Ref. Ares(2025)4647378 - 11/06/2025

Project Acronym:	SEADOTs
Grant Agreement number:	101156488 (HORIZON-CL5-2023-D6-01-03 – HORIZON Innovation Actions)
Project Start Date	1st September, 2024
Project Full Title:	Social-Ecological Ocean Management Applications with Digital Ocean Twins





Funded by the Horizon Research & Innovation Programme part of the EU Missions of the European Union

DELIVERABLE

D1.1 – SEADOTs list of identified datasets

Dissemination level:	PU -Public
Type of Document	Deliverable
Contractual date of delivery	06.2025
Deliverable Leader	SINTEF OCEAN
Status - version, date	Draft – v0.2, 12/2/2025
WP / Task responsible	WP1 / SINTEF Ocean
Keywords:	Ocean data, social-ecological data, geospatial data, qualitative data, data inventory, metadata, socio-economic data

This document is part of a project that has received funding from the European Union's Horizon Europe Research & Innovation Programme part of the EU Missions under Grant N° 101156488. It is the property of the SEADOTS consortium and shall not be distributed or reproduced without the formal approval of the SEADOTS Management Committee. The content of this document reflects only the authors' view. The European Commission is not responsible for any use that may be made of the information it contains.

Executive Summary

SEADOTs (Social-Ecological Ocean Management Applications using Digital Ocean Twins) has the objective of advancing holistic, just and sustainable ocean management by bringing a predictive component for social-ecological aspects into comprehensive digital ocean twins (DOTs). These DOTs will combine digital twins of the ocean (DTO) with human activities in the ocean and combine socio-ecological and socio-economic data with ocean data, ecosystem data, and a variety of models. By creating and demonstrating applications in the Norwegian North Sea, the German North Sea and the Baltic Sea that address current challenges and developments and can simulate the intricate interactions between human activities and marine ecosystems, SEADOTs aims to facilitate and inform political decision making, marine spatial planning and adaptive management. SEADOTs' ambition is to help safeguard ocean ecosystems, promote sustainable resource use, and enhance social and economic well-being.

The project will leverage developments from ongoing Mission and Green Deal projects partners are involved in, including the European Digital Twin projects Iliad and EDITO, OLAMUR and CLIMAREST and demonstrate Ocean Management Applications with Digital Ocean Twins on the EU Digital Twin Ocean (DTO) infrastructure as well as distributed platforms for socio-ecological, socio-economic and political endpoints. For that purpose, SEADOTs will work with data acquisition beyond the state of the art. The objective is to provide spatially-explicit social-ecological data and data interoperability with geospatial ocean data also after the project period in suitable repositories, through stakeholder capacity building, and through collaboration with the co-funded projects of this call.

The project relies on the use of the Social-Ecological Systems (SES) framework, first developed by Elinor Ostrom, for describing each case study's social-ecological system. In SEADOTs, the framework is used for selecting the general data streams (e.g., ecological, social, and economic) to include in demonstrator DOTs. The methodology for collecting data for the data inventory was therefore shaped by defining the three case studies' social ecological systems. Thus, the three case studies were defined following the SES framework prior to the collection of data for the data inventory. Stakeholder mapping was also initiated in parallel to the creation of the data inventory to ensure that the needs of DOT end users were accounted for when identifying relevant data sets. Indeed, bringing local stakeholders the information needed for data-driven decision making is the central focus of SEADOTs. Therefore, stakeholders inform the collection and processing of data (e.g., creation of indicators) for the development of each DOT to ensure their usability and relevance of applications.

A core element of the SEADOTs project is the development of applications which are integrated into DOTs, which require the use of data from a range of sources including statistics agencies, ocean monitoring programs, and data collected *in situ*. As the collection of data by sensors was not in the scope of SEADOTs, data outlined in the inventory on oceanographic/physical variables is sourced from reliable European, national, or regional sources (e.g., Copernicus, EMODnet). Furthermore, economic data is largely sourced from national statistical agencies (e.g., Statistics Norway) to ensure the quality and standardization of data to be used in DOTs.

To be included in the data inventory, all data sets were assessed for quality and applicability to each DOT. Here we outline the data quality assessment framework used for prioritizing the use of datasets in applications (based on the ISO 19157 criteria for geospatial data quality assessment). Further, we outline mitigation strategies for including potentially relevant datasets which may not be usable in their original format (i.e., through processing such as the creation of indicators or adaptation for use in the SES framework). These mitigation strategies include standardizing data for use within the EU interoperability framework to allow for ingestion of data into digital twins.

The data inventory provides a prioritized list of data sources for each of the three SEADOTs case studies based on a quality assessment framework that accounts for not only the data source, but applicability and usability of each data set. As SEADOTs focuses on the inclusion of social-ecological data, the inventory includes data from diverse disciplines and requires the adaptation of existing quality assessment frameworks. The data inventory is

a living document as the prioritization and expansion of data sources will change as each case develops and the development of applications for each case study progresses. Inventory creation was initiated in the first stages of the SEADOTs project.

This deliverable is connected to D6.1, the data management plan (and D6.2, which is the final version of the DMP), which provides detailed information on how data will be handled within SEADOTs (e.g., according to FAIR and TRUST principles), D2.1, Description of the SES systems (and the final version D2.4), which was needed to be initiated prior to identifying datasets to inform which data may be relevant to each case study, D1.2, interoperability framework, due to the importance of FAIR data use within SEADOTs and for making data usable within the framework of digital twin development, D3.1, requirements and use case documents, as the applications and requirements for applications developed in SEADOTs is partially shaped by the data sets available for each case study, D1.3, data catalogue, which will be the final version of data used in applications made openly available to users (following FAIR and TRUST principles), D3.2, decision support algorithm, as the decision support tools developed in SEADOTs require the use of the data collected in the data inventory, D3.3, SEADOTs DTO framework documentation, as the data and applications developed in SEADOTs will be used within the EU DTO (and must follow the EU interoperability framework), and D3.4 execution of demonstrators and evaluation of results as the creation of the demonstrator DOTs require the use of the data inventory.

WP1 in SEADOTs (of which this deliverable is an output) addresses the first two aspects of the interoperability framework, technical and semantic interoperability, while WP6 and the data management plan aim to ensure the other two, organisational and legal interoperability. To establish trust by users of DOTs and their underlying data, information, and knowledge for decision making, one has to establish trust in technology and data, through transparency about data provenance, including the processes applied to the data (like outlier removal and anonymisation), and applied models and AI support.

