Data Handling in Knowledge Infrastructures: A Case Study from Oil Exploration

MARIUS MIKALSEN, SINTEF Digital and Norwegian University of Science and Technology marius.mikalsen@sintef.no
ERIC MONTEIRO, Norwegian University of Science and Technology

Abstract: Offshore oil exploration is concerned with subsea geological reservoirs that are numerous kilometers below the seabed. These reservoirs are knowable only through a knowledge infrastructure of interconnected technologies that are applied to diverse instrument-generated data. Noise, holes, and inaccuracies are inherent in the data, which depend on the technology producing it. We conducted an interpretative case study of data handling work in the exploration unit of a European oil company. Our findings show how data handling involves the skills needed for managing data identities and ownership, a variety of technologies, and contingent negotiations of data needs. We use the notion of repair to analyze this data handling work and discuss how the concept of repair in data handling involves keeping the knowledge infrastructure navigable and attending to countless details. Our research contributes to the literature on repairing infrastructures by considering how repair relates to data work.

Author's addresses: M. Mikalsen, SINTEF Digital, P.O. Box 4760, Strindveien 4, Trondheim, Norway; E. Monteiro, Norwegian University of Science and Technology, Department of Computer Science, NO-7491, Trondheim, Norway.

1 INTRODUCTION

Offshore oil exploration is inherently data centric; data sets are large, historical, and sensor generated (including seismic, radioactive radiation, electro-magnetic readings, down-hole pressure, temperature, and drill torque), and these data suffer from poor quality (e.g., the typical lifespan of a pressure sensor is two years before the readings become skewed) and are highly heterogeneous (time-stamped structured data, but also text, images, graphs, equations, and simulations).

Off the coast of Norway, the oil reserves are thousands of meters below the seabed and are inherently difficult and expensive to reach physically, even when using state-of-the-art instruments. The only way of proving the presence of a hydrocarbon reservoir is by drilling an exploration well, which can cost up to 100 million USD. Such endeavors are naturally only conducted when the prospects are properly justified and have a high enough probability of commercial viability. Therefore, most prospects are substantiated on indirect data, such as seismic surveys. Seismic surveys are also expensive to collect, so they often rely on existing, older surveys. Additionally, they use existing well data from wells that are close to the area or that are from an area presumed to have similar geological characteristics.

Finding suitable prospects has three particularly interesting characteristics. The first characteristic is that the explorationists (a term describing geologists and geophysicists working on exploration projects) need to come up with new and unproven (i.e., supported only by indirect data) concepts. Risky prospects are important because moving into new and unknown areas has the potential to create vast economic opportunities. The second characteristic is the close cooperation between the explorationists and data managers (a term describing the technical staff tasked with finding, importing, managing, and storing the data). This cooperation is required because of the third characteristic: their reliance on instrument-generated, numerous, and indirect data. Because data collection is expensive, explorationists must use historical data. In this domain, there are large volumes of data from over four decades of oil exploration and production on the Norwegian continental shelf. However, data are rarely at hand and instead must be found, imported, quality assured, reworked, and reinterpreted to be applicable in a current analysis. Doing this work requires the explorationists and data managers to deal with different databases, file systems, file formats, standards, and data management technologies.

In light of these characteristics, we can consider oil exploration as a knowledge infrastructure. In Computer-Supported Cooperative Work (CSCW), the work on knowledge infrastructures refocuses how work is organized in large and dispersed scientific collaborations [2], [11]. Infrastructures are characterized by how they involve different users; the number of interconnected modules or systems and practices that are being shaped and used across many different locations; and how they endure over extended periods of time (decades rather than years) [22]. Particular forms of knowledge infrastructures employ different forms of sensor networks to "[…] study phenomena at spatial and temporal scales that were previously impossible to achieve" [19]. A key novelty of these knowledge infrastructures lies in the assembly of methods, technologies, and skills developed to *handle* (i.e., format, disseminate, retrieve, model, and interpret) numerous and diverse sensor-generated data [17].

Sustaining infrastructures involves articulation work [36], which covers the invisible activities that support and enable the interaction of the distributed parts of large-scale systems. In infrastructure studies, articulation has been conceptualized as work that becomes visible when the infrastructure breaks down [33], [36]. Although all infrastructures will experience major breakdowns, a salient feature of the oil and gas knowledge infrastructure is how the methods, technologies, and skills of data handling are not commoditized and, hence, are not at hand. Reiterating Star and Ruhleder's [35] concern regarding the image of infrastructures as simply "sinking into the background", here, knowledge and insight relies on effectively handling numerous and diverse historical data and the ongoing ebb and flow of new data that emerges through a continuous assembly of methods, technologies, and skills [2], [23]. Although invisibility certainly is one aspect of infrastructures, "[…] it is only one and at the extreme edge of a range of visibilities that move from unseen to grand spectacles and everything in between" [15]. The data handling in knowledge infrastructures of instrument-generated data push toward the more visible end of this range.

To capture the active contribution of data handling, we apply the notion of repair. As Jackson [9] explains, the world is constantly breaking, its infrastructures included. Here, repair is about the extension and safeguarding of capabilities that are in danger of decay. Important to grasping data handling in knowledge infrastructures, however, is that repair also accounts for the concept of renewal. When connecting the old to the new—such as historical data with current exploration—repair is generative. These connections are built into the knowledge infrastructures through collective acts of repair [8]. How these connections play out in practice, particularly in a business setting with an infrastructure that is basically always broken, still needs empirical refinement. Therefore, we ask the following research question: *How do organizations execute repair work on a knowledge infrastructure that is always incomplete?*

Empirically, we report on a European-based, internationally operative oil and gas company's exploration unit (referred to as OilCorp for the purposes of anonymity). We analyze the collaborative work between explorationists, data managers, their technologies, and the data. We find that data handling is characterized by repair, and we discuss how such repair work unfolds. In the current paper, we aim to make two contributions to the literature. First, we analyze and discuss data handling as repair. In so doing, we add to the studies that consider repair as a source of technological innovation [3], [8], [20], [33], [34]. Our focus is on how repair in data handling contributes to data analysis, knowledge, and insight. We also analyze how data handling as repair involves keeping the infrastructure navigable, requiring attention to countless details. Second, we empirically analyze data handling practices in business organizations, adding to the CSCW's better-studied cases from scientific settings [11], [13], [19], [23].

