5-16-2023

STUDYING DYNAMICS AND CHANGE WITH DIGITAL TRACE DATA: A SYSTEMATIC LITERATURE REVIEW

Sandro Franzoi

Thomas Grisold

Jan vom Brocke

STUDYING DYNAMICS AND CHANGE WITH DIGITAL TRACE DATA: A SYSTEMATIC LITERATURE REVIEW

Research Paper

Sandro Franzoi, University of Liechtenstein, Vaduz, Liechtenstein, sandro.franzoi@uni.li
Thomas Grisold, University of Liechtenstein, Vaduz, Liechtenstein, thomas.grisold@uni.li
Jan vom Brocke, University of Liechtenstein, Vaduz, Liechtenstein, jan.vom.brocke@uni.li

Abstract

Digital trace data offer promising opportunities to study dynamics and change of various socio-technical phenomena over time. While we see a surge of empirical and conceptual articles, we lack a systematic understanding of why, how, and when digital trace data are or can be used to study dynamics and change. In this article, we present the findings of a systematic literature review to uncover common approaches, motivations, findings, and general themes in the existing literature. We systematically reviewed 40 studies that were published in premium outlets in the information systems field. Our review sheds light on (1) underlying purposes of such studies, (2) utilized data sources, (3) research contexts, (4) socio-technical phenomena of interest, (5) applied analytical methods, and (6) measures that are being used. Building on our findings, we point to several implications for research and shed light on avenues to advance this field in the future.

Keywords: digital trace data, dynamics, change, computational, process, temporal, literature review process theorizing, process research

1 Introduction

Digital trace data research is an emerging genre in the information systems field (Berente et al., 2019; Lindberg, 2020; Miranda et al., 2022). Broadly speaking, digital traces are residuals that are left behind when actors perform activities with digital technologies (Howison et al., 2011). Digital traces offer novel opportunities for research that studies temporal dynamics and change to explain how socio-technical phenomena take shape, persist, and dissolve *over time* (Burton-Jones et al., 2015; Langley & Tsoukas, 2017; Pentland et al., 2021). This is because digital traces are typically equipped with temporal information, indicating when a given activity has been performed (Pentland et al., 2021). Furthermore, they enable a fine-granular view of how actions are carried out in both private and work-related contexts (Pentland, 2015). Despite the increasing interest, and a growing number of conceptual and empirical articles (e.g., Berente et al., 2019; Miranda et al., 2022; Pentland et al., 2020; vom Brocke et al., 2021), we lack a systematic understanding of how, why, and when this form of research is conducted.

In this paper, we review the literature on research using digital trace data to study dynamics and change in socio-technical phenomena. Drawing from research published in renowned outlets in the information systems field, we analyze the motivations, approaches, and findings that are associated with this stream of research. By systematically reviewing 40 articles, we shed light on and discuss the following questions:

1. For what purposes are digital trace data used in studies that examine dynamics and change of socio-technical phenomena?
2. What type of data are being used, which sources do they come from and what research contexts are they embedded in?
3. How are digital traces analyzed to obtain insights into dynamics and change?
4. What are trends, themes, and patterns in research using digital trace data for studying dynamics and change?

We provide systematic answers to these questions. At the core of our research are the following observations. First, digital trace data are used to support theory building and theory testing, and they provide the grounds for developing or advancing computational techniques. Furthermore, these data are collected from different data sources, such as workflow systems, social media, or online communities. In consequence, they can be used to study socio-technical phenomena in all kinds of research contexts, including healthcare, innovation, and finance, among others. Various techniques are used to analyze these data, such as sequence analysis, econometric modeling, or process mining. Finally, digital trace data are analyzed with regard to different measures, such as complexity, frequency, and user behavior, among others.

Our review points to several implications for digital trace data research. Overall, we call for more research that appreciates and leverages how digital trace data can provide novel explanations for dynamics and change in the digital age. Along these lines, we see promising opportunities for combining different measures and methods to study various socio-technical phenomena. To embrace these opportunities, the information systems field can build up and intensify synergies with adjacent research domains—such as computer science or data science—in order to integrate new types of data sources as well as computational techniques to analyze them.

2 Research Background

2.1 Digital Trace Data Research in Information Systems Research

Digital trace data-based research is attracting increasing attention in information systems research (Miranda et al., 2022) and beyond (Lazer et al., 2020; Simsek et al., 2019). Digital trace data are broadly defined as residuals of events or activities that are left behind when actors interact with digital technologies (Freelon, 2014; Howison et al., 2011). They typically appear in large quantities and can be collected in various contexts. Since almost everything we do is mediated by or enabled through digital technologies, digital trace data are thought to provide far-reaching insights into various phenomena (Lazer et al., 2020). Furthermore, digital trace data are typically unbiased with regard to existing data collection strategies and research designs (Golder & Macy, 2014). The interest in digital trace data—both in private as well as work-related contexts—has been further fueled by considerable advancements of computational techniques to analyze them. To this end, an ever-growing frontier of pattern recognition techniques offers a variety of perspectives to analyze, explain, and predict the development of various phenomena based on digital trace data sources (Oliver et al., 2020).

In the information systems field, we see an emerging body of knowledge that addresses different aspects of digital trace data research (Berente et al., 2019; Lindberg, 2020; Miranda et al., 2022). A growing number of empirical studies has been using digital trace data to examine various socio-technical phenomena (Hukal et al., 2019; Müller et al., 2016; Pentland et al., 2021; Weinmann et al., 2022). Furthermore, methodological guidance has been provided to frame and analyze these data from different perspectives (Berente et al., 2019). Also, there have been works that examined and discussed the quality and features of digital trace data sets (Howison et al., 2011; Pentland et al., 2020; Vial, 2019). Moreover, there have been discussions about the opportunities and threats of digital trace data (Grover et al., 2020; Østerlund et al., 2020).

2.2 Studying Dynamics and Change with Digital Trace Data

Recent claims have stressed that information systems research should embrace the dynamics and change of the digital age (Benbya et al., 2020; Mousavi Baygi et al., 2021). These arguments emphasize how socio-technical phenomena dynamically evolve and change over time – often in unexpected ways. To get at these dynamics, there have been calls for adjusting established research approaches and theorizing practices (Baiyere et al., 2023; Grover & Lyytinen, 2022).

Research that foregrounds dynamics and change is typically conducted in the spirit of 'process research' (Langley et al., 2013). Generally speaking, process research refers to (empirical) studies that emphasize dynamics and change in phenomena as they take shape, persist, and dissolve over time (Langley et al., 2013; Poole & van de Ven, 2004). This stream of research is inherently concerned with temporality (Cloutier & Langley, 2020) and studies are traditionally carried out employing qualitative methods, such as interviews and/or ethnographies (Langley, 1999).

