johnson, 2010), and contact the website owners for potential permission. Most major online platforms have a detailed policy for using their data for research purposes and many allow fair and non-commercial uses of their data, however, often in a restricted or altered format. For example, while posts and comments on companies' Facebook public walls are freely available, private conversations on such social media cannot be fetched. Others will require written consent, which may sometimes be difficult to gain, as the use of online data is relatively new and professional (non-media) organisations, in particular, are cautious given the lack of established guidelines. For example, data from closed LinkedIn group discussions would require user consent. Lastly, it has been argued that it is, in fact, the users generating the content, who have the least control over their texts (Felt, 2016). Retaining the anonymity of individual authors and users is particularly important when using the online material since it is easier to be traced and identified than using traditional (non-online and non-public) data.

A second pitfall of doing online research relates to the representativeness and validity of the sample. It is often difficult to argue that a specific online sample represents the general (or traditional) population of interest, as the demographics of online users may be biased in terms of, for example, industry, age or technological orientation (Woodfield *et al.*, 2013). Although an increasing number of HR practitioners maintain a professional online presence, it is clear that not all participate in online discussions and a considerable number of them are “lurking” rather than being active contributors. Yet, the very large amount of data and the

number of users who have generated the data allow for richer data and greater diversity in the sample than traditional smaller-scale samples are able to do (Venturini *et al.*, 2014).

In a similar vein, given that individual authors are likely to have different reasons for participating (be that pushing an agenda, sharing experiences or looking for advice on people-related issues), individual posts will likely be biased in one way or the other. For example, the authors may treat the examined phenomenon in either an extremely positive or negative manner in order to promote their own vested interests, such as their expertise as consultants or their firm's services and products. In turn, the threat of this to validity and reliability will depend on the research question (see below), and the issue of what the informants share with the researchers is a challenge not only for online, but all forms of data. At the same time, individual-level biases are less salient and will be "averaged out" in large bodies of data.

Third and related to this, while online data are appropriate for certain types of research questions – particularly those that look at behavioural and social patterns over time and across different sub-groups – it cannot be used in a meaningful way for all types of research questions. More specifically, research questions that seek to establish causality between variables will be challenging. While this is an obvious limitation, all types of data have advantages and disadvantages in terms of what kinds of research questions they can accommodate, and it is important to see online data as a complement – a source that allows us to examine questions that have thus far been difficult to get to – rather than a replacement of traditional data sources. As an example, Angouri and Tseliga (2010), in their research on examining how members of online forums use the perceptions of impoliteness and disagreement in constructing the group's norms, combined online and traditional sources of data which included observations of the online domain, members' posts in the forum, survey and interviews with members. A good practice for HRM researchers who want to use online data is to establish the boundaries of their research based on well-informed research questions and be upfront with regards to their limitations. We will now move on to describe different methods for analysing online data, focussing on network text analysis as a method particularly suitable for quantifying large sets of unstructured textual data.

Analysing online data

Online data can be analysed with traditional qualitative methods in much the same way as any other types of textual data. Through a range of processes and methods, qualitative data analysis aims at providing an understanding and interpretation of what individuals or groups talk about in relation to the examined phenomenon. This typically involves the identification of themes in the data, with the help of coding and grouping evidence into aggregate and abstract categories. To this end, many qualitative coding practices and techniques are strongly influenced by the grounded theory approach. Other qualitative methods that are potentially appropriate for online data analysis include discourse analysis, which focusses on the structure and linguistic features of the text (for work in the HRM domain, see e.g. Harley and Hardy, 2004), and narrative analysis, which is particularly interested in the chronology and sequence of events in the stories that people share (Barry and Elmes, 1997).

At the same time, the sheer volume and unstructured nature of online data challenge the limits of using traditional qualitative methods in a meaningful way. These same characteristics also make it difficult to employ traditional quantitative (multivariate) research methods for online data, as it cannot be easily coded. Recent methodological advances have addressed these obstacles by developing new types of methods that use computation to process large sets of "big" textual data, blurring the traditional boundaries between qualitative and quantitative methods (Klüver, 2015; Light, 2014; Marciniak, 2016; Pollach, 2012). Such methods may include, for example, text mining (e.g. Moro *et al.*, 2015;

Sunikka and Bragge, 2012), topic modelling (e.g. DiMaggio *et al.*, 2013) and network text analysis approaches such as map analysis (e.g. Carley and Palmquist, 1992; Carley, 1993), semantic networks (e.g. Sowa, 1992) and mental models (e.g. Carley, 1997). While each of these methods employs different techniques, they all seek to quantify large sets of unstructured textual data by statistically computing relations between different textual elements, most commonly the frequency and co-occurrence of words or concepts. The basic underlying assumption of these methods is to identify patterns and underlying structures in the text (Diesner and Carley, 2005). To date, probably the most influential and commonly used model is latent dirichlet allocation (LDA) (Blei *et al.*, 2003), a model designed to iteratively infer the concepts present in a body of text. However, new models beyond LDA have been created to overcome higher levels of complexity of latent thematic structures, such as concept hierarchies, text networks and temporal trends in themes, providing researchers with new ways to analyse, visualise and explore data (Blei, 2012).

