Journal of the Association for Information Systems (2020) **21**(1), 1-22
doi: 10.17705/1jais.00590

RESEARCH ARTICLE

ISSN 1536-9323

Building an Apparatus: Refractive, Reflective, and Diffractive Readings of Trace Data

Carsten Østerlund[1], Kevin Crowston[2], Corey Jackson[3]

[1]Syracuse University, USA, costerlu@syr.edu
[2]Syracuse University, USA, crowston@syr.edu
[3]University of California, Berkeley, USA, coreyjackson@ischool.berkeley.edu

Abstract

We propose a set of methodological principles and strategies for the use of trace data, i.e., data capturing performances carried out on or via information systems, often at a fine level of detail. Trace data comes with a number of methodological and theoretical challenges associated with the inseparable nature of the social and material. Drawing on Haraway and Barad's distinctions among refraction, reflection, and diffraction, we compare three approaches to trace data analysis. We argue that a diffractive methodology allows us to explore how trace data are not given but created through the construction of a research apparatus to study trace data. By focusing on the diffractive ways in which traces ripple through an apparatus, it is possible to explore some of the taken-for-granted, invisible dynamics of sociomateriality. Equally important, this approach allows us to describe what distinctions emerge and when, within entwined phenomena in the research process. Empirically, we illustrate the guiding methodological principles and strategies by analyzing trace data from Gravity Spy, a crowdsourced citizen science project on Zooniverse.org. We conclude by suggesting that a diffractive methodology helps us draw together quantitative and qualitative research practices in new and productive ways that allow us to study and design for the entwined and dynamic sociomaterial practices found in contemporary organizations.

Keywords: Sociomaterial, Diffractive Methodology, Citizen Science, Learning, Trace Data

Dubravka Cecez-Kecmanovic was the accepting senior editor. This research article was submitted on April 18, 2017 and underwent two revisions.

1 Introduction

Information systems have become pervasive platforms for work and life, capturing data about organizational and everyday practices at a fine level of detail (Abbasi, Sarker, & Chiang, 2016; Chen, Chiang, & Storey, 2012, Zuboff 2019). As they are used, systems capture what has been referred to as digital trace data, defined 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" (Howison, Wiggins, & Crowston, 2011, p. 769). As opposed to other forms of data commonly used in information systems research (e.g., surveys and interviews, summary data or post hoc reflections), trace data are generated through routine system usage, and thus track events as they unfold over time. In this way, information systems may serve as research apparatuses, instrumenting and capturing data about a wide range of performances. And like all advances in instrumentation, trace data open new areas of study with vast potential for discovery.

At the same time, trace data raise a number of methodological challenges. First, utilizing trace data demands a deeper exploration of not only the social but also the material performances that go into their production. It is impossible to untangle the data from the technical nature of the information infrastructures capturing the traces (Hanseth & Lyytinen, 2010). Trace data are typically "big data", with high variety, volume, and velocity that pose challenges to analysis. Often heterogeneous and with fine levels of granularity, trace data can include transaction logs, version histories, institutional records, conversation transcripts, and source code, to give a few examples. Trace data tend to be semistructured: a mix of structured metadata fields (e.g., a post in a discussion forum may include the date and time, the ID of the poster, the name of the forum, a previous message being replied to, ratings by other readers, etc.) and possibly additional unstructured data (e.g., the subject or content of the post). Equally important, trace data can rarely be accepted as evidence that is ready for analysis. Researchers tend to put significant time into preparing trace data before they can dive into a deeper investigation. Trace data are created, not given.

Second, one finds a number of different theoretical approaches to trace data, spanning from positivist to interpretive-oriented methodologies. In the big data debate, many scholars approach trace data as a "lens" into organizational life (e.g., Aiden & Michel, 2014). For example, a number of studies have used posts on discussion fora as trace data of user participation (e.g., Goggins, Galyen, & Laffey, 2010; Yoo, 2010; Phang, Kankanhalli, & Sabherwal, 2009). These studies emphasize how the traces offer a lens to user behaviors and not how they are created or co-constituted. At the interpretive end of the spectrum, we find, for example, trace ethnography seeking to draw qualitative insights into the interactions of users. In this and related approaches, trace data allow researchers to reactively reconstruct specific actions at a fine level of granularity (Geiger & Ribes, 2011; Whelan et al., 2016; Loukissas, 2017). Once decoded, traces can be assembled into rich narratives of interactions associated with coordination practices, situated routines, or other organizational phenomena. But again, we find an emphasis on how traces reflect interactions and not so much on the production of trace data and its methodological implications.

Third, the lively information system (IS) sociomateriality debate offers a promising perspective with its attention to the entwined nature of the social and technical (Orlikowski & Scott, 2008; Cecez-Kecmanovic et al., 2014). Despite its relevance, the existing literature provides little methodological guidance for quantitative- and qualitative-oriented trace studies. As highlighted by Cecez-Kecmanovic et al. (2014), the IS field still needs to articulate methodologies illuminating the flow of social and material entanglements, specifically in ways that do not assume the existence of pregiven social and technical entities or that rely solely on social actors to account for how technologies act in complex assemblages. This methodological charge leaves us with a conundrum. If we assume the social and material to be ontologically inseparable, how do we make distinctions? Where in the research process do distinctions emerge?

We address these challenges by developing a set of guiding methodological principles and strategies for trace data studies. Drawing on the notion of apparatus, as well as Haraway's (1991, 1997) and Barad's (2003, 2007) distinctions among refraction, reflection and diffraction, we argue that trace data studies involve the building of an apparatus. Barad (2007) defines an apparatus as "the material conditions of possibility and impossibility of mattering; they enact what matters and what are excluded from mattering" (p. 148). As one constructs an apparatus, the phenomenon of interest emerges, which allows exploration of the boundaries and central distinctions of the phenomenon. These distinctions, or cuts, matter because traces diffract through the apparatus. For instance, when a participant contributes to a crowdsourcing site, such as Wikipedia or a citizen science project, their work is not simply reflected back to them on the screen. Instead, their activities diffract through the system in different ways. Some entries may get structured as visible articles or discussion posts while other practices end up as less visible traces in the apparatus. These performances matter in different ways.

Our sociomaterially informed trace methodology offers a number of benefits. First, a focus on the apparatus and the way it enacts boundaries and distinctions in a phenomenon allows us to understand when in the research process distinctions emerge. We can insist that the social and material are ontologically inseparable but study how distinctions materialize as one builds an apparatus and explores the multiple patterns that emerge as traces ripple through the apparatus. Second, our trace method integrates quantitative and qualitative techniques that previously flourished in different scholarly communities. Finally, our emphasis on the apparatus, its construction and performance bring the methodology into dialog with design studies (Hanseth & Lyytinen, 2010; Bjørn & Østerlund, 2014).

This essay is organized as follows: We introduce our diffractive methodology for trace data and show how it fits into the existing sociomateriality debate and positivist- and interpretivist-oriented methodologies. We then develop our methodological guidelines by illustrating how refractive, reflective, and diffractive methodologies would approach the study of learning among newcomers in a large online citizen-science project. Finally, we discuss the guidelines and note avenues for future research.

2 Theory

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, Berente, Lyytinen, & Yoo, 2014). The recent sociomaterial turn shines a spotlight on these relationships (Orlikowski & Scott, 2008; Cecez-Kecmanovic, et al., 2014; Kautz & Jensen, 2013). Within this broader debate (Jones, 2014), we take our point of departure in the position that the social and material are ontologically inseparable (Orlikowski & Scott, 2008; Orlikowski 2010, 2012; Scott & Orlikowski 2014; Beane & Orlikowski, 2015). The world does not come divided into pregiven substances carrying self-sufficient properties that we as individuated subjects can observe from the outside. Traces do not reflect people or things with inherent characteristics. Instead, we have to look to *relations, practices,* and *performances* if we hope to understand the processes through which people and things gain their qualities and identities.

Relations constitute the world, including traces. It is through relations that people and things gain their properties. Their form, attributes, and capabilities emerge through practice. Like points or lines in a geometric space, subjects and objects derive their significance from the relations that link them, rather than from some intrinsic features of individual elements (Swartz, 1997). Thus, traces do not come with pregiven qualities, properties and identities that are either purely social or material; they emerge through practice.

Practices of all stripes constitute the fundamental building blocks of reality. Rather than seeing the world as made up of predefined substances external to one another, this approach grasps the world as brought into being through everyday activities. Practices produce and reproduce reality, make distinctions, and draw boundaries (Østerlund & Carlile, 2005; Feldman & Orlikowski, 2011). Trace data are no different. They are produced and reproduced through organizational practices and, in the process, delineate the activities of, for example, employees, information systems, or artificial intelligence.

Trace data are *performative*. Not merely records of performance, they also contribute to the constitution of the reality that they trace (Callon, 1998). Organizational members use traces to coordinate and render accountable many of their activities. In crowd systems, e.g., Wikipedia, Facebook, and many citizen science projects, traces left through prior performances compose the organization. The pictures and posts shared with family, friends, crowds, and "algorithmic configurations" (Callon & Muniesa, 2005) on social media co-constitute those very networks.

Grounding our approach to trace data in an ontology of *inseparability*, which highlights the primacy of relations, practices, and performances, does not in and of itself solve our conundrum about how distinctions emerge. How do we know on what relations, performances, and practices to focus?

Heidegger (1949) proposed an early answer to this question with his phenomenological and hermeneutic approach focusing on our conscious subjective experiences and reflections to explain distinctions. In his answer, which has inspired many interpretive scholars since, Heidegger rejects the separateness of human and material entities from an ontological perspective by inverting the primacy of reflection over practical engagement (Riemer & Johnston, 2017). We might believe that we experience the world in dualist terms as a disembodied ego viewing an independent world made up of pregiven objects but, for the most part, Heidegger argues, we are absorbed in practices in a nondeliberative way that does not separate our self from other materials or beings. In other words, the equipment involved in practice is not a collection of self-sufficient entities; rather, they draw their being from a chain of practical involvement. We do not draw our recognition of an object from its properties; rather, we understand its properties based on our practical engagement with it, as something *for* something (Reimer & Johnston, 2017, p. 1066). A computer is truly encountered only when it is not experienced, when we are absorbed in a practice.