Deliverable Leader:	SINTEF OCEAN
Contributors:	Lara Veylit, Piotr Zaborowski, Ute Brönner, Javier Arturo Cantillo Acosta, Katherine Crosman
Reviewers:	OGC, SU
Approved by:	Piotr Zaborowski, Emilie Lindqvist

Document History			
Version	**Date**	**Contributor(s)**	**Description**
0.1	2024-11-01	Lara Veylit	Initial document from template
0.2	2025-02-12	Piotr Zaborowski, Ute Brönner	First content from task description, WP2 & WP3 case descriptions, interoperability standards
0.3	2025-05-26	Lara Veylit, Javier Arturo Cantillo Acosta, Katherine	Information on datasets, figures, text on stakeholder engagement work

		Crosman, Vibeke Stærkebye Nørstebø	
0.4	2025-06-10	Lara Veylit, Piotr Zaborowski, Emilie Lindqvist, Ragnar Árnason	Integrated feedback from internal reviewers

Table of Contents

Executive Summary 2
Table of Contents 5
Table of Figures 7
List of Tables 8
Definitions, Acronyms and Abbreviations 9
Introduction 11
Objectives of the deliverable 11
Overview of the three case studies 12
Norwegian North Sea 12
German Bight 13
Gotland 13
Data inventory creation 14
Data quality assessment framework 14
Identified datasets 17
Social data streams 17
Survey data 17
Policy documents 17
National statistics 18
Interview data 18
Micro-social data 18
Ecological data streams 19
Agent based modelling outputs 19
Catch and landing data 19
Biodiversity assessments 19
Economic data streams 19
Multi-national databases 20
National / country-specific statistics 20
Other data sources 21
Oceanographic/physical variable data streams 22
Ocean models 22
Sediment mapping 22
Hydrological data 22

Weather data / atmospheric data 22
Human impact and use data streams 22
Marine traffic 23
Wind park placement 23
Infrastructure 23
Data integration, transformation, and interoperability 24
Modern data management practices 25
FAIR, CARE, and TRUST principles 25
GDPR 25
Challenges for making data interoperable 25
Integration of Data into DOTs and DTO 26
Mapping users and user needs for development of DOTs 26
Case study DOT users 26
Users from the EU DTO 27
Conclusions 28
References 29
Appendix 30

Table of Figures

Figure 1 High-level overview of data use through the SEADOTs project. 11
Figure 2 Mapping of the ISO19127 framework to the SEADOTs data quality assessment framework, and proposed mitigation measures to ensure data usability in applications and DOTs. 16
Figure 3 Overview of interoperability framework. 24
Figure 4 The central role of local stakeholders in driving the development of the applications and DOTs. Stakeholder needs additionally inform the use of data streams (e.g., datasets used in applications and DOTs) as well as the analyses and processing of data. 27

Definitions, Acronyms and Abbreviations

Acronym/Abbreviation	Title
CARE principles	Data management principles for managing data from indigenous peoples. These principles are: Collective benefit, Authority to control, Responsibility, and Ethics.
CF metadata conventions	The CF metadata conventions are designed to promote the processing and sharing of files created with the NetCDF API. The conventions define metadata that provide a definitive description of what the data in each variable represents, and the spatial and temporal properties of the data.
CSV	Comma separated values, common file format, human readable
DOT	Digital Ocean Twin
DSS	Decision Support System
EDITO	EDITO (a European funded project) serves as the EU public infrastructure for the EU DTO.
EU DTO	The European Digital Twins of the Ocean initiative.
FAIR principles	Modern data management principles related to the sharing of data outlining that data should be Findable, Accessible, Interoperable, and Reusable.
GDPR	The General Data Protection Regulation is a data protection law enforced across the European Union (EU) and European Economic Area (EEA).
NetCDF	NetCDF is a set of software libraries and self-describing, machine-independent data formats that support the creation, access, and sharing of array-oriented data. The format is established in the climate forecast and ocean community, and they have established their own *CF Metadata Conventions.*
SES	Social-ecological system
SES framework	The social-ecological systems framework as outlined by Elinor Ostrom can broadly be described as a multidisciplinary conceptual framework which describes how physical, environmental, and social systems interact to effect resources.
Social ecology	Social ecology is the study of how individuals interact with and respond to the environment around them, and how these interactions affect society and the environment as a whole.
TRUST principles	The TRUST principles for data management promote responsible data use. The five principles are: Transparency, Responsibility, User-focused data use, Sustainable use of data, and the employment-appropriate Technology.
WP	Work Package
Zarr	Zarr is a cloud-friendly format for storage of chunked, compressed n-dimensional arrays, supported by various backends, making it a versatile choice for scientific and large-scale data.

Introduction

Objectives of the deliverable

SEADOTs aligns and connects with the European Digital Twin Ocean (EU DTO) The vision of the EU DTO is to make ocean knowledge readily available to citizens, entrepreneurs, scientists, and policy makers. The objective is to provide them with an innovative set of user-driven and interactive tools, empowering them to collectively share responsibility for marine and coastal habitat management and act on their restoration, to support a sustainable blue economy and to mitigate and adapt to climate change, fostered by digital transformation for political and collective decision making.

To achieve these aims, data must be shared with EU DTO in a way that supports this vision and goals. In SEADOTs, we adhere to the FAIR principles for data sharing, i.e.

- Data is findable through use of catalogues
- Data is accessible through use of suitable repositories, formats and APIs
- Data is interoperable through common data models and semantic metadata enhancement
- Data is reusable through the use of suitable licenses

Task 1.1 and Deliverable D1.1 create a data inventory of the required data to implement the three SEADOTs demonstrator cases in Norway (Utsira), Germany (German Bight), and Sweden (Gotland). Each data set will be identified with respect to data availability, data provenance, data FAIRness and interoperability according to the criteria above. Additionally, suitable TRUSTed repositories will be identified for data sharing beyond the project's duration, and CARE principles applied and described where applicable.

Data streams that encompass ecological, economic, and social dimensions of each case (German Bight, Utsira, Gotland) will be fed into models or directly used within the Social Ecological Systems (SES) framework (as defined by Elinor Ostrom, see Ostrom 2009). Processed data, data transformed to comply with interoperability standards, and data integrated across domains (i.e., social, ecological, economic) will be made available through EU data infrastructure (e.g., EMODnet Human Activities where applicable). These transformed data streams will then be used in the development of applications integrated in Digital Ocean Twins (DOTs-see Brönner et al. 2023), which will be made more widely available through the EU DTO through the infrastructure provided by EDITO. For a visual representation of the use of data through SEADOTs see Figure 1.