The present paper is organized as follows: Section 2 explains the theoretical conceptualization of repair. Studies on articulation in infrastructures have focused on getting things back on track after breakdowns. The concept of repair builds on this work, narrowing in on the methods, technologies, and skills applied to handle data in the infrastructure. Section 3 outlines the empirical background of the current study and explains the interpretative case study, data collection, and analysis. Section 4 presents the repair work in an oil exploration infrastructure. Section 5 discusses data handling in the exploration infrastructure in light of the infrastructure and repair literature. Section 6 offers concluding remarks.

2   DATA HANDLING IN KNOWLEDGE INFRASTRUCTURES AS REPAIR

Strauss [37] applies the concept of articulation work to understand how people support each other in heterogeneous work, which is further expanded on in CSCW as "work that gets things back on track in the face of the unexpected, and modifies action to accommodate unanticipated contingencies" [32]. Articulation work was originally studied as part of co-located work practices [21] and control rooms [7] but has been extended to also describe various types of distributed collaboration [1].

In infrastructure studies, articulation has traditionally been conceptualized as invisible work that only becomes noticeable when the infrastructure breaks down [33], [36]. Although all infrastructures experience major breakdowns, a striking feature of knowledge infrastructures—or infrastructures that allow one to study a phenomenon otherwise impossible [19]—is that they stubbornly resist falling into the background. As such infrastructures emerge more as grand spectacles [15], articulation work moves out of the margins, becoming entangled with the core business processes.

As articulation work moves to the level of large-scale infrastructures, attention must be given to the inter-relationship of the diverse data and technologies that emerge [29]. The study of infrastructures in CSCW is a "…profound refocusing on how work is organised in larger, more geographically dispersed and not rigidly hierarchal collaborations, and how these are organised and evolve as sociotechnical endeavours" [11]. These endeavors entail a necessary redistribution of labor between humans and technology [29], and studies have shown how the social interactions needed to sustain scientific work decrease as the work and agency are delegated to technology [30]. Recent work on "infrastructuring" has engaged with this type of relational and in-the-making perspective of how infrastructures emerge [25]. Karasti et al. [13], for example, show how information managers who are tasked with managing technology over the long term focus on providing ongoing, reliable, and sustainable information environments as opposed to the short-term and project specific goals of project workers. Pipek and Wulf [26] suggest using infrastructuring as a concept to understand system renewal, which would help avoid confusion with classic notions of design as "design-before-use," which is commonly performed by professional designers.

Other studies have focused on data handling in knowledge infrastructures. These studies have examined data sharing and reuse in large-scale organizations, such as the social practices required to share research data [16] and how data sharing must be balanced between standardization and situated understandings [40]. Therefore, the contextualization of meta-data has been investigated to discover how data can be usable outside the place and time of the data's collection [12].

To date, however, less has been researched about the assembly of data, technology, and skills that is at work in the analysis of data, making this area an important one for CSCW [11]. A key novelty of the knowledge infrastructures of instrument-generated data is the methods, technologies, and skills developed to handle all the different data required for analysis work [17].

We use the notion of repair to analyze the active contribution of data handling in knowledge infrastructures and to grasp its relationship to data analysis work. Repair focuses on the active work of repairing, maintaining, and appropriating technology [3]. Here, our work agrees with Cohn's stance that highlights the active work needed to make sociotechnical arrangements function and persist over time [3]. Recent research on repair has highlighted the remarkable and diverse ability of humans to make imperfect systems work [10] and how the negotiated identification of breakdowns and collaborative definitions of worth [31] are important for contingent

problem solving. Repair embeds modes of human interaction with technology that surface valuation of data and technology as contingent and ongoing accomplishments [9]. As Jackson explains [9], repair can be considered a form of articulation work and vice versa. A key facet of repair that makes it an illuminating lens for viewing data handling work in knowledge infrastructures is its emphasis on bridging the past and future. Repair in knowledge infrastructures inherits a multitude of databases, systems, and data, which continuously must be combined and re-combined with emergent systems and data to allow data analysis to work.

Thus, beyond keeping infrastructures from breaking [6], repair has generative and productive connotations. Breakdowns are not necessarily barriers, nor catastrophic from a repair perspective. Rather, it is "[…] precisely in moments of breakdown that we learn to see and engage our technologies in new and sometimes surprising ways" [9]. To clarify how repair contributes to renewal, Sims [33] distinguishes between repair-as-maintenance and repair-as-transformation. Infrastructures can break down in different ways, all of which involve a mismatch between what a system was designed to do and what it can do. Through degradation, technology ages to the point that it no longer performs as intended. Through obsolescence, demands and standards change so that certain interests are no longer being met. Thus, infrastructural repair involves determining the state of a system to create a shared understanding between diverse interests. Achieving a working knowledge infrastructure is less about making grand engineering gestures and more about "getting countless details right" [33].

In the sections that follow, we build on these infrastructuring and repair principles to describe the data handling work in OilCorp's exploration knowledge infrastructure.

3   CASE AND METHOD

3.1   The Knowledge Infrastructure of Exploration

OilCorp bids for Norwegian government-chartered production licenses in the North Sea. A license allows a company to operate, for example, to conduct new seismic surveys or drill an exploration well. OilCorp´s bidding depends on the prospoects produced by their exploration unit.

