DIGITAL TRACES OF INFORMATION SYSTEMS: SOCIOMATERIALITY MADE RESEARCHABLE

Completed Research Paper

Jonas Hedman
Department of IT Management,
Copenhagen Business School,
Howitzvej 60, 2000 Frederiksberg.
Denmark
jh.itm@cbs.dk

Nikhil Srinivasan
Department of MIS,
Terry College of Business,
310 Herty Drive, Athens GA.
U.S.A
nsrini@uga.edu

Rikard Lindgren
Department of Applied Information Technology,
University of Gothenburg,
Forskningsgången 6, Patricia building, Lindholmen.
Sweden
rikard.lindgren@ait.gu.se

School of Business and IT,
University of Borås,
501 90 Borås.
Sweden

Abstract

In this paper, we point to the potential and implications of digital traces as novel data source in the study of contemporary activities and behaviors. We do this to raise awareness of IS researchers of such traces in increasingly complex sociomaterial practices. We develop a two-dimensional framework of data sources (subjective/objective and digitalized/non-digitalized) for analyzing a six-year literature survey comprised of five leading IS journals. The analysis positions current data sources employed within the framework, and sheds light on the under utilization of digitalized data sources. This disconcerting result suggests that IS researchers must pay more attention to the changing landscape of data sources. To motivate and guide fellow colleagues to establish the credibility and reliability of digital traces, we develop a future research agenda that covers both opportunities in theory generation and challenges in data collection.

Keywords: Digital traces, sociomateriality, research method, data sources

Introduction

The increasing pervasiveness of emerging technologies reduces the gap between the digital and physical worlds (Lyytinen et al. 2002). While tech visions like the "Internet of Things" have begun to materialize, users are now able to store, mobilize, and interpret information sources that were unavailable in the past. Handheld digitalized devices that interact with sensors tracking their movements enable sharing of location-based information among friends. Computing capabilities embedded into doors, walls, windows, and floors combined with computers in pens, watches, phones, desks, and cars leverage monitoring processes for sustainable buildings. Trucks and goods equipped with a GPS chip, two-dimensional barcodes, and RFID help logistic partners to reshape and optimize their integrated supply chains by recognizing alterations in inventory levels, market demands, and transport constraints.

Such impending developments in IT may be about to change contemporary practices entirely (Orlikowski et al. 2008). In particular, these practices tend to become increasingly complex and multifaceted due to the abundance of technologies involved. That is, understanding and changing individual technological components will be more difficult because they are integral parts of assemblages. A consequence of being embedded in this way is an increased invisibility, which ultimately makes them disappear from the immediate attention of users. Yoo (2010) argues that an increasingly technologically and economically feasible product digitalization will inevitably make such challenges even more prominent over the years to come.

We can only expect that the sociomaterial practices these digital environments render are likely to significantly change the behaviors of individuals, groups, and organizations (Lindgren et al. 2008; Yoo 2010). However, in what ways is difficult to say because the performance of such an environment and its assessment are not merely discursively constructed, but also materially produced through enactment in practice. Given that the social and material dimensions are inherently inseparable, enactment of technology in practice involves mutually intertwined discursive and material production (Orlikowski 2007; Orlikowski et al. 2008). Such sociomateriality complicates the collection of rich data and meaningful insights (Orlikowski et al. 2008), thereby instigating a quest for adapted or even new data generation and analysis techniques (Venturini et al. 2010; Watts 2007).

Luckily enough, the "smart artifacts" of today's digital environments remember where they were, who used them, and the outcomes of the interactions (Yoo 2010). Artifacts produce massive amounts of data of their conditions, movements, and interactions with others, which makes them traceable in time and space. Such "digital traceability" (Venturini et al. 2010) presents a window of opportunity by recording and archiving digital traces of sociomaterial activities and interactions over time for real time or subsequent review or viewing by those not present (Grudin 2002; Jessup et al. 2002).

The point of departure for this paper is the observation that digital traces represent novel sources of empirical data about sociomateriality and its consequences for social and material properties alike. As such, they invite IS researchers to explore and exploit data material of different origins. Taking advantage of these traces indeed resonates well with calls for IS research that places the IT artifact at center stage (Benbasat et al. 2003; Orlikowski et al. 2001). While we have seen an emerging artifact orientation in the discipline, still scant attention has been paid to the role of IT in the research process itself and in particular as a source of evidence. Historically, there have been only a few notable exceptions that draw on digital traces to come closer to the intertwined nature of sociomaterial practices (Lindgren et al. 2004; Lindgren et al. 2003; Stenmark 2001).

Contemporary IS research into the use of information systems and information technologies, however, is relying increasingly on digitalized environments. One recent example is Chesney et al's (2009) study of a virtual world to explore negative behavior or 'griefing'. The use of data from these environments represents a data source that may supplement data from interviews or surveys, which are the most common data sources in IS research. This digital data promises to enable new forms of theorizing and research within the IS discipline (Yoo 2010). Indeed, researchers in HCI have for long used such traces to investigate the design, implementation, and uptake of different technologies.

The purpose of this paper is to highlight the lack of attention paid to digital traces in the IS research process, and thereby help broaden the scope of acceptable and credible empirical data sources. Given our

intent, we develop a two-dimensional framework of data sources (subjective/objective and digitalized/non-digitalized) to analyze a six-year literature survey of five premier IS journals. Our analysis shows that additional knowledge development is needed to promote systematic usage of digital traces. In response, we propose a research agenda for digital traces that we hope will have a positive impact on the IS field by motivating and guiding future work.

Sociomaterial Practice

We have historically thought of humans and technology as mutually shaping entities, i.e., each is changed by its interaction with the other, but still maintaining their ontological separation. However, it has been suggested that contemporary realities will require us to stop treating materiality as invisible, or expressed differently, separating technology from human affairs. The main point is that the distinction between humans and technologies is merely analytical, and as Nyberg (2009, p 1181) puts it "the social and the material have been inseparable ever since Homo sapiens first used a stone to crack a nut".

Challenging a priori distinctions between humans and non-human actors, the 'practice' literature views the latter as an outcome of practices being granted some form of agency to interact with the former (e.g., Orlikowski 2000). Given that they are produced materially and socially in arrangement as the outcome of the practices, the boundaries that define them are fluid and temporal (Nyberg 2009). Meaning is therefore derived from the contextual ongoing performance of the actors. This suggests there are simply no finished products. Practice theorists in organization studies like Callon (1986) and Latour (1987) have pursued the idea of giving non-human actors agency by drawing upon actor network theory. For them, humans and technologies are not only reciprocally interdependent, but also symmetrically relevant, i.e., equivalent participants in a temporal network aligned to achieving certain effects. Technological artifacts therefore take an active part in creating, and simultaneously are, the products of practices by enabling, facilitating, mediating, mitigating, constraining, and resisting the creation of the sociomaterial world (Suchman 2000; Suchman 2005).

Arguing that "there is no social that is not also material, and no material that is not also social", Orlikowski (2007, p 1437) views the social and the material as inextricably related and coemerging in situationally specific ways. This view asserts that materiality is intrinsic to organizing as well as everyday life (Nyberg 2009), recognizing that the social and the material are constitutively entangled (Suchman 2006). Most things we do entail material means (e.g., clothes, devices, telephones, tools, etc.), which are produced through human practices. Such 'sociomaterial' perspective may help us reconfigure our taken-for-granted assumptions, notions, and practices, and instead direct our attention to the emergent, multiple, and shifting assemblages that increasingly constitute today's society (Orlikowski 2007).

