UNIVERSITY OF CALGARY

Virtual Exploration through a Philosophical Tool

by

Dohyun Kwon

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

GRADUATE PROGRAM IN CHEMICAL ENGINEERING

CALGARY, ALBERTA

MARCH, 2021

Abstract

As the era of big data emerges with digitalized devices, innovative data storage systems, and instant access to data, big data analysis technology has become a ubiquitous tool in our life opening up possibilities to resolve problems in various industries, services, and disciplines and experience our existence richly. However, in the era of big data, as technology evolves rapidly and becomes complicated and sophisticated to deal with big data, technology has been difficult for humans to understand and utilize it to analyze, interpret, and interact with data. This imbalanced phase between development of technology and the alienation of humans from technology is shrinking the possibilities rather than expanding.

To reconcile technology and data with humans and transform technology into a mediator, this research presents a) a philosophical framework that mediates between technology, humans, and data in the form of an interactive exploration pipeline, b) a self-explanatory virtual reality (VR) application system, and c) an immersive human-centered VR environment for problem-solving and cognitive tasks. To discover, identify, and provide new possibilities of applying the philosophical framework to human-computer interaction, we conduct exploratory evaluations on the VR environment. Built upon reflections on the evaluation results, this thesis ends with a remark of a potential impact of this thesis on future exploration in VR research.

Acknowledgements

I was honored to be accepted by the FORED-UCalgary Collaborative Scholarship Program and as a master's student of Dr. Zhangxing Chen, successfully commencing my thirties in my life. Since I have explored multidisciplinary research along with petroleum engineering, I could not have stepped forward without the endless support and help of my supervisor and my VR research team.

First of all, I would like to express my gratitude to my supervisor, Dr. Zhangxing Chen, for giving me an opportunity to navigate a new possibility of various fields of study, such as the philosophy of technology, AI, and VR, to petroleum engineering. Furthermore, I can never thank my VR research team enough, Stephen Cartwright, Steven Samoil, and Die Hu. Stephen has provided all the necessary resources for, constructive advice on, and broad perspective of my dissertation research. Steven has always been willing to help me whenever I am stuck with technical issues and kindly shed light on my predicaments in AI, VR, and my dissertation research. Amy, Die Hu, has spared no effort to help me. I greatly appreciate all the unconditional support and help guys! I will never forget it! In addition, I would like to thank the squash club for being a happy memory of my life in Calgary.

Lastly, I am sincerely grateful to my family for believing and supporting that I can disclose my potential to the world. With my parents' support, I can relish my life journey and focus on what I think is important in my life. A special thanks to my niece, 장예진, for your innocent smile. You are the only one that brings me happiness during my study :) 마지막으로 외할아버지, 친할머니, 그리고 둘째 큰아버지 하늘 나라에서 편히 쉬고 계시길 바랄게요.

Table of Contents

List of Tables

List of Figures

Chapter 1

1. Introduction

1.1. Problem Statement

As the era of big data emerges with digitalized devices and innovative data storage systems, the oil and gas industry has been utilizing emerging technologies to achieve production optimization of hydrocarbons [11]. The emerging technologies, such as big data analytics, artificial intelligence, augmented reality, and virtual reality, offer the promise of transformational solutions to domain problems by analyzing, interpreting, and visualizing collected data from geophysical, geological, operational, and production data sources. As a result, a key strategic initiative of the oil and gas industry is to develop a comprehensive data analytics tool with high accuracy, efficiency, and speed.

Although a comprehensive data analytics tool as an automated tool provides innovative solutions to domain problems, it is human experts that are central to analysis, interpretation, and interaction with data for the purpose of finding optimal solutions among suggestions from the automated tool. The deficiency of human aspects in technology exacerbates the imbalanced phase between development of technology and the alienation of human beings from technology. The detachment prevents human beings from disclosing the potential of themselves for shaping, manifesting, augmenting, and expanding their possibilities to think. To reconcile the imbalance, this research adopts the philosophy of technology as a foundation of the framework, mediating between technology, human beings, and complex domain data.

Domain 3D spatial data comprises a reservoir model with millions of cells containing reservoir properties (static and dynamic data) in a reservoir simulation stage. The multivariate, multiscale 3D spatial data manifests high dimensionality and heterogeneity, making it difficult for domain engineers to analyze, interpret, and visualize a relationship between static and dynamic data in a reservoir model. In addition, the problems raised by the large and complex 3D spatial data are expensive computational costs, inconsistent results with algorithms, and statistical biases.

Furthermore, the visualization of the domain 3D spatial data with desktop computer interfaces, such as a keyboard, mouse, and monitor, has fundamental limitations. When a reservoir model is geologically complex and heterogeneous and visualized in a desktop computer, it is difficult for human experts to interact with such high-dimensional data in the desktop interfaces. This is ascribed to the data that exceeds the human space perception ability to analyze, manipulate, and interpret the data. The traditional desktop-based interfaces are spatiotemporally restrictive for domain engineers and decision makers for exploring the large 3D volume data with high dimensions, collaborating with various disciplines, and making spatial judgements effectively in the limited information space. The above limitations require a comprehensive data analytics tool for big 3D spatial data with high dimensions and heterogeneity and interactive visualization with effective user interface technology.

1.2. Research Objective

Built upon the philosophical consideration in a relationship between technology and human beings, the main focus of this research is to design, implement, and evaluate an interactive exploration pipeline where a user can autonomously perform exploratory data analysis in a virtual world by supporting sensemaking processes to discover new knowledge and gain insights from 3D spatial data with high dimensions and heterogeneity. The whole framework of this thesis mediating data, a human being, and a virtual world enables the insights gained from the pipeline to expand an intellectual and experiential horizon of each individual user. To achieve this objective, there is a need to embrace a paradigm shifting design that integrates a multidisciplinary approach. This research embodies the in-house black oil simulator (BOS), artificial intelligence (AI), virtual reality (VR), the philosophy of technology, visual analytics, and human-computer interaction (HCI).

To design such an interactive exploration pipeline, this research starts with investigating what data is (Section 2.1.1) and current trends and features of data (Section 2.1.2) since every visualization pipeline begins with data and ends with a user. The following sections characterize each component of the pipeline in order to approach a defined problem effectively and select optimal methodology among many possible options [17]. The characteristics of the domain data (Section 2.1.3) necessitate interactive data visualization for knowledge discovery (Section 2.2.1), cluster analysis for data simplification (Section 2.2.2), and virtual reality for an effective user interface (Section 2.2.3). For smooth transition from reality to a virtual world and improvement of task performance in a virtual world, careful design of human-computer interaction is essential for users to gain insights from data. In this regard, this research adopts the philosophy of technology as a bridge between data, VR technology, and users (Section 2.3). To design a sound and successful visualization pipeline, user tasks are analyzed in Section 3.3.3 from interactive data visualization and domain perspectives (reservoir engineering). Section 3.5 implements all of the components of the visual pipeline in a VR environment. To determine the usability, robustness, and precision of the pipeline, the VR environment is evaluated through qualitative study (Chapter 4).

1.3. Background

1.3.1. Reservoir Simulation

A *digital simulation* is a world where human beings can emancipate and surpass their physical, cognitive, and interactive constraints imposed on them by the actual world. It enables humans to design and rationalize their hypotheses and ideas in relation to the actual world without actual risks, allowing them to develop virtual, alternative, and practical scenarios that are physically impossible or dangerous to implement in the actual word [35]. In this regard, reservoir simulation predicts performance of a reservoir naturally situated on the subsurface containing hydrocarbons under various operating conditions. To reflect actual flow behavior in a reservoir, a reservoir simulator is based on physics, mathematics, reservoir engineering, and computer programming [68]. Figure 1.1 shows the predictive technologies and reservoir simulation modeling utilized to predict reservoir performance.

There are three computational components comprising reservoir simulation: a preprocessor, reservoir simulator, and post-processor. After identifying available resources and setting viable objectives and basic strategy, a pre-processor as input of a reservoir simulator builds a reservoir model. A reservoir model as input includes a geological structure, heterogeneous and anisotropic rock properties, variations of reservoir fluid properties, recovery mechanisms, a coordinate system, the number of dimensions in space, history matching, and the number of phases. Once a category of the reservoir simulator input has been decided, a reservoir simulator tailored to the category runs a reservoir model to evaluate various production plans and various conditions of reservoir parameters and perform sensitivity analysis of the various production plans and parameters [69]. A post-processor enables reservoir engineers to understand the results in an intuitive and interactive manner by accessing, analyzing, and visualizing the results.

As interactive visualization technologies advance [70], a post-processor traditionally consisting of desktop-based interfaces has adopted alternative interfaces, such as AR, VR, and CAVE, to provide more powerful, intuitive, and collaborative interaction and improve task performance. In this regard, a post-processor as the component of reservoir simulation allows more intuitive exploration of big volume data and its complexity in increased information space. This feature provides reservoir engineers with new insights leading to improvements in their analytics capabilities. Although a human expert as a decision maker is the end user of reservoir simulation, the oil and gas industry has focused on technical issues of a post-processor, failing to consider human aspects of the post-processor. This motivates this research to embrace the philosophy of technology as a solution to the problem of the inadequate HCI design.

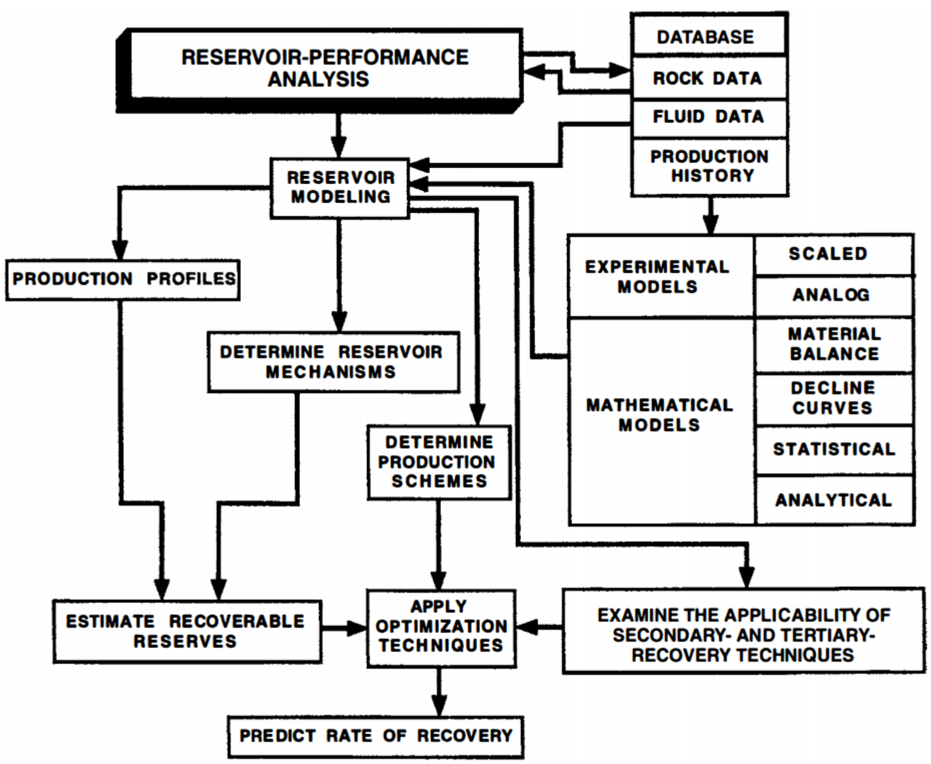

Figure 1.1: Predictive technologies and reservoir simulation modeling utilized to predict reservoir performance [69].

1.3.2. Philosophy of Technology

Having its roots in ancient Greek philosophy, the philosophy of technology is a subfield of philosophy that critically reflects on technology. Since the industrial revolution had a huge impact on various aspects of contemporary society, such as the economy, technology, ethics, and politics, in both a positive and negative manner, the philosophy of technology has been concerned with the influence of technology on the human aspects of society rather than with technology itself [71, 72]. This more human-oriented trend was named by Mitcham the "humanities philosophy of technology" in 1993. From this perspective, the objectives and values of humans initiate and create technology, not vice versa [75].

Instead of understanding and analyzing technology itself, the traditional humanities philosophy of technology studied the relationships between technology and the humanities. In particular, the German philosopher Martin Heidegger focused on the relation of technology to human beings from a metaphysical perspective. He viewed technology as an autonomous force trying to control mankind and viewed human beings as being involved with this autonomous force. With human beings depending on technology, they fail to reveal their potential to surpass their cognitive and operative limitations. In this classical phenomenological approach to the traditional humanities philosophy of technology, human beings take the reductive way of perceiving the world as raw materials at their disposal [35, 71].

Unlike the techno-pessimistic and classical phenomenological perspective, extending Heidegger's view in a more neutral manner, Ihde and Verbeek developed postphenomenology. The postphenomenological approach to the philosophy of technology viewed technology as mediators, unveiling its potential for and impacts on human beings. This approach understands that technology is not a reductive drive leading human beings to isolation from the possibilities associated with participating in abundant existence, but a bridge mediating between human beings and the possibilities. Technology is thus identified as a bridge for human beings to disclose these possibilities, construct their own understanding from reality, and be constructed in return [35, 73, 74].

Embodying the humanities philosophy of technology, this research materializes this philosophy in a virtual world as a mediator of a simulation through VR technology. A simulation opens up a virtual world with possibilities for humans to experimentally manifest, shape, extend, and expand their perceptual and cognitive processes. Embracing VR technology, a virtual world mediates between the simulation and the user, enabling the user to interactively frame, augment, and apply their thoughts on the simulation. Thus, the relationship between a simulation, virtual world, and user shows them to be interactively coexisting and mutually affecting each other. In this research, this relationship will serve as a philosophical framework of VR design.

1.4. Methodology

Built upon the philosophical framework and the characterizations of each component of the interactive exploration pipeline, this thesis represents a philosophized VR environment for exploratory interaction of 3D spatial data. With the foundation of Martin Heidegger's philosophy as the philosophical framework, the visualization pipeline consists of initial, middle, and final stages. Figure 1.2 describes each stage (the scope of 'Computer', 'VR', and 'Human', respectively), the whole components of the pipeline, and how the VR application works. The philosophical framework mediates between a computer, virtual world, and human being in a harmonious manner.

The initial stage consists of data, visual mapping, cluster analysis, and task analysis. Utilizing The Ninth SPE Comparative Solution Project as a dataset [41], the in-house BOS [42] builds a reservoir model and outputs the visualization files constructed by the Visualization Toolkit (VTK) format representing grid information and physical variables at each time step. Cluster analysis is implemented in the VTK static reservoir model to involve a user in the process of pattern discovery. Hierarchical Density Based Spatial Clustering of Application with Noise (HDBSCAN) is selected for cluster analysis in terms of domain dataset size, run time, scalability, and noise identification [46, 54, 57]. In order for a user to perform exploratory data analysis effectively, task analysis is performed from both the domain (reservoir engineering) and interactive data visualization perspectives.

The middle stage is designed for immersive analytics in the big volume of 3D spatial data. Unity renders the VTK spatial and non-spatial data with HDBSCAN and materializes the task analysis in a VR environment [76]. The VR environment is a VR information space which has a relationship between object space (3D spatial data), attribute space (its properties), data structure space (clustering algorithm), and user experience space [65]. Oculus Quest is utilized for user interaction with the VR environment. To mediate between the complex 3D objects (the static and dynamic models) and a user in the VR environment, this research embraces postphenomenology and ontology reified in the form of a game-based approach.

The final stage represents the integrated components of the pipeline in the VR environment. The VR environment becomes a world where users can interact, manipulate, and control the 3D spatial data and experimentally shape, expand, and apply their understanding to generate or validate their hypotheses and theory. To guide users to initiate the perceptual and cognitive processes in the VR environment, the game-based approach is implemented in the form of a world with a theme and goal-oriented interaction. A world with a theme helps a smooth transition from reality to the virtual world. This transition enables users to be immersed in the virtual world, eliciting autonomous user behavior to understand a goal, data, and user tasks of the VR application. Built upon the immersion and information, the goal-oriented interaction activates cognitive and problem-solving activities, which lead to knowledge discovery through sensemaking from the scientific visualizations [23, 39, 64].

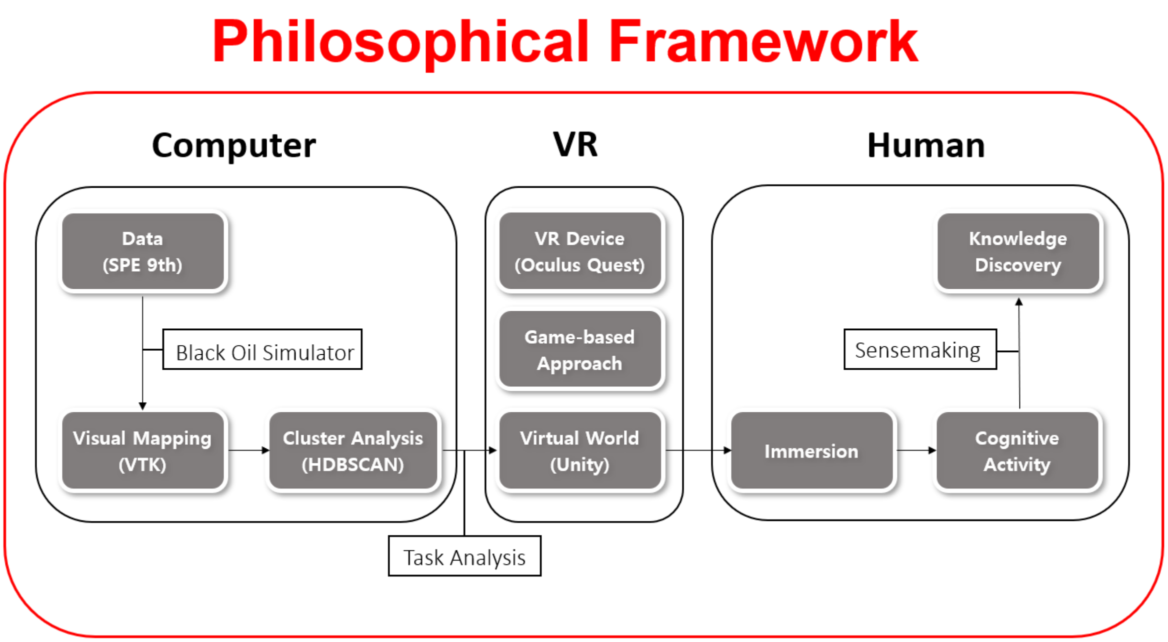

Figure 1.2: Design overview of the interactive exploration pipeline with the associated technologies.

1.5. Contribution

The main contributions of this thesis are:

- A philosophical framework that mediates between technology, human beings, and data.

- A self-explanatory VR application system in the fields of science and engineering.

- An immersive human-centered VR environment for problem-solving and cognitive tasks.

1.6. Thesis Overview

This introductory chapter presented that the stated problem necessitates the development of the interactive exploration pipeline embedded with the philosophy of technology for reservoir simulation in order to mediate between VR technology, human beings, and complex domain data. The following chapters will describe each component of the interactive exploration pipeline (Chapter 2), the embodiment of the philosophy in the pipeline and the implementation in a VR environment (Chapter 3), the evaluation of the pipeline (Chapter 4), and the conclusions (Chapter 5).

Chapter 2: Background – Regardless of any field of study, data is a foundational element and essential step for research. Considering the significance of data, the question of “what is data?” is generally neglected and insufficiently explored since the concept of data always remains abstract, enciphered, and ambiguous. In this regard, Chapter 2 first investigates the above question in order to design a successful and robust visualization pipeline. Alongside this, current trends and features of data (i.e., big data) are introduced. Built upon this foundation resulting from the investigation of what data is and big data, this chapter characterizes domain data to approach a defined problem (detailed in Chapter 2) effectively and select optimal methodology among various possible options. The results of the domain data characterization comprise each component of the pipeline. After introducing the initial, middle, and final stages of the pipeline, this chapter describes a philosophical framework as the foundation of this research, why this philosophical framework is essential for HCI, what this philosophy means for the relationship between a human being, mediator, and virtual world, and how this philosophical framework shapes, materializes, and applies to human-computer interaction.

Chapter 3: Virtual Exploration through a Philosophical Tool – Chapter 3 starts with characterizing the initial stage of the visualization pipeline. Data as the input of the pipeline is described and characterized, and then HDBSCAN for cluster analysis is introduced and demonstrated as the optimal solution for the data simplification. Based on the results of the analysis, this chapter investigates and characterizes user tasks in a VR environment from both domain (reservoir engineering) and interactive data visualization perspectives. For the middle stage, this chapter describes why the philosophical framework is essential for the design of the pipeline and how the philosophical framework is materialized in a VR environment. A game-based approach is introduced as the embodiment. It enables a user to be immersed in and interact with the VR environment and perform cognitive, perceptual, and intellectual activities in the VR environment. The final stage materializes the integrated components of the pipeline in the VR environment where a user autonomously performs exploratory data analysis.

Chapter 4: Evaluation – To discover, identify, and provide new possibilities of applying Heidegger's philosophy to human-computer interaction in a virtual world, Chapter 4 conducts exploratory evaluation on the philosophical framework of this thesis through a qualitative study (i.e., a questionnaire, interview, and observation of user behavior in the virtual world). There are three evaluation criteria of the usability of the VR application: the structural mediator (i.e., game-based approach and self-explanatory environment), functional mediator for user tasks performance (i.e., foraging loop and sensemaking loop), and being-in-the-world for whether this VR application expands experience and knowledge of users in real life. Finally, this chapter provides reflections on the findings.

Chapter 5: Future Work and Conclusions – The future work section discusses three points that will improve this research: VR application features based on user feedback, a quantitative study for evaluation of the VR application, and machine learning applied in a visualization pipeline to optimize the interactive exploration pipeline. Built upon reflections on the era of big data, the conclusion section presents and illustrates main contributions of this thesis: a philosophical framework that mediates between technology, human beings, and data; a self-explanatory VR application system in the fields of science and engineering; and an immersive human-centered VR environment for problem-solving and cognitive tasks. Furthermore, comparing Martin Heidegger with René Descartes, the conclusion section shows that Heidegger's philosophy can resolve the imbalanced phase between development of technology and the alienation of humans from technology. Finally, the conclusion section ends with a remark of potential impact of this thesis on future exploration in other research.

Chapter 2

2. Background

2.1. What Is Data?

Regardless of any field of study, data is a foundational element and essential step for research. Considering the significance of data, the question of “what is data?” is generally neglected and insufficiently explored since the concept of data always remains abstract, enciphered, and ambiguous. Before embarking on a discussion of the main problem of this research, this thesis first describes how data emerges, why this illumination is crucial, and what implications offer to analysis of the problem. Alongside this, current trends and features of data are introduced. Then, by applying the above approaches to domain data, Section 2.1.3 characterizes it in order to design a successful visualization pipeline.

2.1.1. Where Does Data Come from?

There are countless definitions of data, but they generally converge into a dichotomy: data exists in a processed or raw form. From the perspective of a processed form, the appearance of data unveils at the last step of the cognitive, systematic, and automatic process. To be specific, a processed form indicates that data exists preceding knowledge and information. Figure 2.1 describes the reversal of the hierarchy relationship between knowledge, information, and data. On the other hand, the raw form of data indicates a hierarchy relationship between data, information, and knowledge.

From the perspective of a traditional hierarchy (raw form), data is rudimentary and unprocessed. Data consists of unorganized numbers and structures which need to be interpreted. It is waiting to be processed, meaningful, and useful. In this sense, data describes nothing but its existence. In other world, there is nothing rawer than data. People have to define and manifest the potential of data. The concept of information is built upon the above assumptions about data. After the potential of data is defined, information is constructed out of data as its materials, and then forms a specific correlation between data. Once information establishes the relationships between data, it reveals its value as the basis to describe a specific phenomenon, problem, or state. When information accumulates, knowledge as an implicit agreement is accepted as the basis of analysis on a specific phenomenon, problem, or state and is not subject to doubts about its authenticity. As a result, knowledge originates from data and information [4, 5, 6, 7].

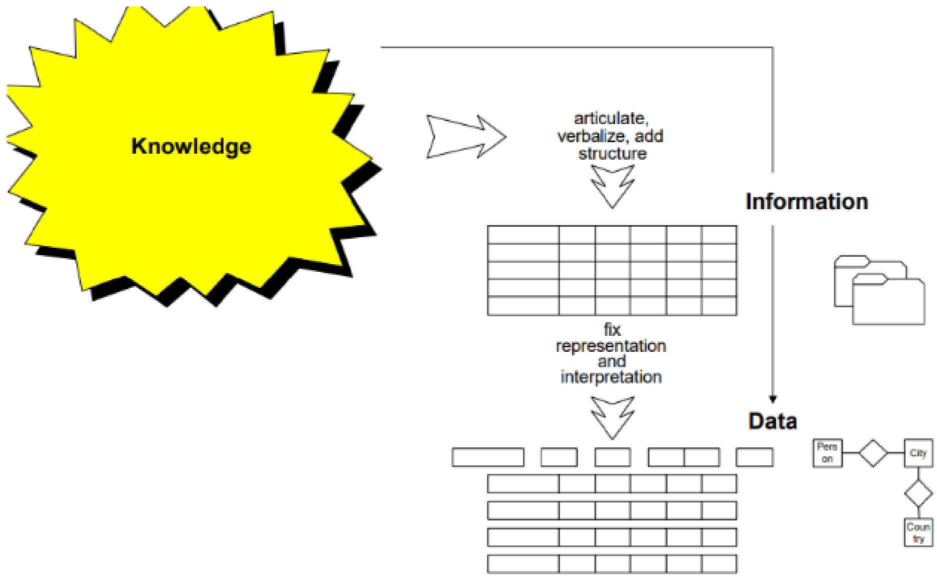

Figure 2.1: Reversal of hierarchy between knowledge, information, and data [7].

In contrast to the above traditional framework, from the perspective of the reversed hierarchy, data is the result of the reflection of knowledge and information. Knowledge and information are stored in the form of data through series of a cognitive process. Data situates knowledge and information as its foundation. Knowledge is the rudimentary element of data. Knowledge consists of proven facts and scientific laws resulting from experiments and observations of phenomena. In science, knowledge is facts accepted by a majority of scientists, often in the form of mathematical descriptions. For example, Newton's Law of Gravity as knowledge is a fact described by a mathematical relationship. When rudimentary elements of knowledge are reasonably combined and constructed, cognitive efforts shape the logical structure of the elements. The integrated structure of knowledge becomes information [7].

To transform information into data, information as a structure of knowledge is destructured through automatic and systematic processing under the direction of syntax and semantics. Syntax reveals how information should be broken down into fragments. Semantics serializes information in terms of interrelationships between fragments [39]. For example, by combining relevant knowledge of scientific laws from fluid dynamics and rock mechanics, a permeability measurement tool is created. The tool digitally measures a rock's ability to transmit fluids, resulting in permeability data. In this regard, embedding the relevant knowledge, the tool integrates and structures the relevant knowledge into information, which defines what permeability is. The tool systematically and automatically stores the stream of the data in the form of a file. The file consists of the sequence of the permeability data. From the perspective of the hierarchy relationship, the data seems rudimentary, unprocessed, and raw without the background of the relevant knowledge and the tool. This example illustrates that building a structure of knowledge, information is materialized in a measurement tool. The result of the measurement is data and stored in a digital architecture containing the knowledge and information. Thus, from the perspective of the reversed hierarchy, data is created from knowledge and information.

In this section, the hierarchy relationship and the reversed hierarchy were described to investigate the nature of data. Section 2.1.2 describes the nature of big data to understand current trends of data. Built upon the reversed hierarchy and the features of big data, Section 2.1.3 demonstrates how knowledge and information produce domain data to better understand and characterize domain data.

2.1.2. Big data

The era of big data emerges with digitalized devices, innovative data storage systems, an ecosystem of instant access to data, and big data-processing technologies. By taking advantage of the abundance of data, the era unlocks gates leading to help everyone perceive how the world works. For example, various industries, services, and disciplines utilize big data: business, advertising, marketing, data measurement, education, online dating services, political science, public health, retailers, sports, social media, and social networks. Furthermore, the big data trend changes the decision-making process. Decisions should be made based on a data-first, data-guided, and data-driven process [8, 9, 12, 14, 15].