In the North Sea, there are hydrocarbon-producing basins, which are depressions in the crust of the earth where sediments have accumulated; under intense pressure and heat over millions of years, they have become oil-producing rock. Oil seeps up through the permeable rock layers and is sometimes trapped in a rock layer that has the right shape and form. The exploration unit makes prospects of such traps, which are assumed to contain commercially producible hydrocarbon reservoirs. Commercially producible means that there are sufficient hydrocarbon reserves in place to justify the significant investment needed to recover them. Therefore, a prospect positions the reservoir and estimates the contents of it (oil, gas, and water ratio), volumes, and risks associated with these estimates.

Creating prospects happens through and with a knowledge infrastructure of instrument-generated data. Data come from reflection seismology, which is a method of exploration geophysics and where seismology is used to estimate the properties of the subsurface. In this sense, seismology is similar to sonar and echolocation, requiring a seismic source, such as an air gun, and detectors to register the returning reflections to construct the seismic record. Traditional offshore seismic surveys were shot using ships that tow one or more cables, known as streamers. 2D seismic surveys are shot using one streamer, and 3D surveys using up to 12 streamers. More recently, ocean bottom cables that capture reflections are laid out by remotely operated underwater vehicles, using a separate vessel to generate the seismic source. This enables 4D seismic surveys, that is, a time series of 3D seismic surveys, which are useful, for example, when observing how a reservoir is changing during production. Following collection, seismic data are processed and handed over to the explorationists for seismic interpretation and geologic modeling.

Seismic surveys are relatively crude, with granularity in the neighborhood of 100 m^3. More detailed information accurately reflecting depth is gathered by drilling wells and sinking measuring equipment down into these wells ("well logging"); the well log is a detailed record of the geologic formations and may be based on visual inspections of samples brought to the surface (geological logs), such as coring actual samples of rock and mud-logging, which describes the rock or soil cuttings brought to the surface. Physical measurements are made by instruments (e.g., electric and acoustic measurements) lowered into the borehole (geophysical logs). Geophysical well logging can be done during drilling, completing, producing, or abandoning. Most recently, logging while drilling has been applied, where the logging tools are incorporated into the drill string, and measurements are streamed in real time to the surface.

Data are organized in several databases; hundreds of terabytes of data are in the corporate data store alone. Data are brought to explorationists primarily from two main sources: the Diskos National Data Registry (Diskos)

and OilCorp's main data store (the actual names of OilCorp's data stores are removed for anonymity). Searching for public data (e.g., production licenses, exploration wellbores, discoveries, fields, development wellbores, and business areas) is available through Diskos.[1] OilCorp and other members in the Diskos consortium have their own tools to access Diskos, and companies can upload, search for, download, and trade data through that system. The corporate data store holds OilCorp's proprietary interpreted data, such as seismic interpretations, interpreted well data, production logs, and maps. Business information, such as license areas, infrastructure (for production), and business associates, is also included.

The everyday work practices of explorationists centers around piecing together credible prospects from and grappling with uncertainties in the available data. Given the importance of combining an understanding of geology with different types of data, such as the well log and seismic data, explorationists must use skills from both geology and geophysics. Explorationists work in teams that are assigned to, for example, regional exploration in an area or on exploration in licenses. Specialists, such as geochemists, support the explorationists and are organized in excellence units and perform specialist studies. Sometimes, the specialists can be embedded into the explorationist teams, for example, when doing basin modeling.

Given the intricacies of searching for and accessing different data, project data managers (PDMs) are co-located with the explorationists. The reason for the co-location is because the PDMs need to be embedded in the workflow to have a better understanding of the explorationists' needs in terms of data.

Continuous interaction occurs between the PDMs and explorationists. The PDMs do various tasks, but the prime one is to understand the exploration projects and what data may be relevant to those projects; they know how to search across various databases and use a variety of tools (depending on where the data are and where the data go) to load data into the explorationists' interpretation tools, including various forms of quality control of data upon loading. Others work on organizing and visualizing data, e.g. into maps in geographical information systems (GIS). The PDMs also develop tools for explorationists, such as search tools for use across databases, file structures to browse well data that are accessible outside web browsers, and tools that keep track of the status of data at decision gates in the funnel model.

The PDMs work with central data managers (CDMs), who are responsible for keeping the databases synchronized and accessible. Presenting a certain combination of data in a GIS map, for example, may involve complex database queries that are formulated by the CDMs, followed by the formatting and integration of data via Python scripts carried out by the PDMs.

We studied OilCorp's exploration unit as it navigated promising areas in the North Sea and produced prospects.

3.2   Research Strategy and Data Collection

We conducted a 39-month interpretative case study at OilCorp [38]. The case study is interpretative in that it assumes the epistemological view that our "[…] theories concerning reality are ways of making sense of the world, and shared meanings are a form of intersubjectivity rather than objectivity" [39]. We traced the actors in oil exploration and collected data from eight consecutive phases (February 2013 – May 2016). Different data sources were used throughout the phases. In phases 1 to 4 we conducted interviews and did participatory observation. In phases 5 and 7, we participated in workshops where new tools for data access were explored. Phases 6 and 8 were again used for interviews and observations to clarify the open questions from previous phases. While present at OilCorp, we obtained access to archival data. Important documents included technical specifications that helped us understand various data, data sources, and the tools used during exploration. In between the phases, we accessed public data sources that are used by the explorationists to find, trade, and access relevant data; these sources were the NPD Factpages (http://factpages.npd.no) and the Diskos DB (http://www.diskos.no). Following the advice of Ribes and Polk [26], we observed and talked to those working with the diversity of the infrastructure, here focusing on the data managers. All interviews were audio-recorded, and the interviews with the explorationists and data managers were transcribed. We also made extensive use of field notes [18].

