The Hermeneutics of Trace Data: Building an Apparatus

Carsten Østerlund, Kevin Crowston & Corey Jackson

Syracuse University School of Information Studies
Hinds Hall, Syracuse, NY 13244 USA

{costerlu,crowston,cjacks04}@syr.edu

Abstract

When people interact via information systems, the data is captured by the systems as a side effect of the interaction. These data are increasingly interesting and available for research. In a sense, these systems become a new kind of research apparatus, and like all advances in instrumentation, open up new areas of study with the potential for discovery. While at first glance, such "big data" analysis seems to be most suitable for a positivist quantitative research approach. However, a closer inspection reveals that interpretive research strategies may better support the challenges associated with digital trace data. By merging insights from hermeneutics and sociomateriality, we argue that trace data analysis entails the building of a research apparatus. Hermeneutic principles play a key role in the application of this apparatus and allow researchers to make sense of the often partial traces left by online participants. Drawing on longitudinal trace data from a study of citizen science practices the paper illustrates the value of merging insights from hermeneutics with sociomaterial insights. The approach allows researchers to account for not only the material dynamics of digital trace data but also the temporal dimension of online practices.

1 Introduction

As people increasingly interact via information systems, the data captured by such systems becomes increasingly interesting and available for research. Data captured as a side effect of system use—what we refer to as trace data—offer great potential insight into the actual behaviours of many users. Researchers usually face a tradeoff between the number of subjects studied and the volume of data about each one, but trace data potentially provides the ability to see everything each individual does. In a sense, these systems become a new kind of research apparatus, and like all advances in instrumentation, open up new areas of study with the potential for discovery.

While at first glance, such "big data" analysis seems to be most suitable for a positivist quantitative research approach, we argue in this paper that it is instead the insights of interpretivist research that are most necessary. By merging insights from the literatures on hermeneutics and sociomateriality, we argue that digital trace data analysis entails the building of a research apparatus. Key to this apparatus are the techniques for interpretation of intent and meaning behind the observable performances. In short, as research data, trace data are observations of behaviour, a kind of data that qualitative researchers, ethnographers in particular, are best equipped to analyze. In this paper, we discuss how application of the hermeneutic circle helps identify likely problems in making sense of trace data.

Howison et al. (2011) define digital trace data as "records of activity (trace data) undertaken through an online information system (thus digital). A trace is a mark left as a sign of passage; it is recorded evidence that something has occurred in the past." For example, many studies have used posts on discussion fora as data. They identify three characteristics of trace data that set them apart from the kinds of data often used in information systems research, such as surveys: "1) it is found data (rather than produced for research), 2) it is event-based data (rather than summary data) and 3) as events occur over a period of time, it is longitudinal data". Though not included in Howison et al.'s (2011) definition, a further characteristic is that trace data are typically semi-structured, with a number of structured metadata

fields (e.g., for a post in a discussion forum, the date and time, the ID of the poster, the name of the forum, possibly a previous message being replied to, ratings by other readers, etc.) and possibly some unstructured data (e.g., the subject or content of the post).

These first two properties (found data and event-based data) taken together mean that the data are not intended as measures of a concept of theoretical interest, such as a user's attitudes or beliefs, but are rather records of actual behaviours on the system that have to be interpreted to make a conceptual connection. Such data are thus amenable to interpretivist researchers who are used to having to make inferences from observed behaviour. In this paper, we use the hermeneutic circle to organize our discussion of the practice of analyzing trace data.

The third characteristic, trace data being longitudinal data, raises questions about the temporality both of the data and of the phenomena of interest. Evidence from events spread over time have to be aggregated to describe the system at a particular point in time. Such aggregation can be problematic if the constructs of interest evolve or change over time.

The second part of Howison et al.'s (2011) definition of **digital trace data** is that the data are both produced through and stored by an information system. Trace data can be produced through direct observation, as in traditional ethnographic research that might record the events in some work environment. However, online interaction has lead to an increase in the use of data captured about these interactions. To correctly interpret such *digital* trace data requires a deep understanding about the details of the specific system technology that captured the data. Of course, any study requires an apparatus for collecting and managing data. But researchers employing interviews or observation (for example) may take the apparatus for granted, being familiar with those approaches and the challenges they present. In this way, trace data bring to the forefront issues about the sociomateriality of the apparatus for data analysis.