In the information systems field, process research emerged as one distinctive lens to understand socio-technical phenomena (Burton-Jones et al., 2015). Digital trace data provide new and promising opportunities for process research and, in particular, for studying dynamics and change of socio-technical phenomena (Grisold et al., 2020; Pentland et al., 2021; vom Brocke et al., 2021). Since these data tend to be equipped with temporal information, they provide fine-granular insights into which action was taken at which point in time (Pentland et al., 2020). They are, thus, thought to enable researchers to discover promising insights into the rapid change and emergent dynamics that are associated with the digital age (Benbya et al., 2020; Grisold, Kremser et al., 2023). Recent works have sought to advance process research by employing appropriate computational techniques, for example, by integrating methodological knowledge from adjacent fields, such as business process management (Kallio et al., 2022; Mendling et al., 2020; Pentland et al., 2021).

However, while we see a surging interest in the use of digital trace data to study dynamics and change of socio-technical phenomena, we lack systematic knowledge of this stream of research. Understanding how researchers in the information systems field study dynamics and change with digital trace data, what kinds of phenomena they study, and to what extent the use of digital trace data advances theoretical knowledge is important to further guide the development of the field.

3 Research Method

We conducted a systematic literature review to analyze and categorize prior work that uses digital trace data to study dynamics and change in socio-technical phenomena. In our literature review, we drew on the established approaches of vom Brocke et al. (2009) and Webster and Watson (2002).

The framework of vom Brocke et al. (2009) outlines five distinct phases of a systematic literature review. First, the underlying scope of coverage of the review should be defined, which can either be exhaustive, exhaustive and selective, representative, or central (Cooper, 1988; vom Brocke et al., 2009). We apply a representative and central review that typifies large groups of articles by using a sample. In the second phase, key terms and concepts have to be defined (vom Brocke et al., 2009), which we included in our research background (see section 2). The third step pertains to the literature search process itself, where we defined search parameters and queried databases (vom Brocke et al., 2009). We conducted a keyword search as well as forward and backward searches. In the keyword search, we used the search string *"trace data" OR "temporal data" OR "sequence data" OR "time-stamped"* to comb the full text of articles.

We placed particular emphasis on the notion of 'trace data', because this term is tightly associated with a new research genre in the information systems field (Berente et al., 2019; Miranda et al., 2022). An emerging body of knowledge has been using this term to denote a specific form of data that are recorded and collected on the grounds of interactions between digital technologies and their users (Howison et al., 2011). Research in this field includes empirical studies (e.g., Pentland et al., 2021; Weinmann et al., 2022) as well as methodological works that provide guidance on how to theorize with such data (e.g., Berente et al., 2019; Grisold et al., 2020; Pentland et al., 2020). Hence, by emphasizing the term 'trace data', and by further searching for time and sequence-related features, we expected to be able to represent this discourse.

We queried the database EBSCOhost to search through premium information systems journals. To this end, we drew from the AIS Senior Scholars' Basket (AIS, 2011). The associated journals are listed in Table 1. Our search covers articles published until the 10th of October 2022. The fourth step of the framework comprises the analysis and synthesis of the collected literature (vom Brocke et al. 2009), which we describe in our findings section (see section 4). The fifth and last phase revolves around the development of a research agenda (vom Brocke et al., 2009) and directions for future research, which we pick up in our discussion section (see section 5).

As a result of this systematic literature search process, we identified a total number of 120 articles ('hits' cf. Table 1), which were obtained through the keyword search and additional forward and backward searches. In an initial analysis, the first author scrutinized the respective titles, abstracts, and keywords to remove articles that did not concern research on dynamics and change using digital trace data. When a case was not clear, it was discussed among the co-authors to determine the article's relevance. Specifically, we included a study in our analysis if it fulfilled two criteria: (1) the study uses digital trace data (not necessarily termed this way) that were collected from people's interactions with digital technologies (Grover et al., 2020) and (2) the study has a temporal component (i.e., dynamics and change over time). After eliminating articles that were deemed irrelevant, we were left with a final sample of 31 relevant studies ('relevant' cf. Table 1). Additionally, we included 9 articles from other outlets that we deemed relevant through forward and backward searches. All relevant studies included in our literature review are marked with an asterisk (*) in the reference list.

Table 1 summarizes the respective number of retrieved articles ('hits') as well as the final sample of relevant studies that we reviewed in detail ('relevant'). After finalizing our sample, we coded all relevant articles in our review. We coded the studies according to the categories purpose, data type, data source, research context, method of analysis, and measures to find an answer to our four research questions (see section 1). The results of our analysis are outlined in the next section.

Journal	**Hits**	**Relevant**
European Journal of Information Systems	1	0
Information Systems Journal	10	1
Information Systems Research	14	6
Journal of the Association of Information Systems	21	7
Journal of Information Technology	0	0
Journal of Management Information Systems	27	6
Journal of Strategic Information Systems	0	0
Management Information Systems Quarterly	38	11
Total in keyword search	*111*	*31*
Added through forward and backward search	9	9
Total	**120**	**40**

Table 1. Number of articles identified and screened in the literature review

4 Findings

We departed from four questions that are central to our research endeavor. These questions revolve around (1) the purpose of digital trace data for research on dynamics and change, (2) the data types, data sources, research contexts, and phenomena under scrutiny (3) the analytical methods and measures, and (4) general themes and patterns that emerged throughout our analysis and synthesis. To address these questions, we conducted an in-depth analysis of all articles that we deemed relevant from our systematic literature review. Table 2 showcases three diverse studies from our review and their respective coding along the dimensions of interest. These studies are intended to illustrate the broad range of research using digital trace data to study dynamics and change and to shed light on the categories that we coded for all relevant articles. We structure our findings along the four broad areas of interest. We discuss them in the following.

Study	**Purpose**	**Data**	**Research Context**	**Analysis**	**Measures**
Nan & Lu (2014)	Theory building	Posts from a university online forum	Self-organization in online communities during a crisis	Content analysis and temporal progression of posts	Coding categories; time; frequency of posts
Weinmann et al. (2022)	Theory testing	Mouse-cursor movement data	Online fraud decisions	Regression analysis	Mouse movement speed and deviation; Fraudulent responses
Zhang et al. (2022)	Method advancement	Data collected from Wikipedia	Repetitive collaboration patterns (routines)	Sequential pattern mining analysis and clustering analysis	Frequency of sub-sequences

Table 2. Coding of exemplary studies

4.1 Purpose

The first dimension that we scrutinized for all studies during our analysis and synthesis phase is their respective *purpose*. Our goal with this analysis was to find an answer to the question: *For what purposes are digital trace data used in studies that examine dynamics and change of socio-technical phenomena?* Our analysis revealed three central purposes that studies pursue: theory building, theory testing, and methodological advancement.