Of these, network-based text analysis (also referred to as textual network analysis) provides a particularly promising method for examining under-explored research questions that tap into discourses, behavioural patterns and social processes within and across online communities. Network text analysis incorporates analytical techniques of the classic content analysis, such as frequency and co-occurrence of concepts or keywords, but moves beyond them in that it also allows the interpretation of linkages between concepts, much in the way social network analysis does (for an overview of social network theory and analysis, see e.g. Kildruff and Tsai, 2003).

Network text analysis identifies the extent to which different ideas, people, practices or events feature in the overall body of text across time or in different subsets of the data, giving us insights concerning the relative position and strength of different concepts within the corpus (Popping, 2000; Roberts, 2000; Smith and Humphreys, 2006). We can also measure the sentiment of how a specific topic is discussed, what other concepts are discussed in connection with the focal one, and how strongly and to which direction (positive or negative) they are associated (Godbole *et al.*, 2007; Feldman, 2013; Pang and Lee, 2008). Longitudinal research can be conducted by comparing conceptual networks across time, in which, similarly to content analysis, it is assumed that “a change in the use of words reflects at least a change in attention” (Duriau *et al.*, 2007, p. 6). In addition, the network text analysis method can be used to conduct both inductive and deductive research depending on the research question (Duriau *et al.*, 2007), in that the connections between different concepts can be explored and established either openly without pre-assumptions or for specific concepts of interest.

Network text analysis is based on the assumption that one can “construct networks of semantically linked concepts” (Popping, 2000, p. 97), and consequently, texts are represented as networks of concepts (Diesner and Carley, 2005). These are extracted from the texts by examining two sentences at the time with the help of software, and graphically illustrated with concept maps. Software packages, such as Leximancer (Smith and Humphreys, 2006), AutoMap (Carley, Columbus, Bigrigg and Kunkel, 2011; Carley, Reminga, Storrick and Columbus, 2011) or ORA (Carley, Columbus, Bigrigg and Kunkel, 2011; Carley, Reminga, Storrick and Columbus, 2011), enable an unbiased and reproducible coding of large amounts of data (Young *et al.*, 2015).

A network of concepts (or a conceptual network) consists of four basic objects: concepts, relationships, statements and maps (Carley and Palmquist, 1992). A concept is a single word such as “analytics”, or a composite expression such as “performance management”; these are presented as nodes (see next section). Grammar and words such as basic verbs, prepositions and articles are not taken into consideration (Light, 2014). A relationship refers to a “tie that links two concepts together”, represented by a line between the concepts; a statement, in turn, has to do with “two concepts and the relationship between them”, whereas a map is “a network of concepts formed from statements” (Carley, 1997, pp. 538-540).

Despite the transparency of the network analysis method, its main weakness lies within the focal unit of its analysis, the concept, which is operationalised as a word or expression. As Hannigan (2015, p. 4) puts it, there is the “issue of word sense disambiguation” suggesting that a single word may have multiple meanings (referred to as polysemy), that may or may not be in accordance with the intended meaning. For example, the word “culture” may be used to refer to corporate or national culture. It is, therefore, important to contextualise the meaning of focal words by carefully reading the extracts around them, triangulating and validating the network-based analysis with qualitative evidence. This process is described by Light (2014, p. 119) as “moving from the words to the network and back”.

Network text analysis process

In this section, we suggest a four-step process for network text analysis, which we believe to be generalisable and applicable to different projects utilising (large sets of) textual data of any type. Indeed, although network text analysis is particularly useful for analysing large volumes of text and we have discussed it here in connection with online textual data, it is important to note that it can be used for any type and volume of textual data. We illustrate the process using the Leximancer network text analysis software (Fisk *et al.*, 2009; Smith and Humphreys, 2006), on a body of text that concerns the intersection of information technology and HRM, commonly referred to e-HRM in the literature (see e.g. Bondarouk *et al.*, 2016). Please note, however, that we will in no way cover or discuss the topic of e-HRM *per se*; the purpose of this example is rather to describe the methodology and research process.

Depending on the research question, text network analysis can be conducted either inductively, which involves open-ended exploration of prominent concepts and relationships in an (often vast) body of text, or deductively focussing on specific concepts of interest (Duriu *et al.*, 2007). We will focus here on the deductive approach, and suggest that it should be thought about as a process consisting of four distinct but interlinking steps. In brief, the first step narrows down the total body of data to cover a subset of material that discusses the focal topic, such as the intersection of HRM and information technology in our example. The second step seeks to identify the most relevant focal concepts in the subset of data, in order to facilitate a meaningful network analysis. The third step takes these concepts as input and identifies the most prominent ones in a corpus of text, their potential sentiments, and relationships among the concepts, with the purpose of explaining possible relations in the data. Finally, the fourth step focusses on the interpretation of the results, which consequently involves the extraction of meaning on a higher theoretical level. These four steps are illustrated in Figure 1 and discussed in more detail below.

Step 1: keyword search to narrow the frame

The goal of the first step is to narrow down the total, often vast, body of data to cover only those texts that discuss or mention the focal topic in a meaningful way. This involves a simple keyword search that identifies texts in which the focal topic is covered. Specific care should be employed to make sure that relevant terminology is used. For example, using the term “e-HRM” does not typically yield results, as academic and practitioner audiences tend to use different language to discuss the intersection of HRM and information technology. Instead, the term “HR technology” is commonly used. Further alternate terms can be used to supplement the most prominent one, such as “software” or “web-based” in this case.

