Kunnskap for en bedre verden

Department of Computer Science

TDT4259 - Applied Data Science

# A Multi-criteria Site Selection Analysis for Optimal Placement of an Underwater Data Warehouse in Vestland County, Norway


*Group 11*

| | |
|---|---|
| Marcus Hjertaas | 539571 |
| Martin Wighus Holtmon | 522441 |
| Olivia Foshaug Ingvaldsen | 543894 |
| Jakob Lindstrøm | 522804 |
| Tørris Sandsæter | 504654 |
| Elena Espenes Tjoland | 539454 |


CC-BY 27.11.2023

# Executive summary

This report is the result of a geospatial data science project. The topic of the project is site selection for an underwater data warehouse, and it has a three-sided objective. First, define a set of criteria for the optimal place to deploy an underwater data warehouse. Secondly, use this definition to identify locations in Vestland county that fulfil this set of criteria. And last, perform a consequence analysis of the resulting alternatives, and make a site-recommendation for stakeholders.

The first part of the problem was solved through qualitative measures. An optimal location for deployment was characterised by a set of criteria, such as being located within a certain distance from the power grid, or avoiding places where there are existing maritime infrastructures. After a definition was established, the project focused on identifying such optimal locations in Vestland county. The site selection process was solved using Geographical Information System (GIS) software such as ArcGIS Pro and QGIS, and relevant Python libraries. A multi-criterion decision analysis (MCDA) was implemented as the preferred method, where Python and QGIS was primarily used for exploratory data analysis, and ArcGIS Pro was used for performing the final integrated analysis.

After creating the definition and deploying it to the quantitative part, the analysis came up with two locations that fulfilled all pre-defined criteria. The two alternatives were located in the northern parts of Vestland county, in shallow waters outside Selje- and Nordfjordeid, respectively. There might be other factors of importance that was not captured by our definition. Such missing factors might represent flaws in the analysis, in addition to any information that is missing, unavailable or inaccurate. These factors must be taken into account by all decision makers using this report for site-selection purposes.

The output from this project is two potential sites for placing an underwater data warehouse in Vestland county, Norway. This output - given that the quality of the analysis is good - can be of value for relevant decision makers. This analysis is impacted by several uncertainties and risks, from both a qualitative and quantitative perspective. Such uncertainties and risks might derive from lack on data, resistance from the local population, regulatory barriers and environmental risks.

# Table of Contents

# List of Figures

# List of Tables

# 1 Introduction and problem definition

## 1.1 Context

Data centres produce large amounts of heat, which must be disposed of to keep the components functional and running. As a consequence of the components' cooling needs, it is estimated that over 40% of a data centre's energy consumption is spent on cooling its components (X. Zhang et al., 2017). By leveraging the often very cool ocean water for temperature transport, the concept is aiming to increase the cooling efficiency of data warehouses (Roach, 2023). This includes taking advantage of the superior heat-conducting properties of water compared to air and other gaseous media (CTM Magnetics, 2016).

As the energy consumption of data centres grows, due to the development of more energy-intensive hardware technologies. As well as there being an increased demand for server infrastructure, reducing energy consumption is key. Taking advantage of the water to cool servers is one such strategy. This means the energy required to cool the data warehouse does not have to come from electricity, but can rather come from the natural heat transfer of water. Additionally, underwater data warehouses could potentially be deployed in non-centralised locations, offering services with higher speed and lower latency in areas with poor internet infrastructure. This would allow more people to access higher-quality services.

The goal of underwater data warehouses is primarily to reduce energy consumption by utilising the naturally cold water and provide better access to services by allowing for potentially more distributed deployment of data centres (Roach, 2023; Tamblyn, 2018).

However, while there are advantages to these kinds of data centres being flexibly deployed, not all sites are equal. There are places where these cannot or should not be deployed. A combination of environmental variables such as ground conditions, access to required infrastructure, conflicting interests, and political/legal variables are some examples of those reasons. As with any construction project on land, it has to be planned in advance.

However, underwater data warehouses are relatively new (M. Zhang, 2022), and are therefore lacking in research. This applies to research about the underwater data warehouses themselves, as well as the logistics surrounding their deployment. As there are already projects looking into the feasibility and design of these data warehouses, we opted to focus on the deployment location and its criteria. This was also a problem we could solve using data, which we will describe in detail later.

However, in order to perform a good analysis, we needed to have a clear definition of what constitutes a good site. In this report, we have defined which criteria to use. This analysis aims to help decision-makers identify potential deployment sites that meet our definition of a good site. We approached this problem as external consultants hired by Equinor to provide recommendations for possible underwater warehouse sites in Vestland county, Norway. Applying the definition on the research area in order to locate possible deployment sites. We applied the definition on Vestland county in order to locate possible deployment sites. This project also had a goal of demonstrating the feasibility of using data analysis as an approach to locate deployment sites, by showing that the criteria can actually be used in practice to locate good sites.

## 1.2 Motivation

This project's objective is to identify possible areas for underwater data warehouses. This problem requires a solution that combines both qualitative and quantitative geospatial analyses. This was to ensure that our project not only addresses the problem of where a site could be, in a specific geographical area, but also to formulate a reusable definition for locating sites outside our limited geographical scope.

If underwater data warehouses continue to gain traction, more deployment sites have to be located sooner or later to facilitate further distribution of new such warehouses. This is a data-driven problem, as the sites can be located by applying some criteria to one or more data sets to narrow down the possible deployment locations. However, those criteria have to be defined.

Herein is our first motivation; by defining the criteria for a good site, we also lay the foundation for the analysis, making it easier for shareholders to decide on a deployment site. Additionally, when the criteria are defined, it is easier to do the analysis, as it adds constraints to what data is needed. It also ensures that any new teams attempting to locate a good site have a criteria list that they can follow in attempt to reduce the chances of achieving undesirable outcomes.

These criteria are demonstrated in practice in the second part of our project. Here, we applied the criteria to geographical data from Vestland county in Norway to demonstrate the criteria. This demonstrates the feasibility of using these criteria on geographical data to locate good candidates for deployment sites, and that the problem of finding a good site can be solved with a data-oriented approach. In other words, the goal was to demonstrate the feasibility of using data to simplify the process for decision makers interested in deploying an underwater data warehouse.

The overarching motivation behind the topic of the report is to further drive the deployment of these data warehouses, particularly given their potential for more sustainable energy consumption in regard to cooling requirements. Taking a data-oriented approach allows us to provide optimal suggestions for deployment sites that can help make underwater data warehouses a reality. Additionally, as previously mentioned, by also defining the criteria in a reusable manner, the outcome of the analysis can simplify the decision-making process around underwater data warehouse deployment.

## 1.3 Scope

Solving this issue with the use of data analysis allowed for a complete breakdown, accounting for all relevant variables in the pursuit of optimal deployment site(s). The process of searching for possible locations became more efficient through the use of data and estimated constraints, rather than a physical or qualitative approach. Applying data analysis also allows for precise spatial modelling and visualisations of the relevant locations. This is applied both in presenting the projects results in addition to considerations and analysis throughout the report.

Utilising data-driven methods also allowed for scalability and future-proofing by defining a reusable definition that can be run on arbitrary, yet applicable, data sets. Additionally, the simplicity of making changes to a data analysis opens up its usage in future projects.

As already mentioned, we used geographical data on a list of criteria for a good site to locate potential deployment locations in Vestland county. However, this list had to be defined, before it could be used, as the analysis for possible deployment sites is just as much a demonstration of the criteria in practice as it is about locating these sites.

The data analysis required sufficient data that covered all constraints and requirements. This allowed the optimisation problem to be subjected to all necessary aspects, resulting in locations that satisfy the characteristics. Without high-quality data, the analysis would produce subpar results that either breached certain requirements or have high degrees of uncertainty tied to them. This also limited the project, ensuring that the data is of sufficient quality.

There are certain limitations to the data analysis, the most prominent being the large computational resources needed when applying the search to a sizeable area, and the availability of data material for large continuous areas. By restricting the search to Vestland county, the computational and data requirements are reduced. Reducing the computational load was a decision we made to keep the analysis step practical to perform for the group members. Our decision to limit the search to Vestland county was based on the county's lengthy coastline, which translates to a higher proportion of suitable areas within a smaller geographic search area. Additionally, the county's power grid has a high density of power plants, resulting in a surplus of energy that we could potentially leverage (Statistics Norway, n.d.).

Additionally, as we will get back to in Section 6.1.2, it is not a guarantee that a deployment project will be allowed to connect to the grid. Picking an area with a power surplus is a smart choice to minimise potential problems when it is time to connect to the grid.

## 1.4 Team roles and responsibilities

The team that produced this report consists of three students from NTNU Business School and three students from the department of Computer Science. Given the diversity of backgrounds and technical/academic qualifications, we were able to bring a broad set of skills and perspectives to the project. All group members have experience in data science, with some also having prior experience with geomatics and Geographical Information System (GIS) software. This collaboration allowed us to approach the project with a comprehensive understanding of a data-driven project's business and technological aspects. The primary responsibilities of each member are outlined in Table 1.

| **Member** | **Role(s)** |
|---|---|
| Marcus | Research, data exploration and GIS analysis |
| Martin | Background research and Report writing |
| Olivia | Report writing and research |
| Jakob | Data collection and presentation |
| Tørris | Data exploration and GIS analysis |
| Elena | Report writing and analysis |

**Table 1:** Project Team Roles

The team started by collectively outlining the scope of the project and researching factors influencing the placement of an underwater data centre. This initial approach ensured a thorough understanding of the conditions and requirements of establishing a data centre in an underwater environment and laid the foundation for the subsequent data-sourcing stage. Next, the team sourced as much relevant data as possible, assigning the responsibility of each factor among the members to increase the efficiency of this phase.

Once the data sourcing phase was completed, the preprocessing and analysis phase begun. In this phase, the team split into a group with focus on the report, and a group with a focus on the data preprocessing and GIS analysis. Within the GIS group, members were given specific data sets to analyse and prepare the different data layers required for the final integrated analysis.

We opted to do this as the preprocessing phase is difficult to do in parallel, or work together with, due to the highly individual nature of the geospatial tools. Distributing the geographical data presented a convenient way to distribute responsibilities, without adding an excess amount of work. The distribution also allowed us to strategically join the data together in preparation for the final analysis.

Naturally, all members were still included in the analysis process, and everyone still had input throughout the course of preprocessing and analysis phases. All evaluations regarding the data were performed collectively, even if a subset of the group was assigned the work of preparing the data.

Throughout the project, we had weekly work sessions varying between 2-4 hours in duration, where the first part consisted of stand-ups to discuss progress and identify blockers. We did so in order to ensure that everyone in the group continued to stay up-to-date with what everyone else was doing and to ensure any problems were handled collectively. This also allowed us to discuss other topics that aren not necessarily blockers, but still need or benefit from discussion.

The remaining time was used to work on the project and open up for discussions collectively. Before concluding the sessions, we assigned individual tasks to each team member to ensure progress between meetings. Towards the end, the session length and frequency increased as more discussion and coordinated work were required to finalise the project, with a high focus on writing the report and detailing the results of the analysis.

# 2 Background

Effective decision-making for complex business goals often depends on having a well-defined strategy. These objectives frequently involve large amounts of data, making manual analysis difficult and impractical. As a result, having a well-structured data strategy is essential for producing and deploying useful results for the company.

In this chapter, we outline the objectives of the project and how they were resolved using our data resources. We follow a data science project management approach based on the CRISP-DM framework. Additionally, we use the Design Thinking to complement the data strategy and have a creative approach to solving our problem. We will also discuss why this strategy aligns with the project's goals.

## 2.1 Objectives

As previously mentioned, the goal was to find the best possible locations for an underwater data warehouse in Vestland county. We have defined the main objective of our work as: Identify the optimal location for an underwater data warehouse. The objective will produce site recommendations, which can be narrowed down through a qualitative discussion.

By identifying optimal locations, we aim to assist Equinor in finding the most suitable site. This leads us to two main objectives:

1. Defining what constitutes a good site for underwater data warehouses.
2. Finding suitable sites meeting these criteria within Vestland county.

## 2.2 Project management strategy

In our project management strategy, we adopt a comprehensive approach that integrates both the CRISP-DM framework and Design Thinking principles. The dual-framework strategy aims to address the challenges of identifying the optimal location for an underwater data warehouse. By combining these frameworks, we aim to not only just handle the complexities of data analysis but also to come up with creative and "user-friendly" solutions. The combined approach makes for a robust management strategy that takes into account both the technical and user-focused aspects of our goals.