Through reflection, it is possible to experience the world as though we have stepped outside of it. But such reflections are grounded in our life-words: holistic, material, social, and embody practices that go largely unnoticed in our day-to-day life. To make any kind of distinction requires a background experience of being-in-the-world (Reimer & Johnston, 2017, p. 1063). However, through a hermeneutic process, we can separate entities out of a larger whole and reflect on their roles and properties. In other words, to understand traces, we would have to step outside of our practice and reflect on the role of these traces from the point of view of our position in a particular life-world.

But, why pay so much attention to our human ability to reflect and make distinctions, if we hope to understand the distinctions performed by highly technical trace data? The recent posthumanist literature in science and technology and feminist studies address this issue through a different take on ontological inseparability, one that emphasizes the role of materials and apparatuses to explain how cuts emerge in the research process.

Barad (2003, 2007) articulates such a posthumanist agenda by shifting the focus from the human as a reflexive being to the role of an apparatus in defining a phenomenon. In doing so she attempts to "meet the universe halfway" by neither assuming a pregiven world out there for us to observe nor relying on social actors to account for entanglements and possible distinctions (Barad, 2007). The apparatus sits in between and negates the dichotomy between the world and the human observer. In Barad's words, apparatuses are "the material conditions of possibility and impossibility of mattering; they enact what matters and what is excluded from mattering" (Barad, 2007, p. 148).

The concepts of apparatus and agential cuts allow Barad to explain the emergence of distinctions associated with phenomena. Here, the apparatus implies not a mere observing instrument but rather boundary-drawing practices that define a phenomenon. Apparatuses perform "agential cuts," i.e., marking particular distinctions, boundaries and properties within a phenomenon in practice (Orlikowski, 2010). The properties and boundaries associated with a phenomenon are not ontologically prior but become determinate and meaningful only in relation to the specificity of an apparatus. But apparatuses and their agential cuts do more than make distinctions; they can enact causal structures among components of a phenomenon by marking "measuring agencies" ("effects") by the "measured objects" ("cause") (Barad, 2007, p. 140).

Traces play an integral role by taking part in the performance of particular cuts in a phenomenon. Traces enact what matters and what is excluded from mattering. Accordingly, trace data are neither purely social nor material, neither a pregiven part of phenomena nor the apparatus tracing it. Through ongoing sociomaterial performances that produce distinctions and effect, trace data gain their properties and attributes.

In summary, to build an IS trace methodology on a sociomaterial foundation requires increased attention to how distinctions and boundaries emerge out of a particular apparatus associated with a specific phenomenon. Instead of approaching traces from a phenomenological and hermeneutic position emphasizing human reflection, Barad's agential realism allows us to explore traces as part of an apparatus that performs agential cuts and bounds phenomena. Inspired by the way Barad reads work by quantum physics and STS scholars *through* one another, we will attempt the same—reading the IS methodology literature through Barad's diffractive approach to the research apparatus. In other words, we do not intend to provide a true replica of Barad's work but rather take key insights from her thinking to illuminate issues associated with trace data.

2.1 Apparatus: Refraction, Reflection, and Diffraction

To explore methodological possibilities, we draw on three metaphors introduced by Haraway (1997) and extended by Barad (2007): refraction, reflection, and diffraction. All three are optical phenomena. Yet, the first two can be explained using geometrical optics, where, e.g., a lens or mirror mimics an object. Refraction and reflection reproduce "the same elsewhere" and often serve as metaphors for scientific objectivity. In contrast, Haraway argues that "diffraction does not produce 'the same' displaced, as refraction and reflection do," (Haraway, 1997, p. 273). Diffraction is an example of physical optics that records the patterns of differences caused by the movement of light through a prism or screen. In other words, where refraction and reflection bracket the nature of light, diffraction can be used to study both the nature of light and the source of the light. It can tell you about an object and its traces at the same time. Our discussion of these approaches is summarized in Table 1.

Refraction describes light's change in direction as it passes through the boundary of a medium; it is the explanation for the optical properties of lenses. While Haraway (1997) and Barad (2007) mention refraction only in passing, grouping it with reflection, we note that a commonly applied metaphor in social science for trace data is that of a "lens" (e.g., Aiden & Michel, 2014) through which researchers can see what is happening in the world in great detail (see Figure 1). We find this metaphor useful to describe a positivist-leaning view of data, or what Orlikowski and Scott (2008) refer to in IS as "Research Stream 1."

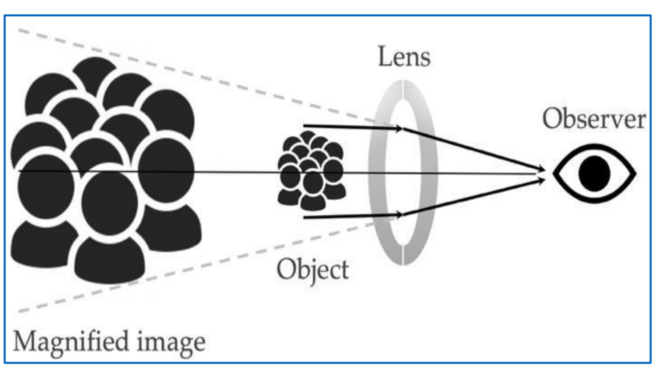

Figure 1. Refraction

Scholars with this bent strive to accurately observe physical reality as discrete entities in their data. From this perspective, trace data produced by an information system are akin to pregiven objects magnified by a microscope ; the lens in the microscope does nothing to the light passing through except to provide an enlarged view. Observing substances through a lens, we assume that these substances are pregiven, with clear and predefined boundaries. Objectivity is associated with methodological practices that produce homologous copies of the original entities, free of distortion.

Reflection is a representation of an object produced by a mirror (see Figure 2). When looking in a mirror we no longer look directly at objects, but rather at a representation of them in the mirror. Furthermore, a mirror may capture only a partial image of a broader context or an image with distortions that need to be accounted for. We find this metaphor useful when describing the methodological approaches of interpretivist and critical scholars (Orlikowski and Scott's Research Stream 2), who argue that knowledge is best understood as a reflection of mutually dependent ensembles. Interactions in these ensembles produce distortions that blur the reflections researchers can produce. Objectivity from this position is still about pregiven substances but recognizes that the image is partial or blurred and thus in need of interpretation—indeed, "reflection" undertaken by the researcher—to discern its meaning. The mirror effect emphasizes the importance of the researcher's position in relation to the object of study. For instance, by going through a process of triangulation, the researcher may examine reflections from different positions.

To Barad, reflection serves as a particularly apt metaphor for science and technology scholars applying interpretive and reflexive methodologies. Even practice-oriented scholars taking a relational view of reality often fall into a reflective view of the word that displaces "the same" elsewhere. They might not argue that their interpretations produce mirror images, but by arguing that their interpretations of objects build on a subject's social position, background knowledge, or life-world, they end up reflecting not objects but pregiven social and cultural categories through their methods. In other words, by giving the human and its reflections such a prominent role, the methodology turns a blind eye to the role of materiality—in particular, the materiality of the traces and the differences they make. Observers end up reflecting their pregiven social and humanistic categories back onto the world.

In both refraction and reflection, though, it is the image's likeness to the substance that matters, not the nature of the light producing the image or the apparatus of observation, i.e., the lens or the mirror. Empirical entities are seen as pregiven, what Haraway (1992) described as "'the same' displaced." Both cases hold the world at a distance (Barad, 2007). To put it differently, a refractive or reflective approach supports what Cecez-Kecmanovic (2016) describes as a substantialist metaphysics concerned with "what there is." Only if one envisions the primary unit of reality as self-contained and bounded substances can one adopt a refractive or reflective methodology.

Diffraction concerns, in contrast, the bending and spreading of waves when they combine or meet an obstacle. Light and sound both exhibit diffraction under the right circumstances. Figure 3 depicts a classic example of diffraction in physics. In this experimental setup, light from a source on the left of the figure passes through two slits in the barrier in the middle of the figure and the beams of light from the two slits interfere with each other, leaving a diffraction pattern of light and dark on the screen beyond the slits to the right of the figure. This pattern does not appear if the light shines directly on the screen or if there is only one slit. Thus, the diffraction pattern records not only differences in the source waves, but their history and interferences encountered on the way to the screen. The metaphor offers a process perspective concerned with "what is occurring" and "ways of occurring" (Cecez-Kecmanovic, 2016). The primary unit of interest is not an image reflected on to a screen, but the processes of configuring meaning and matter.

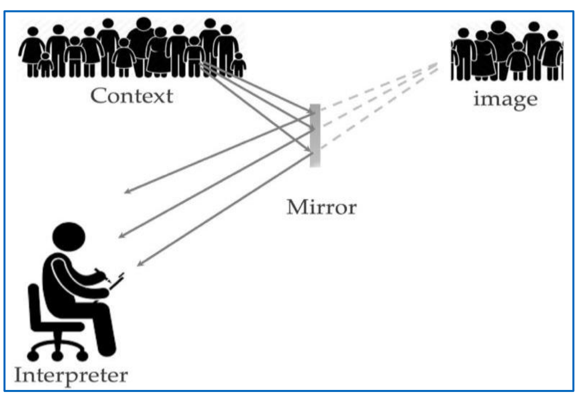

Figure 2. Reflection

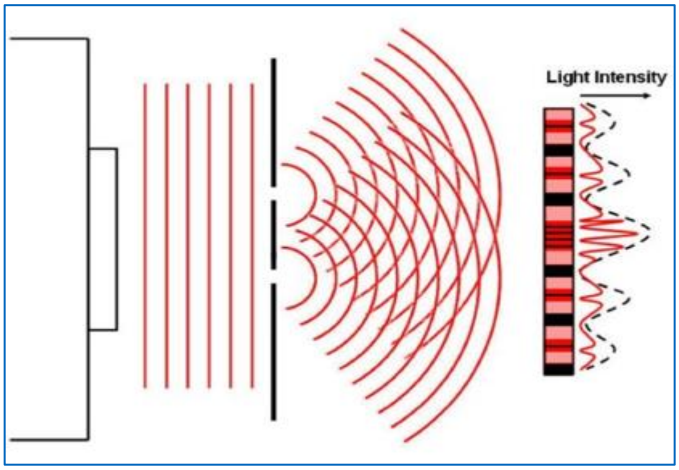

Figure 3. Diffraction Pattern of Light from a Two-Slit Experimental Setup

The apparatus takes on a central position in a diffractive methodology. Barad argues that one cannot disentangle a phenomenon and the apparatus that performs it. Instead, the apparatus plays a constitutive role in the production of the phenomenon by enacting specific boundaries in our sociomaterial reality. That is, online systems do more than record traces of human actions and interaction: they actively shape them. The apparatus is not a simple inscription device installed before the action happens nor is it a neutral probe, measuring preexisting entities, mere reflections of a self-contained reality. Instead, the apparatus stands out as an open-ended practice constantly producing and reproducing the phenomenon that it records.