Orlikowski and Scott (2008) assert, however, that it is easier said than done to study, for example, large-scale conversations using Web-based discussion forums, collaborative dynamics within e-Science Grids, and habitats of connectivity formed through usage of wearable mobile technologies. While existing approaches, lenses, and tools are still appropriate for some situations, there will be many more that necessitate a fresh palette of methodological and theoretical resources. For example, finding "ways of establishing a corpus of data under fieldwork conditions that are distributed, constantly reconfiguring, fragmented into enclaves, or restricted by partial access" is difficult (Orlikowski et al. 2008, p 464). In a sociomaterial world, they argue, research participants tend to disappear from view, ultimately disrupting our routine methods of meaning making. Like Kleinberg (2008), Venturi and Latour (2010), and Watts (2007) among others, it is here we see digital traces and traceability as a gift we cannot do without.

Digital Traces

Given that most conversations and transactions in digital environments leave behind a tangible trace (Agarwal et al. 2008; Venturini et al. 2010), the growth of online interaction has led to increased availability of explicitly digital trace data (Howison et al. 2011). Such data are produced through and stored by an information system (i.e., not ephemeral, see e.g., Grudin 2002), and are typically a by-product of activities rather than produced by a designed research instrument. As records of activity (i.e., evidence that something has occurred in the past), they are event-based data, rather than summary-based data. Because the events occur over time, digital trace data are longitudinal data (Howison et al. 2011).

Digital trace data make visible social processes that are usually far more difficult to study in conventional settings (Agarwal et al. 2008). While Kleinberg (2008, p. 66) argues that "science advances whenever we can take something that was once invisible and make it visible", such data allow us to study social interaction processes of millions of people at a resolution that is sensitive to individual level effects (Watts 2007). Made visible through their online manifestation, these real interactions are part of the ephemeral dynamics of everyday life. That is, how they evolve as a consequence of the sociomaterial environments in which they are embedded.

It has been argued that these newfound digital trace data combined with existing knowledge and appropriate analysis techniques may form the foundation upon which a "21st Century science" can be built (Watts 2007). Such science can lead to a revolution in the measurement of collective human behavior by allowing for theory building based on massive data sets of large social systems (Kleinberg 2008). While IS researchers can "unobtrusively" collect digital trace data on usage, flows, and content from a full census of users, however, the tricky question here concerns the actual content of such traces generated through IT use. These traces represent a potentially wide range of activities and behaviors of individuals and systems. Distinguishing the types of digital traces based on the properties that different IT artifacts possess is therefore key to understand how various forms of traces may influence research strategies (i.e., the properties are key enablers within the digitalized artifact that produce the digital trace data). We draw upon Yoo's (2010) seven material properties of digital artifacts to develop a list of properties. The resulting list is by no means exhaustive, but still it indicates what digital traces we can expect to get from different technologies. We believe that these traces can help establish a general platform for better understanding associated research opportunities and challenges.

Programmability is a property of a digitalized artifact to accept new sets of logic, thus allowing it to modify its behaviors and functions. Examples of such artifacts are embedded software systems (e.g., CAN bus) and artifact control systems (e.g., iTunes). Digital data traces from these include behavioral characteristics of the individual embracing the technology, modifications to device/system functionality, patterns of usage of people interacting with the application, and many more.

Addressability is a property of a digitalized artifact that enables it to respond to a message sent to many similar artifacts. That is, the digitalized artifact can be uniquely identified, thereby allowing it to interact with other components enrolled into an information infrastructure. This property thus concerns both the identification and communication with other available digitalized devices being temporal parts of a network structure. In this way, messages (e.g., SMS messages) sent through standardized protocols, including GSM, IP, and MAC, entails data about action that should be taken by an artifact. Here digital data traces refer to, for example, identification of points of origination and termination of a communication stream between devices.

Enabled by sensors combined with embedded software, *sensibility* concerns the ability of a digitalized artifact to monitor and respond to changes in the context of the surrounding environment. To illustrate, a thermometer sensor paired with a GPS chip can help monitor temperature exposure of temperature-sensitive food products during transportation and storage. Digital data traces emanating from such supply chain process include, for example, information from bar code scanning, communication between sensors and terminal systems, interaction with payment systems, and route planning/navigation decisions.

Given that a digitalized artifact in a network can be uniquely identified, it is able to interact with other artifacts, actors, and infrastructures through embedded communication capabilities. *Communicability* thus concerns the interaction and sharing between social and material entities, which will enable new forms of dependencies and relationships. The digital data traces that come from such sociomaterial processes may reveal communication patterns emerging in digitalized environments. For example, today individuals are increasingly connecting various types of sensors to their Twitter accounts to inform colleagues and friends about what they are doing.

Either on their own or through a network, most of today's digitalized artifacts have some type of memory capacity. This makes them *memorizeable*. That is, an individual artifact is capable of remembering where it was, who used it, and the outcomes of the interactions, etc. The resulting digital data traces can tell us a lot about what information it generated, sensed, and communicated. To illustrate, for learning purposes an echo driving application enabled by embedded systems combined with navigation software is not only able to reveal exactly how an individual driver behaved during a working day, but also suggest what

behaviors and actions should be changed.

Senseable and memorizeable digitalized artifacts are traceable in time and space. The *traceability* property refers to the ability of a digitalized artifact to chronologically interrelate events and entities over time. While this property requires a unique identifier of these events and entities (such as a time and location stamp) along with the actual substantive information of the events and entities stored in the memory, it promises to render digital data traces that may shed light on the nature of conditions, movements, and interactions in sociomaterial practice. With such traces like the time line in Facebook we therefore can examine evolutionary aspects of technology usage more easily.

Associability is a property of digitalized artifacts that enables information related to actors, artifacts, events, and places to become associable. While this property requires commonly shared attributes, typically materialized in the forms of interfaces and standards, to facilitate the smooth interaction among distributed actors and artifacts, the types of digital data traces it generates include tags, keywords, and affiliation patterns that social and material entities leave behind. These traces represent points or events of interaction, thus serving as the platform through which they collaborate.

Apparently, different types of IT systems generate different digital trace data that spur new research opportunities (Yoo 2010). In particular, such rich and plentiful data on sociomaterial practice can close an important feedback loop and allow us to develop/evaluate complex models of social phenomena that may inform the design of new digital solutions (Kleinberg 2008). However, IT systems acting as data collection tools also generate limitations and challenges (Howison et al. 2011). For example, once a digital trace reaches a network, it may surface anywhere at any time (Grudin 2002). If there is an intended audience for the information, however, there is no guarantee that it will spread more widely. To actually benefit from digital traces, therefore, there is a pressing need for the IS community to equip its fellow members with adequate skills and appropriate tools. Only then we will be able to tap into these traces and exploit them productively. By not only assessing current utilization of digital traces but also identifying both challenges and opportunities surrounding them, we are strong in our belief that our paper can play an important role here.

