Journal of the Association for Information Systems (2021) **22**(6), 1715-1735
doi: 10.17705/1jais.00722

RESEARCH ARTICLE

ISSN 1536-9323

Acting with Inherently Uncertain Data: Practices of Data-Centric Knowing

Marius Mikalsen[1], Eric Monteiro[2]

[1]Norwegian University of Science and Technology / SINTEF, Norway, marius.mikalsen@sintef.no
[2]Norwegian University of Science and Technology, Norway, eric.monteiro@ntnu.no

Abstract

Data-driven data science challenges our conceptualization of "data." Significantly beyond capturing a given phenomenon, data, increasingly, *are* the phenomenon. Data may be iteratively manipulated algorithmically, undermining the "faithfulness" of data to any originating phenomenon. Crucially, data that are not "faithful" are inherently uncertain, as data risk becoming meaningless symbols. We empirically study how a community of commercially based geoscientists grapples with the phenomenon of offshore oil and gas reservoirs located kilometers below the seabed. The data available about these reservoirs are algorithmically manipulated sensor-based Internet of Things data. Our main contribution is the articulation of three patterns of work practices detailing *how* inherently uncertain data are woven into consequential work practices: (1) *Accumulating* is the cumulative process of supporting and triangulating one set of data with supplementary data sets; accumulating captures the conservative approach of backing up existing interpretations of the data. (2) *Reframing* describes the process where existing interpretations are contested by new data or models; reframing captures that there are limits to how far data can be manipulated. (3) *Prospecting*, or the cultivation of competing, incompatible data interpretations; while the former two patterns essentially attempt to regulate uncertainty, prospecting is about embracing it. Our concept of *data-centric knowing* is constituted by these three interwoven, ongoing practices.

Keywords: Data, Work Practices, Uncertainty, Empirical, Case Study, Knowing, Big Data, Data-Driven, Data Science, Internet of Things, Commercial

Magnus Mähring was the accepting senior editor. This research article was submitted on January 12, 2019 and underwent three revisions.

1 Introduction

Easily mistaken for a purely philosophical concern, data's representational capacity, i.e., data's capacity to represent a phenomenon (Zuboff, 1988; Burton-Jones, 2014; Kallinikos, 2007; Borgmann, 1999), is increasingly recognized as being at the core of discourses about big data, data science, and data-driven machine learning (Alaimo et al., 2020; Alaimo & Kallinikos, 2021; Markus, 2017, Zuboff, 2019). Initially referring to a physical object, process, or quality, data are iteratively sliced, recombined, and algorithmically manipulated, taking them increasingly away from their originating physical referents (Kallinikos et al., 2013; Orlikowski & Scott, 2016; Lusch & Nambisan, 2015). Increasingly, data are the phenomena, i.e., "signs" (Knorr Cetina, 1999) or "symbols" (Bailey et al., 2012) of events that are otherwise inaccessible.

However, data's representational capacity is exactly that—a capacity. It may, but most certainly need not, be actualized in the sense that data are woven into everyday, data-driven, consequential work practices (Günther et al., 2017). With data's representational capacity but modestly exercised, the conditions for actualization are reasonably well-understood: data need to be a "faithful" (Burton-Jones & Grange, 2013) representation with a "tight coupling" with the phenomenon (Bailey et al., 2012, p. 1500) because "seeing is believing" (Leonardi, 2012, p. 14). What remains unaccounted for, is how data that no longer faithfully represent a phenomenon are actualized (Kallinikos, 1999; see also Kitchin, 2014). This paucity in the literature, of increasing empirical relevance and significance with expanding datafication (Newell & Marabelli, 2015), is the focus of our paper. Rather than delegating this to a new role of "data translators" (Henke et al., 2018), we analyze the work practices involved.

Crucially, data no longer faithfully representing phenomena are inherently uncertain, as their meaning is not (yet) fixed (Alaimo et al., 2020). Actualizing data thus involves practices of sense-making (Weick, 1985) in situations mired in uncertainty. We pose the following research question: How and under what organizational conditions, are inherently uncertain data actualized in consequential work practices?

We empirically study how a commercially based community of geoscientists grapples with the phenomenon of offshore oil and gas reservoirs located kilometers below the seabed. For all practical purposes, the data and algorithmic representations of the reservoirs are the reservoirs in everyday work practices. The available data about the oil reservoirs are largely algorithmically manipulated sensor-based Internet of Things (IoT) data with inherent epistemic uncertainty. We analyze the work practices of geoscientists dealing with incomplete, inconsistent, and inherently uncertain data. Their work practices emerge from conflicting tensions where professionally acquired quasi-scientific approaches run up against the commercial push for operational decision-making.

We contribute with the articulation of three patterns of work practices, detailing how inherently uncertain data are actualized. First, accumulating, which is the cumulative process of supporting and triangulating one set of data with supplementary data sets. Accumulating captures the conservative approach of backing up existing interpretations of the data. Second, reframing, which is the process where existing interpretations are contested by new data or models. Reframing captures how there are limits to how far data can be manipulated. Third, prospecting[1], which is the process of cultivating competing, incompatible data interpretations. With the former two patterns essentially attempting to regulate uncertainty, prospecting is about embracing it. The concept of data-driven knowing, then, is shorthand for the interleaving of these three ongoing patterns of work practices.

The remainder of our paper is organized as follows. Section 2 develops our theoretical framework on understanding the concept of data. We review and discuss relevant perspectives on pressing data's representational capacity beyond that of faithful representation. Section 3 provides context to our case together with an account of research methods. Our empirical findings are presented in Section 4, organized around the three patterns of practices developed in the data analysis from the preceding section. The discussion in Section 5 pursues two threads. First and foremost, drawing on existing literature, we discuss the theoretical implications of the three interwoven patterns of work practices constituting data-centric knowing by critically discussing their enabling conditions. Second, in further pursuing the conditions for data-centric knowing, we analyze the institutional fabric necessary to actualize data. To conclude, Section 6 reflects on the relevance of and boundary conditions for data-centric knowing beyond our case, in addition to offering comments on future research.

2 Conceptualizing Data

2.1 Data Are "Cooked"

Traditionally, a representational view has data corresponding directly with some given, preexisting physical object, process, or quality. Such a view, Jones (2019) reminds us, is still evident, albeit in an implicit and diluted form. For instance, a textbook defines *data* as "raw facts that describe a particular phenomenon" (Haag & Cummings, 2009, p. 508), while the Royal Society (2012, p. 12) defines *data* as "numbers, characters, or images that designate an attribute of a phenomenon" (both definitions are cited in Jones, 2019). However, such a naive, referential view of data has several problems, as we will discuss.

Contrary to big data hype, in which "the numbers [data] speak for themselves" (Andersen, 2008), considerable, ongoing work is needed to craft data *into* data. The notions of "gathering" or "collecting" data are misleading in that they promote the misconception that data speak for themselves. This downplays to the level of nonexistence the way that data provenance—i.e., the methods, procedures, and technologies employed to generate the data—shapes data use and interpretation (Porter, 1996). As Gitelman (2013, p. 3) succinctly puts it, data "are always already 'cooked' and never entirely 'raw.'"

[1] Our use of the term prospecting comes from its analogy with geological prospecting, which is an open-ended, conflictual search for competing, very often mutually inconsistent geological interpretations. In contrast, Slota et al. (2020) use the term to denote the rendering of data amendable to data science methods.

Thus, both collecting and curating data involve work, such as developing and maintaining procedures for cleaning, filtering, and "massaging" data (Leonelli, 2014). Edwards (1999) examined the comprehensive data-gathering process informing climate-change research and reports that measurement devices, such as thermometers, must be constantly calibrated to ensure the validity of their readings. In this context, maintaining calibration involves adhering to protocols that compare a given thermometer with a master device and systematically adjust historic measurement values when it is discovered that a thermometer is uncalibrated.

A significant source of "big" data, not the least of which is in industrial settings, is sensor-based IoT data (Singh et al., 2014). Contrary to appearances, sensors' measurements do not "capture" physical reality in a straightforward manner. IoT measurements are highly constructed renderings of selected aspects of a physical situation, fitted for designated purposes. Anything but "natural," the data gained from sensors is highly mediated—materially as well as epistemologically (Monteiro & Parmiggiani, 2019; Helmrich, 2019).

2.2 The Representational Capacity of Data

Data are inherently editable, recombinable, and subject to repurposing (Kallinikos et al., 2013; Alaimo et al., 2020). This underpins discussions about datafication (Newell & Marabelli, 2015; Markus, 2017; Sugimoto et al., 2016). Numerous observations and concepts tap into the same idea, including, but not limited to, Kallinikos's (2007) "increasingly self-referential rendition of reality," Orlikowski and Scott's (2016) "algorithmic phenomenon," Baskerville et al.'s (2020), "digital objects," and Nambisan and Lusch's (2015; see also Monteiro & Parmiggiani, 2019) "liquefaction," which denotes data's capacity to decouple from their originating physical objects, processes, or qualities.

Taking a step back, the idea here is to conceptualize data's representational capacity beyond the realm of naive realism outlined above. Zuboff (1988) was one of the first to influential analyze data's representational capacity (see also Weick, 1985). Her work generated a wealth of interest, primarily regarding the conditions for an engaged, learning-oriented relationship with technology ("informate", as opposed to "automate" in her vocabulary). This is a pity, Burton-Jones (2014, p. 72) notes, as "[Zuboff, 1988] may have an even stronger story to tell now than it did when first published," given the proliferation of empirical cases of datafication, i.e., cases exercising data's representational capacity beyond faithful representations (Alaimo et al., 2020; Alaimo & Kallinikos, 2021; Burton-Jones, 2014; Burton-Jones & Grange, 2013; Monteiro & Parmiggiani, 2019).

Knorr Cetina (2009) offers a helpful way to approach the issue of data's representational capacity. Her seminal book is based on an ethnography of high-energy physicists' work at the CERN particle collider in Geneva, Switzerland, an example of knowledge work completely immersed in the data of phenomena, but indirectly observable (see also Venters et al., 2014). Knorr Cetina (1999) differentiates (analytically, not necessarily empirically) between three manifestations of data. Physical phenomena, first and traditionally, may be *staged* to produce data that "correspond" with phenomena directly. Second, physical conditions are manipulated to yield *processed, partial versions* of data that are "equivalent" or similar. Third, and most radically, physical phenomena are mere "signatures" and "footprints" of events, yielding data as *signs*. Bailey et al. (2012) arrive at essentially the same taxonomy of three manifestations of data. Drawing on Peirce's semiotics, they identify data that are *indices* (a direct correspondence, similar to Knorr Cetina's staged data), *icons* (similar or equivalent to Knorr Cetina's processed, partial data), and *symbols* (no link to referents, similar to Knorr Cetina's data as signs).