The resulting body of texts must be examined manually in terms of relevance and meaningfulness. Online data typically include a lot of noise and what to be considered depends on the research question. For example, if we are interested in how managers and

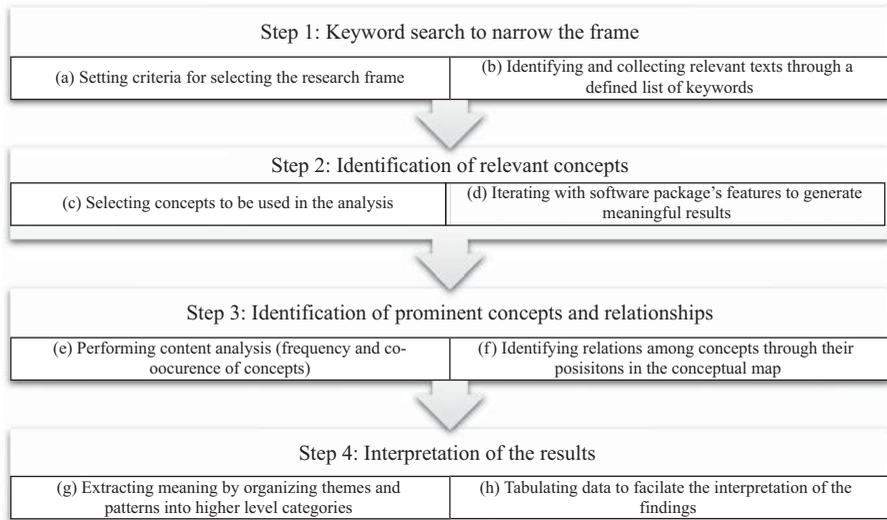


Figure 1.
Four-step analysis
process

professionals discuss the intersection of HR and technology, then advertisements, and invitations to or descriptions of events, conferences and awards may be non-relevant material that should be excluded.

Step 2: identification of relevant concepts

The second step has to do with identifying and selecting the concepts that would be used in the subsequent network analysis. To do this, we can use the software to produce an open-coded list of all words that are frequently used in the focal text corpus. This initial open-coded frequency-based list will likely include many words that are not relevant to the topic and research question; for example, generic words such as “important”, “use/using”, “terms”, “need/needs” and “job/jobs” are commonly used in professional texts. Such non-relevant or ambiguous words should be excluded from consequent analysis (see Figure 2 in Step 3).

The second task, which is often necessary, involves further refinement of search criteria, such as defining the number of concepts, merging words that bear the same meaning, and adding concepts for greater focus. For example, plural forms of words, British and American spelling, as well as synonyms (e.g. providers/vendors and hiring/recruiting) should be merged into one concept. Lastly, Leximancer allows the creation of compound words (e.g. performance management) for more meaningful analysis of the maps and data. Table I outlines central HR and technology-related concepts featuring in the illustrative example.

Step 3: identification of prominent concepts and relationships

The third step aims at analysing the prominence of the concepts and their proximity to each other, with the goal of identifying relationships among concepts. The Leximancer software uses the Bayesian theory to perform content analysis in two stages, semantic and relational. Specifically, it produces a ranked list of concepts, providing numerical measures of relative frequency, strength, prominence, association and sentiment, and a concept map which is a visual representation of the global corpus of text (for more information, see Smith and Humphreys, 2006). Their algorithm transforms the concepts occurring in the

text into a “picture of cognitive impressions” (Purchase *et al.*, 2016 p. 4). The size and brightness of the concept dot indicate the concept’s strength while the brightness and thickness of connections between concepts show the frequency of concepts’ co-occurrence (Indulska *et al.*, 2012). The proximity of the concepts on the map represents the degree to which the concepts “travel” together in the original text.

The interpretation of the concept map is further facilitated by larger themes, their prevalence determined by the number of concepts present in the theme. The colour of the theme indicates its prominence, with those being close to red constituting the most prominent themes. The size of the theme circle, on the other hand, indicates its boundaries only, but has no bearing as to its prevalence or importance in the text. Figure 2 provides an illustrative example of a concept map, together with frequency and relevance data, produced by the Leximancer software. In this example, social media is the most common theme discussed within the text corpus (i.e. the “hottest” circle, denoted by the colour red). The relevance of a concept represents the percentage frequency of the two-sentence segments in which the concept can be found, relative to the frequency of the most frequent concept in the list (in this case, unsurprisingly, “HR”). The most frequent concept will always be noted as 100 per cent but does not mean all text segments will contain the concept. As can be seen from Figure 2, an unforced frequency list will include many generic concepts, such as “favourable” and “company” in this case, and depending on the research question these often need to be removed (see Step 2).

Step 4: interpretation of the results

The fourth stage focusses on organising the emerging themes and patterns into aggregate categories which are juxtaposed with theory. The purpose of this fourth step is the extraction of meaning on a higher theoretical level, in a very similar way that social network analysis results are typically interpreted (see e.g. Kildruff and Tsai, 2003). It is important to note, that given that the software is based on the Bayesian theory, themes are subject to changes and the maps should only be used as a guide, and the interpretation of findings should be based on the relative importance of concepts and their relationships, and validated through qualitative evidence supporting the identified relationships.