A Two-Dimensional Framework

It is of central importance that IS researchers critically reflect upon the data collection techniques and the data sources being utilized. The simple reason is that they determine the relevance and rigor of any individual research effort. The ways in which researchers collect data and the data sources they choose to include have a significant impact on the models and theories that can be developed (Scandura et al. 2000). They also affect the conclusions that can be drawn as well as the validity of their claims (Jick 1979; Yin 1989).

Research activity generally extracts data about individual, group, and organizational activities. Such data are usually collected from observed interactions and unobserved behavior. Observed interactions are gathered through ethnographic techniques such as in-situ observations, videotapes, and participation at the study site. Unobservables in the research process are typically behaviors and psychological states of the individual in various activities. Researchers rely on devices like surveys and interviews to tap into these unobservables. These data extraction devices attempt to take the implicit information about the individual and make it explicit. However, nowadays IT artifacts provide viable interfaces through which information about behaviors can be gathered and inferences about socio-psychological states can be inferred. Indeed, IS researchers are well-positioned to benefit from the utilization of such digital traces because they are closely intertwined with the IT artifact in both the practice and study of technology (Yoo 2010). Apparently a wide variety of data collection techniques are available to IS researchers. The current status of interview-based IS research has been recently examined (Schultze et al. 2011) and discussed Marton 2013). Comparatively little is known about the extent IS researchers transform the digital traces that technologies leave behind into value adding data sources, and how they actually do it. We next develop our two-dimensional framework through which we subsequently analyze contemporary use of data sources in IS research.

The Subjective/Objective Dimension

The subjective/objective dimension concerns the '*role*' of the researcher in the data collection effort. This dimension has long been recognized, simply because the researcher is an intricate part of the research process itself. Well-cited articles such as that of Schultz (2000) provide in-depth and detailed accounts of the role of the researcher at the study site including interactions with the informants. These accounts highlight the complex role that researchers play in the research process where they have to be the provocation for the data, but at the same time be separate and aloof from the source of data.

Berger and Luckmann (1967) made the early distinction between a subjective and an objective construction of reality and knowledge. Burrell and Morgan (1979) have adapted their perspective to help characterize and understand existing social theories based on four major paradigms. The subjective/objective distinction was used to characterize individualistic versus structural social theories respectively. This was a methodological distinction adopted by researchers who approached the research problem in distinctly different ways. Objectivists employ techniques and tools to identify and explain relationships between various elements, whereas subjectivists focus on the identification of the way in which reality is created and modified by individuals. Silverman (2006, p. 201) notes: "Using research interviews (or focus groups) involves actively creating data which would not exist apart from the researcher's intervention (researcher-provoked data). By contrast, observation or the analysis of written texts, audio-tapes or visual images deals with activities which seem to exist independently of the researcher. This is why we call such data naturally occurring: they derive from situations which exist independently of the researcher's intervention."

We adopt this distinction and develop the subjective/objective dimension to characterize the continuum of data sources employed in IS research. Subjective data sources refer to data of provoked or generated nature stemming from situations where the researcher interacts with the source of data and thus influence the data being collected. That is, the data is created to answer specific research questions and does not exist by itself. Examples of such data sources include interviews (Glaser et al. 1967; Miles et al. 1994; Yin 1989), questionnaires (Silverman 2006; Yin 1989), and observations (Denzin et al. 2005). Subjective data sources are thus those sources of data that are researcher-provoked. In contrast, an objective data source refers to data that is naturally occurring and exists in nature independent and is collected by the researcher without provocation. Examples of this data type include artifacts (Yin 2003), board logs, e-mails (Silverman, 2006), project documentation (Yin 2003), and screen diagrams (Silverman, 2006).

The Digitalized/Non-Digitalized Dimension

The digitalized/non-digitalized dimension concerns whether the '*actions of the users*' (individual, group, organization or technology) are reflected in the data itself, or rather inferred from the source of the data. This dimension stems from, or is conditioned by, the ability of technology to record, store, and convey traces of its performance. Expressed differently, in digitalized environments individuals leave behind tangible traces. Venturi and Latour (2010, p. 6) explain that a digitalized data source may capture directly individual intentional behavior: "The interest of electronic media lies in the fact that every interaction that passes through them leaves traces that can be easily recorded, massively stored and inexpensively retrieved. Each day, researchers discover new pools of digital data: public and private archives are sucked into computer memory, economic transactions migrate online, social networks take root on the web. Digital mediation spreads out like a giant roll of carbon paper, offering the social sciences more data than they ever dreamt of. Thanks to digital traceability, researchers no longer need to choose between precision and scope in their observations: it is now possible to follow a multitude of interactions and, simultaneously, to distinguish the specific contribution that each one makes to the construction of social phenomenon."

Historically, the social sciences have prioritized the spoken word over the written word and the written word over the nonverbal communication to understand human action. As Hodder (1994) notes, however, a full sociological analysis cannot only rely on interview data, but must also consider the digital traces. The analysis of digital traces is often depicted as the study of mute evidence, or artifacts created and left in the wake of our existence. Such artifacts exist beyond the moment of their creation and use, thus allowing them to be transported in time and space. This is an opportunity for researchers wishing to study action and behavior without being there to witness them. Such study takes place without the active commentary

of the creator and/or user of the artifact though (Hodder 1994; Markham 2004).

Digital traces allow new light to be shed on topics, and different facets of problems to be explored (Bloor 1997). Involuntary traces are those that users of online environments leave without meaning or intending to. On the contrary, voluntary traces display themselves more clearly as communicative acts. Being a by-product or produced intentionally digital traces cannot be viewed as neutral instruments. All digital traces represent reflections of values and motives of the creator and/or user (Macdonald et al. 1993). They must therefore be understood in the context of their development and use. To obtain an adequate understanding of a complex phenomenon through the study of digital traces requires that a plurality of sources be examined.

Research Approach

Our ambition was to assess to what extent and in what ways digital traces as a data source are utilized in IS research. To this end, we reviewed articles in five leading IS journals that publish rigorous scientific research during the period between 2004 and 2009. We chose the following five journals: Information Systems Journal, Information Systems Research, Journal of the Association for Information Systems,, Journal of Strategic Information Systems, and MIS Quarterly. This selection was motivated by the assumption that these journals promise to include the state of the art reporting on every aspect of research methodology, including data collection and analysis. With regard to the review process, we started off by creating an EndNote library based on the ISI Web of Knowledge. One issue emerged in relation to JAIS though. Because it was not included in the Social Science Citation Index until 2006 some of the articles were missing in the ISI Web of Knowledge. We had therefore to manually add the missing references. All articles were subsequently downloaded in part with the assistance of research assistants, and structured according to journal, volume, and, issue.