Theory testing-related studies typically derive and test hypotheses from extant theory. Studies often follow a quantitative approach and consist of, for instance, experimental studies, various statistical analyses, or econometrics studies that examine socio-technical phenomena *over time*. For example, Spohrer et al. (2021) utilize digital trace data from an eHealth application to evaluate an experiment on combining behavior change techniques to foster stress alleviation. Other studies use digital trace data in an econometric research approach to investigate causal relationships of content sampling for video-on-demand services (Hoang & Kauffman, 2018) or employees' blog posts and blog readership (Aggarwal et al., 2012). Furthermore, such studies utilize digital traces from online communities (e.g., Faraj et al., 2015; Kim et al., 2018). For example, digital trace data from Wikipedia was used to investigate gender bias among contributors (Young et al., 2020), or data from Twitter to analyze information diffusion patterns (Cheng et al., 2011). Furthermore, Weinmann et al. (2022) use mouse-cursor movement data to scrutinize online fraud decisions as they are happening over time. Taken together, these studies use digital trace data to evaluate quantitative research approaches, investigate causal relationships, or collect the data as a primary source for theory testing.

The second main purpose we identified in the literature is *theory building*. In this category, we subsume studies that instantiate, modify or extend theory (Grover & Lyytinen, 2015). For example, research relies on data from dermatology clinics to theorize about the dynamics of organizational routines (Pentland et al., 2021). Vaast et al. (2017) use digital trace data collected from a microblog after an oil spill; they extend affordance theory by introducing the concept of connective affordances. Nan and Lu (2014) instantiate complex adaptive systems theory in the context of self-organization during a crisis, where they utilized trace data from an online community forum to explore the temporal dynamics of self-organization after an earthquake. Østerlund et al. (2020) discuss different approaches to the analysis of digital trace data and showcase them through data that depicts the development of a crowdsourced citizen science project over time.

The third main purpose that we identified is *methodological advancement*. Here, we can distinguish between two broad streams. In one stream, works shed light on the process of theorizing. For example, Howison et al. (2011) discuss validity issues for temporal digital trace data and provide recommendations to resolve them. Berente et al. (2019) draw on grounded theory as well as computational theory discovery to derive a general method to use digital trace data for what they refer to as 'computationally intensive theory development'. Their approach consists of four iterative processes, namely sampling, synchronic analysis, lexical framing, and diachronic analysis (Berente et al., 2019). Building on this work, Miranda et al. (2022) discuss what forms of theoretical contributions can arise from digital trace data research, stressing that outcomes can include, for example, patterns that were identified in the data set. Lindberg (2020) suggests how digital trace data can be used for theory development through an iterative process of combining human and machine pattern recognition. Finally, Hartl et al. (2023) propose how change can be explained by applying temporal bracketing to the analysis of digital trace data. Another stream is concerned with the development of computational methods for the analysis of digital trace data. For example, Yeshchenko et al. (2022) develop a method to detect changes in business processes by identifying process drifts. Kallio et al. (2022) propose a method that combines process mining with variance modeling to uncover research questions and opportunities for IS researchers.

4.2 Data, Research Contexts, and Phenomena

The second dimension in our literature review is data, research contexts, and phenomena. Here, we aim to answer the following question: *What type of data are being used, which sources do they come from and what research contexts are they embedded in?* We refer to the data types and sources, the overall research context (i.e., research setting) in which digital trace data have been collected and the phenomena being scrutinized.

Data types constitute the first central aspect. In general, digital trace data consist of large quantities of fine-granular information about activities executed through an information system (Howison et al., 2011; Lazer et al., 2020). Research that studies dynamics and change with digital trace data typically capitalizes on event logs to examine the temporal dynamics of a given phenomenon over time (Pentland et al., 2014; Pentland et al., 2020; van der Aalst, 2016). An event log includes time-stamped data about process activities and it can entail additional information such as identifiers for cases or resources that were used (Pentland et al., 2020). As a general observation, we found that data is collected either through organizational information systems (e.g., workflow systems) or open-source platforms (e.g., GitHub data). On a closer look, we find a large variety of digital trace data types that are utilized to examine dynamics and change of socio-technical phenomena. First, several studies in our review leverage data from workflow systems to theorize, test, or investigate organizational phenomena, such as organizational routines (e.g., Gaskin et al., 2014; Hartl et al., 2023; Pentland et al., 2021; Wurm et al., 2021). Second, digital trace data from online communities are used to conduct research on temporal relationships between posts and/or users (e.g., Kim et al., 2018; Vaast & Pinsonneault, 2021). Such data comprise, for example, posts and messages recorded in online communities such as GitHub (Lindberg et al., 2016) or online communities that are dedicated to specific topics, such as health (Chen et al., 2019) or cybersecurity (Benjamin et al., 2016). One popular source is Wikipedia, where digital trace data are collected to study the temporal dynamics of emergent roles (Arazy et al., 2016), gender bias (Young et al., 2020), or repetitive collaboration patterns (Zhang et al., 2022). Third, studies work with digital trace data from social networking platforms such as Twitter (Vaast et al., 2017) or other types of blogs (e.g., Aggarwal et al., 2012). These studies rely on posts and tweets to analyze the dynamics of various phenomena such as the dissemination of swine flu news (Cheng et al., 2011). Fourth, digital trace data research capitalizes on sensor data such as geospatial and GPS data from taxis (Zhang et al., 2020) or data from people's mouse movements (Weinmann et al., 2022) to infer users' decisions, intentions, and motives over time. Studies also utilize digital trace data from web applications, such as eHealth apps (Spohrer et al., 2021) or video-on-demand platforms (Hoang & Kauffman, 2018).

The second central aspect is the *research context* in which digital trace data were collected. In our review, the most commonly studied research context is healthcare, where digital traces are leveraged to investigate process event logs (Pentland et al., 2020) or the use of eHealth apps (Spohrer et al., 2021), among others. Furthermore, the research context of innovation is of focal interest in studies on online communities (e.g., Kyriakou et al., 2022) as well as studies that examine collaborative challenges across organizations (Majchrzak & Malhotra, 2016). Another research context of digital trace data research is crisis management. While Vaast et al. (2017) study microblogging use during an oil spill in the Gulf of Mexico, Cheng et al. (2011) build a recommendation framework for swine flu news on Twitter. Adjacent research contexts include contact tracing with trace data from mobile phones during the coronavirus pandemic (Oliver et al., 2020) and mouse movement-based fraud detection (Weinmann et al., 2022). Moreover, Hartl et al. (2023) study temporal changes in an onboarding process in the research context of a financial institution. The idiosyncratic features of digital trace data such as timestamps and large quantities also lend themselves to the research context of repetitive and routinized work, such as in manufacturing (Wurm et al., 2021). Lastly, digital trace data research is commonly embedded in the research context of corporate blogging (e.g., Aggarwal et al., 2012; Lu et al., 2015) and microblogging (Vaast et al., 2017).