2 Sociomateriality and Hermeneutics

Going back to Marx and the Tavistock studies, scholars have gathered and analyzed traces of organizational practices in ways suggesting that technologies, people and discourses come together in dynamic and reciprocal assemblages (Gaskin et al. 2014). The recent sociomaterial turn shines a bright light on this relationship by insisting that the material and social are inseparable in organizational action. Boundaries between humans and the material blur. The two are inseparable, constitutively entangled (Barad 2003; Orlikowski et al. 2008). Sociomateriality highlights the nexus of doings, materialities, and discourses that people carefully enact to support certain practices (Law 2004; Suchman 2007), and offers an analytical gaze under which neither artefacts, people, nor practices can stand naked and alone, revealing some inherent properties. Instead these are bound into one entity where only concrete interactions occur between artefacts, people, and practices. This goes beyond people's mere doings, a sociomaterial lens highlights the *performative* character of action in which objects are constituted, bodies shaped, words formed, and things described.

Applied to the work of research and data analysis, this perspective makes clear that traces are not pre-given entities, packaged and waiting to be picked off a shelf. Rather, traces are the product of specific sociomaterial arrangements, what Barad (2003) terms an apparatus. For instance, a person's position in space and time only gains meaning when a rigid *apparatus* exists and establishes a fixed frame of reference for the measurement of position. Barad argues that one cannot disentangle the phenomenon and the apparatus, so any future measurements of position by this apparatus will be part of the phenomenon studied and inseparable from the object of interest. The apparatus plays a constitutive role in the production of the phenomenon by enacting specific boundaries in our sociomaterial reality. In this way, the apparatus is not an inscription device installed before the action happens. It is not a neutral probe, measuring pre-existing entities. Instead, the apparatus stands out as an open-ended practice constantly producing and reproducing the phenomenon. These practices are open to rearrangements. They are reworked and adjusted. The

creativity of scientific practices includes the skill of making the apparatus work for a particular purpose. Elements are reworked and adjusted, leading to adjustments of the boundaries and cuts performed by the apparatus and the nature of the phenomenon. An apparatus can itself become the phenomenon, the focus of attention. This can happen as researchers turn their attention to the boundaries performed or by engaging the process in which the apparatus intra-acts with other apparatuses. These relations are only locally stabilized phenomena that are part of specific performances.

In this paper, we use hermeneutics as a framework for analyzing issues in the practice of trace data analysis. Hermeneutics has long been a trusted pillar of qualitative IS research. Boland (1985)—inspired by Edmund Husserl's phenomenological perspective and Gadamer's work on hermeneutics (Gadamer 1975)—was among the first scholars to introduce hermeneutics to IS research. Hermeneutics takes its point of departure in a *text*, not the apparatus for gathering a text, and emphasizes *interpretation* as opposed to performativity. In classic hermeneutics, the text constitutes an object of study, which is to be understood on the basis of its own frame of reference (Kvale & Brinkmann 2009). The interpretation, for its part, aims to bring to light an underlying coherence or sense to an otherwise incomplete, cloudy, or contradictory text (Myers 1995).

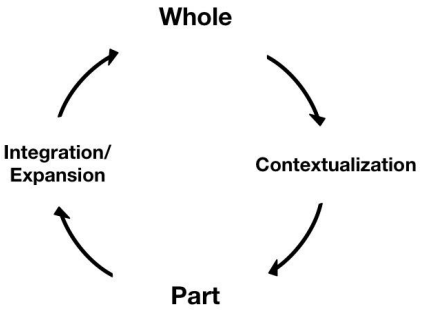
At first sight, the text conceived as a predefined entity seems to be counter to a sociomaterial approach emphasizing how open-ended practice constantly produces and reproduces a phenomenon. Given the emphasis on *performativity* and *apparatus,* one may wonder what role, if any, a sociomaterial move leaves for hermeneutics. However, there are different schools and approaches to hermeneutics. In IS, the conservative school with a focus on given texts is less common, while a rich IS literature has emerged focusing on pragmatic and critical approaches to hermeneutics (e.g., Lyytinen & Klein 1985; Myers 1995; Butler 1998; Cole & Avison 2007). Unlike in literature, for IS researchers, texts are rarely pre-given. IS researchers often take an active part in not only the selection of organizational documents to be analyzed but also in the very production of the text, be it an interview transcript or field note. This application of hermeneutics has a strong parallel to Barad's approach.