Table 1. Data Sources and Description

Digital data sources (tools and archival data)	Internal documents (workflows, architecture documents, and status reports), public documents

---

[1] Available at http://www.diskos.no/public-portal (accessed 01/4/2018).

	(strategies results, and yearly accounts), and tools (NDP Factpages and the Diskos National Data Repository)
Interviews and workshops	Twenty-five interviews, including central data managers, project data managers, explorationists, process owners, and IT management
Direct observation	Sixteen full days in the OilCorp exploration unit; accessing documents, presentations showing systems and processes, observing explorationists and data manager's interactions, system demonstrations and discussions, and attending meetings

3.3 Data Analysis

Although the first author collected most of the data, the second author was also partially involved in the data collection process, allowing us to compare our interpretations [5]. Data analysis was performed in iterations of inductive and deductive steps, following Klein and Myers' [14] principle of dialogical reasoning. We started with a perspective of knowledge infrastructures and the invisible work involved within. By gradually identifying the themes in the data (by annotating transcripts and making timelines), we found that repair was more visible and ongoing in the infrastructure of oil exploration. Using Klein and Myers' [14] principle of multiple interpretations, we searched for perspectives on repair from different actors, such as explorationists and data managers. Explorationists were primarily concerned with the quality and relevance of data, such as seismic surveys and well data, along with the technology to interpret this data. Data managers were primarily concerned with the technology required to find, import, manage, and store these data, but they also did quality assurance of the data as the data were located and loaded into interpretation tools. They assessed well data, for example, to calibrate depth measures for use in interpretation tools.

4   FINDINGS

The data managers reported on how working with instrument-generated data was dealt with in the infrastructure. Our findings illustrate several aspects of this work. First, there is the manual work involved in managing data identities and ownership. Second, there is how the sheer complexity of the data, databases, and technologies requires an assembly of skillsets. Third, there is a contingent negotiation of data needs.

4.1   Keeping File IDs Synchronized

Exploration data are stored in different databases in the infrastructure. The Norwegian national data repository for petroleum data (Diskos DB) stores seismic, well, and production data, per Norwegian petroleum legislation. The member oil companies of the Diskos joint venture can buy and trade confidential data and share non-confidential data. Additionally, OilCorp has several internal databases that store interpreted and quality-controlled data that they are not obligated to share with Diskos. Furthermore, each exploration project has workspaces where intermediate interpretations are stored.

When exploration targets a new geographic area, it is essential to obtain an overview of all the relevant data for that area. Because of the different data vendors (e.g., rig companies) and varying naming schemes in different systems at various times in the exploration process, there are divergent identifiers (IDs) for the same data (e.g.,

well data) across the databases. A well log can have a different ID in Discos than in OilCorp's internal database. To identify the same data set across databases, data managers must have intimate knowledge of the IDs across systems.

We observed OilCorp's data managers and a DBCompany (pseudonym) consultant as they worked on ID mapping. At the time, DBCompany supplied an interpretation tool, the workspace database, and Diskos, and the task was to harmonize IDs for the seismic files in Diskos and the workspace databases. The reason behind harmonizing internal IDs with Discos IDs was to make sure that all the data OilCorp had access to in Diskos would be available for interpretation in the explorationists' project workspaces. The data managers had more intimate knowledge of the IDs and the organization than the external consultant did, as shown in the following fieldnotes:

> The consultant walks through a spreadsheet containing the workspace IDs and Diskos IDs, as well as information such as "not Diskos" or "not workspace database." Because the consultant is not an expert in the contents of the seismic files, he had compared IDs in the project workspace to those in Diskos. The consultant says he thinks he has gotten 95% of the mismatches and fixed them. They discuss what to do about the remaining 5%. They agree that the data managers should go through the remaining mismatches. As they do, the data managers look at the file metadata. Although the "required files" are missing from the metadata, they use clues, such as when the file was created, by whom, and the file ending used, to determine if the files are similar. (Fieldnotes)

4.2 Determining Ownership of Data

We found how data handling extended the technical work by using identification when we observed a team that worked on determining the ownership of seismic surveys. Again, the issue was managing historic data, but this time, it concerned ownership of valuable seismic data. OilCorp was to produce a merge survey (a seismic survey that merges several existing surveys into a new one) as part of a new exploration license. A license is granted by the government to a consortium of companies working together to share the costs and risks. To determine who can legally access the new survey, OilCorp had to determine if the partners in the new license had legal access rights to the existing surveys. There are potential lawsuits and economic repercussions if companies without the correct access right use the new survey. A data manager explain how they work to determine the ownership of data:

> The devil is in the details. A seismic survey company is commissioned by an oil company to shoot a survey. The oil company commissions the work because it is operator in a license. This chain of information is often lost in the data. So to know… And then there is the naming of surveys in Norway. The two first out of three letters are short names for the companies. Then there are the numbers, which is the year the survey was shot. 98 is 1998, 04 is 2004. And then there are three numbers at the end. These numbers indicate what kind of seismic it is. Is it the usual license seismic, or a site survey, that is, seismic shot as they are drilling a well. The Diskos DB has an overview of all surveys shot on the Norwegian shelf. It has the name of the company that was responsible for collecting it and has applied for the permit. Often, if it is license seismic, there are external companies that have done this on behalf of the license. It is important for us to know which license has collected this data because then we can know which companies are allowed to access the data. Also, we can see who in OilCorp is responsible for this dataset. (Interview with project data manager)

Working to determine ownership includes knowing about the finer details of the license processes and how to find and investigate reports, followed by seeking out resolutions with the pertinent stakeholders. Figuring out ownership illustrates how data management goes beyond purely mechanical tasks and involves a great deal of care, as a data manager explained:

> In order to figure this out, we need to go into old reports for the survey or for the processing of the survey. What we do now is to go through, survey for survey, to see what is registered at Discos and on our internal databases. If we are lucky, it says that the survey was done in this and that license, but sometimes, it does not say at all, and then we do not know. And when we do not know, we must see if we find names of people that is in OilCorp, and must check with them if they remember. (Interview with project data manager)

4.3 Develop Tools to monitor the State of Data

Historical well and seismic data are a prerequisite for exploration. Making data reusable requires preparing the data before archiving it. In exploration, data are continuously produced in the form of new interpretations. Tracing seismic horizons, for example, is the tracing of an underground rock surface by selecting ("picking") well data and then interpolating between well data points using seismic data. One explorationist showed an example

of the amount of horizons in an exploration project: "[…] you can find hundreds of versions of a seismic horizon in a project, and there is really no way of knowing which one to trust" (Interview, explorationist).

At certain predefined decision gates in the exploration process, for example, when deciding to drill a prospect, the data and interpretations that are underpinning the decision are made "QC" (their term for official, quality-controlled interpretations at these decision gates). Quality-controlled data are archived in OilCorp's internal database, becomes available outside the project workspace, and are considered an official OilCorp interpretation. In addition to document decisions, the quality-controlled data facilitates reuse by explicitly indicating what is the best available interpretation available.

The data managers are responsible for archiving the QC interpretations. But because they did not do the interpretations, they could not determine which one of a hundred horizons, for example, was the official one. In the beginning of our case study, the data managers had to "nag" the explorationists to get an answer. During our case study, OilCorp worked to ensure that the QC data were properly archived; a key part of this was when the data managers developed a dashboard tool that would allow the exploration managers (who are responsible for several exploration teams in a geographical area) to monitor the status of data at decision gates in the projects they were responsible for. The development of this tool shows how data management becomes increasingly entangled with exploration management and how data management does not only include using technology in innovative ways, but also in developing new technology to solve relevant problems, as a data management leader points out:

> It is the exploration managers who are responsible for the data. With the dashboard tool, they can easily see what seismic and well data they have in their area and what the status is of the interpretations that are produced […]. They can see what data that has gone into a decision gate, and if it has been quality controlled. (Interview with data management leader)

4.4 Diverse Technology and Contingent Procedures

Different types of data management skills come together to connect historical data and current exploration. Collaboration is required because making these connections involves a diverse set of technologies. A central data manager (a term for a data manager who supports the project data managers and works with the core database management systems) demonstrated the process of populating a new working project with data. Typically, a new project copies data from an existing, nearby project to obtain all the relevant data. But in every new area, additional data not yet loaded can potentially be relevant. Searching across different working projects required involving central data managers and the well spider tool, as shown by the following fieldnotes:

> By default, project data managers do not have access to all the data [for data privacy reasons]. They have this tool, the well spider, where you can search for well data across working projects. However, it is not used that much, partly because they [the project data managers] do not know about it and partly because it is a very complicated tool. What often happens is that a project data manager will contact a central data manager to learn who is responsible for data in a certain area. (Fieldnotes from a tool demonstration by a central data manager)

Accessing the databases, filtering database result sets, and presenting answers require diverse tools and skills, which all starts at the most basic level by querying the underlying databases. A member of a team working on automating access to the underlying databases explained it as follows:

> The queries that run against the [OilCorp] databases are very, very complex. There are optimized queries that were made during [corporate database project] in [year], and it is really hard to understand them, so it is best to use the queries that are available through the [name of application programming interface]. It is hard to define new database schemas, and we use the database schemas that were used in the [corporate database project] defined by [vendor name]. [Another vendor] now has the contract for this solution, and it is maintained by them. (Fieldnotes)

The explorationists are eager to get their hands on new well data as soon as possible. In order to achieve this, the data managers must adapt the technologies and methods used depending on who is doing the drilling. A data manager explained how it was done in OilCorp-operated wells as follows:

> In OilCorp operated wells we get real-time data as the well is drilled. Data come in from geo-operations, the guys that do the drilling for exploration. The data go to geo-physicists, who do log splicing and quality control, and when they are done, they send it to Diskos via [quality control company for Diskos]. Simultaneously, they send it to the data manager for loading. (Interview, data manager)

Different procedures and technologies are applied if an external vendor drills the well, as shown in the following interview:

> During drilling operations [external vendor] puts all the data into [name of database], which they use for all the wells they drill, and then, OilCorp staff get access to the wells they are working with. The data manager also gets access and can extract the well data that is needed. This is before it is quality controlled and added to Diskos. (Interview, data manager)

Once the data are found, loading it into working projects is not a purely mechanical task and it can involve using several different tools. Loading well data into a project typically takes four hours, depending on the skill level of the data manager and the quality of the data. The quality of the data varies based on when it was produced, with older data typically being of a lower quality. The data managers conduct basic quality control of the data when they load it, such as checking that the "mean sea level" and "sea bottom" on the well data are properly calibrated.

4.5   Negotiating Data Needs

Explorationists have intricate and shifting needs in terms of data. They can be interested in what wellbore data are available in an area, but also with what wellbore data exist elsewhere:

"[…] due to a lack of access or loading, [then] trade it or load it!" (From documents).

Explorationists may need to know which wellbores have core samples (requiring one to search across different databases) and which cores are intersecting one or more stratigraphic units. Additionally, they can be interested in queries such as the following: "Does a rock core analysis exist for the unit? Does a sample analysis of permeability and porosity measures exist? Are there aggregated statistics of the selected stratigraphic units?" (From documents).