Data as signs (Knorr Cetina, 1999) or symbols (Bailey et al., 2012) demonstrate the potential—but crucially not necessarily the actualization—of data's representational capacity. Several scholars have addressed the conditions under which data as signs or symbols are actually woven into work practices, i.e., the conditions under which data become more than mere symbols i.e., what Kallinikos (1999) calls "referential attribution." In her empirical study of digital transformation of pulp factories from experience-based, embodied, tactile handcraft—smelling, tasting, and feeling the temperature of the pulp—into a remotely operated, digitally enabled control room, Zuboff (1988) notes the unease stemming from data "replacing a concrete reality" (p. 63) and from how data "replace the sense of hands-on" (p. 65) and seek to "invent ways to conquer the felt distance of the referential function [i.e., the decoupling of data from the physical referent]." A lack of sensory feedback undermines the expertise and knowledge that can be acquired from physical, hands-on interaction with technology. Similarly, Turkle's (2009) work emphasizes the dangers of simulation-based renditions of reality, with their strong, seductive capabilities (see Baudrillard, 1994). As users are gradually immersed in simulations, "familiarity with the behavior of [simulation data] can grow into something akin to trusting them, a new kind of witnessing" (Turkle, 2009, p. 63). Also, Burton-Jones and Grange (2013), with their focus on data representations being *faithful*, voice concerns about the limits of representational capacity, as data need to "faithfully represent some domain because they provide a more informed basis for action than unfaithful representations do" (2013, p. 636).

2.3 Toward Data-Centric Knowing Under Uncertainty

To take stock, data's representational capacity is reasonably well understood as long as this capacity is but modestly exercised, i.e., as long as data is a faithful representation of the "tight coupling" (Bailey et al., 2012) with their physical referent. The problem, however, is that the datafication of our lifeworld—the slicing, dicing, and algorithmic manipulation of data that undermine the faithfulness of the data—leaves an expanding empirical phenomenon inadequately accounted for. The aim and ambition of our concept of data-centric knowing is to address this paucity in the literature, a theoretical paucity with growing empirical relevance. Our aim is in line with the call made by Lyytinen and Grover (2017, p. 229) "to critically evaluate how we approach and think about data, its provenance, privacy, and related organizational practices."

Several theoretical problems, resonating with our subsequent empirical analysis, motivate our development of data-centric knowing. Navigating in a situation with, as it were, no stable or fixed ground, data-centric knowing is centered around *fallible* knowing practices, hence tenets of pragmatic action (Dewey, 1930). Without faithful representation, inherent uncertainty exists regarding what data signify, if anything (Alaimo et al., 2020). A principal task, then, for data-centric knowing is to detail *how* users navigate with inherent epistemic uncertainty.

Ours is a case of data-driven work practices. A defining, somewhat ironic, aspect is that users drown in data while knowing is always underdetermined by the data. So how do users cope? Pragmatism, again, offers a starting point for data-centric knowing.

Abductive reasoning, i.e., neither inductive nor deductive, is particularly relevant in navigating with epistemic uncertainty (Dunne & Dougherty, 2016). Abduction involves short-cutting searches for "innumerable possible hypotheses all accounting for the data at hand" (Brown, 1983, p. 401). The parsimony or satisficing principle (March, 1994) regulating abduction sets boundaries (e.g., resources, time) for an otherwise open-ended process. Good-enough solutions ensure the arrival at a decision within set limits. However, the satisficing principle assumes that you know what you are looking for. In many situations, you are not looking for "known unknowns," but rather "unknown unknowns"[2] (Loch et al., 2011).

Thus, data-centric knowing addresses what remains underspecified in the extant literature—i.e., detailing the interleaved patterns of work practices that go into grappling with different levels of and forms of uncertainty in the data at the basis of operational action- and decision-making. Our case provides a particularly vivid empirical illustration to develop an understanding of data-centric knowing.

3 Method

3.1 Case Context

Our case studies *exploration*, i.e., the practice of searching for commercially viable oil and gas reservoirs through a European-based, internationally oriented, upstream oil operator dubbed OilComp. Distinctly different from its historic, roughneck origins, oil exploration in our case is a decisively knowledge-intensive, data-driven endeavor that represents the most significant investment in OilComp, typically 10-20% of total investments. Oil exploration is fiercely competitive. Exploration is *the* most strategically important activity for an oil operator, and strongly influences long-term viability and global competitiveness.

Empirically, we study the community of "explorationists," a term that they use to refer to themselves collectively (in Norwegian: *tolkere*). Explorationists comprise about 2,000 of OilComp's 20,000 employees worldwide. They are organized into projects of 7-10 people each, and we followed three projects. Explorationists comprise several professional disciplines within the geosciences, including geology, geophysics, reservoir engineering, petroleum engineering, and petrochemical engineering. We focus empirically on explorationists working in areas already identified by OilComp as commercially interesting.[3] Co-located with the explorationists are several "data managers" that support explorationists' work. The data managers help locate, prepare, and present the geodata required by the explorationists (Mikalsen & Monteiro, 2018). There is approximately one data manager for every 10 explorationists.

The hydrocarbon reservoirs explored by OilComp in our case study lie 3-5 kilometers below the seabed and are knowable largely through sensor-based IoT data, notably *seismic data* (acoustic reflection measurements and processing), *well logs* (electromagnetic and radioactive measurements of rock properties), and *production data* (real-time measurements of flow volume, temperature, pressure, and chemical composition). The work practices of explorationists rely on a portfolio of specialized digital tools for algorithmically manipulating the sensor-based IoT data. Prominent tools include seismic processing (for velocity determination, 2D and 3D seismic imaging

[2] Then-Secretary of Defense Donald Rumsfeld (in)famously used the phrase during his briefing on the Iraq situation.

[3] This corresponds to so-called license exploration. Prior to this, there is screening, i.e., deliberations about whether or not OilComp should enter into an area.

in either time or depth), geological modeling (to correlate well logs, build cross-sections and create geological maps), petrophysical tools (to load and manage well logs), and simulation tools (to estimate present and future hydrocarbon reservoirs, but also to interpret, model and validate traps).

For all practical purposes, the physical phenomenon the explorationists struggle to know in their everyday work practices—oil and gas reservoirs kilometers below the seabed hence not directly accessible—is a data-driven algorithmic phenomenon (Bond, 2015). *What* the explorationists know is *how* they know it (Monteiro & Parmiggiani, 2019), which is through sensor-based IoT data thoroughly manipulated algorithmically.

The explorationists know only too well that there are inherent epistemic uncertainties based on the lack of completeness, accuracy, and consistency in sensor-based IoT data, but they have no option but to rely on them. Consider accuracy. With a sigh, one explorationist explained that a down-hole pressure sensor's lifespan "is about two years," before "calibration will be off." Consistency across data types is challenging for several reasons, including the fact that data granularity varies. Explorationists draw heavily on seismic data. The attraction of seismic data for the explorationists is that they cover wide geographical areas (several square kilometers), thus providing a much-needed overview of geophysical conditions. The problem, however, is that seismic data are also crude in the sense that the resolution cannot distinguish between entities smaller than a cube with 100-meter-long sides (i.e., entities smaller than the size of a 15-story building). In contrast, well log data are fine-grained with a resolution of down to a meter, but necessarily only cover the well's pinpointed location (Figure 1 illustrates these data types).

Data quality is a chronic concern, not only because of error-prone IoT measurements, but also because data are shaped by the purpose of their collection. For instance, a couple of decades ago, well-logging focused on deep levels, as these corresponded to the geological era of identified interest, Jura.[4] However, more recently, explorationists have become interested in earlier geological eras too, i.e., well logs' shallower stratigraphic layers: "but when we go back in time, the shallow levels were not logged properly [i.e., data quality is poor here], [in contrast to] the deep levels."

Oil exploration entails data-driven predictions about, in their vocabulary, a "prospect,"[5] i.e., a candidate for an oil reservoir in a particular geographical location and geological formation. Verifying predictions by actually drilling an oil well may take many years, if it is even possible, with the cost of drilling approximately USD$100 million. Acquiring new seismic data to learn more about prospects is also costly, albeit less so than drilling. Explorationists' work, then, revolves around identifying, evaluating, and prioritizing these predictions by working with the data at hand in the hope that one day their prospects will indeed be validated by actual drilling.

Oil exploration is a search for particular geological conditions, what they call a *play*. A play fulfills three conditions: a source rock (from organic material, geothermally transformed into hydrocarbons), a migration path (avoiding the fate of most hydrocarbons, to evaporate or dissolve), and a trap (a rock sealing hydrocarbon into a reservoir).[6] The search takes one of two distinctly different forms. Akin to a search for the proverbial needle in a haystack, one starts with "proven plays" in the area, i.e., particular configurations of rocks and formations that, through drilling, have already yielded oil discoveries. The other type of search entails working with an "unproven play," in which you need to develop the concept first, essentially an understanding containing the three necessary conditions above. As one explorationist explained, "[it's] about [traveling] into the unknown, toward a new concept."

Oil exploration in OilComp is regulated by a formally defined, staged funnel model (see Figure 2). The first stage entails deciding what region (an area the size of about 100 square kilometers) to consider, then zooming in on (several) potential prospects before deciding whether to drill one or more of the prospects. Finally, if a significant discovery is made, the discovery is appraised, which can entail the drilling of delineation wells to determine the size of the oil field more accurately and evaluate how best to develop it and produce oil cost-effectively. A plan for development then must be submitted to national petroleum authorities for approval before operations can commence.

[4] In this area, rivers have transported and sedimented matters in layers. The layers indicate age accordingly.

[5] The naming of our third pattern of data-centric knowing, prospecting, is inspired by the work with geological prospects. However, we use prospecting (the verb) for the third pattern and prospect (the noun) for concrete, empirical prediction the explorationists are working with.

[6] This is true for traditional or "conventional" oil exploration. Hydrocarbons in "unconventionals," such as shale gas produced by fracking, is different. There is no trap in unconventionals. The migrating path is a rock with high porosity where very low permeability traps the hydrocarbons. The explorationists we study are searching for conventional oil only.

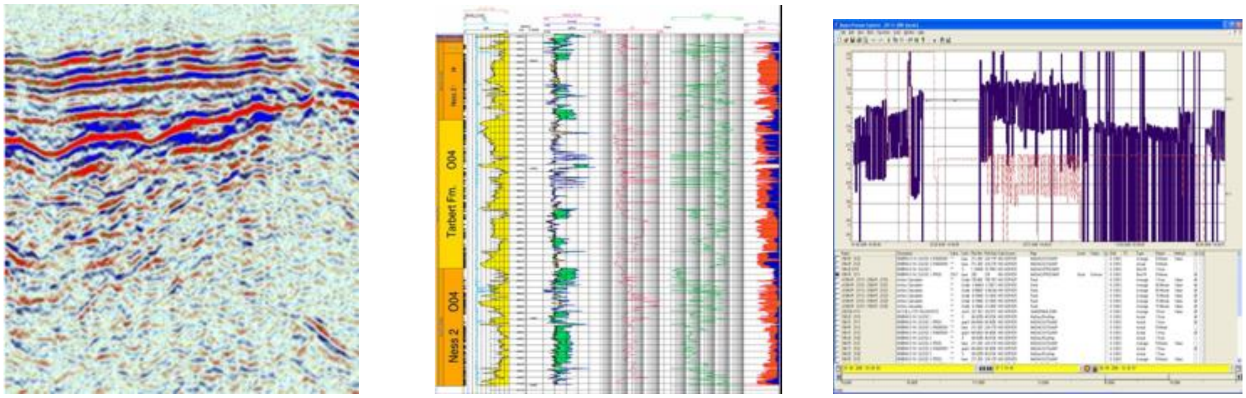


Figure 1. Typical Visual Representations of the Three Principal Types of Data: Seismic Data (left), Well Logs (middle), and Production Data (right)