These research contexts enable the study of a variety of *socio-technical phenomena*. On the organizational level, studies typically foreground organizational change from different perspectives (e.g., Grisold et al., 2020; Hartl et al., 2023). Among others, Wurm et al. (2021), Pentland et al. (2021), and Gaskin et al. (2014) use digital trace data to investigate the dynamics of organizational routines. Other organizational-level phenomena are organizational learning (Avgar et al., 2018), behavior change techniques (Spohrer et al., 2021) or collaboration patterns (Zhang et al., 2022), knowledge-sharing for innovative outcomes (Majchrzak & Malhotra, 2016), or the formation of connective actions (Vaast et al., 2017). A dominant theme is communication, where studies examine communicative genres in online communities (Moser et al., 2013), information diffusion on a microblogging platform (Cheng et al., 2011), or negative communication on blogs and its outcomes on a firm (Aggarwal et al., 2012). In addition, phenomena such as leadership in online communities (Faraj et al., 2015) or job performance (Lu et al., 2015) are also scrutinized with digital trace data. On the individual level, works focus on individual behavior and decision-making, for example, in fraud decisions (Weinmann et al., 2022), in a simulated business game (Kallio et al., 2022), or in the individual-level decisions of taxi drivers (Zhang et al., 2020).

4.3 Analytical Methods and Measures

The third dimension that we emphasized in our review focuses on the analytical methods and measures applied in digital trace data research around dynamics and change. By exploring this dimension, we aim to answer the question: *How are digital traces analyzed to obtain insights into dynamics and change?* In answering this question, we paid specific attention to the analytical methods that are applied as well as the measures and variables of interest (i.e., dependent and independent variables).

Our review showed that scholars study a variety of *measures*. By measures, we mean variables and constructs that are being used in order to operationalize observations regarding dynamics and change. We found that four types of measures are most prominent in research with digital trace data; temporal features, process composition, frequency, and user behavior measures. First, temporal features such as throughput time (Hartl et al., 2023) leverage temporal information (i.e., timestamps) of digital trace data to understand, explain, influence, or theorize about processual phenomena. Second, measuring and analyzing process composition, or utilizing the building blocks of the process (i.e., nodes, edges, etc.), is a common way of conducting research with digital trace data. For example, studies analyze process complexity to study the dynamics of organizational routines or infer the structure of a process by analyzing the number of ways through which a routine or a process can be performed (Pentland et al., 2021; Wurm et al., 2021). Others also used process rules to investigate drift (change over time) of processes (Yeshchenko et al., 2022). Third, studies measure the frequency of certain elements and how the frequency changes over time; examples include the frequency of forum posts (Nan & Lu, 2014), patterns (Zhang et al., 2022), process iterations (Hartl et al., 2023), or systems use (Spohrer et al., 2021). Fourth, user behavior measures such as purchases on video-on-demand platforms (Hoang & Kauffman, 2018), disclosure behavior (Rhue & Sundararajan, 2019), or fraudulent decisions (Weinmann et al., 2022) are used as measures for digital trace data research. Among others, additional measures such as job performance or level of innovation were utilized. All measures are dependent on the respective research context and phenomenon of the study.

These measures are utilized to scrutinize digital trace data with a variety of *analytical methods*. In our review, process mining emerged as a popular method for analyzing process-oriented digital trace data. Process mining has been integrated from the field of business process management (Mendling et al., 2021; van der Aalst, 2016). Scholars leverage various distinct process mining techniques to theorize about change from different perspectives (Grisold et al., 2020; Hartl et al., 2023; Pentland et al., 2021; Wurm et al., 2021). Digital trace data are also analyzed with clustering analysis (e.g., Zhang et al., 2022), content analysis, and sequence analysis techniques (e.g., Lindberg et al., 2016). Apart from these computationally intensive methods of analysis, studies also carry out analyses of digital trace data that build on classic statistical procedures such as comparative statistics and regression analyses (e.g., Majchrzak & Malhotra, 2016) or they apply econometric modeling to determine unilateral causation (e.g., Rhue & Sundararajan, 2019). Furthermore, we observed that studies often apply a mixed methods approach that combines quantitative analyses with qualitative forms of inquiry such as interviews (e.g., Moser et al., 2013) or contextual insights from organizations (e.g., Hartl et al., 2023).

4.4 General Themes and Patterns

In our analysis, we classified and analyzed 40 relevant articles according to their main purpose, data, research context, phenomenon, methods of analysis, and measures. To answer the fourth question of our analysis *What are trends, themes, and patterns in research using digital trace data for studying dynamics and change?*, we identify general themes and trends in our review. Our study maps existing literature along several dimensions and themes, which are summarized in Figure 1.

Dynamics and Change with Digital Trace Data						
Dimensions	**Themes**					
Purpose	Theory Building	Theory Testing	Method Advancement			
Data	Workflow System	Online Community	Social Media	Sensor Data	Wikipedia	
Research Context	Healthcare	Innovation	Blogging	Crisis	Finance	Manufacturing
Socio-technical Phenomena	Process & Routine Dynamics	Organizational Change	Decision -making	Collaboration	Communication	
Analytical Method	Content Analysis	Sequence & Cluster Analysis	Process Mining	Comparative Statistics	Econometric Modeling	
Measures	Temporal Features	Process Composition	Frequency	User Behavior		

Figure 1. Themes in digital trace data research on dynamics and change

With regard to the purpose of the studies in our analysis, studies engage in theory testing, instantiation, modification, or extension of theory (theory building) and studies contribute with methodological advancements. It is important to note that studies can contribute with methodological advancements while also providing a theoretical contribution (i.e., theory building) (e.g., Pentland et al., 2021). Works that promote methodological advancements and theory-building studies tend to apply a mixed methods approach by combining computational techniques with qualitative insights. Studies concerned with theory testing tend to use experiments and various statistical analyses (e.g., Weinmann et al., 2022).

Overall, the most utilized form of digital trace data (i.e., data type) are data from workflow systems (11 studies) and data from online communities (11 studies) followed by data from social media or blogs (4 studies), Wikipedia (3 studies) and sensor data from applications or devices (3 studies). Furthermore, as discussed in section 4.2, digital trace data with temporal information are used to study various phenomena in different research contexts. We found that the most prevalent research contexts are healthcare (5 studies), crisis and fraud (5 studies), manufacturing (2 studies), innovation (2 studies), blogging (2 studies), finance (1 study), and cybersecurity (1 study). In these settings, digital trace data were primarily used to study the phenomena of collaboration, organizational change, process and routine dynamics, communication, and decision-making. Moreover, we discovered that the most frequently employed analytical methods to study dynamics and change with digital trace data are process mining (6 studies), followed by sequence and cluster analyses (5 studies), econometric modeling (4 studies), comparative statistics (4 studies) and content analyses (4 studies). Furthermore, it is important to mention that studies often apply a mixed methods approach to complement digital trace data with other forms of qualitative inquiry to contextualize the findings. We also found that digital trace data can be used to analyze a broad variety of measures and variables ranging from process composition (e.g., process complexity or process rules) (e.g., Pentland et al., 2021; Yeshchenko et al., 2022), temporal features (i.e., throughput time) and frequency to user behavior measures (e.g., purchase behavior or fraud behavior) (e.g., Rhue & Sundararajan, 2019; Weinmann et al., 2022).