Working with geology, geographic information systems (GIS) and maps are the preferred way to have an overview of the data. When queried, accessed, and available to the data managers, data are to be presented coherently in GIS tools. In addition to well and seismic data, there are other types of data the explorationists need to be aware of, such as specialist studies (geochemical, geophysical, etc.), license administrative data, the deadlines for new licensing rounds, and so forth. These data are presented in GIS web maps. A data manager who worked on presenting these data coherently showed us how he formulated queries together with central data managers and used Python scripts to present the results in maps. He stressed the importance of visualizing the data that exist to the explorationists:

> […] there is a challenge that a lot of data exist that are not visible to the explorationists, and hence, they do not know it exists. There are data in the databases, but these are not shown. And if the no one checks, if it is not made visible, the explorationists will miss out on potentially relevant information. (Interview, data manager)

Far short of an automatized system that finds and displays relevant data, the data manager must both know the technology needed to provide an overview and use it to know as much about the data in an area as possible. Because each area and its data are unique, data managers work closely with the explorationists to have a feel of where the exploration is heading, as shown in the following interview:

> It is important for a data manager to work close with the explorationists and to know the area they are working in. Potentially, the data manager can know that there are data in an area that the explorationists do not know exist. (Interview, data manager)

Keeping data updated requires continuous interaction between data managers and explorationists, as a data manager explains:

> Sometimes, the explorationists are frustrated that the well data are not available for them in their project. But often, it is because of a data change in the underlying database, for instance, when a drilling operator change a well log. Then, it needs to be re-loaded into the project database. (Interview, data manager)

5   DISCUSSION

Knowledge infrastructures with instrument-generated data consist of diverse databases, analysis technologies, and diverse historical data that are combined with a continuous flow and production of new data. Whereas articulation work has previously been conceptualized as hidden and only visible when there are breakdowns, we find that knowledge infrastructures are incomplete in ways that necessitate a focus on repairing these small and continuous fractures. This is shown in our analysis that focuses on the data handling work in knowledge infrastructures and further adds to studies that consider the design of infrastructures [13], [25] and of data sharing in knowledge infrastructures [12], [13], [16], [40]. Here, repair work is not concerned with data in isolation, but rather, it includes the databases and various technologies. Guided by our research question concerning how practitioners cope with an incomplete knowledge infrastructure, we explored a case of oil exploration, one characterized by commercial and expensive data gathering, where there is a knowledge infrastructure consisting of instrument-generated data. Our findings indicate how repair in data-handling is concerned with keeping the infrastructure navigable and attending to countless details, as we discuss.

5.1   Repair keeps the infrastructure navigable

Previous insights on data work in knowledge infrastructures have shown how data sharing requires discussions and negotiations between collaborating parties [40] and the use of meta-data [12]. Here, our first contribution is to discuss how repair in data handling keeps the infrastructure navigable.

If sensory generated data in knowledge infrastructures are contradictory, different, or incomplete, what can the practitioners trust? We find that the primary concern in the exploration infrastructure is not simply to find data that can provide an answer; rather, it is the potential to gain new knowledge and insights by keeping the infrastructure navigable. The skills and technologies that enable searching, selection, reinterpretation, and archiving must remain connected. Compensating for the lack of trust in inherently uncertain data is a confidence in the ability to search for an answer within the frames of what the infrastructure permits. As we have seen, the exploration infrastructure drifts toward a state of disconnection. Graham and Thrift [6] argue that disconnection produces learning, yet "[…] disconnection is only possible if connection, or the possibility of connection is present, if a system of forces can be formed…." In our case, data handling extends the traditional technical realms that are typically ascribed to maintenance. Creating a merged seismic survey, for example, went beyond the purely technical work of merging two seismic files and creating a new one; it also involved repair activities, such as clarifying the ownership of the data, requiring accessing reports in different databases, investigating historical licenses, and clarifying the information with the pertinent stakeholders.

Engaging with neglect, errors, and breaks is also a source of innovation. New solutions may be invented [6], and repair allows for engagement with data and technology in new and surprising ways [9]. We saw how the data managers not only applied technology, but also actively developed new pieces of technology themselves, such as the dashboard tool that enabled exploration managers to monitor the status of data at decision gates. Without these new engagements with data and technology, exploration grinds to a halt. Difference and disconnections characterize infrastructures, and acts of repair signify the generativity of the forces that pull data and technology together.

The repair literature has shown how the state of a system requires negotiations with the state of the technology [33] and collaborative valuations of technological worth [8]. Our findings indicate how determining the state of data requires a shared understanding between data managers and data analysts. We found how it is a part of the data management craft to *know about* the data and technology. Data managers can look at meta-data (such as when a file was created and by whom) and know what IDs connect across databases; they are concerned about the sufficient quality control of the data to better connect historical data with current explorations. They know about the technology that can search for and transfer data from one piece of technology to another (such as visualizing data in GIS web maps). Navigating the infrastructure requires such craftsmanship. The data work of explorationists is different, however. Their intimate *knowledge of* the data implies that they can look at a seismic file and see that it does not fit in with the insights from more recent well data. What is notable in the current case is that the object of the work is the same: data. Although a data manager can easily specify the relations of database tables in an excel sheet (as one of the search tools requires), this is not the case for an explorationist. Similarly, a data manager cannot assess which seismic interpretation is to be archived at a decision gate. Considering this in terms of the processes of repair rather than disconnected data practices, it becomes clear that *what* data, knowledge, and insights that materialize and *how* they are brought forward are deeply entangled [8]. Repair within knowledge infrastructures molds a shared understanding between different skills and technologies to form sufficient navigability. Ultimately, this is what the practitioners can trust and what makes a knowledge infrastructure work.

5.2   Data handling is about the countless details

Extending on the studies that have considered cases of the repair of single artifacts [6] or where repair has been applied to recover from catastrophic breakdowns [20] [33], our second contribution is to discuss the ongoing and contingent characteristics of repair that are necessary to attend to the countless details of data handling [33].