Many recent IS scholars have been inspired by Ricoeur (1981) and his use of hermeneutics to study practice and social actions involving case study notes, interviews, organizational files, and other documents related to the particular topic at hand. Here, the researcher's position in relation to the material plays a central role in the investigation by taking into account the researcher's position and history in regard to the study subject. Such a pragmatic and critical approach to hermeneutics allow scholars to overcome many of the shortcomings associated with a purely interpretive approach by paying attention to: 1) the broader context that gives rise to certain meanings and practices, 2) the unintended consequences of actions, 3) structural conflicts within organizations and societies, and 4) the temporal dimension of a social order and how any text must be seen in its historical context (Myers 1995).

In summary, from the IS debate around the application of hermeneutics, we extract seven tenets:

1. Part-Whole Flow: Constant movement exists between the part and the whole.
2. Integration: Partial interpretation is tested against the global meaning.
3. Contextualization: An understanding of the context and the topic is needed to perform high quality interpretation.
4. Presuppositions: The researcher will always bring his or her own perspective and presuppositions to the study of a phenomenon. Asking a question in and off itself presumes presuppositions.
5. Appropriateness: The questions that the interpretation deals with must emanate from the phenomenon. They must be issues raised by the phenomenal.
6. Coherence: The interpretation seeks to reach a good "gestalt," an inner unity that is free of logical contradictions. The interpretation of the phenomenon must present a unified picture and not be contradictory.
7. Suggestiveness: A good interpretation/understanding is suggestive' or fertile in that it raises questions that stimulate further research and interpretation.

The figure below summarizes the the cyclic relationship among the first three tenets in particular: an understanding of the parts is integrated to support a developing understanding of the whole, which in turn contextualizes the parts, leading a more coherent and richer interpretation. Point 4 emphasizes that these interpretations do not emerge tabula rosa, but rather are coloured by the analyst's own experiences. Tenets 5, 6, and 7 address stopping conditions for the cycle.



3 The Hermeneutics of Trace Data Analysis: Building an Apparatus

In this section, we walk through the cyclical stages in a hermeneutic analysis and discuss issues that arise in the process of analyzing trace data. The issues are arranged roughly in the order they might be encountered by researchers working their way around the circle. To illustrate these issues, we will use as a running example the analysis in a recent paper from the *Conference on Computer-Supported Cooperative Work and Social Media*. The paper, "Which Way Did They Go?" (Jackson et al. 2016), used a mix of qualitative interview data and trace data from the Planet Hunters citizen science project. Planet Hunters volunteers carry out many activities on the project website, including annotating images, reading content, and participating in discussion fora (called "talk"). Records of these activities are stored in a number of different database systems. For example, annotations of the images are stored in a MySQL database and comments in talk messages are stored in a Mongodb server. By linking data from these varied data sources, Jackson et al. (2016) reconstructed how users participated on Planet Hunters. They first aggregated the recorded activities into sessions, defined as a sequence of activities separated by no more than 30 minutes. Sessions were then analyzed to identify distinctive patterns of work in a session, e.g., sessions focused primarily on annotating vs. sessions that included a mix of annotation and commenting. Different kind of users were then identified by the mix of session types they exhibited throughout their span of interaction with the system. The researchers thus moved through analysis at the individual activity level, to the session level, and then the user history level in their analysis. The meaning of these sessions and user types were discussed with reference to data from interviews. An example of a finding from the study is that while only a handful of users regularly contribute to talk, many others refer to those discussions and seem to find them useful in orienting towards the work of the project.

While the Jackson et al. (2016) paper serves as our point of departure, description of the cyclical analysis requires us to extend our timeline beyond the production of this particular paper. We must look back at the past three years during which the authors were engaged in the building of their apparatus. This development includes not only assembling traces from different parts of the Zooniverse system but also influencing the development of the Zooniverse by asking developers to expand the apparatus, as will be discussed below.

3.1 Cobbling Together an Apparatus

In the sociotechnical perspective, data are not simply acquired, but rather an apparatus needs to be assembled. Building the apparatus constitutes a first step in a trace data study, as the researchers cobble together material from multiple sources. In doing so we not only generate data, but also create the

phenomena being studied: e.g., citizen science work practices in Zooniverse for Jackson et al. (2016). To do this creation successfully, it is critical to understand the particular details of the system in question and how those shape the traces.