Taken together, we can synthesize the following general observations. First, mixed methods approaches are widely used and typically rely on both qualitative and quantitative insights. Second, digital trace data are collected and analyzed within a wide range of research contexts. Third, we find that dynamics and change, and especially the temporal aspects and effects of change over time, are investigated from different angles. Lastly, there is a stark increase in research on dynamics and change with digital trace data in recent years, which might be based on the growing availability of such data as well as increasingly sophisticated computational methods to analyze them. This trend is also visible in our review, which shows that research on dynamics and change with digital trace data experienced an upsurge from 2011 until today. All relevant studies we analyzed were published in this period of time.

5 Discussion and Implications

In this article, we reviewed the existing literature that studies dynamics and change of socio-technical phenomena with digital trace data (Miranda et al., 2022). Our analysis revealed several aspects and themes that are emerging in this stream of research. In what follows, we discuss the implications of our work and provide avenues for future research.

5.1 Novel Insights Enabled through Digital Trace Data

A number of conceptual arguments stress that digital trace data research provides means to generate novel insights about socio-technical phenomena (e.g., Berente et al., 2019; Miranda et al., 2022; Oliver et al., 2020; vom Brocke et al., 2021). So far, however, we lack systematic knowledge about how the anticipated novelty is represented in actual empirical findings. Based on our analysis, we make several observations.

First, digital trace data allow researchers to investigate change and temporal dynamics on fine-granular levels. To this end, we found that studies draw a nuanced picture of how and when change takes shape over time. Research on organizational routines is a case in point (Feldman et al., 2016). While research in this field has traditionally used manually collected qualitative data, such as interviews and ethnographies (e.g., Dittrich et al., 2016), more recent studies build on digital trace data (Lindberg et al., 2016; Pentland et al., 2021; Wurm et al., 2021). Thereby, these studies locate the dynamics of organizational routines at the level of specific activities that might not be observable by means of traditional research approaches. Certain dynamics of organizational routines, for example, can be detected while those who perform the routines might even be unaware of them (Pentland et al., 2021).

Second, digital trace data offer means of measuring and operationalizing temporal dynamics. This is because they are collected in large varieties, and event logs depict recurrent activities that are consistently labeled in the whole data set. In GitHub, for example, it can be systematically examined how the distribution of a fixed set of activities changes over time (Lindberg et al., 2016). Building on this observation, it has been argued that digital trace data enable more systematic knowledge about theories that were traditionally studied using qualitative approaches, such as interviews. For example, Grisold, Gau et al. (2022) use digital trace data to measure structuration processes in collaboration environments, thus developing in-depth explanations about when, how, and why individual actors influence social structures, and vice versa (Essén & Värlander, 2019). This general observation is supported by the fact that well-known phenomena are increasingly studied with new constructs –such as the dynamics of organizational routines by various perspectives on process composition (Pentland et al., 2021)– and that new constructs are developed to study new kinds of phenomena altogether (Grover & Lyytinen, 2022).

In light of these observations, we encourage future research to identify phenomena that have been studied from abstract angles (such as in structuration theory) and could be studied on a more fine-granular level with digital trace data. Furthermore, new constructs can be developed that provide new perspectives and angles to study socio-technical phenomena.

5.2 Embracing Opportunities for Studying Dynamics and Change

We found that studies leverage a wide range of data sources, analytical methods, and phenomena (see Figure 1). This points to the numerous opportunities that can be embraced when studying dynamics and change using digital trace data. When looking at GitHub, for example, we see that one and the same data source can be leveraged to analyze various socio-technical phenomena, such as organizational routines (Lindberg et al., 2016), or structuration processes (Grisold, Gau et al., 2022).

At the same time, however, there appear to be path-dependent trends and patterns. Studies that focus on a specific phenomenon, for example, tend to use certain data sources and computational techniques. Twitter studies tend to analyze the content of tweets (Vaast et al., 2017); instead of, for example, social relations between users or other use patterns (such as likes or retweets) (Zöller et al., 2020). Similarly, studies using digital trace data from GitHub typically analyze the distribution of activity types (Grisold, Gau et al., 2022; Lindberg et al., 2016); instead of, for example, the evolving code base itself or comments made within specific activity types (e.g., Issue Comment Events). In a similar vein, research on organizational routines (e.g., Pentland et al., 2020; Pentland et al., 2021; Wurm et al., 2021) tends to utilize event logs from workflow systems, such as enterprise resource planning systems. These studies focus on time stamps, activities, and, if available, other resource-related information in the data. Furthermore, across these studies, we find a tendency to use process mining as a computational technique to analyze the data (e.g., Pentland et al., 2020; Pentland et al., 2021; Wurm et al., 2021).

We encourage future research to continue embracing the full spectrum of opportunities that emerge in digital trace data research to focus on dynamics and change. This includes exploring all kinds of data sources along with a variety of available features in data sets. To this end, it seems promising to combine different information in event logs, such as geospatial information, or integrate insights from qualitative data to provide more context to analyses (Oliver et al., 2020; Pentland et al., 2020; Whelan et al., 2016; Zhang et al., 2020). Furthermore, one can leverage methods from other research fields (vom Brocke et al., 2021), such as biology and medicine, and use digital traces for descriptive, explanatory, and predictive research designs (e.g., Karahanna et al., 2018; Oliver et al., 2020). Finally, since digital trace data can be accrued in all kinds of contexts, future research may explore how digital trace data can be used in other research contexts.

5.3 Advancing Research by Promoting Cross-Disciplinary Knowledge Exchange

We found indications that digital trace data research promotes cross-disciplinary dialogues between different research communities (vom Brocke et al., 2021). While the information systems field has a strong tradition in socio-technical research (Sarker et al., 2019), it has been less concerned with explicitly technical matters, such as knowledge that centers around computational methods to analyze digital trace data. This, however, is needed to analyze digital trace data. For example, a recent call for 'process science' (vom Brocke et al., 2021) encourages researchers from information systems, computer science, and organization science to jointly analyze, explain and promote change based on digital trace data sets.