The knowledge infrastructures of instrument-generated data can hardly be considered invisible. Here, repair work occurs at the very visible core of the business. Exploration operates on data–technology assemblies that not only mediate the data, but also "transform [...] that which it makes available" [28]. Exercising sufficient control of these transformations implies that data analysis is hard to separate from the work of making that data available. Our findings indicate how the core services of a knowledge infrastructure are difficult to untangle. This is partly because of the amount of differences and rates of change [2]. Historical data in different forms and formats across databases and archives can potentially be reworked and standardized (still a premise for automation); although, in practice, this is an insurmountable undertaking considering the volume and diversity of historic data. New data adhere to standardized reporting formats because the data must be archived in Diskos. But the tools and methods of data collection are continuously improving. Different vendors still deliver data in different formats and in

different ways, such as real-time data or when data are delivered directly from the vendor database before being quality assured in Diskos. These dynamics continue to intensify the need for repair [11] and necessitate configurations of data and technology that are challenging to plan and standardize (i.e., turn invisible), instead furthering the ongoing problem-solving processes [31].

Jackson [10] argues that repair as "broken-world thinking" signals a shift in our ways of thinking about systems renewal, moving moments of repair to the center of CSCW research and practice. Data handling in knowledge infrastructures symbolizes such a shift. Our findings indicate how renewal, in the form of knowledge and insight, relies on repair work. Consequently, data management gradually took on increased authority in OilCorp's exploration. Consider the dashboard system, for example. The quality assurance of data at the decision gates, previously more informally followed up on by data managers, rose to the level of the exploration managers. Discussing the delegation of tasks that follows from increasingly relying on technology for knowing, Ribes et al. [30] find that reversals of authority and responsibility often are "[...] rooted in the technical demands of computational systems." With the growing reliance on instrument-generated data as a means of knowing, repair work can be considered to move away from a separate and supportive role toward becoming an active part of core business processes.

In sum, the repair perspective developed here adds to Graham and Thrift [6], who caution "[...] the politics of knowledge repair [will try to] deskill repairers," that is, to transfer repairers' knowledge into software. Our analysis shows how such naïve politics of repair can be challenged by the incomplete, dynamic, and disconnected characteristics of the data and technologies involved. Edwards et al. [2] argue that simple technological solutions to what they call ontological incompatibilities are unlikely because "[...] we have not yet fully faced the implication of the basic infrastructural problem of maintenance." Focusing on data handling and being sensitive toward the characteristics of repair enables us to consider how knowledge emerging from knowledge infrastructures is not simply the production of something new, but rather is the result of countless acts of repair by which the "shape, standing, and meaning of objects in the world is produced and sustained" [9]. A repair perspective may be needed to detail how insight emerges through knowledge infrastructures and, ultimately, how exploration is possible.

6 CONCLUSION

In the current paper, we analyzed data handling in knowledge infrastructures as repair work that continuously deals with the inherent shortcomings in the data–technology assemblies that constitute the infrastructure. Through a case study of an infrastructure of instrument-generated data, we discussed how data handling involves repair that keeps the infrastructure navigable and deals with countless details. Keeping the infrastructure navigable implies that repair work extends beyond the pure technical domain by asserting ownership of the data, developing new technology, and connecting data handling and data analysis work. Dealing with countless details is contingent on the particularities of the problem-solving process, can contribute to a reversal of authority, and allows for the problematizing of naïve politics of repair. Future studies of repair in data handling can further detail such assemblies of data, technologies, and skills that are not easily standardized nor removed.

REFERENCES

[1] Pernille Bjørn, Morten Esbensen, Rasmus Eskild Jensen, and Stina Matthiesen. 2014. Does distance still matter? Revisiting the CSCW fundamentals on distributed collaboration. *ACM Transactions on Computer-Human Interaction* 21, 5, 1-27.

[2] Claus Bossen and Randi Markussen. 2010. Infrastructuring and ordering devices in health care: Medication plans and practices on a hospital ward. *Computer Supported Cooperative Work (CSCW): An International Journal* 19, 6, 615-637.

[3] Marisa L. Cohn. 2016. Convival decay: Entangled lifetimes in a geriatric infrastructure. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, 1511-1523

[4] Paul N. Edwards, Steven J. Jackson, Melissa K. Chalmers, Geoffrey C. Bowker, Christine L. Borgman, David Ribes, Matt Burton, and Scout Calvert. 2013. Knowledge infrastructures: Intellectual frameworks and research challenges. Working Paper. Retrieved March 13, 2017 from http://deepblue.lib.umich.edu/handle/2027.42/97552

[5] Kathleen M. Eisenhardt. 1989. Building theories from case study research. *The Academy of Management Review* 14, 4, 532-550.

[6] Stephen Graham and Nigel Thrift. 2007. Out of order understanding repair and maintenance. *Theory, Culture & Society* 24, 3, 1-25.

[7] Christian Heath and Paul Luff. 1992. Collaboration and control: Crisis management and multimedia technology in London underground line control rooms. *Computer Supported Cooperative Work (CSCW) (CSCW): An International Journal* 1, 69-94.

[8] Lara Houston, Steven J. Jackson, Daniela K. Rosner, Syed I. Ahmed, Meg Young, and Laewoo Kang. 2016. Values in repair. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ACM Press, New York, NY, 1403-1414.

[9] Steven J. Jackson. 2014. Rethinking repair. In T. Gillespie, P. Boczkowski, and K. Foot, eds. *Media Technologies: Essays on Communication, Materiality, and Society*. MIT Press, Boston.

[10] Steven J. Jackson, Alex Pompe, and Gabriel Krieshok. 2012. Repair worlds: Maintenance, repair, and ICT for development in rural Namibia. In *Proceedings of CSCW'12*, Seattle, Washington.