The first step of big data analysis is to deal with continuously produced and stored data from various sources at different times and formulated by multiple methods. This indicates that big data consists of high dimensional and massive sample data, which is infeasible to scale in traditional analysis. The nature of the high dimensionality and large samples causes complicated predicaments, which is no longer applicable for traditional technologies to deal with the predicaments. The problems raised by the malfunction of the traditional methods are expensive computational cost, an inconsistent result of an algorithm, and statistical biases. To understand the nature of big data in an interpretable and scalable way, the disciplines of optimization, statistics, and applied mathematics have come together to determine four distinct characteristics of the nature of big data [13]:

1. Heterogeneity

Heterogeneity is diversity resulting from different data sources and a representative feature of big data. It is difficult to discover commonality between the various data sources since each data source has its own distinctive features. This situation makes it more difficult for traditional analysis to handle the heterogeneity. When traditional techniques that work well with a small sample size are applied to big data, the resulting model constructed by the shortage of a sample size considers unique features of each data source as anomalies. In this respect, massive sample sizes of big data alleviate the heterogeneity problem to discover common features across whole populations. When subpopulations of big data become large enough, statistical models can detect the unique features. For example, when large sample sizes are available, finding fatal factors on certain diseases can be more accurately inferred.

2. Noise accumulation

Noise accumulation occurs due to synchronization of analysis on multiple features. Involving numerous parameters at a time causes low prediction or classification rates, especially when analyzing enormous data samples. Most of these parameters may have little impact on the classification. In other words, in high dimensions, only a few features, which are unique and important, may have significant effects on the classification. This could lead to a false classification and aggressively encroach the realm of the true labels when a model simultaneously trains numerous features from different sources of data. To improve the performance of the prediction, variable selections need to be considered by optimizing the most distinctive features out of noise data.

3. Spurious correlation

In big data, there is the possibility to find an unexpected contributing factor to prediction in massive and multiple datasets, although the discovery might turn out to be scientifically trivial and insignificant. This phenomenon is caused by unmeasurable variables that are randomly distributed and may not share any commonality and interconnections in big datasets. The random variables may show a high correlation in high dimensions. But such a correlation may be spurious, which can have an effect on misleading variable selection from big data, resulting in scientifically fictitious explorations in a high dimensional space. The higher the dimensions, the greater the possibility of a high correlation. In high dimensions, a relationship between prominent factors affecting prediction of a model and spurious variables shows significance. This phenomenon indicates that these factors are inseparable from factors that apparently have a strong connection with prediction but actually have no involvement with prediction in terms of scientific discoveries.

4. Incidental endogeneity

In general, an endogeneity problem happens due to unobserved heterogeneity, omitted variables, and measurement errors in a statistical model. To be specific, when analyzing big data with a regression model, endogeneity means there is a correlation between the explanatory variable (X) and the error term. The relationship arises from coincidence. There are two reasons why this problem occurs in big data:

1) To develop a true model with high accuracy rates, a statistical model trains as numerous data as possible. In the process of collecting and training massive data, random X variables might have a high possibility of relationship with the error term. This means that the possibility of the correlation increases as data is collected and the model trains the huge dataset.

2) Big data consists of unrelated and various sources of data. The nature of the mixture results in the increased possibility of an accidental correlation between variables by decreasing the accuracy of the optimal variables selection and measurement processes.

2.1.3. Domain Data Characterization

Reservoir rocks containing oil and gas are naturally situated in the subsurface. To represent the geological information and its features, a reservoir model consists of 3D spatial data with millions of cells containing reservoir properties such as porosity, permeability, saturation, and pressure. The compact spatial data in the cells has numerous parameters and is the target domain of analysis in a simulation stage. Before analyzing the model, characterization of the domain spatial data is a necessary task in order to approach a defined problem effectively and select the optimal methodology among numerous possible options [17].

There are two main characteristics of the domain spatial data: the reversed hierarchy system (Section 2.1.1) and features of big data (Section 2.1.2). Firstly, when characterizing the domain spatial data by utilizing the concept of the reversed hierarchy system, it is important to remember that the sequence of the reversal of the hierarchy relationship is knowledge, information, and data. In this section, each component of the relationship is elaborated to demonstrate how the domain data emerges at the last stage of the reversed hierarchy system based on knowledge and information. After illuminating the data generation process from knowledge and information, the process applies to domain data characterization by demonstrating how the domain knowledge (Black Oil Simulator), information (The Visualization Toolkit), and data (a 3D reservoir model) are correlated.

From the reversed hierarchy system perspective, a system of domain 3D spatial data is structured by a top-down design: knowledge, information, and data. In line with the structure, domain spatial data is derived from domain knowledge and information. Domain spatial data is seen to be simply a result of reservoir simulation in the form of a text file. However, when looking into the data generation process, the text file implicitly contains domain knowledge and information. Automatically implemented by computational logic, the data generation process generates, stores, and manages domain spatial data systematically based on domain knowledge and information. At the end of the reversed hierarchy system, utilizing domain knowledge and information, the automatic and systematic process results in domain spatial data. Since the data generation process is implemented by the automatic processing mechanism, data as the final output of the process looks raw, rudimentary, and unprocessed. This abstractness in data makes most researchers ignore the significance of the implications of data. To help to prevent this negligence and select an optimal solution, this section characterizes domain data by utilizing the reversed hierarchy system.

From the domain perspective, knowledge, information, and data are given by the in-house Black Oil Simulator (BOS) developed by the Reservoir Simulation Group [42], the Visualization Toolkit (VTK), and a 3D reservoir model, respectively. BOS as knowledge contains fundamental principles of multiphase flows in porous media. After the principles are implemented in BOS through computational processes, the processed knowledge becomes information. Information as the output of BOS is represented as VTK. VTK is an open-source class of libraries for scientific and information visualization. VTK manifests and describes spatial data as various types of grids with scalar, vector, and tensor features [18].

Containing spatial and non-spatial information, VTK automatically and systematically serializes the information into numbers under the direction of syntax and semantics. The numbers represent 3D spatial data, such as point data, cell data, and cell scalar data in the form of a text file (described in Section 3.3.1). This data as the end stage of the data generation process is from the processed knowledge and information in the reversed hierarchy system. The data is not only seemingly unorganized, but also intuitively difficult to understand. This problem is ascribed to an automatic and systematic computational process that generates, stores, and manages data by compacting knowledge and information. Considering the abstract and complex nature of the 3D spatial data generation process, an immersive virtual reality (VR) (described in Section 2.2.2) is regarded as the best visual and intuitive interface to represent, interact, and communicate 3D spatial data [19].

For the second characteristic of domain data, domain 3D spatial data shares the features of big data: heterogeneity and noise accumulation (described in Section 2.1.2). Like big data emerging from large samples and high dimensionality, domain 3D spatial data involves large samples and high dimensionality: massive spatial and non-spatial information formulated by multiple methods and generated by different sources of data, such as seismic geophones, well logs, and core samples. The multivariate and multiscale data causes heterogeneity of domain 3D spatial data, which makes it difficult for domain experts to discover knowledge due to the complexity of the spatial information. Furthermore, the multivariate and multiscale data leads to noise accumulation. This indicates that a small number of contributing parameters determine performance of flow behavior simulation. As a result, 3D spatial data with heterogeneity and noise accumulation becomes beyond the grasp of humans to understand, interpret, and analyze. For this reason, a cluster analysis approach is selected as the optimal solution (described in Section 2.2.3) to deal with such data. This approach supports better recognition of big 3D spatial data by extracting hidden patterns and correlations between spatial and non-spatial data [16].

To visualize a reservoir model intuitively and offer spatial information effectively to domain experts, it is essential to carefully design visualization processes, which starts from data and ends with users. The sequence of stages is a *pipeline*. This section characterized domain data to decide what technology should be included in the pipeline to provide an intuitive and tangible VR environment and assist domain experts effectively in the VR information space. The results of the domain data characterization (i.e., interactive data visualization, cluster analysis, VR technology, and human-computer interaction) comprise the sequence of the pipeline. The next sections describe each component of the visualization pipeline that embraces the above elements.

2.2. Visualization Pipeline

In this thesis, the purpose of visualization is to mediate between data and users by supporting them to perform exploratory data analysis. Exploratory data analysis enables users to identify structures, patterns, anomalies, trends, and relationships and discover new knowledge through visualization. To accomplish these goals, design of the visualization pipeline consists of the following components: data, data visualization, cluster analysis, user interface, and human-computer interaction [23, 39].

The visualization pipeline consists of an initial stage (data, data visualization, and cluster analysis), a middle stage (user interface), and a final stage (human-computer interaction). To design a successful and robust pipeline, every visualization pipeline begins with characterization of data. As described in Section 2.1.3, the characteristics of the domain data necessitate intuitive data visualization (Section 2.2.1), cluster analysis for data simplification (Section 2.2.2), and virtual reality as a user interface (Section 2.2.3). Considering the complex nature of scientific visualization and its relation to user tasks materialized in virtual reality, careful design of human-computer interaction as the final stage is essential for users to discover knowledge and gain insights from data. Each component of the pipeline will be described in the following sections.

2.2.1. Data Visualization

Data visualization is an intuitive medium of cognitive and perceptual processes to gain insights from datasets that represent systems, ideas, and events. In the oil and gas industry, visualizing 3D reservoir models situated in the subsurface is increasingly utilized to create better visualizations of reservoirs and develop more intuitive ways to explore them [20, 21, 22]. However, data visualization is always ambiguous and intangible with respect to gaining “insight”. As an initial step to search for “insight”, we can adopt a core idea from philosophy: asking fundamental questions. Philosophy starts with a question to acquire insights from phenomena. Similarly, we initiate questions about visualization to gain insights from datasets. In data visualization, there are two general types of questions about a stated problem [23]: a specific and a fuzzy question.

For a specific and quantitative question about a given problem, visualization can offer intuitive and instant answers. A question to consider is, “what and where is maximum, minimum, or a specific range of values and points in 2D spatial data?” The obtained answers to this question by visualization can lead to insights for data distributions, correlations between different datasets, and data trends. In contrast to 2D spatial data, the value of visualization becomes more salient when dealing with 3D spatial data. The above features of 2D visualization are well described by 3D visualization. 2D visualization consists of pixels, and 3D visualization consists of voxels. Since the volume data can represent one or more data dimensions, it can visualize structure, a continuous phenomenon, and patterns within the data [39]. The visualized information supports the discovery of implicit knowledge not explicitly described in datasets. Furthermore, it leads to new questions and a more in-depth articulation of the original question, thus resulting in a more profound conception of the datasets. Collectively all of these cognitive processes can be referred to as “insight.”

The meaning of a fuzzy and qualitative question in data visualization is obscure. In other words, “the what” and “the how” are unknown. Data structure and its numerous parameters of a complex dataset and dynamic behavior of massive data cause the problem of “the what” and “the how”. In most cases, the problem of “the how” is manageable by employing automatic processing techniques, such as machine learning, data mining, and artificial intelligence [38]. Due to fuzziness, “the what” to explore in a given dataset is difficult to answer without human involvement. Even if the automatic techniques (e.g., pattern recognition, classification, and prediction) sort out a component of a given problem, the final decision is totally impacted by human interaction. For example, a fuzzy and qualitative question is: are there potential areas that may lead to optimal oil production? Cluster analysis can be implemented to discover patterns of a target property in this example. Built upon the result, human experts specify the areas based on their historical knowledge of geology and reservoir engineering. Considering the nature of a fuzzy and qualitative question, it is imperative to find a new interface, which can integrate human experts' knowledge, automatic processing techniques, and interactive exploration.

The problem of "the how" arises in terms of how to gain insights by visualization in a fuzzy and qualitative question. In general, based on a cognitive task analysis, there are two ways to gain insights from visualization in a fuzzy and qualitative question: bottom-up and top-down processes [64]. Figure 2.2 shows a sensemaking loop describing various stages of cognitive activities when a user performs visual tasks. Following the taxonomy of [64], bottom-up processes consist of five stages in Figure 2.2: *Search and Filter, Read and Extract, Schematize, Build Case,* and *Tell Story.* These processes start from data to theory. Firstly, users explore visualized data and collect evidence to make inferences, or support or invalidate theory. In the process of *Search and Filter, Read and Extract,* and *Schematize*, users can form new hypotheses. Utilizing the collected evidence, cognitive processes, and a given visual application, users build a case or theory to support or invalidate their hypotheses.

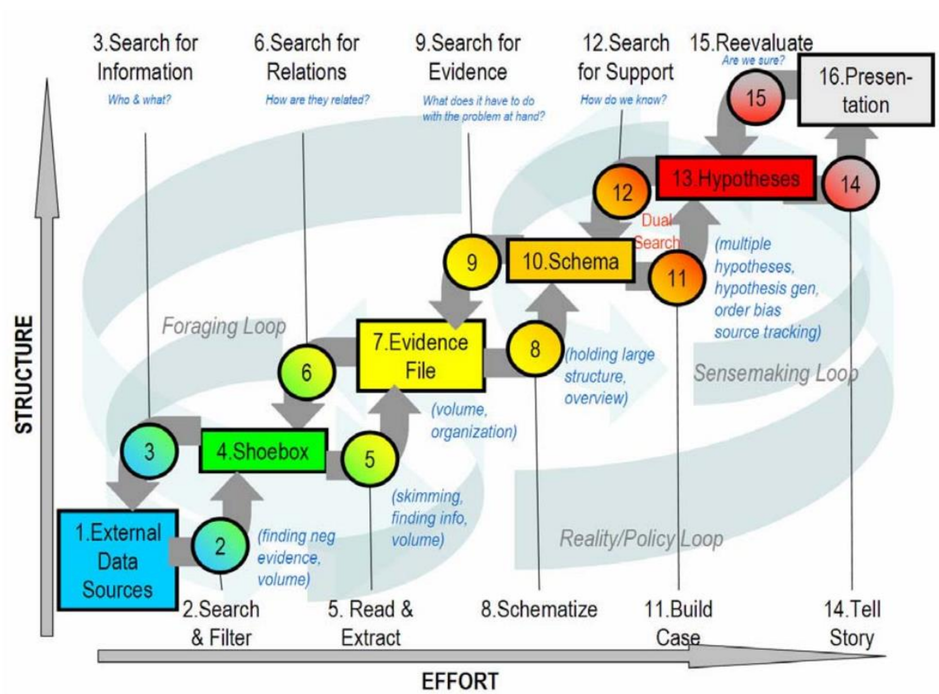

Figure 2.2: Sensemaking loop from bottom-up processes to top-down processes [64].

Top-down processes start from theory to data and consist of five stages: *Re-evaluate, Search for Support, Search for Evidence, Search for Relations*, and *Search for Information*. If users have their concrete theory, hypothesis, or assumption to support or invalidate or want to test alternative theories, this is the first stage of *Re-evaluate.* Users can systematically break the other four stages down by re-examining cognitive processes, collected evidence, piece of evidence, and visualized data [64]. Implementing the bottom-up and top-down processes by harnessing the visual application can help to extract unidentified and unknown facts and correlations from the datasets or generate new hypotheses. This is considered to be the second “insight” from a fuzzy and qualitative question. Following [64], the Pirolli-Card Sensemaking Loop (i.e., Figure 2.2) will serve as a foundation of this research to describe various stages of cognitive activities.

Considering the above different approaches to the specific and fuzzy questions from visualization in order to acquire insight, it is paramount to design an exploratory visualization environment, especially when involving human-computer interactions. In order to achieve this goal, firstly, cluster analysis is implemented in the environment for simplifying data. Secondly, philosophical tools (postphenomenology and ontology) are embedded in the environment for effective human-computer interactions (Section 2.3).

2.2.2. Cluster Analysis

In terms of data analysis, machine learning is a set of tools for understanding data to solve a practical problem. The tools can be categorized into four types: supervised, unsupervised, semi-supervised, and reinforcement learning. Cluster analysis falls into unsupervised learning. The goal of unsupervised learning is to discover unknown relationships and structure from data. Although there are various goals using cluster analysis [2], this thesis utilizes clustering analysis for finding interesting patterns in data. For this purpose, exploratory data analysis is selected.

Unsupervised learning is difficult to evaluate the quality of the results for two reasons. Unsupervised learning models only need observations, which indicates there are no labels to evaluate the quality of the models. Furthermore, unlike supervised learning, since it is impossible for unsupervised learning models to utilize labels representing true answers and the predicted outputs of the models, constructing an evaluation mechanism like cross-validation is a challenging task. Despite the problem, unsupervised learning modelling is a growing field with huge potential [37, 38].

The distinctive feature of cluster analysis is that it intuitively simplifies data by grouping it into clusters that show similar behavior. In order for data to be soundly clustered, understanding characteristics of domain data is essential. This enables domain experts to define the meaning of similar behavior as a cluster. Built upon this understanding, the experts can select optimal clustering algorithms and their parameters. Considering selection of clustering algorithms, distribution of data is an important factor. Different algorithms have different assumptions about data distributions [37, 46].

Since the goal of exploratory data analysis is to discover homogeneous behavior from unknown data distributions, assumptions that clustering algorithms make on data distributions can affect the performance of the algorithms [38]. With this consideration, first, types of clustering algorithms are described to select which algorithms are best for exploratory data analysis. Second, based on each pro and con, an optimal algorithm is selected.

In general, clustering algorithms can be described by three categories: partition clustering, density-based clustering, and hierarchical clustering [1, 2, 46, 49].

1. Partition Clustering

Partition clustering assumes clusters have the shape of a hypersphere. Partition clustering algorithms, such as k-means, assign every data point into specific groups, which is efficient and simple. However, a problem arises if there is the presence of noise and outliers that do not belong to any group. Furthermore, if the distribution of data points is complex, the result is poor since the method cannot discover clear clusters within the data points.

2. Density-based Clustering

Density-based clustering algorithms explore arbitrary shapes of clusters based on the density of data with few assumptions about distribution of clusters. Since the method groups a higher density of points into a specific cluster, the boundary points of the cluster are considered as noise. Although the method can effectively handle noise within data, it is inflexible in terms of dealing with a variable density within data.

3. Hierarchical Clustering

Hierarchical clustering decides the number of clusters to create, unlike partition clustering, which has a predefined number of clusters. Hierarchical clustering results in a dendrogram which is one single upside-down tree representing data points as leaves and branches. A dendrogram intuitively represents the possible number of clusters by cutting the dendrogram based on the height of the dendrogram. However, selecting how numerous clusters to create at a certain height is arbitrary. This feature can lead to less accurate performance than partition clustering.

Considering each feature of the clustering algorithms and the goal of this thesis, exploratory data analysis, a combination of density-based and hierarchical clustering is considered to be the best approach in terms of detecting noise and variable density (detailed in Section 3.3.2).

2.2.3. Virtual Reality

When compared with traditional desktop interfaces, virtual reality (VR) has distinctive features. VR emancipates us from physical and mental constraints of reality. While virtually exploring phenomena, data, or tasks, users can dynamically manipulate and control a scale, location, position, and time in a VR environment. Furthermore, interactively having relationships with avatars and sensory experiences in a VR environment affects users' cognition and perception. The freedom from constraints imposed by reality has huge implications for exploratory data analysis.

Unlike VR technology, desktop computer interfaces have fundamental limitations for visualization of domain 3D spatial data. When a reservoir model is geologically complex and heterogeneous, it is difficult for the desktop interfaces to handle such high dimensional data. This is ascribed to the nature of the data that exceeds human space perception ability to analyze, manipulate, and interact with the data. In addition, traditional desktop-based interfaces are spatiotemporally restrictive for domain engineers and decision makers. The design of the interfaces, such as a keyboard, mouse, and monitor, imposes limitations on collaborating with various disciplines, discovering knowledge in such big 3D volume data, and making spatial judgement effectively in the limited information space. Finally, as described in Section 2.2.1, the nature of a fuzzy and qualitative question revealed the limitation of the desktop-based interfaces in terms of interactive exploration. Overcoming the above limitations of the desktop-based interfaces requires an effective user interface for 3D spatial data with high dimensions and heterogeneity. The following section briefly reviews how different disciplines in petroleum engineering utilized VR technology to the above problems. After this review, some related applications of VR technology and new interfaces are introduced which have inspired this thesis.

In the field of the oil and gas industry, VR technology and a novel interface have been applied in various ways: well control training, drilling operation training, well planning, well placement optimization, enhanced oil recovery, collaboration with various disciplines, visual exploration of high dimensional data, and visual analytics [19, 24, 25, 26, 27, 28, 29, 30, 31, 32, 36]. Among these various applications, the two keywords (visual analytics and exploration of high dimensional data) influenced the design of the visual pipeline of this research. From a visual analytics perspective, designing a visualization pipeline starts with investigating nature of data that the pipeline deals with for successful applications and characterizing the data to select an optimal solution. Since a visualization pipeline begins with data and ends with a user, it is essential for designers to scrutinize what user tasks domain experts will perform with the data. Visualization and interaction with data should be tailored to the needs of the domain. In this respect, task analysis from both domain and interactive data visualization perspectives is a key to design a sound and successful visual pipeline. This approach inspired this research in terms of the design of the visualization pipeline [27, 32, 36].

The different approaches of the VR technology to the different problems have in common human-computer interaction (HCI). Although HCI is the final stage of a visualization pipeline and an indispensable factor of VR technology, it has not been adequately addressed in the literature. When exploring a VR environment, domain experts perform visual analytics: recognize tasks, actively get involved and interact with data, discover insights, and make decisions. To effectively achieve the cognitive, problem-solving sequence from the low level to the high level in a virtual world, this research adopts immersive analytics by supporting sensory, cognitive, and perceptional interactions between VR technology, machine learning, and a human expert [84]. Since immersive analytics is implemented in the final stage of a visualization pipeline (i.e., human-computer interaction), materializing successful immersive analytics in VR technology requires the investigation on HCI from a fundamental and philosophical perspective.

2.3. Human-Computer Interaction

The Human-Computer Interaction (HCI) is to restore, improve, and design a relationship between humans and technology. To reconcile the detachment, it is imperative to know the relationship between individual entities. Built upon the understating, the interaction can be effectively and efficiently designed. This section describes interrelation between human beings and VR technology by embracing the philosophy of technology. This section starts with current problems of HCI. To resolve the problems, from Martin Heidegger's perspective, this section illustrates the construction of the correlation between human beings, VR technology, and interaction. Following Heidegger's philosophy, this section represents philosophical tools as mediators between human beings and VR technology. Finally, this section introduces a game-based approach as a design decision of the relationship between VR technology, human beings, and mediators.

The HCI in a VR environment is a creative activity. While exploring and interacting with virtual objects, humans perform a series of perceptive and cognitive processes that range from establishing goals to interpreting a system based on the goals and intentions of the users [78, 79]. In the fields of science and engineering, the creative and interactive exploration of a VR environment is challenged by inadequate interaction designs and lack of understanding complex nature of data that the domain deals with, such as massive data populations, high dimensional data, and large and complex 3D volume data. A poor interaction design fails to deliver functional user experiences and elicit capacities to experimentally construct, manifest, and expand user perception of the system. The characteristics of the scientific and engineering data outlined above are beyond the realm of human ability to process, making it difficult to explore data intuitively in a VR environment. Furthermore, visualizing the characteristics involves complexity, abstraction, and dynamic behavior. The compacted information represented in a VR environment causes difficulty in analyzing and interpreting the data. To handle these problems, HCI has been expanding its spectrum by incorporating ideas from different fields of study, such as cognitive science, ethnomethodology, and Activity Theory [80].

However, despite immersing the integration of the different fields of science and practice, HCI has omitted quintessence: a human being. The underlying deficiency of HCI is its failure to consider a human being. Since a human being is a subject (user) who perceives and interacts with an object (VR environment), with the human factor missing in the process of HCI design, HCI is inadequate. The omission causes alienation of the subject from the object, which veils the potential capacity of the subject to discover knowledge. To put the subject and object on the same horizon, there is a need to investigate the mutual constitution of a human being and VR environment (technology): how human beings perceive a VR environment and how the VR environment shapes human beings in return.

With respect to the alienation of human beings in VR design, in his work *Being and Time*, the German philosopher Martin Heidegger argued that the cause of the isolation of beings from their surroundings can be ascribed to a dichotomous approach to a relationship between beings and the world. From the philosophy of technology perspective, Heidegger's philosophy can be applied to VR technology: a virtual world created by VR technology turns into an environment that has a harmonious relation with a human. The embedding of such philosophy in a virtual world not only assists users in a problem-solving process, but also helps the user sensory experience and perception that lead users to immersion in a virtual world. With this understanding of the mutual constitution, a VR environment enables users to reveal possibilities to open a new intellectual horizon by offering users autonomy to interactively customize, modify, and redefine the VR environment.

In order for human beings to explore and interact with a virtual world effectively, it is imperative to design an interactive VR environment and mediators. To design such a VR environment and mediators, the understanding of a relationship between a human being, mediator, and the world must to be investigated in a fundamental manner. Section 2.3.1 first introduces Martin Heidegger's philosophy as a philosophical framework of this research and demonstrates that the relationship of a human being and the world is aligned with a virtual world. Then, the mediators are described in terms of how postphenomenology and ontology play their roles as structural and functional mediators, respectively, in the relationship. Built upon these demonstrations, the directionality of VR application design serves as a foundational philosophical framework of this thesis.

2.3.1. Being in the World

Before introducing Heidegger's philosophy, there is a need to clarify what interactive tools are. There are two different types of interactive tools, and the interactive tools are categorized into a structural mediator (postphenomenology) and a functional mediator (ontology). One as a structural mediator is a metaphysical tool mediating between a human and virtual world, such as a game world and game story line. The game world and the story line provide users with the context of a virtual world, making them feel comfortable with the new world. The other as a functional mediator between a human and data is an instrument for users to customize, modify, and interactively redefine data. These interactive tools, from now on, refer to philosophical tools since the interactive tools enable users to extend their experience, perceptions, and cognitive capabilities while exploring a virtual world. In this regard, the interactive tools disclose new possibilities for richness of existence from a philosophical perspective. These philosophical tools including Heidegger's philosophy (being-in-the world) shape the whole design of a virtual world in this research.

It is important to note that Martin Heidegger's philosophy serves as a cornerstone of a philosophical framework of this thesis. In Heidegger's book *Being and Time*, he proposed a concept of being-in-the-world that describes a relationship between a human being (subjectivity) and world (objectivity) [34]. According to Heidegger's existential phenomenology, the relationship can be interpreted in this way: with a being tenaciously exploring and perceiving the disguised world with numerous possibilities, a new dimension within the world is revealed to the being. The ceaseless, cognitive activity enables the being to be immersed in the new dimension. With the being immersed, such an activity allows the being to recognize and access the new dimension that partially reveals itself from the world [35]. This philosophical framework can be aligned with construction of a virtual world. A virtual world is a world where beings can explore, access, and interact with the environment using continuously cognitive activities as an interactive tool to discover knowledge and gain insights. Thus, a virtual world needs not be merely a problem-solving world where human beings are alienated from a virtual world. It has to be a coexisting world where human beings and a virtual world exist in a harmonious manner. This philosophical consideration is reflected in the design of a virtual world in this thesis. The following shows how postphenomenology and ontology frame a virtual world and how the materialized virtual world with mediators influences a human cognitive process.

2.3.2. Philosophical Tools as Mediators

In this section, extending the concept of being-in-the-world from Heidegger's existential phenomenology, postphenomenology and ontology shape a virtual world in a constructive and harmonious manner. These theories as mediators in this thesis provide human beings and a virtual world with copious experience and recognition. To clarify the roles of postphenomenology and ontology as mediators in a virtual world, this thesis redefines postphenomenology as a structural mediator and ontology as a functional mediator. From now on, the terms "the world" and "a virtual world" will indicate the same thing since the 'Being in the World' section demonstrated the philosophical framework is aligned with construction of a virtual world.

1. Postphenomenology: Structural Mediator

Unlike traditional phenomenology, postphenomenology describes the structure of being-in-the-world as united, inseparable, and mutual. When human beings interact with the world, the relationship is intrinsically mutual [35, 73, 74]. The traditional, 'alienated' (non-mutual) approach to the relationship leads to thrownness from the world. In the world, human beings only focus on an imminent task without connection with or immersion in the world. Traditional phenomenology offers the world that disguises possibilities for human beings to experience, failing to elicit potential of human ability within the world. This approach has been applied to VR applications in the oil and gas industry. However, if VR applications embrace postphenomenology as a structural mediator, this mediator can guide human beings to a virtual world in a more natural way to interact with a virtual world.