The interpretation of the results is guided by a theoretical framework, which reflects the specific perspectives, research approach (inductive or deductive) and philosophical assumptions guiding a specific research process. One can create different figures and tables that allow the identification of patterns over time or across subsets of data. As a simple example, the longitudinal analysis of concept relevance in the illustrative example (see Figure 3) suggests that interest in social media peaked in 2012, but its popularity in the HR discourse has decreased considerably since then.

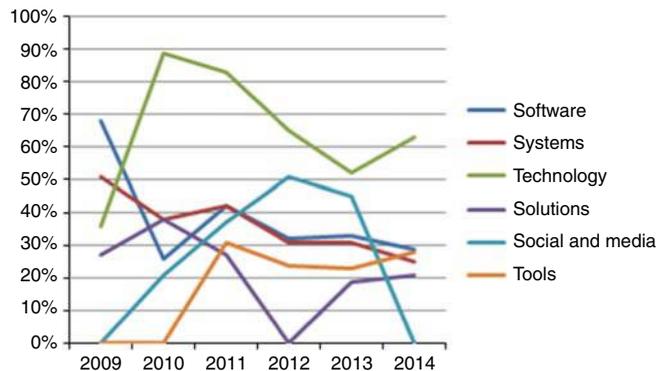


Figure 3. Illustrative example of the over-time evolvement of some HR and technology – related concepts, as per concept relevance

Finally, we recommend that any patterns observed in the network text analysis should be subjected to consequent qualitative analysis in order to gain the most insight. Such quantification of qualitative data allows us to “first zoom out from rich but disordered qualitative data to detect quantifiable general patterns, and then to zoom back in to explore these patterns in depth” (Barner-Rasmussen *et al.*, 2014, p. 14).

Discussion: new directions for HRM research

In this paper, we have argued that online data provide a promising yet currently largely untapped empirical context for HRM research, identified key sources of online data for HRM, discussed potential value added and challenges, introduced network text analysis as a key research method for analysing online material, and suggested a four-step process for the analysis of such data. In what follows, we outline some potentially interesting new research avenues for HRM that we believe online data and network text analysis can contribute to.

To be able to exploit the benefits of online data, it is important to begin by defining the “right” research questions and acknowledging their limitations. Both aspects were discussed above in more general terms, and next, we will contemplate some interesting research questions addressing which online data and network text analysis are particularly promising, followed by some important limitations that need to be considered (see Table II for a summary). In terms of the level of analysis, online data can potentially contribute to answering questions at the macro- (the HR profession, countries and industries), meso- (groups, networks and communities) and micro- (individual) level, and importantly, also multi-level questions spanning all three (Renkema *et al.*, 2016).

First, at the macro-level, online data may enable us to map different discourses within the HRM field more broadly than we have not been able to before, and look at their emergence and evolution both over time and across different groups, industries or countries. Online textual “Big Data” can reach new professional populations and/or the “wisdom of the crowd” related to HRM questions, and push beyond studies that rely on a single source of data. It can broaden our research focus from HRM within organisations to HRM as a profession; we can look at the development of professional discourses over time, and potentially identify how and why different topics, opinions and sentiments (Thelwall and Buckley, 2013), or management innovations, fashions and fads (Birkinshaw *et al.*, 2008)

Research questions	Data collection	Data analysis
<p>Macro-level: professional and managerial discourses and dominant logics within the HRM field; HRM as a profession; evolution of the HRM field</p> <p>Meso-level: interactions and influence tactics among HR actors, including practitioners, consultants, academics, and opinion leaders; legitimization of HR practices</p> <p>Micro-level: ethnographic research focussing on what HR professionals and practitioners really do; HR professional identity; HR roles at the individual level</p> <p>Multi-level: cross-level analyses of the above topics</p>	<p>Online data as a complementary source to traditional data sets</p> <p>Strengths: (a) real-time, behavioural and naturally occurring communication data; (b) longitudinal data; (c) no researcher’s intervention; (d) “collective knowledge-sharing” (Scott and Orlikowski, 2009, p. 2); (e) easy and inexpensive to collect</p> <p>Limitations: (a) ownership and copyright; (b) representativeness of the sample; (c) validity and reliability in relation to sample bias; (d) potential ethical issues related to the anonymity of individual authors; (e) difficult to establish causality among variables</p>	<p>Select the “right” software package based on the objectives of the research</p> <p>Four-step analysis: 1. keyword search to narrow the frame</p> <p>2. Identification of the relevant concepts</p> <p>3. Analysis of prominent concepts and identification of possible relations among concepts</p> <p>4. Interpretation of the results based on the conceptual framework and epistemological assumptions</p>

Table II.
Some key considerations for using online data and network text analysis in HRM research

become popularised and influence HRM practice. Mapping how traditional HRM practices, such as performance or talent management, for example, are discussed over time and in different institutional, cultural or industry contexts may give us new insights into comparative HRM (e.g. Mayrhofer *et al.*, 2011). Online discourses may also help us understand the development of the HRM field or the HR profession from novel or complementary perspectives, such as cognitions and dominant institutional logics (e.g. Kaplan, 2011), legitimisation (e.g. Suddaby and Greenwood, 2005), or credibility, power, and influence (e.g. Bouquet and Birkinshaw, 2008).