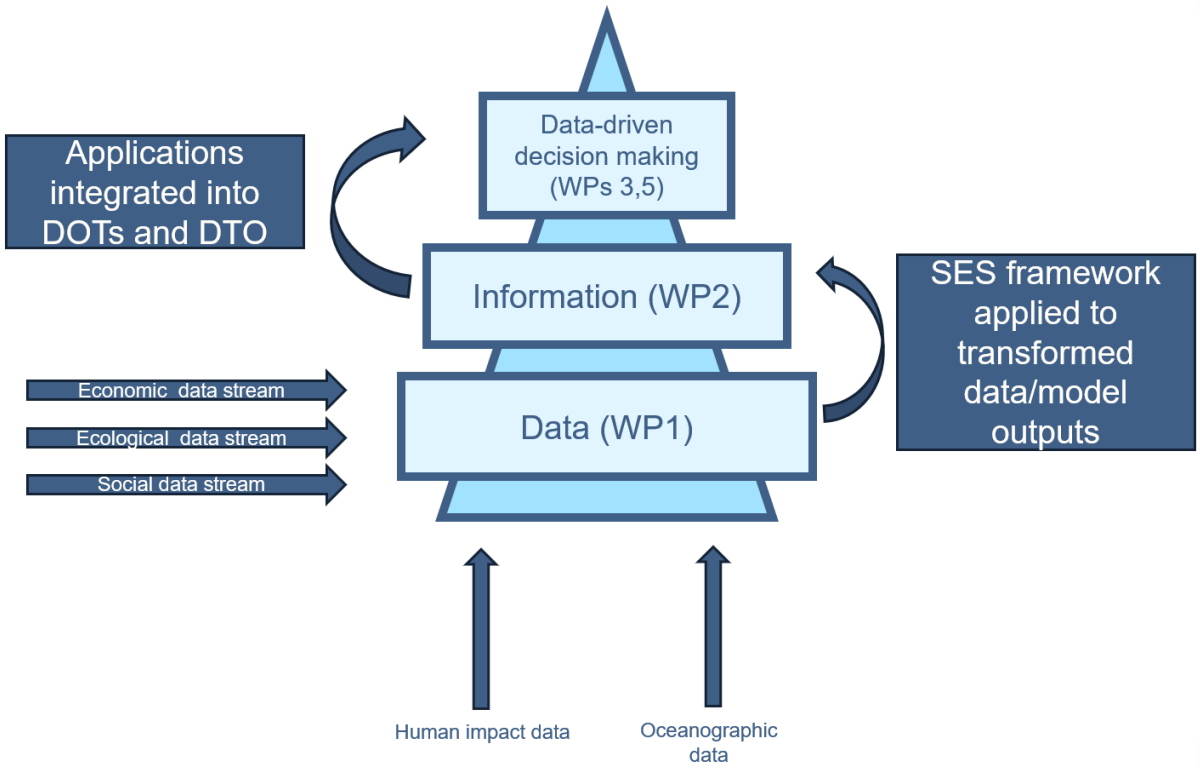


Figure 1 High-level overview of data use through the SEADOTs project.

Overview of the three case studies

Norwegian North Sea

The Norwegian North Sea case concerns the development of Utsira Nord, which is planned to be the world's largest floating offshore wind park. Offshore wind will enhance renewable power production, foster industry growth, and drive technological development and innovation. A floating offshore wind project at Utsira Nord will be pivotal in Norway's offshore wind strategy to meet the government's target of designating area for 30 GW by 2040. The case is of social interest due to it representing a prestige project for Norway as it is both expected to be of low ecological impact (i.e., it will not interfere with fishing activities, expected to have mitigation strategies in place for impacts to seabird populations) and high social impact (i.e., through technological development and increases to employment in the area). The park is expected to be developed in Q2 of 2025 but may be postponed (there may be delays depending on the finalization of the subsidy model and other regulatory requirements). This case study will rely heavily on **social data streams** into the DOT that will be developed in SEADOTs. These data streams include existing datasets but will also encompass datasets collected in the SEADOTs project, including **micro-social data** on changing local perceptions of the wind park as the development process unfolds.

German Bight

The German Bight case is generally focused on the integration of low-trophic aquaculture in an existing wind park located near Helgoland, Germany. Located about 20 km from Helgoland, the existing wind park is maintained by Helgoland port. The park has been operating since 2015 and has initiated wind park tourism. The area is being explored for multi-use by low trophic aquaculture. The park is associated with artificial reefs, and it is planned to incorporate aquaculture in the area of the wind park. Challenges include technical challenges (to aquaculture), and regulatory hurdles for multi-use of the area. Wind farms may also have effects on sedimentation on the coast by reducing wind and wave intensity. The wind farm replaced fishing in this area. Test installations for mussel production are underway. Aquaculture is expected to improve the ecological footprint of renewable energy production by increased efficiency of use of space, removal of excess nutrients, and forwarding environment-friendly food production. Since a great portion of the energy produced in the region is exported by the large energy companies, aquaculture production may increase the share of local businesses such as aquaculture producers, maintenance industry and tourism. This may positively affect local business and welfare and support acceptance of offshore wind energy production by coastal dwellers. This case will focus most on the integration of **economic data** into the case's DOT due to the importance of evaluating the economic viability of integrating low-trophic aquaculture into the existing wind park.

Gotland

The main interest of the Gotland case (located in the Baltic sea, off Sweden's largest island) lies in the trade-offs and synergies between pelagic fisheries and offshore windfarms (see Blomquist et al. 2024, ICES Acoustic data portal [accessed 2025-06-09]). In the Swedish Baltic Sea, commercial fishing mostly targets pelagic fisheries as Gotland's surrounding waters are relatively abundant in herring and sprat (ICES Acoustic data portal). In this area, commercial fishing is mainly conducted by Swedish fishing fleets, but other nationalities (e.g. Finnish, Polish, Danish, German) also operate (Blomquist & Waldo 2024). Furthermore, several zones have been designated of specific interest for future offshore wind energy development (Ibid.). In this case, several scenarios will be assessed to explore the impacts of various technical characteristics of wind parks (e.g. size and density), the interactions with marine spatial plans (e.g. increased marine protected areas, moving trawling limit) and the cumulative impacts of climate change (e.g. through marine heatwaves) on fisheries. Moreover, national regulations under consideration and the political landscape highly impact wind park development. For instance, areas designated for energy production can face trade-offs with national security. Overall, the Gotland case will have a special focus on the use of socio-ecological data streams in the DOT.

Data inventory creation

In milestone 3, each case study collected a list of potentially useful data sources that may be used in the development of applications and DOTs. Those involved in each case study were asked to provide

a list of any available data, to create a comprehensive overview of the current state of data available for each local area (e.g., German Bight, Norwegian North Sea, Swedish Baltic). For all case studies, data on social, economic, ecological, and oceanographic factors is included in this overview. These datasets range in usability, with some datasets in ready to use formats (e.g., csv, netCDF) with other datasets being found in less usable formats (e.g., pdf). Furthermore, datasets range in their local (e.g., only covering the local area under consideration) versus general (e.g., covering an entire country) spatial coverage. Finally, some datasets may require processing for addressing the specific needs of the SEADOTs applications; i.e., where data is not entirely fit-for-purpose, but still applicable, or in the creation of indicators. These raw datasets also may require processing for interoperability. The initial and comprehensive data inventory can be found in the Appendix. This data inventory is a living document, as the case studies develop new data sets may be identified as relevant or may become available over the course of the project. Furthermore, not all data that are needed may be available or have been collected. Therefore, there may be data gaps which may or may not be filled over the course of the project.