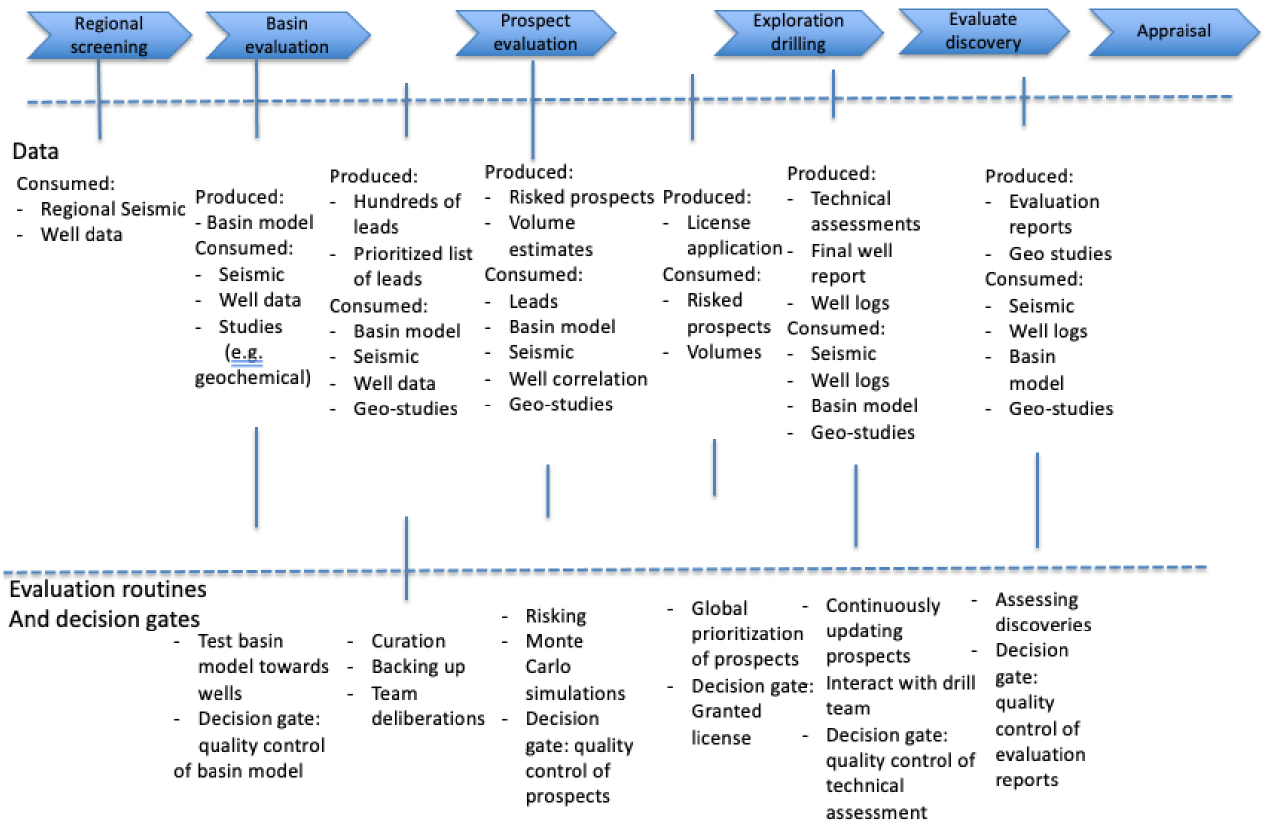


Figure 2. An Overview of the Funnel Model for the Life Cycle of Prospects with Foundational Data

At each stage ("decision gate") in OilComp's funnel process, there are different, formal requirements for actualizing data from exploration and the character of the decision-making. When proceeding further into the funnel—and, thus, closer to a potential heavy investment—the decision-making process, not unexpectedly, becomes more elaborate, as we illustrate below.

3.2 Access to Case

Access to our case was nontrivial and negotiated. The oil and gas sector in Northern Europe is increasingly global, both in its own activities as well as through international collaborations and ventures. Traditionally open to research collaborations, the companies, especially larger ones such as OilComp, have gradually regulated and tightened collaborations by imposing more formal managerial approval procedures. As a consequence, access to the case on which we report in this paper had to be negotiated. Access was granted by "packaging" it as one element within a larger research center involving several researchers (including the second author) and industrial partners (including OilComp). Piggy-backing onto the technological prototyping that the research center focused on, our access came as a response to the research center's need to understand the demand side of their prototypes supporting oil exploration. Our access to the case accordingly relies on two pillars. The second author's long-standing relationship with the research center's partners was crucial. In addition, we had to make ourselves useful within the research center by facilitating the communication of concerns, demands, and experiences from users (explorationists and data managers included) back to researchers engaged in prototyping. Thus, we prepared presentations of preliminary results and participated in meetings and workshops.

3.3 Data Collection

Our data collection spanned more than four years (February 2013-October 2017) and proceeded in rounds. Starting broadly, we studied the work processes and key concerns of actors implicated in oil exploration and their mode of collaboration. We derived a sense of the cross-pressures that oil exploration gets caught up in, such as defending tall investments, working with deadlines, and dealing with incomplete data of variable quality. Following a suggestion from van Maanen (1988), we took extensive field notes, ensuring that we separated informants' data from our own comments, reflections, and questions. We gradually focused on the explorationists' work practices, including their use of digital tools and collaboration patterns. During the final round, we focused specifically on *how* the explorationists use their data, growing steadily more aware of the convoluted relationship between data and decision-making in oil exploration.

We relied on three types of data (see Table 1 for an overview). The first author collected most of the data, but the second author also participated, especially during the later rounds. Notably, the second author also drew on background data from more than a decade of sustained research on digitalization in the oil and gas sector.

First, we used participant observations. As the details of explorationists' strategies are core assets, we were initially given office space with data managers and thereby embedded within the exploration unit. The field researcher spent time getting to know the data managers, asked questions when things happened at the office, took breaks with them, interviewed them, had lunch and dinners with them. Since data managers are tasked with finding, preparing, and presenting the required data to explorationists, this provided an effective entry into the explorationists' practices. We observed data managers' everyday work routines, including their close interactions with explorationists. Most office walls are decorated with maps and geological illustrations of 2D seismic and well logs, and a constant buzz from informal conversations can be heard. Explorationists regularly strolled over to the data managers for a cup of coffee or popped into someone's office to explain what data were urgently needed for a certain purpose.

From our initial home with the data managers, we also gradually got to know the explorationists. Accordingly, participant observations of data managers served as a resource for identifying and recruiting new informants. Spending time with our informants during their everyday work allowed us to explore nuances, unclarities, and questions that lingered after interviews. In informal conversations over coffee or lunch, we could pick up questions or puzzling issues that we were unable to pursue within the more fixed boundaries of our semistructured interviews. For instance, we inquired into why and how searching, in the age of Google, was complicated by different naming conventions across oil fields and professional subcommunities of explorationists. We also conducted participant observations through a variety of meetings, workshops, and seminars. Some formal, but most less so, these events gave us a chance to observe how explorationists backed up their interpretations, how they were challenged, and how agreement on how to proceed emerged.

Second, we conducted semistructured interviews that lasted 45-90 minutes each, with most lasting a little over an hour. Interviews were transcribed. As indicated above, we exploited the interweaving of participant observations and interviews. We conducted interviews in the informants' workspaces, which allowed informants to use their tools to give examples of geodata problems and concerns. Typically, they would point to their screens while explaining it all to us.

Third, we leveraged both electronic and paper-based documents, collecting both internal and external documents. Internal documents included memos, slide presentations, and reports. External documents covered public information on drill results (drill operations, results, and tests conducted), reclassified interpretations (final well reports, core photos, and well logs), and public reports on OilComp discoveries and recoverable reserves. The documents, especially the internal ones, e.g., from presentations, were particularly useful in identifying concerns and discussions. Examples of internal documents include prospects, workflow descriptions, tool screenshots and guides, data types, databases, and procedures and issues concerning database querying and the presentation of results. We used 150 pages of internal documents in our analysis.

Table 1. Summary of Data Collection

Participant observation: 25 full days, 60 machine-written pages of field notes		
Everyday work practices of explorationists and data managers		25 full days of observations and informal conversations with explorationists and data managers in their offices, around coffee tables, and during their lunch breaks
Participation in workshops, meetings, and seminars		Six one-day events with 19 participants altogether (explorationists, data managers, process owners, and IT management)
Semi-structured interviews: 27 interviews, 45-90 minutes each, transcribed		
15 explorationists		Geologists, geochemists, and geophysicists
12 data managers in exploration		Project and central data managers
Documents: electronic and paper-based		
Internal	Documents	Descriptions of routines and work practices, manuals for tools, example prospects, internal reports and memos, meeting minutes, overviews of challenges with querying databases, and presented results
	MS SharePoint team sites	Project reports, discussions, and slide presentations
Public	OilComp information	Drilling operations, tests, and results, reclassified interpretations, and reports on discoveries and recoverable volumes
	Norwegian Petroleum Directorate FactPages	NPD FactPages (http://factpages.npd.no/factpages/) contain information regarding petroleum activities on the Norwegian continental shelf (see example in Figure 3). The information is synchronized with the NPD's databases on a daily basis.
	Diskos Database	The Diskos National Data Repository (NDR) is Norway's national data repository for petroleum data. Its index is open to the public and contains, in principle, all geodata (seismic, well, and production). Figure 1 provides examples.