Second, at the meso-level, using online data from social network platforms and discussion forums, we can examine how different types of actors and protagonists interact and influence each other. Such actors include not only HR managers, but also professional associations, consultants, academics and other opinion leaders. This way, we could gain more fine-tuned insights to how ideas travel and disseminate (Hauptmeier and Heery, 2014), how different HR actors seek to legitimise their ideas (e.g. Pohler and Willness, 2014) and use influence tactics, often without formal authority. For example, and as discussed above, LinkedIn enables researchers to collect data on users, professional networks, trends and textual content, and Twitter's "retweets" can explore the virality of ideas, while hashtags are relevant for specific topic search and analysis.

A substantial part of the HRM professional discourse is moving towards online spaces forming distinct communities of practice. Broadly speaking, communities of practice have been defined as "groups of people who share a concern, a set of problems, or a passion about a topic, and who deepen their knowledge and expertise in this area by interacting on an ongoing basis" (Wenger *et al.*, 2002, p. 4). Online data, and more specifically, blogs and discussion forums, allow us to focus on examining issues that relate to how these communities are formed and developed over time and identifying the strategies that the members use in order to negotiate their position within the community. This could potentially generate valuable insights regarding the key actors that drive change and innovation within the HRM field.

Third, at the micro-level, there is a possibility for HRM researchers to do ethnographic research by following HR professionals on a variety of online media such as personal blogs, Twitter accounts or LinkedIn profiles (referred to as digital ethnography; Murthy, 2008 or network ethnography; Howard, 2002). This type of research can address research questions concerning what HR professionals do (Welch and Welch, 2012), how they define their role and competencies explicitly and implicitly (Brockbank *et al.*, 2012), and how they construct their professional identity (Alvesson and Kärreman, 2007). Following real-time online discussions can give us insights into questions related to the "everyday actions and behaviours of HR professionals" (Björkman *et al.*, 2014, p. 133), without hindsight bias or researcher intervention.

Lastly, many of the above questions can be examined across different levels of analysis, as individual online entries often take place in collective forums, which can be aggregated by group (e.g. managers vs consultants), industry and even country, across different time periods. This is important because complex organisational phenomena are typically socially embedded (Sanders *et al.*, 2014), have microfoundational drivers (Felin *et al.*, 2012), and are consequently difficult to examine in a rigorous and comprehensive manner without drawing on influences from different levels (Renkema *et al.*, 2016). We can also, potentially, connect people management with broader issues such as technological, economic or political development.

At the same time, online data have some important limitations. Common disadvantages and pitfalls were discussed in detail above and summarised in Table II. These need to be considered carefully, but they may or may not be limiting for all research questions. For example, the ways in which different HR actors discuss relevant topics both reflect their perceptions of HRM at each point of time and collectively provide an appropriate textual

source for examining underlying discourses. In fact, some of the issues – such as participants pushing their own agendas – can become topics of interest on their own sake. Active online participants can be considered important constituent actors, who influence the formation of professional discourse and adoption of organisational practices in the HR field. For example, we may gain a deeper insight into the adoption of e-HRM practices when we examine how technology is discussed among these protagonists (management innovators, thinkers, consultants, service providers and other forerunners who seek to drive and disseminate changes in the field) and users (HR managers who are actively interested in e-HRM-related issues). We could even go as far as arguing that different protagonist agendas have become an important part and driver of change, which has received too little research attention in the past.

In conclusion, we view the online textual material as a complementary source of data that has the potential to provide richer insights into extant HRM phenomena. What is more, the unique characteristics of online data provide us with an opportunity to introduce a number of potentially interesting new research questions that have to a large extent been absent in the HRM literature. We specifically suggest that online data and network text analysis would benefit the HRM field by allowing researchers to identify patterns in large amounts of textual data, both real-time and longitudinally, and across levels of analysis. By doing so, they can contribute to recent calls for quantifying qualitative data (see e.g. Latour *et al.*, 2012; Marciniak, 2016) and for an increased focus on incorporating “Big Data” in management research (George *et al.*, 2014, 2016).

As a final note, it needs to be emphasised that the area is rapidly developing in that new machine learning tools such as Artificial Intelligence algorithms (e.g. neural networks, ant colony) are constantly emerging (George *et al.*, 2016). Yet, human researchers continue to be at the heart of the online Big Data revolution (Mayer-Schönberger and Viktor Cukier, 2013). At the core, online data require not only methods and tools that are able to deal with complexity and ambiguity, but – and importantly – also analytical judgement and skills on behalf of researchers in ensuring appropriate research design, making sure the data and method choices are appropriate for the research question and empirical design, and in interpreting the data and drawing theoretical conclusions.

References

- Alvesson, M. and Kärreman, D. (2007), “Unraveling HRM: identity, ceremony, and control in a management consulting firm”, *Organization Science*, Vol. 18 No. 4, pp. 711-723.
- Angouri, J. and Tseliga, T. (2010), “‘You have no idea what you are talking about!': from e-disagreement to e-impoliteness in two online fora”, *Journal of Politeness Research*, Vol. 6 No. 1, pp. 57-82.
- Barner-Rasmussen, W., Ehrnrooth, M., Koveshnikov, A. and Mäkelä, K. (2014), “Cultural and language skills as resources for boundary spanning within the MNC”, *Journal of International Business Studies*, Vol. 45 No. 7, pp. 886-905.
- Barros, M. (2014), “Tools of legitimacy: the case of the Petrobras corporate blog”, *Organization Studies*, Vol. 35 No. 8, pp. 1211-1230.
- Barroso, L.A., Clidas, J. and Hölzle, U. (2013), “The ‘big data’ ecosystem at LinkedIn”, *Proceedings of the VLDB Endowment*, Vol. 6 No. 11, pp. 1-6.
- Barry, D. and Elmes, M. (1997), “Strategy retold: toward a narrative view of strategic discourse”, *Academy of Management Review*, Vol. 22 No. 2, pp. 429-452.
- Birkinshaw, J., Hamel, G. and Mol, M.J. (2008), “Management innovation”, *Academy of Management Review*, Vol. 33 No. 4, pp. 825-845.
- Björkman, I., Ehrnrooth, M., Mäkelä, K., Smale, A. and Sumelius, J. (2014), “From HRM practices to the practice of HRM: setting a research agenda”, *Journal of Organizational Effectiveness: People and Performance*, Vol. 1 No. 2, pp. 122-140.