Data quality assessment framework

To account for the diversity of types of datasets to be included in the applications developed in SEADOTs, data was assessed using several criteria based on existing frameworks as well as an applicability assessment framework for dataset inclusion in applications. “Data quality” in SEADOTs extended beyond assessing the source and format of data to also include more generally how representative datasets were for each case study (i.e., geographically and in terms of socio-demographic factors).

Each dataset was subject to a primary quality assessment based on dataset source. SeaDataNet, EMODnet, and Copernicus already apply robust quality assessment frameworks to datasets (and metadata) before ingestion. However, in the case of datasets not found in similarly quality assured infrastructures, data-type specific quality assessment frameworks were used for the primary assessment of datasets (e.g., ISO 19157 criteria for geospatial data quality assessment).

To account for the heterogeneity of disciplines represented in SEADOTs, an applicability data assessment framework was developed based broadly on the ISO 19157 criteria for geospatial data quality assessment to allow for the use of data in data spaces (see Noardo et al. 2024). The criteria provided by ISO 19157 are:

1. **Completeness**: Assessing whether all required data is present
2. **Logical Consistency**: Ensuring data adheres to defined rules and structures.
3. **Positional Accuracy**: Measuring the accuracy of the spatial location of features.
4. **Thematic Accuracy**: Evaluating the accuracy of attribute data.
5. **Temporal Quality**: Assessing the accuracy of temporal attributes and the temporal validity of data.
6. **Usability**: Determining the ease with which data can be used for specific purposes.

The quality assessment frameworks and applicability framework outlined below were used for developing a prioritization schema. The prioritization schema follows a "traffic light system", with the highest priority datasets being color-coded as green, medium priority datasets in yellow, and low priority datasets in red. The prioritization schema was applied to the comprehensive data inventory (i.e., provided in milestone 3) in D1.1 to inform the creation of applications in WP3. This framework will be used alongside existing standards, as where domain-specific standards exist, these will be used for the standardization of data.

In SEADOTs, the applicability framework reflects a decision-tree structure (i.e., if the dataset does not follow a criteria de-prioritization is considered, if the dataset does follow the criteria, then the next criteria is assessed):

1. **Relevance:** is the data relevant to the information of interest?
    a. Comprehensiveness of how data stream relates to or can be mapped to information of interest.

    b. Direct or proxy relation to the model requirements

2. **Representativeness**: does the dataset represent the target population/area?

    a. Social science data (e.g., surveys, interview data) should represent a whole population rather than a population segment (or be possible to combine with other datasets to capture the heterogeneity of populations). This is to reduce introducing biases to applications, and DOTs. Data representativeness applies for both data on local populations and on more general demographic data/statistics.

    b. Geographically explicit information should include or be tailored to a given case study. While high quality general information may be contextually relevant, or a data source may be of high quality for an area outside the local region, data included in the applications should be representative of the local area of interest.

3. **Reliability**: is the data reliable both in terms of scientific standards and accurate in terms of positional and temporal accuracy?

    a. The collection of all data used in applications should follow scientific standards (i.e., ideally representative rather than opportunistic sampling).

    b. Data should be generally accurate in terms of position and time i.e., GPS accuracy for sensor data, time stamps should be correct for observations.

4. **Temporal validity**: is the dataset of temporal relevance to the digital twin?

    a. Data should be relevant to the context of the digital twin. This may include historical contextual information, but ideally data is synchronized at a specific frequency.

5. **Ingestibility**: is the dataset useable in the raw or a processed format considering technical and ethical requirements?

    a. The data source should be useable according to both technical and ethical requirements.

b. Accordingly, personal data is anonymized to comply with i.e., GDPR. All social data should be useable in applications.

c. Data should comply with FAIR principles and follow technical interoperability to allow for reuse e.g., by other digital twins.

6. **Reusability in the digital twin framework**: does the data comply with the requirements for use in a digital twin?

   a. The data should be usable for the creation of a digital ocean twin. This includes following interoperability standards and use in the creation of "what if" scenarios.

This data quality assessment framework was used by each case area team to pre-assess the usability of datasets from the data inventory for the development of applications and DOTs.

For a visual representation of this framework as well as mitigation measures for making data usable even if data does not comply with the outlined criteria, as well as a mapping of the SEADOTs framework to the ISO 19157 criteria for geospatial data quality assessment, see Figure 2.

