Project Acronym:	DataBio
Grant Agreement number:	732064 (H2020-ICT-2016-1 – Innovation Action)
Project Full Title:	Data-Driven Bioeconomy
Project Coordinator:	INTRASOFT International

Funded by the Horizon 2020 Framework Programme of the European Union

This project is part of BDV PPP

DELIVERABLE

D3.1 – Fishery Pilot Definition

Dissemination level	PU -Public
Type of Document	Report
Contractual date of delivery	M12 – 31/12/2017
Deliverable Leader	SINTEF Ocean
Status - version, date	Final – v1.0, 20/10/2017
WP / Task responsible	WP3
Keywords:	Fishery, Big Data, modelling, small pelagic fisheries, oceanic tuna fisheries, stakeholders, user requirements, ArchiMate models

Executive Summary

The objective of *WP3 Fishery Pilot* is to demonstrate how Big Data can boost the fishery sector. The Fishery Pilot focus is on two separate types of fisheries in two countries: *Oceanic Tuna fisheries in Spain and small Pelagic fisheries in Norway*. The areas encompassed by these pilots have an annual capture production of above 13 million tons.

Six separate pilot cases have been defined, addressing key concerns as the cost of fuel and vessel maintenance as well as overfishing and selection of correct species. The pilot cases cover these three separate viewpoints: *Immediate operational choices, fishing vessel trip and fisheries planning and fisheries sustainability and value*.

Task 3.1 Co-innovative preparations deals with the specification of user and stakeholders' needs through the specification of user stories to specify the most beneficial areas of interest from different points of view. The potential for different solution technologies were reviewed in this process resulting in a set of scenarios for the fishery sector within the three viewpoints mentioned above. The main objective of the task is to come up with the pilot definition encompassing the goals and strategies for fulfilling the requirements specified from the analysis of these scenarios and setting the direction for the implementation and piloting. The results are the pilot cases definitions including standardized pilot case specifications and motivation, strategy and evaluation plans.

According to the Description of the Action, the organizations participating in this task, and their respective planned work effort in person-months are: TRAGSA (1)/TRAGSATEC (3), VTT (2), SINTEF Fisheries (4), ECHEBF (1).

The deliverable D3.1 Fishery Pilot Definition specifies the pilot case definitions, requirement specifications, as well as implementation and evaluation plans.

Deliverable Leader:	Per Gunnar Auran (SINTEF Ocean)
Contributors:	Göran Granholm (VTT), Per Gunnar Auran (SINTEF Ocean), Karl Gunnar Aarsæther(SINTEF Ocean), Zigor Uriondo (EHU/UPV), Iñaki Quincoces(AZTI), Jørgen Jensen (SINTEF Ocean), Joakim Haugen (SINTEF Ocean)
Reviewers:	Tomas Mildorf (UWB), Jesus Estrada (TRAGSA), Miguel González (TRAGSATEC), Ståle Walderhaug (SINTEF Digital)
Approved by:	Athanasios Poulakidas (INTRASOFT)

Document History			
Version	**Date**	**Contributor(s)**	**Description**
0.1	04/05/17	GG	Initial draft, existing texts inserted
0.2	10/05/17	GG	Content transferred to new template
0.3	16/06/17	PGA	Pilot descriptions based on DX.1 template and best effort ToC adaptation
0.4	20/06/17	PGA	ArchiMate diagrams and elements tables added, minor ToC edits
0.5	14/07/17	PGA	Update based on initial review and D1.1/D2.1 harmonization.
0.6	27/09/17	PGA, KGA, JH, ZU, GG, JE, MGG	Updated ArchiMate figures, BDVA diagrams and added individual pilot roadmaps, edited plan section.
0.7	27/09/17	MGG, JE	Enumerations and formatting before review
0.8	03/10/17	GG	Final edits based on review comments. Figure and table references corrected. Removed empty appendix chapters. Added keywords. Added Conclusions chapter.
1.0	20/10/2017	Quality Manager, AP, Irene Matzakou	Final corrections and format adaptations.

Table of Contents

Table of Figures

List of Tables

Definitions, Acronyms and Abbreviations

Acronym/ Abbreviation	Title
AIS	Automatic identification system used for collision avoidance on ships and by vessel traffic services.
BDVA	Big Data Value Association
BDT	Big Data Technology
CEN	European Committee for Standardization
EO	Earth Observation
ESA	European Space Agency
EU	European Union
FAD	Fish Aggregation Device
FAO	Food and Agriculture Organisation of the United Nations
fAPAR	fraction of Absorbed Photosynthetically Active Radiation
GEOSS	Group on Earth Observations
GPRS	General Packet Radio Service
GPS	Global Positioning Service
HPC	High Performance Computing
IAS	Integrated Automation System
IACS	Integrated Administration and Control System
ICT	Information and Communication Technologies
IoT	Internet of Things
ISO	International organization for Standardisation
ISO JTC WG9	ISO Joint Technical Committee Work Group 9 - Big Data
KPI	Key Performance Indicator
MRU	Motion reference unit for precise measurement of vessel attitude and motion
NDVI	Normalized Difference Vegetation Index
NIR	Near-InfraRed
PC	Personal Computer
PPP	Public-Private Partnership
PU	Public
RGB	Red-Green-Blue
RS	Remote Sensing
SMEs	Small and medium-sized enterprises
SST	Sea Surface Temperature
TRL	Technology Readiness Level
UI	User Interface
UML	Unified Modelling Language, often used in software design.
UVA, UVB	(UV) ultraviolet rays, (A) long wave, (B) short wave

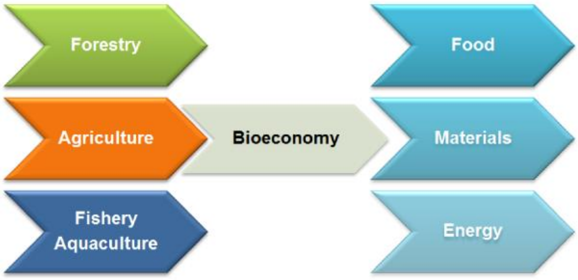
VMS	Vessel Monitoring System
VRA	Variable Rate Application
VTS	Vessel traffic services are a system to monitor traffic at sea, near ports and offshore installations.
WFS	Web Feature Service
WMS	Web Map Service
WMTS	Web Map Tile Service

Term	**Definition**
Big Data	Data sets that are so large or complex that traditional data processing software is inadequate to deal with them and distributed computing technologies are needed.
In situ	Latin phrase translated “on site” or “on position”- it means “locally” or “in place” to describe an event where it takes place
WP (Work Package)	A building block of the work breakdown structure that allows the project management to define the steps necessary for completion of the work.

1 Introduction

1.1 Project Summary

The data intensive target sector selected for the DataBio project is the **Data-Driven Bioeconomy**. DataBio focuses on utilizing Big Data to contribute to improve the efficiency and sustainability of the production of the raw materials from agriculture, forestry and fishery/aquaculture for the bioeconomy industry, in order to output food, energy and biomaterials.



DataBio will deploy state-of-the-art big data technologies and existing partners' infrastructure and solutions, linked together through the **DataBio Platform**. **DataBio Platform** consists of multiple components developed by the Big Data industries of the project which would be combined as appropriate for each pilot requirement. The three identified sectors (**agriculture, forestry and fishery**) can benefit from components already developed that can be often applied to more than one sector or specific pilot (case study). For example, earth observation components providing environmental conditions from satellite or visualization components for graphical visualization of results with georeferences. The execution will be through continuous cooperation of end user and technology provider companies, bioeconomy and technology research institutes, and final end-users from the big data value PPP programme.

DataBio is driven by the development, use and evaluation of six **pilot groups** in the 3 identified sectors. This document describes the selected pilot concepts to facilitate their later implementations utilizing the **DataBio Platform**. This will allow pilots select and utilize the best suitable market ready or almost market ready components (e.g. ICT, Big Data and Earth Observation methods, technologies, tools and services) to be integrated. Based on the pilot results and the new components interactions, new solutions and new business opportunities are expected to emerge.

The DataBio consortium is listed in Table 1. For more information about the project see [REF-01].