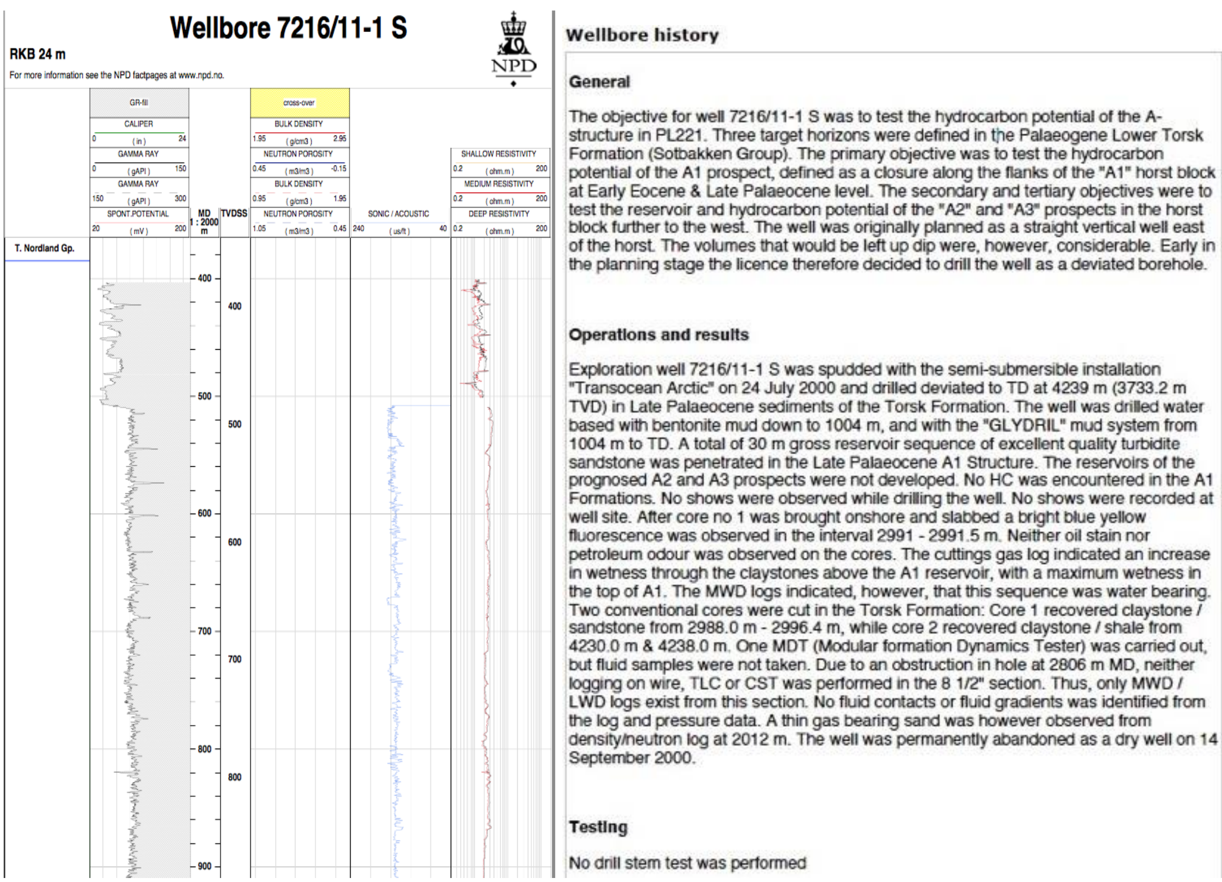


Figure 3. Example of a Wellbore Log and Wellbore History from NPD FactPages

3.4 Data Analysis

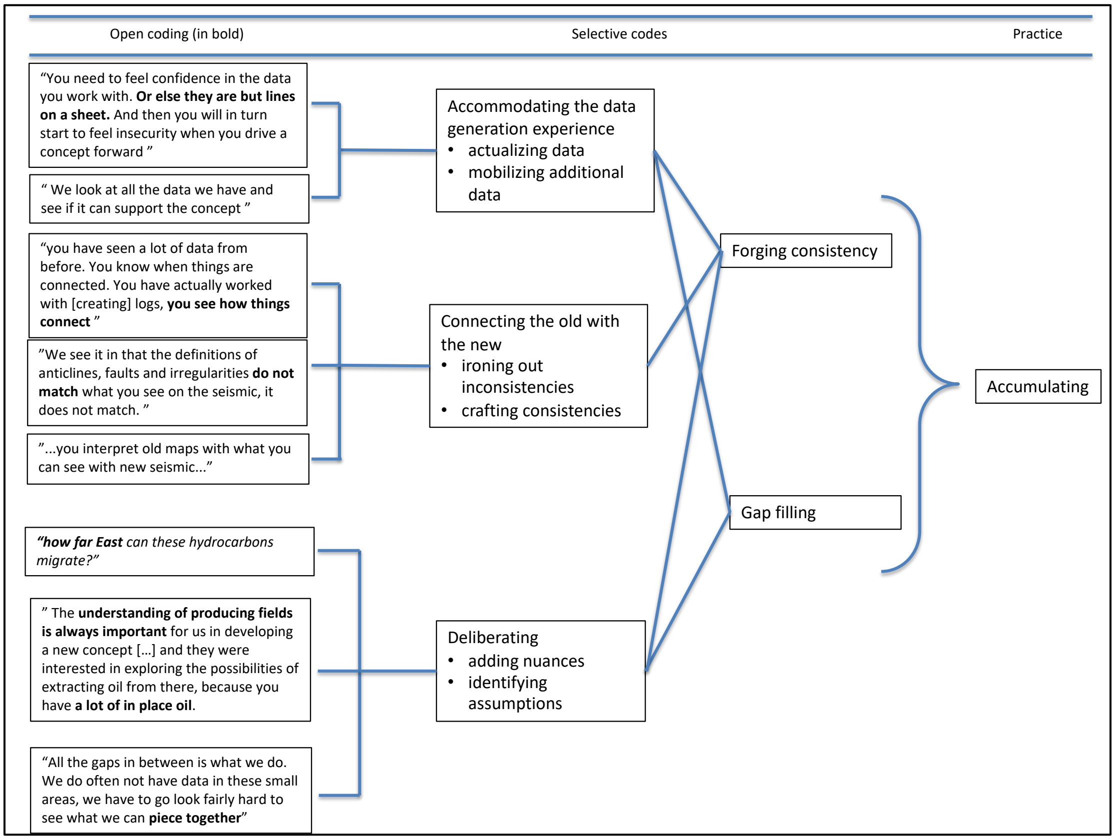
Our data analysis process was interpretative (Walsham, 1995), as we sought to capture the perceptions on data actualization from the actors involved in exploration. Data analysis was iterative. Data collection overlapped with data analysis, thus granting us the flexibility to continuously consider our partial interpretations toward a gradually expanding amount of data and to refine our interpretations together with the actors in the case (Klein & Myers, 1999). Our data analysis may be reconstructed into four main stages.

The first round of coding was open-ended (Wiesche et al., 2017). We sought an overview of relevant actors, organizational routines, and prominent concerns within the unit of exploration. The first author, who collected most of the data, immersed himself in the data. During this round, we attached initial labels, consisting of sentences and paragraphs capturing salient aspects of the work of explorationists. An example of a label is "well and seismic data turn indications into leads" (with empirical excerpts, e.g., "The process is such that you get many ideas when you are looking at the seismic. 100 ideas, perhaps 200. Beneath them we have different branches, and out of them we have been able to concretize approximately 100 to something that can be a precursor to a prospect, what we call a lead, that means, we set a polygon on a map, you can define an area, you can calculate a volume, and you can calculate probabilities", and "We use well and seismic data. It is very rarely gravitational data and magnetometric data, here the level of detail is poorer"). We manually coded data using word processing software. We used bold text for descriptive codes and entered the data under the codes. We collected descriptive codes and illustrative empirical data in tables, developing 432 descriptive codes.

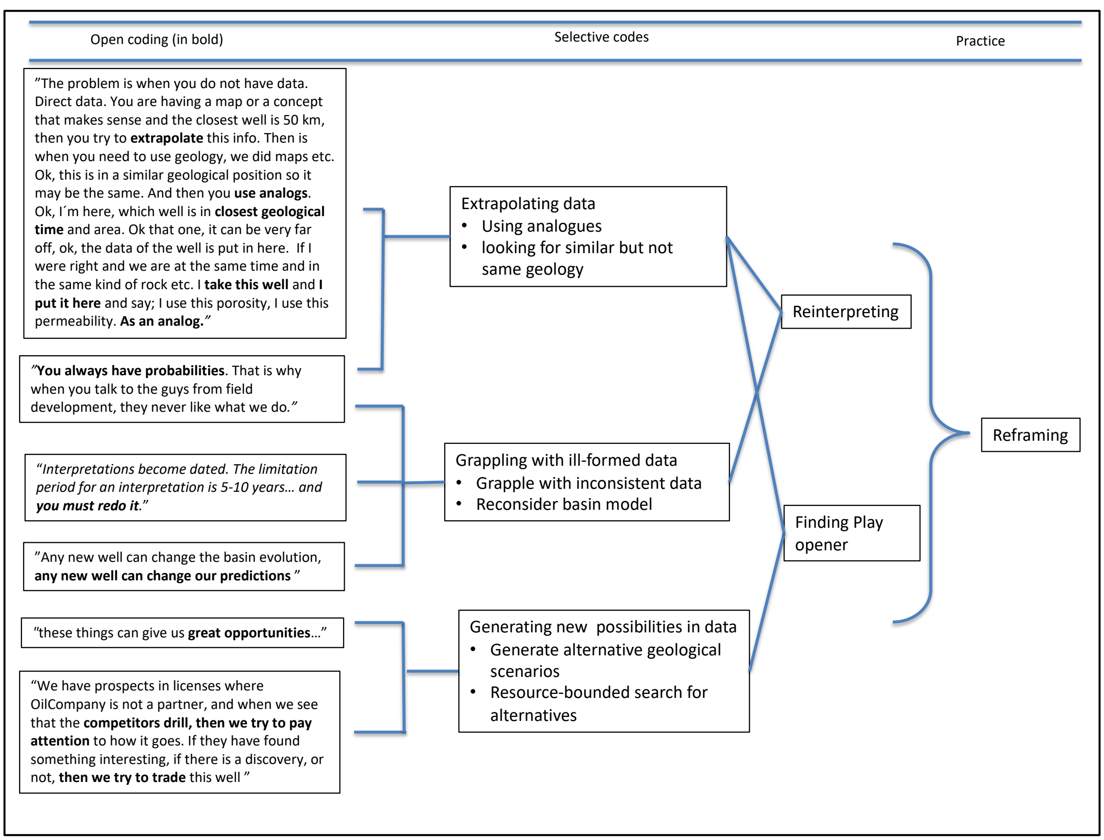
The second round of analysis involved both authors. In this round, we focused on the work practices of explorationists as a particular category of data (Wiesche et al., 2017), looking closer at their use of digital tools, in general, and the sense in which they worked with data, in particular. Using the tables with the open codes from the previous rounds as a starting point, we selectively coded the data. We created a 10-page data analysis document with selective codes and empirical examples. This document served as a basis for discussion between the authors but also with our research group to refine our understanding by challenging our preliminary interpretations. Through 16 iterations, where codes and data were compared and challenged, we gradually unpacked the professional community of explorationists, comprising a heterogenous set of more than ten professional disciplines with different roles and tasks. The heterogeneity of professional disciplines under the umbrella term "explorationist" is mirrored in the extensive list of specialized digital tools used by the explorationists. There are corporate databases for seismic data (including navigation data, faults, horizons and grids), well data (such as drilling data, well logs, geochemical analysis, and core sample images) and production data (volume, pressure, temperature). The three principal data types used by explorationists are illustrated in Figure 1. In addition, there are extensive, public repositories for all oil activities (exploration, drilling, production, maintenance) on the Norwegian continental shelf (see Table 1), a feature of the political and institutional history of North Europe distinctly different from that in North America. We developed an understanding of the collaborative practices within the units in which explorationists work, but also learned why and when they interact with outside specialists or management. We analyzed how explorationists refine and ensure quality control on their predictions i.e., prospects.

The third round focused on *how* explorationists implicate data in their work practices, an underresearched theme in IS research on data science (Günther et al., 2017; Sivarajah et al., 2017). Through a form of memoing (Wiesche et al., 2017), we wrote up concepts, categories, and the relationship between them. We generated, over the life cycle of prospects for oil reservoirs, visual illustrations and tables depicting how the chronological development of a prospect occurs, including the different units involved, their main goals, the data used, the digital tools used, the actors involved, their assumptions, their evaluation procedures—i.e., an overview of the data that explorationists consume and produce at certain times in the exploration process, and how they use it.

In the fourth round, we engaged with theoretical imports to conceptualize patterns of work practices detailing how data are actualized in oil exploration. Anything but clean slates (Suddaby, 2006), our prior experience influenced our analysis. Our long-standing interests in work practices, knowledge work, and organizational change served as resources. This round may be characterized as largely inductive but with formative, deductive injections. Early on, we were puzzled by the sparse and underdetermined data on which exploration is based: logging from a few wells are literally presented like pins on a vast map together with coarse-grained seismic images. How were these data actualized (Knorr Cetina, 1999) in the work practices of explorationists, we asked ourselves. Clearly, it was not because they "faithfully" represent (geological) reality (Burton-Jones & Grange, 2012; Zuboff, 1989). A principal reason, we inductively found, was the gradual support through supplementary data (see Leonelli, 2014).