- Blei, D. (2012), "Topic modeling and digital humanities", *Journal of Digital Humanities*, Vol. 2 No. 1, pp. 8-11.
- Blei, D., Ng, A.Y. and Jordan, M.I. (2003), "Latent dirichlet allocation", *Journal of Machine Learning Research*, Vol. 3, January, pp. 993-1022.
- Bollen, J., Mao, H. and Zeng, X. (2011), "Twitter mood predicts the stock market", *Journal of Computational Science*, Vol. 2 No. 1, pp. 1-8.
- Bondarouk, T., Parry, E. and Furtmueller, E. (2016), "Electronic HRM: four decades of research on adoption and consequences", *The International Journal of Human Resource Management*, Vol. 28 No. 1, pp. 98-131.
- Bouquet, C. and Birkinshaw, J. (2008), "Managing power in the multinational corporation: how low-power actors gain influence", *Journal of Management*, Vol. 34 No. 3, pp. 477-508.
- Brockbank, W., Ulrich, D., Younger, J. and Ulrich, M. (2012), "Recent study shows impact of HR competencies on business performance", *Employment Relations Today*, Vol. 39 No. 1, pp. 1-7.
- Carley, K.M. (1993), "Coding choices for textual analysis: a comparison of content analysis and map analysis", *Sociological Methodology*, Vol. 23, January, pp. 75-126.
- Carley, K.M. (1997), "Extracting team mental models through textual analysis", *Journal of Organizational Behavior*, Vol. 18 No. 1, pp. 533-558.
- Carley, K.M. and Palmquist, M. (1992), "Extracting, representing, and analyzing mental models", *Social Forces*, Vol. 70 No. 3, pp. 601-636.
- Carley, K.M., Columbus, D., Bigrigg, M. and Kunke, F. (2011), *AutoMap User's Guide 2011: Carnegie Mellon University, School of Computer Science*, technical report, CMU-ISR-11-108, Institute for Software Research.
- Carley, K.M., Reminga, J., Storrick, J. and Columbus, D. (2011), *ORA User's Guide 2011: Carnegie Mellon University, School of Computer Science*, technical report, CMU-ISR-11-107, Institute for Software Research.
- Chae, B. (2015), "Insights from hashtag #supplychain and Twitter analytics: considering Twitter and Twitter data for supply chain practice and research", *International Journal of Production Economics*, Vol. 165, July, pp. 247-259.
- Dellarocas, C. (2003), "The digitization of word of mouth: promise and challenges of online feedback mechanisms", *Management Science*, Vol. 49 No. 10, pp. 1407-1424.
- Diesner, J. and Carley, K.M. (2005), "Revealing social structure from texts: meta-matrix text analysis as a novel method for network text analysis", *Causal Mapping for Information Systems and Technology Research: Approaches, Advances, and Illustrations*, pp. 81-108.
- DiMaggio, P., Nag, M. and Blei, D. (2013), "Exploiting affinities between topic modeling and the sociological perspective on culture: application to newspaper coverage of US Government arts funding", *Poetics*, Vol. 41 No. 6, pp. 570-606.
- Duriau, V.J., Reger, R.K. and Pfarrer, M.D. (2007), "A content analysis of the content analysis literature in organization studies: research themes, data sources, and methodological refinements", *Organizational Research Methods*, Vol. 10 No. 1, pp. 5-34.
- Evans, J. and Aceves, P. (2016), "Machine translation: mining text for social theory", *Annual Review of Sociology*, Vol. 42 No. 1, pp. 21-50.
- Fawley, N. (2013), "LinkedIn as an information source for human resources, competitive intelligence", *Online Searcher*, Vol. 37 No. 3, pp. 31-50.
- Feldman, R. (2013), "Techniques and applications for sentiment analysis", *Communications of the ACM*, Vol. 56 No. 4, pp. 82-89.
- Felin, T., Foss, N., Heimericks, K.H. and Madsen, T. (2012), "Microfoundations of routines and capabilities: individuals, processes, and structure", *Journal of Management Studies*, Vol. 49 No. 8, pp. 1351-1374.
- Felt, M. (2016), "Social media and the social sciences: how researchers employ big data analytics", *Big Data & Society*, Vol. 3 No. 1, pp. 1-15.