Data might be generated by "scraping", that is, recording data from publicly visible web pages. With cooperation of system managers, it may be possible to obtain dumps from the databases driving the system, which was the approach taken in the Jackson et al. (2016) study. In either case, capturing and managing a volume of structured data often requires a different kind of data analysis infrastructure than is needed for analyzing textual documents. For example, Jackson et al. (2016) needed their own versions of the various database systems to process the dumps and data analysis required considerable preprocessing to get the data into the correct format.

While it is tempting to expect that the system captures traces of all events in a given time period, this assumption should be carefully examined. The user interface available for scraping may show only a subset of data (e.g., omitting some historical or private data or displaying a sample of a large dataset). Database dumps may also be incomplete: as Howison et al. (2011) point out, systems are subject to many problems that result in data loss (e.g., server outages, disk failures, deleted log files, or truncated database tables), though the problems may not be immediately visible. To address these problems requires developing a deep understanding of the fine details of the technical system. Unfortunately, it can be hard to obtain the necessary exposure without the assistance of those running the system (Howison et al., 2011).

A major caveat to using trace data is that many activities of interest related to system use are not available for integration into the apparatus. That is, there can be a lot of activity leading up to the use of the system that is not captured in any system trace. For example, in studying OSS development, trace data will capture the code check-ins and some of the discussion, but not the individual development work on developers' own computers. The implication is that studies using trace data need to consider critically the entirety of the human activity of interest and what fraction is reflected in traces.

In some cases, pre-existing traces might be sufficient to address the research phenomenon. But when traces are incomplete or fail to completely address the behaviours of interest, additional data are needed. One possibility is creating additional traces. The researcher might play a role as a co-constructor of traces, arranging with software developers to have the system collect new traces. For example, a researcher studyings FLOSS development could expand their apparatus by convincing developers to run work-tracking software on their computers. The expansion can be iterative where the researcher cycles between appreciation of what new data can be collected (or is able to be observed) and consideration of appropriate sources of evidence to address the phenomenon.

In our own work, we have spent considerable time defining specific data related to users' interactions online. As an example, lurking (using a system without visible participation) is a common step in a user's learning how to participate in an online community. However, lurking is often not an observable behaviour in trace data, i.e., the easily visible traces address only posting behaviour. System developers had not considered the possibility of analyzing anonymous users, so the ability to track this behaviour did not exist. Adding a capability to track which web pages users visited created novel trace data that was constituted for a study of learning in Zooniverse. Thus, researchers can play a central role in building the apparatus and its resulting traces that can go beyond cobbling together parts from existing systems.

Many other researchers will not have the same opportunity to influence system design. Nevertheless, even a study of a fixed systems, e.g., analyzing Twitter data, still involves the building of an apparatus. For example, a researcher's apparatus might be narrowed to only include tweets for a single event or a specific time period. Most of the parts of the apparatus are designed by others but the researchers make informed decisions and the process involves the creative skill of making the apparatus work for a particular purpose.

3.2 From Whole to Parts (Contextualization)

Engaging the apparatus allows one to move through the hermeneutic circle, which starts by forming an interpretation of its parts by identifying the various actions and practices recorded in the traces. A first step is simply understanding the connection between system actions and the trace data record. While data may have particular labels (particularly in a database dump), the connection between that label and an action is not always straightforward. For example, a database architect might give a database field a name that is suited to the context of the system's creation, but which may not be meaningful to a researcher. Longitudinal data pose particular challenges because as systems evolve, the use of particular data items may change. One possible strategy is to interact with the system personally and observe how those interactions appear in the traces (though this strategy won't help and might actually be misleading for interpreting older data). Jackson et al. (2016) described how they engaged in Planet Hunters as registered users: creating accounts, completing tutorials, using project resources, annotating images, and posting comments which framed their understanding of how newcomers engaged in the project.

Describing complex events may require data from multiple sources. For example, in the Jackson et al. (2016) study, making sense of an annotation required merging together data from multiple sources, such as user data, metadata, and the image, and these are in different databases and database tables.

Trace data further offer the possibility to work at multiple levels of analysis, e.g., single actions, sessions composed of temporally-adjacent actions, the individual (all actions performed by one person), groups, projects or the whole system. Forming these higher-level units can be useful, as trace records themselves may have very little variability, making comparisons uninteresting, while whole life histories of users may have too much variation for comparisons to be sensible.

Linking traces in either way can pose challenges. In particular, problems can arise in assigning traces to individuals. A single individual may have multiple representations in the traces (e.g., variations in spelling of a name or multiple login IDs in different systems) or conversely, multiple individuals may share a single ID. On some systems, a subset of actions are available without logging in, so the actions associated with a particular user ID may not include every action performed by that individual.