Open coding (in bold)
Selective codes
Practice
"You need to feel confidence in the data you work with. **Or else they are but lines on a sheet.** And then you will in turn start to feel insecurity when you drive a concept forward "
" We look at all the data we have and see if it can support the concept "
"you have seen a lot of data from before. You know when things are connected. You have actually worked with [creating] logs, **you see how things connect** "
"We see it in that the definitions of anticlines, faults and irregularities **do not match** what you see on the seismic, it does not match. "
"...you interpret old maps with what you can see with new seismic..."
*"**how far East** can these hydrocarbons migrate?"*
" The **understanding of producing fields is always important** for us in developing a new concept [...] and they were interested in exploring the possibilities of extracting oil from there, because you have **a lot of in place oil**.
"All the gaps in between is what we do. We do often not have data in these small areas, we have to go look fairly hard to see what we can **piece together**"
Accommodating the data generation experience
• actualizing data
• mobilizing additional data
Connecting the old with the new
• ironing out inconsistencies
• crafting consistencies
Deliberating
• adding nuances
• identifying assumptions
Forging consistency
Gap filling
Accumulating



Open coding (in bold)
Selective codes
Practice
"The problem is when you do not have data. Direct data. You are having a map or a concept that makes sense and the closest well is 50 km, then you try to **extrapolate** this info. Then is when you need to use geology, we did maps etc. Ok, this is in a similar geological position so it may be the same. And then you **use analogs**. Ok, I´m here, which well is in **closest geological time** and area. Ok that one, it can be very far off, ok, the data of the well is put in here. If I were right and we are at the same time and in the same kind of rock etc. I **take this well** and **I put it here** and say; I use this porosity, I use this permeability. **As an analog.**"
"**You always have probabilities**. That is why when you talk to the guys from field development, they never like what we do."
*"Interpretations become dated. The limitation period for an interpretation is 5-10 years... and **you must redo it**."*
"Any new well can change the basin evolution, **any new well can change our predictions** "
"these things can give us **great opportunities**..."
"We have prospects in licenses where OilCompany is not a partner, and when we see that the **competitors drill, then we try to pay attention** to how it goes. If they have found something interesting, if there is a discovery, or not, **then we try to trade** this well "
Extrapolating data
• Using analogues
• looking for similar but not same geology
Grappling with ill-formed data
• Grapple with inconsistent data
• Reconsider basin model
Generating new possibilities in data
• Generate alternative geological scenarios
• Resource-bounded search for alternatives
Reinterpreting
Finding Play opener
Reframing

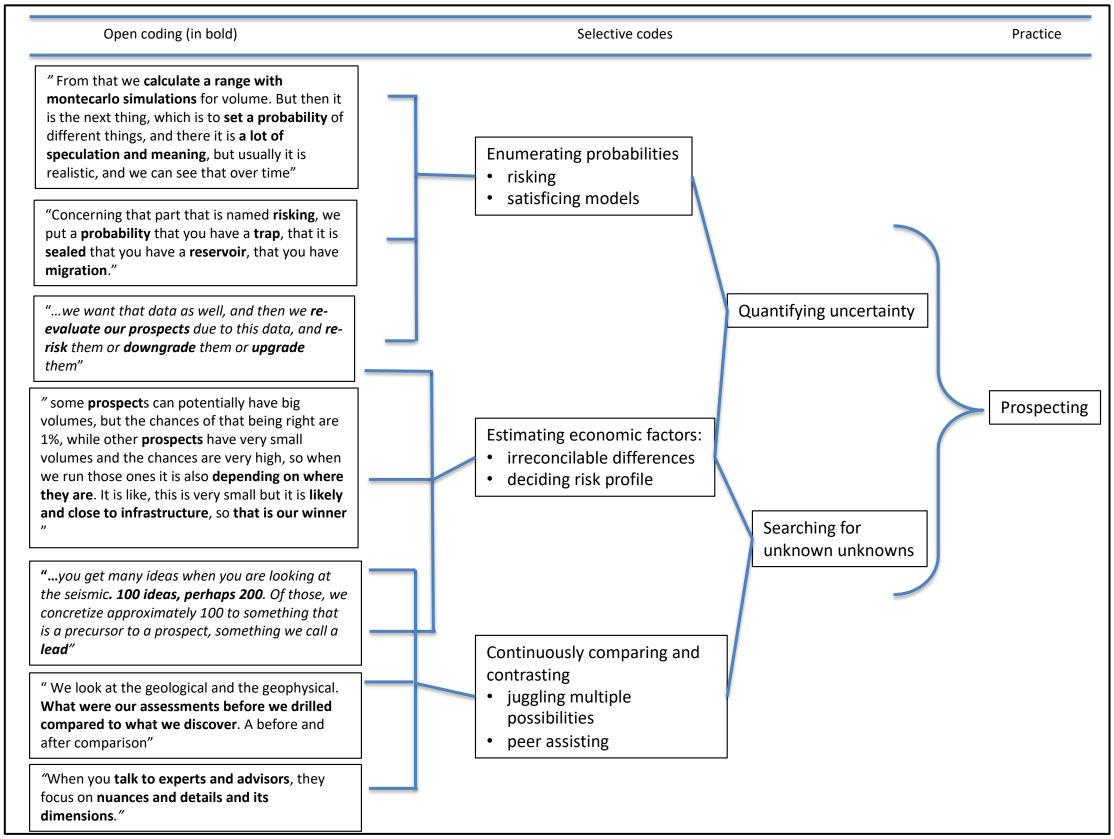


Figure 4. Summary of our Interpretative Template with Three Patterns of Work Practices (Main Constructs) Supplemented with Underlying Codes and Empirical Excerpts

This conservative, cumulative process is captured by the first of our three identified patterns of work practices, *accumulating*. It empirically represents the dominant pattern of working. However, the continuous work of accumulating is occasionally interrupted. Similar to pragmatic action, our second practice, *reframing*, captures how new geodata regularly contest prevalent models and interpretations.

Our third practice, *prospecting*, emerged to capture the strikingly provisional nature of the geological interpretations underpinning the oil prospects. In our experience, agreement was never realized. At every juncture, competing interpretations were voiced and mobilized. However, since OilComp is not a debate club, in the absence of anything close to agreement, provisional agreement was forged for the purpose of deciding what to do (Kellog et al., 2006; Mol, 2002; Oborn et al., 2011). Multiple interpretations are the norm and "agreement" is but a provisional arrangement to solve the what-to-do-next imperative of pragmatic action.

Figure 4 summarizes the interpretative template resulting from our process of data analysis. It outlines our three patterns of work practices with concepts aggregated from the coding as well as illustrative empirical excerpts. In the next section, we organize our findings around these three patterns of work practices.

4 Findings

4.1 Accumulating: Gathering Organizationally Credible Evidence

A team of explorationists works for extended periods of time, typically years, with its "prospects," i.e., its candidates for yet to be discovered oil reservoirs. A prospect is essentially a prediction about an oil reservoir's location—including risk, volume, and value estimates—should the prospect be considered for drilling at a later stage. Working in co-located teams organized in the same corridor of an office building that is part of OilComp's headquarters, explorationists' everyday work routines revolve around refining interpretations of data that justify prospects, interweaving individual work with informal discussions between fellow team members. Few bother to close their office doors, making it easy to drop by for consultations about issues that come up. Small groups of explorationists regularly engage in informal discussions around the whiteboard in someone's office.

For long stretches of time, explorationists work in areas with *proven plays*—for example, areas where they already have producing fields. Producing fields creates assumptions—they know what they are looking for. In a given area, a *proven play* is a particular geological configuration in which necessary (but not sufficient) conditions for an oil reservoir—i.e., source rock, migration paths, and traps—are known to exist (see Section 3.1). Working with a proven play, exploration focuses on *gaps*, i.e., areas where discoveries have not yet been made, but where they believe hydrocarbons may have migrated. As an exploration team leader explained, “in this area [pointing to his screen], we knew that in the southern [name of the basin], which in this case is 250 kilometers north-south, a lot of hydrocarbons have been generated. So, how far east can those hydrocarbons migrate?”

However, turning opportunities into more credible prospects involves actualizing data as evidence that a gap may contain commercially interesting amounts of hydrocarbons. Working with a proven play implies that large amounts of well and seismic data are available (namely all historic discoveries regarding that play). Dealing with the vast amount of available, sensor-based IoT data, the team of explorationists mobilize data and thus accumulate evidence that its prospect fits the proven play. One of our informants was struggling in front of his screen using a petrophysical analysis tool filled with well-log data. The old well data he had available did not support his prospect. Was he wrong or were the data wrong? The data dated from when the well was originally drilled back in the 1970s. As he explained to us, the knowledge that injecting mud into the borehole while drilling influences the temperature readings in the well “was learned only in the 1980s.” Instead of simply contradicting his prospect, he assigned a mark signifying that the data were of low quality to indicate their lack of relevance. To be able to do this, one explorationist explains, experience from parts of exploration work is necessary, e.g., creating well logs: “You need to feel confidence in the data you work with. Or else they are but lines on a sheet. And then you will in turn start to feel insecurity when you drive a concept forward.” Thus, the work practice implicated in accumulating evidence includes ironing out contradictory data.

An important task in determining a prospect’s credibility is to see whether assumptions made fit with several types of data, i.e., what effectively corresponds to a form of triangulation. So-called “well tie-ins” are a particularly important way in which this triangulation operates. Digital interpretation tools are used to determine the relationship between boundaries in the well logs and seismic reflections, consequently producing a relationship between the well logs (measured by depth) and the seismic reflections (measured in time). A well tie-in is an effort to find consistency between the broad but crude overview provided by seismic data and the much more detailed well data that come from measurements from the specific, pinpointed location of an oil well. Perfect consistency between seismic and well data occurs rarely, if ever. Consistency is crafted through well tie-ins, which are labor-intensive endeavors. We sat down with one of the explorationists tasked with a well tie-in and visually superimposed well data onto seismic data (see Figure 1). “We see it in that the definitions of anticlines, faults, and irregularities do not match what you see on the seismic, it does not match,” he noted. He did not despair, however, as this could be explained: “It matters how old the wells are, what types of data were collected, how far away the wells are. If they are close, that is obviously beneficial.” He keeps working. The inconsistency between seismic and well data is compounded by the fact that they are measured with different scales: Seismic data are measured as the time that it takes an echo of a particular sound wave to travel back to the sensors after being refracted by subsurface rocks, while well data are measured relative to the depth (distance in meters) in the well where they were recorded. As the speed of acoustic waves differs with different types of rocks, the time-to-depth conversion is nonlinear and is about gaining a sense of subsurface “nonconformities.” Nonconformities result from geological processes such as fault lines stemming from earthquakes. As our informant explained, “if you have very steep non-conformities, [the nonconformities] can jump several hundred meters back and forth from time to depth.”

4.2 Reframing: Contesting Prevalent Predictions

The account above is one of continuity in the sense that explorationists gradually accumulate evidence to support a prospect (prediction) in a given geological formation with a proven play. Thus, the actualizing of data amounts to crafting the fitting of data to one’s predictions or ironing out inconsistencies. The explorationists painstakingly fill in the gaps to back up their leads and prospects, with data marshaled into (more) organizationally credible evidence. Explorationists “stretch” their data, seeking to strengthen rather than defeat their predictions. This essentially conservative approach is at times punctuated by new data that can neither be accommodated nor dismissed, which leads us to our second pattern of work practices.