As a result, a diffractive methodology offers an analytical approach in which one reads elements of the research setup through other elements by following the multiple patterns traces form as they ripple through the apparatus. It allows us to trace different practices and examine the distinctions they make. This *reading through* is possible because the elements are intertwined; changing the size, number, or position of the slits or the nature of the light source in Figure 3 causes the diffraction pattern to take on a new shape. By studying changes in diffractive patterns, researchers learn about the nature of the light source and the nature of the apparatus the light encounters (e.g., the slits). For example, physicists can study the nature of a chemical element by sending light from that element through a diffraction grating with known properties and observing the resulting diffraction pattern. Reading through can also work in the reverse direction: physicists can study the diffraction grating itself by illuminating it with light with known properties. For instance, one can learn about a crystal used as a diffraction grating by sending an X-ray of a known wavelength through it and studying the resulting diffraction pattern. Following the same line of thinking, information systems researchers can learn about trace data through studying the users of an online system, learn about users through studying their information system, or learn about an information system through studying its traces.

Further, the performances of an apparatus are open to rearrangements. The creativity of scientific practices includes the skill of making the apparatus work for specific purposes. Elements are reworked and adjusted, leading to adjustments of the boundaries and cuts performed by the apparatus and thus the nature of the phenomenon enacted and recorded. An apparatus can itself become the phenomenon, the focus of attention. This shift 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 short, from a *refractive* methodology, trace data serve as a lens projecting images of pregiven objects with sharp boundaries. A *reflective* position mirrors the world, leading to an interpretive stance that deals with trace data as distorted or incomplete reflections of pregiven objects that need interpretation to determine their meaning. In contrast, a *diffractive* methodology emphasizes the apparatus and sees it as constitutionally entwined with the phenomena under study. The apparatus enacts cuts around and within the phenomena and thus is part of the making of boundaries and distinctions that we as researchers apply in our empirical descriptions. Differences emerge in a diffractive methodology but without absolute separation. Trace data diffract through the apparatus as ripples and waves and, in the process, they co-configure the apparatus and phenomena. Traces are thus not given but created. They open a window into both the phenomena and the apparatus by allowing researchers to read them through one another.

Table 1. Refractive, Reflective and Diffractive Approaches

	Refraction	Reflection	Diffraction
	Positivist research stream 1	**Interpretivist research stream 2**	**Sociomaterial research stream 3**
Phenomena (ontological priority)	Discrete entities with clear properties that may interact with one another through causal relationships	Mutually dependent ensembles with emerging properties that coevolve over time	Sociomaterial assemblages with no inherent properties that acquire form and features through interpenetration with an apparatus
Metaphor for the apparatus	Lens (shows objects directly)	Mirror (Shows objects but indirectly)	Diffraction (Enacts cuts around and within phenomena)
Objectivity	About refractions, copies that are homologous to originals, authentic, free of distortion	About reflections, images that may be incomplete or blurred	About diffractive patterns that mark differences and relations that matter. Subjects and objects do not preexist but emerge through practice
Boundaries and distinctions	Pregiven and sharp	Pregiven but fuzzy	Emergent, performed, and fuzzy
Traces	True depiction of the world. Image of pregiven objects; Measure specific features of objects	Distorted and incomplete reflection of pregiven objects that need to be interpreted to determine meaning	Waves and ripples that diffract through the apparatus and in the process co-configure the apparatus and phenomena. Traces are not given but created. Allows one to read the phenomena and apparatus through one another

3 Case Example: Learning in Citizen Science

To illustrate the three different approaches outlined above, we present examples from an ongoing study of learning in an online citizen science project, Gravity Spy, that was based, in large part, on trace data, thus providing examples of the issues discussed above. Citizen science is a broad term describing scientific projects relying on contributions from members of the general public (i.e., citizens in the broadest sense of the term) who volunteer time and effort to advance the goals of the project. There are several kinds of citizen science projects: some have volunteers collect data, while others, including the one we examine here, ask volunteers to analyze already-collected data. Increasingly, the work of volunteers and project organizers take place via the web, e.g., on a site that presents data to be analyzed and collects volunteers' annotations (e.g., www.zooniverse.org). Their work is sometimes described as "crowdsourcing science" and so is relevant to IS researchers. Moreover, citizen science projects are an intriguing example of distributed learning and knowledge production, supported by public engagement in scientific research processes. To be effective over time, the projects must facilitate ways for new users to orient themselves towards the goals and work practice of the project.

How newcomers to a crowd learn to be effective participants thus stands out as a critical issue (Van Maanen & Schein, 1977; Ostroff & Kozlowski, 1992; Klein & Weaver, 2000). In some groups, new members go through formal educational or orientation activities in order to learn group practices, while others rely on informal orientations. Online groups, in particular, often face difficulties with newcomer orientation, as many online groups are composed of members who are not part of a single formal organization and who contribute only in their free time, reducing or eliminating the possibility of formal training. However, technology-supported group interaction makes it possible for distributed volunteers to observe work in progress, thus enabling a form of legitimate peripheral participation (Antin & Cheshire, 2010; Bryant, Forte, & Bruckman, 2012; Halfaker, Keyes, & Taraborelli, 2012).

We draw our examples specifically from the Gravity Spy (Zevin, et al., 2017) citizen science project (http://gravityspy.org/), which is built on the Zooniverse.org citizen science platform. The Gravity Spy project was developed to support the Laser Interferometer Gravitational-Wave Observatory (LIGO). LIGO comprises two detectors that measure minute changes in distance caused by the gravitational waves bending space-time as they travel through it. However, the sensitivity that enables LIGO to detect distant astrophysical events also makes it very susceptible to non-astrophysical instrumental and environmental noise, referred to as “glitches.” Glitches hamper the detection of gravitational wave events, either by blocking events outright or by increasing the number of potential events to be examined. At LIGO’s current sensitivity, detectable astrophysical events are expected to occur only about once a month, while a glitch may occur every few seconds, making a search for true events akin to finding a needle in a haystack.

Similar glitches may have a common cause that can be eliminated if it can be identified; therefore, finding and classifying glitches stand out as core tasks for improving the LIGO detectors. However, with thousands of glitches, the LIGO researchers do not have the manpower to examine them all. Relying on computers alone has thus far also fallen short, as the diversity of glitches defies easy attempts at classification. At present, there are 22 known categories of glitches, but many glitches do not fit any of these categories and so may be examples of as-yet-unidentified classes of glitches. Presently, humans are much better at the visual processing needed to identify similar types of glitches. Given these concerns, the project has developed a citizen science approach to classifying glitches.

When using a citizen science platform such as Zooniverse, volunteers are presented with images and asked to classify them into one of the known categories. Gravity Spy also provides options of *none of the above* or *no image* for images that do not include an event of interest at all. The Gravity Spy system is shown in Figure 4: an image of a glitch to be classified is shown on the left as a spectrograph, with time on the x-axis, frequency on the y-axis and intensity represented as colors ranging from blue to yellow. Possible classes are shown on the right. The initial learning challenge for new volunteers is how to identify the appropriate class for a glitch by matching it to one of the given exemplars. An innovation in this system is that a machine learning (ML) algorithm has been trained to distinguish glitches, and the ML classifications are used to pick images with which to train new volunteers.

The Zooniverse system is instrumented to record several kinds of data. The classification dataset contains the classifications users contributed to the project. Included in the dataset are the glitch class chosen by the user (e.g., blip, whistle, etc.), the timestamp of the classification, and other metadata about the image, such as the image size and glitch type for images that were classified by experts (“gold standard” data). System interaction data contains events of users’ interaction with pages on the site. When a user clicks on a link to access a new page on the website, an event record is stored. In total, 83 different kinds of website events are recorded. The record also contains a timestamp showing when the resource was requested. Data were collected and linked to a user ID; they include no personally identifying or demographic data.

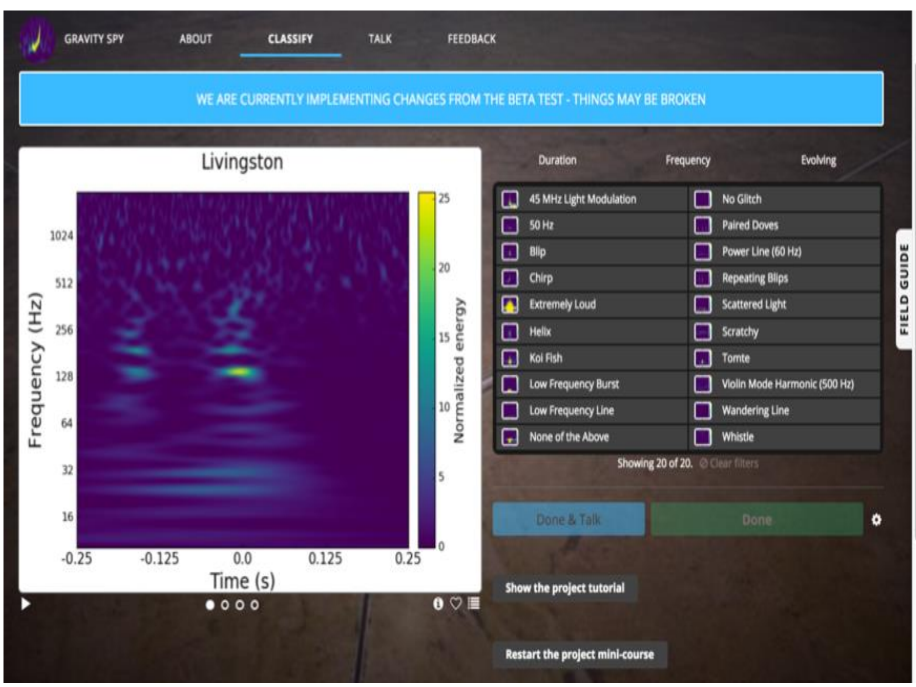

Figure 4. Full Gravity Spy Classification Interface (http://gravityspy.org/)

4 Approaches to Analyzing Trace Data for Learning

In an effort to build a set of guiding principles for a sociomaterial trace data methodology, we next present examples of how learning in Gravity Spy might be defined and studied from the three perspectives developed above. This will allow us to illustrate the assumptions going into each methodology.