The Jackson et al. (2016) study used sessions as an intermediate level of analysis. The intuition is that users will often interact with an online system for some period, resulting in a set of traces, then take a break (e.g., until the next day). Traces of events separated by a short period can therefore be grouped together as a single session, separated by a longer period from the next session. Sessions were useful because there were only a few different kinds of trace records (e.g., annotating or contributing to talk), and one individual's annotation looks much like another's. On the other hand, many users had contributed only one or two annotations, but a few users had contributed thousands. At the level of entire user histories, there are so many differences that it is hard to know where to start in comparing them. Describing sessions required aggregating data from multiple events that were related in time, work that had to be done programmatically.

An important consideration in aggregating data is picking which data to keep to represent the aggregation, and which details to suppress. Jackson et al. (2016) represented sessions by counts of different kinds of actions, but in so doing lost the detail about the ordering of events. In contrast, other researchers apply sequence analysis techniques to analyze order information (Keegan et al 2015). The point is that these representation questions are central to the notion of the analysis as a construction of an apparatus for seeing the world (and for not seeing some details).

Above we described the problems in linking data. Conversely, identifying an event may require parsing a complex document into smaller units. As an example of the latter, Wikipedia talk pages are the forum for discussion in the project, but it is complex to reconstruct the sequence of conversational turns in the discussion from the final page.

Having identified the various events, attention turns to understanding their meaning. As with any hermeneutics analysis, different traces may provide stronger or weaker evidence of theoretical constructs of interest. The evidence may sometimes be direct: e.g., friending someone on social media site suggests a certain attitude towards that person. Often though, the evidence may require more inference.

More importantly, technologies are often used differently than intended by the designers, so it is important to understand how the system is actually used in practice, and what the recorded system actions actually mean to users. For example, what exactly does a user mean when they click "Like" in Facebook? The actual meaning of those events and associated data can differ from a common-language understanding of the label. Geiger and Ribes (2011) call the process of learning the meaning of digital traces "inversion." Complicating things further, different users may mean different things or use a particular feature with different levels of intensity.

A major challenge in interpreting events is that trace data lack situational clues, so it takes work to establish the context of the events. To decode the meaning of a trace, it must be understood within the broader context of the platform that captures the activity. Such contextualization is a key tenet of the hermeneutic approach. It may be useful to compare across time, settings or projects or to position traces in context with other work, perhaps other activities happening at the same time. Finally, trace data may include some unstructured elements, e.g., the text of a post as opposed to the metadata about time and poster. These texts are amenable to well-known methods of qualitative analysis, though the volume may be daunting.

3.3 From Parts to Whole (Integration/Expansion)

Moving along in the hermeneutic circle the researcher now zooms out to form an interpretation of the whole from the pieces, testing a growing overall interpretation against the details. The analyst's presuppositions will likely come to the fore at this stage. As in any hermeneutic analysis, these presuppositions should be acknowledged and made explicit rather than trying to maintain a facade of tabula rosa.

In forming an integrated interpretation, it is typical to initially pick a subset of data on which to focus. For example, in literary hermeneutics, readers typically form an interpretation of a particular passage, say, for example, a chapter rather than the whole Bible. However, with trace data, picking what to read is complicated by the longitudinal nature of the data. System use and the system itself may change over time, so data from early in a project may not be comparable to data from later phases. As well, if participation is skewed, as is often the case in online communities' data, a random sample of users or events will likely include many users who are not very active and may miss particularly active users. An alternative is to group events by user and to select representative users of different levels of activity. Jackson et al. (2016) selected a handful of exemplary users for each of the six clusters they identified and described the behavioural characteristics of users and how they interacted with the system.

Conversely, rather than sampling, an analysis may address the full data set. In this case, there are interesting issues involved in interpreting complete data rather than a sample. Analysts are accustomed to treating observed behaviours as being typical of usual behaviour, that is, to inferring from a sample. For example, when observing one person interacting with another in a sample of interaction, it is typical to assume that that interaction is representative of a pattern of similar interactions. With a complete data set though, there is no need for that inference: the individuals either do or do not interact further. Having a complete record of a person's action means inference is unneeded and perhaps misleading.