Table 1: The DataBio consortium partners

Number	Name	Short name	Country
1 (CO)	INTRASOFT INTERNATIONAL SA	**INTRASOFT**	Belgium
2	LESPROJEKT SLUZBY SRO	**LESPRO**	Czech Republic
3	ZAPADOCESKA UNIVERZITA V PLZNI	**UWB**	Czech Republic
4	FRAUNHOFER GESELLSCHAFT ZUR FOERDERUNG DER ANGEWANDTEN FORSCHUNG E.V.	**Fraunhofer**	Germany

5	ATOS SPAIN SA	**ATOS**	Spain
6	STIFTELSEN SINTEF	**SINTEF ICT**	Norway
7	SPACEBEL SA	**SPACEBEL**	Belgium
8	VLAAMSE INSTELLING VOOR TECHNOLOGISCH ONDERZOEK N.V.	**VITO**	Belgium
9	INSTYTUT CHEMII BIOORGANICZNEJ POLSKIEJ AKADEMII NAUK	**PSNC**	Poland
10	CIAOTECH Srl	**CiaoT**	Italy
11	EMPRESA DE TRANSFORMACION AGRARIA SA	**TRAGSA**	Spain
12	INSTITUT FUR ANGEWANDTE INFORMATIK (INFAI) EV	**INFAI**	Germany
13	NEUROPUBLIC AE PLIROFORIKIS & EPIKOINONION	**NP**	Greece
14	Ústav pro hospodářskou úpravu lesů Brandýs nad Labem	**UHUL FMI**	Czech Republic
15	INNOVATION ENGINEERING SRL	**InnoE**	Italy
16	Teknologian tutkimuskeskus VTT Oy	**VTT**	Finland
17	SINTEF FISKERI OG HAVBRUK AS	**SINTEF Fishery**	Norway
18	SUOMEN METSAKESKUS-FINLANDS SKOGSCENTRAL	**METSAK**	Finland
19	IBM ISRAEL - SCIENCE AND TECHNOLOGY LTD	**IBM**	Israel
20	MHG SYSTEMS OY - MHGS	**MHGS**	Finland
21	NB ADVIES BV	**NB Advies**	Netherlands
22	CONSIGLIO PER LA RICERCA IN AGRICOLTURA E L'ANALISI DELL'ECONOMIA AGRARIA	**CREA**	Italy
23	FUNDACION AZTI - AZTI FUNDAZIOA	**AZTI**	Spain
24	KINGS BAY AS	**KingsBay**	Norway
25	EROS AS	**Eros**	Norway
26	ERVIK & SAEVIK AS	**ESAS**	Norway
27	LIEGRUPPEN FISKERI AS	**LiegFi**	Norway
28	E-GEOS SPA	**e-geos**	Italy
29	DANMARKS TEKNISKE UNIVERSITET	**DTU**	Denmark
30	FEDERUNACOMA SRL UNIPERSONALE	**Federu**	Italy

31	CSEM CENTRE SUISSE D'ELECTRONIQUE ET DE MICROTECHNIQUE SA - RECHERCHE ET DEVELOPPEMENT	**CSEM**	Switzerland
32	UNIVERSITAET ST. GALLEN	**UStG**	Switzerland
33	NORGES SILDESALGSLAG SA	**Sildes**	Norway
34	EXUS SOFTWARE LTD	**EXUS**	United Kingdom
35	CYBERNETICA AS	**CYBER**	Estonia
36	GAIA EPICHEIREIN ANONYMI ETAIREIA PSIFIAKON YPIRESION	**GAIA**	Greece
37	SOFTEAM	**Softeam**	France
38	FUNDACION CITOLIVA, CENTRO DE INNOVACION Y TECNOLOGIA DEL OLIVAR Y DEL ACEITE	**CITOLIVA**	Spain
39	TERRASIGNA SRL	**TerraS**	Romania
40	ETHNIKO KENTRO EREVNAS KAI TECHNOLOGIKIS ANAPTYXIS	**CERTH**	Greece
41	METEOROLOGICAL AND ENVIRONMENTAL EARTH OBSERVATION SRL	**MEEO**	Italy
42	ECHEBASTAR FLEET SOCIEDAD LIMITADA	**ECHEBF**	Spain
43	NOVAMONT SPA	**Novam**	Italy
44	SENOP OY	**Senop**	Finland
45	UNIVERSIDAD DEL PAIS VASCO/ EUSKAL HERRIKO UNIBERTSITATEA	**EHU/UPV**	Spain
46	OPEN GEOSPATIAL CONSORTIUM (EUROPE) LIMITED LBG	**OGCE**	United Kingdom
47	ZETOR TRACTORS AS	**ZETOR**	Czech Republic
48	COOPERATIVA AGRICOLA CESENATE SOCIETA COOPERATIVA AGRICOLA	**CAC**	Italy

1.2 Document Scope

This deliverable D3.1 Fishery Pilot Definition (due M12) specifies each pilot case general definitions, detailed requirement specifications (e.g. components, data), evaluation plans and interactions with pilot users.

Specification of key performance indicators (KPIs) and business models is primarily left for inclusion in WP7 deliverables.

Reporting the results of the pilots is out of the scope of this document, and will instead be included in D3.2 Fishery Pilots Intermediate Report (M24) and D3.3 Fishery Pilots Final Report (M36).

1.3 Document Structure

This document is comprised of the following chapters:

Chapter 1 presents an introduction to the project and the document.

Chapter 2 gives a general overview of the Fishery Pilot and summarises key points of the pilot cases.

Chapters 3 to 8 describe the individual pilot cases.

2 Summary

2.1 Overview

According to the Food and Agricultural Organization (FAO), the world's marine fisheries expanded continuously to a production peak of 86.4 million tonnes in 1996 but have since stabilised despite high fluctuations of pelagic species due to el Niño events [REF-04]. The global recorded production was 82.6 million tons in 2011 and 79.7 million tonnes in 2012. The fraction of assessed stocks fished within biologically sustainable levels has exhibited a decreasing trend, declining from 90 percent in 1974 to 71.2 percent in 2011, as shown in Figure 1.

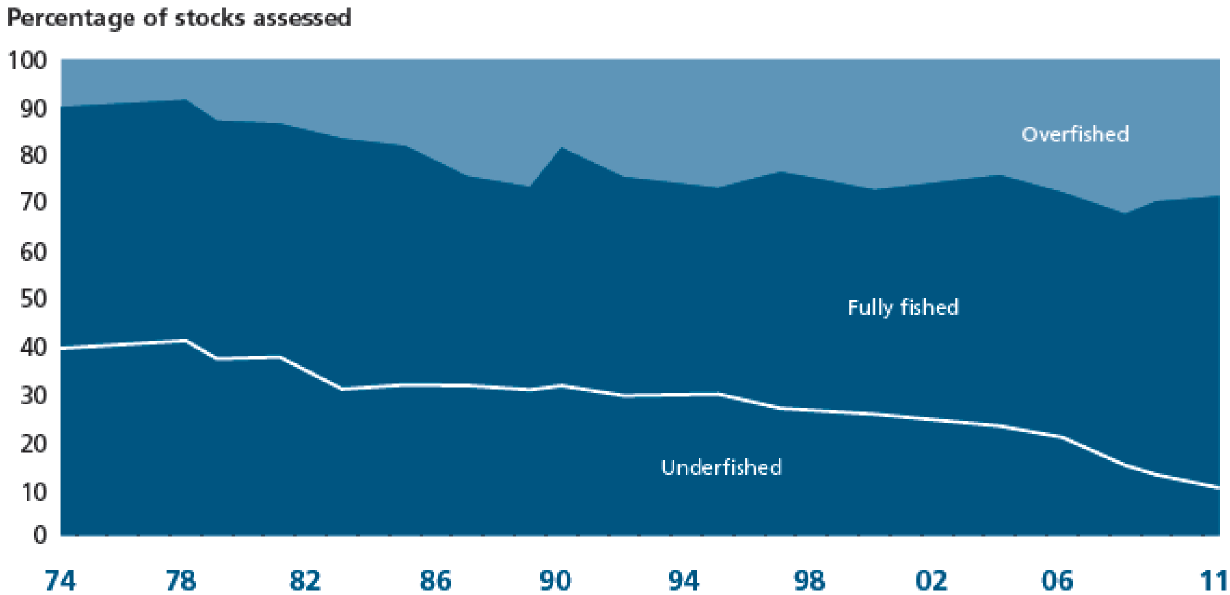


Figure 1: Percentage of marine fish stocks assessed (source: Food and Agricultural Organization)

Of the total number of stocks assessed in 2011, the fully fished stocks accounted for 61.3 percent and underfished stocks 9.9 percent (separated by the line in Figure 1). The underfished stocks decreased continuously from 1974 to 2011, but the fully fished stocks decreased from 1974 to 1989, and then increased to 61.3 percent in 2011. Correspondingly, the percentage of stocks fished at biologically unsustainable levels increased, especially in the late 1970s and 1980s, from 10 percent in 1974 to 26 percent in 1989. After 1990, the number of stocks fished at unsustainable levels continued to increase, albeit more slowly, and peaked at 32.5 percent in 2008 before declining slightly to 28.8 percent in 2011.