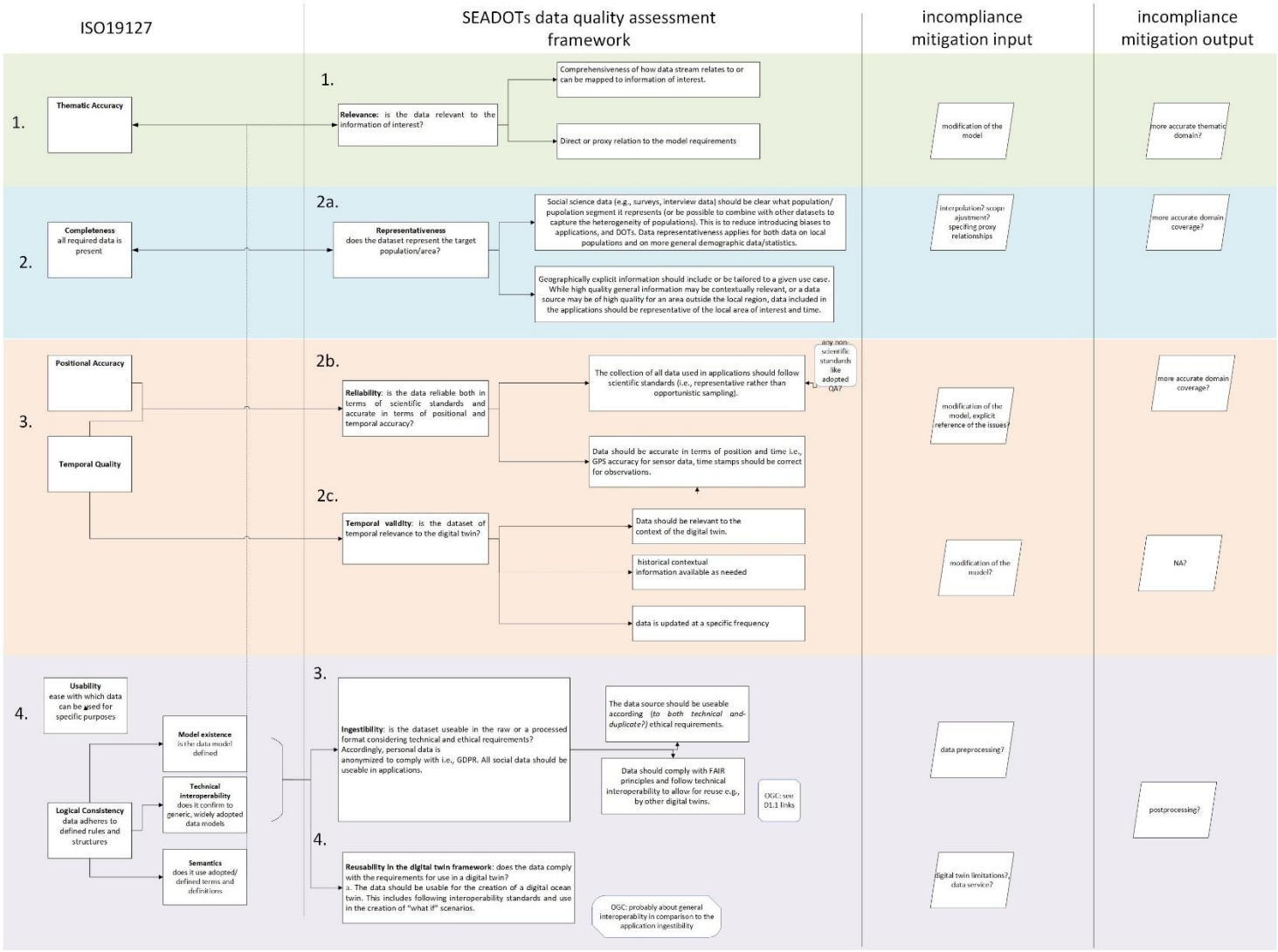


Figure 2 Mapping of the ISO19127 framework to the SEADOTs data quality assessment framework, and proposed mitigation measures to ensure data usability in applications and DOTs.

Identified datasets

All three case studies included datasets that would serve as social, ecological, and economic data streams feeding into applications. Additionally, data on human impacts on oceanographic/physical factors were necessary to include in the inventory due to their importance to developing DOTs. The nature and source of these datasets differed across the case studies based on the availability of data currently, the focus of the DOT to be developed for the case study (i.e., the use of social data is emphasized in the Utsira case, economic data integration in the German Bight case, and ecological data in the Gotland case), and the planned data collection activities for each case study.

Social data streams

To capture the social context and consequences of the developments in each use case, various kinds of social data streams will be used and processed to be useable within SEADOTs applications and DOTs (as well as through the EU DTO). As the majority of social data is in the form of text (e.g., policy documents, public perceptions, survey data, interview data) these data will be used either for providing contextual information or transformed into quantitative data streams to feed into DOTs (i.e., using sentiment analyses, Likert scaling).

Survey data

Data from standard survey items (closed-ended responses) can be nominal (also known as categorical, in which numbers represent unordered categories such as 'your favorite viewpoint'), ordinal (also known as ordered categorical, commonly seen in agree/disagree response scales), interval (ordinal data with equal spaces between values, such as acceptable temperature change expressed in degrees Celcius) or ratio (interval data with a true zero, such as number of wind turbines). Survey data are normally available in .csv or similar formats; they should be accompanied by metadata that provides information about the survey sample, items, administration, and any cleaning, manipulation, or weighting that has been done. Some surveys also include open-ended items, which elicit textual responses; these can be analyzed using qualitative coding, as described below (Policy documents). Survey data are useful to understand broad information from larger populations; for instance, survey data can capture public attitudes towards offshore wind at a certain point in time and be used to identify demographic predictors or correlates of those attitudes. Surveys are planned to be conducted in at least one case study area (Norwegian North Sea).

Policy documents

Policy documents are text documents. The raw text can be transformed into structured data through qualitative coding, in which a coding scheme that assigns topics of interest a numerical code is developed either deductively or inductively, and qualitative coders work through each document, highlighting text and assigning the relevant code(s) as appropriate. This method reveals the underlying structure of the data and allows for a better understanding of the conditions under which different topics of interest co-occur or connect to one another, as well as of the substance of the document. Coding of a policy document relevant to social impact assessment in wind turbine siting, for example, might include codes for the presence or absence of a requirement for social impact assessment, types of impact considered in any assessment – economic, non-economic, etc., whether and how public participation in assessment is structured, assessment funding sources, and so on, with each of these topics having the necessary subcodes to disaggregate relevant information, as well as other topics of interest. Policy documents are especially relevant to understanding governance of the case social ecological systems, where governance is included in the bounds of the case. All three use cases will

use policy documents to inform the creation of “what if” scenarios and to provide contextual information in DOTs.

National statistics

In all three case studies, national statistics will be used to estimate factors such as local population size and socio-demographic factors (e.g., age and gender distribution of local populations). These statistics are collected heterogeneously across Europe, however, so information on methodologies for collecting statistics are also being collected.

Interview data

Similar to policy documents, raw interview data are text files, normally de-identified transcripts of in-depth interviews on topics of interest. Transcripts are also qualitatively coded as described above (policy documents), with coding revealing data structure that may be particularly obscured by the conversational nature of data collection. Interviews are useful for exploring causal explanations, nuance, and detail – for instance, how a key informant perceives co-siting of wind and lower-trophic level aquaculture, the potential benefits and costs they see, and why. These data will be collected *in situ* or from secondary sources and processed for use in SEADOTs applications.

Micro-social data

In order to integrate near real-time social data into DOTs, micro-social data will be collected and processed through the use of sentiment analysis (see Khan et al. 2016). Micro-social data (coined by combining the concepts of "microdata" and "microsociology") refers to social data collected in near real-time from direct engagement with stakeholders. This will be done through a web application with simple questions to collect a time series on social perceptions of developments in the case studies (e.g., a response to “how do you feel about the development of Utsira Nord?” may be “since the Fred Ohlson consortium is no longer interested in the windfarm, I feel more negatively about the development of the wind park in Utsira”, which may serve as micro-social data alongside a time stamp of when the user recorded their response to the survey on the application). In SEADOTs, micro-social data will feed into the DOT from data collection through a web application to capture local perspectives on the wind park and subsequent sentiment analyses. The use of these web applications will be advertised in the local case studies to encourage user engagement and allow for the regular collection of data on perceptions of developments by local stakeholders. Micro-social data streams will be collected in the relevant case studies in SEADOTs and be processed for use in applications.

Ecological data streams

Data on the ecological context of each case study are included in the data inventory. These data are collected from secondary sources as the deployment of sensors in the case study areas is outside the scope of SEADOTs. Due to the key role of fisheries in the Gotland case and the historic reliance of the Utsira community on fisheries, there is a focus in these two cases on catch and landing statistics. In all three case studies, some information on biodiversity (either data collected in the local areas or from models) are included due to potential impacts of wind farm installations.