To facilitate classification, we created an MS Access database and imported all articles. The first step involved classifying whether an article was empirical or not. The qualifying criterion for the sample was that the article included data from empirical sources (such as articles, documents, and e-mails, and interviews). We did not rely on a perusal of the abstracts. We instead browsed every article to determine whether they included any empirical data sources. Then the individual data sources in each article were listed and classified according to our two-dimensional framework. The names that are used for each of the data sources are those employed by the authors of the paper as they describe the sources of data. Consequently, in order to keep the analysis of data sources grounded, the data sources described are based on author use of terminology. The classification was executed by two of the authors in concert. The joint classification strategy was chosen to create a shared understanding of the empirical data as suggested by Saldaña (2009) and Guest and MacQueen (2008). In total, we reviewed 821 articles of which 524 (63.82 percent) were found to be empirical (comprising one or several data sources). Table 1 below summarizes the number of reviewed articles and the number of empirical articles by journal.

| Table 1. Published articles with empirical data							
	Year						
	2004	2005	2006	2007	2008	2009	Total
Journal							
ISJ	14 (20)	13 (21)	11 (19)	13 (24)	21 (32)	21 (29)	93 (145)
ISR	16 (21)	17 (22)	19 (24)	19 (25)	23 (28)	22 (31)	116 (151)
JAIS	6 (18)	7 (14)	18 (34)	15 (36)	18 (34)	21 (36)	85 (172)
JSIS	12 (22)	16 (22)	13 (18)	16 (22)	15 (20)	12 (18)	84 (122)
MISQ	16 (30)	21 (34)	29 (46)	26 (35)	28 (39)	26 (47)	146 (231)
Total	64 (111)	74 (113)	90 (141)	89 (142)	105 (153)	102 (161)	524 (821)

Classification of Data Sources

The cumulative count of data sources in empirical IS articles was 1045. Of these 1045 data sources, 346 were unique. The number of data sources in individual articles ranged from one to nine. Questionnaires, interviews, observations, and secondary data sources are the most commonly used data sources in IS research (see Table 2). In addition, there are several other data sources being used too. These sources are aggregated into the 'Other' category. While this category itself forms a large part of the overall range of data sources, the individual data sources that comprise it form only a small fraction of the data sources used in the articles. For example, auction and bidding data from sites such as Ebay are employed in articles but are used only twice in two articles across separate journals. Similarly, sources such as the Amazon database, clickstream data, and others form a small part of the overall use of data sources in IS research.

Table 2. The most used data sources in leading IS journals						
	Journal					
	ISJ	ISR	JAIS	JSIS	MISQ	Total
Questionnaire	28	48	32	21	74	203
Interviews	47	19	22	56	35	179
Observations	10	2	5	9	6	32
Secondary data	4	7	5	3	5	24
Documents	7	4	4	2	4	21
Other	107	108	90	140	141	586
Total	203	188	158	233	263	1045

The 1045 data sources identified in the articles were then used in the analysis. Drawing on our prior theoretical development, each data source was classified as subjective, objective, digitalized, and non-digitalized (see Figure 1). The figure depicts the distribution of data sources in IS research along the dimensions of subjective/objective and digitalized/non-digitalized. The analysis indicates that the dominant forms of IS research rely on data sources that are oriented toward non-digitalized and subjective understandings of individual, group, and organizational activity and behavior. There are also a significant number of articles that employ objective sources of data. However, digitalized sources of data are underrepresented in both subjective and objective dimensions. Data sources that are mentioned only once across the journals and years sampled have been aggregated into an "Other" category.

Figure 1. Classified data sources
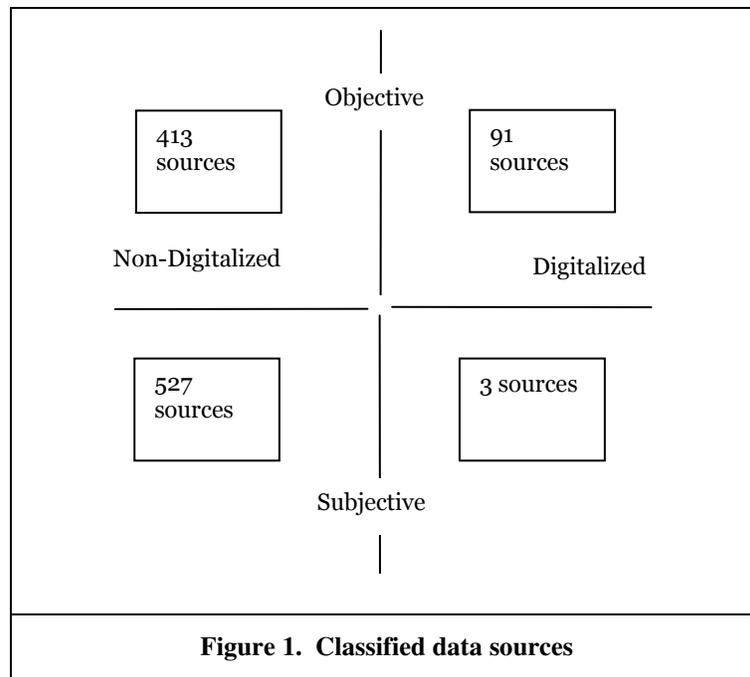
Data Sources in Use

Whereas we above classified the various data sources and the manner in which they were employed across the four quadrants (see Figure 1), this analysis will scrutinize the top five data sources within each quadrant. The objective is to highlight and contrast the varying data sources employed across our framework. Table 3 highlights the most frequently used data sources for each quadrant (objective/non-digitalized, subjective/non-digitalized, objective/digitalized, subjective/digitalized) from Figure 1.

Table 3. Data sources in use				
Order desc.	Objective/non-digitalized		Subjective/non-digitalized	
	Data Source	Count of use	Data Source	Count of use
1	Secondary data	24	Questionnaires	203
2	Documents	21	Interviews	179
3	Documentation	18	Observations	30
4	Simulated data	15	Web-based questionnaires	14
5	Articles	14	Focus groups	9
	Objective/digitalized data		Subjective/digitalized	
	Data Source	Count of use	Data Source	Count of use
1	E-mail	10	Observations	1
2	Web site data	6	Blogs	1
3	Log data	6	Conversation threads	1
4	E-mail messages	5		
5	Transaction data	4		

Objective/non-digitalized data sources

The objective/non-digitalized data sources quadrant describes those data sources that the researcher has little involvement in as a source of data, and also represents those data sources that do not have a digitalized component. There are 413 such sources of data reported in the IS research articles analyzed. 227 of these 413 sources of data are unique, i.e. 186 instances of reported data sources are non-unique.

The primary sources of data are thus secondary sources, documents, documentation, simulated data, and articles. While documents and documentation may refer to similar type of data, they are mostly used for developing rich context descriptions of the organization, environment, industry, group, process, etc. These frequently used data sources typically support interviews, focus groups and/or case studies. The number of times these top five data sources have been reportedly used represents 22.27 percent of all such data sources in this quadrant.

Subjective/non-digitalized data sources

The subjective/non-digitalized data sources quadrant describes those data sources that the researcher has significant involvement in as a source of data, and also represents those data sources that do not have a digitalized component. There are 527 such sources of data reported in the IS research articles analyzed. 66 of these 527 sources of data are unique, i.e., 461 instances of reported data source are non-unique.

The most frequently used data sources are thus questionnaire, interviews, observations, web-based questionnaires, and focus groups. Questionnaires and web-based questionnaires are distinguished because they employ different techniques for collecting data. Overall, these five data sources represent 82.54 percent of all reported data sources within this quadrant. These data sources are primarily used in interpretative, case study, survey, experimental research, and constructs generated from these sources are employed in a wide variety of ways.

Objective/digitalized data sources

The objective/digitalized data sources quadrant describes those data sources that the researcher has little involvement in as a source of data, and also represents those data sources that have a digitalized component. There are 91 such sources of data reported in the IS research articles analyzed. 58 of these 91 sources of data are unique, i.e. 33 instances of reported data source are non-unique.