### 2.2.1 CRISP-DM framework

To achieve these objectives, we will employ a data-driven approach, leveraging the CRISP-DM framework, which stands for Cross-Industry Standard Process for Data Mining. This framework provides a structured approach to planning and executing data mining projects and has been widely adopted in various industries due to its flexibility (Wirth & Hipp, 2000).

The frames is structured into six phases: *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modelling*, *Evaluation*, and *Deployment* (Hotz, 2023). The methodology outlines the tasks associated with each phase and their relationship. Figure 1 represents a life cycle for a data-driven project and allows for flexibility in transitioning between the phases when necessary.

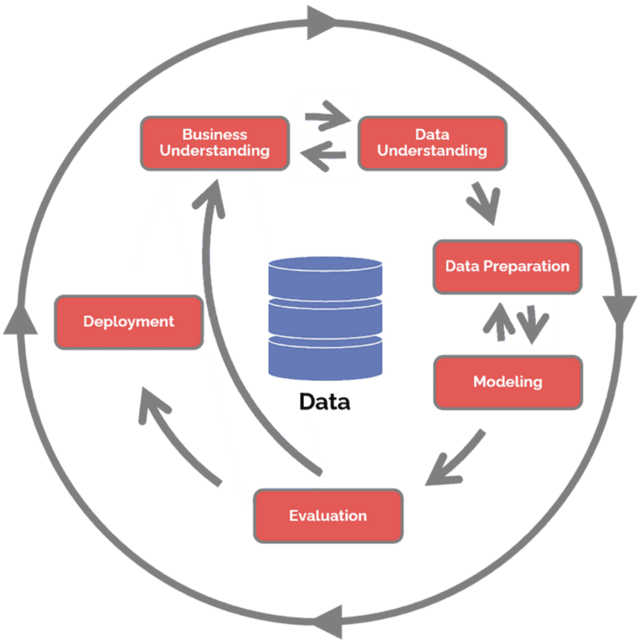


**Figure 1:** CRISP-DM Diagram.
Source: Hotz, 2023

The CRISP-DM framework can be perceived as either Agile or Waterfall methodologies, depending on how it is used. It can be viewed as a rigid, waterfall-like process due to its structured phases, but by iterating quickly and focusing on incremental deliverables, the framework allows for agile-like iteration and adaptation as the project progress (Hotz, 2023). This framework is useful in this project as it is flexible. In the case of locating a good site for underwater data warehouses, there are lots of uncertainties tied to the project.

First of all, this project relies on a high volume of data to perform a site selection analysis. The framework provides a systematic approach to handling and analysing the data required for this analysis. Even though there are a lot of considerations and constraints, it provides a structured approach to ensure we cover most of the aspects necessary for defining a good site. The ability to iterate quickly and focus on gradual milestone aligns well with our project's complexity. The process might force us to revisit previous phases as we can gain more insights or encounter new constraints. The framework allows us to do this efficiently.

The framework also outlines distinct stages, from understanding the business problem to the final deployment. This ensures we proceed through all phases of a data science project and avoid missing any critical steps in the process.

Finally, given that our project required the analysis of data from various sources, the data understanding and data preparation phases are particularly relevant. The CRISP-DM methodology provided guidance on how to handle these tasks effectively.

### 2.2.2 Design Thinking

Design thinking is an addition to our project strategy, serving as an iterative and human-centric approach that complements the CRISP-DM framework. This methodology reframes complex problems into human-centric challenges, emphasising the focus on what is most important to Equinor. The key stages of design thinking are *Empathise*, *Define*, *Ideate*, *Prototype* and *Test* (Interaction Design Foundation - IxDF, n.d.).

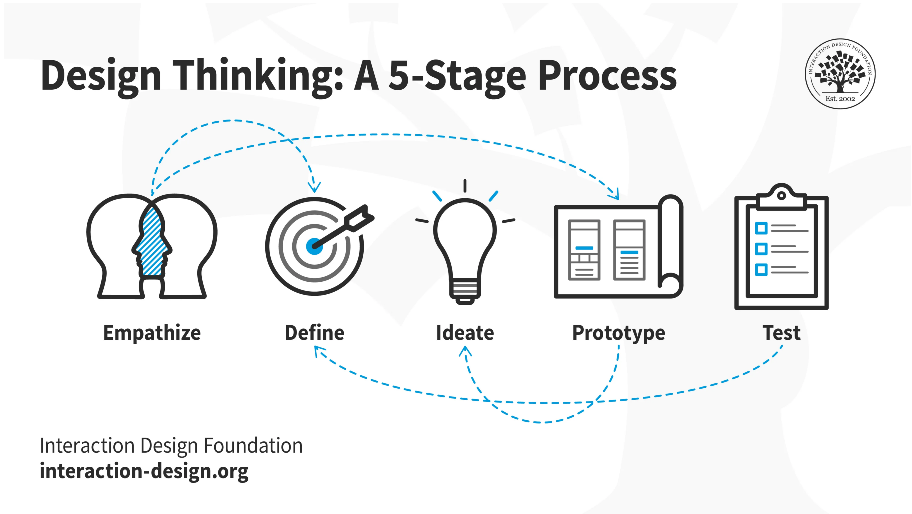


**Figure 2:** Design Thinking.
Source: Interaction Design Foundation - IxDF, n.d. CC BY-SA 3.0

The first stage, *empathise*, involves research to gain insights into the view of Equinor and other stakeholders involved in the decision-making process. This means understanding the challenges, preferences, and needs through interviews with key staff in Equinor, research on similar projects like Natick in Scotland, and exploring our initial data. These insights are important in shaping our Data strategy, as detailed in Section 2.3.

The second stage, *define*, focuses on creating a human-centred problem statement based on the insights gathered during the *empathise* phase. Although our primary objective was clear from the outset, we made the decision to only look at Vestland county, as discussed in section 1.3. This stage also involved identifying criteria for "no-go zones" along the coastline to rule out locations that present issues, regulatory barriers, or other significant challenges for deployment.

In the *ideate* phase, we identify innovative solutions through brainstorming sessions. This involves generating ideas on how to approach the problem, such as determining the optimal data strategy and choosing the appropriate tools for the analysis.

The fourth phase, *prototype*, involved testing out different methods in a defined test area, allowing us to experiment and evaluate the feasibility of our approach in a smaller environment. This phase allowed us to explore different approaches to site selection, and display potential challenges with our approach, giving a more informed decision for our final methodology.

Lastly, the *test* phase was used in evaluating the analysis results. Confirming the outputs of recommended areas are in line with the criteria and are avoidant of any other factors that could impact their credibility. This is expanded on in section 5.4.

Incorporating Design thinking into our project strategy enhances the effectiveness of the CRISP-DM framework. The problem-solving approach ensures that our final recommendations not only meet technical criteria but also align with Equinor's practical needs and experiences.

## 2.3 Approach

Our approach to defining a good site will involve comprehensive research to understand the environmental and technological factors that influence the suitability of underwater data warehouse locations. The environmental factors cover considerations such as oceanography, climate, and ecosystem impact, while technological factors involve infrastructure compatibility and accessibility. Additionally, there are several economic factors that impact the plausible areas. Our site selection analysis aims to provide a comprehensive understanding by thoroughly exploring these aspects.

For the identification of suitable sites, we gathered a diverse set of data, including geospatial, environmental, and infrastructural information for Vestland county. This data will include relevant details, ensuring a good understanding of each criteria outlined in Table 3. For instance, geographical features, climate patterns, and existing infrastructure will be examined to provide a good perspective on the potential sites.

The data gathering was performed after establishing all plausible factors that could impact the location of a data warehouse. After data was collected, it was used in exploratory analysis, which is expanded on in section 4.1. This allowed for a better understanding of which variables were necessary and exactly how they would impact our project and analysis.

This data will be analysed using spatial analysis techniques and multi-criteria decision analysis (MCDA) to evaluate and rank potential sites (Malczewski & Rinner, 2015).

### 2.4 Data strategy

Our data strategy is centred around the collection, processing, and analysis of relevant data. We prioritised open-source geographical data and attempted to access proprietary data where necessary. The data was processed and analysed using a combination of Geographic Information Systems (GIS) and data analytic tools. The data strategy was designed to complement the CRISP-DM framework, focusing on understanding the business understanding (defining a good site) and data understanding (identifying suitable sites) phases.

To illustrate the potential impact of our project, we can look to examples such as the Microsoft Project Natick, which demonstrated the feasibility and environmental benefits of underwater data centres (Microsoft, n.d.). By using geographical data to do site selection for the deployment of such data centres, we can reduce energy consumption and improve internet latency, impacting both the environment and global connectivity.

## 3 Method

In this section, we will present the geographical data sets along with some initial insights. We will provide a detailed description of the applied preprocessing techniques and an overview of the tools and methodologies employed in the analysis.

### 3.1 Data- and modelling introduction

This data science project performs a site selection through a geospatial analysis. Such projects differentiate themselves from the more common data science projects, which make use of statistics or machine learning (source). The primary distinction between machine learning models and geospatial analysis lies in the way they handle and utilise the data. Prevalent machine learning models use data sets that typically include numerical, categorical, textual, or image data. These models focus on finding patterns, relationships, and predictions based on statistical and algorithmic methods (A. Zhang et al., 2023). In contrast, geospatial analysis specifically deals with spatial data, which includes location information such as coordinates, geographical areas, and spatial relationships near the surface of the earth (AWS, 2023).

This project utilises various data files in different formats and data types. The large amount and variation of data require a lot of preprocessing. Those processes have been done using the Python programming language[1] and the GIS tools QGIS[2] and ArcGIS[3]. Due to the nature of the data and the project's desired outcome, a rule-based algorithm has been chosen as the modelling technique.

[1]https://www.python.org/
[2]https://qgis.org/en/site/
[3]https://www.esri.com/en-us/arcgis/about-arcgis/overview

## 3.2 Description of the data sets

The geographical data sets have been collected through a variety of websites, mostly governmental. Since the data are collected from a variety of websites, the type of data has also been varying in terms of the file formats, data types, and projections[4]. The project started out with 11 data files. However, several more files have been created due to the preprocessing phase.

### 3.2.1 Data collection

The table below shows the data files we collected to perform our analysis. The table provides information concerning features, the Coordinate Reference System (CRS), file format, whether it is engineered, and the source of the data. The engineering column refers to whether the data have been changed in comparison to its original attribute straight form its source.

| Description | CRS | File Format | Engineered | Data supplier |
|---|---|---|---|---|
| Domestic Fibre Infrastructure | - | JPEG | Yes | Norwegian Datacenter Industry, 2023 |
| Maritime municipality borders | UTM33 | GeoJSON | Yes | GeoNorge, n.d.-c |
| Vestland county borders | UTM33 | GeoJSON | Yes | GeoNorge, n.d.-a |
| Harbours | UTM33 | GeoJSON | Yes | Kystinfo, n.d. |
| Powerlines (sea and land) | UTM32 | GeoJSON | No | Norges vassdrags- og energidirektorat, n.d. |
| Powerplants (hydro and wind) | UTM32 | GeoJSON | No | Norges vassdrags- og energidirektorat, n.d. |
| Nature reserves | UTM32 | GML | No | GeoNorge, n.d.-b |
| Maritime infrastructure | UTM32 | GML | No | GeoNorge, n.d.-d |
| Coastal traffic | UTM32 | shapely | Yes | Barents Watch, n.d. |
| Digital (submarine) terrain model | UTM32 | TIFF | Yes | Kartverkets Sjødivisjon, 2022 |

**Table 2:** The following is an overview of the data sets collected

Additionally, a data set concerning coral reefs was considered, as it would reveal more possible conflict areas the warehouse should not be deployed on. Since the search area had been narrowed down to Vestland county, we did not include the data set because there are no coral reefs within Vestland county's administrative borders (Fiskeridirektoratet, n.d.). However, future analysis with a larger geographical scope should include this data set, as other counties are affected.