Agent based modelling outputs

The Gotland case will draw on both ecological and social data as inputs for an agent-based model representing the interactions between offshore windfarms and pelagic fisheries (sprat and herring), notably in the context of heatwaves. Ecological data will inform the growth rate, spatial distribution and stock size of the pelagic fish population. The model will also take as inputs water temperature and other bio-chemical data to inform the evolution of fish stocks. Social data will inform the modelling of vessel agents, including their technical characteristics and patterns of movement, as well as the market conditions, such as producer prices for fish or fuel prices. The model will produce time series data on the profitability of fishing fleets (e.g. evolution of costs and revenue), on the abundance and distribution of pelagic fish populations, as well as on the energy production and revenue associated with the offshore windpark.

Catch and landing data

Catch and landing data represent an intersection between ecological and economic domains. These data provide insights into species abundance and fishing pressure from an ecological perspective, and they also serve as foundational inputs for economic analyses—informing assessments of revenue generation, and the economic viability for instance.

Relevant data sources are Eurostat, Annual Economic Reports on the EU Fishing Fleet, ICES, Statistics Sweden, SLU (for Sweden), DataFangst (For Norway). These data sources are further described in the section “Economic data streams.”

Biodiversity assessments

To describe the impact of the three wind farms on the natural environment, it was essential to include datasets on local biodiversity to the data inventory for each case study. Biodiversity information is sourced by either international repositories (OBIS, HELCOM biodiversity database, LifeWatch ERIC, GBIF) or national sources (e.g., DOME for Swedish ecosystem-level data, MAREANO for sea floor biodiversity data). In addition to general information on biodiversity, data on specific groups (e.g., fish, birds) were collected based on the topic of each case study. Information on fish species were sourced from datasets related to the exploitation of fisheries (i.e., ICES, EMODnet fisheries data) as well as from ecological surveys (e.g., the Swedish SHARK database).

Economic data streams

To understand the economic dimensions and implications of developments in each use case, a variety of economic data streams will be collected, processed, and integrated into SEADOTs applications and DOTs. These streams will include indicators such as economic activity, employment and investments relevant to offshore energy, fisheries, and aquaculture, catch and landing data, prices etc. Data can be both quantitative and qualitative in form of text documents, as for the social data streams. Data will typically be used to support scenario analysis, economic impact assessments, and policy evaluations within the DOTs.

Multi-national databases

FAO, EUMOFA, EUROSTAT are databases relevant for all three cases in general:

- FAO, through the FishStat tool/database, provides country specific statistical data on global fisheries and aquaculture production. Data includes production (catches and harvest of species for different purposes) measured in weight terms available from the year 1950, global trade statistics (imports and exports) available from the year 1976, and statistics of apparent consumption organised by country and species group, available from year 1961. The data is aggregated per country, and not to the specific case studies.
- EUMOFA (European Market Observatory for Fisheries and Aquaculture Products) present yearly data of landings for European countries, including values in monetary terms (Euro) and Volume (kg). Data is disaggregated according to the main commercial species, the presentation type and the preservation. Data is available for different years, including recent data. Moreover, there is available weekly, monthly and yearly information on consumption, first sales, imports and exports with the same level of disaggregation.
- The International Council for the Exploration of the Sea (ICES) collects annual data on nominal catches of more than 200 species of fish and shellfish in the Northeast Atlantic region, in collaboration with the Statistical Office of the European Communities (EUROSTAT). Data is available from the year 1903 and is disaggregated by species, fishing area, and country, and the catches are published in tonnes of live weight.
- EUROSTAT has databases of the catches per species in fishing areas from 2000 and onwards and also includes historical data from 1950-1999. Values are measure in term of Tonnes of Live weight. There is also data on landings considering presentation form of the fish, fish destination and use and fish species. Values are expressed in monetary terms (EUR) and volume (tonnes of live weight). There is also available historic data on number of fishing vessels per country.

For Germany and Sweden cases, the 2024 Annual Economic Report on the EU Fishing Fleet published by the European Commission is relevant. It includes data on landings and value per species, effort (e.g. fishing days, days at sea, energy consumption), employment, capacity (e.g. mean age of vessel, total vessel power and tonnage), income, expenditure, capital (e.g. investment, subsidies, gross debt). Data is available from 2008 to 2024.

Additional comment:

Data of “Tourism” from Eurostat was not included in this section, as it is considered out of scope.

National / country-specific statistics

Some data is only relevant for one case-study/country and is typically found in national statistics.

Norway:

- For the Norwegian case, data in ‘DataFangst’ (https://datafangst.orcalabs.no/live) present aggregated monthly and yearly statistics of gears used, species harvested and vessel lengths for fisheries in Norway from 1999 to 2025. There is also disaggregated data by vessel/company, including name of vessel, date, travel time, delivered catch (including economic and weight values), fuel consumption, type of fishing gear.

- FAO presents weekly prices for more than 350 fisheries products across the European market, including the Norwegian market. Prices are disaggregated according to species, preservation method, product form, price source (first-sale, wholesale, etc).
- Various data on historical sector-wise economic activity, employment, and demographic can be found in Statistics Norway (www.ssb.no)

Additional comment:

Data in "Barentwatch/Coastal Administration" is related to vessels operations, identification of relevant areas (fairway areas, protected areas), speed limits, weather, maritime boundaries rather than of economic data. Most of the data is on real-time basis.

Germany:

There is no specific data for the German case in terms of economic data.

Sweden:

- The Swedish Agency Marine and Water Management has data on catch statistics since 1999, including data on catch per fish species and type of gear, as well as sold catch (first-hand value in SEK) per fish species. The information includes also the catch area, landing area and type of vessel.
- Statistics Sweden have available information on fishing activities in marine waters by commercial fishermen up until 2015. Data includes catches by fishing, quantities landed and first-hand value by month and year. Also, there is data about the number of vessels and vessel capacity per year.

Other data sources

Data from surveys, interviews and policy documents, as described in 'Social data sources' might also include relevant economic data.

Oceanographic/physical variable data streams

To describe how the physical environment interacts with each case study, data and model outputs will be integrated into each DOT. These data streams provide important contextual information that impact the local areas selected for the development of wind farms, the location of fisheries, and the possible integration of low-trophic aquaculture (i.e., in the German Bight case).

Ocean models

Several ocean models (e.g., NEMO, CROCO) exist that can provide detailed data on physical, chemical, and biological processes in the marine environment, supporting a wide range of applications and analyses. Examples of data these models can provide include temperature (SST), ocean currents and circulation patterns, salinity levels, nutrient concentrations (e.g., nitrates, phosphates). Data sets in the inventory include global, low-resolution outputs from ocean models (e.g., accessible through Copernicus such as European North West Shelf - Ocean Physics Analysis and Forecast for the German case study) as well as outputs from higher resolution models run for specific local areas (e.g., Baltic

Sea Physics Analysis and Forecast for Sweden, data from Norkyst800 which includes 1 hour operational ocean model data with forecasts).