The top five most frequently used data sources are thus e-mail, web site data, log data, e-mail messages, and transactions data. E-mail and e-mail messages have been separated out as they represent different types of data captured. The former has been employed to characterize data that uses both the text of the e-mail and the information about sender and receiver, while the latter represents typically the content of the message alone. These sources of data have been cumulatively used 34.07 percent of the time in reported IS research, and are typically used as causal, moderator/mediator or dependent constructs. They are also usually used in quantitative analysis or an analysis of processes.

Subjective/digitalized data sources

The subjective/digitalized data sources quadrant describes those data sources that the researcher has significant involvement in as a source of data, and also represents those data sources that have a digitalized component. There are three such sources of data reported in the IS research articles analyzed. Each data source within this quadrant represents 33% of the reported data sources. Overall, the data sources within this quadrant are underrepresented compared to data sources in other quadrants.

Discussion

Assessing Current Behavior

Applying our two-dimensional framework, we have analyzed a six-year literature survey covering five

leading IS journals. The analysis clearly shows that these journals primarily published empirical studies that build on well-recognized data sources collected through conventional methods. We believe this result signals what characterizes the dominating behavior of our fellow colleagues who are engaged in empirical research activities. However, the analysis also indicates that IS researchers have at least occasionally turned to digital traces as a source of data, which represents a new type of instrumentation that captures individual, group, organizational, and technological activities and behaviors. While such tangible traces possess the potential to help reveal the nature of these activities and behaviors, both objective and subjective digitalized sources have been employed in IS research. Indeed, our analysis demonstrates the utility of these sources for theorizing the design and use of IT.

Reflecting further on our survey analysis, we found a large number of objective digitalized data sources, of which e-mail logs and other communication related sources represent the largest portion. Besides these sources, transaction logs have been frequently utilized in IS research. One prominent early example is Stenmark's (2001) study of how to capture tasks carried out on an Intranet through published documents, submitted search engine queries, or web server log files. By exploiting a user's everyday actions in an unobtrusive manner, the Intranet activities a user is already engaged in can be aggregated and turned into an organizational benefit, thus revealing otherwise invisible patterns. Given that a corporate Intranet contains work related material, information retrieval systems may reveal people's tacit knowledge by making salient their search patterns, Stenmark argues. In the same vein, Lindgren et al. (2003, 2004) propose agent-based recommender system supportive of information seeking activities as means for competence identification in knowledge-based organizations.

Our analysis identified three research articles that employ subjective digitalized sources of data. Analyzing supply chain processes in disaster management efforts, Day et al. (2009) rely on data, including interactive personal blogs, from a wide range of relief providing and demanding organizations. Along the same line, Abbasi and Chen (2008) assess their design framework for information representation through the utilization of experiences emanating from discussions in various digitalized forums. The perhaps most intriguing example is Chesney et al. (2009) who explore the negative behavior of "griefing" in online environments by analyzing data sources being collected within the virtual world of second life itself. As participants of the community, their immersive data generation approach involved interviews, observations as well as focus groups. These three papers indicate that IS researchers increasingly enter into the digital space seeking closer engagement with the study object.

It is well worth repeating: IS researchers mainly use non-digitalized and well-established data sources that have been instrumented across several scientific disciplines. Ambitiously and accurately documented in textbooks (Cook et al. 1979; Denzin et al. 2005; Yin 1989) the experience of applying them has greatly contributed to the cumulative tradition of the IS discipline. The instrumentation provides rigor in data collection and analyses with scientifically established quality measurements of analytical results (e.g., R2) and coding reliability (e.g., Cronbach Alfa). At the same time, however, these benefits unfortunately also serves as an effective filtering mechanism that directs our attention towards already established and tried sources of evidence, ultimately blinding us to the rich plethora of data sources increasingly populating our sociomaterial world. Observing this phenomenon, Venturi and Latour (2010) assert that we are entering an age of abundance of digitalized data. It is therefore important that we better understand the emergence of digital traces and its consequences for IS research. We must leave our comfort zones and reshape the landscape of IS research methods, so that our discipline can take a central role in theorizing of sociomaterial practice.

To deal productively with digital traces requires profound changes in the ways we organize our empirical studies in the future. We posit that digital traces instigate a quest for adapted or even new research skills and methods (see also Venturi and Latour 2010; Watts 2007). The IS literature has so far rarely considered how digital traces affect an individual researcher's strategic choices and subsequent actions. This is unfortunate because we believe IS researchers are perfectly positioned to take the lead here by capitalizing on our intellectual heritage and core capabilities. Our objective with this paper is therefore to provide a starting point for a deeper conversation around digital traces among IS scholars by formulating glimpses of a research agenda. To this end, we identify viable theoretical opportunities and outline significant research challenges.

Theorizing Digital Traces

It goes without saying that new sources of data will influence not only IS theorizing in general (Scandura and Williams 2000), but also the validity of our claims (Jick 1979; Yin 2003). Our reference disciplines have provided us with a rich pedagogy on how to develop and test theories (Eisenhardt 1989; Glaser et al. 1967; Yin 1989), which explains a wide range of behavior and phenomena relevant to IT domains. However, this pedagogy pays very little or no attention to the notion of digital traces. With such traces we mean trails that individuals, groups or organizations leave behind in digitalized environments, including their usage of mobile devices, embedded sensors, and web applications etc. (see e.g., Yoo 2010). As such, they promise to enable more accurate theorizing of increasingly sociomaterial activities and processes like information personalization, service innovation, and targeted marketing. This is not to say that traditional data sources are no longer useful or necessary, but rather that digital traces represent a timely addition making our research methods portfolio much more complete.

Here we draw upon Yoo's (2010) seven material properties of digital artifacts to further discuss data sources. They are programmability, addressability, sensibility, communicability, memorizability, traceability, and associability. A specific data source does not need to rely on one single property, but may rather be the product of a multitude of such properties. That is, the correspondence between them is thus one to many. Note that we not found any data source that does not correspond to any of the seven material properties presented by Yoo (2010). Drawing on the types of traces previously identified, we next elaborate on how and why they by themselves or combined with conventional data sources may advance IS theorizing.

First, programmability enables novel sets of logics, behaviors, and functions that create new data sources embedded within the digitalized artifact itself. Such digital traces can be productively used to examine, for example, the ways in which responsibilities are transferred and offloaded to IT artifacts in project management. Here it is especially interesting to see how and why these artifacts take on new roles and tasks, ultimately affecting the existing work arrangement and its constituent parts. Indeed, task programmability is a vital component of agency theory that is also relevant to the project management context (Eisenhardt 1989). It refers to the degree to which an appropriate behavior may be ascribed to an agent supposed to execute a specific task (Mahaney et al. 2003). While this makes the task more or less programmable, however, the ability of a digitalized artifact to cater for the responsibilities of such task increases its programmability. Digital traces capturing novel logics and routines inscribed into digital artifacts therefore spur theory development by allowing us to identify and analyze innovative changes to process behavior and their consequences in project management.

Second, addressability gives birth to digital traces of the erection and establishment of temporal device networks as well as the interactions going on within them. In the distributed cognition discourse, for example, such traces may play a pivotal role in that they can help us understand how to distribute an individual's cognition across multiple digital artifacts (Rogers et al. 1994). Without a detailed understanding of the user's immediate interaction with and relation to a set of digital options, however, the design task of productively assembling them becomes virtually unmanageable. We posit that the theoretical importance of digital traces here is significant, especially because the construction of such cross-channel ecologies is receiving more and more attention.