Along these lines, we see that recent moves have been incorporating new computational methods from other fields. For example, we found that process mining (van der Aalst, 2016) is increasingly used as a method to visualize and analyze temporal dynamics of organizational routines based on digital trace data (Pentland et al., 2021). This method has been adopted from the field of business process management. This field of research is concerned with developing methodological knowledge to visualize and analyze the performance of business processes (Mendling et al., 2021). Adapted to dynamics and change in the information systems field, it is used to reveal the temporal dynamics of organizational routines, which in turn, informs the development of theoretical explanations (Berente et al., 2019; Miranda et al., 2022).

In light of this example, we expect that researchers will find additional means to study digital trace data when they intensify efforts to connect the information systems field with other scientific communities. Computational methods that are used in public policy-making, for example, serve various purposes, such as creating awareness of a societal problem or evaluating cause-and-effect relationships after an intervention was implemented (Oliver et al., 2020). Such methods can support innovative theorizing about novel socio-technical phenomena (Grover & Lyytinen, 2022) or evaluating design science research projects (vom Brocke et al., 2020). In general, the information systems field might benefit from initiatives that make such methods accessible to a wider audience (Grisold, Kremser et al., 2023).

5.4 Limitations and Outlook

Our literature review comes with certain limitations. In our literature search, we focused on research that centers around 'trace data'. Our motivation was that this term is associated with a new research genre in the information systems field that advances empirical knowledge as well as methodological knowledge (Berente et al., 2019; Lindberg, 2020; Miranda et al., 2022). To this end, our analysis surfaced several trends and themes in this emerging stream of research. At the same time, other keywords (e.g., 'big data' or 'computationally intensive theorizing') may have led to additional complementary insights. While we counteracted any potential biases through forward and backward searches, we will expand our systematic literature review with more keywords in the future. Furthermore, we observe that digital trace data-based research is gaining prominence in other fields, such as the social sciences in general terms (Lazer et al., 2020), and the management field more specifically (Simsek et al., 2019). There were even calls to establish cross-disciplinary research agendas (Oliver et al., 2020; vom Brocke et al., 2021). While other research fields (such as biology) may reveal themes and topics that are relevant for information systems research, we did not systematically integrate them in our analysis. Moreover, reviewing the development of digital trace data research in the information systems field in general, rather than exclusively focusing on dynamics and change, seems promising. We are planning to expand our study in future works by including different fields and adopting a broader focus. Lastly, we see an opportunity in expanding our study with quantitative analyses (e.g., scientometrics). Such approaches could corroborate our findings and allow for tracing and unpacking how this evolving field has been unfolding over time. For example, it would be interesting to investigate how different authors, studies, and/or ideas have been forming, and shaping digital trace data research. We plan to integrate quantitative analyses to enhance the understanding of research on dynamics and change with digital trace data in future work.

6 Conclusion

There is a surging interest in the information systems field around digital trace data research and computationally intensive theorizing. Our study is the first to systematically analyze the state of this field with regard to research that investigates dynamics and change of socio-technical phenomena over time. Drawing on a total number of 40 articles published in high-quality information systems outlets, we observe that this stream of research pursues different interests and themes. To this end, we found that studies center around different (1) research purposes, (2) data types, (3) contexts, (4) phenomena, (5) methods, and (6) measures. Our findings allow us to map current trends and patterns, and we provide a number of directions for future research.

Acknowledgements

This research was carried out, among others, within the framework of the Schöller Senior Fellowship on “Process Science – The Interdisciplinary Study of Continuous Change”. The research has also been funded by the ERASMUS + program of the European Union (EU Funding 2021-1-LI01-KA220-HED-000027575 “Developing Process Mining Capabilities at the Enterprise Level”) and the Liechtenstein Research Fund (grant number 705076 “Process Science: Conceptual foundation for the interdisciplinary study of continuous change”).

References[1]

*Aggarwal, R., Gopal, R., Sankaranarayanan, R., & Singh, P. V. (2012). Blog, Blogger, and the Firm: Can Negative Employee Posts Lead to Positive Outcomes? *Information Systems Research*, *23*(2), 306–322.

AIS. (2011). *Senior Scholars' Basket of Journals*. https://aisnet.org/page/SeniorScholarBasket

*Arazy, O., Daxenberger, J., Lifshitz-Assaf, H., Nov, O., & Gurevych, I. (2016). Turbulent Stability of Emergent Roles: The Dualistic Nature of Self-Organizing Knowledge Coproduction. *Information Systems Research*, *27*(4), 792–812.

*Avgar, A., Tambe, P., & Hitt, L. M. (2018). Built to Learn: How Work Practices Affect Employee Learning During Healthcare Information Technology Implementation. *MIS Quarterly*, *42*(2), 645–659.

Baiyere, A., Berente, N., & Avital, M. (2023). On digital theorizing, clickbait research, and the cumulative tradition. *Journal of Information Technology*, *38*(1), 67–73.

Benbya, H., Nan, N., Tanriverdi, H., & Yoo, Y. (2020). Complexity and Information Systems Research in the Emerging Digital World. *MIS Quarterly*, *44*(1), 1–17.

*Benjamin, V., Zhang, B., Nunamaker, J. F., & Chen, H. (2016). Examining Hacker Participation Length in Cybercriminal Internet-Relay-Chat Communities. *Journal of Management Information Systems*, *33*(2), 482–510.

*Berente, N., Seidel, S., & Safadi, H. (2019). Research Commentary—Data-Driven Computationally Intensive Theory Development. *Information Systems Research*, *30*(1), 50–64.

Burton-Jones, A., McLean, E. R., & Monod, E. (2015). Theoretical Perspectives in IS Research: From Variance and Process to Conceptual Latitude and Conceptual Fit. *European Journal of Information Systems*, *24*(6), 664–679.

[1] Articles included in the systematic literature review are indicated by an asterisk (*)

*Chen, L., Baird, A., & Straub, D. (2019). Fostering Participant Health Knowledge and Attitudes: An Econometric Study of a Chronic Disease-Focused Online Health Community. *Journal of Management Information Systems*, *36*(1), 194–229.

*Cheng, J., Sun, A., Hu, D., & Zeng, D. (2011). An Information Diffusion-Based Recommendation Framework for Micro-Blogging. *Journal of the Association for Information Systems*, *12*(7), 463–486.

Cloutier, C., & Langley, A. (2020). What Makes a Process Theoretical Contribution? *Organization Theory*, *1*(1).

Cooper, H. M. (1988). Organizing knowledge syntheses: A taxonomy of literature reviews. *Knowledge in Society*, *1*(1), 104–126.

Dittrich, K., Guérard, S., & Seidl, D. (2016). Talking About Routines: The Role of Reflective Talk in Routine Change. *Organization Science*, *27*(3), 678–697.