With a bit of drama, one explorationist exclaims, “Any new well can change the basin evolution; any new well can change our predictions!” He knows that a change in the basin evolution is a radical change. The basin model that the explorationists rely on when searching in an area is effectively the prevailing understanding of that area’s geological history—the result of extensive efforts—and represents significant sunk investment in terms of earlier work. One key activity in substantiating a prospect is modeling ("basin modeling") the history of the area's geological evolution. Once a basin model is conceptualized, it is tested against existing well data for consistency. However, in practice, consistency is never fully achieved. Working on a basin model, one explorationist explains how the team selects 200 reference wells out of a sample of 1,000 wells to support this consistency check. Well data is inconsistent, so they use heuristics such as the well's age (assuming new wells have better data quality than older ones), and then consider how much work went into calibrating the data, noting that often, "we must go in and calibrate the well to the seismic [i.e., well tie-in]. And if it is a bad calibration, if things do not match, then the logs are poorly collected." Poor quality can be tied to a variety of reasons, e.g., "things that happened on the rig that are not documented well enough, that give a sloppy [well] log."

In a similar vein, another explorationist recounted his struggles with deeply inconsistent data on a field. The seismic data indicated that there should be sand throughout the field, but the well data told a different story:

> *I have a well here [pointing to the screen] that hits sand, and I have a well here [pointing] that does not hit sand. And then I have a seismic processing [pointing to another location on the screen] that shows me it should be sand all over. Then I need to decide: No, that [pointing] is not sand; this [pointing] is sand.*

To account for different probabilities, data sometimes need to be extrapolated from geographic areas that are less known into geographic areas that are more well-known geologically—for example, because they have drilled more there, or shot more seismic surveys. Data then are extrapolated, as one explorationist explains, talking about a well:

> *Ok that one, it can be very far off, ok, the data of the well is put in here. If I were right and we are at the same time and in the same kind of rock etc. I take this well and I put it here and say, I use this porosity, I use this permeability. As an analog.*

In the world of explorationists, data are king. You can make the most elaborate models, but, if they do not hold up against the data, they never make it into everyday work practices. One explorationist warns:

> *I do not fall in love with my models, I mean, they are wrong by definition. Some people get really personal, and if new data goes against it, they try to go all around to try to avoid the data. If the data go in an opposite direction of your concept, it is better to just kill it.*

Explorationists have an unquenched thirst for new data. Well data, with its fine-grained measurements, are particularly appreciated. With coarse-grained seismic data, well data are the closest that explorationists come to "hard" evidence. Given the considerable financial costs of drilling new wells, OilComp invests in the drilling of a few dozen wells in a typical year in the area reported on in our case study. The explorationists' thirst for new well data makes them cut corners in formal procedures. Rather than use the formal, time-consuming process of quality-control data from an ongoing well-drilling operation lasting a month or two, they import the data directly from the drilling database. Two purposes motivate explorationists' keen interest in new well data. First, they provide immediate validation: Was there an oil discovery as they had predicted? However, new data also provide a much valued resource for considering unproven plays and alternative geological scenarios: "When we have a new well, it is not like we do not care anymore [whether there was oil or not]," one explorationist told us, "we use it for future exploration … I care about the data. Data from the well is key."

The operational reality of operating within a highly competitive business environment is internalized among the explorationists. Working with prospects is always resource bounded; thus, they are neither exhaustive nor perfect. Searching for new opportunities in the form of unproven plays is no exception. The resource-demanding nature of assessing the credibility of new concepts (unproven plays) forces settling for good enough, rather than elaborate assessments:

> *We often do not have time to work out all [the concepts]; it takes too much time. We very often have limited time to drive concepts forward. It can be a matter of a few months. During that time, a lot of data must be pieced together, [and] a model needs to be built and to run basin simulations. In sum, it is a bit hard.*

4.3 Prospecting: Cultivating Alternatives

The essentially abductive processes described above depict the practical constraints on resources (economy) that regulate and format explorationists' work, underscoring that prospects are satisficing, i.e., good enough to comply with the institutionalized decision-making process. The explorationists comply to produce the required input. Explorationists estimate the amounts of oil that a well might produce. In one case, volume was estimated from variables such as rock porosity (estimated with well logs and/or core samples), oil saturation in the rock (estimated using electrical resistivity well logs), and the recovery factor (estimated from reservoir permeability and oil viscosity). As a means of bracketing uncertainty, the team performed

Monte Carlo simulations (i.e., a statistical approach to risk analysis in which numbers are selected from likely ranges of input data and through iterative calculations to determine a range of probable outcomes) using risk-assessment software. However, the version fed to management as part of formal decision gates radically undercommunicates the prevalence of the multiple, competing possibilities known to the explorationists. For purposes of arriving at a managerial decision, the extent and role of multiple, divergent predictions are bracketed. However, among the explorationists there is a healthy interest in entertaining multiple possibilities at the same time.

Earlier, we emphasized how highly explorationists regard well data—the closest they have to "hard" data. This should not be misconstrued as suggesting that explorationists trust well data at face value, as they regularly provide deeply ambiguous results that feed divergent possibilities, neither of which can be put to rest by the data themselves. Despite the presence of "big" data, the explorationists' prospects (predictions) are underdetermined by the data. One explorationist illustrates the dilemma. One aspect or type of well data is the analysis of hydrocarbons' chemical compositions. Hydrocarbons from different oil reservoirs have varying compositions. Each has a unique chemical profile that allows for the discernment of hydrocarbons from two different oil reservoirs. Normally, one would assume that two wells in close proximity to each other would draw from the same oil reserve. The explorationist is puzzled:

> *In one of the fields in our area, each well is different when it comes to the [origins of the] hydrocarbons. They have different chemical compositions, which is really strange. They are so close by, you would think they are all the same, but they are not. [The geology] is very complex in some areas.*

What he refers to with "complex" geology, is that there are multiple irreconcilable—given available data—interpretations depending on the assumptions about the geological history (erosion, faults). This ambiguity or multiplicity is not so much resolved as relegated to a nagging uncertainty that, in later situations, may turn into a salient rather than a latent possibility.

Digital tools for managing and interpreting well logs, processing and interpreting seismic data, doing seismic well tie-ins, plus basin modeling and simulation are crucial to explorationists' ability to juggle multiple possibilities. Increased computing capacity and new digital seismic processing and interpretation tools make the creation of 3D seismic cubes (3D seismic data sets made from other multiple seismic data), which was previously prohibitively time-consuming, more practical. One explorationist told us over lunch how he was working in a field with 200 variants of the same 3D seismic cube. From the outside, there was no way of knowing the purpose of all 200 variants. The one officially quality-controlled variant shed little light on the other 200. As he was interested in a particular subsurface level in the project, he looked into it. Perhaps there was an underlying implicitly assumed idea that he had missed. He asked himself: "What was this idea? Why? It is not apparent in that 'pick' [their term, implying interpretation of a subsurface of the seismic level in light of subsurfaces picked from well data]. You have some new data that do not fit. How is it connected?"

Coping with multiple possibilities is fundamentally collective. In formal discussions, but more often and more importantly in informal, peer-based discussions, explorationists collectively deliberate multiple possibilities: "When you talk to experts and advisors, they stress the nuances, and the details in it, and especially the dimensions in it." Peer-based discussions are vital to avoid tunnel vision that working strenuously with a prospect easily might create. As one explorationist confessed, after a while, "you begin to think [your work] is great, [so] we have to drill it."

The prevalence of multiple and radically different interpretations is internalized by explorationists as part of their professional identity. However, institutional constraints make it organizationally and politically necessary at times to bracket this inherent multiplicity. Multiplicity is not resolved or eliminated as much as put temporarily on hold for purposes of passing one of OilComp's decision gates. The task of "risking" (their term) a prospect is illustrative. *Risking* is the quantification of qualitatively manipulating the prospect. One explorationist commented: "Concerning that part that is named risking, we put a probability that you have a trap, that it is sealed that you have a reservoir, that you have migration." Crucially, quantified measures for variables such as rock porosity and permeability, oil saturation, viscosity, and volumes are assigned with each prospect. Despite estimates, risking contains "a lot of speculation and (subjective) opinions," but still necessarily legitimizes OilComp's gated decision processes. The problems with quantifying the probabilities of an oil discovery for different prospects under consideration are particularly pronounced for those with medium-range probabilities, i.e., 10-25%: "Here we are struggling. They diverge in all directions."

5 Discussion

Actualizing data's representational capacity into everyday work processes mired in data uncertainty and ambiguity tests "the limits of meaning" (Weick, 1985, p. 64). Sensor-based IoT data, thoroughly manipulated algorithmically, do not "faithfully" mirror the physical conditions of the geology, i.e., data might easily become mere signs or symbols with little or no relevance to explorationists' work practices. Thus, how do data acquire meaning in the sense of being woven into work practices (i.e., by being actualized)? We discuss the three patterns constituting data-centric knowing relative to existing literature. We emphasize the social and material conditions of data-centric knowing. In addition, we recognize the broader institutional fabric in OilComp embedding the three patterns of practices of data-centric knowing.

The data that inform explorationists' work practices are what Knorr Cetina (1999) identifies as signs and Bailey et al. (2012) as symbols. This makes the problem of "referential attribution" immediate (Kallinikos, 1999). Data are " 'footprints' of [physical] events, rather than ... the events themselves" (Knorr Cetina, 1999, p. 41). They are, in our case, "footprints" of physical geology that are mediated (hence distorted) by sensors and algorithmically manipulated to make their correspondence with the originating geology anything but "faithful" (see Monteiro & Parmiggiani, 2019). Consider the two types of sensor-based IoT data that are by far the most important to explorationists' work practices: seismic and well-log data. From a campaign of seismic shooting, less than 1% of the data is kept. The remainder is removed through a variety of nonlinear mathematical filtering techniques. The seismic data actually used in explorationists' work practices are accordingly a fraction of available data, and mathematically filtered (i.e., algorithmically manipulated). Similarly, well-log data are generated from sensors that capture radioactive radiation, electromagnetic conductivity, and electrical resistivity by lowering measuring equipment down into the drilled well. Particular patterns of values in the sensor-based IoT readings are "footprints" of geophysical properties as, for instance, gamma radiation is higher in shale than in sandstone, and electrical resistivity is higher in oil than in water (see Bowker, 1994).

With data never being stable—they are transient, dynamic, contingent, and subject to aggregation, slicing, and other manipulation (Kallinikos et al., 2013)—there is, by implication, an "inherent epistemic uncertainty" (Alaimo et al., 2020) that makes crucial the question of *how* the weaving of data into work practices unfolds. The chronic dilemma that explorationists grapple with is how the data become more than mere symbols, and how data are drawn into consequential action and decision making. Our three patterns of work practices—accumulating, reframing, and prospecting—as demonstrated in the previous section, are key. We discuss how our three patterns resonate with existing literature.