Third, digital traces created by the sensibility property of a digitalized artifact reveal the individual user's action, behavior, and movement. These traces can shed light on how and why the artifact itself becomes an integral part of the process where an individual makes sense of and respond to contextual changes in the environment (Weick et al. 2005). IS researchers can benefit from them because of their potential to expose the nature of sensemaking enabled by technology support. In particular, we can thanks to them carefully scrutinize the ways in which sensemaking emerges through the modifications an individual user makes to better tailor a digitalized artifact (Germonprez et al. 2007).

Fourth, communicability enables a digitalized artifact to interact with other entities by sending and receiving messages, which makes it part of a temporal network comprised of loosely coupled devices and systems. Communication through multiple such devices and systems is at the heart of contemporary organizations, and therefore this property is central to theories of organizing. The digital traces made available through it allow us to closely examine, for example, evolution of organizing behavior and outcomes of communication patterns. Indeed, communicability plays a central role in sociomaterial

networks because it forms the foundation of their relationship structures (Borgatti et al. 2003). Given that individuals largely manage the emergence of such networks (Kane et al. 2008), the digital traces they leave behind can help exploring the formation and resolution of them. These traces can therefore influence theorizing of the anatomy of sociomaterial networks.

Fifth, memorizability allows a digitalized artifact to record and store generated, sensed, and communicated information. Sociomaterial work practice can benefit from such property because individuals must maintain memory systems of knowledge and resources (i.e., transactive memory) to cope with requirements critical for organizational performance (Hollingshead 2000; Moreland et al. 2000). We know that digitalized artifacts complement transactive memory systems managed by individuals in organizations. The digital traces that follow utilization of these artifacts promise to shed light on not only how to develop such systems, but also the actions through which individuals manage them. Such understanding can be productively transferred also to other knowledge management situations. It may help to reveal and explain, for example, the intricate interplay between knowledge seeking and action in situated sociomaterial work practice (Bailey et al. 2010; Phang et al. 2009).

Sixth, the traceability property of a digitalized artifact creates viable opportunities for IS research by interrelating events and entities chronologically over time. For example, the digital traces that emerge from it invite us to enact historic perspectives when exploring outcomes of system development efforts. We can take a close look at the evolution of power struggles, zooming in on critical design decision throughout the process. In addition, by analyzing such traces, we can also scrutinize how and why activities of users change while they interact with the technology under study (Erickson et al. 2002). This means we not need to rely on conventional techniques of requirements engineering/analysis, but can instead follow their material behaviors to determine use patterns (Wesler et al. 2008).

Seventh, the associability property of a digitalized artifact allows it to relate itself to other sociomaterial entities. Enabled by tags and keywords this property opens up new frontiers for studying IT design and use. To illustrate, greater collaboration around sets of tags or keywords may represent strong relationships between users. The outcome of such relationships is often increased social capital, which helps to form viable networks of users. The accumulation of digital traces invites us therefore to examine the aggregation process of social capital (Labianca 2004). This represents an interesting opportunity given our limited understanding of the enabling role of IT in such situations (Montazemi et al. 2008).

Leveraging Future Research

In response to the limited use of digital traces in IS research so far, our previous discussion about viable theoretical opportunities can be seen a deliberate attempt to inspire our fellow colleagues to start exploring a new methodological path. However, to raise their awareness of the potential of researching digital traces is necessary but not sufficient. To leverage future research that enjoys the benefits these traces offer requires also a critical discussion about what challenges we need to overcome. Perhaps the most important challenge concerns access to digital traces. Here it is essential to consider whether the data are public (i.e., open to anyone) or private (open by disclosure) and whether the data reside on the server or client side of the IT artifact.

The issue of access to public and/or private data is not entirely new. In particular, it has been previously discussed in relation to publicly funded research projects (Uhlir et al. 2007). The argument put forward is that data from such research projects should be treated as digital common goods. That is, non-exclusive information and knowledge resources that are collectively created, owned, or shared among a community" (Fuster Morell 2010). Reflective of this argument, EU and OECD have developed a regulatory framework requiring that researchers make data available and publish in open journals (Arzberger et al. 2004). In the research community, we can see efforts like JSTOR (www.jstor.org) to make publications available and accessible. We believe that these and other similar initiatives pave the way for facilitating access to digital traces.

Publicly accessible IT artifacts include, for example, Wikipedia pages, tweets from Twitter conversations, and corporate branded Facebook pages. While these sites are accessed through APIs (application programming interfaces), utilizing them for data collection purposes requires an adapted documentation strategy. The simple reason is that this type of data, unlike interview or survey data typically stored by the researcher, is outside researcher control. In addition, individuals themselves delete or remove entries

from these sites as they wish. This means that the researcher cannot expect to find the same data when returning to the source (Lewis et al. 2008). Clearly, deletions by systems and users alike put new constraints on the integrity of log data (Casey 2002). In response, the researcher can systematically recollect data to determine the degree to which data are changed or altered. Venturi and Latour (2010) view open source channels as a viable solution in that they make data available to others for flexible reuse.

Private IT artifacts pose other challenges, simply because they are closed and do not provide open access to the general public. Digital traces here include, for example, customer data from CRM systems, DNA from gene banks, and SMS traffic from mobile phones/mobile operator systems. Given that the control of these traces typically resides with individuals, organizations or governments, negotiating access rights becomes a central skill for researchers (Girardin et al. 2008; Grudin 2002). They must also be able to take issues of privacy seriously and put them center stage in their research efforts (Jessup and Robey 2002). To spur access to these private digital traces further, researchers may initiate and establish cooperation with the large Internet companies who currently dominate data collection. In such cooperation, researchers will inevitably encounter both privacy and intellectual-property issues (Watts 2007).

Besides the public/private access issue, another issue concerns whether data are stored on the server side or client side. This issue is critical because it relates very much to the level of analysis. Server side data cater for a macro perspective on sociomaterial practice. To deal with such data is challenging not only because of the large quantities of data (i.e., big data) but also because of the exponential growth in data (Lynch 2008). Blair et al. (2011) suggest that new databases or even advanced business intelligence tools might be needed to actually distill and understand such data. On the other hand, client side data provide a micro view of what is happening on the individual device. Cookies represent one such data source that can serve as valuable input to self-ethnographic studies of sociomaterial everyday practice. However, the process of accessing, collecting, and extracting client side data is sometimes difficult, which suggests that novel embedded research tools might be needed to support the task at hand.

To be able to successfully exploit the digital traces that public and private IT artifacts promise to offer it is clear that we must adapt our approaches and rethink out behaviors. The access challenges (public/private and server/client) suggest that IS researchers must either develop the necessary skills themselves or team up with those already having them. Such skills must involve data access approaches, data collection/extraction procedures, and data storage routines. However, we can only expect that the enactment of those skills in research practice will inevitably lead to unanticipated consequences.