4.1 Positive/Research Stream 1: Trace Data as Refraction

Investigations of learning in the tradition of Research Stream 1 consider data to be depictions of the discrete and pregiven entities in the world, such as glitches and Gravity Spy volunteers. In this view, trace data are seen as providing a lens on what volunteers are doing in the system and what and how they have learned (see Figure 1). As noted, the Zooniverse system records data as volunteers contribute to and navigate through a project. Within the system (and the trace data) these actions are well identified, as the clickstream data are discrete units based on materials predefined by the system creators. Data are stored in rows and columns in a data store, embodying a set of identified boundaries. The system defines a user by a persistent user ID and linking records with the same user ID provides a record of the user's interactions with the system. To study volunteer learning, researchers can look for evidence that volunteers' performance on the classification task improves over time (e.g., Crowston, Østerlund, & Lee, 2017), where performance is defined as the correctness of volunteers' classifications, i.e., the agreement of their choice of class with either an expert's choice or the consensus of other volunteers.

Research can further examine which system features lead to quicker or better learning (i.e., higher correctness). For example, some volunteers might have viewed the project tutorial, which describes the classification process, the science of gravitational wave research, and how the data being analyzed by volunteers came into existence. Volunteers may also consult other resources, such as the FAQs and the About page that provide additional context for the project and task, supporting volunteers' comprehension of the project and task. The system records which resources a volunteer has seen, creating for each viewing one or more records, for example the user ID, a timestamp, and other metadata. Statistical analysis of these data can test the relationships between performance and the use of resources and other volunteer-specific factors, thus suggesting which resources are most helpful for learning.

In short, a refractive approach assumes the existence of pregiven objects associated with an observer. Traces are important to the extent that they can serve as a lens to users and their behaviors. In the context of Gravity Spy, a unique user ID represents users and their classification record captures their activities. Overall, the focus lies on the object: volunteers causing some effect in the system, i.e., traces. We find an emphasis on a unidirectional relationship between the object and the observer only mediated by the traces.

4.2 Interpretivist/Research Stream 2: Trace Data as Reflection

Researchers in the tradition of Research Stream 2 assume that data, even quantitative data, do not speak for themselves, but require interpretation. The system serves as a mirror where the recorded data do not project reality, but rather reflect what happened, imperfectly with omissions and distortions (see Figure 2). Such interpretivist research lays a critical eye on trace data and their implications for understanding a phenomenon.

With this approach, the job of the researcher is to make sense of what they are seeing in the mirror of the dataset. Hermeneutics offers a well-articulated approach that has long served as a trusted pillar of qualitative and interpretive 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. In classic hermeneutics, a text constitutes an object of study that is to be understood based on its own frame of reference (Kvale & Brinkmann, 2009). Interpretation aims to bring to light an underlying coherence or sense from an otherwise incomplete, cloudy, or contradictory text (Myers, 1995). The hermeneutic cycle summarizes the basic analytic process in which a researcher repeatedly moves back and forth between the whole corpus and its parts.

From this perspective, trace data becomes a text requiring an interpretation. The need for an interpretive approach is clearest when dealing with textual traces. For example, we might be interested in how volunteers draw on posts on discussion boards (known in Gravity Spy as "talk") to support their learning (Mugar et al., 2014). Just counting posts (as described in Stream 1) is unlikely to be satisfactory. Some posts might have more relevance for learning than others. Instead, the researchers would read and reread messages to form an interpretation of the kinds of messages and their function and then test that growing understanding against a larger set of messages and the overall context of volunteer learning. For example, research could examine how a volunteer calls attention to some feature of a glitch and how other volunteers respond, building a theory of communal learning (e.g., Mugar et al., 2014). Such an analysis might also lead to a redefinition of learning, e.g., moving from a focus on accuracy to consideration of how volunteers engage with scientific practice. In this case, the hermeneutic approach is applied much as in any qualitative study.

While the need for interpretation is clear for qualitative data, we note that an interpretivist approach can help discern the meaning of quantitative trace data taken from an online system. At the most basic level, the researcher needs to understand the mapping of actions that volunteers can take on the system to the data that are recorded in the traces. While data may have labels (e.g., in a database dump), the connection between that label and an action is not always straightforward.

Further, to understand the import of data about user actions requires understanding the purpose and meaning of the captured interactions in the overall context of a volunteer's engagement with the system. Because technologies are often used differently than intended by the designers, it is important to recognize how volunteers enact the system in practice and what the recorded system actions mean to volunteers. For example, in Gravity Spy, what the system records about interactions are the specific links that a volunteer clicks on the web page. To understand the meaning of clicks, we must form interpretations of this action in terms of user behaviors. For example, the system might record that volunteers clicked on the link for a discussion board. However, we do not know for sure that the volunteers actually read a particular post on the discussion board. It might be that the volunteer navigated to the board, intending to create a new post rather than read. Complicating things further, different volunteers mean different things by their use, or use a feature with different levels of intensity. And yet, to assign meaning to the trace data, these nuances must be understood.

A key point of a hermeneutic approach is that to decode the meaning of a trace, it must be understood within the broader context of the work being done. However, trace data often lack situational clues; thus, it takes work to establish the context of the events. 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.

In summary, an interpretive approach operates with pregiven categories that are reflected through the information system in the form of traces. These pregiven categories do not have to be well-defined objects; rather, they can reflect social and cultural classifications or practices that interact with the information system and coevolve over time. Researchers have only a partial view of the broader context. and it requires interpretation to discern how the reflected traces fit into this larger phenomenon. The position of the researchers becomes important as does the hermeneutic process through which they compare partial views of one another and a larger context.

4.3 Sociomaterial/Stream 3: Trace Data as Diffraction

Finally, developing an understanding of learning through a diffractive methodology by following Stream 3 goes hand in hand with building an apparatus and exploring how practices ripple through the system. Investigating the apparatus cannot be separated from an exploration of the phenomena. In asking the question, "What is learning?", we notice the two sides to the question: "*what* is learning" and "what is *learning*." Both sides come into play as we build an apparatus.

4.3.1 Demarcate the Phenomenon and Apparatus

From a diffractive perspective, we turn our attention to the apparatus by exploring its boundaries and intra-actions with the phenomena. As noted above, refraction and reflection take the objects comprising the phenomenon as given—for example volunteers, glitches, and classifications. However, a diffractive reading helps us realize that these objects emerge out of the performances going into the apparatus. Here, we provide several examples: First, as researchers, we tend to assume that volunteers exist and thus look for them in our data (i.e., traces linked by a common user ID), but it is the distinctions and boundaries enacted by the apparatus that call them into play. Second, glitches are created in the preprocessing of data obtained from LIGO. Whether a particular piece of signal is considered a glitch or not depends on whether it passes an arbitrary signal-strength threshold; decreasing that threshold creates more glitches to be added to the system. The spectrograph displayed in the system is also created as part of the preprocessing, and the appearance of the image depends on a number of parameters, which can be varied. Finally, correctness of a classification, a key variable in a study of learning, is determined by comparing a volunteer's classification against the "correct" answer for a glitch. For most glitches though, "correct" is taken as the consensus of volunteer classifications, meaning it is itself a product of the system. In the absence of consensus, correctness cannot be determined. A few glitches have classification given by LIGO experts ("gold standard data"), but classification is a practice, and even these expert decisions are occasionally called into question. In summary, upon closer inspection, the sharp distinctions assumed in the refractive and reflective analyses discussed above turn out to be entwined with the apparatus.

Looking at boundaries more broadly as a citizen science project, Gravity Spy plays a role in a much larger apparatus. It includes detectors with four-kilometer-long arms in Washington state and Louisiana, recently joined by a third smaller detector in Italy named VIRGO. Hundreds of researchers across the world actively work on these instruments and, in the process, apply large IT infrastructures to store and analyze the data produced. Gravity Spy, with its tens of thousands of citizen scientists, constitutes just a small part of this larger effort. But Gravity Spy is hosted on Zooniverse, a citizen science platform with more than 80 active projects and millions of volunteers. Where does the apparatus stop? Should our apparatus account for the machine learning unit built into Gravity Spy? Or should we simply demarcate the apparatus as our locally stored and curated database of Gravity Spy trace data?

Our answers to these questions and thus how we demarcate the apparatus have consequences for the phenomena: namely, learning. Accounting for the detectors and their international research team suggests learning processes that go beyond volunteers' rather limited activities. The entire LIGO apparatus points us towards large-scale societal knowledge production and teaches us how research communities learn about the universe and its fundamental processes. This type of learning clearly motivates many volunteers, who eagerly search out additional readings about gravitational waves and the instruments, capable of detecting changes in space-time of about 10^{-19} meter, or less than one-thousandth the diameter of a proton. The larger apparatus would lend itself to a conception of learning that fits into science and technology studies or the 90s debates about organizational learning (Suchman, 2007; March, 1991).

Limiting our view to Gravity Spy work would allow us to define learning more narrowly around the volunteers' activities on the system. However, restricting our apparatus to Gravity Spy alone is easier said than done, as boundaries remain fuzzy. Gravity Spy volunteers look at glitches produced by the detectors and interact with LIGO researchers in the discussion boards, but they also interact outside of the system and its traces, e.g., by reading LIGO blog posts. Given that Gravity Spy is part of the Zooniverse platform, many of the volunteers participate in multiple projects spanning the fields of history, biology, medicine, and astronomy. Despite our best intentions, bounding the phenomena and apparatus will always be a work in progress; claiming otherwise would require us to turn a blind eye to important performances.