- Fisk, K., Cherney, A., Hornsey, M. and Smith, A. (2009), *Rebuilding Institutional Legitimacy in Post-conflict Societies: An Asia Pacific Case Study – Phase 1A*, Air Force Office of Scientific Research, Asian Office of Aerospace Research and Development, University of Queensland, Brisbane.
- George, G., Haas, M. and Pentland, A. (2014), “Big data and management”, *Academy of Management Journal*, Vol. 57 No. 2, pp. 321-326.
- George, G., Osinga, E., Lavie, D. and Scott, B. (2016), “Big data and data science methods for management research”, *Academy of Management Journal*, Vol. 59 No. 5, pp. 1493-1507.
- Godbole, N., Srinivasaiyah, M. and Skiena, S. (2007), *Large-Scale Sentiment Analysis for News and Blogs, Proceedings of the International Conference on Weblogs and Social Media (ICWSM 07)*, Boulder, CO.
- Granello, D.H. and Wheaton, J. (2004), “Online data collection: strategies for research”, *Journal of Counselling and Development*, Vol. 82 No. 4, pp. 387-393.
- Hannigan, T. (2015), “Close encounters of the conceptual kind: disambiguating social structure from text”, *Big Data & Society*, Vol. 2 No. 2, pp. 1-6.
- Harley, B. and Hardy, C. (2004), “Firing blanks? An analysis of discursive struggle in HRM”, *Journal of Management Studies*, Vol. 41 No. 3, pp. 377-400.
- Hauptmeier, M. and Heery, E. (2014), “Ideas at work”, *The International Journal of Human Resource Management*, Vol. 25 No. 18, pp. 2473-2488.
- Hewson, C. and Laurent, D. (2008), “Research design and tools for internet research”, in Fielding, N.G., Lee, R.M. and Blank, G. (Eds), *The Sage Handbook of Online Research Methods*, Sage, London, pp. 58-78.
- Hookway, N. (2008), “Entering the blogosphere’: some strategies for using blogs in social research”, *Qualitative Research*, Vol. 8 No. 1, pp. 91-113.
- Howard, P.N. (2002), “Network ethnography and the hypermedia organization: new media, new organizations, new methods”, *New Media & Society*, Vol. 4 No. 4, pp. 550-574.
- Indulska, M., Hovorka, D.S. and Recker, J. (2012), “Quantitative approaches to content analysis: identifying conceptual drift across publication outlets”, *European Journal of Information Systems*, Vol. 21 No. 1, pp. 49-69.
- Kaplan, S. (2011), “Research in cognition and strategy: reflections on two decades of progress and a look to the future”, *Journal of Management Studies*, Vol. 48 No. 3, pp. 665-695.
- Kildruff, N. and Tsai, W. (2003), *Social Networks and Organizations*, Sage, London.
- Klüver, H. (2015), “The promises of quantitative text analysis in interest group research: a reply to Bunea and Ibenskas”, *European Union Politics*, Vol. 16 No. 3, pp. 456-466.
- Latour, B., Jensen, P., Venturini, T., Grauwin, S. and Boullier, D. (2012), “‘The whole is always smaller than its parts’ – a digital test of Gabriel Tardes’ monads”, *The British Journal of Sociology*, Vol. 63 No. 4, pp. 590-615.
- Light, R. (2014), “From words to networks and back: digital text, computational social science, and the case of presidential inaugural addresses”, *Social Currents*, Vol. 1 No. 2, pp. 111-129.
- Lipizzi, C., Iandoli, L. and Ramirez Marquez, J.E. (2015), “Extracting and evaluating conversational patterns in social media: a socio-semantic analysis of customers’ reactions to the launch of new products using Twitter streams”, *International Journal of Information Management*, Vol. 35 No. 4, pp. 490-503.
- Marciniak, D. (2016), “Computational text analysis: thoughts on the contingencies of an evolving method”, *Big Data & Society*, Vol. 3 No. 2, pp. 1-5.
- Markham, A. and Buchanan, E. (2012), “Ethical decision making and internet research recommendations from the AoIR ethics working committee (Version 2.0)”, Association of Internet Researchers, available at: <http://aoir.org/reports/ethics2.pdf> (accessed 10 December 2016).
- Mashey, J.R. (1997), *Big Data and the Next Wave of Infrastrass. Computer Science Division Seminar*, University of California, Berkeley, CA.