2. Ontology: Functional Mediator

In the mutual structure of being-in-the-world, a being tries to explore and understand the world by using a cognitive tool that a being has possessed and developed. Utilizing the tool, a being organizes its general understanding of the world. Ontology is the cognitive tool that serves as a way for a being to understand and interact with the world [35]. The first step for the interaction is to use the tool and discover things that exist in an unperceived state as they are in the world. The ontological tool as a functional mediator enables things in the world to become reachable, accessible, and understandable for an individual being to have a relationship with part of the revealed world. Once human beings disclose the possibilities of things by using the ontological tool, they construct their perceived, subjective world within the objective world [35]. From this perspective, the ontological tool will be designed for human beings in order to help explore, understand, and discover the disguised virtual world.

In this thesis, built upon the structural mediator that helps users naturally mingle with a virtual world that consists of a deluge of scientific information, the role of the ontological tool as the functional mediator is to provide an environment in which users can develop their capacity to perceive, relate, and interact with a virtual world. In this environment, utilizing the functional mediator, users can open a new dimension in which they can understand, organize, and categorize both a virtual world and themselves into user-dependent worldviews (their own private perspectives). This careful design of the VR environment will lead to impacts on user cognitive, operational, and decision-making processes when human beings explore a virtual world.

2.3.3. Design of Virtual World

Before embarking on how postphenomenology and ontology as mediators can act in a virtual world, there is a need to identify what a simulation is. A *simulation* is a world where human beings can emancipate and surpass their physical, cognitive, and interactive constraints imposed on them by the actual world without actual risks. Simulation supports people in designing and rationalizing their hypothesis in relation to the actual world. This support allows them to gain virtual, alternative, and practical scenarios that are physically impossible or dangerous to implement in the actual world. In science and engineering, a digital simulation represents processes or systems and provides a tool that enables humans to validate their theory or research questions or find new knowledge. Furthermore, it provides interactive relationships between a digital simulation and humans that enable them to transcend, manipulate, and adjust physics, such as time, scale, and location. In line with these features, a virtual world shares the characteristics of a digital simulation equipped with novel, immersive technologies [35, 84].

As explained in the previous sections, a virtual world is the culmination of an integration between the concept of being-in-the-world, philosophical tools as mediators (structural and functional mediators), and characteristics of a digital simulation. The big questions for VR designers are: how can postphenomenology and ontology as mediators act in a virtual world and how can VR designers get users to clearly understand the concept of a scientific simulation and get involved and immersed in a virtual world? The answer to both these questions is to let users play with a scientific simulation in a virtual world like a video game. Most VR applications in science and engineering encourage users or experts to focus on problem-solving tasks directly without giving them any information about the VR application, goals, tasks, and data. This rigid VR design not only makes users feel alienated from a virtual world but also hides numerous potentialities of a virtual world and users themselves. To let users play in a virtual world and make them feel comfortable with the new interface, this research designs a game-based approach: a world with a theme and goal-oriented interaction. Figure 2.3 describes each characteristic of the game-based approach.

1. World with a Theme

In video game design, a game world has its own background and story line, for example, a spaceship mission to explore a new planet. A world with a theme delivers concepts and goals of the game world, explains rules and tasks, and provides narratives that enable a user to get immersed in and autonomously explore the game world. These guidelines help inform the users of the worldview and encourage them to have self-determination, equipping them with strategies on how to explore the world and achieve the goals and tasks. In this regard, a world with a theme provides concrete and legitimate purposes for users to be able to get immersed in a virtual world and enables autonomous behaviors for them to explore a virtual world.

The immersive interaction with the world is the goal of a world with a theme. A world with a theme shares the same purpose of postphenomenology: to connect users with the world in a harmonious manner to disclose the potential of the world itself and human capacity within the world. A world with a theme as the reification of postphenomenology is thus aligned with postphenomenology as the structural mediator. The characteristic of a world with a theme will be materialized in a virtual world in this research.

2. Goal-oriented Interaction

Goal-oriented interaction refers to an intended interaction that users perform with game objects by utilizing tools to achieve their goals and tasks. For example, in a video game, users employ weapons to kill monsters to get money or fulfill a given mission. Before users act on their mission, a world with a theme (postphenomenology) informs them of the worldview, goals, and tasks. With them being immersed in the world and having a harmonious relation to the world, they have become equipped with strategies on how to play and be ready to implement their actions and mission. What initiates user actions is an interactive tool. The interactive tool is a key concept that enables goal-oriented interaction. In the sense that ontology as the functional mediator enables users to reach, access, disclose, and interact with the objects in the world by utilizing an interactive tool, goal-oriented interaction materializes ontology in the form of an interactive tool.

In order for the goal-oriented interaction to operate efficiently and open up a virtual world to users, a virtual world and its interactive tools need to adequately simulate aspects of reality. In terms of phenomenology, reality is a stream of persistent perception of beings within being-in-the-world. Considering this characteristic of reality, virtual experiences happen from a steadily perceivable and logically understandable world. With users experiencing, exploring, and manipulating the virtual world by utilizing its interactive tools that simulate aspects of reality, the interactions and self-organizing subworlds open countless possibilities that lead users to new dimensions by having a causal relationship with the present state of that specific virtual world [35]. To design such a virtual world effectively, this thesis takes the game-based approach by materializing a world with a theme and goal-oriented interaction in a virtual world.

The game-based approach practically implements the abstract concept of Heidegger's philosophy in a virtual world by providing a worldview of a virtual world for a smooth transition from reality to a virtual world, clear direction in exploration of a virtual world, and physiological immersion in a virtual world. These foundations will elicit virtually real experiences, derive autonomous behaviors, and foster sensemaking processes.

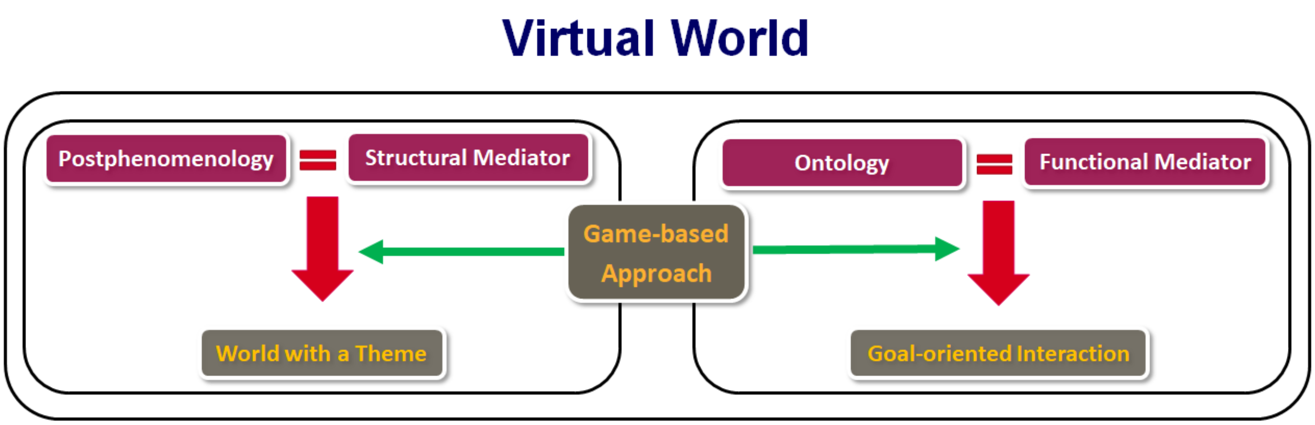

Figure 2.3: Elements of the game-based approach.

Chapter 3

3. Virtual Exploration through a Philosophical Tool

Chapter 2 described each component of the visualization pipeline from data to human-computer interaction. Chapter 3 shows how the components are integrated and reified in a philosophized VR environment for exploratory interaction of 3D spatial data. Firstly, this chapter characterizes data, a user experience goal, interaction types, and user tasks from reservoir engineering and interaction design perspectives. The VR environment embraces and represents the results of the characterizations implicitly and explicitly. Secondly, it describes the philosophical framework of the research and its mediators as philosophical tools for the relationship between data, a virtual world, and a user in the form of the game-based approach. Finally, it materializes the integrated components of the pipeline in the virtual world.

3.1. Overview

The reified VR environment is a VR information space which has a relationship between object space (3D spatial data), attribute space (its properties), data structure space (clustering algorithm), and user experience space [65]. This environment is designed for a user to perform virtual exploration of big 3D spatial data. Since the VR environment involves these various factors to consider, most VR application designs in the fields of science and engineering have omitted and neglected human beings in the designs, directly engaging the users in a problem-solving process without the users being informed about the interface, goal, data, and user tasks of the VR applications. Furthermore, the omission of human beings in VR applications fails to immerse the user in the VR environment, alienating the user from the environment and deteriorating task performance. In this regard, VR designers have to remember that it is a human being that analyzes, interprets, and interacts with the environment. With regard to these shortcomings, this thesis embraces the philosophy of technology as the foundation of the VR design to overcome them.

The philosophical framework shapes a harmonious relationship between VR technology, mediators, and human beings in order for human beings to open up and elicit their potential by emancipating themselves from constraints on disguised existence imposed by inadequate design of technology. The mediators materialized in the form of the game-based approach provide human beings and their world with copious experience and recognition. Postphenomenology as the structural mediator constructs a virtual world, offering users with the context of the virtual world and guiding them to be immersed in the virtual world. Materializing the results of task analysis, ontology as the functional mediator is a cognitive and interactive tool that enables users to perform perceptual and intellectual activity in the virtual world by reaching, accessing, and interacting with the virtual world (e.g., complex 3D objects). Constructing their perceived, subject world within the virtual world, a user discloses the possibility to manifest, extend, and expand their understanding of the virtual world by performing exploratory data analysis.

3.2. Methodology

A big question is raised before designing an interactive exploration pipeline as the methodology of this thesis: can a VR application materializing complex 3D spatial data support a user to autonomously think in a virtual world? This big question is divided into two perspectives (i.e., data and VR). From a data perspective, since the nature of oil and gas data is multivariate and multiscale 3D spatial data manifesting high dimensionality and heterogeneity, desktop-based interfaces are restrictive to visualize and interact with the 3D spatial data due to the limited information space (i.e., cognitive, interactive, and physical constraints). From a VR perspective, a VR environment is 1) a word in which every single aspect of a virtual word has to be designed and constructed and 2) complex information space with millions of cells containing reservoir properties. The amount of the 3D spatial information with a new interface for reservoir engineers is beyond human capacity to understand, analyze, and interpret. These predicaments necessitate a mediator for human-data interaction as a central aspect of the pipeline design. In this regard, Martin Heidegger's philosophy (being-in-the-world) as a philosophical framework mediates between a computer (data), virtual world (VR technology), and human being (engineers) in a harmonious manner.

With the foundation of the philosophical framework, as Figure 1.2 represents the design overview of the interactive exploration pipeline, this research starts with characterizing the initial stage of the pipeline (the scope of 'Computer' in Figure 1.2) to design a successful and robust pipeline. The components of the initial stage are data, visual mapping (data visualization), cluster analysis, and task analysis (Section 3.3). The dataset utilized in this thesis is The Ninth SPE Comparative Solution Project [40, 41]. Employing this dataset, the in-house Black Oil Simulator (BOS) [42] builds static and dynamic reservoir models and outputs the visualization files constructed by the Visualization Toolkit (VTK) format representing grid information and physical variables at each time step. For better understanding the VTK dataset of BOS and designing optimal visualization systems, the architecture of the VTK dataset is described, and the main components of the architecture are characterized. These features necessitate HDBSCAN for cluster analysis in terms of dataset size, run time, scalability, and noise identification [46, 54, 57]. To enable the reservoir models in the VR environment to elicit user immersion in their research questions or hypotheses and elicit autonomous user behavior to perform exploratory data analysis, task analysis is investigated and designed from both the domain (reservoir engineering) and interactive data visualization perspectives.

The middle stage (the scope of 'VR') is designed for interactive 3D visualization. Activiz as the VTK C# API is utilized to render the 3D reservoir models in Unity Framework (the C# .NET environment). For interactive data exploration in the big volume of spatial data, cluster analysis in Python is implemented in the static model by involving a user in the process of pattern discovery. The static model with cluster analysis and a dynamic model are visualized in a VR environment in order for human experts to interact, manipulate, and analyze the models. In this VR environment, a user can perform exploratory data analysis to explore the implications of the static model on the dynamic model. To mediate between the complex 3D objects (the static and dynamic models) and a user in the VR environment, the game-based approach of this thesis embodies postphenomenology and ontology by materializing a world with a theme and goal-oriented interaction.

The final stage (the scope of 'Human') is designed for knowledge discovery in a VR environment. Postphenomenology as the structural mediator shapes the virtual world, providing users with the context of the virtual world. This enables users to be immersed in the virtual world, eliciting autonomous behavior to develop strategies of exploration with their hypotheses and understand goals and user tasks. Based on visual analytics and cognitive task analysis [23, 39, 64], postphenomenology materialized in a VR avatar offers three ways of sensemaking for users to perform exploratory data analysis in the VR environment. In this manner, the VR environment becomes a world that enables a user to interact with 3D spatial data by utilizing an interactive tool. Materializing the results of the task analysis in the interactive tool, ontology as the functional mediator is a cognitive tool that enables users to reach, access, and interact with the virtual world (e.g., complex 3D objects). Constructing their perceived, subject world within the virtual world, users disclose the possibility to manifest, extend, and expand their understanding of the virtual world. These processes lead to knowledge discovery, expanding an intellectual and experiential horizon of users through this VR application.

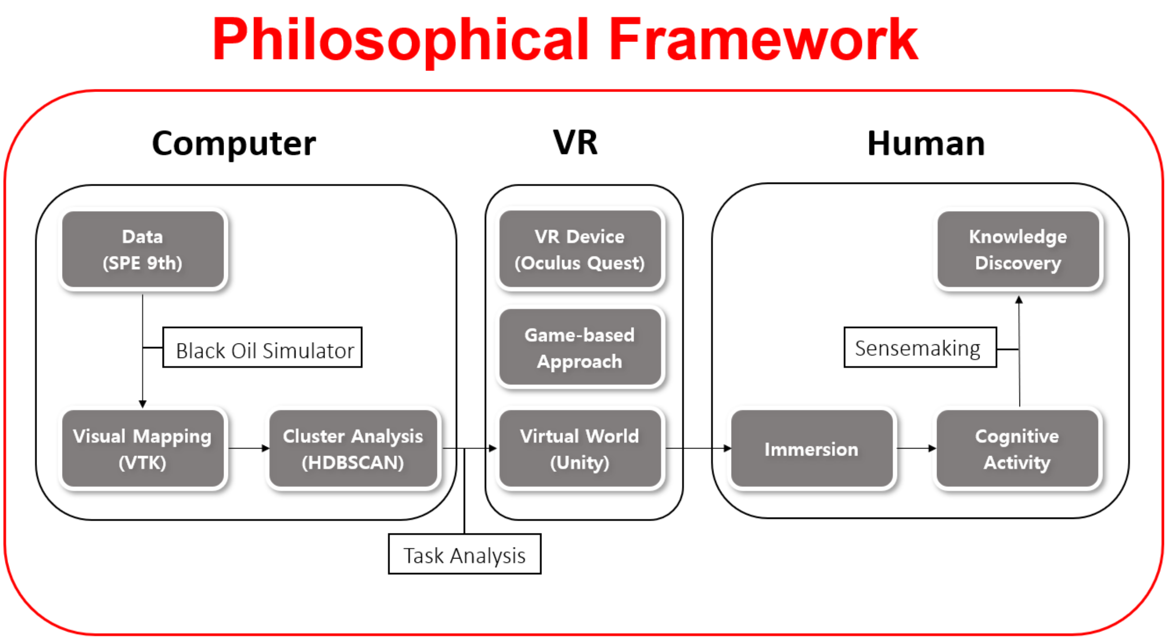

Figure 1.2: Design overview of the interactive exploration pipeline with the associated technologies.

3.3. Problem Characterization

3.3.1. Data Description

3.3.1.1. Data description and Structure

The dataset used for this thesis is The Ninth SPE Comparative Solution Project and open-source data. The purpose of the project is to re-examine black-oil simulation based on a black-oil reservoir model with 9,000 cells and a highly anisotropic and heterogeneous permeability generated by a geostatistically-based permeability field. The model consists of a 24 x 25 x 15 grid and a 10-degree dipping-angle in the *X*-direction. The model dimensions are 300 x 300 x 359 (ft). The model has a rectangular-shaped geometry with a simple geological structure (e.g., no faults). The model uses a Cartesian coordinate system, and the first cell in the coordinates (0,0,0) starts at a depth of 9,000 feet subsea [40, 41].

All relevant data of the project as simulation inputs are provided, such as permeability, porosity, relative permeability, PVT properties, initial conditions, end time of simulation, and well configurations [40, 41]. Utilizing the dataset, the Black Oil Simulator (BOS) developed by the Reservoir Simulation Group builds a reservoir model. BOS is a parallel simulator and employs the in-house parallel platform consisting of distributed-memory computers. This feature solves large-scale problems effectively in terms of scalability, computing time, and computational cost [42]. The output files of BOS written in a text-based format are runtime, well production summary, field and numerical summaries, and visualization. This thesis utilizes the visualization output for 3D spatial data visualization and interaction. The 3D visualization output file constructed by the Visualization Toolkit (VTK) format represents the grid information and physical variables at each selected time step. The number of VTK output files is 127 files, representing 900 days of the simulation.

A dataset of VTK consists of its structure and attributes. Figure 3.1 describes the architecture of a dataset. Cells and points comprise the structure. Cells indicate topology, which refers to geometric transformations such as rotation. Points indicate geometry, specifically point coordinates. Since a dataset generated by digital computers uses points to represent the dataset, points represent an element of the dataset and cells represent interpolation between the points. The typical data attributes associated with the structure are scalars, vectors, texture coordinates, and tensors. The data attributes are assigned to each cell [43]. The structure and data attributes are integrated into one file. The total VTK files of BOS are 127 files, and each file represents time steps for 900 days of simulation. Figure 3.2 describes the structure of a VTK data file format. The open source software Paraview reads the VTK files to display 3D visualization data and time step animation.

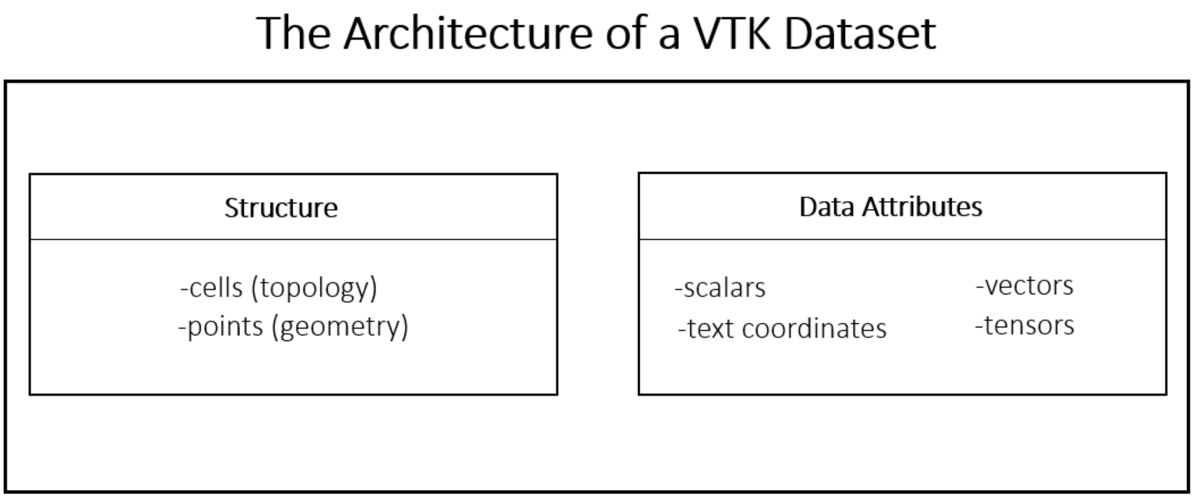

Figure 3.1: The Architecture of a VTK dataset.

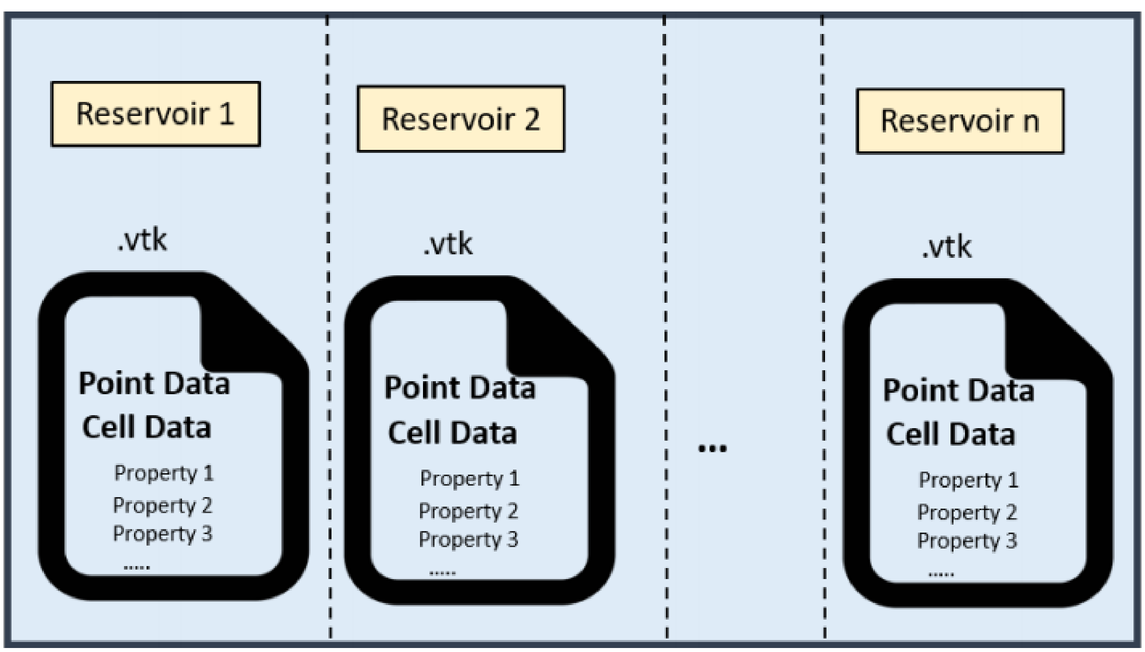

Figure 3.2: The structure of a VTK data file format [27].

3.3.1.2. Data Characterization

For better understanding a VTK dataset of BOS and designing optimal visualization systems, characterizing a dataset is essential. An important characteristic of the VTK dataset is the type of the structure. The structure is categorized into structured and unstructured data. This thesis is using unstructured data since it is more flexible when representing irregular patterns of a model. A second characteristic of the VTK dataset is the type of the data attributes. The attributes fall into static and dynamic properties. Static properties are geological relevant parameters from geostatistics or well logs, such as porosity, permeability, and facies. Dynamic properties are time-varying parameters from reservoir flow simulation or production wells, such as pressure, production, and temperature [45].

Characterized by a type of structure and data attributes, the VTK file is structured by unstructured data. The unstructured data describes the geometry of a model and data attributes associated with the model, (i.e., the static and dynamic properties). A vtkPointData class represents coordinates of the model. Each point index of the coordinates represents cells by interpolating the point indices. Each cell consists of eight-point indices. After the model is constructed, a vtkCellData class describes static and dynamic properties associated with each cell. The static reservoir model includes the permeability in the *x*, *y*, and *z* directions and the constant porosity of each layer (total 15 layers). The dynamic reservoir model includes 16 properties: oil pressure, oil saturation, water saturation, gas saturation, bubble point pressure, oil phase relative permeability, water phase relative permeability, gas phase relative permeability, oil phase viscosity, gas phase viscosity, gas capillary pressure, water capillary pressure, oil phase density, gas phase density, water phase density, and oil-solution gas density.

By visualizing static and dynamic reservoir models, this thesis focuses on exploratory data analysis of highly heterogeneous permeability and its implications for dynamic properties. There are two reasons for the focus. The first reason is that the purpose of The Ninth SPE Comparative Solution Project is to scrutinize an effect of a highly anisotropic and heterogeneous permeability on performance of black-oil simulation. Although porosity is one of the important static properties for reservoir simulation, the porosity values of the project are constant for each layer. Compared to the values of the porosity, the distribution of permeability is highly heterogeneous. To support exploratory data analysis from multiple perspectives, this research includes the distribution of porosity values of each layer in a reservoir model alongside the distribution of permeability.

The second reason is that permeability as an essential fluid property for reservoir simulation is a criterion to determine how much fluid a rock can transmit through networks of pores in the rock. In this respect, identifying and analyzing the spatial distribution pattern of permeability is imperative for achieving the following four goals: optimization of recovery processes, potential areas of production operations, design of reservoir development planning, and pragmatic reservoir performance predictions [44]. To discover the spatial distribution pattern efficiently, cluster analysis is implemented in the static model. Built upon the 3D pattern recognition analysis, the virtual reality information space supports domain experts to analyze the pattern, find correlations between the pattern and dynamic properties, and make decisions based on their knowledge discovery in terms of the above four goals.

3.3.2. HDBSCAN

A typical reservoir model consists of millions of cells with many reservoir properties. The large scale necessitates simplifying the complexity of the model using an efficient computational tool. In this thesis, cluster analysis is implemented to find patterns of heterogeneous permeability data in a 3D reservoir model from The Ninth SPE Comparative Solution Project. The goal is to discover clusters from a skewed population of the permeability data (variable density clusters) in order to visualize potential behavior of oil flows inside the model and analyze the behavior and its implications for estimation of oil production. Furthermore, cluster analysis enables correlating the behavior with dynamic properties to understand how the behavior affects them with a time-series simulator.

There are three criteria used to choose a clustering algorithm in this thesis: the purpose of cluster analysis (exploratory data analysis), the permeability data distribution (skewed data), and the large spatial data size, which consists of 9000 cells and its attributes related to each cell. The dataset for cluster analysis consists of six features, (i.e., *x*, *y*, and *z* coordinates and permeability values in the *x*, *y*, and *z* directions). When shedding light on the criteria, Hierarchical Density Based Spatial Clustering of Application with Noise (HDBSCAN) represents better performance compared to other clustering algorithms in terms of run time, a dataset size, the number of parameter selections, scalability, noise identification, and assumptions about data distribution [46, 54, 57]. Although Density Based Spatial Clustering of Application with Noise (DBSCAN) shows comparable performance to HDBSCAN, HDBSCAN is a better solution for exploratory data analysis in terms of scalability, a small number of intuitive parameters, and detection of variable density clusters [46, 53].

Since the HDBSCAN algorithm is an extension of the DBSCAN algorithm, there is a need to describe DBSCAN first. For clarity and consistency with a statistical interpretation of clustering, DBSCAN is modified from the standard DBSCAN algorithm, which involves the concept of border points [58]. There are two parameters of DBSCAN. Epsilon, ε, is a distance scale, and a minimum number of points, *k*, is a density threshold. ε defines a distance between points to be considered part of a cluster. *k* represents a minimum number of points that a cluster should contain in a neighborhood. To deal with the problem of variable density clustering, DBSCAN is extended to HBSCAN by building a hierarchy of the DBSCAN algorithm. This allows HDBSCAN to find stable clusters over epsilon and avoid selecting the epsilon parameter. Below are steps showing how HDBSCAN works [46, 54, 59]:

1. Define a new distance metric between points.

Let $X = \{x_1, \ldots, x_n\}$ be a dataset of points, and let D be an $n \times n$ matrix containing the pairwise distances $d(x_p, x_q)$, $x_p, x_q \in X$, for a metric distance $d(\cdot\ ,\ \cdot)$. To find higher density points from lower density points (noise), three concepts need to be defined: a core point, core distance, and mutual reachability distance. A core point and core distance are defined with respect to ε and *k*. A point x_i is considered a core point if a circle with radius ε contains at least *k* many points from X. A point is considered noise if it is not a core point. A core distance ($dcore(x_{\mathrm{i}})$) is a distance from the core point to its k-nearest neighbor. To handle variable density clustering, DBSCAN is extended to HDBSCAN by defining a new distance metric between two core points (x_i and x_j) in X. With a given value *k* fixed, a new distance metric, mutual reachability distance (MRD), is defined as $dmreach(x_i, x_j) = \max\ \{dcore(x_i),\ dcore(x_j),\ dcore(x_i, x_j)\}$. Since value *k* is the single input parameter of HDBSCAN, the result of the estimate of density depends on the selection of *k*. A larger *k* value means a cluster includes more points.