Essén, A., & Värlander, S. W. (2019). How Technology-Afforded Practices at the Micro-Level can Generate Change at the Field Level: Theorizing the Recursive Mechanism Actualized in Swedish Rheumatology 2000-2014. *MIS Quarterly*, *43*(4), 1155–1176.

*Faraj, S., Kudaravalli, S., & Wasko, M. (2015). Leading Collaboration in Online Communities. *MIS Quarterly*, *39*(2), 393–412.

Feldman, M. S., Pentland, B. T., D'Adderio, L., & Lazaric, N. (2016). Beyond Routines as Things: Introduction to the Special Issue on Routine Dynamics. *Organization Science*, *27*(3), 505–513.

Freelon, D. (2014). On the Interpretation of Digital Trace Data in Communication and Social Computing Research. *Journal of Broadcasting & Electronic Media*, *58*(1), 59–75.

*Gaskin, J., Berente, N., Lyytinen, K., & Yoo, Y. (2014). Toward Generalizable Sociomaterial Inquiry: A Computational Approach for Zooming In and Out of Sociomaterial Routines. *MIS Quarterly*, *38*(3), 849–871.

Golder, S. A., & Macy, M. W. (2014). Digital Footprints: Opportunities and Challenges for Online Social Research. *Annual Review of Sociology*, *40*(1), 129–152.

Grisold, T., Gau, M., & Yoo, Y. (2022). Studying the Co-evolution of Individual Actions and Emergent Social Structures using Digital Trace Data. *Proceedings of the International Conference on Information Systems (ICIS).*

Grisold, T., Kremser, W., Mendling, J., Recker, J., vom Brocke, J., & Wurm, B. (2023). Keeping Pace with the Digital Age: Envisioning Information Systems Research as a Platform. *Journal of Information Technology*, *38*(1), 60–66.

*Grisold, T., Wurm, B., Mendling, J., & vom Brocke, J. (2020). Using Process Mining to Support Theorizing About Change in Organizations. *Proceedings of the Hawaii International Conference on System Sciences (HICSS).*

*Grover, V., Lindberg, A., Benbasat, I., & Lyytinen, K. (2020). The Perils and Promises of Big Data Research in Information Systems. *Journal of the Association for Information Systems*, *21*(2), 268–293.

Grover, V., & Lyytinen, K. (2015). New State of Play in Information Systems Research: The Push to the Edges. *MIS Quarterly*, *39*(2), 271–296.

Grover, V., & Lyytinen, K. (2022). The Pursuit of Innovative Theory in the Digital Age. *Journal of Information Technology*, *43*(4).

*Hartl, S., Franzoi, S., Grisold, T., & vom Brocke, J. (2023). Explaining Change with Digital Trace Data: A Framework for Temporal Bracketing. *Proceedings of the Hawaii International Conference on System Sciences (HICSS).*

*Hoang, A.-P., & Kauffman, R. J. (2018). Content Sampling, Household Informedness, and the Consumption of Digital Information Goods. *Journal of Management Information Systems*, *35*(2), 575–609.

*Howison, J., Wiggins, A., & Crowston, K. (2011). Validity Issues in the Use of Social Network Analysis with Digital Trace Data. *Journal of the Association for Information Systems*, *12*(12), 767–797.

Hukal, P., Berente, N., Germonprez, M., & Schecter, A. (2019). Bots Coordinating Work in Open Source Software Projects. *Computer*, *52*(9), 52–60.

*Kallio, H., Malo, P., Lainema, T., Bragge, J., Seppälä, T., & Penttinen, E. (2022). Generating Research Questions from Digital Trace Data: A Machine-Learning Method for Discovering Patterns in a Dynamic Environment. *Communications of the Association for Information Systems*, *51*(1).

*Karahanna, E., Benbasat, I., Bapna, R., & Rai, A. (2018). Editor's Comments: Opportunities and Challenges for Different Types of Online Experiments. *MIS Quarterly*, *42*(4), iii–x.

*Kim, Y., Jarvenpaa, S. L., & Gu, B. (2018). External Bridging and Internal Bonding: Unlocking the Generative Resources of Member Time and Attention Spent in Online Communities. *MIS Quarterly*, *42*(1), 265–283.

*Kyriakou, H., Nickerson, J. V., & Majchrzak, A. (2022). Novelty and the Structure of Design Landscapes: A Relational View of Online Innovation Communities. *MIS Quarterly*, *46*(3), 1691–1720.

Langley, A. (1999). Strategies for Theorizing from Process Data. *The Academy of Management Review*, *24*(4), 691-710.

Langley, A., Smallman, C., Tsoukas, H., & van de Ven, A. H. (2013). Process Studies of Change in Organization and Management: Unveiling Temporality, Activity, and Flow. *Academy of Management Journal*, *56*(1), 1–13.

Langley, A., & Tsoukas, H. (Eds.). (2017). *The SAGE Handbook of Process Organization Studies*. SAGE Publications.

Lazer, D. M. J., Pentland, A., Watts, D. J., Aral, S., Athey, S., Contractor, N., Freelon, D., Gonzalez-Bailon, S., King, G., Margetts, H., Nelson, A., Salganik, M. J., Strohmaier, M., Vespignani, A., & Wagner, C. (2020). Computational Social Science: Obstacles and Opportunities. *Science*, *369*(6507), 1060–1062.

*Lindberg, A. (2020). Developing Theory Through Integrating Human and Machine Pattern Recognition. *Journal of the Association for Information Systems*, *21*(1), 90–116.

*Lindberg, A., Berente, N., Gaskin, J., & Lyytinen, K. (2016). Coordinating Interdependencies in Online Communities: A Study of an Open Source Software Project. *Information Systems Research*, *27*(4), 751–772.

*Lu, B., Guo, X., Luo, N., & Chen, G. (2015). Corporate Blogging and Job Performance: Effects of Work-related and Nonwork-related Participation. *Journal of Management Information Systems*, *32*(4), 285–314.

*Majchrzak, A., & Malhotra, A. (2016). Effect of Knowledge-Sharing Trajectories on Innovative Outcomes in Temporary Online Crowds. *Information Systems Research*, *27*(4), 685–703.

Mendling, J., Berente, N., Seidel, S., & Grisold, T. (2021). The Philosopher's Corner: Pluralism and Pragmatism in the Information Systems Field: The Case of Research on Business Processes and Organizational Routine. *ACM SIGMIS Database: The DATABASE for Advances in Information Systems*, *52*(2), 127–140.

Mendling, J., Pentland, B. T., & Recker, J. (2020). Building a Complementary Agenda for Business Process Management and Digital Innovation. *European Journal of Information Systems*, *29*(3), 208–219.

*Miranda, S., Berente, N., Seidel, S., Safadi, H., & Burton-Jones, A. (2022). Editor's Comments: Computationally Intensive Theory Construction: A Primer for Authors and Reviewers. *MIS Quarterly*, *46*(2), iii–xviii.