[11] Marina Jirotka, Charlotte P. Lee, and Gary Olson. 2013. Supporting scientific collaboration: Methods, tools and concepts. *Computer Supported Cooperative Work (CSCW): An International Journal* 22, 667-715.

[12] Helena Karasti and Karen S. Baker. 2008. Digital data practices and the long term ecological research program growing global. *International Journal of Digital Curation* 3, 2, 42-58.

[13] Helena Karasti, Karen S. Baker, and Florence Millerand. 2010. Infrastructure time: Long-term matters in collaborative development. *Computer Supported Cooperative Work (CSCW): An International Journal* 19, 3-4, 377-415.

[14] Heinz K. Klein and Michael D. Myers. 1999. A set of principles for conducting and evaluating interpretive field studies in information systems. *Management Information Systems Quarterly* 23, 1, 67-93.

[15] Brian Larkin. 2013. The politics and poetics of infrastructure. *Annual Review of Anthropology* 42, 1, 327-343.

[16] Charlotte Lee, Paul Dourish, and Gloria Mark. 2006. The human infrastructure of cyberinfrastructure. In *CSCW '06 Proceedings of the Conference on CSCW*, ACM, Banff, Alberta, 483-92.

[17] S. Leonelli. 2014. What difference does quantity make? On the epistemology of Big Data in biology. *Big Data & Society,* April-June, 1-11.

[18] John Van Maanen. 1988. *Tales of the Field: On Writing Ethnography*. University of Chicago Press, Chicago, IL.

[19] Matthew S. Mayernik, Jillian C. Wallis, and Christine L. Borgman. 2013. Unearthing the infrastructure: Humans and sensors in field-based scientific research. *Computer Supported Cooperative Work (CSCW): An International Journal* 22, 1, 65-101.

[20] Colin McFarlane. 2010. Infrastructure, interruption, and inequality: Urban life in the global south. In S. Graham (ed.) *Disrupted Cities – When Infrastructure Fails*. Routledge, New York, NY, 131-44.

[21] Naja Holten Møller and Pernille Bjørn. 2011. Layers in sorting practices: Sorting out patients with potential cancer. *Computer Supported Cooperative Work (CSCW): An International Journal* 20, 123-153.

[22] Eric Monteiro, Neil Pollock, Ole Hanseth, and Robin Williams. 2013. From artefacts to infrastructures. *Computer Supported Cooperative Work (CSCW): An International Journal* 22, 4-6, 575-607.

[23] Gerard Oleksik, Natasa Milic-Frayling, and Rachel Jones. 2012. Beyond data sharing: artifact ecology of a collaborative nanophotonics research centre. In *CSCW '12 Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, 1165-1174

[24] Andrew Pickering. 1995. *The Mangle of Practice: Time, Agency and Science*. The University of Chicago Press, Chicago, IL.

[25] Volkmar Pipek, Helena Karasti, and Geoffrey.C. Bowker. 2017. A preface to "infrastructuring and collaborative design." *Computer Supported Cooperative Work (CSCW): An International Journal* 26, 1.

[26] Volkmar Pipek and Volker Wulf. 2009. Infrastructuring: Toward an integrated perspective on the design and use of information technology. *Journal of the Association for Information Systems* 10, 5, 447.

[27] David Ribes and Jessica Polk. 2014. Flexibility relative to what? Change to research infrastructure. *Journal of the Association for Information Systems* 15, 5, 287-305.

[28] David Ribes. 2014. The kernel of a research infrastructure. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM Press, New York, NY, 574-587.

[29] David Ribes and Charlotte P. Lee. 2010. Sociotechnical studies of cyberinfrastructure and e-research: Current themes and future trajectories. *Computer Supported Cooperative Work (CSCW): An International Journal* 19, 3-4, 231-244.

[30] David Ribes, Steven Jackson, Stuart Geiger, Matthew Burton, and Thomas Finholt. 2013. Artifacts that organize: delegation in the distributed organization. *Information and Organization* 23, 1, 1-14.

[31] Daniela K. Rosner and Morgan Ames. 2014. Designing for repair? Infrastructures and materialities of breakdown. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, ACM Press, New York, NY, 319-331.

[32] Kjeld Schmidt and Liam Bannon. 1992. Taking CSCW seriously: Supporting articulation work. *Computer Supported Cooperative Work (CSCW): An International Journal* 1, 7-40.

[33] Benjamin Sims. 2009. A sociotechnical framework for understanding infrastructure breakdown and repair. Paper presented at the *Annual Meeting of the Society of Social Studies of Science*. http://public. lanl. gov/bsims

[34] Benjamin Sims and Christopher Henke. 2012. Repairing credibility: Repositioning nuclear weapons knowledge after the Cold War. *Social Studies of Science* 42, 3, 324-347.

[35] Susan Leigh Star and Karen Ruhleder. 1996. Steps toward an ecology of infrastructure: Design and access for large information spaces. *Information Systems Research* 7, 111-134.

[36] Susan Leigh Star and Anselm Strauss. 1999. Layers of silence, arenas of voice: The ecology of visible and invisible work. *Computer Supported Cooperative Work (CSCW): An International Journal* 8, 1-2, 9-30.

[37] Anselm Strauss. 1988. The articulation of project work: An organizational process. *Sociological Quarterly* 29, 2, 163-178.

[38] Geoff Walsham. 1995. Interpretive case studies in IS research: Nature and method. *European Journal of Information Systems* 4, 2, 74-81.

[39] Geoff Walsham. 2006. Doing interpretive research. *European Journal of Information Systems* 15, 3, 320-330.

[40] Ann Zimmerman. 2008. New knowledge from old data: The role of standards in the sharing and reuse of ecological data. *Science, Technology, and Human Values* 33, 5, 631-652