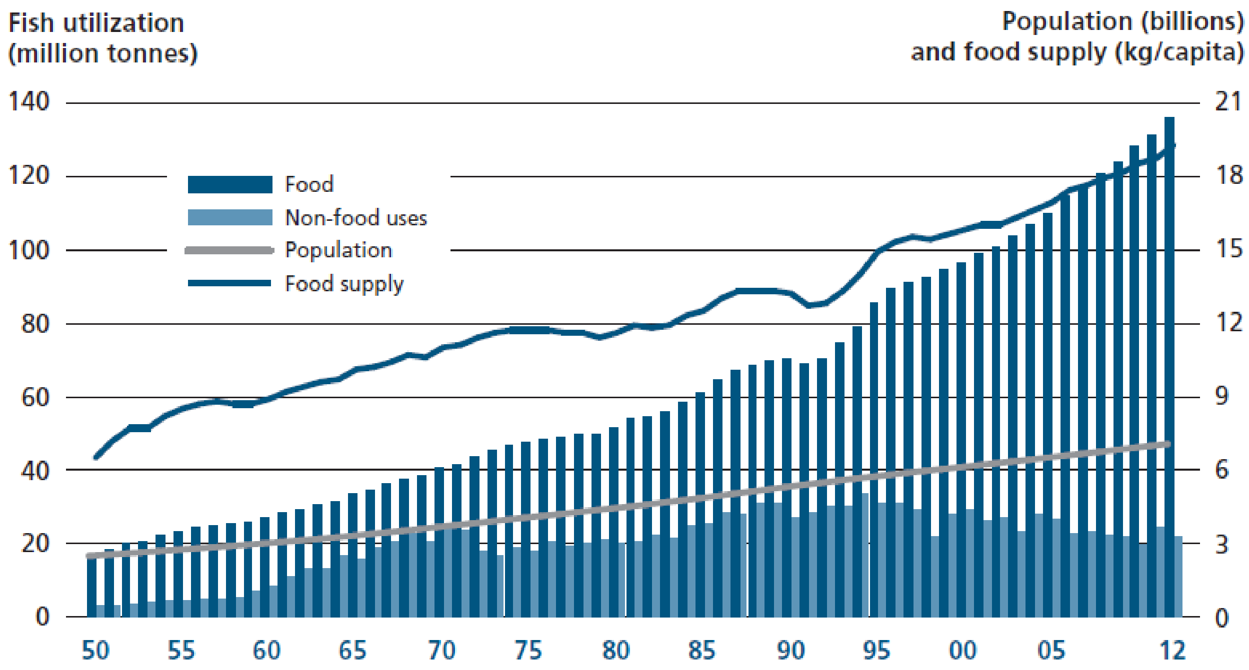


Figure 2: Fish utilisation and population growth.

Figure 2 and Figure 3 show how fish utilization and the global capture and aquaculture production have developed the last 60 years. As it can be seen from Figure 3, the capture production has stagnated, while the aquaculture production is still increasing.

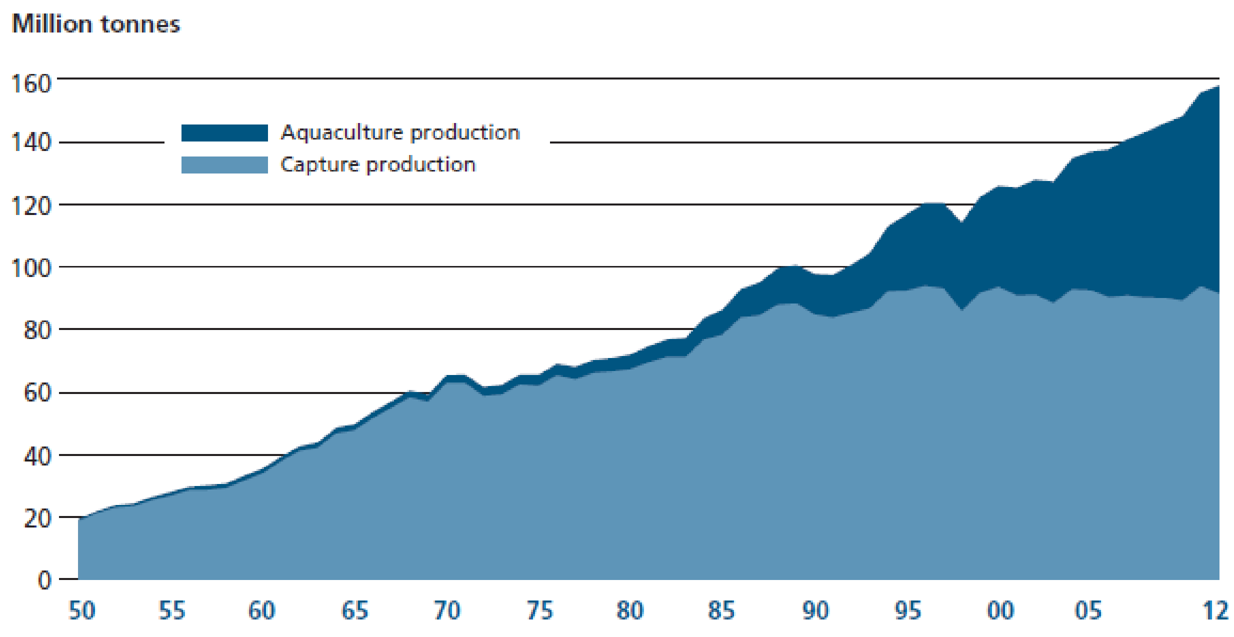


Figure 3: Development of global capture and aquaculture production.

According to the World Bank and the FAO, fisheries are an underperforming global asset. It is estimated that its production could be increased by $50 billion per year, if one could achieve better management and less overcapitalization of the fishing fleets. This means that a further growth in the blue economy would benefit from better monitoring and management of the sea environment and resources, as well as reducing the fisheries effort in terms of e.g. costs and energy consumption.

2.2 Pilot introductions

Big data technology (BDT) is a new technological paradigm that is driving the entire economy, including low-tech industries like agriculture, forestry and fishery. The adaptation of BDT in the fishing industry has been slow, even when compared to the other two primary sectors. However, the potential is untapped and this is fuelling increased interest and investments into Big Data fishery projects, with the Global Fishing Watch project being a recent example of a very large-scale processing of AIS data for detecting commercial fishing activity [REF-03].

Global fisheries need to be sustainable both from an economical and biological viewpoint, and BDT is both applicable for optimizing fishery revenue and for gaining deeper insight in the ocean ecosystem and the relationships between marine species. The cost of fuel, vessel maintenance as well as overfishing and selection of correct species to catch are crosscutting concerns that will be addressed by using BDT as enabling technology in the Fishery Pilot. The focus is on two separate types of fisheries in two countries: *Oceanic Tuna fisheries in Spain and small Pelagic fisheries in Norway*. The areas encompassed by these pilots have an annual capture production above 13 million tons.

In terms of technology readiness level (TRL), the agriculture pilots are mostly positioned at a wide range between the second and sixth, depending upon the pilot and the topic. For example, the Pilot B1: Oceanic Tuna Fisheries shows a TRL of 6, regarding Energy efficiency, but a TRL of 2 in routing taking in account buoys data.

The required skills are already available in the organizations participating in the project, and the organizations are ready to develop and improve the different business processes, which is a key factor for adopting the new technology.

2.3 Overview of pilot cases

Six separate pilot cases have been defined. The pilot cases cover three separate viewpoints:

(A1-A2) Immediate operational choices aiming at improving energy efficiency and preventive maintenance through data supported decisions on board the vessel,

(B1-B2) Fishing vessel trip and fisheries planning, covering route optimization, fish observation and selection of methods, and

(C1-C2) Fisheries sustainability and value, dealing with fish stock and market assessment to better target the demand and improving traceability to promote responsible and environmentally fishery.

The pilot cases are presented in Figure 4.