2. Build a minimum spanning tree.

To classify dense data from noise, the graph theory and MRD are utilized. The graph theory considers data points as vertices and an edge between two data points. With vertices and edges, it builds a weighted graph by dropping edges according to varying threshold values. This approach results in a graph of connected vertices. However, the approach is not efficient since, if there are N data points, the approach takes $O(N^2)$ run-time. For an efficient computation, Prims' algorithm [77] finds a minimum spanning tree for a weighted graph by the MRD. Using the MRD, Prims' algorithm builds a tree by connecting vertices that consist of the lowest weight edges.

3. Build a single linkage cluster tree.

Built upon the MST (minimum spanning tree), we can construct a single linkage cluster tree by converting the MST into the hierarchy of connected components. There are two steps involved. Firstly, the edges of the MST are sorted by the MRD between a pair of points an edge span (i.e., weight). Secondly, using a union-finding data structure [60], we can build a single linkage cluster tree. The union-finding data structure creates a merged cluster for each edge by increasing the MRD (weight), which is the bottom-up fashion. To be specific, at each step, the structure merges two clusters into a cluster by combining their closest pair of points. The process results in a cluster hierarchy, which is a single linkage cluster.

4. Condense a cluster tree.

To find variable density clusters, first, a single linkage tree needs to shift from a distance-based tree to density-based tree. λ is introduced by inverting a core distance (ε), which is the distance from a core point to its k^{th} nearest neighbor. Using the $\lambda = \frac{1}{\varepsilon}$ value, the single linkage tree is constructed in terms of varying density. Second, the single linkage tree cuts into a condensed tree in terms of simplification of the hierarchy. Since the single linkage tree is a large cluster hierarchy, the tree needs to be split into small parts of a cluster from the root downward. The minimum cluster size (MCS) as the single input parameter of HDBSCAN handles this process. The parameter decides whether data points should split from a cluster or not at the given λ value, which enables us to pinpoint a location of split points and a split distance value. When cutting the cluster hierarchy, at each split, the parameter checks if each new cluster split from the parent cluster has fewer data points than the MCS. There are three cases:

-Case 1: if any child cluster is the case, it is considered as spurious data points separating from the parent cluster at the given λ value.

-Case 2: if only one child cluster has a larger size of data points than the MCS, it persists from the parent.

-Case 3: if more than a child cluster has a larger size of data points than the MCS, it is considered as a true cluster split.

At the end, with clusters shrunk by increasing λ from the root downward, a condensed tree is created.

5. Extract clusters.

The goal of HDBSCAN is to find dense clusters out of noise. The last step here is to select dense clusters using λ values. To extract a flat set of clusters, there are three concepts to be defined: stability of a cluster, $\lambda_{max,\,\mathrm{C}_i}(X_j)$, and $\lambda_{min,\,\mathrm{C}_i}(X_j)$.

1) Stability of a Cluster

For points $(X_1, X_2, \ldots, X_j, \ldots, X_n)$ in a cluster $(C_1, C_2, \ldots, C_i, \ldots, C_n)$, stability of a cluster is defined by the sum of the range of λ values:

$$S(C_i) = \sum_{\mathrm{X}_j \in \mathrm{C}_i} (\lambda_{max,\,C_i}(X_j) - \lambda_{min,\,C_i}(X_j))$$

2) $\lambda_{max,\,C_i}(X_j)$

$\lambda_{max,\,C_i}(X_j)$ is the λ value at which the point X_j separates from the cluster C_i. This results in either a spurious point or true cluster split from the parent cluster.

3) $\lambda_{min,\, C_i}\left(X_j \right)$

$\lambda_{min,\, C_i}\left(X_j \right)$ is the λ value at which the point X_j persists in the cluster C_i.

The stability of a cluster can be described as an optimization problem with a constraint:

$$maximize \sum_{i \in I} S(\, C_i),\ I \subseteq \{1,2, \dots ,n\}$$

subjected to $C_i \cap C_j = \emptyset$, for all $i, j \in I$ with $i \neq j$

In this fashion, flat clusters are selected in the condensed tree.

3.3.3. Task Analysis

Given the enormous and complex amount of the 3D spatial information for users to deal with, as shown in Section 3.3.1 and Section 3.3.2, users can be overwhelmed by the information and feel alienated from the VR information space, making it difficult to find a starting point in the VR environment. To mitigate this problem, the VR information space is conceptualized in terms of user experience: who will use, what data users deal with, and how they interact with the data [81]. In accordance with the interaction design, the researcher investigates interaction types and user tasks from both domain (reservoir engineering) and interactive data visualization perspectives.

To guide users to tasks clearly and initiate their autonomous behavior to explore the VR environment efficiently, this research categorizes visual tasks that reservoir engineers perform into seven user tasks: Overview, Detail-on-demand, Filter, Cluster, Relate, History and Extract, and Time. These seven categories are typical user tasks when reservoir engineers explore and analyze 3D reservoir models [27, 36, 39, 50, 51, 62]. The seven user tasks materialized in the form of virtual interfaces, such as buttons, dropdown menus, and sliders, enable users to navigate, manipulate, and adjust the 3D models. While focusing on the tasks, these virtual interfaces enable users to start immersing themselves in the VR environment by experimentally utilizing the interfaces. These processes encourage users to expand their understanding of the models and explore the information space more deeply. Furthermore, performing exploratory data analysis, they can find answers to their research questions, hypotheses, or theories. The following are typical visual tasks of reservoir engineers.

1. Overview

Navigating an overview of the entire collection of data offers a user a general understanding of the information space and available tools to perform tasks. The navigation provides the user with information about how the information space is structured and contextually relevant tools that lead to a more in-depth articulation of the navigation. While moving around the visualization of 3D spatial and non-spatial data in the information space and experimentally utilizing tools, the user can gain basic intuitions and pattern distribution of non-spatial data related to each cell in a 3D model. Furthermore, wandering around the space, the user can gain a spatial sense of a 3D model, such as its structure, scale, and location by adjusting the position and rotation of a 3D model. An overview of the information space enables motivating a user to reach, access, and explore a 3D model.

2. Detail-on-demand

While exploring information space, to gain a more profound understanding of data, a user typically focuses on some part of the data and gets the details by selecting options from menus or simply touching them. For example, when hovering over a cell, a user can see pop-up windows showing reservoir properties and its value. Furthermore, a user can select a region of interest by choosing group or individual items based on the historical knowledge to see the pattern or relationship between the items. A UI menu can enable a user to browse and access specific information easily if properly designed. In the menu, by simply clicking a submenu, a user can effectively access details about features or values of interests on the screen. Detail-on-demand supports the discovery of implicit and explicit knowledge in datasets.

3. Cluster

Clusters intuitively show groups that share similar behavior in dataset. Since 3D spatial data with many attributes is beyond the grasp of humans to understand, interpret, and analyze, it is important to simplify data by grouping it into clusters. Clustering permeability differentiates a reservoir model into portions that show similar behavior. The patterns of permeability are an important indicator of initial estimations of oil production volume since permeability determines performance of flows in porous media. 3D cluster analysis offers a user insights of correlations between cells and their property. Built upon the clusters, experts can analyze the patterns and their implications in order to estimate oil production.

4. Filter

By filtering out uninteresting data, users can specify their interests. A filtering technique allows them to reduce the size of data and focus on specific data by removing noise, outlier, or unwanted data. A typical reservoir model consists of millions of cells related to the corresponding reservoir properties. The complexity makes it difficult for domain experts to focus on their interests. In a VR environment, by adjusting a range of a slider, users can stretch and shrink a 3D spatial model and immediately see the changes. The direct manipulation on the screen with the immediate feedback affects users' task performance. Furthermore, controlling the 3D model enables users to feel that they are in control over the system. These user experiences encourage them to quickly concentrate and work on their tasks. Specifying a specific range of data facilitates task performances by eliminating uninteresting data.

5. Relate

In scientific and engineering research, investigating cause and effect is an important step to reach a robust conclusion. In terms of reservoir simulation, a relationship between cause and effect is aligned with a relationship between static and dynamic properties. Since static properties affect performance of dynamic properties, visualizing the two properties in one scene helps users connect the correlation. The exploration in one scene enables users to involve the processes of information gathering and sensemaking, such as problem structuring, evidentiary reasoning, and decision making [64].

In a VR environment, a dropdown menu for static properties and their clusters offers an effective navigation tool for investigating the nature of dynamic properties. While exploring and selecting static properties and their patterns from the static dropdown menu, such as permeability and porosity, users can select a specific dynamic property from the dynamic dropdown menu, such as oil and gas productions, to logically connect the relationship between the static and dynamic property. By actively exploring and relating the cause and effect in one scene, users can interactively collect information and based on the foraged information, perform sensemaking.

6. History and Extract

Since information exploration involves various cognitive steps, it is important to store a history of the information processing steps. Supporting an annotation and recording functions are an example. While replaying and browsing history, users can combine and redefine the information that the users acquired or find new knowledge. This process enables users to retrace their logic, actions, and decisions. Once users go through the process, extracting, saving, and sending the process or data in a format to other users is important. This function would be useful for a collaborative project that involves many experts. The information archive system facilitates remote work and a reasonable decision among experts.

7. Time

A time exploration tool helps users intuitively understand a model by offering rich analytical support. The visualization tool shows trends, expected and unexpected patterns, and relationships in data by developing a model over time. By selecting and developing a different time-series model, users interactively investigate the details and gain instant visual feedback. In a VR environment, users can select a dynamic property among 16 different dynamic properties and its time step amount 127 time-series models from the dropdown menus. This function supports options for selecting a property in a specific time step. Along with the selection of a specific time step, by pressing a play button, a user can see the changes of the property over time. Users can utilize the tool to select a different time step of a property for investigating the details and visualize time steps of a property for pattern recognition.

3.4. Design Rationale

3.4.1. Why Game-based Approach?

As described in Section 3.3, VR users deal with the complexity of scientific visualization tasks, which require a large amount of background knowledge of the domain and comprehensive information processing skill to analyze big 3D spatial data with numerous reservoir properties. From the perspective of eliciting maximal VR user task performance, most important factors are whether VR users are familiar with a VR interface and well informed about the general context of a VR application, such as user tasks and viable analysis functions. However, in the oil and gas industry, most engineers are new to AR/VR interfaces. In this regard, considering the complex nature of a VR environment, engineers can be overwhelmed by the amount of the information to handle in the VR environment. The complexity necessitates a mediator to narrow the gap between engineers and a new VR interface with its deluge of 3D spatial data information for improving task performance.

Task performance of reservoir engineers in a VR interface can be determined by three factors: familiarity with a VR interface, the user tasks, and operational skill of VR visual tools. In most VR applications in the field of science and engineering, due to the complex nature of data, VR designers must inform users of the three factors verbally before users enter a VR environment. A great amount of time is spent explaining a VR interface (how to use the controllers of the VR devices and what will happen when users put the devices on), tasks (the goal of the VR application and what tasks they will perform), and VR visual tools (how to interact with the visual tool in the VR environment). This approach is not only inefficient and time consuming but also fails to take advantage of the potential of a VR interface. There is an inherent limitation insofar as VR designers are unable to explain everything about a VR environment. VR designers should fulfill the potential of a VR interface by designing a self-explanatory VR environment for users to autonomously explore a VR environment.

To design such a self-explanatory VR environment, this thesis embraces and reifies Martin Heidegger's philosophy in a virtual world. Figure 3.3 describes how the sequence of interactions between a user, mediator, virtual world, and the world works in a self-explanatory VR environment. Firstly, the philosophical framework constructs a virtual world where mediators (i.e., postphenomenology and ontology) enable human beings to reach, access, and interact with a virtual world in a harmonious way. In this construction of a virtual world, what initiates interaction between a user and a virtual world is being-in-the-world. Being-in-the-world is a perspective of an individual on the world, which can be background knowledge of domain or skills that users have developed. Postphenomenology and ontology as mediators help being-in-the-world (i.e., users) open up a new dimension in which they can understand, organize, and construct a virtual world and themselves with their own private perspectives. In return, the new dimension within the virtual world shows the human beings possibilities to expand their being-in-the-world within the world through the self-explanatory VR environment. This series of the interactions enables the user to expand their understanding of the virtual world, themselves, and the world. This cognitive activity ignites knowledge discovery.

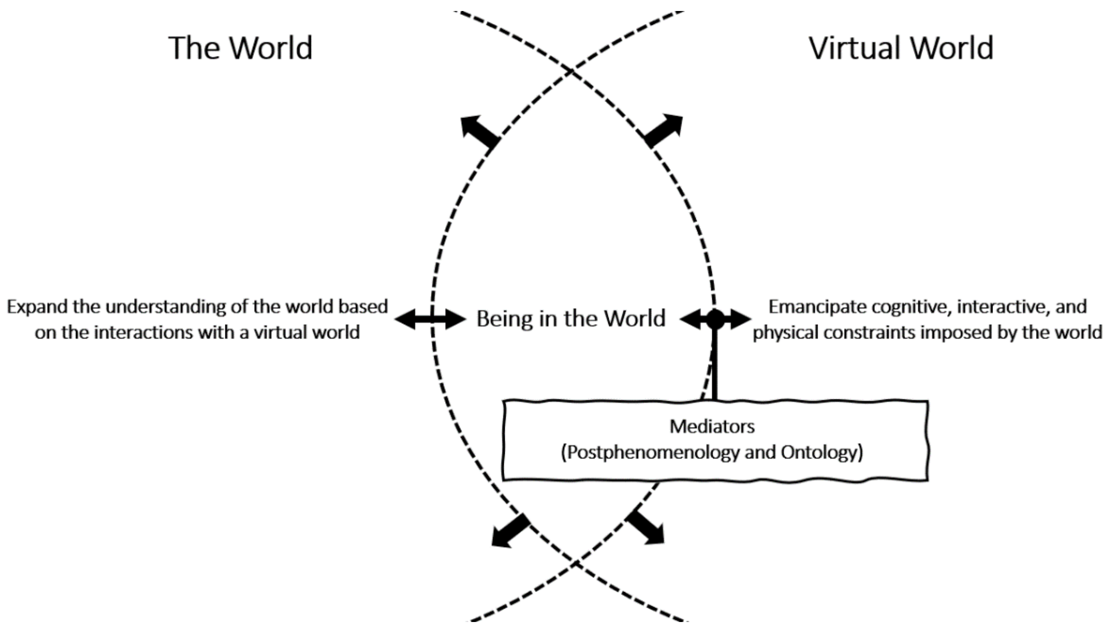

Figure 3.3: Interactive relationships between a user (i.e., being-in-the-world), mediators, a virtual world, and the world.

The game-based approach pragmatically instantiating Martin Heidegger's philosophy shapes a virtual world in terms of interaction design. From an interaction design perspective, to design an efficient system, the nature of user experience should be illuminated first. Since user experience of a system involves emotions and felt experiences, it is subjective [81]. In this research, the design decision on a virtual world takes into account the nature of a game: an engaging, exciting, and entertaining VR environment for user immersion. To realize the decision, a VR environment sufficiently informs users of contextual information about the VR environment, such as where they are, what available tools they have in order to interact with the environment, and why they are here. Once the immersion takes place, it enables a user to equip themselves with strategies for performing high-level cognitive, problem-solving tasks (i.e., data analytics processes for 3D spatial data). In this regard, the game-based approach effectively fulfills the design decision and efficiently represents the VR environment as a self-explanatory environment [83, 84, 85]. Built upon the user experience goal, the game-based approach embodies mediators in the form of a world with a theme and goal-oriented interaction.

Mediators help users to perform cognitive activities in a VR environment by informing users of the general context of a VR application. Delivering the background of a virtual world, such as the concept, goal, and task, a world with a theme is a first mediator when users enter a virtual world. In this research, the background of the world is a space station on a new planet, and there is an exploration mission for users to fulfill. When users approach a virtual avatar, the avatar explains the worldview. This mediator supports smooth transition to a virtual world. In this manner, instead of feeling alienated from a virtual world, users feel that they are having strong psychological connection with a virtual world. The harmonious relationship guides users to not only immersion in a virtual world, but also autonomous action to explore. The entire series of these cognitive processes is postphenomenology as the structural mediator materialized in a world with a theme.

Once immersion into a virtual world takes place, users are prepared for developing strategies on how to achieve a goal of a virtual world with their research questions, hypotheses, and theories. In the case that users have no clue about the exploration, as described in Section 2.2.1, this research suggests three ways of performing exploratory data analysis in terms of visual analytics and sensemaking [23, 64]: a specific and quantitative question, bottom-up processes, and top-down processes. To support exploratory data analysis in a virtual world effectively, user interface (UI) design as a component of game design is selected (Section 3.5.1.2). The purpose of UI design is to make games work better. UI is a starting point of games for users to be informed about a virtual environment, tasks, and how to use tools. In this regard, UI design can facilitate the goal-oriented interaction practically in a virtual world.

Equipped with their strategies, users initiate their action toward their target, such as data, scientific visualization, and 3D virtual objects, in a VR environment to achieve a goal of a VR environment by utilizing an interactive tool. The interactive tool enables them to reach, access, and interact with the target. This research embodies the results of task analysis (Section 3.3.3) in interactive tools for users to analyze, manipulate, and adjust static and dynamic models. By interacting with the models, a user can open up a new dimension in which a user discloses possibilities for expanding the understanding of the virtual world and users themselves, leading to knowledge discovery. In this regard, the goal-oriented interaction materializes ontology as the functional mediator.

The philosophical framework shapes a virtual world to establish a harmonious relationship between a computer, mediator, and human by materializing the structural mediator (postphenomenology) and the functional mediator (ontology) in a virtual world. In the form of the game-based approach, the structural mediator is materialized in a world with a theme, and the functional mediator is materialized in goal-oriented interaction. The following sections describe how the game-based approach implements a world with a theme for user immersion and UI design for effective goal-oriented interaction in a virtual world.

3.5. Virtual Exploration through a Philosophical Tool

The main focus of this research is to design, implement, and evaluate the interactive exploration pipeline where a user autonomously performs exploratory data analysis by supporting sensemaking processes to discover new knowledge and gain insights from 3D spatial data with high dimensions and heterogeneity. To achieve this goal efficiently, Section 3.3 characterized the initial stage of the interactive exploration pipeline (i.e., data, visual mapping, cluster analysis, and task analysis). Section 3.4 described the middle stage as a philosophical mediator between the initial stage and a user materialized in the form of a self-explanatory VR environment. In this section, underlying the initial and middle stages, the final stage presents a thought-experiment VR environment where a user forages information and performs sensemaking to construct or validate their assumption, theory, and hypothesis.

3.5.1. Implementation of Game-based Approach

A virtual world is a space where the initial stage of the pipeline is represented, and the space mediates between the initial components and a user. To help users get familiar with, informed about, and immersed in a VR environment, this thesis implements a world with a theme and user interface (UI) design in the form of the game-based approach. The section below presents how the game-based approach achieves the above goals and instantiates them in a virtual world.

3.5.1.1. World with a Theme

A world with a theme as the structure of a virtual world enables users to feel like they are having a mutual relationship between the theme and the virtual world by providing the context of the virtual world. The role of a world with a theme is not only a smooth transition to a virtual world but also a starting point for connection with a virtual world. As long as a world with a theme is persistently perceptible and constantly provides the spatial and temporal context of the virtual world, while users are playing, it leads to users' autonomous behavior and immersion in a virtual world. In this respect, a world with a theme embodies postphenomenology as the structural mediator between a virtual world and users.

Considering the purpose of this thesis (exploration of 3D spatial data from reservoir simulation), a spaceship exploration mission on a reservoir of a new planet is selected as a world with a theme. The spaceship exploration mission is aligned with exploration of reservoir models to gain insights from them. The world consists of two scenes. The first scene is a spaceship station on a new planet, and the second scene is an environment to explore reservoir models. The first scene consists of a robot, red button, station, and spaceship described in Figure 3.4. The first thing that a user faces in the world is a robot. A robot as a spaceship crew member explains the world and how to use VR controllers for a user who has less experience in the VR interface. Furthermore, the robot is a mediator to connect the first scene with the second scene to inform the user that the second scene is the extension of the first scene. When a user approaches the robot, the following script automatically pops up and destroys (see Figure 3.5). A user can skip the sentences in the panel by pressing a "A" button on their controller.

Welcome to our base camp!

You are on a reservoir exploration mission to a new planet!

I am Jarvis and here to explain about the mission.

Our Mission is to explore static and dynamic reservoir models with interactive tools!

I will give you guidelines where to start later!

Whenever you feel you are well informed about the mission, hit the red button on the left side of you.

Then you will teleport to the mission scene!

Don’t worry! I will always be with you, so ask me anything!

Firstly, I would like you to get familiar with how to control your controllers.

On the right controller, you can use a joystick to move around the scene.

You can select something by using A button on the right controller.

On the left controller, you can use a joystick to change your view.

Feel free to test your controllers and look around our base camp!

Good job! Now, I will explain about your mission!

When you hit the red button on your left side, your mission will begin!

Let’s meet there and I will explain everything about the mission!

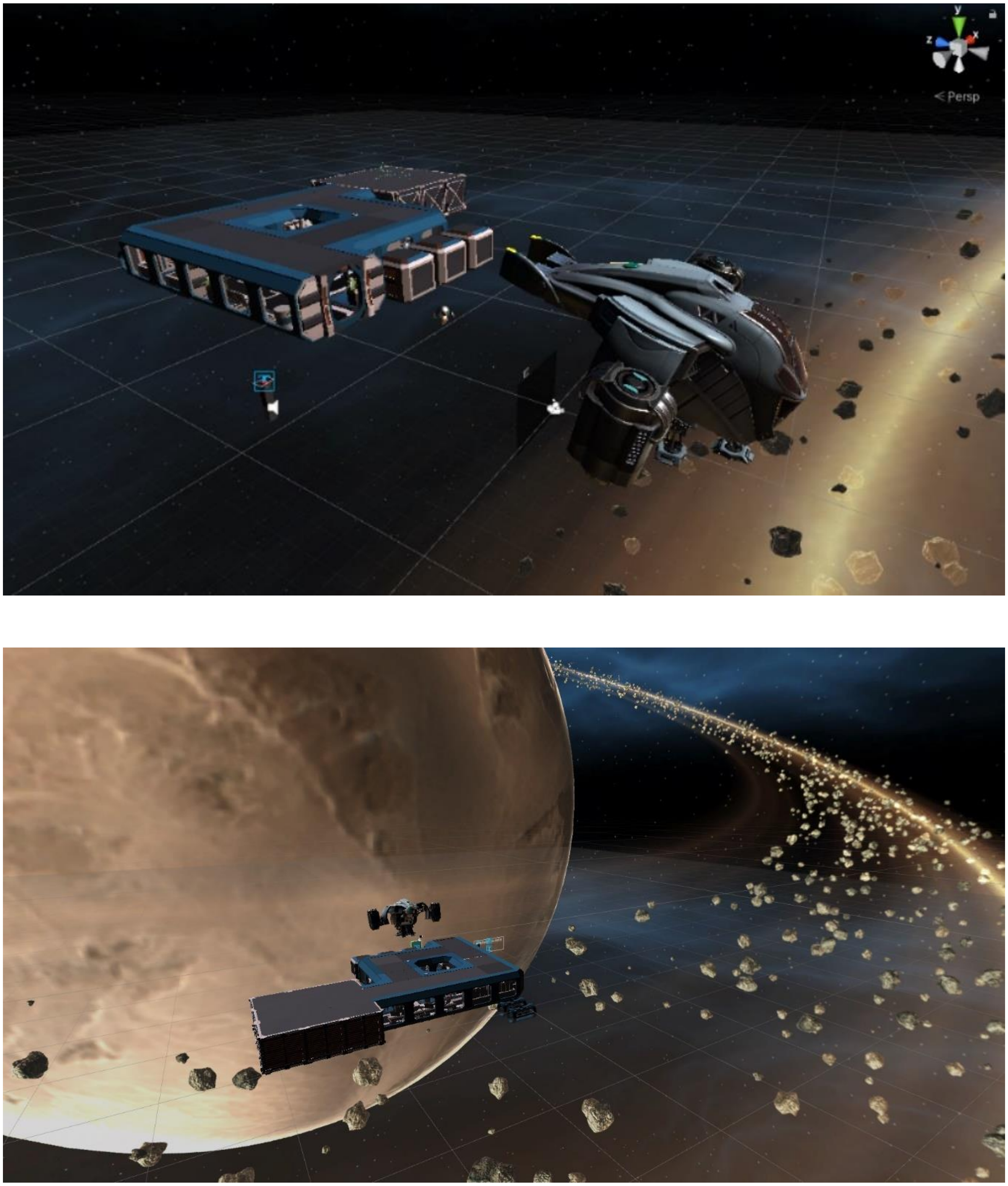

Figure 3.4: Elements of the first scene: a spaceship, robot, red button, and station on a new planet.

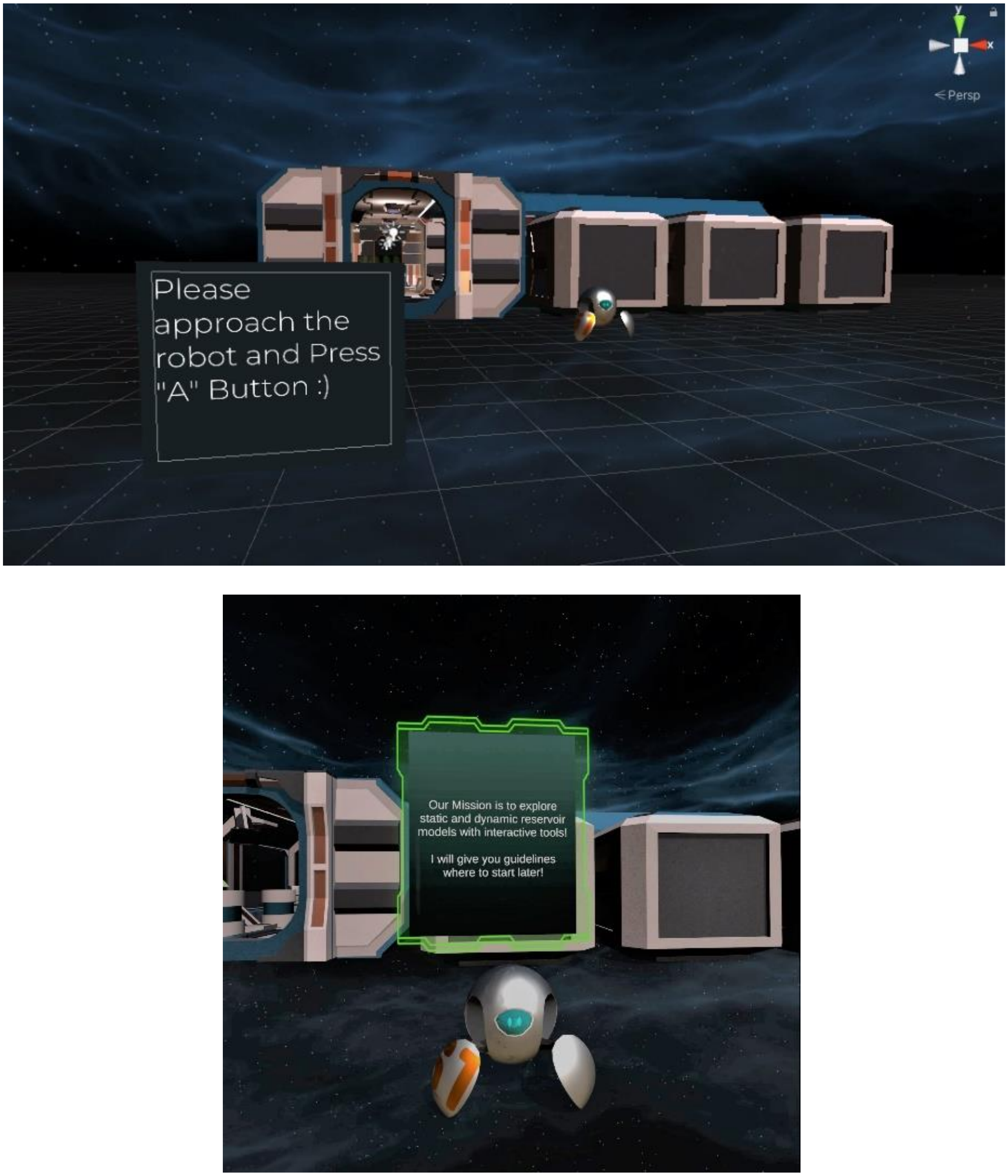

Figure 3.5: When a user enters this virtual world and approaches the robot, the robot talks to the user.