Our accumulating pattern captures the constant hum of mundane work that goes into making the data amendable and accessible *as* data. Scholars have pointed out what Edwards (2011) calls "data friction," which includes washing, calibrating, and slicing up data (see Leonelli, 2014). There is ample evidence of data friction in our case as well (e.g., the efforts involved in "well tie-ins" or quality-assuring data). However, because of the particular epistemic uncertainty of data, there is a considerable amount of work involved in *supporting and triangulating one* kind of data by connecting and hence grounding it relative to other supporting data. In isolation, data literally are a symbol. As Morgan (2010, p. 4) points out, "we depend upon systems, conventions, authorities and all sorts of good companions to get [data] to travel well." By supporting and confirming them, data are grounded in additional data similar to the way triangulation works (see Weick (1985) underscoring the importance of triangulation in all human sensemaking). How, then, does this form of accumulating confirming otherwise fragile data work? A principal manner is that data are assigned *different* levels of epistemic uncertainty. All data suffer from epistemic uncertainty, but some more acutely than others (see Chang, 2004; Østerlie & Monteiro, 2020). In our case, well-log data are viewed as more reliable than seismic data because of better resolution. An important way to back up otherwise inconsistent (thereby potentially dismissed) data is by connecting them to other types of data, as pointed out by Kallinikos (1999). In our case, this is illustrated when each of the relevant seismic sections is carefully connected to neighboring wells' fine-granular well-log data, a manual process known as well tie-in (see details in the previous section).

The accumulating pattern's modus operandi is that of confirmation, i.e., the conditions and processes for supporting data that otherwise risk being mere symbols. The efforts that go into supporting and triangulating data are investments that risk creating path dependencies. Our second pattern, reframing, addresses the purposeful contesting of accumulated (i.e., supported and triangulated) data. The pattern of reframing accordingly addresses situations at the boundaries of the accumulating work pattern's reach. Efforts covered by the accumulating pattern to iron out wrinkles, inconsistencies, and outliers in the data are attainable only up to a certain level. The pattern is contested by the arrival of a new type or new data set triggering and abductively searching for new ways to make sense of all the data, new and old. In our case, this amounts to coming up with a new geological narrative in the form of a sequential process of shifting tectonic plates, up- and down-lifting, and erosion and faults—conditions that make hydrocarbons plausible (see Wylie, 2002, who argues for the importance of a narrative understanding in archaeology). The drilling of a new exploration well—with new, highly valued well data—as illustrated in the previous section ("any new well can change the basin evolution"), provides a possibility to challenge the entrenched, path-dependent understanding that results from the accumulating work pattern.

Abductively challenging the accumulating pattern's everyday hum, the search for new interpretations and geological models is, as already emphasized by Peirce (1931, p. 5.600), regulated and bounded. The work pattern is not an open-ended search, but rather is bounded by time and resource constraints aimed at a plausible, imperfect solution, as pointed out by Lyytinen and Grover (2017). In our case, new data come with a hefty price tag. Well data from drilling, in particular, but also from new seismic surveys, represent significant economic investments. The abductive nature of the work pattern of reframing is, accordingly, directional and goal-seeking. An example is provided by Dunne and Dougherty (2016, p. 132), who demonstrate how scientists in the biopharmaceutical industry apply abductive reasoning as a "deliberate and methodological" social process to "navigate in the labyrinth" of drug innovation. Adding to such studies that point out how clues enable practitioners working abductively to "conceive of a whole design almost at once" (Dunne & Dougherty, 2016, p. 151), we find that explorationists challenge clues by rubbing them up against historical data. Leads or clues, in the commercial environment of OilComp's explorationists, come with economic returns and risk, because, as one regional explorationist noted, "we are in competition with thousands of others, and there are many other very skilled geologists and geophysicists around who have seen and are aware of all the well-known stuff." Reframing is thus generative in the sense that it involves being able to justify the unknown—potentially with bigger economic returns—and is simultaneously limited because it requires necessary backing in historical data, i.e., more well-known data.

As a direct consequence of their disconnect with their originating, physical referents (Alaimo et al., 2020; Borgmann, 1999; Kallinikos, 1999; Kallinikos et al., 2013; Knorr Cetina, 1999), data as signs or symbols come with inherent epistemic uncertainties. Both of the preceding patterns of work practices are aimed at eliminating or regulating this uncertainty: The accumulating pattern covers practices of conservatively supporting and triangulating otherwise vulnerable interpretations of data, while the reframing pattern covers the abductive refactoring of an earlier interpretation triggered by the arrival of new data that are incompatible with the old. However, our third pattern of work practices is different. It addresses how explorationists, rather than trying to eliminate epistemic uncertainty, cultivate and encourage a multiplicity of interpretations of data, thereby embracing rather than resolving epistemological uncertainty while simultaneously avoiding halting operational decisions. In contrast to both of the preceding patterns of work practice, prospecting has not been addressed in the context of data and datafication.

To develop the prospecting work pattern, it is helpful to compare it to the notion of *search* (Stark, 2009). Explorationists search for something not yet recognized as a category. In our case, the well-defined search for proven plays (i.e., geological configurations demonstratively yielding hydrocarbons in the areas under scrutiny) could, and regularly does, spill over into the ill-defined search for unproven plays (i.e., potential, but not yet demonstrated, geological configurations for yielding hydrocarbons). The data radically underdetermine this search (see Loch et al., 2011). The efforts covered by the former two patterns of work practices bracket the uncertainty, but it may resurface.

The prospecting work pattern also fills productive, organizational roles, as it is "through divergent or misaligned understandings that problematic situations can give way to positive reconstructions" (Stark, 2009, p. 192). Keeping an eye open for new, unproven plays is crucial for OilComp. Ambiguity exists over whether the data at hand support proven plays or might, in fact, indicate something radically different. The multiplicity of interpretations is regulated in what is fundamentally a *collective:* They are played out, deliberated, and regulated in collective arenas that result in partial agreements (see Oborn et al., 2011). In our case, the formal and informal peer-based feedback sessions manifest this collective, in ways similar to how medical physicians are trained to always be open to secondary, alternative diagnoses in their treatment of patients (Timmermans & Berg, 2000).

The prospecting pattern underscores how consensus is never arrived at, but rather is worked around in temporal and local arrangements, resonating with Mol's (2002) study of how medical specializations such as surgery and pathology—despite radical differences in routines, theories, vocabulary, and instruments—forge temporary agreements about how to treat atherosclerosis in patients. Underscoring, as we have done, the productive role of the prospecting pattern begs the question of how the practical problem par excellence is decided, i.e., what to do next. As Kellogg et al. (2006, p. 38) point out: "Instead of … shared meanings and common knowledge, organizational actors juxtapose their diverse efforts into a provisional and emerging collage of loosely coupled contributions." In short, the pattern of prospecting does not undercut decision-making and action-taking, but temporal, partial arrangements, rather than full-fledged consensus, are required.

The three patterns discussed above that comprise data-centric knowing flesh out the work practices of explorationists. As scholars have pointed out (Kallinikos, 2004), practice-oriented perspectives risk becoming near-sighted in the sense of downplaying broader historic and institutional contexts that go beyond the "here and now" (see also Monteiro et al., 2014). We thus discuss tenets of the institutional fabric that underwrite data-centric knowing. This is in line with Porter (1995, p. 44; see also

Poovey, 1998) who argues that meshing data with institutional routines is central to their actualization: "Given the ways that [data] measures can be undermined … we may doubt that they correspond to anything in the world. But a plausible measure backed by sufficient institutional support can nevertheless become real."

A key element of the institutional fabric embedding of data-centric knowing in OilComp is the compliance with a formally defined, sequential process known as a "funnel" (see Figure 2). The increased level of formal requirements for data as decisions move through OilComp's decision gates come with markedly ritual connotations. As one explorationist laughingly put it, "all our models are wrong," an insight that explorationists often believe has been lost on "the guys upstairs [management]." In the funnel, when zooming in on candidates for commercially viable oil reservoirs ("prospects"), a need exists to manage the *portfolio* of prospects. There are typically tens if not hundreds of prospects, so the operational concern at this decision gate is how to prioritize efforts when pursuing prospects. Briefly, prioritization relies heavily on the quantification of prospects' estimated risks, costs, and revenues. However, as Porter (1995) convincingly documents, trust in numbers comes at one's own peril, as numbers hide as much as they reveal. Nevertheless, the conflation of rich geological interpretations with quantified estimates feeds business operational needs for making decisions (Scott, 1998), as illustrated by the use of Monte Carlo simulations and so-called "risking" (see preceding Findings section).

To summarize, the crafting of institutional facts at OilComp is not a steady march from uncertain, error-prone data to solid facts. What we see is a formalized and sequential process in which, for operational needs to move forward, the epistemological uncertainty of data is *provisionally bracketed* to reach a decision. However, uncertainty is never eliminated. Away from the formal decision gates at the managerial level, professional deliberations persist and thrive among explorationists. Data engage, as it were, in two different language games: the explorationist community vs. the managerially governed decision gates (Mol, 2002; Oborn et al., 2011).

6 Conclusion

A program launched decades ago (Knorr Cetina, 1999; Zuboff, 1988), the problem of "referential attribution" (Kallinikos, 1999)—how data that lack any immediate correspondence or similarity with physical objects, processes, or qualities acquire meaning by being woven into everyday work practices—is gaining empirical relevance and significance with ongoing datafication of our lifeworld. Key to this project is empirical grounding (Günther et al., 2017; Lyytinen & Grover, 2017). Clearly, our articulation of the three patterns of interwoven work practices constituting data-centric knowing emerges from a particular case study. What, then, is the relevance of our analysis to other empirical domains? As Kallinikos (1999, p. 289) points out, "the project … needs, therefore, to pass through both the investigation of other empirical contexts and even involve the more successful integration of the relevant available literature."

In translating the analysis underpinning data-centric knowing to other domains, the concept will inevitably be appropriated and hence modified. Theoretical concepts travel via not despite appropriation (Walsham, 1995). Our case is characterized by three salient aspects that significantly shaped our analysis: inherently uncertain data and interpretations, quasi-scientific approaches, and a corporate, operational logic. We expect generalization qua translation to other domains with similar characteristics. (1) *Inherent uncertainty*: other domains also evolve around uncertain data—e.g., security analysts tasked with predicting the value of stocks to investors deal with inherent uncertain data-driven interpretation (Beunza & Garud, 2007) or Gartner group's analysts predicting future industry trends (Pollock & Williams, 2016). (2) *Quasi-scientific communities*: the search for and openness toward the unknown, captured in our prospecting pattern, is likely to be found in communities with strong scientific identities such as medicine (Timmermans & Berg, 2010), biology (Leonelli, 2014), and high-energy physics (Knorr Cetina, 1999), as scientific models are regularly underdetermined by data. (3) *Corporate logic*: the time- and resource-bounded nature of satisficing search, key to our reframing pattern, is likely to show up across a variety of corporate settings (see e.g., Passi and Jackson's [2018] study of data analytics at a telecom vendor). In addition, the cross-pressure surrounding operational decision-making stemming from competing agendas of professional norms, formal rules, and operational demands has been identified in safety-critical, operational settings (Perin, 2006; see also Fine, 2009).

In closing, our hope is that the analysis articulated in the three patterns of work practices may provide a fertile and generative breeding ground for pursuing a research program that "gets under the hood" of data science in IS through practice-oriented studies.

Acknowledgments

This paper is the result of a long journey. It has evolved over a process of several years. A number of colleagues have commented on earlier, even ancient, versions of the paper. These include but are not limited to: Petter Almklov, Samer Faraj, Ole Hanseth, Vidar Hepsø, James Howison, Neil Pollock, Georg von Krogh, and Youngjin Yoo. We are indebted to the particularly helpful feedback from the *JAIS* reviewers and, last but not least, the editor. This research was partially financed by Research Council of Norway grants #237898 and Digital Class #309631.