### 3.2.2 Data Preprocessing

We utilised multiple data sets in this project, some as standalone geographical layers, and some where merged together to create a new layer of data. This is due to our analysis using many different factors, and these do not exist as a single data set but rather multiple data sets from different sources. This comes with multiple complications:

1. The data sets were not all available in the same file formats. The majority were GeoJSON and Esri SHAPE, but we also got GML, SOSI, TIFF and other file types.
2. The data sets describes very varying features, and with different resolutions. Joining these together to cohesive data requires a lot of work.
3. Some data sets allow different coordinate systems to be used, which also has to be handled.
4. Some of the files have different coordinate reference systems that need to be normalised.

[4] A map projection is a method used to represent the three-dimensional surface of the Earth on a two-dimensional plane, with various types emphasising different aspects such as area, shape, distance, or direction.

It is important to make all the data readable and usable for the two GIS software used in this project. This has been done by changing the file formats, types, and coordinate reference systems to the same format. We have been handling a variety of data formats, including GeoJSON, GML, SHAPE, TIFF and XYZ[5]. Most of these formats are compatible with at least one of our GIS software, with the exception of XYZ. Therefore, we converted the incompatible format to GeoJSON using Python scripts.

Most of the data sets required for our analysis were available in vector format, with the exception of bathymetry data, which is only available in raster format. Vector data consists of points, lines and polygon entity types defined by coordinate geometry (x, y), and can store many variables in addition to reference coordinates. The second datatype is raster data, which is characterised by it is grid structure of regularly spaced cells, where each cell holds a single value representing a continuous or categorical spatial variable.

Furthermore, the actual data have been handed out in different coordinate reference systems (CRS), referred to as CRS from here on. The three most common CRS are UTM32[6], UTM33, and latitude & longitude. To visualise the data in a reasonable way, all the data need to have the same CRS. If the data do not share the same CRS, it will look precarious. All the data have been preprocessed to have a CRS equal to UTM32 to avoid such a result. These processes have been handled using Python, QGIS, and ArcGIS. Finally, all the data files have been reformatted into the .gpkg file format. The individual preprocessing for each data set will now be described:

The fibre data came as an an illustration map of Norway with the domestic fibre infrastructure represented as red points with connecting lines, we extracted the red points on the map along the coastline to generate a synthetic data set. To convert the points on the map to coordinates, we plotted three reference points, Oslo, Stavanger, and Trondheim, and used those points to calculate the coordinates of the extracted points. The result was exported as GeoJSON files so that it could be used in the GIS software. Given that this was a very manual process, and the data set is an illustration map, the result should be looked at as a synthetic data set, and should not be given particular weight in the analysis.

Marine municipality borders was a data set that contained maritime borders for all of the municipalities in Norway. The first process was to remove the municipalities that did not belong in Vestland county. Afterwards, the CRS had to be changed from UTM33 to UTM32. The two next data sets Vestland county borders and harbours faced the same problem, and was solved using the same method. Both of the data files was given in UTM33 coordinates, which was changed in the layer properties in QGIS.

The six next data files: power lines (sea and land), power plants (hydro and wind), nature reserves and maritime infrastructure did not go through any changes. All these files were available in GIS-compatible file formats projected in the UTM32 coordinate system.

The coastal traffic data was a very large file, which contained all the common routes for commercial coastal traffic in the whole of Norway, and all routes in and out of Norway. Most of this data is outside the scope of Vestland county, therefore only the routes intercepting with Vestland county were used in the final analysis. In order to do this the overlay intercept tool, is used in the buffer analysis in section 4.3.1.

The DTM 50 bathymetry data set of Norway's economic zone exists as 1860 separate TIFF files in a mosaic data set. In order to prepare a data set covering the fjords and continental shelf of Vestland county, 113 files were merged into one separate TIFF file with the *Mosaic to New Raster* tool in ArcGIS Pro 3.2. This allowed for the DTM model to be handled as one object in the GIS analysis.

The *conflict zones* is a variable made up of several factors all taken from the maritime infrastructure data set table 2. The factors that went into conflict zones were split into polygon and line types mostly they cover the same infrastructure types with some exceptions. The factors in the polygon variable are ship wreckage, sub-sea cables, seaplane ports, rooted aquaculture, floating docks,

[5].xyz is a fileformat for storing x-, y- and z coordinates.
[6]UTM is an acronym for Universal Transverse Mercator, it is a grid system for map projection.

dumping grounds, bridges and nature conservation areas collected from the nature reserves data set table 2. The line variable contains sub-sea cables, rope-ways, pipelines, high-voltage cables, docking cables and bridges.

## 3.3 Methods and tools

For this project, QGIS and ArcGIS were used in collaboration for the main analysis. The site selection is built on a rule-based algorithm in a multi-criteria analysis. These GIS tools are built specifically to handle the analyses of geographical data and provides simple navigation and visualisation options, lowering the computational need and time invested. As procedures like changing projections and coordinate reference systems were considerably simpler than in more familiar tools like Python, it allowed for more time to be spent on the actual analysis. As mentioned in the preprocessing paragraph, python was used in many instances to clean and alter the data. However, performing the entire site selection analysis there would be a lot more time-consuming than necessary.

The rule-based algorithm consisted of several criteria that defined what a good site was. These criteria are presented in full later in section 4.2.1. Site selection is a process where the analyst strives to determine the optimal location of something while satisfying a set of criteria. It is similar to an optimisation problem in that it has several constraints limiting the position of, in this case, underwater data warehouses. What makes it different in this project is that we are not optimising the location based on any given factor, rather we make binary decisions that are either plausible or not. A site selection is in principle a complex decision-making problem, and in order to perform site selection one needs to use appropriate complementary methods (Rediske et al., 2021). In this project, we decided on a rule-based system or a multi-criteria decision-making. In concept alone site selection can be a purely qualitative analysis, finding the location of a data centre based on a set of criteria or considerations.

The foundation of a multi-criteria decision analysis or MCDA is to support a decision maker in choosing the most preferable variant of many possible options (Wątróbski et al., 2019). The method does this by taking in several criteria characterising the acceptability of decision variants. MCDA comes in two main types, continuous and discrete. The analysis in this project uses a combination of both types as this site selection requires a form of linear programming and preference and indifference. There are a lot of different approaches to MCDA that differ in how they tackle the continuous and discrete constraints and how well they handle quantitative and qualitative data (Wątróbski et al., 2019). The most notable of these multi-criteria approaches are fuzzy membership approaches. In this project none of these methods are specifically used, rather the problem is approached rather bluntly. Utilising constraints with binary results to find a plausible area and several preferential variables to optimise a site based of this opportunity area. Whether this approach is similar to a defined MCDA method is beyond the research done into MCDA and has been down prioritised. Compatibility to the project thesis was more important than using a predefined analysis method. Since the definition of a good site and by extension the evaluation criteria are a mixture of objective observations and subjective assumptions and conclusions, it was natural to perform an analysis that was easily controllable and explainable.

Working with spatial data the methods used for narrowing down our opportunity area were different vector overlay methods in both QGIS and ArcGIS. Mainly the intercept method, extracting areas where the layers overlapped, i.e., areas that fulfil all criteria. The difference method, subtracting overlapping conflict areas from the plausible area was used to rule out specific areas in accordance with our pre-determined selection criteria. These are simple geometry operations and provide precisely what is needed to narrow down the plausible area. The selections were made based on distances to or from the different variables. Mostly these distances where measured using geometric buffers that extended the required distance, using the overlap to either intercept or differentiate out the plausible area. Opportunity areas unrelated to distances were fed into these geometric analyses from separate sub-analyses.

# 4 Analysis

As prefaced in the introduction, the analysis is organised into qualitative and quantitative sections. This is a slight simplification of the process. In reality the analysis has four sections the first is finding an appropriate research area. This was done in the introduction 1.3, limiting the search area to Vestland county.

The next phase is an additional qualitative section, were the aim is creating a definition of what a good site is. This includes a thorough review of all important characteristics for a deployment location. This includes essential features such as water depth, surface topography and access to power. In addition it consists of beneficial features such as power surplus, avoiding conflict zones and proximity to civilisation.

Based of this research area and the definition we can proceed to the quantitative analysis where the features defining a good location will be reformulated as constraints allowing for a visualisation of all feasible locations found in the designated area.

Lastly, the feasible areas can be analysed resulting in recommended locations, these positions are deemed as the best location for a potential underwater data warehouse in the designated area. This process is a combination of qualitative and quantitative features, both taking into account the optimal values of the essential and beneficial features.

## 4.1 Exploratory Analysis

To ensure a good understanding of the relevant data sets, an exploratory data analysis was performed. For this purpose, both Python and the GIS tools were utilised. The exploration itself consisted of plotting the data with Python scripts and in GIS software, as well as familiarising with the variables attached to each data object in their attribute tables. Data visualisation gives a good understanding of the different data types, formats and their precision. The data sets were initially visualised separately for detailed scrutiny and quality evaluation, before the they were explored further in unity. Understanding the data through exploratory analysis is crucial for planning and evaluating their utility in more complex analyses.

The relevant data-sets employed in this project are presented in Figure 3. The map extents only display information contained within the borders of Vestland county, and are visualised on top of the Open Street Base-map in ArcGIS Pro 3.2

## 4.2 Definition of a good site

There are a number of factors at play when determining a good site for placing a underwater data centre. These considerations aim to ensure the feasibility and efficiency of the data warehouse deployment. Some factors include "no-go zones", areas strictly regulated under Norwegian law, and conflict zones involving marine activities. Some of the important ones we decided to account for were:

- "No-go zones", specifically zones where building or other human activity is strictly regulated under Norwegian law
- Conflict zones with other marine activities, including fisheries and offshore wind farms
- Depth
- Seabed topography (slope)
- Distance from the shore
- Access to power and internet infrastructure

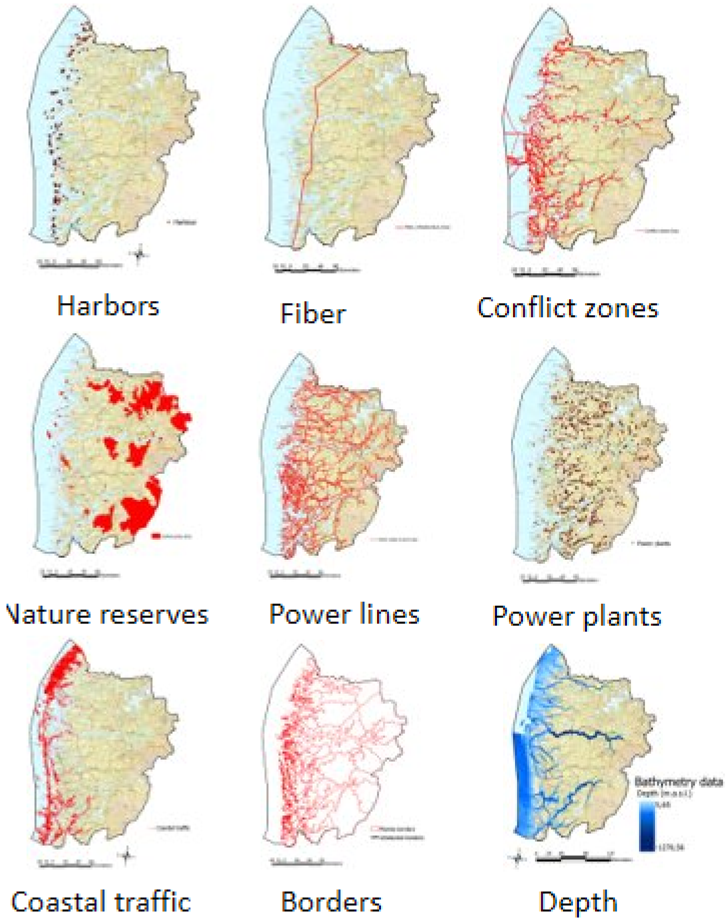


**Figure 3:** Univariate preprocessed data analysis.

- Power surplus
- Ship traffic
- Distance to nearest major city, for access to qualified personnel for maintenance (if necessary)
- Seabed sediments to ensure slope stability

### 4.2.1 Decision criteria

In the following section a table of all the variables used in the quantitative GIS analysis is included with their respective constraints. These decision criteria effectively define the requirements and constraints of a good location for placing submarine data warehouses.

| Variable | Criteria | Consideration | Comment |
| --- | --- | --- | --- |
| Distance to shore | $35 < x \leq 1000m$ | Economic | Placement is within manageable distance for maintenance |
| Distance to power grid | $\leq 1000m$ | Economic | The further away the costlier |
| Distance to harbour | $\leq 1000m$ | Economic | A large distance will be costly during deployment and maintenance |
| Fibre cable | $< 5000m$ | Economic | The further away the costlier |
| Water depth | $35m < x < 120m$ | Environmental | Security distance to ships |
| Seabed slope | $\leq 2.5°$ | Environmental | Needs a stable and flat surface |
| Size of area | $\geq 22500m^2$ | Environmental | A security distance, securing that we have a sizeable area to deploy |
| Conflict zones | $> 50m$ | Planning | Security distance to avoid other marine infrastructure |
| Ship Traffic | $> 150m$ | Planning | Security distance from recorded ship traffic |

**Table 3:** Overview of the site selection criteria

Most of the variables that define a good site was used in the quantitative GIS analysis. However, a few criteria, such as seabed sediment types and maximum distance to nearest power plant and city, were given a secondary or qualitative roles. Partly because they should not receive too much emphasis for the quantitative screening of opportunity areas - as they are not as critical in terms of their immediate importance. Moreover, we consider them to be more general consideration that are valuable for the more subjective consequence analysis of the final opportunity area alternatives. In the case of seabed sediment, we had some initial qualms about the data as it was provided in point format. In essence this brought uncertainty how accurate it would be across the entire county, resulting in it being used as a argument in recommendation rather than the analysis itself.