By placing a spaceship and station in the first scene, the first scene helps a user offer the general context of the virtual world. This allows the user to look around the virtual world, getting familiar with the world and the interface. With the user immersed in the world, the user is ready to explore complex 3D spatial data. Once the user feels comfortable with the world and the interface, the red button teleports the user to the second scene to explore reservoir models by pressing the red button (see Figure 3.6). Then, the countdown begins. The role of the red button is to get the user ready to perform tasks and to offer autonomy to control the transition to the second scene.

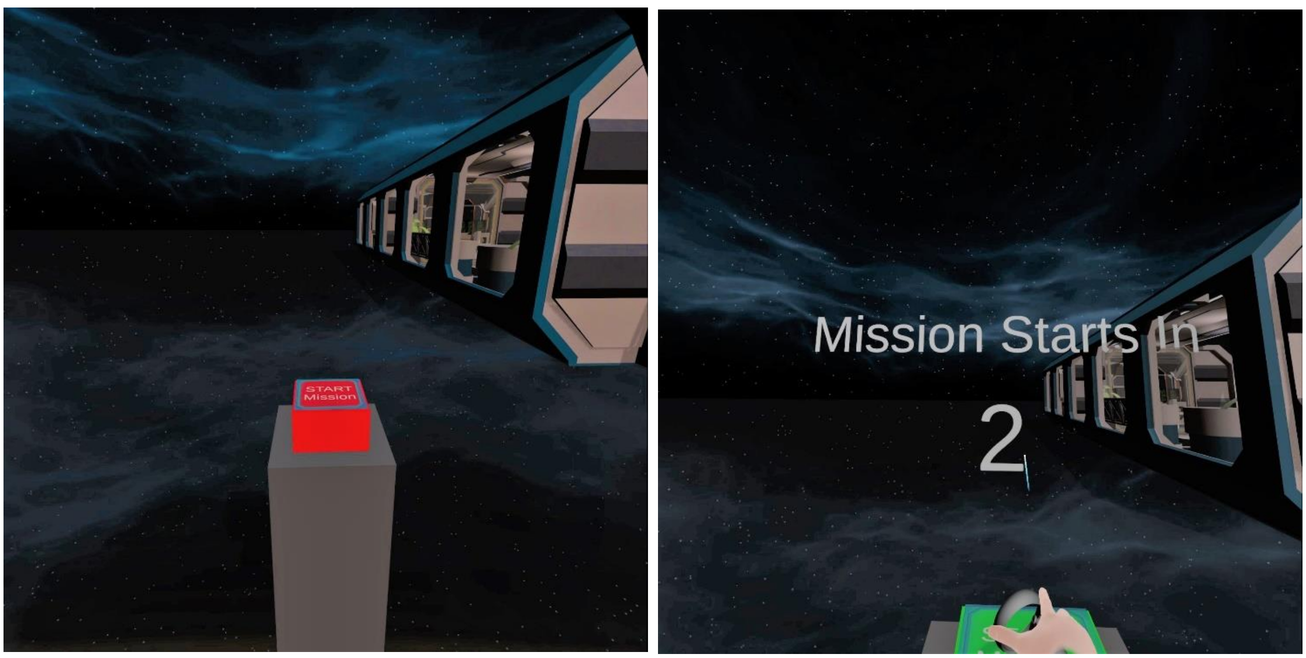

Figure 3.6: When a user presses the red start button, the button teleports the user to the second scene.

The second scene as the extension of the first scene is the subsurface of the new planet and an exploratory data analysis environment to explore the reservoir of the new planet with analysis tools. The second scene consists of a static model, dynamic model, interactive tools (i.e., touch screens, recording balls, and grabbable text blocks), and the same robot in the first scene (see Figure 3.7). The robot as a mediator between the first and second scenes explains the structure and goal of the second scene, static and dynamic models, interactive tools, and three ways of performing exploratory data analysis for a user to start with. The guideline of the three ways is for those who still do not know where to start with the deluge of 3D spatial data information. Although a user is well informed about the VR interface, interactive tools, and user tasks, a user can be overwhelmed by the deluge of the information. This overwhelming feeling can be ascribed to the complexity of exploratory research. Since exploratory research rarely provides clear questions and answers about data, this thesis tries to support a navigational guideline for a user to gain insights from the data.

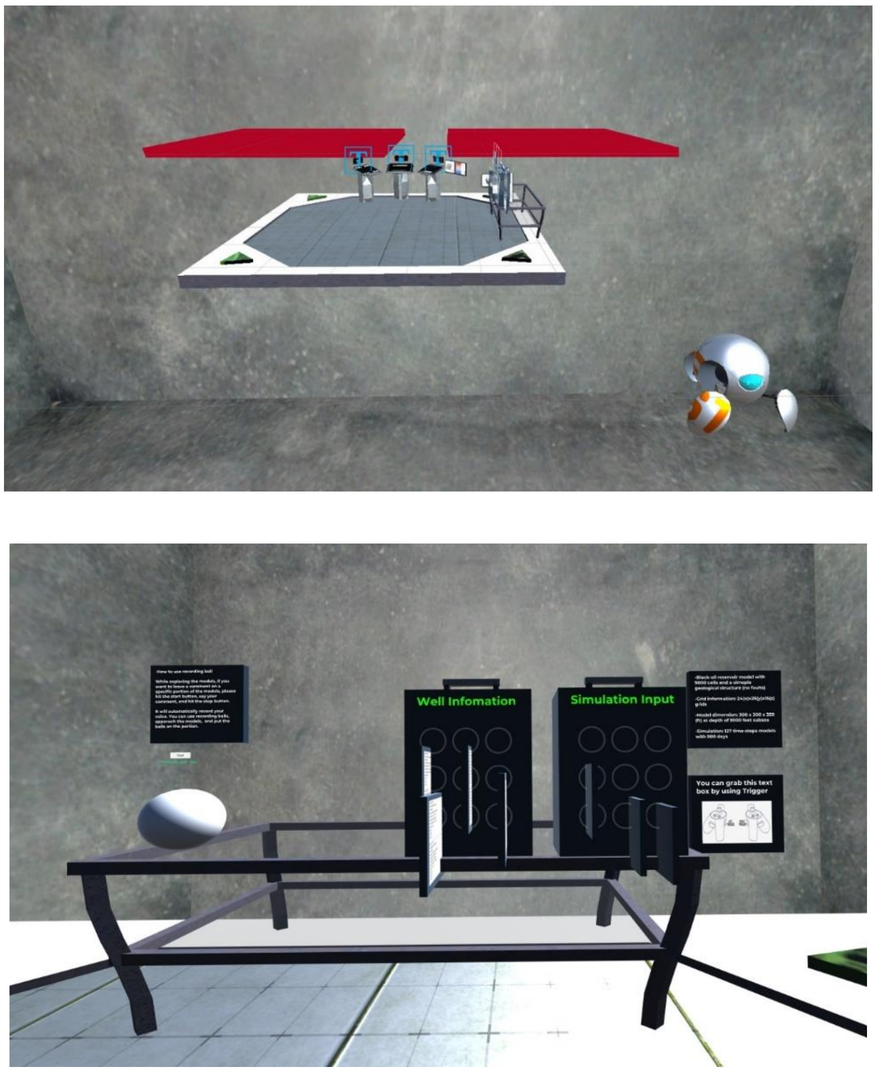

Figure 3.7: Elements of the second scene: a static model, dynamic model, interactive tools (i.e., touch screens, recording balls, and grabbable text blocks), and the same robot in the first scene.

There are three navigation ways to explore the 3D spatial data (detailed in Section 2.2.1). Firstly, a user can start with a specific and quantitative question. For example, questions to consider are “how much oil is produced during a specific time range?” or “how many groups show similar behavior in the data?”. These questions might lead to subsequent or new questions by discovering implicit knowledge described in the data. Secondly, there are two ways to ask a fuzzy and qualitative question for sensemaking by taking bottom-up and top-down approaches. For example, one question would be “which areas of a static model can potentially affect optimal oil production of a dynamic model?”. With the background of the domain and interactive tools, this question supports a user to explore a VR environment clearly and efficiently.

The next description is a script for the robot to explain the structure and the goal of the second scene, the static and dynamic models, the interactive tools and the three ways of exploratory data analysis. This research designs the script as simple, short and effective as possible to convey key contents.

Good to meet you again!

Reservoir exploration mission consists of a static model (right), dynamic model (left), touch screens to control and manipulate the models, and interactive tools, such as touch screens, text blocks and recording balls on a table.

The goal of the mission is to explore the models and the implications of the static model for the dynamic model by using the touch screens.

The details about the simulation inputs are described in the text blocks on the table.

Furthermore, while exploring the models, if you want to leave a comment on a specific portion of the models, you can use recording balls and put the balls on the portion.

The balls enable you to trace your logic, actions, and decisions and store the history of the information processing steps.

Let me explain about the static model first!

On the static touch screen, you can select porosity, heterogeneous permeability and the patterns of the integrated permeability in the *x*, *y*, and *z* directions implemented by cluster analysis.

The patterns show groups of permeability that share similar behavior.

There are three different pattern parameters to select: a conservative (small) number of clusters, middle number of clusters, and progressive (large) number of clusters.

The dynamic model consists of 16 different dynamic properties and 127 time-series models.

You can select a specific property and time step or play time-series visualization.

Also, you can adjust the size of the dynamic model and specify the region of your interest!

Furthermore, there is a 2D graph button showing oil, gas, and water productions over time.

You don't have a clue where to start?

There are three ways to explore the models!

Firstly, you can ask a specific and quantitative question!

For example, how much oil is produced during a specific time range?

Secondly, you can take bottom-up processes!

While playing with the touch screens and gathering information, you can expand your understanding of the data. Then you can make inferences or come up with your hypothesis or theory!

Lastly, you can take top-down processes!

If you have your assumptions, hypotheses, or theories to support or invalidate or want to test alternative theories, that can be the first step to start with!

Let's get this reservoir exploration mission started!

Good luck to you!

3.5.1.2. User Interface Design

The purpose of User Interface (UI) design is to analyze a way how to efficiently deliver information to users. The components of UI design are movement of UI, saturation, typography, readability of text, interactivity, and intuitiveness [61]. This thesis adopts readability of text, interactivity, and intuitiveness. Readability of text is about not only straining users' eyes but also conveying key information as simply and effectively as possible. Interactivity is to get users involved in a virtual world by interacting with interactable objects, such as a button and dropdown menu. Intuitiveness is to put a VR designer in a user position and to try to understand how a user perceives information from UI. From this perspective, the VR designer reflects the user perspective on a virtual world. With these three components, this research designs a user-centered UI.

Design decisions on UI take into account a world with a theme and familiarity with reality. From a phenomenology perspective, to build a successful immersion and VR application, a virtual world and its virtual objects sufficiently simulate aspects of reality. In this regard, considering a spaceship exploration mission as a world with a theme, the virtual objects are designed to represent the characteristics of the world view. On the basis of the characteristic, familiarity with reality articulates the virtual objects. For example, spaceship touch screens are selected as UI to control and manipulate 3D spatial models. A spaceship robot informs a user of the VR application. The design decisions enable augmenting user experiences in a virtual world, leading to user immersion and autonomous behavior for user tasks.

Including the dialogue box of the first and second scenes, there are four UIs that users can interact with: Help, Size, Static, and Dynamic (see Figure 3.8). Help is a UI in which the robot in the first and second scenes talks to a user when a user approaches the robot and a dialogue box automatically pops up. The purpose of Help is to provide the general context of the virtual world whenever a user needs help. In the first scene, Help mediates between the virtual world and a user by explaining the context of the world (e.g., goal and mission). In the second scene, Help describes the structure of the second scene, static and dynamic models, interactive tools (four UIs), and three ways of performing exploratory data analysis for a user to start with. Size, Static, and Dynamic are touch screens of a real-world analogue with a spaceship theme. This design decision takes usability of intuitive data manipulation into account. When a user hovers over the screens with their virtual right hand, a laser beam automatically detects the screens, which allows a user to interact with and select dropdown menus, a slider, and buttons.

Size allows a user to adjust a dynamic model in the *x*, *y*, and *z* directions by controlling a slider. In this manner, a user can specify the scope of data to investigate a specific part of the model. Removing unwanted data supports a user to focus on their interests within the high dimensional and heterogeneous data. Static allows a user to access a static model. Static consists of a dropdown menu, which enables a user to select permeability in the *x*, *y*, and *z* directions, constant porosity values for 15 layers, and HDBSCAN pattern parameters. HDBSCAN pattern parameters consist of three types: a conservative (small) number of clusters, an intermediate number of clusters, and a liberal (large) number of clusters. The three types are the different number of precalculated clusters generated by different HDBSCAN parameter selection in Python. These options offer a user the ability to select the number of clusters without having any background concerning HDBSCAN and the data. In this regard, interactive data exploration is adopted by involving a user in the process of pattern discovery.

Dynamic allows a user to access a dynamic model. Dynamic consists of two dropdown menus (a time-step model selection menu and property selection menu), a play button, and a 2D graph button. A time-step model selection menu is comprised of 127 time-step models that represent the development of a reservoir property over 127 time-steps. There are 16 reservoir properties, and a user can select a specific reservoir property and model in a specific time step. When a user selects a reservoir property and hits a play button, a model with the reservoir property changes over time. The play button shows trends, expected and unexpected patterns, and relationships in data. A 2D graph button includes oil, water, and gas production rates over time and oil, water, and gas cumulative productions over time to help a user intuitively understand data.

Utilizing the four UIs (Help, Size, Static, and Dynamic), a user is ready to perform exploratory data analysis on the static and dynamic reservoir models. The following shows how a user explores 3D spatial data by employing the interactive tools in a virtual world.

Figure 3.8: Touch screens of a real-world analogue with a spaceship theme.

3.5.2. 3D Spatial Data Exploration

An interactive VR environment is a VR information space that has a relationship between object space (3D spatial data), attribute space (its properties), data structure space (clustering algorithm), and user experience space [65]. In this environment, utilizing the interactive tools (Help, Size, Static, and Dynamic) in the second scene, a user can reach, access, and interact with the VR information space. This exploration encourages a user to understand, forage, and organize information systematically based on the domain background and the understanding of the virtual world. The systematic information processing as the basis of sensemaking is a bridge leading a user to perform high-level problem-solving tasks. To efficiently support the series of the cognitive processes, it is important to investigate domain-specific tasks and practically design and implement tools for the tasks.

The results of the task analysis (Section 3.3.3) are embodied in the interactive tools in order for a user to explore 3D spatial data effectively and perform sensemaking efficiently. These results are Overview, Detail-on-demand, Filter, Cluster, Relate, History and Extract, and Time. In Figure 3.9, this research categorizes the interactive tools into three functions: Navigation (Overview and Filter), Augmentation (Relate, Time, Detail-on-demand, and Cluster), and Elaboration (History and Extract). In the game-based VR environment, the interactive tools are materialized in the form of interactable objects, such as buttons, dropdown menus, and a slider, which are components of the UIs described in Section 3.5.1.2. The present section describes how a user can utilize the interactive tools and how the tools initiate cognitive processes to perform exploratory data analysis in the VR environment.

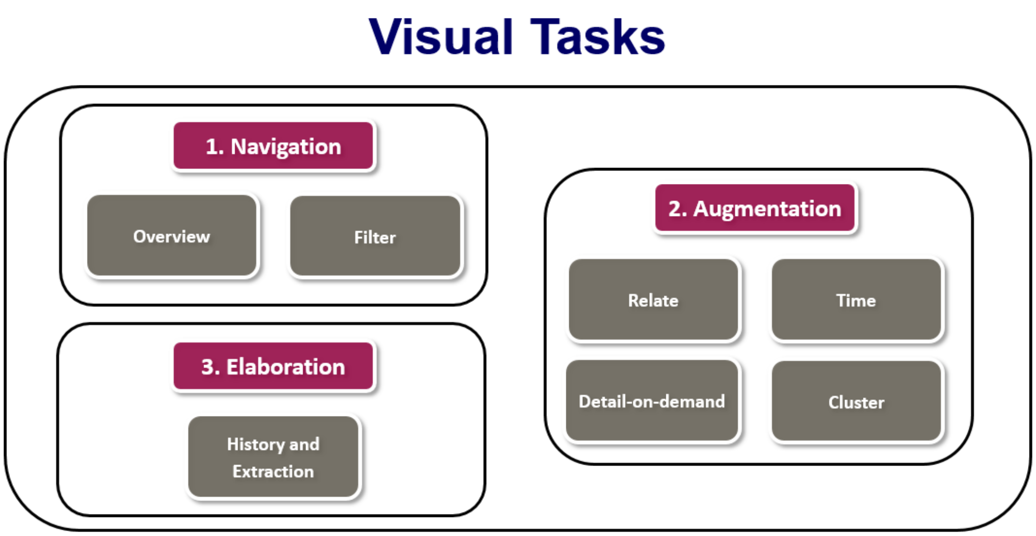

Figure 3.9: Visual tasks designed from a reservoir engineer perspective.

3.5.2.1. Navigation (Overview and Filter)

The VR environment framed by postphenomenology and ontology consists of the two scenes for a user to navigate. These philosophies are materialized in the form of the game-based approach: a world with a theme (i.e., spaceship exploration mission) and goal-oriented interaction (i.e., interactive tools). Reifying postphenomenology as the structural mediator for user immersion, the first scene is designed to provide a user with the context of a virtual world and instruction on how to use controllers, prepare user tasks, and immerse themselves in a virtual world. This design decision is ascribed to the amount and complexity of information in scientific visualization, which can overwhelm and alienate a user from a virtual world. This problem can lead a user to lose interest in exploration and fail to involve transition from the low-level to the high-level problem-solving tasks. Navigating the first scene enables a user to feel comfortable with the new virtual world and elicits autonomous user behavior to explore and forage information from static and dynamic models in the second scene. If a user wants to skip the first scene, by pressing a red button directly, a user can teleport to the second scene.

Built upon the concrete immersion in the first scene, embodying ontology as the functional mediator, the second scene support users to initiate sensemaking processes from static and dynamic models. To support a user with information gathering from the visualization models, text blocks describe the input parameters of the Black Oil Simulator, such as initial conditions, water-oil capillary pressures, and porosity values (see Figure 3.10). Since the Black Oil Simulator outputs 3D spatial data, which is mapped to the static and dynamic reservoir models, information about the simulation inputs can help a user navigate the models by providing the context of the static and dynamic models in terms of the cause-effect relationship. By grabbing and placing the text boxes, a user can interact with the physical artifacts and extract necessary information. Furthermore, moving around the scene with the text blocks, a user can increase their understanding of the simulation mechanism visually and the relationship between the models.

The distinctive feature of VR technology compared to traditional desktop-based interfaces is that in a VR environment, it is easy to change perspective of navigation (zooming in and out) without constraints on views, such as quality of resolution, camera location, viewing angle, and the shape and size of the viewing frustum [65]. This allows a user to freely explore all possible angles of the models without spatial limitation. In addition, a VR environment allows a user to have easier visual access to large amounts of information. This accessibility virtually enables physical interactions with visualization models. This spatial sense of the 3D models from the interactions provides a user with a better general understanding of the models, basic intuition about the reservoir property distribution, and a basis for subsequent exploration in a more in-depth investigation. In these regards, the layout of the two models in one scene supports a user with spatially unbounded, virtually physical navigation for information gathering.

With the overall understanding of the 3D models, a user can specify a range of interest of a model by using a slider (see Figure 3.11). The filtering slider enables a user to adjust the range of the dynamic model and eliminate the uninteresting region of the model by reducing the volume on the VR space. This function assists in focusing on specific data that a user is interested in. Narrowing the volume down into the region of interest can improve task performance by concentrating on their interest and immerse a user in the VR space by giving the autonomy in the data, which leads to subsequent exploratory data analysis. Utilizing the collected information, a user is prepared for transition to high-level cognitive operations on the models.

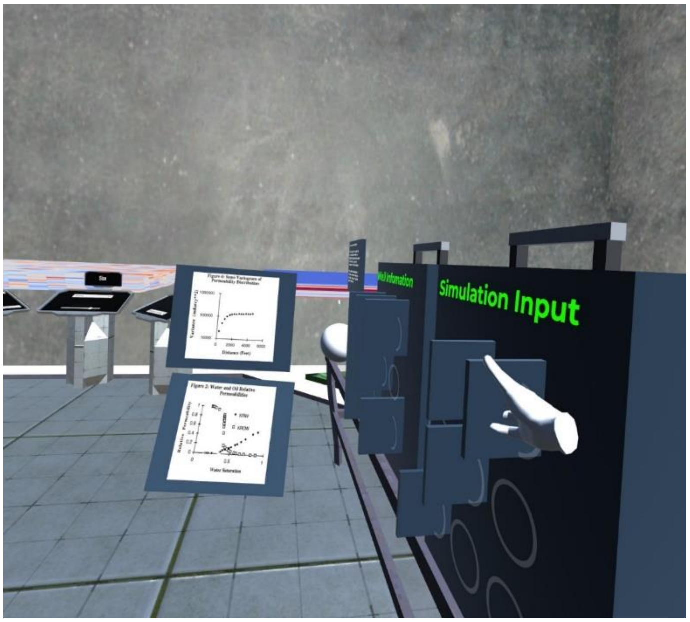

Figure 3.10: A user can grab and place text blocks on the space to read and understand the system of black oil simulation.

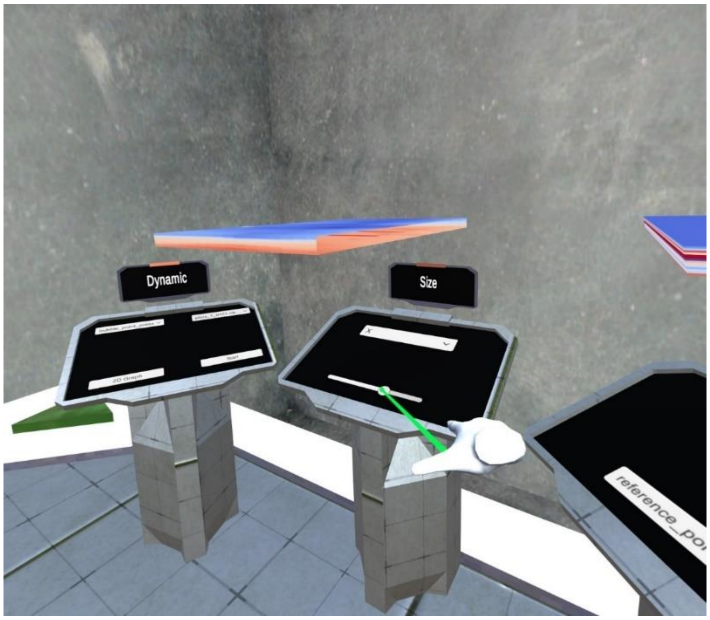

Figure 3.11: A user can change a range of the dynamic model by setting the direction of the model (i.e., *x*, *y*, and *z*) and utilizing the slider.

3.5.2.2. Augmentation (Relate, Time, Detail-on-demand, and Cluster)

Underlying the overall understanding of the 3D models achieved through the navigation in the second scene, a user can augment the navigational information. The augmentation involves two cognitive processes: a foraging process and sensemaking process. The foraging process is stages of expanding the understanding of the static and dynamic models, and it consists of three stages: exploring, enriching, and exploiting [64]. By utilizing the interactive tools, such as a play button, a filtering button, and dropdown menus, a user explores the VR information space, enlarges the range of new information, and reflects on the new information in the exploratory data analysis process. By narrowing the focus of interest and improving the levels of the precision of the exploratory information, a user enriches the collected information. Exploiting the enriched information, a user can extract new knowledge, discover patterns, and make inferences [64].

Based upon the foraged information, a user can perform sensemaking. The sensemaking process consists of problem structuring, evidentiary reasoning, and decision making. Problem structuring is a process of analysis to formulate hypotheses from the foraged information. Evidentiary reasoning is conducted by collecting evidence that is selected from the foraged information to generate, support, or invalidate the hypotheses. As the final stage of the sensemaking process, the decision-making proceeds to select an alternative from numerous possibilities and take action [64]. To help a user initiate the foraging and sensemaking processes, a robot in the second scene briefly suggests three ways of exploratory data analysis, such as raising specific and quantitative questions and exploring both bottom-up processes (foraging processes) and top-down processes (sensemaking processes). Equipped with these strategies and utilizing the interactive tools, a user can initiate exploratory data analysis by relating static and dynamic models.

The purpose of relating static and dynamic models is to provide a user with a better foundation for reservoir simulation [45] by enabling a user to explore the input and output of the 3D spatial data in one scene. The most intuitive way to understand the cause-effect relationship between static and dynamic properties is to visualize the recovery processes with time-series visualization. Time is a key component showing trends, expected and unexpected patterns, and relationships in the data. Supporting basic time-based analysis, the “play” button allows a user to simulate the distribution of a dynamic property over time. The time-series visualization of a dynamic property intuitively exhibits how heterogeneous permeability affects dynamic properties over time. To navigate and investigate the complications of the recovery processes, a user can specify a specific dynamic property and select a certain time-step model among 127 time-step models. The time-series visualization tool provides instant visual feedback and supports further exploratory data analysis.

Furthermore, a user can select and change a specific static and dynamic property by utilizing dropdown menus. Since 3D spatial data with numerous attributes is a compacted information display, a user wants to explore not only the overview of the information for general understanding, but also a subset of the data in detail. Visualizing a subset of data can be useful for a user to obtain and focus on specific information. In order for a user to browse and access details to the models, a 2D graph for the dynamic model and cluster analysis implemented in the static model are designed. A 2D graph shows oil, water, and gas production rates over time and oil, water, and gas cumulative productions over time to help a user intuitively understand the behavior of the dynamic properties over time. The 2D graph button is designed for a user to link the oil, gas, and water productions with the permeability patterns from cluster analysis as an indicator for estimation of the dynamic properties. Cluster analysis is employed to simplify non-spatial data for pattern discovery and support exploratory data analysis by linking the patterns and their impact on dynamic properties. Figure 3.12 shows the dynamic touch screen elements: the 2D graph button, dynamic property dropdown menu, time step dropdown menu, and time-series visualization play button.

Simplifying non-spatial data (permeability in the *x*, *y*, and *z* directions) corresponding to each cell, cluster analysis is performed in terms of both the domain (reservoir engineering) and interactive data exploration perspectives. From the domain perspective, permeability is key data for the initial estimation of oil production volume. Since clustering permeability is utilized to differentiate a reservoir model into portions that show similar behavior, a user can perform exploratory data analysis based on the patterns and domain background. To achieve interactive data exploration in cluster analysis, visual analytics and human-computer interaction are integrated. From these perspectives, involving a user in the process of pattern discovery allows the resulting patterns to be more relevant and interesting to user tasks [67]. The static dropdown menu enables a user to participate in the process by selecting and changing the resulting patterns. The interactive data exploration in cluster analysis makes it possible for a user to discover expected or unexpected patterns and extract relationships in the large amounts of the data [47, 66].