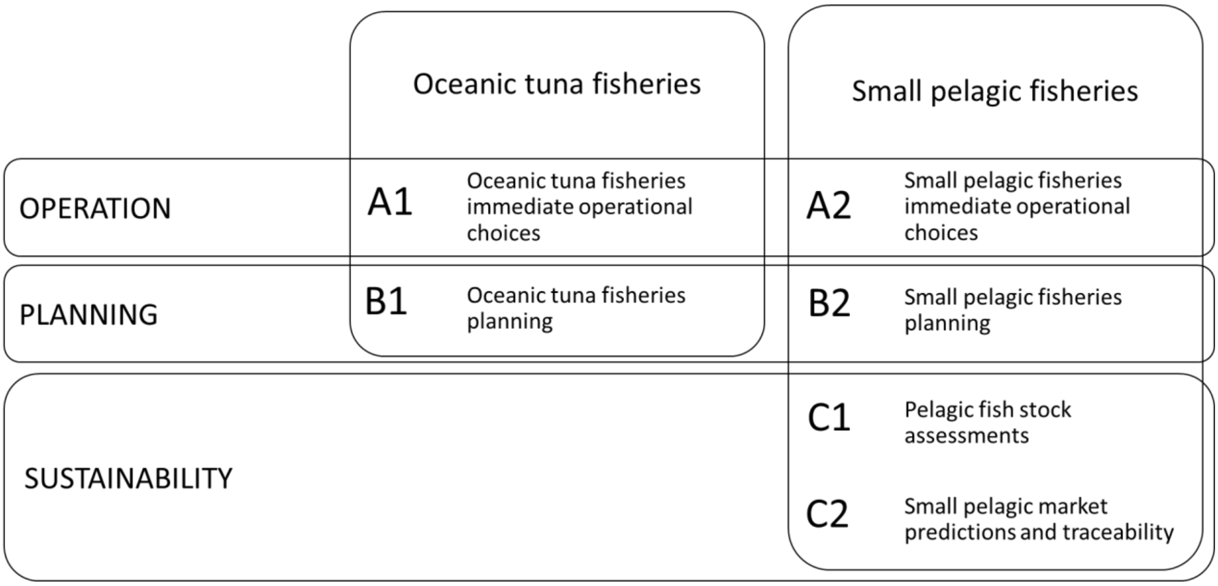


Figure 4: Summary of fishery pilot cases

Each pilot case is briefly introduced below with more detailed descriptions given in Chapters 3 to 8.

Fishery Pilot A1: Oceanic tuna fisheries immediate operational choices

In addition to fuel consumption, costs and downtime associated with vessel maintenance and breakdown are important for both the vessel economy and the environmental impact. This pilot develops technologies to improve vessel energy efficiency and engine preventive maintenance by providing support for operational choices such as:

- Vessel loading (In order to reduce the hull resistance and reduce fuel consumption).
- Weather routing (Reduce fuel consumption by taking weather conditions into account).
- Condition based maintenance (Proactive maintenance based on machinery sensors)

A central challenge is to develop models to reduce fuel consumption from the interaction between engine data, propulsion data, meteorological data and the vessels design by means of big data approaches. The second one is to use algorithms in real time to estimate the expected lifetime of different parts of the engine and propulsion system, or to learn when a part of the engine is close to failure and advice the technical staff in order to have the necessary parts and technicians at port and reduce the downtime for unexpected failures.

Another central challenge is to establish a common data management and analysis system that combines data from the fleet in real time (engines, propulsion, route and speed of the vessel, destination), data coming from public institutions with different types of meteorological models and model outputs and finally data from Earth observation for sea surface currents and other oceanographic parameters.

Fishery Pilot A2: Small pelagic fisheries immediate operational choices

Many new pelagic fishing vessels provide a greater flexibility in how they are operated. The crew must typically decide on propulsion mode (diesel-electric, diesel-mechanic and various hybrid configurations) as well as power generation (use of shaft generator and auxiliary engines). The best choice depends on desired speed, thrust and electric consumption, as well as the environmental conditions. This flexibility gives good possibilities for energy efficient operations, but it increases complexity and requirements on the crew.

Operational data are beginning to provide valuable insight in how the operational choices affect fuel economy for fishing vessels. To increase the usefulness of such data, the considerable effect of parameters such as waves, wind, and load condition must be compensated. This pilot will combine on-board measurements with available meteorological and oceanographic data, so that the underlying connection between more parameters can be more accurately modelled. This increased measurement coverage makes it possible to reduce noise from unmodelled effects and study the effect of operational parameters with more confidence. This includes, but is not limited to, how to operate the vessel in the most economical way in terms of propulsion mode, loading and use of auxiliary and main engines.

Fishery Pilot B1: Oceanic tuna fisheries planning

Purse seine is the surface gear that contributes most to the catch of yellowfin and skipjack tuna globally. One of the main fishing strategies are the use of fish aggregating devices (FADs). FADs are floating objects that have been modified and placed in the fishing areas by the fishermen to attract fish, and to facilitate their aggregation and capture. Additionally, FADs are often outfitted with a buoy to help fishers locate them. The energy efficiency of the fishing vessels is highly dependent on in which order the FADs are visited. This must be based on the expected tuna aggregation at each FAD, as well as how the choice of route impacts energy consumption. Such decisions are today based on the skills and knowledge of the skipper, and are one of the main areas which may improve the profit of the fishing industry without harming the state of the stocks. This pilot will handle two possible scenarios based on how data is shared:

1) No sharing between vessels.
2) Sharing of data within the vessels of one company.

Fishery Pilot B2: Small pelagic fisheries planning

A deciding factor for the energy efficiency of fishing vessels, is the ability to find the best suitable fishing grounds, combined with the best suitable fishing methods and tools. Such decisions are today based on the skills and knowledge of the crew based on experience and tradition. The goal of this pilot is to provide the crew and ship owners with information which benefits fisheries planning. It will run in two phases. The first phase will focus on providing an interactive end user tool for browsing compilations of past history, while the second level will provide the end user with a tool which will do predictions of where and when fish can be caught in an energy efficient manner. This pilot will collect and combine as many data sets as possible which may be of interest for predicting/planning of future fisheries effort. Extensive data sets within fisheries activity and catch statistics will be combined with information from that time and history of the same, such as meteorological and oceanographic data, moon phase and time of day and year.

Fishery Pilot C1: Pelagic fish stock assessments

Fish stock assessments are to a large extent based on scientific surveys conducted at dedicated research vessels. The objective of this pilot is to demonstrate that the combination of available information from existing sources, such as catch reports, oceanographic measurements, oceanographic simulations, stock simulations and stock observations, can be used for improving assessment of fish stocks and their distribution. Some of this data can be supplied by remote sensing, such as ocean surface currents, ocean surface temperatures and weather information. Some will have to be measured locally, such as fish species and densities. These data will be provided by fishing vessels equipped with the necessary sensors and communications abilities. Several projects are being initiated with the goal of providing such data. Most of the data will pass through several simulation models before being used in the predictive biological model. An important part of this system will be the use of historical data and hindcasts for testing model on historical data as well as retraining of the model.

Fishery Pilot C2: Small pelagic market predictions and traceability

The goal of this pilot is to provide information for predicting the development of various market segments, so that the fisheries may be targeted against the most beneficial fisheries. This pilot will also act as a basis for providing the consumers with information about the products, so that he/she can take into consideration quantitative aspects such as sustainability, environmental impact, energy consumption per kg fish etc. This could be presented either directly or through certification labels. The basis for the market predictions will be to combine as many data sources as possible which are assumed to be related to the market situation. Using historical data, machine learning will be employed to model consumption habits in certain areas, as well as the relation between market development and other factors. These models will then be used for providing predictions for how various market segments will develop in the future.

2.4 Fishery datasets utilized in pilots

The datasets used by the fisheries pilots can be coarsely divided into four different categories:

- *Local measurements* are data obtained by sensors situated onboard vessels and buoys, typically machinery and hydroacoustic data.
- *Remote measurements* are measurements which may cover a greater geographical area, such as measurements from satellites.
- *Model data* are data which are provided by simulation models, typical meteorological, oceanographic and biological models.
- *Reports* are data which originates in manual or automatic reports, such as fish landings and quality reports.

An overview of the expected data sources is given in Table 2 and

Table 3 below.