4.3.2 Investigate the Apparatus

Working with a particular apparatus involves an ongoing investigation of its performances starting with the question: *What does the apparatus trace?* And, *what does the apparatus not trace?* While it is tempting to expect that the system captures traces of all events, data storage is itself a practice, and the assumption of completeness must be carefully examined. Activities of interest may be unavailable for study. For example, the Zooniverse platform primarily supports science tasks. When we first began our study, it recorded only the annotations done and not activities such as volunteers' tutorial use, which the designers did not consider to be data.

Other important activities might take place outside the apparatus. Trace data does not capture the work done by volunteers drawing on non-Zooniverse servers. For instance, one volunteer created a web scraper to quickly capture the images without having to go through the regular annotation procedure. The software crawled the Gravity Spy site by generating a URL based on the subject-ID naming conventions Zooniverse uses for images on the server. The volunteer would then visually inspect the retrieved images to see if they fit the category he was interested in collecting. Other volunteers sometimes provide the URL of external resources (e.g., academic papers, notebooks detailing alterations to the instrumentation at the detector sites) in a post, demonstrating that they are actively seeking additional knowledge. However, there is no systematic trace data record of when they do so or how those resources are used.

Finally, one should keep in mind that systems are subject to many problems that result in data loss (e.g., server outages, disk failures, deleted log files, or truncated database tables), meaning that trace data—even from database dumps—can be incomplete, though the problems may not be immediately visible (Howison et al., 2011). To address these problems, the researcher should develop a detailed understanding of the apparatus. From a learning perspective, it makes a big difference whether one has access to annotation work only or a range of other activities, such as discussions among volunteers or external resources people might utilize to support their work on Gravity Spy.

Not only does the apparatus include and exclude certain practices in the traces produced, it also performs certain cuts. These distinctions play an essential role in demarcating key categories. For instance, we discussed above what encompasses the "learner" in Gravity Spy trace data. We assume in our analyses that a user ID represents an individual, but it is not inconceivable for groups to utilize a single user ID, such as a group of students working on Gravity Spy in their physics class or a family engaged with the project after dinner. In contrast, participants may have multiple user IDs or work anonymously on the system without logging in, which means that they can have significant experience with the system that the trace data does not capture. Again, the apparatus does not draw sharp distinctions and therefore requires additional work if one hopes to define an individual within the trace data. Similar questions may be asked about other categories and practices central to learning, such as what constitutes a science team member engaging in a project, or how central is the machine learning unit to the Gravity Spy project.

The boundaries and cuts performed by the apparatus change over time. A genealogy of the apparatus helps one understand how distinctions and boundaries gradually emerge in this sociomaterial system. The Zooniverse platform started out with the Galaxy Zoo project, which initially included only an annotation system. Volunteers were presented with an image to annotate and, to avoid groupthink, they had to perform their own assessment before being able to access other participants' work on the same image. Soon after, a discussion board feature was added (originally a stand-alone open-source discussion forum package). Gradually, user profiles, collections, and search capabilities followed. Major funding from the Sloan Foundation and later Google allowed Zooniverse to create a more integrated project-builder platform, permitting research teams to easily set up citizen science projects. Not only did all of these changes lead to alterations to the apparatus, they also mark important cuts. For instance, the current Zooniverse project makes a rather sharp distinction between annotation work and discussions; they take place in different parts of the system and their relations are carefully managed.

4.3.3 Extending the Apparatus

Performing trace analyses further changes the apparatus. In other words, the apparatus and its traces are not pregiven. Additional cuts get added as researchers work with the trace data. These changes can take many forms, including, among others, the building of trace databases, conducting statistical analyses, experimental interventions (e.g., A/B splits), and interviews.

We turn to the question of databases first. To study a phenomenon as complex as learning requires us to pull data from multiple sources, such as records of use data and other metadata. These may be stored in different databases and database tables. In our study of learning, the available traces were not sufficient to address our questions. Zooniverse gathered traces about participants' annotation of science data but little else. After months of lobbying and joint funding, we persuaded the software developers to add new trace features to the system so we would know when people had used various tools such as tutorials, science pages, collections, discussion boards, and user profiles. The expansion can be iterative: researchers cycle between appreciating the available traces and adding new traces to further flesh out and define the phenomenon.

The work doesn't end here. The newly constructed databases often leave us with a big unruly pile of traces, making it difficult to discern what differences matter. Constructing the apparatus involves further processing. For example, to understand how learning evolves over time, we divide volunteer traces into sessions (i.e., we perform additional cuts). The intuition is that volunteers will often interact with an online system for some period, creating a temporally adjacent set of traces, then take a break (e.g., until the next day). Traces of events separated by a short gap can be grouped together into a single session, separated from the next session by a longer gap. This analysis approach provides a way to bound and separate traces to a format that acknowledges the temporality of Gravity Spy performance. We selected a set of traces to comprise a session. Prior work on Wikipedia has defined a gap of one hour between activities as indicating the start of a new session (Geiger & Halfaker, 2013), but given our own experiences annotating items in Gravity Spy and observing others do the same, we chose a gap of 30 minutes for our understanding of Gravity Spy annotation work, that is, the sequence of activities separated by less than 30 minutes was considered a session.

Applying statistical packages further extends the apparatus. Each analytic technique bundles and slices the trace data in new ways, and with it the phenomenon of learning. A session might be represented by counts of different kinds of actions (e.g., classification, reading or posting to discussion boards, consulting the field guide) that contribute to learning. For example, applying computational approaches such as linear regression allows us to model learning through use of these resources. However, analyzing counts loses information about the order of events. An alternative strategy applies sequence analysis techniques that focus on the order of events (e.g., Keegan, Lev, & Arazy, 2015). Cluster analysis can also be used to identify sessions with similar patterns of activities. However, decoding these clusters requires a diffractive reading of the quantitative analysis and calls for an exploration of how traces ripple through the apparatus.

4.3.4 Diffraction: Explore How Traces Ripple Through the Apparatus

An apparatus does more than produce metadata about practices associated with its use. As depicted in Figure 3, traces ripple through the apparatus. In Gravity Spy, annotations done by volunteers feed into algorithms deciding how many other volunteers need to see the image before it is retired; it serves the user profile to help participants know how much work they have done on the project. After a volunteer has annotated a glitch, it is possible to leave a note with the particular image. As mentioned above, Zooniverse projects allow volunteers to see other volunteers' annotations and provide access to discuss traces about an object only after the user submits an annotation to avoid propagation of user biases. These restrictions to the way that traces ripple through the system make it hard for newcomers to observe and learn from more advanced volunteers' work practices. However, we find that many volunteers compensate for this lack of legitimate peripheral participation (Lave & Wenger, 1991) by spending significant time looking over experienced participants notes in the "Talk" feature. These advanced notes act as proxies for practice for less experienced participants (Mugar et al., 2014; Jackson et al., 2015). In other words, the traces do not refract or reflect users' behaviors, but instead ripple through the apparatus and feed other practices. Some of these traces ricochet back to the participants in the form of user profile stats or Talk posts.

To make sense of the activity clusters generated statistically, we follow how participants' behaviors rippled through Zooniverse. For example, we applied cluster analysis to the sessions mentioned above. One prominent cluster captured performances restricted to the annotation feature. Participants did one annotation after the other over a short time span with no traces left, suggesting use of other features. We named this type of session "light work." A less prominent but still significant cluster involved traces of activities indicating that a volunteer after each annotation would check if other people had left notes on that image. Often, they spent a long time going through these communal discussions, but rarely left any notes themselves. We named this cluster "careful annotation." Another cluster named "talking and annotating" included a lot of discussion board traces with a few detours rippling into the annotation system. From the sequencing of the traces, we discerned that in some sessions volunteers spend most of their time engaging in the discussion board or collection features, with only periodic visits to the annotation task (Jackson et al. 2016).

For each user ID, we organized these session types sequentially and found that most participants engaged in light work sessions only. More dedicated participants oscillated between light work and "heavier" sessions where they either engaged with the community through posts and discussions or spent a lot of time diving into each image and other people's annotations of those glitches. A small number of participants had sessions focusing on individual images, building collections of unusual images, and reading science notes. In short, to explore diffractive patterns, one traces paths through the apparatus. Again, it is important to vary the unit of analysis. We move between following a single trace, following clusters of traces, temporal ordering traces and sequences of sessions, and grouping participants with similar session sequences. By dialing up and down (Gaskin et al., 2014) on the size and order of trace bundles, we explore multiple performances, patterns, and learning phenomena and identify how they change over time.

More explicit design changes to the apparatus further allow one to explore what differences matter by testing multiple rippling effects. In the diffraction experiment depicted in Figure 3, for example, we can change the light source or the slits the light passes through to see how it changes the way that traces ripple and the diffractive patterns they form. Similarly, as part of our study of learning in Gravity Spy, we implemented a scaffolded progression of tasks to support newcomers' learning. Volunteers annotate glitch images into the 22 known classes of glitches, but rather than providing all classification options to new users, the system introduces them a few at a time. New volunteers start at Level 1, a simplified version of the classification interface, in which they are presented with glitches to classify that are expected to be of one of only two distinctive classes—"blips" vs. "whistles" or "none of the above." Once the volunteer can successfully classify glitches of the initial two classes (currently assessed by accuracy in classifying gold-standard data), the volunteer can advance to the next training level, in which they see glitches of additional classes. In other words, to scaffold volunteer learning, the system gradually expands the number of classes presented to the volunteers. The glitches to be presented in each level are selected by a machine learning (ML) algorithm. The ML classifies all glitches added to the system into one of the known classes, with an accompanying confidence in the classification. Glitches with a high ML confidence are given to new participants as training. Once volunteers have become experienced with more glitch classes, they are presented with images with lower- and lower-ML confidence.