Interpretation is also challenged by the temporality of the data. We noted the problems of possible system change above under sampling. In some cases, there may be a clear divide in the data, e.g., before and after a system change. There may be external events that can be included in the interpretation process, adding to the temporal components of the analysis. But often the process of change is gradual, and it is hard to draw

boundaries in a continuous process of learning, evolution, and adaptation, e.g., between novices and experienced users.

A similar process arises in interpreting overall patterns of user behaviour. As noted above, trace data often come from online systems with highly skewed distributions of activity: many individuals have few activities, and only a few are extremely active. It is clear the ends of the distribution do not represent the same kind of behaviour (e.g., heavy users are unlike those who tried the system just once), but in a continuous distribution, it is difficult to identify where to draw lines between different kinds of users.

To support interpretation of the mass of data, data visualizations can be very helpful. However, trace data can be a large volume of messy data that are hard to represent succinctly. Another approach to summarization is development of narratives. These representations can be useful in reporting the work. Jackson et al. (2016) analyzed the traces of 481 users in Planet Hunters, but only visualized the trace data for six exemplary cases that show how a user's contribution patterns change over time.

3.4 Repeating the Cycle: Reaching for Coherence and Suggestiveness

As the study repeats the movement through the hermeneutic cycle the researcher seeks coherence in the analysis and appropriate and suggestive findings that can lead to new and refined questions. This process involves addressing contradictions between the interpretations of the parts and the whole. Contradictory evidence can lead to changes in the interpretation of the pieces or the whole. A challenge here is that with a large diverse population, there are likely to be a few cases or individuals that do not fit the pattern, raising the question of how broadly an interpretation has to cohere to be acceptable.

Given the limited scope of trace data, a particularly useful step is to triangulate the interpretation with other data. Operating with multiple units of analysis allows such triangulation. For instance, Jackson et al. (2016) compared analyses made at the unit of the individual action, session, individual participant, and whole system. It can be useful to triangulate the analyses using different data sources as in any other study. Participant observation further allow researchers to compare personal experiences with the trace record. Member checking can be used similarly. Trace data may also be a helpful starting point for follow-up interviews, grounding interview questions in actual recorded behaviour. A challenge here is that is may be difficult to connect traces to a particular individual being interviewed.

Jackson et al. (2016) found interviews with experienced users to be useful to understand why users contributed as they did. In one case a cluster of sessions was described as Light Work, defined by a focus on a small number of activities. From interviews, the authors were able to associate these behaviours with users who described their work as contributing annotations and did not want to engage in other project activities

A challenge to developing an overall interpretation is that with a large diverse population, there are bound to be a few cases or individuals that do not fit the pattern, raising the question of how broadly an interpretation has to fit to be acceptable vs. needing to revise the interpretation of the parts or the wholes.

Finally, as noted above, with the cooperation of system developers, it is possible to change what the system traces, thus changing the apparatus (though within limits of what participants do on the system). These expansions might be sparked by apparent contradictions uncovered at this stage. These new data will then spark further rounds of data acquisition and interpretation.

4 Discussion and Conclusion

Facing a continuous torrent of trace data, IS researchers confront a number of methodological challenges. We have argued that researchers will be well served by cross-pollinating insights from the literatures on hermeneutics and sociomateriality. Trace data stand out not merely as *text* to be collected and dissected but

the product of an *apparatus* cobbled together by the researcher in an effort to demarcate a phenomenon of interest. Likewise, the interpretive work associated with trace data takes on a performative character. Building the apparatus associated with trace data analysis highlights the nexus of doings, materialities, and discourses researchers enact to support their investigation. There is a design element to trace analysis (Bjørn & Østerlund 2014).

Digital trace data inevitably calls our attention to their sociomaterial nature. Multiple systems, some with long histories, others created just for the purpose of the study, challenge the researcher to tinker with their frame of reference and the boundaries they enact. Seeing these issues for trace data challenges us to rethink methods more generally. In more traditional qualitative methods, e.g., interview studies, similar performances take place: traditional interview techniques embody equally sociomaterial practices. However, it has been easier to ignore the material side of an interview by fronting the participants' oral exchanges and neglecting the rest of the apparatus made up of recorders, transcription techniques, coding systems, and archival infrastructures. As we have argued for trace data, the interviewer does not simply interpret a pre-given text but rather plays a central role in the production of the text: the interview. The carefully crafted interview techniques become equally difficult to disentangle from the object of interest: the phenomenon under investigation.