A lot of the criteria distances are estimations, either based on costs, environmental or physical limitations or security distances tied to planning and deployment. Variables like distance to power grid, harbour, fibre cable and the shoreline are effectively arbitrary numbers. However, they represent a maximum cost that the distance to these infrastructures entails. When deploying a data warehouse there are direct costs tied to drawing fibre and power cables to the location. The further a data warehouse is placed from the shore or harbours, the higher the deployment costs. This impacts the financial viability of this type of project. Therefore the optimal position should be relatively close to these infrastructures, hence the 1km constraints. Unlike the other data the fibre cable has a larger distance criteria despite it is direct costs. This is due to the data being a unreliable estimation, there is a general uncertainty as to whether this data is close to the actual fibre cables.

The next type of constraints are environmental or psychical factors limiting the opportunity area of the data warehouse. Such factors cannot be tweaked or worked-around that much by the decision makers. These include the water depth and local seabed topography of the ideal site, as well as the physical extent/size of the area. The seabed slope limits the range of opportunity areas, as deploying the warehouse in a steep underwater seabed formation could lead to technological malfunctions, risk submarine landslides or other unforeseen incidents. Therefore, by limiting the search to slopes $\leq 2.5°$, with a minimum size of 22500 $m^2$, we can ensure that area surrounding the warehouse is approximately flat. The water depth is also a limiting factor with important considerations. It is partly a technical constraint, as data warehouses deeper than -120m has never been tested, nor are they planned in the near future (Microsoft, n.d.). Furthermore, the vessel would need to withstand the additional pressure that comes with increased water depth. A third consideration is the minimum required depth to ensure clearance from the hulls of passing ships. Based on these considerations, a range from -35 meters to -120m in depth seems appropriate and practically achievable. This includes a safety margin to ship hulls, as well as avoiding large structural or technological changes from previous pilot projects.

Lastly, there are some decision criteria that ensures security distances from different marine infrastructure or activity. In this category, constraints like ship traffic and other marine activities are included. For most of the factors included in this group of criteria, they have no physical considerations restricting the position of the warehouse, but rather it simplifies the planning and insures a successful deployment if these areas are avoided with a sizeable security distance. Conflict zones are complex, as they include several different activities, stakeholders and interests that we do not wish to interfere with 4.2.1. What these factors are have been reviewed previously. By implementing a security distance to all major marine activities, we avoid unnecessary complications and risks.

## 4.3 Analysis using GIS Software

This section compiles the sub-analyses and explains how we implemented the methods to generate the results.

### 4.3.1 Buffer analysis

To implement the analysis parameters specifying proximity or distance to specific infrastructures, activities, and areas, buffer zones were created around these map layers as specified in Table 3. Buffer zones representing opportunity areas define a distance interval where the underwater data centre can be placed with criteria of the type "less than". While those specifying distance "more than" indicate an interval for the minimum required distance to conflict areas or zones that needs to be avoided.

Power cables, harbours, ship-traffic routes, marine activity conflict zones with belonging buffer zones are presented in Figure 4, 5 and 6, respectively. With the exception of shorelines, all buffer zones are displayed with grey colour-coding. The original data-elements are represented by either point, line or polygon features with red colour-coding. Figure 4 is included as an example of how the buffer layers looks in a larger figure.

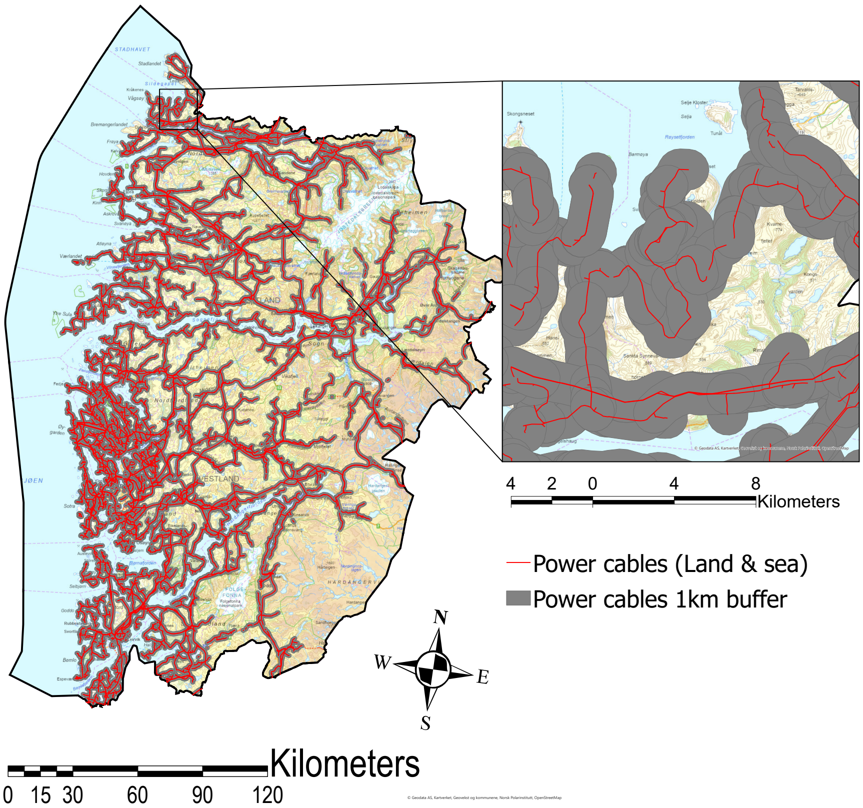


**Figure 4:** Data layer containing power cables with a 1km buffer zone.

The selection criteria "distance to shore", "distance to closest harbour" and "distance to power grid" all specify a given proximity to infrastructures that are vital for the operation and maintenance of an underwater data centre. These selection criteria were implemented in the analysis by creating 0-1km buffer zones around these data elements. After the buffer zones were created and loaded into the study area's map extent, areas that met all three criteria were identified with the *intersection tool* in QGIS. Areas that met none, or just one or two of the criteria, were thus excluded from the opportunity area. With the proximity criteria in place, the next step was to implement areas to avoid, i.e., no-go zones specifying a minimum required distance.

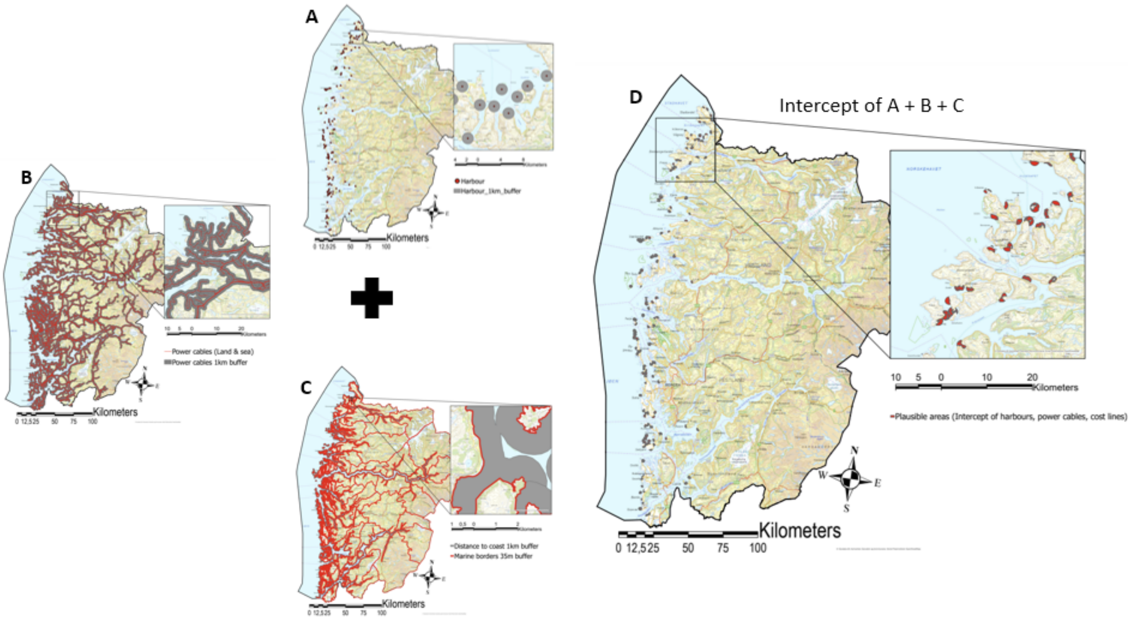


**Figure 5:** Intercept layer analysis: Overlapping the buffer layers for the (A) variables harbour, (B) power cables and (C) coastline in order to (D) extract the intercept of these.

By creating a 150m buffer zone around all ship traffic routes, areas within said distance of such activity was differentiated as no-go zones for the underwater data centre. The same process was repeated for marine activity conflict zones and the shoreline, where the buffer distance was set to 50m and 35m, respectively. The shoreline data-set that was employed for the latter buffer-zone was coarse and inaccurate with respect to the shorelines of the base-map layer, such that a certain level uncertainty was introduced into the shore-line buffer.

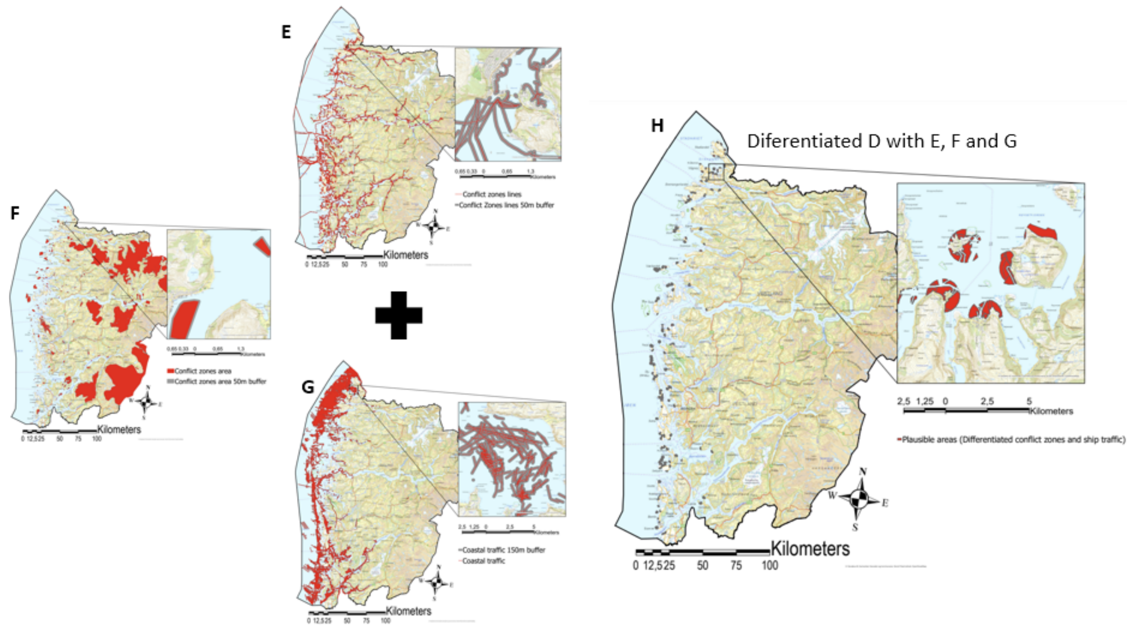


**Figure 6:** Differentiation layer analysis: Differentiating out (E) ship traffic, (F) conflict zones (polygon) and (G) conflict zones (lines) from (D) the intercepted plausible area.