There are three pattern options that a user can select and change: a conservative number of clusters, intermediate number of clusters, and liberal number of clusters. These three pattern options as calculated labels of HDBSCAN in Python are determined by the number of clusters. The pattern options offer user discretion to choose the different patterns depending on the user's judgement of the data. Although a user might not have prior knowledge about the principle and input parameters of the clustering algorithm, a user can easily utilize the three pattern options to involve the process of pattern discovery (see Figure 3.13). To contextualize the three pattern options in the virtual world, the robot informs a user of the purpose, meaning, and manipulation. To support subsequent exploration in a more in-depth investigation with the pattern discovery, the VR environment supports the simulation input text boxes, the filter, the time-series visualization tools, and the 2D graph. This exploratory data analysis reveals the relationship between the selected static and dynamic properties and information about the phase behavior of a specific dynamic properties, the recovery processes, and the reservoir performance.

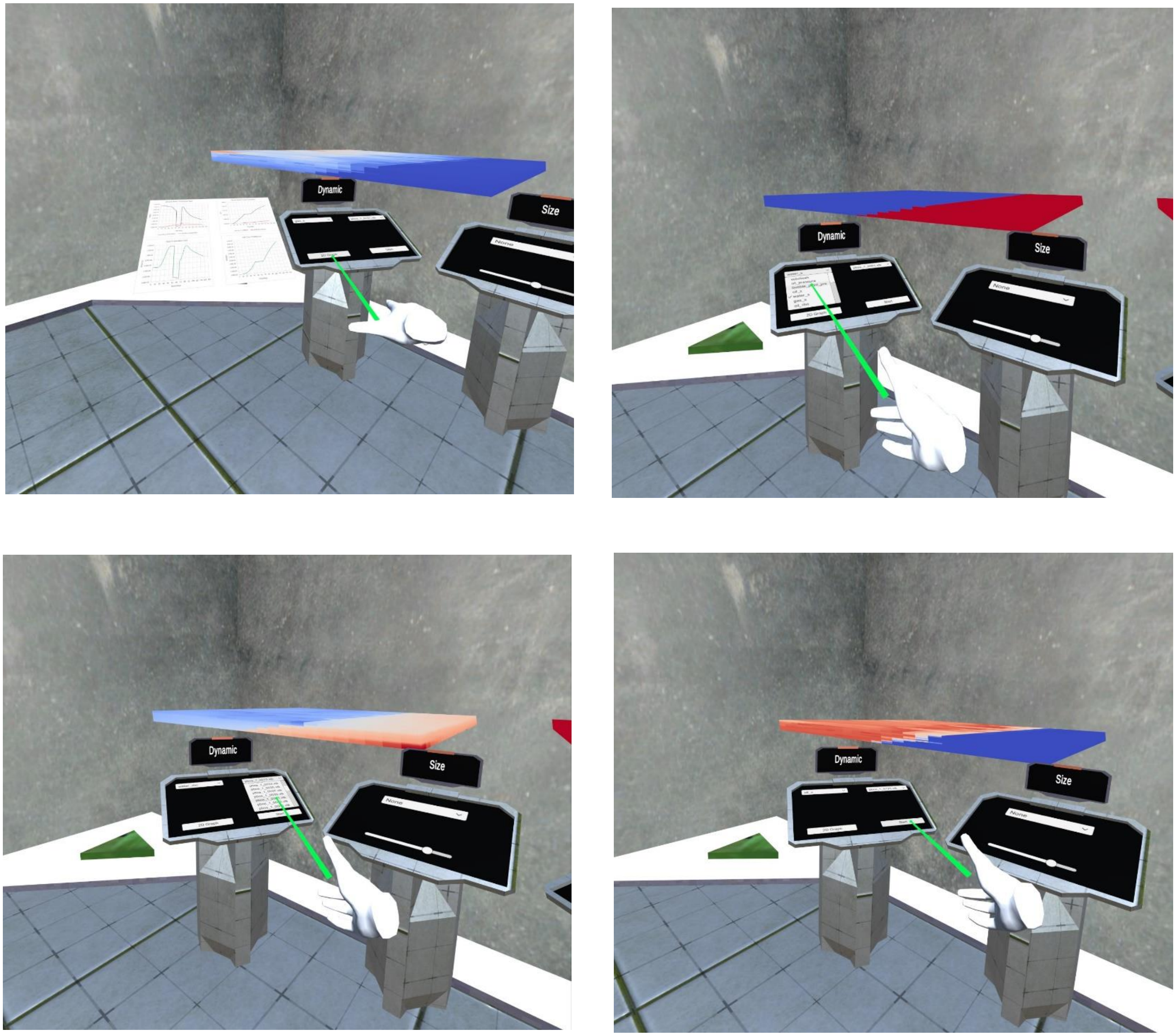

Figure 3.12: Elements of the dynamic touch screen: the 2D graph button (top left), dynamic property dropdown menu (top right), time step dropdown menu (bottom left), and time-series visualization play button (bottom right).

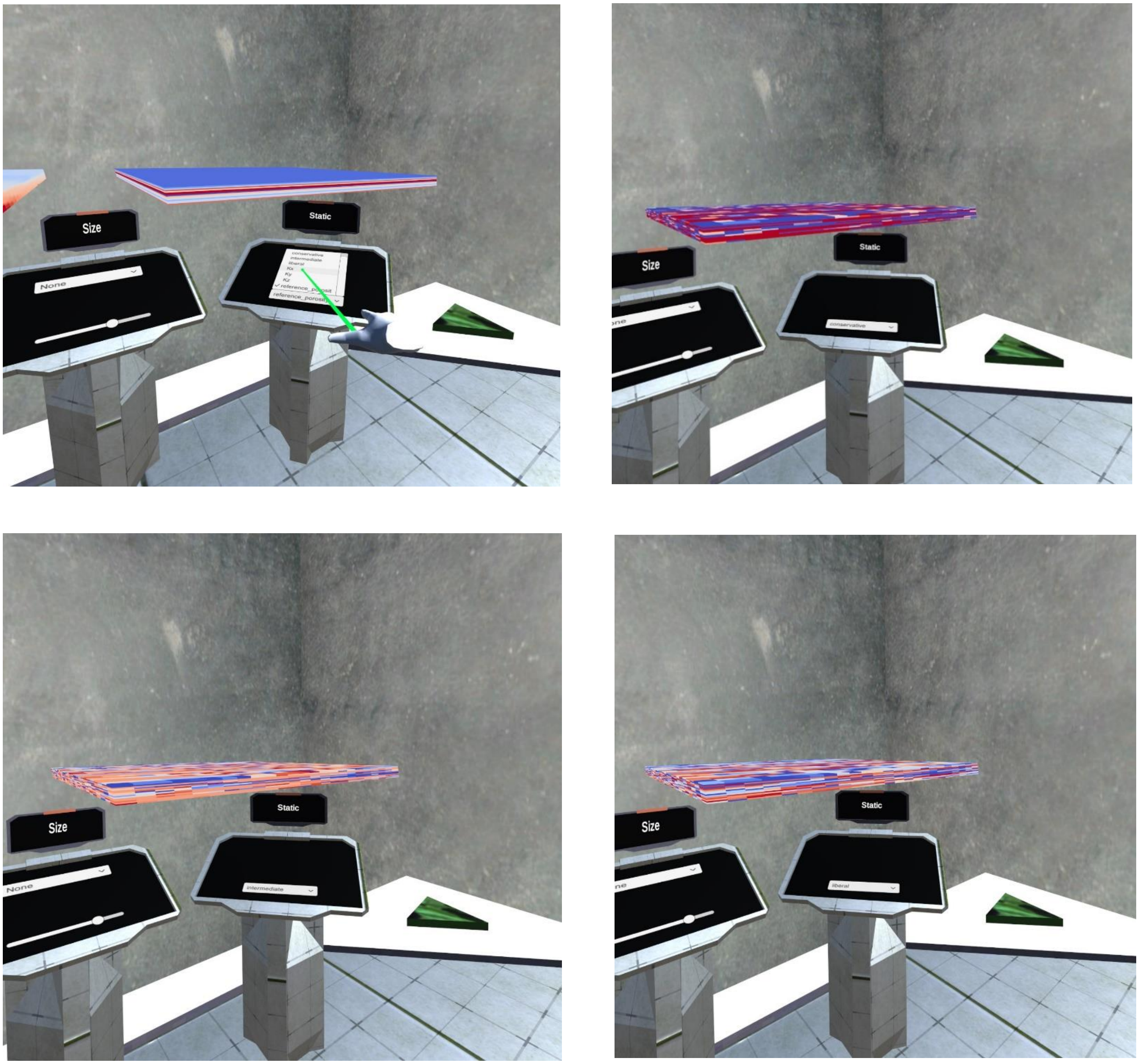

Figure 3.13: Three permeability patterns on the static touch screen: the dropdown menu (top left), conservative (top right), intermediate (bottom left), and liberal number of clusters (bottom right).

3.5.2.3. Elaboration (History and Extract)

The VR environment has supported immersive analytics for user engagement in the first scene and higher-bandwidth communication between VR, AI, and a user in the second scene to elicit human capacity to think based on collected data. Informing the goal, user tasks, and interactive tools of the VR environment, the second scene has enabled a user to perform information gathering through Overview, Filter, Relate, Time, Detail-on-demand, and Cluster. By organizing the collected information, a user can transfer to a higher-level cognitive task of the sensemaking processes, such as schematizing, problem structuring, and evidentiary reasoning in accordance with the Pirolli and Card model (detailed in Section 2.2.1). In the process of performing the higher-level cognitive tasks, more in-depth exploration and evidentiary reasoning are involved. On the other hand, in the case that a user has their own hypothesis or theory, the information gathering can be performed to support or validate them.

For either the bottom-up or top-down processes, since these processes involve various cognitive steps, recording balls can support a user with an information archive system to retrace the logic, actions, and decisions. While exploring and relating the models, a user can immediately utilize the balls and record their voice by grabbing the ball and pushing the start button on the recording ball. To recall the cognitive steps easily, a user can put the ball on a specific area of the models if a user finds new information, patterns discovered from the models, or the area of important implications during the exploration (see Figure 3.14). The ball automatically saves the recording file, which enables a user to replay and browse the history. Furthermore, since the recording file can be shareable, this function can facilitate collaborative and remote works in which experts freely exchange their decision-making processes to reach a consensus.

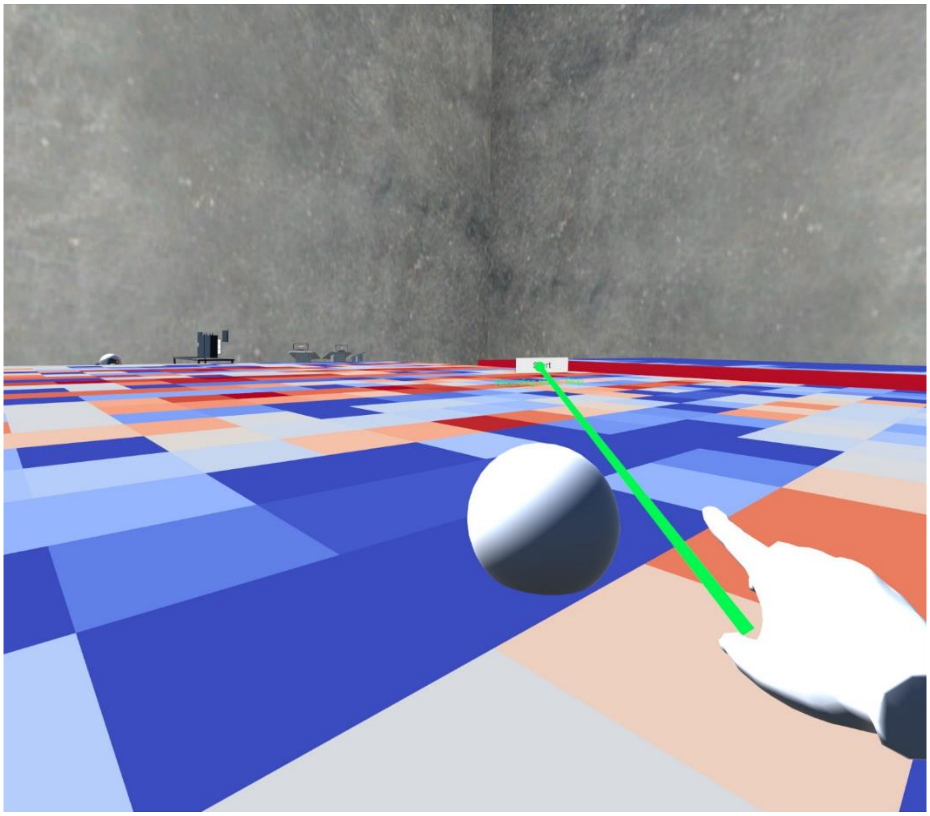

Figure 3.14: A user can put the ball on a specific area of the model and record the voice by grabbing the ball and pressing the start button on the recording ball.

Chapter 4

4. Evaluation

This chapter aims to conduct exploratory evaluation [82], which is to discover, identify, and provide new possibilities of applying Heidegger's philosophy to human-computer interaction in a virtual world. To achieve this goal, this chapter investigates how Heidegger's philosophy reified in this VR application affects interactions with a virtual world, immersion in user tasks, and cognitive processes when exploring complex and large 3D spatial data through a qualitative study (i.e., a questionnaire, interview, and observation of user behavior in the virtual world). This chapter starts by describing the evaluation procedure (Section 4.1) and academic and professional background and VR experiences of all participants (Section 4.2). The following describes the evaluation results (Section 4.3) built upon the qualitative study which is designed to evaluate the usability of this VR application in three categories: the structural mediator for a self-explanatory VR environment, the functional mediator for user tasks performance, and being-in-the-world for expansion of knowledge and experiences of users by this VR application. Finally, the discussion section (Section 4.4) provides reflections on the findings.

4.1. Evaluation Procedure

The evaluation procedure proceeded with introduction, VR demo, and post-interview sessions. The procedures and questionnaire can be found in the Appendix A. The total sessions took approximately 1 hour. During the introduction session, I explained the purpose of the evaluation and the philosophical framework of the research to users. The purpose is to test this VR application as a self-explanatory VR environment in which a user can perform from low-level to high-level cognitive tasks.

The structure of the philosophical framework encompasses the world and a virtual world (detailed in Figure 3.3). A user is right in the middle of the realms in which is transferable to both the realms at any time through a VR device. The realm of the middle is called "being-in-the-world". From the philosophy of technology perspective, being-in-the-world as humans is their private and structured scope having their potential for extending their understanding of the scope to the world through technology. In this regard, user experiences and interactions with a virtual world constructed by VR technology can affect their being-in-the-world, providing an opportunity for the extension of their being-in-the-world to the world. To reify the philosophical framework in the VR application as a self-explanatory VR environment, this research has adopted postphenomenology and ontology in the form of the game-based approach. These philosophical tools not only support a smooth transition from reality to an exciting and entertaining virtual world to elicit user immersion, but also mediate between the virtual world, humans, and 3D spatial data to assist sensemaking processes.

Following the introduction, before proceeding with the VR demo session, I asked participants whether they have had VR experience. If a participant had no VR experience, I only informed the participant of how to use controllers. It took less than 30 seconds. If a participant had VR experience, the VR demo session began based on the information from the introduction session. Thus, for the VR demo session, there were no provided training or instruction as to the VR demo. The idea of the design of a self-explanatory VR environment in this thesis is from observations when oil and gas companies visited our VR research lab. In a VR demo session, while engineers were exploring the virtual world with a VR headset and controllers, I observed their interactions, information processing steps, and thought process through the desktop monitor. These observations serve as the basis of identifying, discovering, and generalizing user behavior patterns for their sensemaking processes.

The post-interview session proceeded with a questionnaire and verbal interview. The questionnaire implicitly asked the participants to evaluate postphenomenology as the structural mediator, ontology as the functional mediator, their sensemaking processes, and being-in-the world. The interview consisted of a predefined set of questions and a tailored set of questions based on the answers to the questionnaire. While proceeding with the interviews, the contents were recorded for further analysis and clarification of their comments with the consent from all the participants.

4.2. Participants

There were six participants recruited. This research investigated background, specialization, and VR experience of participants to validate a hypothesis of this thesis: these factors (i.e., familiarity of domain knowledge and level of VR experience) would have a significant impact on sensemaking performance in a virtual world (i.e., task performance in a virtual world). Table 4.1 describes the six participants. The subject

4 has a M.Sc. degree in Geophysics and is currently working in the IT industry. For the subject 5 and subject 6, they both have B.Sc. and M.Sc. degrees in Computer Science. One of them is a Ph.D. student in Petroleum Engineering, and the other is professional in the oil and gas industry and human-computer interaction with over fifteen years experiences in the IT industry. From the background and specialization of the participants, the VR application was evaluated in the various aspects: human-computer interaction, reservoir engineering, computer science, and geophysics.

This research asked all the participants if they had VR experience. This research divided VR experience into three cases: VR experience in both VR games and academic VR applications, VR experience in either VR games or academic VR applications, or no VR experience. This question was designed to elicit different opinions on this evaluation from different VR experiences. For the subject 3 and the subject 4, they had no VR experience before this study. For the novice users, I only informed the participants of how to use the controllers. There was no training, instructions, and explanation as to the VR application. For the experienced users, the participants put on the VR headset without any information and training about the VR application after the introduction session.

Participant	Background	Specialization	VR Experience
Subject 1	Petroleum Engineering	Ph.D. Petroleum Eng Student	VR Games Academic VR applications
Subject 2	Safety Assessment of Pipeline	Ph.D. Petroleum Eng Student	VR Games
Subject 3	Pipeline Engineering	M.Sc. Petroleum Eng Student	X
Subject 4	Software Development	Geophysics	X
Subject 5	Software Development	Ph.D. Petroleum Eng Student	VR Games Academic VR applications
Subject 6	Computer Science	Oil and Gas Technology	VR Games Academic VR applications

Table 4.1: Background, specialization, and VR experience of the participants.

4.3. Evaluation Results

This section represents the usability evaluation results of this VR application based on a questionnaire, interview, and observation of user behavior in a virtual world. To clarify what aspects of the VR application will be measured, the terms ‘usability’ and ‘evaluation’ are defined first. Usability indicates that it includes all related virtual artifacts that have affected human use of a virtual artifact. Evaluation is to measure features of the usability of an application. In this regard, the purpose of usability evaluation is to identify usability problems of an application and redesign and improve the application built upon user preference and user task performance [83]. Underlying usability evaluation, this section presents three evaluation criteria: the structural mediator for the game-based approach and a self-explanatory VR application (Section 4.3.1), the functional mediator for various stages of information seeking and sensemaking activities in an immersive environment (Section 4.3.2), and being-in-the-world for expansion of an intellectual and experiential horizon of users by the VR application (Section 4.3.3).

4.3.1. Structural Mediator

For a smooth transition from reality to a virtual world, user immersion in a virtual world and user tasks, and a self-explanatory VR application, this research has designed the structural mediator reified in the form of a world with a theme. A world with a theme constructs a virtual world in a manner that a user can feel comfortable with, familiar with, and exciting about a virtual world. Underlying the design decision, the architecture enables eliciting user immersion in a virtual world and delivering the context of a virtual world to a user effectively. This design decision becomes salient for novice users. A world with a theme provides enough time for a transition from reality to a virtual world, information about a virtual world, and psychological immersion in a virtual world to a user before a user embarks on exploratory data analysis. For experienced users, the virtual world gives discretion as to skipping a world with a theme by using a button taking a user to a second scene directly.

Considering the purpose of this VR application (i.e., exploration of 3D spatial data from reservoir simulation), a spaceship exploration mission on a reservoir of a new planet is selected as a world with a theme. When entering a virtual world, the first thing that a user encounters is a text block saying “Please go and talk to the robot by pressing ‘A’ button”, and the exploration mission begins. This text block is a precaution to prevent user non-linear behavior from failing to inform a user of information about this scene from the robot. In this setup, all the participants started with the same condition. Built upon this controlled variable, this section represents the usability evaluation of the structural mediator (i.e., game-based approach and self-explanatory VR environment) in terms of various aspects of the user preference.

4.3.1.1. Game-based Approach

Reified in the first and second scenes, the structural mediator is designed for mediating between a virtual world and a user. This first scene shapes a virtual world by utilizing a world with a theme (i.e., spaceship exploration mission on a reservoir of a new planet) as one of elements of the game-based approach. The first scene consists of a station with the background of a new planet, and a spaceship, a robot for information delivery, and a start button for teleportation to the second scene. Underlying this theme, the second scene (i.e., reservoir exploration scene) materializes the aspects of the world with the theme for the consistency of the world and successful user interaction with the virtual world. For example, the interactive tools of the second scene represent real aspects of touch screens of the theme for goal-oriented interaction (described in Section 4.3.2) to operate efficiently. This structural mediator of the second scene constructs the logically understandable virtual world for a user to experience, explore, and interact.

The anticipation was that the game-based approach materialized in the first and second scenes enables a smooth transition from reality to a virtual world especially for novice users, which leads to user immersion in a virtual world and use tasks and aids in goal-oriented interaction in the second scene. In this respect, this research describes the user evaluation of the usability of the game-based approach, its applicability to engineering and scientific problems, and its possibility for improvement in task performance.

For the usability of the game-based approach of the first and second scenes, the most frequent (73.2%) and relevant (97.6%) words from the questionnaire were “interesting way” and “new method”. All the participants said that the game-based approach is engaging, simple, and intuitive to work with a VR application for research. From this perspective, the participants reached consensus on a point that for those who are not familiar with AR/VR technology, this approach provides a comfortable and entertaining environment due to the tutorials on the VR demo and the user-friendly design of the virtual artifacts, such as the dialogue panels, buttons, text blocks, which enables a user to intuitively utilize and control them. In this respect, a participant mentioned that “I think the use of real-world analogues is a good idea for people that are not that familiar with AR/VR" and "I had an impression that I was doing something fun. I thought as if I was waiting to play a game".

For the question that how do the elements of the game-based approach in the first scene affect users in terms of performing exploration, interaction, and tasks in the second scene?, the overall reaction was that since from the experience of the first scene, they could expect a similar theme and interactions in the second scene, the consistency made the second scene understandable, intuitive, and familiar to explore and interact with 3D spatial data. This consistency elicited user immersion in the second scene and helped users focus on performing tasks and achieving the goal of the application. More specifically, the same robot in the first and second scenes provided clear instructions on each scene, facilitated the use of the application, and helped advance what will happen next. Furthermore, touch screens with a physical analogue in reality and the information about how to use the screens from the robot allowed a user to easily and quickly interact with the screens, which led to facilitating task performance.

In order to evaluate the applicability of the game-based approach to engineering and scientific problems and its possibility for improvement in task performance, the questionnaire asked each participant if they fall into a category: VR experience in both VR games and academic VR applications (case1), VR experience in either VR games or academic VR applications (case2), or no VR experience (case3). This question was designed to eliciting different opinions on this evaluation from different VR experiences. For the participants categorized into the case1 who have experience in both using VR games and developing and using academic VR applications, they all agreed that the game-based approach was easy to use, learn, and interact with VR demo, its tools, and data. A participant from the case1 mentioned that "this feature allows the user to focus on the tasks rather than fighting the application" and "this feature saves valuable and expensive time on tasks, and also reduces training costs and costs related to encouraging users to use the application".

From the above perspective, the insightful feedback was "In my experience with academic data visualization applications, the most successful approach is to take lessons learned from games". The participant added that "Games must be engaging and intuitive to be successful, as there is no other purpose to them than to be enjoyable to use. Utilizing various aspects of games to create the right "feel" and user experience is a crucial success factor for academic or scientific applications". In this respect, the participant stated that this VR application achieved the stability between the purpose of games and a scientific application.

On the other hand, there were arguments against the game-based approach regarding a world with a theme (i.e., the spaceship mission concept of the first scene) applied in engineering and science fields. One of the arguments was that the implementation of a world with a theme may need to be carefully balanced. In this respect, a participant mentioned that "For applications I am experienced with, I prefer to just drop into the data and have access to all of the tools right away. For applications I am a beginner, having a game-based introduction to the software feels like a simple and intuitive way to learn the different tools." Aligning with this argument, another participant commented that "Many older engineers or scientists may not find it "serious" enough. However, that is more of a cultural and social consideration than a functional one". The participant suggested on this issue that if a world with a theme is matched with the background of the data, the older engineers or scientists would be able to understand the simulated contextual environment. For example, if the source of the data is from an oil field in Alberta, the first scene of a world with a theme can consist of and simulate the background of the oil field.

For the case2 and the case3 who have a few VR game experiences or no VR experience, the overall opinion on the applicability of the game-based approach (i.e., a world with a theme materialized in the first scene) to engineering and scientific problems was that it was extremely helpful for novice VR users. Before embarking on engineering or scientific problems in the second scene, the first scene provided enough time, information, and entertainment with novice users to adjust and explore the VR environment. This preparation enabled the novice users to expect the second scene. Participants commented that "The first scene gave me interest in the second scene about how this VR demo would look like and what I can do in the second scene" and "I think the first scene was helpful in that I could be immersed in the virtual world and the impression of the first scene continued on the second scene, which helped me understand what will happen in the second scene and how to interact with the models".

4.3.1.2. Self-explanatory VR Environment

The purpose of the design of the self-explanatory VR environment is to support immersive analytics in exploration and interaction with 3D spatial data effectively by providing sufficient information about the VR application, such as the goal, data description, simulation input, and available tools to utilize. This design decision is ascribed to the fact that it is almost impossible for a designer to deliver all the detailed information before a user enters a virtual world. The anticipation was that the self-explanatory VR environment enables reducing costs and time on training for the VR application and eliciting autonomous action in the virtual world.

In this respect, this research describes the user evaluation of the usability of the self-explanatory VR environment in terms of exploration in the virtual world and interaction with the 3D spatial data.

The overall opinion on the self-explanatory VR environment was that since the virtual artifacts, such as a robot with a dialogue panel, touch screens, buttons, and dropdown menus, and a 2D graph, were intuitively designed, the participants could utilize the artifacts easily and expect what they can do with the artifacts and in the virtual world without any help from the designer. The important aspect here is that the artifacts enable users to explore the virtual world "autonomously". From my observations when petroleum engineers from different oil and gas companies who have no VR experience visited our lab in several times and experienced academic VR applications, although we informed the engineers of the VR applications in detail, they seemed confused about what to do with the applications and how to use the tools. While putting on the VR headset, after we added more explanations about the applications, the engineers tried to perform the tasks. However, the tasks that they performed were low-level interactions, such as moving around the VR environment and touching a virtual artifact.

Regarding the above problem, this research demonstrated the usefulness and effectiveness of the self-explanatory VR application for the case1, case2, and case3 users. The case1 user commented that "The virtual world does provide some common experience with the real world, making the experience more approachable. Other data visualization apps can bring a user into a visualization, however, users (especially novice users) can feel disoriented and unconnected to the application and how to use it. Providing a simulated contextual environment with real-world analogues makes the application much more self-explanatory". The case2 user mentioned that "I can find every information in the instruction and interact with the objects. I can do what the application wants me to do. It is quite easy for me to know what input and output parameters are". The case3 user said that "The explanation coming from the robot was helpful and it was good to have another place first before going into the simulation space because it allowed me some time to explore the virtual world in advance. In the simulation space, the guide of the robot was a great introduction to the simulation system. I could easily control those panels in the touch screen with its guide".

There was a mixed response to the dialogue panel. Compared with plain text of guidance, the dialogue panel with the conversation concept from the robot made novice users feel comfortable with the exploration. However, some participants preferred audio guidance to text guidance since they felt that reading text was demanding. Furthermore, they suggested that since the text was quite long for them, if some of the text was selective information, it would facilitate their exploration process in the virtual world. For example, instead of including all the text information in the dialogue panel, if text information was assigned to an artifact in the form of a pop-up panel explaining what users want to know about the artifact, the user could selectively choose information. In this respect, a participant suggested that if a user hovers over terminology or a unit, when a pop-up panel describes them briefly, this function would be helpful for the user to better understand the system of the application.

4.3.2. Functional Mediator

Built upon the user immersion in the virtual world elicited by the structural mediator, the functional mediator enables the users to disclose the virtual world that is logically understandable, perceivable, and accessible to the goal of the virtual world. The journey to the goal requires not only low-level but also high-level interactive, cognitive, and problem-solving tasks. In order to encourage and enable a user to perform the tasks and accomplish the goal, this research has designed goal-oriented interaction underlying the virtual world constructed by a world with a theme. Once a world with a theme shapes every artifact that simulates aspects of a world with a theme, which leads to user immersion, interactable artifacts (i.e., robots, text blocks, touch screens, and a recording ball) systematically open up a new dimension within the virtual world in which they autonomously explore, manipulate, interact with the virtual world in order to achieve the given goal.