A second theme in our analysis is temporality. Temporality becomes a particularly important concern in trace analysis. Trace data help track evolving activities on a citizen science site like Zooniverse. The temporal order of social phenomena is well documented in the sociomaterial literature that often focuses on the continuous entanglement of the human and the material over time (Cecez-Kecmanovic et al 2014). Equally important, the apparatus often changes over time in ways that are not always under the control of the researcher. This forces researchers to pay careful attention to the provenance of the traces and the history of the apparatus they build.

Despite our move in attention away from the interpretation of text and toward performances associated with an apparatus, hermeneutics offers a number of valuable insights central to trace data analysis. The continuous cycling between the whole and the part allows researchers to explore the materiality, temporality, and performativity of a phenomenon. As the researchers move through the cycle they may find it necessary to not only consider the relationships between parts and whole but also question the best boundary between the parts and the whole. The research process might lead the researcher to expand the apparatus by including new trace sources, adding interviews and participant observations. This allows scholars to triangulate results from multiple units of analysis by zooming in and out of multiple temporal, spatial, and sociomaterial configurations (Nicolini 2013). Other hermeneutic tenets remain equally relevant in contemporary trace analysis. Even when qualitative researchers start building elaborate databases, they still arrive at the scene with presuppositions. The research questions must be appropriate for the phenomenon. Their analytical work strives to reach a good "gestalt" free from logical contradictions. The emerging results should be suggestive by raising fertile questions that stimulate future research. Our hope is that applying the lens of hermeneutics to trace data will help researchers adopt a more self-critical perspective on their work, raising its quality and the value of its insights.

5 References

Barad, K. (2003). Posthumanist performativity: Toward an understanding of how matter comes to matter. *Signs*, *28*(3), 801-831.

Bjørn, P., & Østerlund, C. (2014). *Sociomaterial-Design:* Bounding Technologies in Practice. Switzerland: Springer.

Butler, T. (1998). Towards a hermeneutic method for interpretive research in information systems. *Journal of Information Technology*, *13*, 285-300.

Cecez-Kecmanovic, D., Galliers, R. D., Henfridsson, O., Newell, S., & Vidgen, R. (2014). The sociomateriality of information systems: current status, future directions. *MIS Quarterly*, *38*(3), 809-830.

Cole, M., & Avison, D. (2007). The potential of hermeneutics in information systems research. *European Journal of Information Systems*, *16*(6), 820-833.

Gadamer, H. G. (1975). Truth and Method, trans. *W. Glen-Dopel, London: Sheed and Ward*.

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.

Geiger, R. S., & Ribes, D. (2011). Trace ethnography: Following coordination through documentary practices. In *Proceedings of the Hawaii International Conference on System Sciences (HICSS)*. doi: 10.1109/HICSS.2011.455

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. Available from: http://search.proquest.com/docview/916253418?accountid=14214

Jackson, C., Østerlund, C., Maidel, V., Crowston, K. & Mugar, G. (2016). Which Way Did They Go? Newcomer movement through the Zooniverse. In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*.

Keegan, B. C., Lev, S., & Arazy, O. (2015). Analyzing Organizational Routines in Online Knowledge Collaborations: A Case for Sequence Analysis in CSCW. *In Proceedings of the 2016 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '16). ACM, New York, NY, USA, 257–266.*

Kvale, S., & Brinkmann, S. (2009). *Interviews: Learning the craft of qualitative research interviewing*. London: Sage.

Leonardi, P., Nardi, B., Kallinikos, J. (2012). Materiality and Organizing: Social Interaction in a Technological World. Oxford University Press, Oxford.

Lyytinen, K.J. and Klein, H. (1985) The critical theory of Jurgen Habermas as a basis for a theory of information systems. In Research Methods in Information Systems, Mumford, E., Hirschheim, R., Fitzgerald, G. and Wood-Harper, T. (eds) Amsterdam:Elsevier Science Publishers

Myers, M. D. (1995). Dialectical hermeneutics: a theoretical framework for the implementation of information systems. *Information systems journal*, *5*(1), 51-70.

Nicolini, D. (2012). *Practice theory, work, and organization: An introduction*. New York: Oxford University Press.

Orlikowski, W., Scott, S. (2008). Sociomateriality: challenging the separation of technology, work, and organization. Acad. Manage. Ann. 2(1), 433–474.

Ricoeur, P., & Thompson, J. B. (1981). *Hermeneutics and the human sciences: Essays on language, action and interpretation*. Cambridge university press.