Table 2: Local measurement data

Source	Type	Volume (GB)	Velocity (GB/year)
Ervik & Saevik	Local measurement (Operational data)	60.17	30.09
Ervik & Saevik	Local measurement (Motion data)	45.41	22.71
Kings Bay	Local measurement (Operational data)	35.95	17.98
Kings Bay	Local measurement (Motion data)	45.41	22.71
Eros	Local measurement (Operational data)	34.82	17.41
Eros	Local measurement (Motion data)	45.41	22.71
Heroyhav	Local measurement (Operational data)	28.19	28.19
Heroyhav	Local measurement (Motion data)	45.41	22.71
Echebastar fleet	Remote measurement (FAD/Buoys data)	73.12	24.37
Echebastar fleet	Local measurement (Engine sensors)	13.28	6.64
Echebastar fleet	Local measurement (VMS)	0.01	0.00
Echebastar fleet	Local measurement (AIS)	0.34	0.17
Echebastar fleet	Reports (Logbooks)	0.00	0.00
Echebastar fleet	Reports (Observers)	0.01	0.00
Total		***427.54***	***215.68***

Table 3: Remote measurements

Source	Type	Volume (GB)	Velocity (GB/year)
Earth observation	Geolocated Raster	2 499.27	499.85
Earth models	Geolocated Raster	4 998.54	4 998.54
Total		***7 497.81***	***5 498.39***

Note that these summaries do not account for hydroacoustic and sonar data, which will also be investigated in the Fishery pilots, in particular the C1 pilot described below. Modern multibeam sonar as the SIMRAD MS70 may deliver data *at a rate of 2GB per second*, which will significantly increase the data volume. However, for this sensor type realtime data filtering and feature extraction will be very important to both decide which data to keep (local storage limited to tens of Terabytes) and which information to send onshore (limited bandwidth offshore). Therefore, it is not tabulated here what the net volume and rate will amount to as it will depend on the pre-processing time.

2.5 Representation of pilot cases

Each pilot is described in the following structure:

- PILOT OVERVIEW
  - Introduction
  - Overview
- PILOT CASE DEFINITION
  - Stakeholder and user stories
  - Motivation and strategy
- PILOT MODELLING WITH ARCHIMATE
  - Motivation view
  - Strategy view
- PILOT EVALUATION PLAN
  - High level goals and KPIs
  - Evaluation view
- BIG DATA ASSETS

Standardized descriptions are used to describe the pilot use cases (ISO JTC1 WG9 use case template), models (ArchiMate 3.0) and data assets (extended BDVA reference model).

2.6 Pilot modelling framework

The pilot cases are modelled using the ArchiMate 3.0 modelling framework. Figure 5 summarizes the overall ArchiMate 3.0 framework. The figure also depicts the input provided by the domain WPs (WP1, WP2, WP3 and their pilots) and that provided by the technology WPs (WP4, WP5), which will be correlated in the next stages of the modelling process.

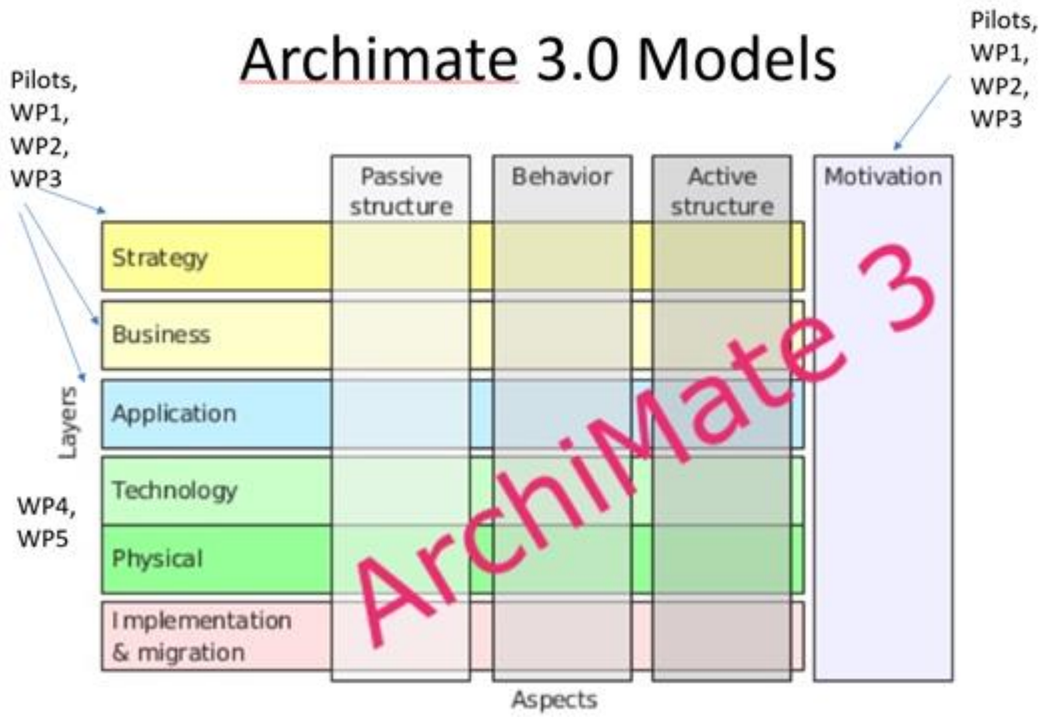


Figure 5: ArchiMate 3.0 modelling framework.

The modelling presented in this deliverable focuses on the “Motivation” and “Strategy” views. The “Motivation” view models the reasons that guide the design of the architecture. The “Strategy” view adds how the course of action is realized. Table 4 provides an extended description of the two views. (The “Application” and “Business” views are also described in the table as they were optionally included in some of the pilot models.) After the completion of this deliverable, the plan is to extend the modelling with other views, while investigating the correlations with the technology WP input.

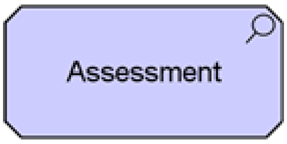
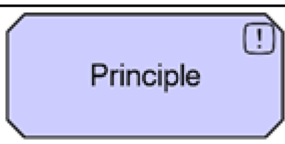
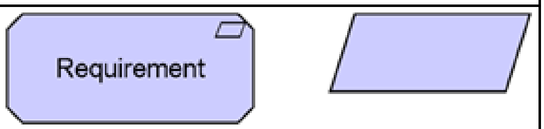
Table 4: ArchiMate Motivation and Strategy views

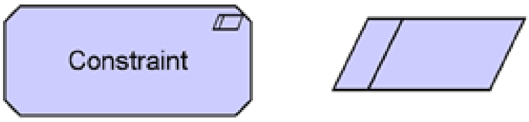
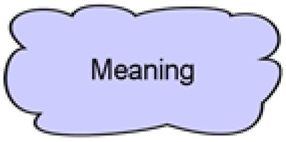
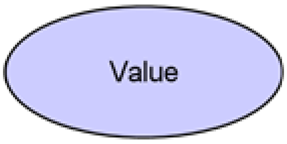
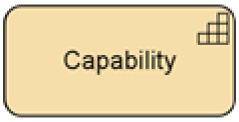
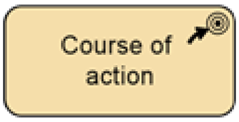
View name	Description
Motivation view	Motivation elements are used to model the motivations, or reasons, that guide the design or change of an Enterprise Architecture. It is essential to understand the factors, often referred to as drivers, which influence other motivation elements. They can originate from either inside or outside the enterprise. Internal drivers, also called concerns, are associated with stakeholders, which can be some individual human being or some group of human beings, such as a project team, enterprise, or society. Examples of such internal drivers are customer satisfaction, compliance to legislation, or profitability.
Strategy view	The key elements of the strategy view are resources, capabilities and course of action, and this diagram illustrates how these elements are combined to provide outcomes supporting the goals from the motivation view. ● A resource represents an asset owned or controlled by an individual or organization. Resources are analyzed in terms of strengths and weaknesses, and they are considered when implementing strategies. ● A capability represents an ability that an active structure element, such as an organization, person, or system, possesses, where the focus is on realizing outcomes to support some goal or deliver value. ● A course of action is an approach or plan for configuring some capabilities and resources of the enterprise, undertaken to achieve a goal, i.e. what to do to support the goals.
Business view (optional)	The Business Layer is typically used (often in conjunction with the strategy elements) to model the business architecture of an enterprise, as a description of the structure and interaction between the business strategy, organization, functions, business processes, and information needs. The key elements are the business process, functions and interactions, including stakeholders and resources.

Application view (optional)	The Application Layer is typically used to model the information systems architectures of the enterprise, including the application architecture that describes the structure and interaction of the applications. The key element is the application component.