The above cognitive mechanisms are initiated by the robot in the second scene informing a user of the goal, tasks, data, and interactive tools when a user approaches the robot automatically talking to the user. This foundation enables equipping the user with strategies on how to explore the data, utilize the tools, and perform the tasks for achieving the goal. Typical visual tasks for reservoir engineering are investigated through task analysis. The results of the task analysis are materialized in the form of text blocks, touch screens, and a recording ball to support various stages of interactions and information seeking and sensemaking activities in an immersive environment.

The text blocks provide the contextual information of the reservoir simulation, which is the simulation input to support investigating the cause-effect relationship between the static and dynamic 3D spatial data. The user can grab the blocks and put them in the space to read and extract information. Employing the extracted information and three touch screens (i.e., Static, Size, and Dynamic), the user can explore, manipulate, and control the 3D spatial data. The three screens consist of dropdown menus, a slider, and buttons. If the user points their virtual right hand to each screen, a line is automatically rendered. By utilizing the line, the user can select and manipulate the artifacts on each screen.

Collecting and organizing all the information, the user can build a hypothesis or assumption (i.e., a bottom-up approach). On the other hand, if the user has their own hypothesis or assumption, the user can validate or reject them by exploring and searching the artifacts for evidence (i.e., a top-down approach). The recording ball enables the user to record these series of low-level to high-level interactive, cognitive, and problem-solving activities, which are automatically stored in a file and shareable to other researchers.

This section represents the usability evaluation of the functional mediator in terms of user task performance. The user task performance is categorized into Foraging Loop and Sensemaking Loop in accordance with the Pirolli-Card model (detailed in Section 2.2.1). Each loop consists of a series of cognitive steps that describes how people create a model from data (i.e., i.e., a bottom-up approach) and evaluate the model from theory (i.e., a top-down approach). In this regard, this section describes various stages of user cognitive activities in an immersive environment when they perform exploratory data analysis in 3D spatial data of the second scene by utilizing the artifacts, which reify the results of the task analysis. Built upon the descriptions, the goal of this section is to evaluate how well this VR application supports low-level to high-level cognitive tasks in exploration of 3D spatial data.

4.3.2.1. Foraging Loop

After entering the second scene from the first scene by utilizing the start button, all the participants were informed about the same goal of the VR application by the dialogue box from the robot (i.e., the goal of the mission is to explore the models and the implications of the static model for the dynamic model by using the touch screens). Furthermore, the robot provided information about the 3D spatial data (i.e., static and dynamic models) and how to use the touch screens, the text blocks of the simulation input, and the recording ball with all the participants. Built upon this controlled variable, the VR environment put all the users on the same starting line with the same amount of the information. In this VR environment, the users began their exploration mission by collecting the information of the 3D spatial data.

Table 4.2 described the result of the foraging loop of each participant. The anticipation was that all the participants will initiate their exploratory data analysis in the 3Dspatial data by utilizing the text blocks of the simulation input in order to inform themselves of the system. However, a participant from the case 1 (subject

1), a participant from the case2 (subject 2), and a participant from the case 3 (subject 3) deviated from the instruction from the robot. They seemed confused about what to do, tried controlling and selecting the dropdown menus, the slider, and the buttons on the touch screens, and stopped experiencing the VR demo. In this case, they failed to gather the information from the text blocks. This absence of the basic information about the simulation system led to preventing the further exploration and investigation of the 3D spatial data. After the VR demo session, this study proceeded with the interview asking the users that “can you recall or understand the goal of the second scene?”. They could not reply to this question clearly. This obvious evidence indicated that if users can not grasp the goal of a VR application, it is impossible for the VR application to fulfil the purpose and guide them to high-level cognitive activities.

The other three participants initiated the foraging loop by reading the text blocks of the simulation input. This group showed similar behavior while utilizing the text blocks. They sent three to four minutes reading each text block and trying to focus and understand the simulation system. Based on the extracted information, they connected the information with the static model. For example, with a participant from the case 3 (subject 4) holding the text block of the porosity layer information, the subject 4 performed the comparison of the value of each layer and the color mapping in the static model to confirm how this information was matched with the visualization of the static model. Participants from the case 1 (subject 5 and 6) skimmed and put the text blocks next to the static touch screen and interacted with and controlled the screen by utilizing the dropdown menu.

The above behavior can fall into part of the schematizing process in which a user improves the understanding of the system, refines extracted information, and connects the information to build potential scenarios. This process is the cornerstone that enables a user to perform sensemaking. Regarding the mechanism that a user collects information about the simulation system with the text blocks, the subject 6 suggested the improvement of the interaction with the text blocks. The participant commented that “To reduce the work of the individual, the text blocks with information could be pre-arranged in a way that key pieces of information are visible, and the user does not need to sort through them. Perhaps configurable pads that are iPad analogues could be used that can display whatever data the user wants could be used”.

Participant	Understand the goal?	Read the text blocks?	VR Experience / Specialization
Subject 1	X	X	VR Games & Academic VR applications / Ph.D. Petroleum Eng Student
Subject 2	X	X	VR Games / Ph.D. Petroleum Eng Student
Subject 3	X	X	X / M.Sc. Petroleum Eng Student
Subject 4	O	O	X / M.Sc. Geophysics
Subject 5	O	O	VR Games & Academic VR applications / Ph.D. Petroleum Eng Student
Subject 6	O	O	VR Games & Academic VR applications / Oil and Gas Technology

Table 4.2: Result of the foraging loop of each participant.

4.3.2.2. Sensemaking Loop

Built upon the foraged information, the subject 4, subject 5, and subject 6 embarked on exploratory data analysis in 3D spatial data. By interacting with the touch screens, they started to know what available options the screens have and how they can manipulate and control the 3D spatial data from the screens. While exploring the screens, they experimentally selected different reservoir parameters from the static and dynamic screens, adjusted the size of the dynamic model, and played the time-series visualization and the 2D graph. With these experiences and the foraged information, they began to develop strategies on how to construct their thought process in order to build a hypothesis or an assumption, which potentially leads to knowledge discovery. Before discovering new information or knowledge about the relationship between the static and dynamic models, the precedent stage of the cognitive activity is to design thought processes and formulate a hypothesis or an assumption to validate based on the collected information. This section describes this sensemaking process of the subjects in the virtual environment.

Table 4.3 describes the result of the sensemaking loop of each participant. In accordance with the cognitive stages of the Pirolli and Card model, regarding the transition from *Schematize* to *Hypotheses*, while subject 6 could not explicitly build a hypothesis, the subject 4 and the subject 5 developed a hypothesis and tried to validate it while exploring the 3D spatial data. This research investigated the thought process of formulating a hypothesis of each subject and whether they validated the hypothesis. The most interesting aspect of the cognitive transition of the subject 4 was that although the subject did not have both VR experience in academic VR applications and VR games and background in petroleum engineering, the subject reached the closest cognitive stage of validating a hypothesis. The thought process of the subject (hypothesis) was to connect the three different permeability patterns (i.e., conservative, intermediate, and liberal numbers of clusters) with the dynamic time pattern of the oil pressure in order to find hidden patterns in the dynamic model. There were two factors that this VR application failed to support the implementation of the idea and further investigation.

Firstly, the subject could not understand what the unit of the 2D graph and terminology means. The 2D graph as the key information of the simulation result and its terminology show the oil, gas, and water productions over time. However, the subject could not take advantage of this information to perform in-depth investigation. There was a suggestion about this issue that if a user hovers over the unit, the small pop-up window explains the meaning of the unit. Secondly, the subject could not recognize the time changes in the dynamic model due to the lack of a function that supports changing multiple angles. The angles where the subject could see the changes in the dynamic model were too narrow to notice the changes. It would improve user task performance if the VR application supported a function that enables a user to adjust the location and rotation of the static and dynamic models. This function will provide different aspects of the two models in different angles, allowing a user to interact with the models easily.

In the case of the subject 6, the subject was clearly informed of the goal of the VR application, did collect the basic information about the simulation system from the text blocks, and interacted with the touch screens effectively. However, there was no further cognitive transition to build a hypothesis. This phenomenon can be ascribed to a degree of preference of the game-based approach (i.e., a world with a theme). In the questionnaire and interview with the subject, the subject preferred to go straight and skip the first scene and directly interact with the data in the second scene. Considering the user preference of the game-based approach illustrated by the questionnaire and interview with the subject, there is a possibility that the subject was not deeply immersed in the virtual world enough to activate the cognitive and perceptual processes. This preference can potentially lead to a huge difference in terms of user performance when compared to the subject 5.

Both the subject 5 and the subject 6 fall into the case 1 that they both have experience in development of academic VR applications in petroleum engineering and VR games. During the VR demo session, the subject 5 enjoyed exploring the first scene of the world with the theme and spent enough time to systematically understand the virtual world by following the guidance of the robot. This tenacious and ceaseless concentration provided by the world with the theme helped the subject get engaged in the virtual world and autonomously open up and explore the virtual world to achieve the goal. This series of the experience from the first scene could stimulate the cognitive process, enabling the subject 5 to build and validate a hypothesis. This different result from the subject 5 and 6 indicates that systematic and continuous user immersion that guides users from low-level to high-level cognitive activities is a key for a successful VR application.

The hypothesis that the subject 5 developed during the exploration in the 3D spatial data was that the clustered areas of the permeability in the static model would have an impact on the oil saturation in the dynamic model. The subject described the cognitive process of verifying this hypothesis in the questionnaire: "I laid out the quantitative data from the text blocks to get an idea of the numbers. Then I displayed the conservative clustering of the static permeability data, and then played the time series simulation of the oil saturation data and looked at the three different clustering methods". By implementing the idea, the subject tried to discover new information about the relationship between the clustered permeability data and the oil saturation patterns. The subject concluded that the clusters did not show anything predictive about the oil saturation.

There were three factors that underlay the logic of the above decision making: color mapping, limited angles, and a lack of an analysis feature. From a data visualization perspective, the subject supposed that the color map that was utilized in this VR application for scientific visualization failed to represent some scalar values to colors. This obscurity could not reveal the hidden patterns of the oil saturation. Regarding the color map issue, the subject described the thought processes when performing the sensemaking. "I was left wondering what the color encoding is, and if there needed to be a transfer function between the data values and coloring, because maybe some of the relationships are hidden by how the data is colored within intervals in the data. I would make sure the color mapping is clearly defined, and also what the function is, to ensure that the coloring varies appropriately to match the data distribution." Regarding this issue, this research adopted a diverging color map that performs well in scientific visualization applications [86].

In addition to the color mapping, limited angles of views from the touch screen prevented the subject from seeing the changes in the dynamic model. On this issue, the subject suggested a function on a touch screen that enables a user to adjust, view, and slice the models in different angles. Furthermore, instead of walking around the models to see the changes in a property, the subject suggested that teleportation can facilitate exploratory data analysis. Finally, this VR application failed to support the subject to efficiently see the interior structure and the cells of the models. The subject had to approach the models and see the interior cells. Regarding this functional issue, the VR application can be improved by implementing an x-ray shader button that enables a user to see the inside of the models. The three factors were the main flaws that failed to support the sensemaking processes of the subject. The suggested features will be included in the VR application for the future work.

Participant	Total Exploration Time	Build a hypothesis?	Validate the hypothesis?
Subject 4	More than 10 minutes	O	X
Subject 5	5 minutes	X	X
Subject 6	More than 10 minutes	O	O

Table 4.3: Result of the sensemaking loop of each participant.

4.3.3. Being in the World

As the core framework of this research, the ultimate goal of being-in-the-world is to augment and expand knowledge and experience of users in real life by mediating between a virtual world, users, and the world. Utilizing the structural and function mediators, being-in-the-world provides a harmonious and coexisting virtual environment in which users can feel comfortable, connected, and immersed in the virtual environment. This user immersion elicits autonomous behavior to explore the virtual environment and activates cognitive activities to focus on a goal and tasks provided by the virtual environment. The cognitive activities lead users to discover new knowledge and information by understanding, constructing, and organizing the virtual world with their own private perspectives. This experience from the virtual world can extend a world of each individual user within the world in different ways. This section investigates how this philosophical framework of the VR application discloses an intellectual and experiential horizon of users.

From an intellectual horizon perspective, participants expanded their realm of knowledge of VR technology to its utility for research in two aspects: user experience and task performance. Built upon the philosophical framework reified in the form of the game-based approach, this VR application materialized the input and output of the black oil simulator and supported visual analysis in the reservoir models. Compared to desktop-based interfaces, participants have recognized the utility of VR technology for research in terms of how VR technology economically saves training costs and valuable and expensive time on tasks, vividly simulates aspects of reality, and efficiently connects them with a virtual environment. This utility led to improvement in user experience and user task performance. In this respect, a user commented that "This VR application expands my knowledge and experiences, because it is a nice introduction to how a good game-influenced user experience can provide a much more accessible and engaging tool for serious work". This utility of VR technology helped users understand data easily and focus on user tasks effectively.

From an experiential horizon perspective, participants expanded their sensory experience and imaginative capability through this VR application. What made these kinds of the expansion possible is ascribed to the nature of data from oil and gas fields. Since the data is a variety of different samples collected from the subsurface, which makes the data abstract, complicated, and uncertain, visualizing the data in an immersive environment enables users to interact with the data intuitively. In this respect, a participant commented that "The experiments in the oil field, especially for the subsurface experiments, are implemented usually by simulations in computers, which sometimes give experimenters impressions of obscureness. In this VR experiment, I could have the feeling that I would get from the chemical experiments. It is always easier for me to remember what I experiment with my own hands. Also, I felt I had more controls over the simulation and I could explore the results amply, which was followed by the confidence over the experiment". The VR environment augmented user their sensory experience and improved their task performance by converting the abstract data to the intuitive and tangible artifacts and utilizing easily the manipulated interactive tools.

Aligning with the above perspective, another participant could enhance an ability to mentally construct a concrete image from abstractness. The participant described that "It was a very interesting new experience to see and interact with the reservoir data in a fully immersive 3D environment. In this context, it definitely expanded my knowledge and experience of what is possible to do with virtual environments. It helped me mentally visualize the data in a 3D way and going forward it may help with how I mentally picture the data I am working with". The VR environment encouraged using the imaginative capability to understand the complicated and intangible 3D data. This experience from the VR environment potentially contributed to expanding the being-in-the-world and the real life in terms of an imaginative capability.

4.4. Discussion

This research designed and implemented the interactive exploration pipeline underlying the philosophical framework in which the VR application supports exploratory data analysis in complex 3D spatial data and sensemaking without any training and instructions as to the VR application. To discover, identify, and provide new possibilities of applying Heidegger's philosophy to human-computer interaction in an immersive environment, the exploratory evaluation of the pipeline was conducted in the three categories: the structural mediator (i.e., game-based approach and self-explanatory environment), the functional mediator (i.e., foraging loop and sensemaking loop), and being-in-the-world (whether this VR application expands experience and knowledge of users in real life).

When a user enters the virtual world, the structural mediator makes the user feel comfortable and provides the tutorials about the VR environment, leading to user immersion. The user immersion enables eliciting user autonomous behavior to explore the VR environment. The functional mediator supports interaction with the artifacts for what the user tries to fulfil. The interaction opens up a new dimension within the virtual world in which the user can understand, discover, and manipulate the new dimension with their own private perspective. With the user immersed in their dimension, to achieve the goal of the VR application, the user deeply focuses on their cognitive tasks in the form of foraging and sensemaking loops. Depending on what kind of a new dimension users disclosed and which stages of cognitive tasks the users performed, the experiences lead to different being-in-the-world in real life. In this respect, what initiates the difference is the functional mediator.

In the functional mediator environment where all the users were immersed in their dimension and informed about the goal, data, tasks, and available tools through the robot, all the users initiated their exploration of the complex and large 3D spatial data with the same amount of the information. Considering differences in academic background, VR experience, and domain knowledge of users, each user showed different behavior and thought processes and developed different strategies to achieve the goal of the VR application. Examining these observations and the qualitative data of each user, this section suggests two factors that failed to guide users to high stages of cognitive tasks, such as *Schematize* and *Build Case* in the Pirolli and Card model. Since the Section 4.3.2.2 analyzed the factors that the VR application functionally failed to support the subject 4 and the subject 6 for further sensemaking processes, this section illustrates why the subjects 1, 2, 3, and 5 stayed the foraging loop and the transitional point to *Schematize* based on the two factors: an amount of information and effort.

1. Amount of Information

Since the VR application visualized the complicated and large 3D spatial data resulting from the in-house black oil simulation, this research adopted UI design (i.e., readability of text, interactivity, and intuitiveness) detailed in Section 3.5.1.2 to relieve the burden of the amount of the information users will digest. For readability of text, one paragraph consists of three sentences in the dialogue panel, and the panel is designed to represent information as succinctly as possible. For interactivity and intuitiveness, when users approach the robot, the dialogue panel automatically pops up and talks to users, and users can skip the sentences in the panel when they press a button on their controller. This VR application guides users by providing tutorials from the robot in the form of conversation. By reading the contents, users are informed about the entire information of the virtual world.

Compared to the subjects 4, 5, and 6, in the case of the subjects 1, 2, and 3, the total time spent exploring the VR application was approximately less than 5 minutes. Furthermore, according to the questionnaire and interview, the subjects thought that the contents were too long to read. Based on the observation and the results of the questionnaire and interview, the subjects 1, 2, and 3 spent less time focusing on reading and understanding the contents from the robot. In this regard, in the case that a VR application provides effective analysis features to explore a virtual world, if users do not focus on understanding what a VR application offers, before users utilize analysis functions, they will be lost in the virtual world and do not know what they can do. Without users understanding the system, the VR application fails to guide users to support the foraging and sensemaking loop and focus their tasks.

Despite the UI design, for the subject 1, 2, and 3, the text length made the subjects lose their interest in the VR application. In this respect, the VR application can be improved by shortening the amount of the information and utilizing an audio function instead of text information. However, most VR applications in the fields of science and engineering represent complex, large, and abstract systems of phenomena. Furthermore, in this research, the self-explanatory VR environment provides all the information to users without a training session about the VR application. These characteristics require a lot of information to understand for interacting with the VR application and performing cognitive tasks.

Considering the above characteristics, even if the VR application reduces the information as little as possible, a certain amount of information is necessary for users to understand the system of the self-explanatory VR application. It is important to provide an adequate amount of information to users through a balance between audio and text information. However, before experiencing VR applications, users have to be prepared for dealing with complex and abstract scientific and engineering subjects. This is because adopting the game-based approach for a research purpose, VR applications might be considered as a VR game. The preparation from participants will help them more focus on VR applications and improve task performance when performing high level of problem-solving and cognitive tasks.

2. Effort

Considering the X axis of the Pirolli and Card model is Effort, it is obvious but easily neglected to recognize effort as a critical factor of a VR application. If a VR application implements useful and efficient analysis features, effort can have a huge impact on determining destiny of a VR application. In order to help users take less effort to understand and digest the amount of the information and immerse themselves in a virtual world, this research designed and implemented the UIs and systematic immersion in a user-friendly fashion. The structural mediator in the form of the game-based approach elicits a comfortable, entertaining, engaging virtual world by designing real-world analogue, a friendly robot with tutorials talking to users, and storyline of a virtual world (i.e., a world with a theme). These elements can lead to the systemic immersion.

Guiding users to systemic immersion requires a little effort. It is easy for users to follow the system and get immersed in the virtual world. However, cognitive tasks require an enormous effort for users to think. In accordance with the questionnaire, interview, and observation, although the subject 1, 2, and 3 were immersed in the virtual world, they did not understand the goal of the VR application due to the amount of information. The absence of understanding the basic information led to preventing the subjects from performing the further cognitive tasks. Compared to the subjects 4 and 6, the subject 5 relatively spent less time performing cognitive tasks and stayed the transitional point to *Schematize*. The criterion of effort can indicate spent time during the exploration in the second scene. The subjects 4 and 6 approximately spent over 10 minutes understating the system, collecting data, and performing sensemaking. The subject 5 approximately spent 5 minutes to process all the information and perform sensemaking.

There might have been functional flaws and malfunction of the VR application when the subjects 1, 2, and 3 stayed in the foraging loop, and the subject 5 was at the transitional point to *Schematize*. Even considering these issues, based on the spent time of the subjects 1, 2, and 3 (approximately less than 5 minutes), the subject 5 (approximately 5 minutes), and the subjects 4 and 6 (approximately over 10 minutes) in the same starting line with the same amount of information, the researcher could confirm the Pirolli and

Card model that the level of cognitive tasks is proportional to effort. In addition, what makes the confirmation more robust is the subject 1 who had both VR experience in academic VR applications and VR games and background in petroleum engineering and the subject 4 who did not have both VR experience in academic VR applications and VR games and background in petroleum engineering. With the above conditions, the subject 1 spent less than 5 minutes exploring the virtual world, while the subject 4 spent over 10 minutes understanding the VR system and building a hypothesis. Thus, on the premise that VR applications support systematic immersion and useful and efficient analysis functions for domain specific research, effort can be a significant factor of a VR application.

Chapter 5

5. Future Work and Conclusions

5.1. Future Work

To improve user experience and usability of this VR application and provide better support for sensemaking processes, this section discusses some directions for future work. This section represents three directions: VR application features based on user feedback, a quantitative study for the evaluation of the VR application, and machine learning applied in a visualization pipeline to optimize the interactive exploration pipeline.

5.1.1. VR application Features

There were numerous future directions suggested by the users for the VR application features after the users interacted with the application. This section presents three critical feedbacks that will support sensemaking processes constructively in a virtual world: teleportation, controlling the views of the models at the platform, and data interaction with the text blocks.

The most frequent suggestion was to adjust angles of the static and dynamic models when the users explores the second scene. The VR application was designed for users to freely move around the second scene by employing the joysticks of the controllers, but most users did not want to move around the scene. They preferred to move by teleporting or controlling the views of the models while staying at the platform of the second scene. For teleportation, since the Oculus Integration in the unity asset store provides the teleportation function, it can be a simple solution to implement it in the VR application. This function allows users to point at a spot where they want to teleport and see the part of the models that they want to investigate.

For controlling the views of the models at the platform, we received a number of comments. When designing the VR application, we expected that moving their virtual body to see the models closely would provide a better user experience than teleportation and manipulation of the views of the models at the platform. However, this design decision was unconstructive and prevented users from performing further sensemaking processes. The angles where the subject could see the changes in the dynamic property was too narrow to notice the changes. A potential approach to this problem is to allow users to adjust and manipulate the location and rotation of the static and dynamic models on a touch screen with a simple slider. This function will provide different aspects of the two models in different angles, allowing a user to interact with the models easily and facilitating exploratory data analysis.

Furthermore, the above problem of moving their virtual body led to a new problem. Since users did not want to move their virtual body to see the models, this phenomenon made them stay at the platform and lose their will to investigate the inside of the models. This opportunity cost resulting from the unconstructive design decision can be potentially huge in this research since the interior structure and the cells of the models can indicate important implications. Regarding this issue, the VR application can be improved by implementing an x-ray shader button that enables a user to zoom in and out the inside of the models at the platform.

When it comes to the data interaction with the text blocks, the text block caused ineffectiveness. In the VR application, users had to grab each text block to see and read the black oil simulation input. This interaction type did not give users an opportunity to be selective about information that they want to read and utilize. In order to collect the information, users had to grab all the text blocks and spread and put them in the space to see which information will be useful. Although the space itself can inherently support for document clustering to group similar information in clusters or to place important information in order [87], this problem can prevent users from performing information gathering and potentially making a transition from the foraging loop to the sensemaking loop. On this issue, a user suggested an iPad analogue artifact which contains the simulation input in the one artifact and helps users to browse, select, and retrieve information easily and efficiently.

5.1.2. Quantitative Study

We conducted the usability testing to evaluate the VR application based on the qualitative study (i.e., the questionnaire, interview, and observation). When Section 4.4 discussed Effort, we presented that user task performance can be proportional to effort regardless of the familiarity with domain knowledge and VR technology built upon the qualitative study. To statistically confirm or reject the hypothesis and make a prediction of the relationship, a quantitative study can be conducted by utilizing multiple regression analysis.

We can assume that there is a linear relationship between performance of sensemaking activities and the familiarity with domain knowledge and VR technology. In this regard, a dependent variable is performance of sensemaking activities (y_i). Independent variables are the familiarity with domain knowledge ($X_{i,1}$), and VR technology ($X_{i,2}$). Error (e_i) is nonlinearity of user behavior in a virtual world. With the multiple independent variables, the multiple regression model is $y_i = \alpha + \beta_1 X_{i,1} + \beta_2 X_{i,2} + e_i$, for $i = 1, 2, 3, \ldots, n$. The increment of the i value refers an increase in the familiarity of each independent variable.

The controlled variables is all the participants are informed about and understand the same goal of the VR application by the dialogue box from the robot (i.e., the goal of the mission is to explore the models and the implications of the static model for the dynamic model by using the touch screens) while performing exploratory data analysis in complex 3D spatial data in the virtual world. We can estimate the score of user task performance (y_i) on a scale of 1 to 10 in accordance with the Pirolli and Card model. Furthermore, we can estimate the scores of domain knowledge ($X_{i,1}$), and VR technology ($X_{i,2}$) by utilizing a questionnaire, interview, and observation or asking users to specify the scores on a scale of 1 to 10. Utilizing this method, we can estimate the variables of the multiple regression model.

With the above data, we can build a multiple regression model and determine whether there is a significant relationship between user task performance (y_i) and the familiarity with domain knowledge ($X_{i,1}$), and VR technology ($X_{i,2}$). Since there are the two independent variables, the null and alternative hypothesis can be set up. If there is no a liner relationship between user task performance (y_i) and the familiarity with domain knowledge ($X_{i,1}$), and VR technology ($X_{i,2}$), the null hypothesis is H_0: $\beta_1 = \beta_2$. If there is a liner relationship between user task performance (y_i) and the familiarity with domain knowledge ($X_{i,1}$), and VR technology ($X_{i,2}$), the alternative hypothesis is $H_1\text{: At leat one } \beta_j \neq 0$. In the case of the null hypothesis, we can confirm that user task performance can be proportional to effort regardless of the familiarity with domain knowledge and VR technology. In the case of the alternative hypothesis, we can determine the contribution of an independent variable to the multiple regression model. This quantitative study will increase the confidence of the hypothesis of this research.

5.1.3. Machine Learning Applied in a Visualization Pipeline

In this research, machine learning (ML) included in the interactive exploration pipeline was utilized for the visual representation of the permeability data. The ML technique simplified the large and complex 3D spatial data and supported users for pattern recognition. Furthermore, to help users interact with the ML-assisted visualization and assist exploratory data analysis, this VR application provided user discretion to choose the three different options of the permeability patterns (i.e., conservative, intermediate, and liberal) depending on the user judgement of the data. This user experience design encouraged users to explore the patterns easily without prior knowledge about the principle and input parameters of the clustering algorithm (HDBSCAN) and involve the process of pattern discovery.

In addition to application of ML techniques in visual representations and human-computer interaction, ML-assisted visualization can be applied in six processes of a visualization pipeline: data processing, data presentation, insight communication, style imitation, user interaction, and user perception [88]. From a visualization-centered design perspective, which is how ML techniques can solve visualization-related problems, this research paper summarizes the utility of ML techniques in each visualization process. Built upon this utility, the paper includes the six visualization processes in and make two modifications from an existing typical visualization pipeline. One of the modifications is “style”. *Style* from style imitation refers to aesthetically generate color of a visualization for memorability and engagement of a visualization. Utilizing ML techniques, style imitation automatically generates high-quality color ramps. The style imitation can be applied in this VR application for future work to optimize the color ramps, remove the obscurity of the color map, and increase information communication.

5.2. Conclusions

As the era of big data emerges with digitalized devices, innovative data storage systems, and instant access to data, data analysis technologies have been improving to process big data and find optimal solutions. In this era, decisions should be made based on a data-first, data-guided, and data-driven process. Although the technologies can provide innovative solutions, it is human beings that utilize them to analyze, interpret, and interact with data. The deficiency of human aspects in technology can conceal its potential for human beings to discover and employ. To reconcile technology with human beings, this research adopted the philosophy of technology as the foundation of the framework, mediating between technology, human beings, and complex domain big data.