After both a intercept and differentiation analysis the extracted areas where still missing a few vital variables like slope and depth which underwent a slightly different process in the analysis.

Ultimately, these analyses leaves us with figure H or the plausible area pre-raster data, as it is referred to in later parts of the analysis, Figure 9.

## 4.4 Slope and depth analysis

A digital terrain model (50×50m) was used to identify ocean floor with 0 - 2.5° slope at -35 to -120m depth in the fjords and continental shelf of Vestland county (see Figure 7A). This data set was created by merging bathymetry data from all the municipalities of Vestland county using the *Mosaic To New Raster* tool. Next, the depth data was reclassified to a binary TIFF file with the *reclassify* tool, such that all raster-cells with depth values ∈ (-35, -120m) received the new value 1, while cells outside the opportunity depth range received a value of 0 (see Figure 7B).

The slope layer in Figure 7C was calculated from the bathymetry data set by the *Slope* spatial analyst tool. This algorithm calculates the slope of each raster-cell based on the depth values of the neighbouring cells (Esri, 2023b). The slope layer was reclassified to binary as described above, such that the opportunity areas received a value of 1 (0 - 2.5°), while other areas received 0 (see Figure 7D). By multiplying the two raster layers cell-by-cell, the product of raster-cells fulfilling both criteria was 1 ($1 \times 1 = 1$), while all other values was 0. Afterwards, the product raster-layer was converted to vector-format with the *Raster to Polygon* tool. From the new vector-layer, all polygons with a value of 0 and area $< 22.5\ km^2$ were removed by *Select by Attributes > delete*. The remaining opportunity areas shown in yellow in Figure 7E were exported as .SHAPE-file to be implemented with the other vector layers described in section 4.3.1.

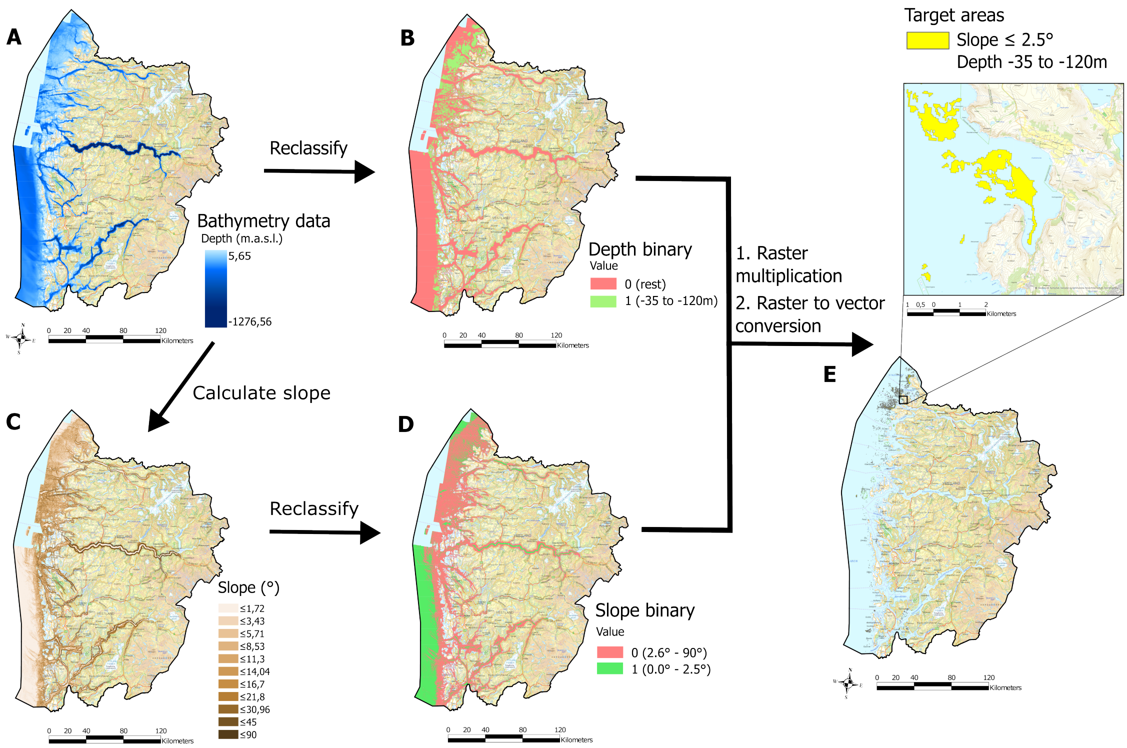


**Figure 7:** Slope and depth analysis. Areas fulfilling the selection criteria of 0 - 2.5° slope, -35 to -120m depth and area > 22.5 daa were modelled from the bathymetry data set of Vestland county. All analysis steps were performed in ArcGIS Pro 3.2 GIS software.

## 4.5 Seafloor sediment analysis

The Geological Survey of Norway (NGU) has performed an extensive mapping of the seafloor sediments in Norwegian waters, analysing 18650 point samples in Vestland county alone (see Figure

8A). These samples are classified into 10 sediment-classes based on grain-size (mm), as shown in Table 4. Although this data set consist of evenly to irregularly spaced point samples, these geo-referenced data-points can be used to predict the spatial distribution of these sediment types by spatial interpolation (Sibson, 1981).

| Classname | Grainsize (mm) |
|---|---|
| Clay | 10 |
| Silt | 20 |
| Sandy clay | 30 |
| Fine sand | 95 |
| Sand | 100 |
| Coarse sand | 105 |
| Sandy gravel | 160 |
| Gravel | 170 |
| Rock and block | 180 |
| Biologic material | 500 |

**Table 4:** The sediment type classification employed in the seafloor sediment type data set by The Geological Survey of Norway (NGU).

ArcGIS Pro 3.2 houses a myriad of different geospatial analysis tools, including a number of spatial interpolation methods. The most widespread being Inverse Distance Weighted (IDW), Empirical Bayesian Kriging, Spline and Natural Neighbour interpolation (Esri, 2023a). Since Natural Neighbour interpolation performs well in preserving abrupt changes in the data - even with irregularly spaced data - this interpolation method is well suited for predicting the distribution of sediment types - which can vary substantially within short distances (Rousi et al., 2011). The latter quality was pivotal when choosing a suitable interpolation-method for the NGU data set, as the density of sampling sites varies a great deal throughout the county's fjords and coastal areas.

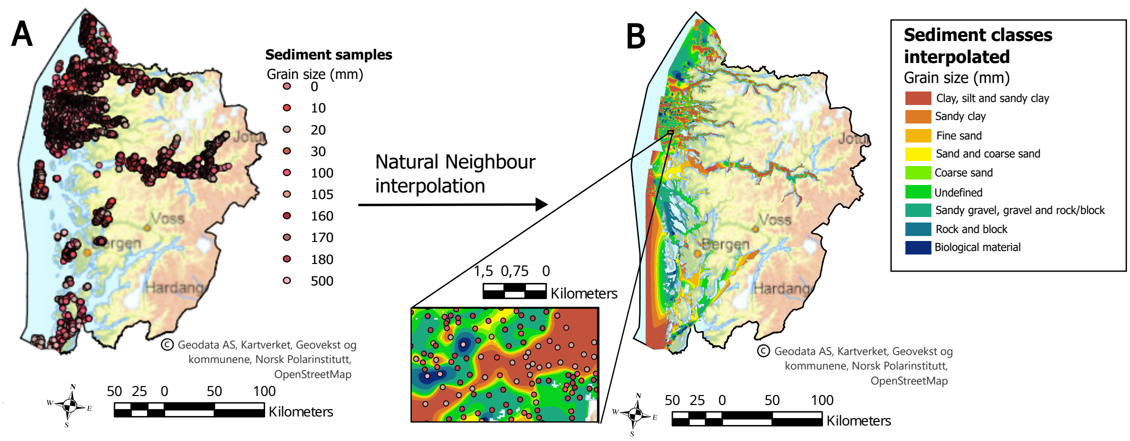


**Figure 8:** Prediction of seafloor-sediment distributions in Vestland county from the NGU Marine Bunntyper data set. The sediment distribution map was interpolated using Natural Neighbour interpolation in ArcGIS Pro 3.2.

The seafloor sediment distribution map obtained using the *Natural Neighbour* tool is shown in Figure 8B. The inset map shows the predicted spatial distribution of the 10 sediment classes in a smaller area outside of Dalsfjorden in Vestland county. This inset clearly demonstrates how similar sediment types are distributed in unique colour-coded areas, as can be read from the sediment class legend in Figure 8B. The predicted seafloor-sediment distribution map can prove a valuable tool when choosing between several opportunity sites, as the bottom sediments which the underwater data centre will sit on can have consequences for it is stability (Hance, 2003).

As briefly mentioned in section 4.2.1, the accuracy of the predicted sediment distributions rely on the data material from which the predictions were made. As shown in Figure 8A, the 18650 samples of the NGU dataset are concentrated in a few location in the Sognefjord and the northern parts of the county, with the rest of the county being only scarcely mapped. Therefore, this data layer is included mostly as a possible evaluation tool for recommendations, rather than a rigid criteria in the analysis.

### 4.6 Integration of sub-analyses

After importing the vector layer with opportunity areas based on the depth and seabed slope criteria, it was overlaid with the buffer layers described in 4.3.1. Extracting the intercept of these layers, i.e., areas fulfilling all selection criteria, we had our final opportunity areas. The only criteria remaining was distance to the fibre cable and the size of the area. The former was done by selecting areas within a 5000m distance of the fibre cables with the *Select by Location* tool in ArcGIS Pro 3.2. The latter was achieved by filtering the remaining sites according to polygon area only extracting the sites with an area equal or larger than 22 500m$^2$.

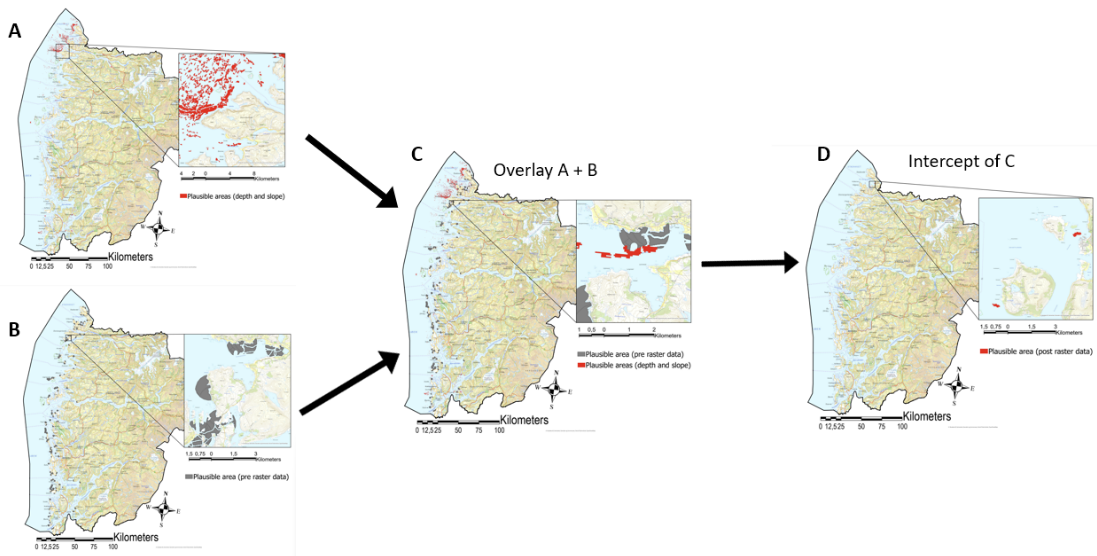


**Figure 9:** Overlapping the plausible areas from the (A) slope and depth variables and (B) the results of the buffer analysis.

After filtering the opportunity areas according to distance to fibre cables and total area, our GIS analysis had created an output of two areas that met all selection criteria and constraints. This marked the end of our quantitative GIS analysis, all that remained was performing a consequence analysis and evaluation of the two opportunity areas pertains to qualitative analytic considerations.

## 5 Evaluation and interpretation

In this part we will present our results. We will discuss pros and cons of the locations and furthermore give a recommendation based on qualitative considerations.