For example, as soon as we start making data available and open to others, issues of self-plagiarism and co-authorship are likely to appear in a new light. As for the former, that kind of behavior is not encouraged by the research ethics of the IS community. In a situation where you make your data available to others, however, it is reasonable that you have the right to reanalyze/reinterpret that data too. With regard to the latter, open data is likely to change the rules of the game, simply because data collectors/providers might start to claim co-authorship. Indeed, in disciplines like biology, medicine, and physics this is legitimate or routine behavior. An example of how this could work comes from Marketing Science. While the journal requires that researchers make data available, it has opened up for a database report as a new type of submission. Such report is a dataset being published in paper or electronically that is subject to proper referencing. This not only gives credit to the data collector, but also allows for replication studies. We believe such thinking opens up for discussions in IS about publications outlets focusing on open data provision and on-line data repositories that any community member could query.

Conclusions

It is fair to say that the IS field is one of the most fast-moving disciplines. Simply because what we study evolves over time, the methods we use to approach it must change too. Recent IS research suggests we need to stop treating the social and the material as distinct and largely independent spheres of contemporary life. Rather we must acknowledge the inherent inseparability between the social and the material. By doing so we may open up novel avenues for IS research.

However, the weapons currently stockpiled in the arsenal of data collection techniques are not only inadequate, but also quite inappropriate for researching new and emerging research fronts such as sociomateriality. While this is a bold claim that is difficult to establish, we have at least tried to make the

case here. Our six-year literature survey, comprised of five leading IS journals, suggests that most of our fellow colleagues keep employing the data sources that have dominated the field since the 1960s. Unfortunately, we rarely engage in serious discussions about the nature of prevailing data collection techniques. We have therefore not seen big methodological breakthroughs lately.

Sociomateriality renders discontinuous change that requires new research methods and techniques. There is a rapidly diminishing window of opportunity in which to develop and establish these though. Bluetooth communications, Facebook updates, Internet purchases, and Twitter interactions promise to reveal what contemporary life is all about, but IS researchers seem both unable and unwilling to adapt their actions to the emerging possibilities coming our way. It is time therefore that we start asking some difficult questions about what it takes for new sources of empirical data to become acceptable and credible.

Taking the sociomateriality of everyday and organizational life seriously, we propose that digital traces offer the possibility of restructuring the study of such situations. While these traces can reveal how the social and material co-emerge in situationally specific ways, the trick is to successfully incorporate them into the research process itself as a potential source of evidence. We have offered suggestions and ideas for fellow colleagues who share our aspirations to establish digital traces as an acceptable and credible new data source. Even though it is difficult to speculate on exactly what the future will be, we believe our proposed agenda addresses at least some of the issues we must deal with to benefit from the opportunities that digital traces thankfully present.

Acknowledgement

References

Abbasi, A., and Chen, H. 2008. "CyberGate: A Design Framework and System for Text Analysis of Computer-mediated Communication," *MIS Quarterly* (32:4), pp. 811-837.

Agarwal, R., Gupta, A. K., and Kraut, R. 2008. "The Interplay between Digital and Social Networks," *Information Systems Research* (19:3), pp. 243-252.

Arzberger, P., Schroeder, P., Beaulieu, A., Bowker, G., Casey, K., Laaksonen, L., Moorman, D., Uhlir, P., and Wouters, P. 2004. "An International Framework to Promote Access to Data," *Science* (303:5665), pp. 1777-1778.

Asur, S., Huberman, B. A., Szabo, G., and Wang, C. 2011. "Trends in Social Media: Persistence and Decay," in *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*.

Bailey, D. E., Leonardi, P. M., and Chong, J. 2010. "Minding the gaps: Understanding technology interdependence and coordination in knowledge work," *Organization Science* (21:3), pp. 713-730.

Benbasat, I., and Zmud, R. W. 2003. "The identity crisis within the IS discipline: Defining and communicating the discipline's core properties," *MIS Quarterly*, pp. 183-194.

Berger, P. L., and Luckmann, T. 1967. *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*. Anchor, Garden City, NY.

Blair, G., Kon, F., Cirne, W., Milojicic, D., Ramakrishnan, R., Reed, D., and Silva, D. 2011. "Perspectives on Cloud Computing: Interviews with Five Leading Scientists from the Cloud Community," *Journal of Internet Services and Applications* (2:1), pp. 1-7.

Bloor, D. 1997. *Wittgenstein, Rules and Institutions*. New York, NY: Routledge.

Borgatti, S., and Foster, P. 2003. "The Network Paradigm in Organizational Research: A Review and Typology," *Journal of Management* (29:6), pp. 991-1013.

Burrell, G., and Morgan, G. 1979. *Social Paradigms and Organizational Analysis: Elements of the Sociology of Corporate Life*, London: Heinemann Educational.

Callon, M. 1986. "Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fisherman of St Brieuc Bay," in *Power, Action and Belief: A New Sociology of Knowledge? Sociological Review Monograph*, J. Law (ed.), Routledge and Kegan Paul, London.

Casey, E. 2002. "Error, Uncertainty, and Loss in Digital Evidence," *International Journal of Digital Evidence* (1:2), pp. 1-45.

Chesney, T., Coyne, I., Logan, B., and Madden, N. 2009. "Griefing in Virtual Worlds: Causes, Casualties

and Coping Strategies," *Information Systems Journal* (19:6), pp. 525-548.

Cook, T. D., Campbell, D. T., and Day, A. 1979. *Quasi-experimentation: Design & Analysis Issues for Field Settings*. Boston, MA: Houghton Mifflin.

Day, J. M., Junglas, I., and Silva, L. 2009. "Information Flow Impediments in Disaster Relief Supply Chains," *Journal of the Association for Information Systems* (10:8), pp. 637-660.

Denzin, N. K., and Lincoln, Y. S. 2005. *The Sage Handbook of Qualitative Research*. Thousand Oaks, CA: Sage Publications.

Eisenhardt, K. 1989. "Building Theories from Case Study Research," *The Academy of Management Review* (14:4), pp. 532-550.

Erickson, T., Halverson, C., Kellogg, W. A., Laff, M., and Wolf, T. 2002. "Social Translucence: Designing Social Infrastructures that make Collective Activity Visible," *Communications of the ACM* (45:4), pp. 40-44.

Fuster Morell, M. 2010. "Governance of Online Creation Communities: Provision of Infrastructure for the Building of Digital Commons," in *Department of Political and Social Sciences*, European University Institute, Florence.

Germonprez, M., Hovorka, D., and Collopy, F. 2007. "A Theory of Tailorable Technology Design," *Journal of the Association for Information Systems* (8:6), pp. 351-367.

Girardin, F., Calabrese, F., Fiore, F. D., Ratti, C., and Blat, J. 2008. "Digital Footprinting: Uncovering Tourists with User-Generated Content," *Pervasive Computing, IEEE* (7:4), pp. 36-43.

Glaser, B. G., and Strauss, A. L. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago, IL: Aldine Publishing Company.

Grudin, J. 2002. "Group Dynamics and Ubiquitous Computing," *Communications of the ACM* (45:12), pp. 74-78.

Guest, G., and MacQueen, K. M. 2008. *Handbook for Team-Based Qualitative Research*. Altamira Press, Plymouth, UK.

Hodder, I. 1994. "The interpretation of documents and material culture," in *Handbook of Qualitative Research*, Denzin, N. K. and Lincoln, Y. S. (eds.), Thousand Oaks, CA: Sage Publications.

Hollingshead, A. B. 2000. "Perceptions of Expertise and Transactive Memory in Work Relationships," *Group Processes & Intergroup Relations* (3:3), pp. 257-267.