Sediment mapping

To describe grain size distribution and sediment concentration, data from CMEMS (Copernicus Marine Service) are used, together with sediment maps. The CMEMS dataset provides high-resolution biogeochemical data for the Northwest Shelf at a spatial resolution of approximately 1 km and a daily temporal resolution. These data are particularly relevant for the German case.

Hydrological data

For the German case in particular, hydrological data will also be utilized. Historical river discharge data are obtained from the ´Abflussdaten´ database, while riverine nutrient loads are sourced from the ´European Riverine Nutrient Loads dataset.

Weather data / atmospheric data

Atmospheric data, including forecast variables, will be provided by the COSMO-EU model. This high-resolution regional weather model delivers data on meteorological parameters such as wind speed and direction, air temperature, precipitation, and atmospheric pressure, supporting both real-time applications and scenario analysis.

Human impact and use data streams

Each case study is subject to unique impacts by human use, which will be shown in the DOTs. Data streams on human impacts and use are particularly needed to inform the creation of “what if” scenarios and to provide contextual information in DOTs (e.g., other infrastructure that may impact the area selected for developing a wind park).

Marine traffic

There are different sources that will be used to get information on marine traffic. This is especially relevant for the Norwegian and Swedish case studies. Data sources include AIS data, Statistics Norway, EU Fleet register, and for fisheries specifically, Ministry of Fisheries and Global Fishing Watch organisation data.

Windpark placement

Exact data on wind park placement will typically be provided by the operator/owner of the wind park. They can also provide non-confidential information on physical characteristics of wind farms. In addition, public authorities (e.g., NVE the Norwegian water and energy directorate) have published documents describing possible and/or planned areas for offshore wind development.

Infrastructure

In addition to the placement of wind parks, it is also relevant to include information on other physical infrastructure that are in the same area in the DOTs. Relevant infrastructure includes for instance

aquaculture facilities, ports, location of cables and pipelines. Information can both come from surveys, interviews and documents, as well as from databases such as EMODnet Human Activities.

Data integration, transformation, and interoperability

Data collected from secondary sources and *in situ* will be integrated into applications in each case study DOT, and each DOT will be integrated into the EU DTO. Thus, both data collected in the project and datasets collected from secondary sources and transformed for use in DOTs will be shared openly through the EU DTO. Data sharing must also be in line with the European interoperability framework, especially when it comes to sensitive data like personal data, ethnicity, or political opinions (see Figure 3).



Figure 3 Overview of interoperability framework.[1]

Technical interoperability covers the applications and infrastructure that connect systems and services. It includes elements such as interface specifications, interconnection services, data integration and exchange, data presentation and secure communication protocols.

Semantic interoperability ensures that the precise format and meaning of exchanged data and information is preserved and understood throughout exchanges between parties, in other words 'what is sent is what is understood'.

Organisational interoperability refers to the way in which organisations align their business processes, responsibilities and expectations to achieve commonly agreed and mutually beneficial goals.

Legal interoperability is about ensuring that organisations operating under different legal frameworks, policies and strategies are able to work together. This might require that legislation does not block the establishment of European public services within and between Member States and Associated Countries and that there are clear agreements about how to deal with differences in legislation across borders, including the option of putting in place new legislation.

[1] European Interoperability Framework:
https://ec.europa.eu/isa2/sites/default/files/eif_brochure_final.pdf

Modern data management practices

FAIR, CARE, and TRUST principles

In order for data to be ingestible into the applications, DOTs, and the EU DTO, data should follow the FAIR and TRUST principles. Data that is not fit-for-purpose for use in applications and DOTs (e.g., not following the quality assessment framework outlined above) due to a lack of ingestibility may be transformed following the mitigation measures described in Figure 2. Data generated by SEADOTs and data transformed from its original format are planned to be included in the final **SEADOTs data catalogue** (D1.3) which will be made openly available at the end of the project. In addition, data will be openly available through the **SEADOTs Zenodo community** when possible (i.e., complying with ethical handling of personal data).

GDPR

It is planned that data from interviews and surveys will be collected in the case studies, which may include personal data. These data will be handed according to GDPR (as outlined in the data management plan) prior to being made available. The project coordination team has an open dialogue with Sikt, the Norwegian Agency for Shared Services in Education and Research, to ensure all personal data is handled correctly throughout the project lifetime and made accessible according to regulations such as GDPR.

Challenges for making data interoperable

Where the methods for collecting similar information differ, the integration of data will require either a lack of integration of data or processing of data (e.g., for the creation of indicators). For example, data collected from national statistical agencies may be collected following dissimilar methodologies, necessitating the use of standardized data (e.g., sources from EUROSTAT). However, standardized datasets may lose resolution – requiring in some instances datasets are not combined. Furthermore, where data is in a format where clear standards do not exist (e.g., policy briefs), some transformation/processing is required to make data quantitative for use following the framework for digital twins. One example of this process is the collection of social perception data on the wind park development in Utsira, where qualitative data will be transformed using Natural Language Processing methods into quantitative data. Background data may also be used to provide text as context for the use of the applications developed in SEADOTs and DOTs as not all qualitative data can (or should) be transformed into quantitative metrics.

Integration of Data into DOTs and DTO

The datasets selected for use in each DOT will be informed by the needs of both local stakeholders from the areas surrounding each case study and stakeholders from the EU DTO landscape. Local stakeholder groups each have distinct needs, which necessitated the use of stakeholder mapping to determine which needs to prioritize (based on a Power-Interest Matrix approach). EU DTO user needs were assessed based on engagement in arenas such as EU Ocean Days, One Ocean Week in Bergen, Norway, and in other Mission Oceans and Waters events. These arenas allowed for partners to discuss the needs of stakeholders from the EU DTO landscape (e.g., others developing digital ocean twins in EU funded projects, developers of infrastructure such as EDITO regarding interoperability of data and data products in the EU DTO).

Mapping users and user needs for development of DOTs

Case study DOT users

Partners leading each case study are in the process of developing case-specific stakeholder maps. Case stakeholders are those who can influence the specific case context (e.g., regulatory agencies with oversight of wind turbine siting), are likely to be affected by case developments (e.g., wind developers, coastal residents, fishers) and other interested parties (e.g., environmental non-profit organizations). Understanding case stakeholder landscapes is critical to developing applied understandings of case dynamics that are grounded in local expertise and responsive to local needs, as well as for identifying potential DTO users.