The main elements used in the above views are explained in Table 5. Their relationships are shown in Figure 6 and Figure 7. For further information see [REF-02].

Table 5: Elements used in the ArchiMate Motivation and Strategy views

Element	Definition	Notation
Stakeholder	The role of an individual, team, or organization (or classes thereof) that represents their interests in the outcome of the architecture.	
Driver	An external or internal condition that motivates an organization to define its goals and implement the changes necessary to achieve them.	
Assessment	The result of an analysis of the state of affairs of the enterprise with respect to some driver.	
Goal	A high-level statement of intent, direction, or desired end state for an organization and its stakeholders.	
Outcome	An end result that has been achieved.	
Principle	A qualitative statement of intent that should be met by the architecture.	
Requirement	A statement of need that must be met by the architecture.	

Constraint	A factor that prevents or obstructs the realization of goals.	
Meaning	The knowledge or expertise present in, or the interpretation given to, a core element in a particular context.	
Value	The relative worth, utility, or importance of a core element or an outcome.	
Resource	An asset owned or controlled by an individual or organization.	
Capability	An ability that an active structure element, such as an organization, person, or system, possesses.	
Course of action	An approach or plan for configuring some capabilities and resources of the enterprise, undertaken to achieve a goal.	
Optional view elements	The business and application view elements are found in the ArchiMate 3.0 standard [REF-02].	These elements are omitted here to keep the element overview focused on the required diagrams.

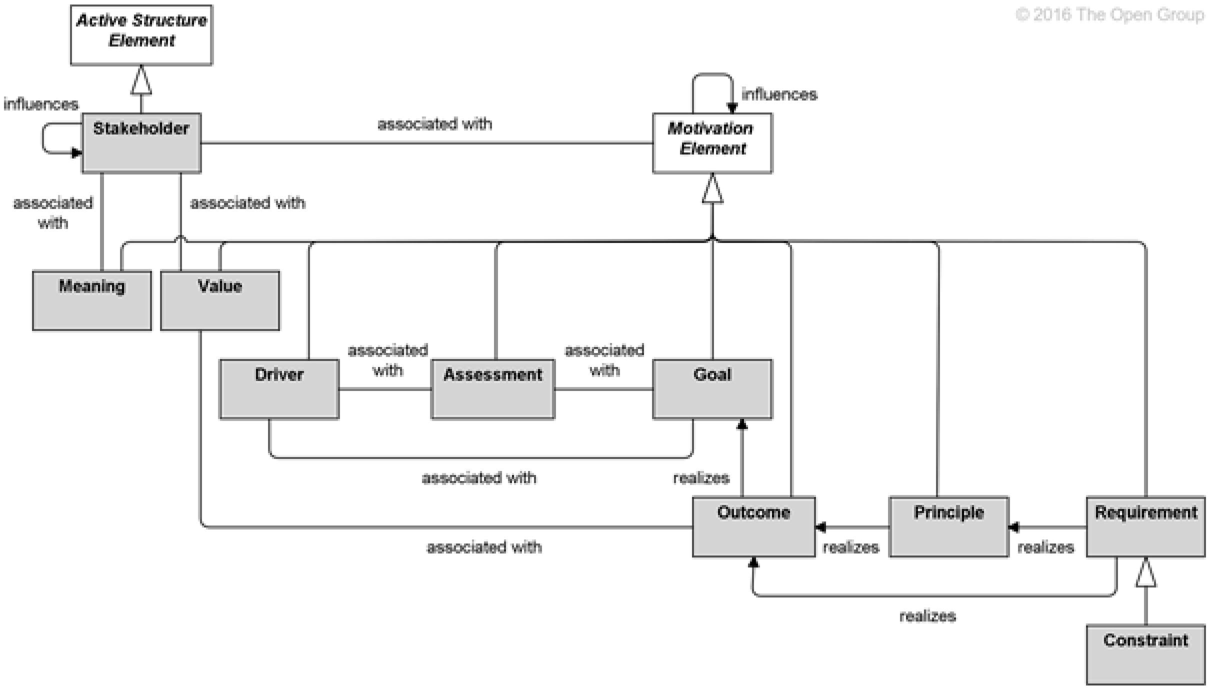


Figure 6: Relationships of the Motivation elements

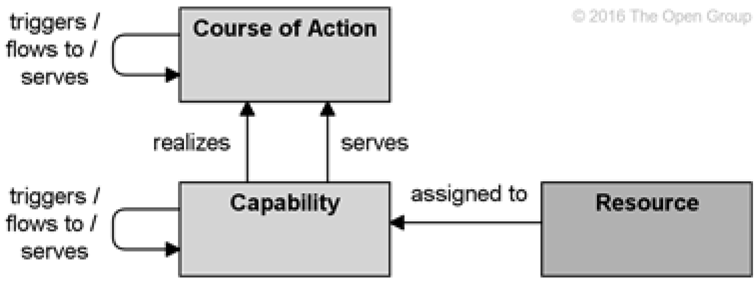


Figure 7: Relationships of the Strategy elements

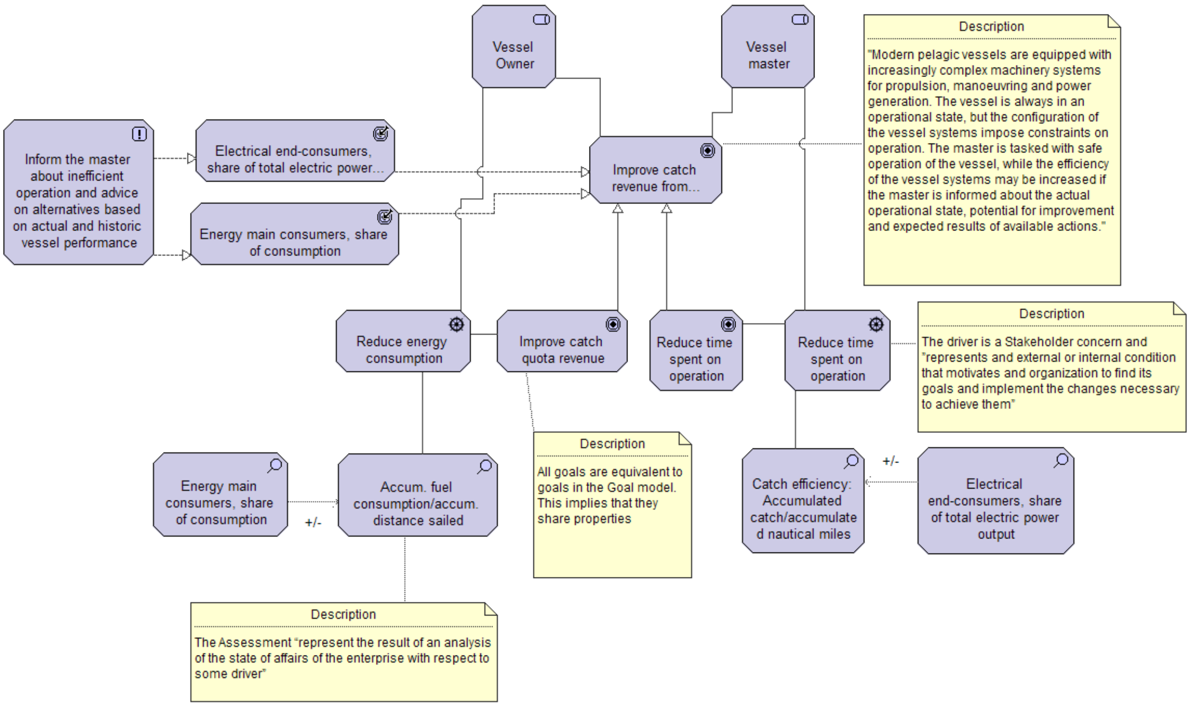


Figure 8: DataBio Fishery Pilot A2. Example motivation view

For example, the models developed for the different pilots and scenarios would be similar to Figure 8 from the Fisheries A2 Pilot used as the ArchiMate modelling showcase in DataBio. It explains how the motivation of the vessel master and vessel owner is to reduce operational costs by reducing time expenditure and fuel consumption.

2.7 Overall pilot roadmap

A general roadmap with important milestones for the fishery pilots is included below in Figure 9. It has been adapted to the two scheduled iterations of the DataBio services and component platform development and depends on these internal project deliveries from work package 4 (WP4). This is an overall roadmap common to the Fishery pilots and each pilot will have a customized and more detailed version of it for the pilot implementation.