Howison, J., Wiggins, A., and Crowston, K. 2011. "Validity Issues in the Use of Social Network Analysis with Digital Trace Data," *Journal of the Association of Information Systems* (12:12) art. 2.

Jessup, L. M., and Robey, D. 2002. "The Relevance of Social issues in Ubiquitous Computing Environments," *Communications of the ACM* (45:12), pp. 88-91.

Jick, T. D. 1979. "Mixing qualitative and quantitative methods: Triangulation in action," *Administrative Science Quarterly* (24:4), pp. 602-611.

Kane, G. C., and Alavi, M. 2008. "Casting the Net: A multimodal Network Perspective on User-System Interactions," *Information Systems Research* (19:3), pp. 253-272.

Kleinberg, J. 2008. "The Convergence of Social and Technological Networks," *Communications of the ACM* (51:11), pp. 66-72.

Labianca, G. 2004. "Group Social Capital and Group Effectiveness: The Role of Informal Socializing Ties," *Academy of Management Journal* (47:6), pp. 860-875.

Latour, B. 1987. *Science in Action: How to follow Scientists and Engineers through Society*. Cambridge, MA: Harvard University Press.

Lewis, K., Kaufman, J., Gonzalez, M., Wimmer, A., and Christakis, N. 2008. "Tastes, Ties, and Time: A new Social Network Dataset using Facebook. com," *Social Networks* (30:4), pp. 330-342.

Lindgren, R., Andersson, M., and Henfridsson, O. 2008. "Multi- Contextuality in Boundary- spanning Practices," *Information Systems Journal* (18:6), pp. 641-661.

Lindgren, R., Henfridsson, O., and Schultze, U. 2004. "Design Principles for Competence Management Systems: A Synthesis of an Action Research Study," *MIS Quarterly* (28:3), pp. 435-472.

Lindgren, R., Stenmark, D., and Ljungberg, J. 2003. "Rethinking Competence Systems for Knowledge-Based Organizations," *European Journal of Information Systems* (12:1), pp. 18-29.

Lynch, C. 2008. "Big data: How do your data grow?," *Nature* (455:7209), pp. 28-29.

Lyytinen, K., and Yoo, Y. 2002. "Research Commentary: The Next Wave of Nomadic Computing," *Information Systems Research* (13:4), pp. 377-388.

Macdonald, K., and Tipton, C. 1993. "Using documents," in *Researching Social Life*, Gilbert. N. (ed.), London, UK: Sage Publications, pp. 187-200.

Mahaney, R. C., and Lederer, A. L. 2003. "Information Systems Project Management: An Agency Theory interpretation," *Journal of Systems and Software* (68:1), pp. 1-9.

Markham, A. N. 2004. "Internet Communication as a Tool for Qualitative Research," in *Qualitative Research: Theory, Method and Practice*, Silverman D. (ed.), pp. 95-124.

Marton, A. 2013, forthcomming. "Purposive Selection and the Quality of Qualitative IS Research", in *the Proceedings of the 2013 ICIS*.

Miles, M. B., and Huberman, A. M. 1994. *Qualitative Data Analysis: An Expanded Sourcebook*. London, UK: Sage Publications.

Montazemi, A., Siam, J. J., and Esfahanipour, A. 2008. "Effect of Network Relations on the Adoption of Electronic Trading Systems," *Journal of Management Information Systems* (25:1), pp. 233-266.

Moreland, R. L., and Myaskovsky, L. 2000. "Exploring the Performance Benefits of Group Training: Transactive Memory or Improved Communication?," *Organizational Behavior and Human Decision Processes* (82:1), pp. 117-133.

Nyberg, D. 2009. "Computers, Customer Service Operatives and Cyborgs: Intra-Actions in Call Centres," *Organization Studies* (30:11), pp. 1181-1199.

Orlikowski, W. J. 2000. "Using Technology and Constituting Structures: A Practice lens for Studying Technology in Organizations," *Organization Science* (11:4), pp. 404-428.

Orlikowski, W. J. 2007. "Sociomaterial Practices: Exploring Technology at Work," *Organization Studies* (28:9), pp. 1435-1448.

Orlikowski, W. J., and Iacono, C. S. 2001. "Research Commentary: Desperately Seeking the "IT" in IT Research—A Call to Theorizing the IT artifact," *Information Systems Research* (12:2), pp. 121-134.

Orlikowski, W. J., and Scott, S. V. 2008. "Sociomateriality: Challenging the Separation of Technology, Work and Organization," *The Academy of Management Annals* (2:1), pp. 433-474.

Phang, C. W., Kankanhalli, A., and Sabherwal, R. 2009. "Usability and Sociability in Online Communities: A Comparative Study of Knowledge Seeking and Contribution," *Journal of the Association for Information Systems* (10:10), pp. 721-747.

Rogers, Y., and Ellis, J. 1994. "Distributed Cognition: An Alternative Framework for Analysing and Explaining Collaborative Working," *Journal of Information Technology* (9), pp. 119-119.

Saldaña, J. 2009. *The coding manual for qualitative researchers*, London, UK: Sage Publications.

Scandura, T. A., and Williams, E. A. 2000. "Research Methodology in Management: Current Practices, Trends and Implications for Future Research," *Academy of Management Journal* (43:6), pp. 1248-1264.

Schultze, U. 2000. "A Confessional Account of an Ethnography about Knowledge work," *MIS Quarterly* (24:1), pp. 3-41.

Schultze, U., and Avital, M. 2011. "Designing Interviews to Generate Rich Data for Information Systems Research," *Information and Organization* (21:1), pp. 1-16.

Silverman, D. 2006. "Interpreting qualitative research," London: Sage Publication Limited.

Stenmark, D. 2001. "Leveraging Tacit Organizational Knowledge," *Journal of Management Information Systems* (17:3), pp. 9-24.

Suchman, L. 2000. "Organizing Alignment: A case of Bridge-Building," *Organization* (7:2), pp. 311-327.

Suchman, L. 2005. "Affiliative Objects," *Organization* (12:3), pp. 379-399.

Suchman, L. 2006. *Human-machine Reconfigurations: Plans and Situated Actions*. Cambridge University Press, New York.

Uhlir, P. F., and Schröder, P. 2007. "Open Data for Global Science," *Data Science Journal* (6), pp. 36-53.

Venturini, T., and Latour, B. 2010. "The Social Fabric: Digital Traces and Quali-quantitative Methods," in *Proceedings of Future En Seine 2009: The Digital Future of the City*, Digital, C. (ed.), Paris, France, pp. 30-15.

Watts, D. J. 2005. "A twenty-first Century Science," *Nature* (445:7127), pp. 489-489.

Weick, K. E., Sutcliffe, K. M., and Obstfeld, D. 2005. "Organizing and the Process of Sensemaking," *Organization Science* (16:4), pp. 409-421.

Wesler, H., Smith, M., Fisher, D., and Gleave, E. 2008. "Distilling Digital Traces: Computational Social Science Approaches to Studying the Internet," *The Sage handbook of online research methods*. Sage: London. pp. 116-140.

Yin, R. 1989. *Case Study Research: Design and Methods*, Sage Publications, London, UK.

Yoo, Y. 2010. "Computing in Everyday Life: A Call for Research on Experiential Computing," *MIS Quarterly* (34:2), pp. 213-231.