To see if these differences matter, as compared to typical Zooniverse projects in which individuals access all known classes from the beginning and without ML support, we performed a simple A/B split. New participants were divided into two groups over a period of a few weeks. One group went through the scaffolded system while the second group faced all 22 known glitch classes from the beginning. Subsequent trace analysis suggests that the members of the scaffolded group contributed to the project significantly longer, mastered the task faster, and did more annotation work than the second group. During the experiment, some volunteers in the second group went back through the scaffolded levels that they had bypassed without any prompts from the system.

Recently, we have given advanced participants access to the ML processing to support their search for new glitch classes unknown to the science team. Instead of assigning images to volunteers, the advanced participants use ML to find images similar to clusters of images that they hypothesize belong to a new glitch class. In this way, we hope to learn more about machine-human learning intra-actions, and agential cuts that are significant for such performances. These dynamics cannot be explored without carefully following the ways that traces ripple through the apparatus.

Direct engagement with volunteers offers ways to explore the apparatus and its diffractive patterns. Participant observations, interviews with individuals, and focus groups help explore traces and the way that they ripple. For instance, visualizations of trace data such as the sequences of sessions described above can serve as productive interview prompts. They give the volunteers a view into the apparatus and illustrate the way that their practices ripple through the system; they also offer volunteers an opportunity to describe how these traces relate to other activities not captured by the apparatus. Such interview protocols can span a broad range of traces. We used collections of Talk posts to explore how newcomers use experienced participants' annotations as practice proxies. In other interviews, we shared highly processed trace visualizations of session sequences associated with the interviewee. The method goes beyond traditional triangulation, which tend to assume pregiven entities and test one statement against other statements about an object. One can imagine the interpreter in Figure 2 rolling their office chair around to look at the object from different positions to get a better view of it in context. Instead of relying on the reflection of pregiven entities, trace interview prompts offer ways to learn more about performances and clarify how they do and don't diffract through the apparatus.

4.3.5 Differences that Matter

The diffractive analytic process involving the demarcation of the apparatus and phenomena, exploration of the apparatus, and the way that traces ripple through it add up to a search for differences that matter. This rippling does not refer to a more traditional conception of causality as relations between distinct entities (Barad, 2007). Instead, it explores the effect of specific distinctions and boundings, i.e., agential cuts build into the apparatus. As Barad argues: "Causal relations entail a specification of the material apparatus that enacts an agential cut between determinately bounded and propertied entities within a phenomenon" (Barad, 2007, p. 176). In other words, we have to pay attention to the boundaries enacted by the apparatus in its entwined relationship with the phenomena and the distinctions it makes. Only then can we explore how traces ripple through the apparatus and understand what changes they leave in their wake.

We have found benefits in a circular analytical process where the researcher oscillates between exploring the boundaries of the apparatus/phenomenon and the way that traces ripple through the apparatus/phenomenon. Just as a hermeneutic process cycles between analyzing a whole pregiven text and its parts, we envision a circular movement through a diffractive apparatus. Studying Gravity Spy, one cannot assume that traces scrapped from the system constitute a whole. Instead, the researchers and, in many situations, the volunteers explore the boundaries of the apparatus and may add new features to the configuration. Tracking traces as they ripple through the system allows one to question the distinctions made. For instance, what constitutes learning or what demarcates a volunteer? What type of performances do they engage in and how do they change over time?

Volunteers leave traces behind them like a boat cutting a wake in its path. The traces make up part of the reality that defines the performances. What one sees in Gravity Spy is a product of one's own traces as well as the traces of other volunteers. The boat is rocked by its own wake as it plows through a canal, with each wave diffracting back to the boat after hitting the channel banks. The diffractive patterns of the waves must be read through the rocking of the boat, the structure of the embankments, and the decisions of the pilot trying to avoid spilling his morning coffee. Moreover, the researcher may change the banks of the channel or the shape of the boat to see how the wave patterns change. We should even question whether we are dealing with a captain at the helm or a middle school class supported by ML. The diffraction pattern marks differences that matter.

The dynamic intra-actions between phenomena and apparatus, i.e., boundaries and distinctions emerging as traces rippling through the system, allow us to operate with multiple learning phenomena at the same time. Each form of learning is associated with a different apparatus and co-configuration of how traces ripple through it. Inspired by Sørensen (2009), we distinguish three learning phenomena associated with Gravity Spy: authority-subject, communal, and agent-centered learning.

First, authority-subject learning emerges in an apparatus divided into clear regions and subregions, each associated with clusters of homogeneous and highly structured activities, events, and objects. One can imagine a classroom as a region divided into two subregions. The front of the room, which is occupied by the teacher and the blackboard, and the rest of the classroom inhabited by students, their desks, chairs, all positioned to face the blackboard and the teacher's subregion. The separation between students in their chairs and the teacher at the blackboard thus marks two distinct regions, each associated with particular activities. The tutorial pages and training modules work much like the teacher's subregion, pushing authoritative knowledge from the expert science team to the volunteers' subregion in the annotation system. As in a classroom, the annotation subregion of Gravity Spy constitutes a highly structured environment where volunteers are asked to review one image after another. From time to time, they review a gold-standard image and receive feedback. Did they annotate it as the authority did or not? To understand this form of learning, one could focus on the two subregions of the apparatus of interest and track how traces ripple through them and what differences are significant. The scaffolding experiment described above could allow one to further explore these dynamics. Whether or not a user ID stays active longer and performs annotations with high accuracy after frequenting the tutorial and field guide is what matters.

Second, communal learning forms around a central collective activity, object, or event. All other elements are identified by their resonance with that center. For instance, at a festival or during a communal celebration, the collective develops a joint experience around this shared event. Communal learning takes form as the collective takes shape and extends its performances. Relevant traces could be the folksonomies of shared hashtags that develop in the discussion board and collections feature or new glitch classes developed by participants out of images relegated to "*none of the above*." What matters are the formation of these collective hashtags, the degree to which they are used over time, and how they solidify into new glitch classes used by a range of participants.

Third, one can envision agent-centered learning with no central focal point. Rather, the agents' evolving practices build on one another to form a bricolage that pieces together elements of their participation as they move through Gravity Spy and beyond. Boundaries are fluid, and the apparatus and phenomena are not defined but continuously morph and change as participants develop practices and discourses associated with, e.g., gravitational waves and the LIGO detectors. It is the sequential ordering of traces and the type of resources, discussions, people, and events they link that matter. Piecing together the session types that individuals combine can help develop the understanding of agent-centered learning.

These three forms of learning are not mutually exclusive. As researchers, citizen scientists can approach the apparatus in multiple ways, demarcate their phenomena of interest, and perform certain cuts through their intra-actions with the apparatus. This does not mean that anything goes; we cannot dream up endless forms of learning and project them onto an apparatus. One needs to perform differences that matter—i.e., ripples moving through the apparatus and creating some effect on a phenomenon. Operating with multiple forms of learning does not constitute a contribution. The field has long acknowledged cognitive and situated learning theories side by side (Lave & Wenger, 1991; Miner, Bassoff, & Moorman, 2001; Gherardi, 2006; Levinthal & Rerup, 2006). Rather, a diffractive reading embraces multiple entwined forms of learning, all operating in a dynamic field of possibilities.

5 Discussion

Facing a torrent of trace data, IS researchers confront a number of methodological challenges associated with building an apparatus and understanding how it co-constitutes the phenomena under investigation. Trace data are not given but produced. Thus, they do not refract or reflect some pregiven reality that researchers can project through hard labor onto the pages of their articles. The boundaries defining the phenomena of interest are not prepackaged subjects and objects. Instead, the researcher needs to pay careful attention to how the building of the apparatus demarcates different entities and the way they co-constitute one another. Carefully assembling an apparatus and following the traces rippling through it offer new ways to explore organizational practices. We contribute in this paper by offering a number of methodological principles and strategies for such a diffractive approach to trace data, as summarized in Table 2. These are not bureaucratic procedures to be followed one after another, but rather fundamental questions guiding the research process. We find it helpful to think of the research process as a circular motion in which we track the way that traces ripple through the apparatus. Continuing this iterative process enables scholars to follow how things take shape and to describe how boundaries form and fall apart. By observing and experimenting with rippling traces, the dynamics of our research practices expose the *becoming* of technologies, people, and entities and articulate how their boundaries and properties are reshaped, with what consequences and for whom (Cecez-Kecmanovic, et al., 2014, p. 821). Equally important, the methodology offers a fresh view on divisions in the IS literature. Below, we briefly discuss some implications for future research.

Leading voices in the sociomateriality debate long called for empirical studies investigating how relations and boundaries between humans and technologies are enacted in practice, rather than pregiven or fixed (Orlikowski & Scott, 2008; Jones, 2014; Suchman, 2007; Lave & Wenger, 1991). Before these dynamics can be examined, we need to understand how boundaries and distinctions emerge as part of our research process. Even if we fully accept the relational and inseparable nature of our sociomaterial world, we cannot question all distinctions in every study. It is paramount, however, that we recognize the distinctions that we make and understand where they appear in the research process. We need to acknowledge what Barad (2007) calls agential cuts—differences that matter. Recognizing these distinctions will not catapult us back to a substantialist position. Rather, it will strengthen a process perspective on how distinctions and boundaries emerge in the entanglement of human beings and materials (Cecez-Kecmanovic et al., 2014).