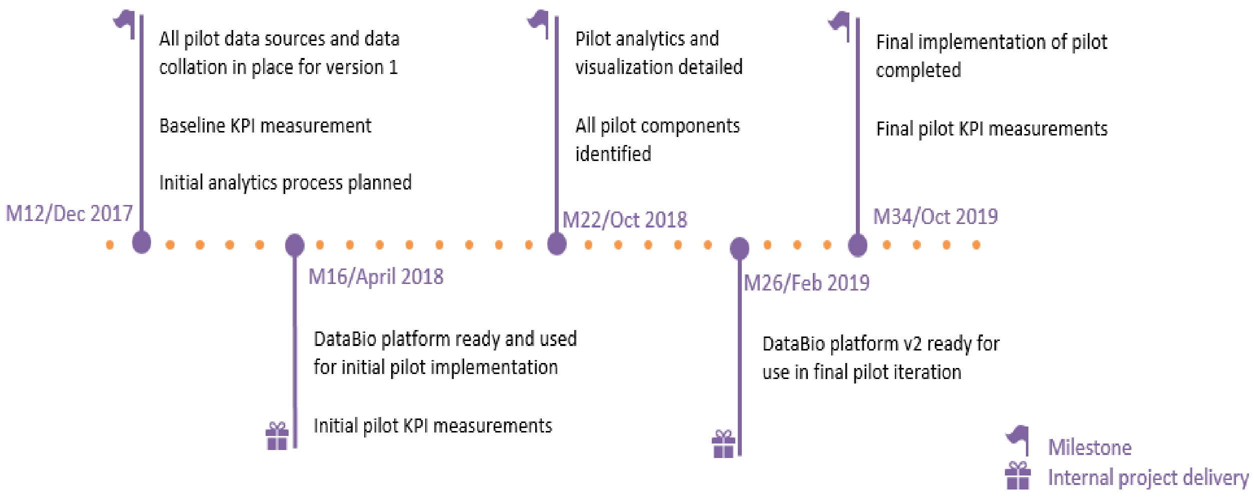


Figure 9: Overall roadmap for the fishery pilot cases

3 Fishery Pilot A1: Oceanic tuna fisheries immediate operational choices

3.1 Pilot overview

In addition to fuel consumption, costs and downtime associated with vessel maintenance and breakdown are important for both vessel economy and environmental impact. This pilot will develop technologies to improve vessel energy efficiency and engine preventive maintenance by providing support for operational choices such as:

- Vessel loading (In order to reduce the hull resistance and reduce fuel consumption).
- Weather routing (Reduce fuel consumption by taking weather conditions into account).
- Condition based maintenance (Proactive maintenance based on machinery sensors)

3.1.1 Introduction

The engines installed in ships used in tropical tuna fisheries are medium and high-speed engines burning distillate diesel oil, which is a high quality and high cost diesel oil. The main engines are medium-speed engines in the range of 4000 to 7000 kW and the auxiliary engines are medium and high-speed engines with a total installed power in the range of 4000 to 6000 kW. Usual propulsion layout is a Controllable Pitch Propeller system with shaft generator combined with smaller diesel engines coupled to electric generators and the hydraulic pumps. The systems are very complicated with several combinations of running conditions, which requires sophisticated control systems for energy optimization and management.

The total fuel consumption per vessel is in the range of 5.000 to 7.000 mT per year. All of which is distillate diesel oil supplied in remote areas, which increases the already high price for such fuel in marine applications.

Fishing is carried out in remote fishing areas, with reduced ship traffic, which means that any main engine failure leaves the ship out of operation for several days. This makes the main engine correct running condition vital for crew safety.

3.1.2 Overview

This pilot will develop technologies to assist the fishing vessel crew in taking the best decisions to improve vessel energy efficiency and engine preventive maintenance.

The specific operational choices which will be addressed are:

- Loading of the vessel in the different possible configurations to reduce the hull resistance and reduce fuel consumption
- Reduction of the fuel consumption by means of taking in account sea surface currents and meteorological conditions to make an intelligent sailing
- Use of the data from the engine and propulsion sensors to predict in advance any mechanical failure for proactive maintenance and saving unproductive days at port and reduce the cost of the reparations (increase also safety of ships and crew on board due to reduced unexpected failures).

Method/description

All the historic available data from 3 tuna purse seines built exactly with the same design in the same shipyard will be used in this pilot. In this case, a model of the hull used in the hydrodynamics channel is available together with the last two years of results of more than 100 sensors of the different engines and the propulsion system of the vessels with a sampling rate of every ten seconds. All this data with the route data (from AIS or vessels own plotters) and the historical data about meteorological conditions, sea currents and engine failures will be used to adjust the more appropriate models (end of life, machine learning algorithms, etc.) for giving immediate operational choices to oceanic tuna fisheries.

During the analysis period, the platforms and mechanisms to acquire in real or near to real time, the data from the engine and propulsion system sensors and for meteorological models will be set up in order to be ready for the second phase of providing immediate operational choices to the skippers or captains of the vessels minutes after sending the engine and propulsion data and receiving the meteorological models.

Challenges

A central challenge is to develop models to reduce fuel consumption from the interaction between engine data, propulsion data, meteo data and the vessels design by means of big data approaches.

The second challenge is to use algorithms in real time to estimate the expected lifetime of different parts of the engine and propulsion system, or to learn when a part of the engine is close to failure and advice the technical staff in order to have the necessary parts and technicians at port and reduce the downtime for unexpected failures.

Another central challenge is to establish a common data management and analysis systems that combines data from the fleet in real time (engines, propulsion, route and speed of the vessel, destination), data coming from public institutions with different types of meteorological models and model outputs and finally data from Earth observation for sea surface currents and other oceanographic parameters.

Relevance to and availability of Big Data and Big Data infrastructure

The connection between operational choices and its consequences are difficult to reveal, since many hard to measure parameters and effects come into play. It is therefore assumed that employing big data methods, such as machine learning, to existing datasets will prove useful. By combining specific vessel measurements with meteorological and oceanographic hindcast, it will be possible to analyse how the vessel loading affects vessel movements and ship resistance.

- Historical records of 117 measurements from the engine and propulsion system every 10 seconds. Access to real time data from the vessel in a second phase.
- AZTI's Marine Datacenter. Currently has implemented different databases for engine and propulsion parameters.
- Integration against other data sources, such as oceanographic and meteorological parameters is planned.

Benefit of pilot

South Atlantic and Indian Oceans are the main fishing grounds for the Basque Oceanic Tuna Purse Seine fleet but Europe has vessels all around the world.

Echebastar Fleet will be the main beneficiary of the developed technology at first, but, if successful, the creation of a company for exploitation of the technology as a service to other vessels owner is expected.

The whole Tuna purse seine industry will benefit from the implementation of such measurements. In the case of the condition based maintenance all the vessels with engines similar to the ones used by tuna purse seines will benefit of optimized maintenance intervals and improved propulsion condition. The improvement of propulsion system condition will not only save energy and money in maintenance; but will also increase safety of ships and crews that sail in remote areas.

Current TRL's: Intelligent sailing and loading: 4, Engine failure prevention: 2-3.

3.2 Pilot case definition

Table 6: Summary of Fishery Pilot A1: Oceanic tuna fisheries immediate operational choices

Use case title	Oceanic tuna fisheries immediate operational choices	
Vertical (area)	Fishery	
Author/company/email	Zigor Uriondo / EHU-UPV / zigor.uriondo@ehu.eus	
Actors/stakeholders and their roles and responsibilities	Skipper / Chief Engineer / Technical staff in the shore office / ship owner Skippers take operational decisions on board and are responsible of catching fish / Chief engineers on board and technical staff in the shore office are responsible to keep ship running and fishing with the minimum downtime and operational cost	
Goals	Develop tools to assist when taking operational decisions regarding fishing operations and maintenance operations	
Use case description	Refer to the pilot case definition section and diagrams in the pilot modelling sections.	
Current solutions	**Compute(System)**	Standard computer operating system. No specific tool employed at the moment.
	Storage	Company Server and ship's computer
	Networking	NO. Human assisted assessment of available machinery performance data
	Software	No software used at the moment apart of standard office tools
Big data characteristics	**Data source (distributed/centralized)**	Some data from on board monitoring system. Some data from weather and sea condition information.