### 5.1 Results

Our analysis has identified two potential locations for the underwater data warehouse. These locations are situated within the fjords outside Nordfjordeid and Selje, both located in the northern region of Vestland county. The selection of these sites is based on an examination of the criteria defined in section 4.2.1, such as environmental, technological, and infrastructure factors.

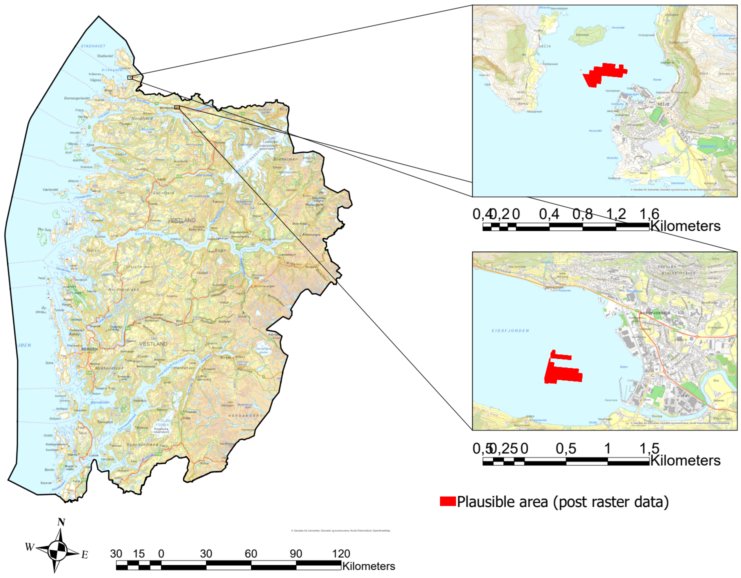


**Figure 10:** Showcasing the plausible areas outside of Sleje and Nordfjordeid

## 5.2 Consequence analysis

In evaluating the suitability of the results, we need to consider several factors aligned with our original business goal. This includes evaluating whether the selected locations meet the specified technical and cost considerations. Additionally, we will reflect on how well these locations align with our initial location requirements. The suitability analysis will consider the strengths and weaknesses of each location, providing an understanding of their potential impact on the project's success.

### 5.2.1 Seafloor sediments at opportunity areas

In order to distinguish the two opportunity areas derived from the GIS analysis, the seafloor sediment types found at these locations were explored using the sediment distribution map produced by Natural Neighbour interpolation in section 4.5. The seafloor-sediment distribution map is presented in Figure 11, where the inset maps display the two opportunity areas located in the Nordfjord and outside of Selje. Based on the sediment samples taken in these areas, both opportunity areas are located on finely granular sediments, i.e., clay, silt or sandy clay. If the underwater data centre were to placed outside Selje, it would sit on top of sandy clay, while a data centre in Nordfjord would be placed on even finer clay and silt or sandy clay.

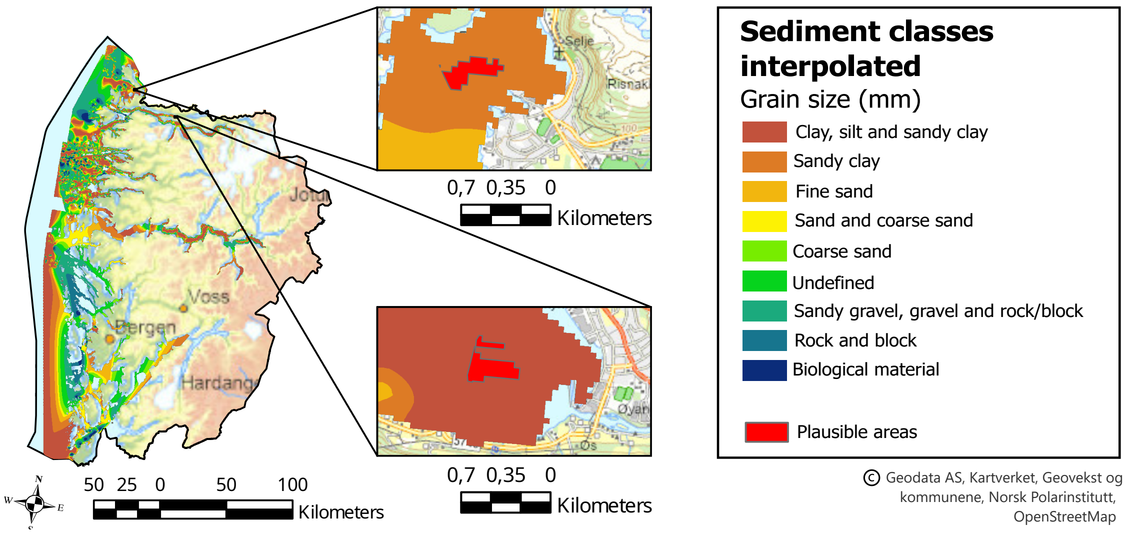


**Figure 11:** Predicted sea-floor sediment distributions at the two opportunity areas for the underwater data centre in Nordfjord and Selje, Vestland county.

Considering that the underwater slope stability increases with the size of sediment grains (Hance, 2003), the opportunity area outside of Selje would in theory be slightly more stable than that of Norfjordeid. This theory is based on the type of sediment found at the sites of 266 registered submarine landslides in the US, where 115 were triggered in clay, whereas 80 and 62 were triggered in silt and sand, respectively (Hance, 2003).

### 5.2.2 Evaluation of Selje

#### Advantages

**Isolation and Low boat traffic:** The location of Selje experiences minimal boat traffic for larger vessels. This could mean reduced disruptions form boat traffic but also minimises the interference in therms of maintenance and physical operations for the data warehouse. The location contributes to a more conducive setting for the smooth functioning of the facility.

**Proximity to Måløy:** Being close to Måløy in the maritime route provides access to a larger port facility. The port is designed for larger vessels, for example Hurtigruten. This makes the distance to larger harbour facilities closer, without being a impediment to boat traffic. This also gives the opportunity to have technical personnel and equipment.

**Fibre Cable Connectivity:** With a distance of approximately 3800 meters to the nearest fibre cable, Selje meets our criteria for connection to this.

**Seafloor sediment type:** Since the seafloor is covered by sandy clay in this particular area, the slope stability should be higher compared to the Nordfjord opportunity area.

#### Disadvantages

**Long Distance to Airport:** The nearest airport is located at a considerable distance, potentially impacting accessibility.

**Limited existing industrial activity:** Selje currently has limited existing industrial activity, which may cause challenges in terms of immediate infrastructure and resources for large-scale projects.

### 5.2.3 Evaluation of Nordfjordeid

#### Advantages

**Airport proximity:** The nearby Sandande Airport offers advantages with easy access to both personnel and equipment requiring air transport.

**Port of Nordfjordeid:** Close proximity to the port of Nordfjordeid provides additional logistical benefits.

**Fibre Cable Connectivity:** With a distance of 4200 meters to the nearest fibre cable, Nordfjord also fulfils criteria for connectivity.

#### Disadvantages

**Ferry traffic:** The harbour experiences frequent ferry traffic, which could cause challenges for a data warehouse.

**Seafloor sediment type:** Since the seafloor is covered by clay, silt and/or sandy clay in the Nordfjord opportunity area, this slope could potentially be unstable, making submarine landslides a plausible risk. The fact that the local slope is $\leq 2.5°$does not mitigate this risk, as submarine landslides are known to trigger and travel far at slopes $\leq 3°$(Hance, 2003).

### 5.2.4 Discussion

Selje presents key advantages for hosting an underwater data warehouse. Its secluded location minimises boat traffic, reducing potential disruptions and ensuring a conductive operational environment. The proximity to Måløy enables access to a larger harbour designed for larger vessels, providing connectivity to essential marine industry and technical resources. However, a notable disadvantage is the long distance to the nearest airport. Additionally, Selje currently has limited existing industrial activity, which may create challenges in terms of immediate infrastructure and resources for large-scale projects. Conversely a additional marine infrastructure such as underwater data centres could provide a opportunity to drive more industry and workers to the small town.

On the other hand, Nordfjordeid has proximity to Sandane Airport which ensures easy access for key personnel and equipment requiring air transport. Close proximity to the Port of Nordfjordeid provides additional logistical benefits. Nevertheless, frequent ferry traffic in the harbour challenges the data warehouse operations.

### 5.2.5 Recommendation

After a thorough assessment of the advantages and disadvantages of the potential locations, we recommend Selje as the optimal site for an underwater data warehouse. Its secluded location away from frequent boat traffic minimises disruptions, while proximity to Måløy provides access to a larger port, especially suitable for larger vessels. The lower cost of the fibre cable, with a distance of only 3800 meters, makes Selje a more cost-effective option. Although Nordfjordeid has benefits such as proximity to the airport and an additional advantage with the harbour, these may not necessarily out-weight the practical and cost advantages offered by Selje. Since the slope stability at Selje should be higher relative to Nordfjord according to the local seafloor sediment types, this factor contributes to tip the scales in favour of Selje. Choosing Selje as the location thus provides a efficient solution for an underwater data warehouse.

Furthermore, Selje, being a location with limited existing industrial activity, presents an opportunity for positive local economic impact. If Equinor requires additional facilities, storage, or other resources in immediate proximity, it could contribute to fostering local economic development in Selje. This potential for creating a positive local ripple effect adds another layer of advantage to selecting Selje as the preferred site for the underwater data warehouse.

## 5.3 Analytical shortcomings

There are several potential shortcomings of this project. The shortcomings that we discuss is related to the data and method used for generating results.

### 5.3.1 Data sets

There are several shortcomings regarding data used in this project. The shortcomings can be subcategorised into two sections, quality of data and lack of data. This section was heavily emphasised by our project mentor Richard Hall, because the analysis is no better than the quality of the data it is built upon.

The data retrieved have been mostly downloaded from governmental or generally recognised websites. Therefore, the source of the data can be determined to be trustworthy. However, the data collected is not necessarily equally trustworthy. One important factor to look at is whether the data file and the actual data are discrete or continuous.

Starting of with the maritime municipalities borders one knows that the actual borders are of continuous nature, but the data file have discrete features. This indicates that the data used in the analysis is not a true representation of reality. The harbour data consist of points, but an actual harbour should be presented as a polygon. However, this difference is not considered to be of importance. The same situation is probably present in the data files for power plants and power lines.

As mentioned earlier, the data for the domestic fibre infrastructure of Norway is an illustration map provided as a JPEG, and we converted it manually to real world coordinates. As a result of this manual process and the limitation inherent in the quality of the data, the resulting data set should be considered synthetic. It is primarily included to supplement the realism of our analysis and should not be weighted highly during the analysis.

The DTM 50 bathymetry data set is a Digital Terrain Model collected in TIFF format. This raster data exists as the result of a 23-year long endeavour to map the seafloor of Norway's economic zone with multi-laser sonar, with The Norwegian Defence and The Mareno Project as main contributors (Kartverkets Sjødivisjon, 2022). As this is an ongoing project, some areas has yet to be mapped. This includes parts of the continental shelf of northern Vestland county, such that these areas could not be included in the analyses due to a lack of data. However, since our opportunity area was restricted to the coastal areas within 1km from the mainland, this fact likely does not impact our analysis. This data exists with a resolution of 5×5m and 25m×25m however, access to such detailed information is restricted to security-cleared personnel. The coarser DTM 50 is a very rough representation of reality, and do clearly not simulate the continuous nature of the seafloor topography. This resolution is sufficient for an initial screening, but if the project were to be implemented, it would be necessary to apply for security clearance to access details within the 50×50m grid. This is probably one of the largest quality shortcomings in our analysis.

The coastal traffic data shows only the common traffic routes for large vessels. The data might give the impression that the routes acts as roads and that everything besides the routes is "untouched" waters. However vessels, large and small, often deviates from these routes and therefore one cannot be sure that the data outside the routes is not subject of frequent coastal traffic.

Nature reserves is probably the only data that is a true representation of reality. The reason for this claim is that nature reserves are quite specifically defined, this means deviations is highly unlikely, and if the appear the geographical data should be updated and reflect the changes.