References

Alaimo, C., & Kallinikos, J. (2021). Managing by data: Algorithmic categories and organizing, *Organization Studies*, *42*(9), 1385–1407.

Alaimo, C., Kallinikos, J. & Aaltonen, A. (2020). Data and value. In *Handbook of digital innovation*, Lyytinen, K., Nambisan, S. & Yoo, Y. (Eds.), Edward Elgar.

Anderson, C. (2008). *The end of theory: The data deluge makes the scientific method obsolete.* Wired. https://www.wired.com/2008/06/pb-theory/

Bailey, D. E., Leonardi, P. M. & Barley, S. R. (2012).The lure of the virtual. *Organization Science, 23*(5), 1485-1504.

Baskerville, R., Myers, M. & Yoo, Y. (2020). Digital first: The ontological reversal and new challenges for IS research. *MIS Quarterly*, *44*(2), 509-523.

Baudrillard, J. (1994). *Simulacra and simulation.* University of Michigan press.

Beunza, D. & Garud, R. (2007). Calculators, lemmings or frame-makers? The intermediary role of securities analysts. *The Sociological Review*, *55*(2), 13-39.

Bond, C. E. (2015). Uncertainty in structural interpretation: Lessons to be learnt. *Journal of Structural Geology, 74*, 185-200.

Borgmann, A. (1999). *Holding on to reality: The nature of information at the turn of the millennium*. University of Chicago Press.

Bowker, G. C. (1994). Science on the run: Information management and industrial geophysics at Schlumberger, 1920-1940. *MIT Press.*

Brown, W. M. (1983). The economy of Peirce's abduction. *Transactions of the Charles S. Peirce Society*, *19*(4), 397-411.

Burton-Jones, A. (2014). What have we learned from the smart machine? *Information and Organization*, *24*(2), 71-105.

Burton-Jones, A. & Grange, C. (2012). From use to effective use: A representation theory perspective. *Information Systems Research*, *24*(3), 632-658.

Chang, H. (2004). *Inventing temperature: Measurement and scientific progress.* Oxford University Press.

Dewey, J. (1930). *The quest for certainty*. G. Allen & Unwin.

Dunne, D. D. & Dougherty, D. (2016). Abductive reasoning: How innovators navigate in the labyrinth of complex product innovation. *Organization Studies, 37*(2), 131-159.

Edwards, P. N. (1999). Global climate science, uncertainty and politics: Data-laden models, model-filtered data. *Science as Culture, 8*(4), 437-472.

Fine, G. A. (2009). *Authors of the storm: Meteorologists and the culture of prediction.* University of Chicago Press

Gitelman, L. (Ed.) (2013). *Raw data is an oxymoron.* MIT Press.

Günther, W. A., Mehrizi, M. H. R., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. *Journal of Strategic Information Systems 26*(3), 191-209.

Haag, S., & Cummings, M. (2009). *Management information systems for the information age.* McGraw-Hill.

Henke, N., Levine, J. & McInerney, P. (2018). *You don't have to be a data scientist to fill this must-have analytics role*. Harvard Business Review. https://hbr.org/2018/02/you-dont-have-to-be-a-data-scientist-to-fill-this-must-have-analytics-role

Jones, M. (2019). What we talk about when we talk about (big) data. *Journal of Strategic Information Systems, 28*(1), 3-16.

Kallinikos, J. (1999). Computer-based technology and the constitution of work: A study on the cognitive foundations of work. *Accounting, Management and Information Technologies 9*(4), 261-291.

Kallinikos, J. (2004). Deconstructing information packages: organizational and behavioural implications of ERP systems. *Information Technology & People*, *171*(1), 8-30.

Kallinikos, J. (2007). *The consequences of information: Institutional implications of technological change.* Edward Elgar Publishing.

Kallinikos, J., Aaltonen, A. & Marton, A. (2013). The ontological ambivalence of digital objects. *MIS Quarterly 37*(2), 357-370.

Kellogg, K. C., Orlikowski, W. J., & Yates, J. (2006). Life in the trading zone: Structuring coordination across boundaries in postbureaucratic organizations. *Organization Science 17*(1), 22-44.

Kitchin, R. (2014). Big data, new epistemologies and paradigm shifts. *Big Data & Society 1*(1), https://doi.org/10.1177/2053951714528481.

Klein, H. K., & Myers, M (1999). A set of principles for conducting and evaluating interpretative field studies in information systems. *MIS Quarterly, 23*(1), 67-93

Knorr Cetina, K. (1999). *Epistemic cultures: How the sciences make knowledge.* Harvard University Press.

Leonardi, P. M. (2012). *Car crashes without cars: Lessons about simulation technology and organizational change from automotive design*. MIT Press.

Leonelli, S. (2014). What difference does quality make? On the epistemology of big data in biology. *Big Data & Society, 1*(1). https://doi.org/10.1177/2053951714534395.

Loch, C. H., DeMeyer, A. & Pich, M. (2011). *Managing the unknown: A new approach to managing high uncertainty and risk in projects.* Wiley.

Lusch, R., & Nambisan, S. (2015). Service innovation: A service-dominant logic perspective. *MIS Quarterly 39*(1), 155-176.

Lyytinen, K., & Grover, V. (2017). Management misinformation systems: A time to revisit?, *Journal of the Association for Information Systems, 18*(3), 206-230.

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

March, J. G. (1994). *A primer on decision making.* Free Press.

Markus, M. L. (2017). Datification, organizational strategy, and IS research: What's the score? *Journal of Strategic Information Systems, 26*(3), 233-241.

Mikalsen, M. and Monteiro, E. (2018). Data handling in knowledge infrastructures: A case study from oil exploration. *Proceedings of the ACM Conference on Human-Computer Interaction*.

Mol, A. (2002). *The body multiple: Ontology in medical practice*. Duke University Press.

Monteiro, E., Pollock, N., Hanseth, O. and Williams, R. (2013). From artefacts to infrastructures. *Computer Supported Cooperative Work*, 22(4-6), 575-607.

Monteiro, E. & Parmiggiani, E. (2019). Synthetic knowing: The politics of the Internet of Things. *MIS Quarterly, 43*(1), 167-184.

Morgan, M. S. (2010). Travelling facts. In P. Howlett, & M. S. Morgan, M.S. (Eds.), *How well do facts travel? The dissemination of reliable knowledge* (pp. 3-39). Cambridge University Press.

Newell, S., & Marabelli, M. (2015). Strategic opportunities (and challenges) of algorithmic decision-making: A call for action on the long-term societal effects of "datification." *Journal of Strategic Information Systems, 24*(1), 3-14.

Oborn, E., Barrett, M., & Davidson, E. (2011). Unity in diversity: Electronic patient record use in multidisciplinary practice. *Information Systems Research, 22*(3), 547-564.

Orlikowski, W J., & Scott, S. (2016). Digital work: A research agenda. In B. Czarniawska (Ed.), *A research agenda for management and organization studies* (pp. 88-96). Edward Elgar Publishing.

Passi, S. & Jackson, S. (2018). Trust in data science: Collaboration, translation, and accountability in corporate data science projects. *Proceedings of the ACM Conference on Human-Computer Interaction*.

Peirce, C. S. (1931-35). *The collected papers of Charles Sanders Peirce* (C. Hartsthorne & P. Weiss, Eds.). Harvard University Press

Perin, C. (2006). *Shouldering risks: The culture of control in the nuclear power industry*. Princeton University Press.

Pollock, N, & Williams, R. (2016). *How industry analysts shape the digital future.* Oxford University Press.

Poovey, M. (1998). *A history of the modern fact: Problems of knowledge in the sciences of wealth and society.* University of Chicago Press.

Porter, T. M. (1996). *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton University Press.

Power, M. (1997). *The audit society: Rituals of verification.* Oxford University Press.

Royal Society (2012). Science as an open enterprise. http://royalsociety.org/policy/projects/science-public-enterprise/report/ 2012.

Scott, J. C. (1998). *Seeing like a state: How certain schemes to improve the human condition have failed.* Yale University Press.

Singh, D., Tripathi, G., & Jara, A. J. (2014). A survey of Internet-of-Things: Future vision, architecture, challenges and services. *Proceedings of the IEEE World Forum on Internet of things* (pp. 287-292).

Sivarajah, U., Kamal, M. M., Irani, Z. & Weerakkody, V. (2017). Critical analysis of big data challenges and analytical methods. *Journal of Business Research, 70*, 263-286.

Slota, S. C., Hoffman, A. S., Ribes, D. & Bowker, G. C. (2020). Prospecting (in) the data sciences. *Big Data & Society, 7*(1). https://doi.org/10.1177/2053951720906849

Stark, D. (2009). *The sense of dissonance: Accounts of worth in economic life.* Princeton University Press.

Suddaby, R. (2006). From the editors: What grounded theory is not. *Academy of Management Journal, 49*(4), 633-642.

Sugimoto, C., Ekbia, H. & Mattioli, M. (Eds.). (2016). *Big data is not a monolith.* MIT Press.

Timmermans, S. & Berg, M. (2010). *The gold standard: the challenge of evidence-based medicine.* Temple University Press.

Turkle, S., Clancey, W. J., Helmreich, S., Loukissas, Y.A. & Myers, N. (2009). *Simulation and its discontents*. MIT Press.

Venters, W., Oborn, E. & Barrett, M. (2014). A trichordal temporal approach to digital coordination: The sociomaterial mangling of the CERN grid. *MIS Quarterly, 38*(3), 927-949.

Walsham, G. (1995). Interpretive case studies in IS research: Nature and method. *European Journal in Information Systems, 4*(2), 74-91.

Weick, K. (1985). Cosmos vs. chaos: Sense and nonsense in electronic contexts. *Organizational Dynamics, 14*(2), 51-64.

Wiesche, M., Jurisch, M., Yetton, P. & Krcmar, H. (2017). Grounded theory methodology in information systems research. *MIS Quarterly, 41*(3), 685-701.

Wylie, A. (2002). *Thinking from things: Essays in the philosophy of archaeology.* University of California Press.

Zuboff, S. (1988). *In the age of the smart machine.* Basic Books

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power.* Profile Books.

Østerlie, T. and Monteiro, E. (2020). Digital sand: The becoming of digital representations. *Information and Organization, 30*(1), Article 100275.

About the Authors

Marius Mikalsen is a senior researcher at SINTEF and an associate professor at the Norwegian University of Science and Technology (NTNU) and holds a PhD in informatics from NTNU. His research focuses on information systems development, use, and organizing across sectors such as health, finance, and energy. His publications have appeared in *Computer Supported Cooperative Work, Proceedings of the ACM on Human-Computer Interaction*, *Communications of the Association for Information Systems*, *International Journal of Medical Informatics*, and in the proceedings of ICIS and ECIS.

Eric Monteiro is a professor of information systems at the Norwegian University of Science and Technology (NTNU) and an adjunct professor at the University of Oslo. He has studied sociotechnical processes of digitalization in a variety of private and public organizations and sectors. Research outlets include *MIS Quarterly*, *Information Systems Research, Journal of the Association for Information Systems*, *European Journal of Information Systems*, *The Information Society*, *International Journal of Medical Informatics*, and *Computer Supported Cooperative Work*.