In SEADOTs, initial stakeholder maps apply a Power-Interest Matrix approach (Mendelow 1991), in which stakeholders are categorized along two intersecting axes according to the level of power they have to influence case developments (e.g., co-siting of wind and lower-trophic aquaculture) and the degree of interest they have in case developments. The result is a four-quadrant map that categorizes stakeholders into relative groups: high-interest/high-power; high-interest/low power; low-interest/high power; and low-interest/low-power. These matrices will be aligned across cases to identify overlapping stakeholders/clusters of stakeholder types. In conjunction with the second SEADOTs General Assembly (October 2025), the aligned stakeholder maps will be collated and extended by the consortium as a whole to define a set of potential users of the DOTs (e.g., public decision-makers, regulatory agencies, industry organizations).

Stakeholder mapping connects to Tasks 2.1 (description of the social-ecological system of each case), 2.2 (development of SES conceptual models), 2.3 (development of SES indicators), 5.1 (stakeholder engagement and dissemination), and 5.3 (capacity building, knowledge transfer and policy making). User identification based on stakeholder mapping connects to Tasks 3.1 (define functional and non-functional requirements for use cases), 3.4 (develop end-user applications and visualization tools), 4.1 (specify demonstrators for validation and testing), 5.1 (stakeholder engagement and dissemination), and 5.3 (capacity building, knowledge transfer and policy making). Stakeholders play a central role to the selection of data, the processing of data, and development of applications and DOTs (see Figure 4).

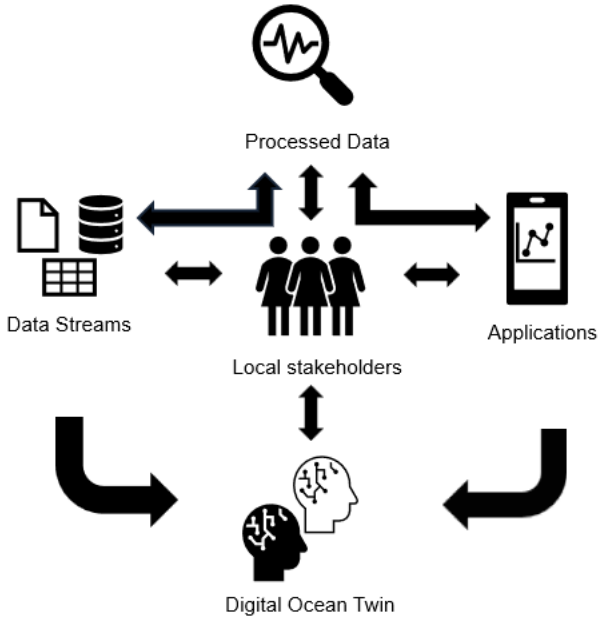


Figure 4 The central role of local stakeholders in driving the development of the applications and DOTs. Stakeholder needs additionally inform the use of data streams (e.g., datasets used in applications and DOTs) as well as the analyses and processing of data.

Users from the EU DTO

Users from the EU DTO include those developing similar DOTs for different case studies (e.g., in the SeaTWINs project cluster – EcoTwin, SURIMI, and SEADITO) and owners of infrastructure sharing data (e.g., EMODnet), data products (e.g., EMODnet, Copernicus), and infrastructure for hosting the EU DTO (EDITO). These stakeholders may also interact with the applications developed in SEADOTs for more technical use than local stakeholders that are looking for information and decision-support. Stakeholders from EU DTO emphasized the need to open and accessible data, software, and data products. Furthermore, discussions with other projects facilitate a more comprehensive understanding and definition of what a digital ocean twin is, as the concept of a digital twin within new fields, such as environmental management and ecology, emerge (see Durden 2025, de Koning et al. 2023).

Conclusions

This document outlines the methodology for collecting data to populate the data inventory which will serve as the basis for developing applications for each case study. These applications will then be integrated into case study-specific DOTs, which aim to serve as data-driven decision-support tools for local stakeholders. These DOTs will furthermore be integrated into the EU DTO. The methodology for the initial prioritization of the data inventory was outlined and follows a data quality assessment framework based on the ISO 19157 criteria for geospatial data quality assessment. Detailed mapping of the ISO 19157 criteria for geospatial data quality assessment to the SEADOTs data quality assessment framework is provided, with mitigation measures for cases where data may not be usable in its current state. The important role of local stakeholders for shaping the data requirements for applications and DOTs is underlined, as the development of the data inventory will be shaped by the needs of the stakeholders which are the target users of the DOTs. The methodology for mapping stakeholders is also provided, as this process is essential for determining the profile of end users and identifying their needs. The next steps of the project will necessitate the use and development of the data inventory as local stakeholders are engaged for inputs on the development process, and to shape the development of applications to be used in each use case DOT. Once DOTs are developed, they will be integrated into EU digital infrastructure (i.e., EU DTO) based on the interoperability framework.

References

Blomquist, J., and Waldo, S.. "Hur påverkas svenskt yrkesfiske av havsbaserad vindkraft?." *Rapport/AgriFood Economics Centre* 2024: 2 (2024).

Brönner, U., Sonnewald, M., & Visbeck, M. (2023). Digital twins of the ocean can foster a sustainable blue economy in a protected marine environment. In *The International Hydrographic Review*, 29(1), 26–40. https://doi.org/10.58440/ihr-29-a04

de Koning, K., Broekhuijsen, J., Kühn, I., Ovaskainen, O., Taubert, F., Endresen, D., Schigel, D., & Grimm, V. (2023). Digital twins: dynamic model-data fusion for ecology. In *Trends in Ecology & Evolution*, 38(10), 916–926. https://doi.org/10.1016/j.tree.2023.04.010OUCI+5Wageningen University & Research+5PubMed+5

Durden, J. M. (2025). Environmental management using a digital twin. In *Environmental Science & Policy*, 164, 104018. https://doi.org/10.1016/j.envsci.2025.104018

ICES Acoustic data portal. ICES, Copenhagen, Denmark. https://acoustic.ices.dk [accessed 2025-06-09]

Khan, M. T., Durrani, M. Y., Ali, A., Inayat, I., Khalid, S., & Khan, K. H. (2016). Sentiment analysis and the complex natural language. In *Complex Adaptive Systems Modeling*, 4(1), 2. https://doi.org/10.1186/s40294-016-0016-9

Mendelow, A. (1991). Stakeholder mapping. In Proceedings of the 2nd International Conference on Information Systems, Cambridge, MA, U.S.A.

Noardo, F., Atkinson, R., Bastin, L., Maso, J., Simonis, I., Villar, A., Voidrot, M.-F., & Zaborowski, P. (2024). Standards for data space building blocks. In *Remote Sensing*, 16(20), 3824. https://doi.org/10.3390/rs16203824

Ostrom, E. (2009). A general framework for analyzing sustainability of social-ecological systems. In *Science*, 325(5939), 419–422. https://doi.org/10.1126/science.1172133

Appendix

D1.1_inventory_20052025.xlsx