The second section regarding data shortcomings is lack of data. The first obvious lack of data concerns seabed sediments. When deploying an underwater data warehouse, or any marine infrastructure the knowledge of what this infrastructure is laying upon is imperative, however we where unable to find sufficient data on the topic. The data we did find on seabed sediments across the Norwegian west coast where lacklustre and only covered parts of the research area. As prefaced earlier both in section 4.2.1 and 5.2.1 including the data would likely not provide any improvement on the results. As we could not conclude on any plausible area, because they would have to undergo sediment test regardless. Based of not including this data, the end results are comprised. However, it is a certainty that sediment tests are needed and we can include it as a shortcoming rather than including data that fills this requirement halfheartedly.

Furthermore, there are probably other factor that might have significance in the analysis, which have not been accounted for.

### 5.3.2 Method, tools and analysis

This project has utilised simple rule based decision criteria as a method to narrow down the plausible areas. There are many different methods that are far more complex and analytical in their approach. Perhaps most notably methods like fuzzy membership, implementing a weight system to the different criteria. Assigning a continuous value to each criteria, like a distance in order to find an optimal site (Arron Beecham, Tony Gregory, Jonathan Michael, 2015). Deciding to go with a simpler rule based system has made the areas more or less binary as either plausible or not. Thus, recommendations are left to subjective qualitative discussions rather than selecting based on weighted criteria. Whether this is a shortcoming or not is relative to the goals of the project, yet it could limit the usage of this research.

Not relying on any specific algorithm allowed for more direct control over the analysis and its results. However, it also allowed for more human error. The buffer analysis was done in a step-by-step process using *intercept* and *differentiate* tools and relying on our understanding of the method, and a subjective visual confirmation of the results. With further research and knowledge of the topic of site selection and GIS analysis, we could have concluded in a more analytical approach, in favour of the brute force rule based approach.

The methods like *buffer*, *intercept* and *differentiate* where all chosen as a consequence of our lack of GIS knowledge. The methods are fundamentally easy to understand, since they are simple geometric operations. It is difficult to conclude whether these methods are appropriate in such a complex site selection problem. However, by using them we can deliver a concise result for the parameters we have defined. However, we are also limiting the analysis to simple binary outcomes where each criteria carries equal weight relative to each other. In a decision making scenario this is likely not the case, different parties would likely weight criteria differently depending on their position. That being the case, the results become redundant and the analysis would have to be repeated.

Some of the same method and analysis concerns apply to the slope and depth analysis, although that process relies a lot more on the innate data values.

These flaws are due to the methods and analysis approach used in this project. We early opted out of creating a interactive map which would have allowed decision makers to weight and toggle criteria based on their preferences. However, with the chosen methods and our GIS knowledge this was far to complex. Thus limiting the final product to a result and recommendation rather than something akin to a product.

The possible faults in both choice and implementation of these methods would bleed into the analysis, and by extension impact how it was performed or the interpretation of the sub-analyses and their results. Despite our beliefs that the methods where used appropriately and correctly this hinges on human intervention which is prone to error.

Looking past possible human error there is a chance the analysis has not been performed as assumed. Though unlikely, it is still a possibility that the areas we perceive as plausible are in fact not, because of a fault in the algorithm or again as a result from human error. Additionally as the fibre cable data is an estimate the degree which this is weighted in the analysis will also has an impact on the results.

## 5.4 Result quality control

In order to ensure the quality of the results, you could use qualified professionals with appropriate domain knowledge. Utilising knowledgeable personnel to approve or methods and approach to the site selection would be a great quality assurance. As this would enhance our confidence in the process leading up to our results. For the data you could apply the same approach contacting professionals like captains or other local marine experts that could inform us about other arguments related to the recommended areas. This could lead to other discoveries that the data did not cover. Another avenue to quality control the data would be a manual research approach. Simply searching the web for more data on the areas of interest or articles or papers that could lead to discoveries that our data did not cover. This could either present itself as a flaw from the data collection or our interpretation of it. As mentioned in previous sections a more accurate estimation of the fibre cable or actual data of it, would improve the quality of the result, likely it would also alter the recommended areas.

In general confirming or denying the suspicions about the project shortcomings would also reaffirm the quality of the results. If we could prove this conjecture it would allow us to better understand what cannot be accounted for and interpret the results more accurately. Or disproving the suspicions and by extension reaffirming the quality and our interpretation of the results.

# 6 Deployment and recommendation

As mentioned in Section 5.2.5, we recommend Selje as the top candidate for a deployment. In this section, we will recommend a deployment plan, look at some limitations regarding our method, and discuss future research options.

## 6.1 Deployment plan

The deployment plan for our recommendations is centred around the deployment scenario of a data warehouse. As out data is a decision foundation for Equinor, our project provides necessary input in a deployment, but cannot be deployed by itself.

Therefore, this deployment plan is centred around the deployment of a data warehouse, based on the recommendations of this report.

### 6.1.1 Quality assurance

The first step after receiving the data is to do quality assurance; checking with experts and other regional stakeholders, if any, to ensure the deployment site does not have any irreconcilable conflicts that were not reflected in data. This might primarily be applicable for sites where there is no data foundation to represent a conflict area. Some possible pitfalls are discussed in Section 6.2.

Human quality control on the analysis itself is important. Having expert's within GIS, geospatial analysis, or similar areas can inspect and validate the results. This is important to avoid issues from errors or shortcomings in the method, or shortcomings with the data. Feedback from these experts, based on the degree of importance, should either be taken into account when interpreting the results or should be used to change the analysis.

The next form of human intervention is detecting new important qualitative information. Such detection can happen in the form of workshops with business agents for the relevant sites. These workshops should aim to gather relevant business agents for the location, and present the plan of deploying and underwater data warehouse. During this activity potential hazards and new relevant information might be detected. Example of such hazards could be that our planned site is on the exact same place as a annual diving competition, or it might be right on top of an common anchoring point.

Generally speaking, having humans verifying the different aspects of the project is crucial. For this geospatial data science project, the analysis must undergo human intervention, because the result is only to be considered a recommendations mainly based on quantitative data. Relevant feedback from humans is desirable. Depending on the importance of the feedback, the analysis should be changed in accordance to the new information. Refining the preciseness- and quality of the analysis will significant improve the recommendations.

### 6.1.2 Ensuring power availability

There are certain logistical challenges that have to be checked. Even if a truly perfect site by all definitions is found, and there are not any problems with it, there can be other challenges with deployment. For example, Statnett has predicted that Norway will have a negative power balance as early as 2027 (Statnett, 2022b). In some of the more extreme regions, meaning northern Norway, all grid capacity for the next few years has already been reserved (Statnett, 2022a; Viseth, 2022).

However, our focus is on Vestland county, where 2027 is the biggest, immediate risk. The expected negative power balance can make it more challenging to connect to the grid in certain regions, or even anywhere in the country. In the event of this, at best, it delays the deployment of a data warehouse by a couple years. At worst, it fully prevents it in favour of sites in other countries with power grids that allow for more rapid deployment.

Alternatively, ensuring that the data warehouses are fully self-contained in terms of energy is also an option. This may be complicated, however, and means that just finding an optimal deployment site will not be enough. The analysis would have to be amended to include sites optimal for various types of renewables, and likely multiple types of renewables to ensure the data warehouse can remain operational 24/7. Building power generation infrastructure does increase the complexity of the deployment phase as well, by requiring more infrastructure before the data warehouse can be deployed.

This means that one major step before deployment is ensuring that the warehouse can be connected to the grid.

### 6.1.3 Deploying the warehouse

When everything has been addressed, the data warehouse can be deployed. There are more details around how this would have to happen, but this also boils down to when and how power and internet access is available, as well as when the data warehouse is designed and ready for deployment from Equinor's side. There may also be other logistical challenges, such as applications requiring approval before the warehouse can be deployed, but the exact details are left to Equinor to determine.

## 6.2 Limitations

While we have defined what we believe to be a good site, there may be factors that we have missed that change the answer. However, some of these factors may require trial and error to discover, so our analysis presents an educated attempt at as many of these factors as possible. Further quality control that these factors address as many major concerns as possible may be of interest in the future.

Additionally, as previously mentioned, grid capacity is a concern, but not one we can trivially measure in data. We can determine whether or not there is available capacity now, but this does not mean an application to connect to the grid would be accepted. NVE could deny it for many different reasons, the notable one being that other things are prioritised over a single data warehouse. This cannot be determined from data alone, but can at worst be the end of a data warehouse project - at least in the short term.

Any flaw mentioned in section 5.3 would limit the deployment of the warehouse. As they all impact the results, and with poor result quality we cannot blindly approve them. Prior to an actual deployment, several tests on all critical features would be run. However, running such tests are not cheap and performing them based of an uncertain analysis is unintuitive and likely a poor financial decision. Something Equinor or any other party with interest in deployment of underwater data warehouses would avoid.

### 6.3 Future research

As previously mentioned, part of the analysis was defining what factors constitute a good deployment site, and the reason was to allow future research with more data and access to more solid hardware. Future research should be in expanding the analysis to other areas, starting with covering the rest of Norway. However, this is primarily a next step once the underwater warehouse technology has gotten to a point where it can be widely deployed.

Going global with the data would make the system serve a much better foundation for selecting data, particularly if this change is combined with the use of fuzzy methodology, as this would enable a global list of potential sites, and for deployment to strategically

Additionally, it may be worth investigating whether there is data for power availability, as this could help avoid areas with a clear power shortage, or where all the capacity has been reserved already. This is primarily useful when expanding to a national or global scope, as it is possible to strategically pick regions that have a power surplus otherwise. Doing this can help avoid planning to build in an area where the data warehouse does not get the required permits to connect to the grid, though this will not fully defend against it. There are other reasons an application to connect to the grid can be declined that cannot be trivially represented in data, which is why this cannot be fully solved. However, including power availability and how much capacity has been reserved could be used to better model sites, or even take it a step further and plan *when* a site is expected to have enough unreserved capacity for a project to be viable.

It is also a good idea to reintroduce sediment analysis into the system, but this requires high-quality data, which we struggled with finding. Sediment analysis is useful for determining whether or not the ground is stable enough to handle the weight of an underwater data warehouse. However, for this to be more useful than just doing a manual sediment analysis with a human in the loop, higher resolution data than what we were able to find is required.

As the data for the domestic fibre infrastructure in Norway was an illustration map of Norway, further research is suggested. It is apparent that the underwater data centres must be connected to the fibre infrastructure, so the data set containing this information is something that must be researched further.

## 7 Monitoring and maintenance

Our results cannot be deployed in the same way as, for example, a machine learning model. The primary intent is for the data to be used by decision-makers to simplify the process of locating a deployment site. As a result, we have chosen to focus on the deployment of an underwater data warehouse, with a specific focus on the site selection process.

### 7.1 Key Performance Indicators

In this section, we will detail the key performance indicators aligned with the evaluation and recommendation metrics for the site selection of an underwater data centre, rather than the metrics used to manage or maintain the data warehouses directly.

### 7.1.1 Business KPIs

**Feasibility and cost-effectiveness** An important part is assessing the economic viability of potential sites.

**Energy efficiency potential** Evaluation of the ability of a site to leverage natural cooling effectively is very important. Due to the goal being offloading cooling capacity into water, the energy efficiency potential is heavily linked to the cooling efficiency at the site.

**Accessibility for maintenance** Ease of access for future maintenance and upgrades is another important factor. This is primarily based on how easy it is to get a maintenance crew out to the underwater data warehouse if something goes wrong. This specifically applies to mechanical faults, and not software faults that can be resolved remotely without physical access to the warehouse. This also applies for when the underwater data warehouse is scheduled for a hardware upgrade.

**Risk assessment scores** This KPI is tied to an analysis of geopolitical risks associated with each site. For example, deploying to an active war zone comes with considerable risk both during the deployment process, and in terms of potential sabotage or other damage to the underwater data warehouse caused by the conflict. Note that this primarily addresses current geopolitical risks, and not potential risks in the future.

### 7.1.2 Model performance KPIs

Note that the model in this case is our list of criteria for a good site. The performance of the criteria is heavily linked to a few key factors surrounding deployment.

**Number of site recommendations confirmed by experts** As mentioned in Section 6.1.1, humans have to verify the data prior to a deployment, to ensure that the recommendations are good deployment candidates. Therefore, the performance of the criteria is directly tied to its false positive rate, meaning the rate of site recommendations that turn out to be bad due to missing factors. A certain false positive rate has to be expected with any model, particularly one facing factors that cannot be represented in data, such as legal barriers, power access, and others.