- Mayer-Schönberger, V. and Cukier, K. (2013), *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, Houghton Mifflin Harcourt, Boston, MA.
- Mayrhofer, W., Brewster, C., Morley, M. and Ledolter, J. (2011), "Hearing a different drummer? Convergence of human resource management practices in Europe – a longitudinal analysis", *Human Resource Management Review*, Vol. 21 No. 1, pp. 50-67.
- Moro, S., Cortez, P. and Rita, P. (2015), "Business intelligence in banking: a literature analysis from 2002 to 2013 using text mining and latent Dirichlet allocation", *Expert Systems with Applications*, Vol. 42 No. 3, pp. 1314-1324.
- Murthy, D. (2008), "Digital ethnography: an examination of the use of new technologies for social research", *Sociology*, Vol. 42 No. 5, pp. 837-855.
- Orton-Johnson, K. (2010), "Ethics in online research", in Hughes, J. (Ed.), *SAGE Internet Research Methods*, Vol. 15, Sage Publications Ltd, London, pp. 305-315.
- Pang, B. and Lee, L. (2008), "Opinion mining and sentiment analysis", *Foundations and Trends in Information Retrieval*, Vol. 2 Nos 1/2, pp. 1-135.
- Pohler, D. and Willness, C. (2014), "Balancing interests in the search for occupational legitimacy: the HR professionalization project in Canada", *Human Resource Management*, Vol. 53 No. 3, pp. 467-488.
- Pollach, I. (2012), "Taming textual data: the contribution of corpus linguistics to computer-aided text analysis", *Organizational Research Methods*, Vol. 15 No. 2, pp. 263-287.
- Popping, R. (2000), *Computer-Assisted Text Analysis*, Sage Publications, London.
- Purchase, S., Rosa, R.D.S. and Schepis, D. (2016), "Identity construction through role and network position", *Industrial Marketing Management*, Vol. 54, April, pp. 154-163.
- Renkema, M., Meijerink, J. and Bondarouk, T. (2016), "Advancing multilevel thinking and methods in HRM research", *Journal of Organizational Effectiveness: People and Performance*, Vol. 3 No. 2, pp. 204-218.
- Roberts, C.W. (2000), "A conceptual framework for quantitative text analysis", *Quality and Quantity*, Vol. 34 No. 3, pp. 259-274.
- Sanders, K., Cogan, J.A. and Bainbridge, H.T.J. (2014), *Research Methods for Human Resource Management*, Routledge, New York, NY and London.
- Scandura, T.A. and Williams, E.A. (2000), "Research methodology in management: current practices, trends, and implications for future research", *Academy of Management Journal*, Vol. 43 No. 6, pp. 1248-1264.
- Scott, S. and Orlikowski, W. (2009), "Getting the truth: exploring the material grounds of institutional dynamics in social media", paper presented at the 25th European Group for Organizational Studies Conference, Barcelona.
- Sillince, J.A.A. and Brown, A.D. (2009), "Multiple organizational identities and legitimacy: the rhetoric of police websites", *Human Relations*, Vol. 62 No. 12, pp. 1829-1856.
- Smith, A.E. and Humphreys, M.S. (2006), "Evaluation of unsupervised semantic mapping of natural language with Leximancer concept mapping", *Behavior Research Methods*, Vol. 38 No. 2, pp. 262-279.
- Sobkowitz, P., Kaschesky, M. and Bouchard, G. (2012), "Opinion mining in social media: modeling, simulating, and forecasting political opinions in the web", *Government Information Quarterly*, Vol. 29 No. 4, pp. 470-479.
- Sowa, J. (1992), "Conceptual graphs as a universal knowledge representation", *Computers & Mathematics with Applications*, Vol. 23 No. 2, pp. 75-93.
- Statista (2015), "Social media statistics", available at: www.statista.com/topics/1164/social-networks/ (accessed 24 August 2017).
- Suddaby, R. and Greenwood, R. (2005), "Rhetorical strategies of legitimacy", *Administrative Science Quarterly*, Vol. 50 No. 1, pp. 35-67.

- Sugiura, L., Wiles, R. and Pope, C. (2016), "Ethical challenges in online research: public/private perceptions", *Research Ethics*, pp. 1-16.
- Sunikka, A. and Bragge, J. (2012), "Applying text-mining to personalization and customization research literature – who, what and where?", *Expert Systems with Applications*, Vol. 39 No. 11, pp. 10049-10058.
- Tetlock, P.C. (2007), "Giving content to investor sentiment: the role of media in stock market", *Journal of Finance*, Vol. 62 No. 3, pp. 1139-1168.
- Thelwall, M. (2006), "Blog searching: the first general-purpose source of retrospective public opinion in the social sciences?", *Online Information Review*, Vol. 31 No. 3, pp. 277-289.
- Thelwall, M. and Buckley, K. (2013), "Topic-based sentiment analysis for the social web: the role of mood and issue-related words", *Journal of the American Society for Information Science and Technology*, Vol. 64 No. 8, pp. 1608-1617.
- Venturini, T., Baya Laffite, N., Cointet, J., Gray, I., Zabban, V. and De Pryck, K. (2014), "Three maps and three misunderstandings: a digital mapping of climate diplomacy", *Big Data & Society*, Vol. 1 No. 2, pp. 1-19.
- Waskul, D. and Douglass, M. (1996), "Considering the electronic participant: some polemical observations on the ethics of on-line research", *Information Society*, Vol. 12 No. 2, pp. 129-140.
- Welch, C. and Welch, D. (2012), "What do HR managers really do? HR roles on international projects", *Management International Review*, Vol. 52 No. 4, pp. 597-617.
- Wenger, E., McDermott, R.A. and Snyder, W. (2002), *Cultivating Communities of Practice: A Guide to Managing Knowledge*, Harvard Business Press, Cambridge, MA.
- Woodfield, K., Morrell, G., Metzler, K., Blank, G., Salmons, J., Finnegan, J. and Lucraft, M. (2013), "Blurring the boundaries? New social media, new social research: developing a network to explore the issues faced by researchers negotiating the new research landscape of online social media platforms: a methodological review paper", National Centre for Research Methods (NCRM) Networks for Methodological Innovation, Sage, University of Oxford, Southampton.
- Young, L., Wilkinson, I. and Smith, A. (2015), "A scientometric analysis of publications in the journal of business-to-business marketing 1993-2014", *Journal of Business-To-Business Marketing*, Vol. 22 Nos 1/2, pp. 111-123.

Corresponding author

Kristiina Mäkelä can be contacted at: kristiina.makela@aalto.fi

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgroupublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com