Table 2. Methodological Principles, Strategies, and Evidence

Principle	**Strategies and questions**	**Evidence from learning in Gravity Spy**
Demarcating the phenomena and apparatus	What are the boundaries of the apparatus? And thus, what are the phenomena?	Demarcating the apparatus call into question: *What* is learning? What is *learning*? For example, including the larger LIGO collaboration leads to a study of societal knowledge production. Restricting the apparatus to Gravity Spy traces may point to performances associated with the volunteers, machine learning unit, or community of participants. Boundaries remain fuzzy and we cannot draw a sharp line between entities, e.g., Gravity Spy and LIGO. We know that volunteers work anonymously on the site and use non-Zooniverse systems. We consider whether those performances play a role in learning.
	What cuts do the apparatus make?	What entities can we distinguish in the learning environment? For instance, can we associate certain performances to volunteers, machine learning units, and science team members, or does the apparatus not allow us to distinguish e.g., humans and machine learning? We explore how a single user ID might represent an individual, a school class or family of four. The same questions should be asked about other central performances attributed to science teams and machine learning units.
	Genealogy of an apparatus: How have the boundaries and cuts changed over time?	Explore how the learning environment changes over time? This helps us detect important distinctions performed by the apparatus. For instance, there is a clear distinction between the annotation system and discussion forums in Gravity Spy.
Extending the apparatus	What additional traces might be helpful?	To analyze Gravity Spy trace data, we built a database merging several datasets. We also persuaded Zooniverse to add tracking capabilities to the platform to record users' interactions.
	What additional cuts might be helpful? For example, statistical tools can be added to the apparatus performing additional cuts	To understand how performances evolve over time, we parse traces into sessions divided by gaps of inactivity. We try out different statistical apparatuses to see if they help distinguish cuts that matter, e.g., k-means clustering. Does one simply regard the number of times a user ID has visited certain features as contributing to learning or does the sequence of performances matter?
Diffraction: Explore how traces diffract (i.e., not refract or reflect).	How do traces ripple through the apparatus?	What performances by other agents are participants allowed to access and when? What consequences do they have for learning? In Gravity Spy participants cannot access other people's annotation work. Instead, participants go to Talk looking for practice proxies, in the form of descriptions of work.
	What intra-actions do the ripples highlight?	By adding cluster analysis to the apparatus, we explored how traces rippled through the apparatus in different ways and formed multiple patterns. Some ripples stayed within the annotation system (e.g., light work), others spanned multiple performances (e.g., talking and annotating).

	What happens if you change the way that traces ripple through the system?	An A/B split in Gravity Spy experimented with two pathways through the apparatus. One group was guided through an ML supported scaffolding of the work and a second group went straight to classify all known classes. By changing people's access to the ML in the apparatus, we can follow how traces ripple differently through the system and evaluate whether they lead to different patterns and performances. Visualizations of traces serve as interview prompts and help explore how performances ripple within and beyond the boundaries of the apparatus. The A/B split and interviews allowed us to look for differences that matter for performances associated with the apparatus.
Differences that matter	How does a circular movement between exploring the boundaries of the apparatus/phenomenon and the way that traces ripple through it help us find differences that matter?	To explore *what* is learning and what is *learning* we move in circular patterns between different apparatuses/phenomena and agential cuts that shape the way that traces ripple through these configurations.
	What differences matter?	A diffractive method allows us to operate with multiple forms of learning that play out in co-configured apparatuses: Authority-subject, communal, agential, and machine learning are all performances associated with Gravity Spy. For each of these learning phenomena, different traces and cuts matter.

The methodological principles and strategies outlined in Table 2 help guide the research process but also articulate the genealogy of boundaries and distinctions. These principles remind us that we, as researchers, are an integral part of the apparatus, not in the sense that we distort some reflection of user behaviors, but rather that our active engagement in building and running the apparatus offers rich opportunities to explore how boundaries and cuts emerge and what can and cannot be known about the ongoing dynamics of becoming associated with the system. Data collection practices are open to rearrangements and the creativity of scientific practices includes the skill of making an apparatus work for a purpose. Elements are reworked and adjusted, leading to adjustments of the boundaries and cuts performed by the apparatus. In ethnographic monographs, it has long been the norm to include a reflection on the researcher's entrance to the field. Future IS publications using trace data might similarly require an appendix describing the building and running of the apparatus in a way that acknowledges the distinctions and boundaries drawn and shows where they emerged in the research process. We would extend our attention beyond a human-centered emphasis on the interpreter and his or her position to include sociomaterial concerns about the apparatus.

Ethical considerations are an appropriate part of these considerations. Instead of framing ethical research as impacting or interacting with human subjects in a way that ensures their rights and welfare, a diffractive approach articulates how the research makes responsible and accountable distinctions and connections to what comes to matter and what is excluded from mattering. Future research could further articulate such approaches to ethics and its consequences for institutional review boards and research practices. Likewise, we have only scratched the surface when it comes to a diffractive methodology. As Barad's work suggests, this methodology allows us to revisit well-worn categories and see them in a new light, including, among other aspects, causality, discourse, measurement, time, and space (Barad, 2007).

A diffractive methodology suggests ways to integrate quantitative and qualitative approaches. Cluster analyses and interviews both have a role to play. As highlighted in our study of Gravity Spy, both methods help explore how traces ripple through the apparatus. Visualizations of trace data can serve as powerful interview prompts that may, in turn, inform changes to the apparatus, which then allows for the tracking of other practices and alterations to the cuts and boundaries. Researchers read insights gained from these different techniques through one another in a cyclical motion as one follows the traces' ripples through the apparatus. It will take additional research to map a broader range of productive combinations of participant observation, interviews, and various statistical techniques.

Our guidelines have practical implications. Building a research apparatus and paying attention to its performances brings diffractive methodology into close proximity with design theory (Hanseth & Lyytinen, 2010) and neighboring disciplines with a design agenda, such as computer supported cooperative work (CSCW) (Bjørn & Østerlund, 2014).

One can envision a joint interest in how the apparatus and phenomenon intra-act and the ways in which distinctions take shape and categories are bounded. The diffractive ways that traces ripple through the Gravity Spy project was as relevant to the designers at Zooniverse as it was to our research and the volunteers. All were hoping to learn about and improve organizational performances.

6 Conclusion

We started out noting how information systems have become pervasive platforms for work and life that capture data about organizational and everyday practices in great detail. Such abundant trace data open new areas of study with vast potential for discovery. But, to leverage these opportunities requires the rethinking of longstanding and trusted methodological principles. We cannot untangle the social and material in these big and heterogeneous data spanning transaction logs, conversation transcripts, and source code. There is no way to tell where the material starts and the social ends, as they are ontologically inseparable. Accepting this basic premise calls into question our long-standing propensity to use visual phenomena as metaphors for thinking and knowledge production, e.g., a method serves as a lens magnifying an object of interest, data reflect parts of an organizational context, or the interpretive scholar applies a reflexive approach to a topic. As noted by Haraway (1997), all this visual imagery produces “the same” displaced. We have come to expect clearly bounded and pregiven substances that we can magnify, mirror, or project in ways that allow us to study them in great detail. Equally important, these visual metaphors inevitably promulgate the observer staring through the lens or the interpreter reflecting on the images produced by their methods. The human agent takes the lead role and leaves technologies largely understudied in organizational research.

To nurture a sociomaterial methodology that takes ontological inseparability as its point of departure, we advance Haraway’s (1997) and Barad’s (2007) conceptions of diffraction and apparatus as central methodological metaphors in IS trace studies. The method attempts to “meet the universe halfway,” as suggested by the title of Barad’s 2007 book. We should not try to peek at the universe through our scientific lenses (Figure 1), nor should we engage in armchair activities, in which a human interprets worldly reflections (Figure 2). Instead, we must meet the world halfway by making the apparatus our pivot (Figure 3). Agential cuts take place here, mark the boundaries of a phenomenon under investigation, and help establish the conditions for causal relationships and agency. When the apparatus changes, so does the phenomenon, and with it, relevant intra-actions. Trace data play a central role, if we hope to understand the workings of an apparatus. The metaphor of diffraction trains our attention on how traces emerge and move through the apparatus and help demarcate the phenomenon under study. Following traces allows us to understand what differences matter. Genealogical analysis of the apparatus shows how distinctions are produced, instead of assuming pregiven substances.

This perspective brings us back to the sociomateriality debate and the apparent tension between an ontology of inseparability and the methodological need to make distinctions and draw boundaries as part of a research study. To overcome this conundrum, we must acknowledge the apparatus and the boundaries, agential cuts and diffractive patterns they perform. Only then can we leverage trace data to explore the sociomaterial nature of organizational information systems’ use. We believe that a diffractive methodology offers a promising approach that allows researchers to draw on trace data in a way that does not presume pregiven entities, but opens up the apparatus and lets us explore organizational and everyday practices in new and productive ways.

Acknowledgments

We thank the many citizen scientists whose engagement have made Gravity Spy possible and our collaborators on the project, including M. Zevin, S. Coughlin, S. Bahaadini, E. Besler, N. Rohani, S. Allen, M. Cabero, A. Katsaggelos, S. Larson, T. Littenberg, A. Lundgren, J. Smith and L. Trouille. Our gratitude also goes to the reviewer team who helped develop this paper from its scruffy beginnings. We are particularly grateful for the invaluable feedback and guidance we have received from Dubravka Cecez-Kecmanovic. A special thank you goes to Susan Scott, who long ago over coffee in Corfu called our attention to the notion of diffraction. Gravity Spy is partly supported by the US National Science Foundation award INSPIRE 15–47880.

References

Abbasi, A., Sarker, S., & Chiang, R. H. (2016). Big data research in information systems: Toward an inclusive research agenda. *Journal of the Association for Information Systems*, *17*(2), i-xxxii.

Aiden, E., & Michel, J. B. (2014). *Uncharted: Big data as a lens on human culture*. Penguin.

Antin, J., & Cheshire, C. (2010). Readers are not free-riders: Reading as a form of participation on Wikipedia. *Proceedings of the ACM Conference on Computer-supported Cooperative*.

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

Barad, K. (2007). *Meeting the universe halfway: Quantum physics and the entanglement of matter and meaning*. Duke University Press.

Barley, S. R. (1986). Technology as an occasion for structuring: Evidence from the observation of CT scanners and the social order of radiology departments. *Administrative Science Quarterly*, *31*, 78-109.

Beane, M., & Orlikowski, W. J. (2015). What difference does a robot make? The material enactment of distributed coordination. *Organization Science*, *26*(6), 1553-1573.

Bjørn, P., & Østerlund, C. (2014). *Sociomaterial-design: Bounding technologies in practice*. Springer.

Boland, R. J. (1985). Phenomenology: A preferred approach to research on information systems. In E. Mumford, R. Hirscheim, G. Fitzgerald, & T. Wood-Harper, T. (Eds.), *Research Methods in Information Systems* (pp. 193-201). North-Holland.

Bourdieu, P. (1977). *Outline of a theory of practice*. Cambridge University Press.

Brown, J. S., & Duguid, P. (2001). Knowledge and organization: A social-practice perspective. *Organization Science*, *12(2)*, 198-213.