**Minimising environmental impact of deployment** The purpose of the underwater data warehouses is to use energy more sustainably. Doing so should not be at the expense of the environment in some other way, meaning an important aspect of the good site criteria should be to prefer sites that minimise other types of environmental impact.

For example, this includes ensuring that areas with vulnerable species remain untouched, and otherwise ensuring that the deployment does not negatively affect the local ecosystem.

**Conflict minimisation** The analysis, with its results and corresponding sites shall be conflict minimising. Conflicts of any kinds are costly and not of interest, therefore it is crucial to minimise conflicts. The conflicts could occur in the form of placing the warehouse in a no-go zone or meeting resistant from the local human population.

## 7.2 Possible deployment faults

There are several faults that may occur during the deployment. These faults might be in relation to the environment, technical-, legal- or stakeholder risks.

### 7.2.1 Environmental risks

One of the major potential faults with deploying based on our analysis is regarding the accuracy of the no-go metrics.

In order to reduce conflict with other marine users, it is important that the no-go metrics represent as many stakeholders that need to be addressed as possible. However, not all these use-cases have data associated with them, presenting the potential for a deployment fault.

This is particularly important if the deployment fails to account for other potential environmental concerns surrounding the site. We have aimed to exclude, for example, nature reserves from the site list, but there may be other environmental factors that are not accounted for. This results in a fault if this also goes past human intervention undetected.

### 7.2.2 Technical risks

As previously mentioned, we were unable to include sediment analysis in our analysis. As a result, the sediment analysis has to be performed separately prior to deployment to ensure the data centre can be deployed on stable ground. Without human intervention, or in the event of an oversight, a massive fault in the data centre deployment would be failing to properly account for sediment types.

### 7.2.3 Legal and regulatory barriers

The project could run into legal or regulatory problems that delay or completely prevent it from being deployed. For example the underwater data warehouse would need to be in line with the marine resource act upheld by the Norwegian Ministry of Industry and Fisheries (Kommunal- og distriktsdepartementet, 2018). The law mostly relates to the process of extracting resources from the ocean and how this can be done fairly. However, it also includes sentiments about maintaining ecosystems which can become important for this topic, both during deployment, but also in how it could impact the environment over time.

### 7.2.4 Community and stakeholder opposition risks

A major risk involved in the deployment is if the local community objects to the deployment. This can happen for a number of different reasons, but the consequences could potentially hinder the project as a whole. Considering the project would likely bring industry to whichever area it is deployed in, this is unlikely but still a risk to consider.

The stakeholder risk is equally probable, considering the lack of research on the topic, the investment could be deemed to costly. There should at least be concerns regarding the projects profitability. Most concepts on this topic are relatively new, therefore raising questions regarding its longevity and benefit could hinder and delay deployment.

## 7.3 Alarms and human intervention

Alarms are not particular useful for this kind of project, because our project is not an AI-model, and cannot be deployed on its own. However, the project would greatly benefit from human intervention, in order to ensure that the recommendations are actual candidates for a deployment. Human intervention can be of great value, since such intervention can detect flaws or missing information that is crucial for either the analysis and/or deployment. The human intervention can take either of the two forms: intervention of results and analysis, or on qualitative information that have not been taken into account in the analysis. The latter can contribute to ensuring the inclusion of factors that cannot be trivially covered in data, such as legal and regulatory barriers, or other concerns that affect the deployment.

Human intervention is considerably more applicable, due to the important role of humans in quality assurance of the data. Human intervention has been described in more detail in Section 6.1.1.

## 7.4 Lessons learned

Throughout the course of the project, we learned a lot about geospatial analysis, and the tools and data involved in them. As a whole, our group was inexperienced with this type of data science project, which was initially why we found the topic interesting. While working with geographical data, familiar procedures like preprocessing were very different. We had to learn about new data formats, both in terms of file types, the CRS and projection types. The preprocessing of the geographical data in preparation for the analysis took a significant amount of the project's time, largely due to many of these differences.

We also learned how to approach data science projects in a systematic and organised way. Methodologies like CRISP-DM have been utilised to support the planning phase for this project. This was quite useful as site selection heavily differs from machine learning approaches. By relying on a framework, we could navigate the new process with less effort.

Additionally, we learn about geographical information systems like QGIS and ArcGIS, both how the software functions and how they can be utilised in a data science project. We learnt more about visualisations using maps and geographical data layers.

We have learned to collaborate with outside partners to achieve satisfactory results. This has allowed us to leverage their expertise and perspectives, enhancing our productivity and the final result.

# Bibliography


Arron Beecham, Tony Gregory, Jonathan Michael. (2015). *Wind Farms: 'I'm a Big Fan!'* https://uwe.maps.arcgis.com/apps/MapJournal/index.html?appid=e1533caf14d541b48f7db92c0380dd7a

AWS. (2023, November). What is Geospatial Data? - Geospatial Data. https://aws.amazon.com/what-is/geospatial-data

Barents Watch. (n.d.). *Fiskeriaktivitet [Dataset]*. Retrieved October 4, 2023, from https://www.barentswatch.no/fiskeriaktivitet/

CTM Magnetics. (2016, August 5). *Air Cooled vs. Liquid Cooled: The Differences*. https://ctmmagnetics.com/general/air-cooled-vs-liquid-cooled-differences/

Esri. (2023a). *An overview of the interpolation toolset*. Retrieved August 12, 2020, from https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/an-overview-of-the-interpolation-tools.htm

Esri. (2023b). *Slope (spatial analyst)*. Retrieved October 12, 2020, from https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/slope.htm

Fiskeridirektoratet. (n.d.). *Korallrev forbudsområder*. Retrieved October 28, 2023, from https://portal.fiskeridir.no/portal/apps/webappviewer/index.html?id=bed4faeea84e4d6dbc548d43cc134c96&layer=Korallrev+forbudsomr%C3%A5der

GeoNorge. (n.d.-a). *Adminstrative enheter fylker [Dataset]*. Retrieved October 4, 2023, from https://kartkatalog.geonorge.no/metadata/administrative-enheter-fylker/6093c8a8-fa80-11e6-bc64-92361f002671

GeoNorge. (n.d.-b). *Naturvernområder [Dataset]*. Retrieved October 4, 2023, from https://kartkatalog.geonorge.no/metadata/naturvernomraader/5857ec0a-8d2c-4cd8-baa2-0dc54ae213b4

GeoNorge. (n.d.-c). *Norske fylker og kommuner illustrasjonsdata 2020 (klippet etter kyst) [Dataset]*. Retrieved October 4, 2023, from https://kartkatalog.geonorge.no/metadata/norske-fylker-og-kommuner-illustrasjonsdata-2020-klippet-etter-kyst/7408853f-eb7d-48dd-bb6c-80c7e80f7392

GeoNorge. (n.d.-d). *Sjøkart - Maritim infrastruktur [Dataset]*. Retrieved October 7, 2023, from https://kartkatalog.geonorge.no/metadata/sjoekart-maritim-infrastruktur/a894ea02-d2dc-4550-ac3e-49230ceed42a

Hance, J. (2003). *Development of a database and assessment of seafloor slope stability based on published literature*. University of Texas at Austin. https://books.google.no/books?id=cjOHuAAACAAJ

Hotz, N. (2023). What is CRISP DM? - Data Science Process Alliance. *Data Science Process Alliance*. Retrieved November 6, 2023, from https://www.datascience-pm.com/crisp-dm-2

Interaction Design Foundation - IxDF. (n.d.). What is design thinking? Retrieved November 25, 2023, from https://www.interaction-design.org/literature/topics/design-thinking

Kartverkets Sjødivisjon. (2022). *Dybdedata fra kartverket*. Retrieved October 12, 2020, from https://www.mareano.no/kart-og-data/dybdedata

Kommunal- og distriktsdepartementet. (2018). H-6/18 Lover og retningslinjer for planlegging og ressursutnytting i kystnære sjøområder. *Regjeringen*. Retrieved November 25, 2023, from https://www.regjeringen.no/no/dokumenter/lover-og-retningslinjer-for-planlegging-og-ressursutnytting-i-kystnare-sjoomrader/id2616581/?ch=13

Kystinfo. (n.d.). *Kystinfo [Dataset]*. Retrieved October 4, 2023, from https://a3.kystverket.no/kystinfo

Malczewski, J., & Rinner, C. (2015). *Multicriteria Decision Analysis in Geographic Information Science*. Springer. Retrieved November 6, 2023, from https://link.springer.com/book/10.1007/978-3-540-74757-4

Microsoft. (n.d.). Project natick. Retrieved November 6, 2023, from https://natick.research.microsoft.com

Norges vassdrags- og energidirektorat. (n.d.). *Nedlastning av fagdata fra NVE [Dataset]*. Retrieved October 5, 2023, from http://nedlasting.nve.no/gis/

Norwegian Datacenter Industry. (2023, November). Networks and Connectivity in Norway. Retrieved November 25, 2023, from https://www.datasenterindustrien.no/connectivity

Rediske, G., Burin, H., Rigo, P., Rosa, C., Michels, L., & Siluk, J. (2021). Wind power plant site selection: A systematic review. *Renewable and Sustainable Energy Reviews*, *148*, 111293. https://doi.org/https://doi.org/10.1016/j.rser.2021.111293

Roach, J. (2023). Microsoft finds underwater datacenters are reliable, practical and use energy sustainably. Retrieved September 14, 2023, from https://news.microsoft.com/source/features/sustainability/project-natick-underwater-datacenter/

Rousi, H., Peltonen, H., Mattila, M., & Back, S. (2011). Impacts of physical environmental characteristics on the distribution of benthic fauna in the northern baltic sea. *Boreal Environment Research*, *16*. Retrieved August 10, 2023, from https://www.borenv.net/BER/archive/pdfs/ber16/ber16-521.pdf

Sibson, R. (1981). A Brief Description of Natural Neighbor Interpolation. In *Interpolating multivariate data* (pp. 21–36). John Wiley & Sons.

Statistics Norway. (n.d.). *Electricity*. Retrieved October 23, 2023, from https://www.ssb.no/energi-og-industri/energi/statistikk/elektrisitet

Statnett. (2022a). Omfattende forbruksplaner fyller opp kapasiteten i strømnettet i nord. Retrieved November 12, 2023, from https://www.statnett.no/om-statnett/nyheter-og-pressemeldinger/nyhetsarkiv-2022/omfattende-forbruksplaner-fyller-opp-kapasiteten-i-stromnettet-i-nord/

Statnett. (2022b). Increased consumption results in a negative norwegian energy balance from 2027. Retrieved November 12, 2023, from https://www.statnett.no/en/about-statnett/news-and-press-releases/news-archive-2022/increased-consumption-results-in-a-negative-norwegian-energy-balance-from-2027/

Tamblyn, T. (2018). Microsoft has dropped a data centre to the bottom of the ocean - here's why. Retrieved November 4, 2023, from https://www.huffingtonpost.co.uk/entry/microsoft-has-dropped-a-data-centre-to-the-bottom-of-the-ocean-heres-why_uk_5b17a224e4b0734a9938fc38

Viseth, E. S. (2022). Statnett: All nettkapasiteten i nord er allerede reservert. Retrieved November 12, 2023, from https://www.tu.no/artikler/statnett-all-nettkapasiteten-i-nord-er-allerede-reservert/521751

Wątróbski, J., Jankowski, J., Ziemba, P., Karczmarczyk, A., & Zioło, M. (2019). Generalised framework for multi-criteria method selection. *Omega*, *86*, 107–124. https://doi.org/https://doi.org/10.1016/j.omega.2018.07.004

Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*. Retrieved November 6, 2023, from https://www.researchgate.net/publication/239585378_CRISP-DM_Towards_a_standard_process_model_for_data_mining

Zhang, A., Lipton, Z. C., Li, M., & Smola, A. J. (2023). *Dive into deep learning*. Cambridge University Press. https://D2L.ai

Zhang, M. (2022). *Underwater Data Centers: Servers Beneath the Surface*. Retrieved November 10, 2023, from https://dgtlinfra.com/underwater-data-centers-servers/

Zhang, X., Lindberg, T., Xiong, N., Vyatkin, V., & Mousavi, A. (2017). Cooling energy consumption investigation of data center it room with vertical placed server [8th International Conference on Applied Energy, ICAE2016, 8-11 October 2016, Beijing, China]. *Energy Procedia*, *105*, 2047–2052. https://doi.org/https://doi.org/10.1016/j.egypro.2017.03.581