The technology utilized in this research was VR technology to effectively visualize, deal with, and interact with multivariate, multiscale 3D spatial data manifesting high dimensionality and heterogeneity. VR technology can deliver better user experience and provide an immersive decision-making environment to the domain data compared to desktop-based interfaces. Utilizing these advantages, the VR environment in this research materialized a VR information space which has a relationship between object space (3D spatial data), attribute space (its properties), data structure space (clustering algorithm), and user experience space. Creating such a VR environment indicated creating a virtual world. In this regard, this research had to design and build a VR environment from the ground up. The philosophy of technology served not only as a mediator for users to utilize a VR information space efficiently but also as a guideline of human-centered design of the space. To create such an immersive environment, this research investigated the design, implementation, and evaluation of the interactive exploration pipeline.

Materializing black oil simulation, the VR information space supported visual analysis on 3D spatial data and sensemaking tools for discovering new information. To assist the low-level to the high-level problem-solving, cognitive processes in the virtual world, the philosophical framework was reified in the form of the structural and functional mediators. The structural mediator provided users with all necessary information about the virtual world and helped them feel comfortable with and get immersed in the virtual world to elicit autonomous user behavior. Built upon the structural mediator, the functional mediator encouraged the users to disclose a new dimension within the virtual world with their own private perspective by investigating and supporting domain-specific visual tasks. Evaluating the abovementioned usability of the VR application (i.e., Section 4.3.1 and Section 4.3.2) is important, but the more important aspect is whether this VR application actually affects real life in terms of expanding an intellectual and experiential horizon of users (i.e., Section 4.3.3). Shedding light on the usability and the expansion of being-in-the-world, this section represents the main contributions of this thesis:

- A philosophical framework that mediates between technology, human beings, and data.

  In Chapter 1, we stated the main problem of this research, deficiency of human aspects in technology, which makes it difficult for human beings to not only utilize technology but also analyze, interpret, and interact with data. To deal with this problem, we adopted the philosophy of technology and embraced Martin Heidegger's philosophy as the philosophical framework of this research. With the framework established, in Chapter 2, we designed each element of the interactive exploration pipeline by starting with investigating the foundational element of research and technology (i.e., data) as the first element of the pipeline. Built upon the results from the investigation of the data, each element of the pipeline was characterized to design a successful and robust visualization pipeline. Utilizing postphenomenology and ontology, the philosophical framework mediated between domain data (initial element), VR technology (middle element), and human beings (final element). Underlying the framework, the reified virtual world materialized the whole element of the interactive exploration pipeline to support sensemaking and decision-making processes. In the final stage of the pipeline, the experience from the virtual world expanded an intellectual and experiential horizon of users.

- A self-explanatory VR application system in the fields of science and engineering.

  In Chapter 2 and Chapter 3, we illustrated that in the oil and gas industry, most engineers are new to AR/VR interfaces and deal with multivariate, multiscale 3D spatial data, which requires comprehensive information processing skills. In this regard, it can be overwhelming for the engineers to perform tasks in a VR environment materializing the complex data and black oil simulation mechanism. This combination of the complexity makes the designers inform users of a VR interface, description of data, user tasks, and analysis tools verbally before the engineers enter a VR environment. To narrow the gap between the new VR interface and the complex nature of domain data and elicit maximal VR user task performance, we designed the self-explanatory VR environment for users to autonomously explore. From an interaction design perspective to design such a self-explanatory VR environment, we adopted the game-based approach to deal with the abovementioned problems. The VR environment sufficiently provided users with the contextual information, such as where they are, what available tools they have in order to interact with the environment, and why they are here, and encouraged them to develop strategies for performing problem-solving, cognitive tasks.

- An immersive human-centered VR environment for problem-solving and cognitive tasks.

In Chapter 2 and Chapter 3, to guide users from reality to a VR environment smoothly and efficiently, this research first delineated how humans perceive and interact with the world, disclose a new dimension within the world, and extend their cognitive span of the world by utilizing post-phenomenology, ontology, and being-in-the-world. This thesis demonstrated that this cognitive structure applied in the world is aligned with a virtual world. Built upon this philosophical framework, this research designed a virtual world by practically reifying postphenomenology and ontology in the form of a world with a theme (i.e., structural mediator) and goal-oriented interaction (i.e., functional mediator). In order to help immerse users in the virtual world and support sensemaking tasks, this thesis systematically described low-level to high-level problem-solving, cognitive processes of humans by utilizing the Pirolli and Card model and conducting task analysis in reservoir engineering. This foundation enabled designing, implementing, and evaluating the immersive human-centered VR environment for problem-solving and cognitive tasks.

In Chapter 4, to discover, identify, and provide new possibilities of applying the philosophical framework to human-computer interaction in an immersive environment, this thesis conducted the exploratory evaluation of the VR application in three categories through a questionnaire, interview, and observation. These categories were the structural mediator (i.e., game-based approach and self-explanatory environment), the functional mediator (i.e., foraging loop and sensemaking loop), and being-in-the-world (whether this VR application expands experience and knowledge of users in real life). Each category elaborated on cognitive, perceptual, and problem-solving processes of participants while performing exploratory data analysis in a virtual world. Finally, we investigated how this VR application helped users expand their intellectual and experiential horizon in real life.

5.2.1. Closing Remarks

Technology has become a ubiquitous tool in our life opening up possibilities to perceive how the world works and experience our existence richly. As technology evolves rapidly and becomes complicated and sophisticated, it has been difficult for humans to understand and utilize. This imbalanced phase between development of technology and the alienation of humans from technology is shrinking the possibilities rather than expanding. In this regard, for a solution to this problem, I was inspired by Martin Heidegger's philosophy explaining a relationship between human beings and the world as a structure of being-in-the-world. From this Heidegger perspective, each entity is inseparable and mutual while René Descartes described the relationship as an independent entity with his "cogito, ergo sum" ("I think, therefore I am"). To transform technology into a mediator of the relationship, which helps overcome difficulties in the fields of science and engineering, I adopted the philosophy of technology in this research.

This thesis has particularly focused on how VR technology can mediate between a virtual world, human beings, and domain data and support opening up possibilities for users to discover knowledge from domain data. There are two main contributions of this thesis. This thesis concretely delineates and pragmatically reifies Heidegger's philosophy in the form of a self-explanatory VR environment for research by utilizing the game-based approach. This thesis demonstrates that underlying the human-centered design, a self-explanatory VR environment provides support for users to effectively perform low-level to high-level problem-solving, cognitive tasks and expand an intellectual and experiential horizon of users in real life. In Chapter 4, this thesis shows the utility of the philosophical framework by illustrating that the way humans understand, interact with, and disclose the world is aligned with the way in a virtual world. This thesis opens a possibility of pragmatic application of philosophy on science and engineering.

While the philosophical framework applied in the VR application was utilized in petroleum engineering, I believe the philosophical framework to be broadly applicable in various fields of study in terms of building a self-explanatory and human-centered VR environment for problem-solving and cognitive tasks. I expect that this thesis can disclose future exploration of how VR technology and ML techniques can help human beings disclose their potential and how the game-based approach (e.g. real-world analogues) can provide systematic user immersion and improve user task performance. At the time when VR technology is growing with infinite potential, I hope that this thesis serves as a bridge between VR technology and human beings.

Appendix A

Recruitment Letter

Virtual Exploration through

a Philosophical Tool

Hello,

I am a master's student in the Department of Chemical and Petroleum Engineering at the University of Calgary. We are currently developing a virtual reality (VR) application that implements a variety of techniques to interact with subsurface data in order to analyze this data. The project title is "Virtual Exploration through a Philosophical Tool."

To improve the user experience in a VR environment, we have designed a VR environment built upon metaphysics, postphenomenology, and ontology. In this VR environment, utilizing virtual interfaces, you will perform typical visualization tasks, such as navigation, refinement, and elaboration of 3D spatial data, by manipulating, controlling, and adjusting virtual objects. We would like to invite you to take part in the design and evaluation of our prototype and share your thoughts, critiques and suggestions on how we could improve interactions, and the user experience of the application.

Sessions will begin by introducing the context of the VR design and its tools to be used. This will be followed by a session interacting with our prototype, followed by a post-task interview. Sessions are expected to take about 1 hour. With your consent, we may photograph and/or videotape the sessions, for later analysis (e.g., for publication, presentation, etc.) according to your preferences indicated in the consent form. A $10 cash gift will be provided as a token of appreciation for your time.

Please reply to ***dohyun.kwon@ucalgary.ca*** *or call* ***587-436-1472*** *letting us know whether you are willing to volunteer to participate on this research. We will do our best to accommodate your availability and welcome any questions or concerns you may have regarding the study. We have COVID-19 protocols in place to ensure your safety.*

This study has been approved by the University of Calgary Conjoint Faculties Research Ethics Board (REB20-1417).

Questionnaire

Post-task Interview

Reviewer Name	
Subject Background/Specialization	
Date of interview	

Immersive visualization hardware such as virtual and augmented reality is currently being examined by a variety of industries for many different applications. These immersive technologies provide both a potential opportunity for more efficient exploration tools as well as potentially poor designs due to the relative lack of research involving interactions in immersive environments compared with more traditional technologies. Understanding how users perceive interactions with data, interfaces and virtual agents such as avatars is crucial to making optimal use of immersive technologies.

The following questions are intended to gather general feedback/comments from the participant and any suggestions he/she may have regarding improving our system.

The participant may ask to clarify any question and more explanation will be provided, and he/she can decline to answer any question.

For your interest, this part may take up to 20 minutes.

Thank you very much for your time and effort.

Post-task questionnaire:

1- Please provide your impressions of the virtual world (e.g., spaceship mission concept of this study and virtual objects, such as a robot, dialogue panel, start button, touch screens on the second scene) you interacted with?

1.1 Do you think that the virtual world is self-explanatory in terms of exploring and interacting with 3D spatial data? Please describe ***how and what aspects***? If not, please describe ***why***?

1.2 Have you had VR experience in VR games or academic VR applications?

Case 1: I have experience in both

Case 2: I have experience in either VR games or academic VR applications.

Case 3: I have no VR experience.

Case 1) If you had VR experience in both VR games and academic VR applications, do you think that adopting a game-based approach for engineering or scientific problems in VR like this study helps engineers or scientists perform tasks effectively? ***Please describe what aspects of the game-based approach and the reasons***. If it is not helpful, please describe ***why***?

Case 2) If you had VR experience in either VR games or academic VR applications (please specify which experience you had before), do you think that adopting a game-based approach for engineering or scientific problems in VR like this study helps engineers or scientists perform tasks effectively? ***Please describe what aspects of the game-based approach and the reasons***. If it is not helpful, please describe ***why***?

Case 3) If you had no VR experience, do you think that adopting a game-based approach for engineering or scientific problems in VR like this study helps engineers or scientists perform tasks effectively? ***Please describe what aspects of the game-based approach and the reasons***. If it is not helpful, please describe ***why***?

2- Of the interactions you experienced, which did you like the best and ***why***?

2.1 ***How*** did the interaction support what you were trying to do (e.g., data exploration, performing tasks, validating your hypothesis or theory, etc.)?

2.2 Do you think that any elements of the game-based approach (e.g., spaceship mission concept of this study and virtual objects, such as a robot, dialogue panel, start button, touch screens on the second scene) affect your interactions with the second scene in terms of exploratory data analysis? ***Please describe how and why.***

3- Of the interactions you experienced, which did you like the least and ***why***?

3.1 ***How*** did the interaction not support what you were trying to do (e.g., data exploration, performing tasks, validating your hypothesis or theory, etc.)?

3.2 Please provide your opinion on how the interaction will be improved?

4- Did you feel that the first scene was helpful in terms of informing you about the general context of the VR system (e.g., the goal of the VR system, general explanation about the virtual world, and immersion in the virtual world, etc.) before performing exploration, interaction, and tasks in the second scene and ***why and how***? If not, please describe ***why***?

5- Did you have your hypothesis, theory, or assumption that you wanted to test or validate while exploring the second scene?

Case1) If yes, please describe your thinking process that you performed to implement your idea with the interactive tools in the second scene.

Case2) If no, did you think that the three guidelines from the robot for the exploration in the second scene give you a starting point to explore the second scene? Please describe ***why*** and ***how***. If not, please describe ***why***.

6- While exploring the models, did you discover new information or knowledge about the relationship between the static model and dynamic model?

Case1) If yes, please describe ***what*** you discovered and ***which interactions*** were helpful to support what you discovered.

Case2) If no, please describe ***why*** and please suggest if you were a designer, ***what you would design for interactions*** to encourage a user to discover new information or knowledge about the static and dynamic models.

7- Do you think that this study or this VR demo helps you expand your knowledge or experiences? If it does, how is the expansion likely to affect your real life in any ways? Please describe ***what aspects and how***. If not, please describe ***why***?

8- Please provide a general comment on the VR demo (e.g., what you liked, what you did not like, things to be improved, things you would like to say about this study or the design of the VR application, etc.)

References

[1] Höppner, F. (2009, July). How much true structure has been discovered?. In *International Workshop on Machine Learning and Data Mining in Pattern Recognition* (pp. 385-397). Springer, Berlin, Heidelberg.

[2] Santhisree, K., & Damodaram, A. (2011). SSM-DBSCAN and SSM-OPTICS: Incorporating a new similarity measure for Density based Clustering of Web usage data. *International Journal on Computer Science and Engineering*, *3*(9), 3170.

[3] Bellinger, G., Castro, D., & Mills, A. (2004). Data, information, knowledge, and wisdom.

[4] Davenport, T.H. and L. Prusak (1998). Working Knowledge: How Organizations Manage What They Know. Harvard Business School Press, Boston, MA.

[5] Ackoff, R. L. (1989). From data to wisdom. *Journal of applied systems analysis*, *16*(1), 3-9.

[6] Wiig, K.M. (1993). Knowledge Management Foundations: Thinking About Thinking - How People and Organizations Create, Represent, and Use Knowledge. Schema Press, Arlington, TX.

[7] Tuomi, I. (1999, January). Data is more than knowledge: Implications of the reversed knowledge hierarchy for knowledge management and organizational memory. In *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences. 1999. HICSS-32. Abstracts and CD-ROM of Full Papers* (pp. 12-pp). IEEE.

[8] Sagiroglu, S., & Sinanc, D. (2013, May). Big data: A review. In *2013 international conference on collaboration technologies and systems (CTS)* (pp. 42-47). IEEE.

[9] Eynon, R. (2013). The rise of Big Data: what does it mean for education, technology, and media research?.

[10] Rendón, E., Abundez, I. M., Gutierrez, C., Zagal, S. D., Arizmendi, A., Quiroz, E. M., & Arzate, H. E. (2011, January). A comparison of internal and external cluster validation indexes. In *Proceedings of the 2011 American Conference, San Francisco, CA, USA* (Vol. 29, pp. 1-10).

[11] Mohammadpoor, M., & Torabi, F. (2018). Big Data analytics in oil and gas industry: An emerging trend. *Petroleum*.

[12] B.Gerhardt, K. Griffin and R. Klemann, "Unlocking Value in the Fragmented World of Big Data Analytics", Cisco Internet Business Solutions Group, June 2012,

[13] Fan, J., Han, F., & Liu, H. (2014). Challenges of big data analysis. *National science review*, *1*(2), 293-314.

[14] Iqbal, R., Doctor, F., More, B., Mahmud, S., & Yousuf, U. (2020). Big data analytics: Computational intelligence techniques and application areas. *Technological Forecasting and Social Change*, *153*, 119253.

[15] Lohr, S. (2012). The age of big data. *New York Times*, *11*(2012).

[16] Koperski, K., Adhikary, J., & Han, J. (1996, June). Spatial data mining: progress and challenges survey paper. In *Proc. ACM SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery, Montreal, Canada* (pp. 1-10).

[17] Sedlmair, M., Meyer, M., & Munzner, T. (2012). Design study methodology: Reflections from the trenches and the stacks. *IEEE transactions on visualization and computer graphics*, *18*(12), 2431-2440.

[18] https://vtk.org/

[19] Ramos Mota, R. C., Cartwright, S., Sharlin, E., Hamdi, H., Costa Sousa, M., & Chen, Z. (2016, October). Exploring immersive interfaces for well placement optimization in reservoir models. In *Proceedings of the 2016 Symposium on Spatial User Interaction* (pp. 121-130).

[20] Somanath, S., Carpendale, S., Sharlin, E., & Sousa, M. C. (2014). Information visualization techniques for exploring oil well trajectories in reservoir models. In *Proceedings of Graphics Interface 2014* (pp. 145-150).

[21] Rocha, A., Mota, R. C. R., Hamdi, H., Alim, U. R., & Sousa, M. C. (2018). Illustrative Multivariate Visualization for Geological Modelling–Supplementary Material.

[22] de Carvalho, F. M., Brazil, E. V., Marroquim, R. G., Sousa, M. C., & Oliveira, A. (2016). Interactive cutaways of oil reservoirs. *Graphical Models*, *84*, 1-14.

[23] Telea, A. C. (2014). *Data visualization: principles and practice*. CRC Press.

[24] Kinsland, G. L., & Borst, C. W. (2015). Visualization and interpretation of geologic data in 3D virtual reality. *Interpretation*, *3*(3), SX13-SX20.

[25] Lidal, E. M., Langeland, T., Giertsen, C., Grimsgaard, J., & Helland, R. (2007). A decade of increased oil recovery in virtual reality. *IEEE computer graphics and applications*, *27*(6), 94-97.

[26] Cartwright, S. (2018). *Secure Collaboration Across the Reality-Virtuality Continuum Using Reservoir Data* (Master's thesis, Science).

[27] Sahaf, Z. (2018). Visual Analytics Framework for Exploring Uncertainty in Reservoir Models.

[28] Sultanum, N. B. (2011). *Exploring novel interfaces for 3d visualization of reservoir simulation post-processing data* (Master's thesis, University of Calgary).

[29] de Almeida, J. M., & Miranda, C. R. (2019, October). Virtual Reality for Visualization of Enhanced Oil Recovery Processes at Nanoscale. In *Offshore Technology Conference Brasil*. Offshore Technology Conference.

[30] Ferrara, P., Ricci Maccarini, G., Poloni, R., Campaci, R., Favaretto, M., & Grasso, T. (2020, January). Virtual Reality: New Concepts for Virtual Drilling Environment and Well Digital Twin. In *International Petroleum Technology Conference*. International Petroleum Technology Conference.

[31] ZHU, K. R., JIA, H. Q., WU, X. M., & Zhang, T. W. (2019). Research on 3D Dynamic Well Control Virtual Reality System1.

[32] Cabral Ramos Mota, R. (2017). *Exploring Immersive Virtual Environments for Well Placement Optimization in Reservoir Models* (Master's thesis, Graduate Studies).

[33] Busse, J., B. Humm, C. Lübbert, F. Schlüter, A. Zeh, M. Moelter, V. Reibold, E. Rewald, T. Seiler, and Tegtmeier. "Actually, What Does "ontology" Mean?: A Term Coined by Philosophy in the Light of Different Scientific Disciplines." Journal of Computing and Information Technology 23.1 (2015): 29-41. Web.

[34] Heidegger, M., Macquarrie, J., & Robinson, E. (1962). Being and time.

[35] Gualeni, S. (2015). *Virtual worlds as philosophical tools: How to philosophize with a digital hammer*. Springer.

[36] Mostafa, A. E. (2013). *PetroVis & FractVis: Interactive Visual Exploration of High-dimensional Oil and Gas Data* (Master's thesis, Graduate Studies).

[37] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112, pp. 3-7). New York: springer.

[38] Burkov, A. (2019). *The hundred-page machine learning book* (pp. 70-82). Quebec City, Can.: Andriy Burkov.

[39] Ward, M. O., Grinstein, G., & Keim, D. (2015). *Interactive Data Visualization: Foundations, Techniques, and Applications*. CRC Press.

[40] Killough, J. E. (1995, January). Ninth SPE comparative solution project: a reexamination of black-oil simulation. In *SPE Reservoir Simulation Symposium*. Society of Petroleum Engineers.

[41] Islam, A. W., & Sepehrnoori, K. (2013). A review on SPE's comparative solution projects (CSPs). *Journal of Petroleum Science Research (JPSR) Volume*, *2*(4).

[42] Kun Wang, Hui Liu, and Zhangxin Chen. A scalable parallel black oil simulator on distributed memory parallel computers. Journal of Computational Physics, 301:19–34, 2015.

[43] Schroeder, W. J., Lorensen, B., & Martin, K. (2004). *The visualization toolkit: an object-oriented approach to 3D graphics*. Kitware.

[44] Abraham, A. B., Evans, A. B., & Thompson, B. E. (2019). Analysis of Spatial Distribution Pattern of Reservoir Petrophysical Properties for Horizontal Well Performance Evaluation-A Case Study of Reservoir X. *The Open Petroleum Engineering Journal*, *12*(1).

[45] Cunha, L. B. (2003, January 1). Integrating Static and Dynamic Data for Oil and Gas Reservoir Modelling. Petroleum Society of Canada. doi:10.2118/2003-219

[46] McInnes, L., & Healy, J. (2017, November). Accelerated hierarchical density based clustering. In *2017 IEEE International Conference on Data Mining Workshops (ICDMW)* (pp. 33-42). IEEE.

[47] Rosalina, E., Salim, F. D., & Sellis, T. (2017, May). Automated density-based clustering of spatial urban data for interactive data exploration. In *2017 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)* (pp. 295-300). IEEE.

[48] Ertöz, L., Steinbach, M., & Kumar, V. (2003, May). Finding clusters of different sizes, shapes, and densities in noisy, high dimensional data. In *Proceedings of the 2003 SIAM international conference on data mining* (pp. 47-58). Society for Industrial and Applied Mathematics.

[49] Swarndeep Saket, J., & Pandya, D. S. (2016). An overview of partitioning algorithms in clustering techniques. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, *5*(6), 1943-1946.

[50] Shneiderman, B. (1996, September). The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings 1996 IEEE symposium on visual languages* (pp. 336-343). IEEE.

[51] Keim, D. A. (2002). Information visualization and visual data mining. *IEEE transactions on Visualization and Computer Graphics*, *8*(1), 1-8.

[52] Dahshan, M., Polys, N. F., Jayne, R. S., & Pollyea, R. M. (2019). Making sense of scientific simulation ensembles with semantic interaction. In *Computer Graphics Forum*.

[53] Campello, R. J., Moulavi, D., Zimek, A., & Sander, J. (2015). Hierarchical density estimates for data clustering, visualization, and outlier detection. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *10*(1), 1-51.

[54] https://hdbscan.readthedocs.io/en/latest/index.html

[56] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, *12*, 2825-2830.

[57] McInnes, L., Healy, J., & Astels, S. (2017). hdbscan: Hierarchical density based clustering. *Journal of Open Source Software*, *2*(11), 205.

[58] Campello, R. J., Moulavi, D., & Sander, J. (2013, April). Density-based clustering based on hierarchical density estimates. In *Pacific-Asia conference on knowledge discovery and data mining* (pp. 160-172). Springer, Berlin, Heidelberg.

[59] Campello, R. J., Moulavi, D., Zimek, A., & Sander, J. (2015). Hierarchical density estimates for data clustering, visualization, and outlier detection. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *10*(1), 1-51.

[60] Tarjan, R. E. (1975). Efficiency of a good but not linear set union algorithm. *Journal of the ACM (JACM)*, *22*(2), 215-225.

[61] https://learn.unity.com/tutorial/design-thinking-your-digital-aesthetics?uv=2018.1&projectId=5d0763baedbc2a001ebefa9f

[62] Bögl, M., Aigner, W., Filzmoser, P., Lammarsch, T., Miksch, S., & Rind, A. (2013). Visual analytics for model selection in time series analysis. *IEEE transactions on visualization and computer graphics*, *19*(12), 2237-2246.

[63] Self, J. Z., House, L., Leman, S., & North, C. (2015). *Andromeda: Observation-level and parametric interaction for exploratory data analysis*. Technical report, Department of Computer Science, Virginia Tech, Blacksburg, Virginia.

[64] Pirolli, P., & Card, S. (2005, May). The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of international conference on* intelligence analysis (Vol. 5, pp. 2-4).

[65] Ward, M. O., & Yang, J. (2004, May). Interaction spaces in data and information visualization. In VisSym (pp. 137-145).

[66] Goldstein, J., Roth, S. F., Kolojejchick, J., & Mattis, J. (1994). A framework for knowledge-based interactive data exploration. Journal of Visual Languages & Computing, 5(4), 339-363.

[67] Van Leeuwen, M. (2014). Interactive data exploration using pattern mining. In Interactive knowledge discovery and data mining in biomedical informatics (pp. 169-182). Springer, Berlin, Heidelberg.

[68] Chen, Z. (2007). Reservoir simulation: mathematical techniques in oil recovery. Society for Industrial and Applied Mathematics.

[69] Ertekin, T., Abou-Kassem, J. H., & King, G. R. (2001). Basic applied reservoir simulation (Vol. 7). Richardson, TX: Society of Petroleum Engineers.

[70] Bowman, D. A., & McMahan, R. P. (2007). Virtual reality: how much immersion is enough?. *Computer*, *40*(7), 36-43.

[71] Franssen, M., Lokhorst, G. J., & Van de Poel, I. (2009). Philosophy of technology.

[72] Dusek, V. (2006). *Philosophy of technology: An introduction* (Vol. 90). MaldenOxfordCarlston: Blackwell.

[73] Ihde, D. (2012). Experimental Phenomenology Multistabilities. Albany: State University of New York Press.

[74] Verbeek, P. P. (2005). *What things do: Philosophical reflections on technology, agency, and design*. Penn State Press.

[75] Mitcham, C. (1994). *Thinking through technology: The path between engineering and philosophy*. University of Chicago Press.

[76] Wheeler, G., Deng, S., Toussaint, N., Pushparajah, K., Schnabel, J. A., Simpson, J. M., & Gomez, A. (2018). Virtual interaction and visualisation of 3D medical imaging data with VTK and Unity. *Healthcare technology letters*, *5*(5), 148-153.

[77] Prim, R. C. (1957). Shortest connection networks and some generalizations. *The Bell System Technical Journal*, *36*(6), 1389-1401.

[78] Dix, A., Dix, A. J., Finlay, J., Abowd, G. D., & Beale, R. (2003). *Human-computer interaction*. Pearson Education.

[79] Ruiz, P. H., & Agredo-Delgado, V. (Eds.). (2019). Human-Computer Interaction: 5th Iberoamerican Workshop, HCI-Collab 2019, Puebla, Mexico, June 19–21, 2019, Revised Selected Papers (Vol. 1114). Springer Nature.

[80] Carroll, J. M. (2010). Conceptualizing a possible discipline of human–computer interaction. *Interacting with Computers*, *22*(1), 3-12.

[81] Preece, J., Sharp, H., & Rogers, Y. (2015). *Interaction design: beyond human-computer interaction*. John Wiley & Sons.

[82] Ellis, G., & Dix, A. (2006, May). An explorative analysis of user evaluation studies in information visualisation. In *Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization* (pp. 1-7).

[83] LaViola Jr, J. J., Kruijff, E., McMahan, R. P., Bowman, D., & Poupyrev, I. P. (2017). *3D user interfaces: theory and practice*. Addison-Wesley Professional.

[84] Marriott, K., Schreiber, F., Dwyer, T., Klein, K., Riche, N. H., Itoh, T., ... & Thomas, B. H. (Eds.). (2018). *Immersive Analytics* (Vol. 11190). Springer.

[85] Adams, E.: The designer's note book: Postmod-ernismandthe3typesommersion(2004), http://www.gamasutra.com/view/feature/130531/the_designers_notebook_.php

[86] Moreland, K. Diverging color maps for scientific visualization (expanded). *Proceedings in ISVC*, *9*, 1-20.

[87] Andrews, C., Endert, A., & North, C. (2010, April). Space to think: large high-resolution displays for sensemaking. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 55-64).

[88] Wang, Q., Chen, Z., Wang, Y., & Qu, H. (2020). Applying Machine Learning Advances to Data Visualization: A Survey on ML4VIS. *arXiv preprint arXiv:2012.00467*.