Bryant, S. L., Forte, A., & Bruckman, A. (2005). Becoming Wikipedian: Transformation of participation in a collaborative online encyclopedia. *Proceedings of the International ACM SIGGROUP Conference on Supporting Group Work.*

Callon, M. (1998). *The laws of the markets*. Blackwell.

Callon, M., & Muniesa, F. (2005). Peripheral vision: Economic markets as calculative collective devices. *Organization Studies*, *26*(8), 1229-1250.

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.

Cecez-Kecmanovic, D. (2016). From substantialist to process metaphysics: Exploring shifts in IS research. *Proceedings of the IFIP Working Group 8.2 Conference on Information Systems and Organizations*.

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, *(36)4*, 1165-1188.

Crowston, K., Østerlund, C., & Lee, T. K. (2017). Blending machine and human learning processes. *Proceedings of the 50th Hawaii International Conference on System Sciences*.

Feldman, S. & Orlikowski, W. (2011). Theorizing Practice and Practicing Theory. *Organization Science*, *22(5)*, 1240-1253.

Foucault, M. (1980). *Power/knowledge: Selected Interviews and other writings, 1972-1977*. Pantheon.

Gadamer, H. G. (1975). *Truth and method*, trans. W. Glen-Dopel. Sheed & Ward.

Gaskin, J., Berente, N., Lyytinen, K., & and 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. *Proceedings of the 44th Hawaii International Conference on System Sciences*.

Geiger, R. S., & Halfaker, A. (2013). Using edit sessions to measure participation in Wikipedia. *Proceedings of the Conference on Computer-supported Cooperative Work*.

Gherardi, S. (2006). *Organizational knowledge: The texture of workplace learning*. Blackwell.

Giddens, A. (1979). *Central problems in social theory: Action, structure and contradiction in social analysis*. University of California Press.

Goggins, S. P., Galyen, K., & Laffey, J. (2010). Network analysis of trace data for the support of group work: Activity patterns in a completely online course. *Proceedings of the 16th ACM International Conference on Supporting Group Work*.

Halfaker, A., Keyes, O., & Taraborelli, D. (2013). Making peripheral participation legitimate: Reader engagement experiments in Wikipedia. *Proceedings of the Conference on Computer-supported Cooperative Work*.

Hanseth, O., & Lyytinen, K. (2010). Design theory for dynamic complexity in information infrastructures: The case of building Internet. *Journal of Information Technology, 25(1)*, 1-19.

Haraway, D. (1991) *Simians, cyborgs and women*. Free Association Books.

Haraway, D. (1997). *Modest_Witness@Second_Millenium.FemaleMan_Meets_OncoMouse: Feminism and technoscience*. Routledge.

Heidegger, M. 1949. *Sein und Zeit*. Neomarius.

Howison, J., Wiggins, A., & Crowston, K. (2011). Validity issues in the use of social network analysis with digital trace data. *Journal of the Association for Information Systems*, *12*(12), 767-797.

Jackson, C. B., Østerlund, C., Mugar, G., Hassman, K. D., & Crowston, K. (2015). Motivations for sustained participation in crowdsourcing: Case studies of citizen science on the role of talk. *Proceedings of the 48th Hawaii International Conference on System Sciences*.

Jackson, C., Østerlund, C., Maidel, V., Crowston, K., & Mugar, G. (2016). Which way did they go? Newcomer movement through the Zooniverse. *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*.

Jones, M. (1998). Information systems and the double mangle: Steering a course between the Scylla of embedded structure and the Charybdis of strong symmetry. *Proceedings of the IFIP WG8.2 and 8.6 Joint Working*.

Jones, M. (2014). A matter of life and death: Exploring conceptualizations of sociomateriality in the context of critical care. *MIS Quarterly, 38*(3), 895-925.

Kautz, K., Jensen, T.B. (2013). Sociomateriality at the royal court of IS: A jester's monologue. *Information and Organization, 23(1)*, 15-27.

Keegan, B. C., Lev, S., & Arazy, O. (2015). Analyzing organizational routines in online knowledge collaborations: A case for sequence analysis in CSCW. *Proceedings of the Conference on Computer-supported Cooperative Work and Social Computing*.

Klein, H. J., & Weaver, N. A. (2000). The effectiveness of an organizational-level orientation training program in the socialization of new hires. *Personnel Psychology, 53*, 47-66.

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

Latour, B. (1986). Visualization and cognition. *Knowledge and Society, 6*, 1-40.

Lave, J. & Wegner, E. (1991). *Situated learning*. Cambridge University Press.

Law, J. (2004). *After method: mess in social science research*. Routledge.

Levinthal, D. A. & Rerup, C. (2006). Crossing an apparent chasm: Bridging mindful and less-mindful perspectives on organizational learning. *Organization Science, 17(4)*, 502-513.

Loukissas, Y. A. (2017). Taking Big Data apart: Local readings of composite media collections. *Information, Communication & Society, 20(5)*, 651-664.

March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science*, *2(1)*, 71-87.

Miner, A., Bassoff, P., & Moorman, C. (2001). Organizational improvisation and learning: A field study. *Administrative Science Quarterly, 46*, 304-337.

Mugar, G., Østerlund, C., Hassman, K. D., Crowston, K., & Jackson, C. B. (2014). Planet hunters and seafloor explorers: Legitimate peripheral participation through practice proxies in online citizen science. *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*.

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

Orlikowski, W. J. (2007). Sociomaterial practices: Exploring technology at work. *Organization studies*, *28*(9), 1435-1448.

Orlikowski, W. J. (2009). The sociomateriality of organisational life: Considering technology in management research. *Cambridge Journal of Economics*, *34*(1), 125-141.

Orlikowski, W. & Scott, S. (2008). Sociomateriality: Challenging the separation of technology, work, and organization. *Academy of Management Annals*, *2(1)*, 433-474.

Østerlund, C., & Carlile, P. (2005). Relations in practice: Sorting through practice theories on knowledge sharing in complex organizations. *The Information Society*, *21*(2), 91-107.

Ostroff, C., & Kozlowski, S. W. J. (1992). Organizational socialization as a learning process: The role of information acquisition. *Personnel Psychology*, *45*, 849-874.

Phang, C. W., Kankanhalli, A., & 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*, 721-747.

Pickering, A. (1995). *The mangle of practice: Time, agency and science.* University of Chicago Press.

Riemer, K., & Johnston, R. B. (2017). Clarifying ontological inseparability with Heidegger's analysis of equipment. *MIS Quarterly*, *41*(4).

Schatzki, T. R. (2002). *The Site of the social: A philosophical account of the constitution of social life and change*. Penn State University Press.

Scott, S.V, & Orlikowski, W. J. (2014) Entanglements in practice: Performing anonymity through social media. *MIS Quarterly*, *38*(3), 873-893.

Shankar, K., Hakken, D., & Østerlund, C. (2017). Rethinking documents. In U. Felt, R. Fouché, C. A. Miller, & L. Smith-Doerr (Eds.), *The handbook of science and technology studies* (pp. 59-86). MIT Press.

Sørensen, E. (2009). *The materiality of learning: Technology and knowledge in educational practice*. Cambridge University Press.

Suchman, L. A. (2007). *Human-machine reconfigurations: Plans and situated actions* (2nd ed.). Cambridge University Press.

Swartz, D. (1997). *Culture and Power: The Sociology of Pierre Bourdieu*. University of Chicago Press.

Van Maanen, J., & Schein, E. H. (1979). Toward a theory of organizational socialization. In B. M. Staw (Ed.), *Research in organizational behavior* (pp. 209-264). JAI.

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

Yoo, Y. (2010). Computing in everyday life: A call for research on experiential computing. *MIS Quarterly*, *34*(2), 213-231.

Zevin, M., Coughlin, S., Bahaadini, S., Besler, E., Rohani, N., Allen, S., Cabero, M., Crowston, K. Katsaggelos, A., Larson, S., Lee, T. K., Lintott, C., Littenberg, T., Lundgren, A., Østerlund, C., Smith, J., Trouille, L., & Kalogera, V. (2017). Gravity Spy: Integrating Advanced LIGO detector characterization, machine learning, and citizen science. *Classical and Quantum Gravity*, *34*(6), 064003.

Zuboff, S. (2019). Surveillance capitalism and the challenge of collective action. *New Labor Forum*, *28*(1), 1-29.

About the Authors

Carsten Østerlund is a professor at the iSchool at Syracuse University. He earned his PhD in management from Massachusetts Institute of Technology. He also studied at University of California, Berkeley, University of Århus, and University of Copenhagen and was affiliated with the Work Practice and Technology Group at Xerox PARC. His research explores the organization, creation, and use of documents in distributed environments with an emphasis on learning and knowledge dynamics in new forms of work. Empirically, he studies these issues through in-depth qualitative studies of everyday work practices in a range of settings, including citizen science, crowdsourcing, distributed science teams, game design and healthcare. Recently, he has been particularly interested in how we can merge qualitative and quantitative methodologies to study trace data associated with AI systems.

Kevin Crowston is a Distinguished Professor of Information Science at the Syracuse University School of Information Studies (aka the iSchool). He received his AB (1984) in applied mathematics (computer science) from Harvard University and a PhD (1991) in information technologies from the Sloan School of Management, Massachusetts Institute of Technology. He currently serves as associate dean for research. His research examines new ways of organizing made possible by the use of information technology. He approaches this issue in several ways: empirical studies of coordination-intensive processes in human organizations (especially virtual organization), theoretical characterizations of coordination problems and alternative methods for managing them, and design and empirical evaluation of systems to support people working together. Specific domains of interest include free/libre open source software development projects, citizen science projects, and research data management.

Corey Jackson is a postdoctoral scholar with the School of Information, University of California, Berkeley. He received a BA in political science from the University of Illinois Urbana-Champaign in 2010, an MS in library and information science from the Graduate School of Library and Information Science at the University of Illinois Urbana-Champaign UIUC in 2012, and a PhD in information science and technology from the School of Information Studies at Syracuse University. His research interest includes phenomena such as contribution behaviors, learning, and motivation of users in large-scale sociomaterial systems.