*Moser, C., Ganley, D., & Groenewegen, P. (2013). Communicative genres as organising structures in online communities - of team players and storytellers. *Information Systems Journal*, *23*(6), 551–567.

Mousavi Baygi, R., Introna, L. D., & Hultin, L. (2021). Everything Flows: Studying Continuous Socio-Technological Transformation in a Fluid and Dynamic Digital World. *MIS Quarterly*, *45*(1), 423–452.

Müller, O., Junglas, I., Debortoli, S., & vom Brocke, J. (2016). Using Text Analytics to Derive Customer Service Management Benefits from Unstructured Data. *MIS Quarterly Executive*, *15*(4), 243–258.

*Nan, N., & Lu, Y. (2014). Harnessing the Power of Self-Organization in an Online Community During Organizational Crisis. *MIS Quarterly*, *38*(4), 1135–1157.

*Oliver, N., Lepri, B., Sterly, H., Lambiotte, R., Deletaille, S., Nadai, M. de, Letouzé, E., Salah, A. A., Benjamins, R., Cattuto, C., Colizza, V., Cordes, N. de, Fraiberger, S. P., Koebe, T., Lehmann, S., Murillo, J., Pentland, A., Pham, P. N., Pivetta, F., . . . Vinck, P. (2020). Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. *Science Advances*, *6*(23),

*Østerlund, C., Crowston, K., & Jackson, C. (2020). Building an Apparatus: Refractive, Reflective, and Diffractive Readings of Trace Data. *Journal of the Association for Information Systems*, *21*(1), 1–22.

Pentland, A. (2015). *Social physics: How social networks can make us smarter*. Penguin Books.

*Pentland, B., Haerem, T., & Khaledi, H. (2014). Using Action Networks to Detect Change in Repetitive Patterns of Action. *Proceedings of the International Conference on Information Systems (ICIS).*

*Pentland, B., Vaast, E., & Wolf, J. R. (2021). Theorizing Process Dynamics with Directed Graphs: A Diachronic Analysis of Digital Trace Data. *MIS Quarterly*, *45*(2), 967–984.

*Pentland, B. T., Recker, J., Wolf, J., & Wyner, G. (2020). Bringing Context Inside Process Research with Digital Trace Data. *Journal of the Association for Information Systems*, *21*(5), 1214–1236.

Poole, M. S., & van de Ven, A. H. (Eds.). (2004). *Handbook of Organizational Change and Innovation*. Oxford University Press.

*Rhue, L., & Sundararajan, A. (2019). Playing to the Crowd: Digital Visibility and the Social Dynamics of Purchase Disclosure. *MIS Quarterly*, *43*(4), 1127–1141.

Sarker, S., Chatterjee, S., Xiao, X., & Elbanna, A. (2019). The Sociotechnical Axis of Cohesion for the IS Discipline: Its Historical Legacy and its Continued Relevance. *MIS Quarterly*, *43*(3), 695–719.

Simsek, Z., Vaara, E., Paruchuri, S., Nadkarni, S., & Shaw, J. D. (2019). New Ways of Seeing Big Data. *Academy of Management Journal*, *62*(4), 971–978.

*Spohrer, K., Fallon, M., Hoehle, H., & Heinzl, A. (2021). Designing Effective Mobile Health Apps: Does Combining Behavior Change Techniques Really Create Synergies? *Journal of Management Information Systems*, *38*(2), 517–545.

*Vaast, E., & Pinsonneault, A. (2021). When Digital Technologies Enable and Threaten Occupational Identity: The Delicate Balancing Act of Data Scientists. *MIS Quarterly*, *45*(3), 1087–1112.

*Vaast, E., Safadi, H., Lapointe, L., & Negoita, B. (2017). Social Media Affordances for Connective Action: An Examination of Microblogging Use During the Gulf of Mexico Oil Spill. *MIS Quarterly*, *41*(4), 1179–1205.

van der Aalst, W. (2016). *Process Mining*. Springer Berlin Heidelberg.

Vial, G. (2019). Reflections on quality requirements for digital trace data in IS research. *Decision Support Systems*, *126*(1), 113133.

vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., & Cleven, A. (2009). Reconstructing the Giant. On the Importance of Rigour in Documenting the Literature Search Process. *Proceedings of the European Conference on Information Systems (ECIS).*

vom Brocke, J., van der Aalst, W. M. P., Grisold, T., Kremser, W., Mendling, J., Pentland, B., Recker, J., Roeglinger, M., Rosemann, M., & Weber, B. (2021). Process Science: The Interdisciplinary Study of Continuous Change. *SSRN Electronic Journal.*

vom Brocke, J., Winter, R., Hevner, A., & Maedche, A. (2020). Special Issue Editorial –Accumulation and Evolution of Design Knowledge in Design Science Research: A Journey Through Time and Space. *Journal of the Association for Information Systems*, *21*(3), 520–544.

Webster, J., & Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*, *26*(2), xiii–xxiii.

*Weinmann, M., Valacich, J., Schneider, C., Jenkins, J., & Hibbeln, M. (2022). The Path of the Righteous: Using Trace Data to Understand Fraud Decisions in Real Time. *MIS Quarterly*, *46*(4), 2317-2336.

Whelan, E., Teigland, R., Vaast, E., & Butler, B. (2016). Expanding the horizons of digital social networks: Mixing big trace datasets with qualitative approaches. *Information and Organization*, *26*(1-2), 1–12.

*Wurm, B., Grisold, T., Mendling, J., & vom Brocke, J. (2021). Measuring Fluctuations of Complexity in Organizational Routines. *Academy of Management Proceedings*, *2021*(1).

*Yeshchenko, A., Di Ciccio, C., Mendling, J., & Polyvyanyy, A. (2022). Visual Drift Detection for Event Sequence Data of Business Processes. *IEEE Transactions on Visualization and Computer Graphics*, *28*(8), 3050–3068.

*Young, A. G., Wigdor, A. D., & Kane, G. C. (2020). The Gender Bias Tug-of-War in a Co-creation Community: Core-Periphery Tension on Wikipedia. *Journal of Management Information Systems*, *37*(4), 1047–1072.

*Zhang, Y., Li, B., & Krishnan, R. (2020). Learning Individual Behavior Using Sensor Data: The Case of Global Positioning System Traces and Taxi Drivers. *Information Systems Research*, *31*(4), 1301–1321.

*Zhang, Z., Lee, H., Yoo, Y., & Choi, Y. T. (2022). Theorizing Routines with Computational Sequence Analysis: A Critical Realism Framework. *Journal of the Association for Information Systems*, *22*(2), 589–630.

Zöller, N., Morgan, J. H., & Schröder, T. (2020). A topology of groups: What GitHub can tell us about online collaboration. *Technological Forecasting and Social Change*, *161*(2).