	Volume (size)	Few terabytes combining all the data sources (propulsion/ship and weather and sea condition data)
	Velocity (e.g. real time)	From 1 Hz to 1/6 Hz. Ship/propulsion data not in real time.
	Variety (multiple datasets, mashup)	Great variety as the intention is to include as many relevant data sources as possible.
	Variability (rate of change)	Same as above, rate of change depends very much on data source/type.
Big data science (collection, curation, analysis, action)	**Veracity (Robustness Issues, semantics)**	The software has to give reliable advice to end user to be considered in normal operational condition
	Visualization	Different graph and tendency curves to show correlation between different parameters and working condition.
	Data quality (syntax)	Filtering will be necessary to remove outliers from faulty monitored parameters.
	Data types	Text files (.csv) from ship propulsion and energy generation on board.
	Data analytics	Predictive analytics for event detection in engine performance and energy consumption prediction based in weather and sea condition predictions using ship energy consumption models.
Big data specific challenges (Gaps)	Predict sea conditions and energy consumption on board with enough precision using available data. The predictions need a precision threshold as high as possible if they have to be used by skippers and ship operators.	
Big data specific challenges in bioeconomy	A system intended for use at open sea faces severe network connectivity challenges, both for onshore data logging and the use of web based services. The vessel system must be able to operate for days without Internet access based on the latest information that was available and also from the logged data from the main engine and propulsion system on board.	
Security and privacy technical considerations	Operational areas and ship position are extremely delicate due to pirate presence in Indian ocean and due to possibility to be used by competitors in their fishing operations.	
Highlight issues for generalizing this Use case (e.g. for ref. architecture)	Ships require complex monitoring system with periodic upload of recorded data. Monitoring system is very complex to be retrofitted once ship in operation.	
More information (URLs)	www.databio.eu <other URLs to be added later if relevant>	
Note: <additional comments>		

3.2.1 Stakeholders and user stories

The end user is a Tuna fishing company and the staff in the company. The interests within the company are not the same for all the members of the company; but all of them could be benefited with a software tool that could optimize fishing effort with a minimum cost.

Table 7: Stakeholders and user stories of Fishery Pilot A1: Oceanic tuna fisheries immediate operational choices

Who (type of user)	**I want to** (can you perform some task)	**Why** (achieve some goals)
Skipper	As a skipper I would like to have a tool that recommends me which buoys would provide me with maximum catches and in the same time reduce me the burden of estimating the minimum fuel consumption for each of the possible buoys with catches.	Catch the same amount of fish reducing sailed nautical miles and uncertainty with fish availability on board.
Chief Engineer / superintendent	As chief engineer / superintendent the improved capability of predicting machinery faults or undesired machinery events will reduce vessel downtime and risks of vessel security associated with machinery faults on board. Such kind of tool would also reduce the stress associated with vessel operation in remote areas.	Reduce downtime and maintenance costs due to unpredicted machinery failures.
Ship owner	As owner I would like to keep the income from fish catches and reduce the operational costs with reduced fuel consumption and maintenance costs.	Cost reduction with same or higher income will increase my profit.

3.2.2 Motivation and strategy

The main motivations for this pilot are:

- Reduce costs from fuel consumption on board while keeping the amount of catches.
- Reduce maintenance costs and downtime of ship with precise event prediction prior to fault occurrence.

The pilot motivation and strategy is summarized using ArchiMate diagrams in the next section, while goals and KPIs are addressed in the succeeding evaluation plan. The pilot motivation and strategy is summarized using ArchiMate diagrams in the next section, while goals and KPIs are addressed in the successive evaluation plan.

3.3 Pilot modelling with ArchiMate

3.3.1 DataBio fishery pilot A1 motivation view

This section describes the "Fishery A1 Motivation view" defined in the "Fishery A1 Oceanic tuna fisheries immediate operational choices modelling with ArchiMate" view point (Figure 10). There are different interests to improve operational information in the company in order to increase the energy efficiency of the ship: skipper, vessel owner and technical staff on shore. The skipper wants better information to maintain catch level while reducing the total of nautical miles sailed and reducing the amount of buoys he has to look for. Machinery technical staff is interested in improved machine running conditions in order to reduce downtimes and high cost caused by unexpected failures. The vessel owner is interested in reducing costs through reduced fuel oil consumption and machinery failure costs, hence, reducing operational costs while keeping income from fish catches.

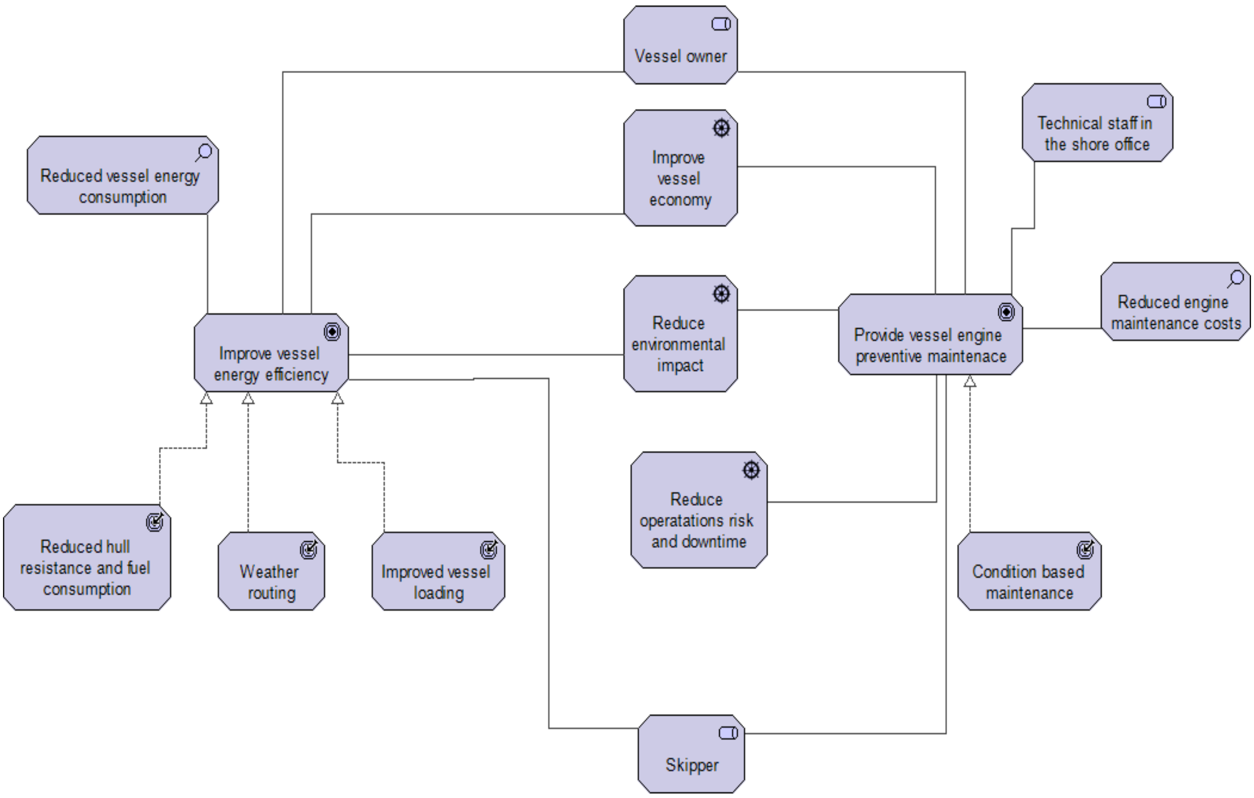


Figure 10: Fishery Pilot A1 motivation view

3.3.2 DataBio fishery pilot A1 strategy view

The strategy view for the A1 fishery pilot is shown in Figure 11. The skipper will receive information about the buoys fishing probabilities obtained from other sources and models about the buoys and the best possible fishing option. Within the calculations and models used to decide the best possible fishing buoy ship's energy efficiency models will be introduced. The ships' fuel consumption models will use information from the vessels in order to select best loading condition on board and best operational choice (constant rpm, shaft generator connected, etc.). Weather routing will be also one of the inputs to the models to give best possible operational choice.

The monitoring systems on-board collect data that will be analysed on board and on shore, in real time and off-line with time delay. The main engine operational data will be analysed and condition will be evaluated with computer tools that will evaluate engine condition and give recommendations to technical staff on board and on shore, preventing unexpected failures and downtimes. Machine learning will be used to generate propulsion system healthy condition models that will be later used to evaluate condition and give recommendations.

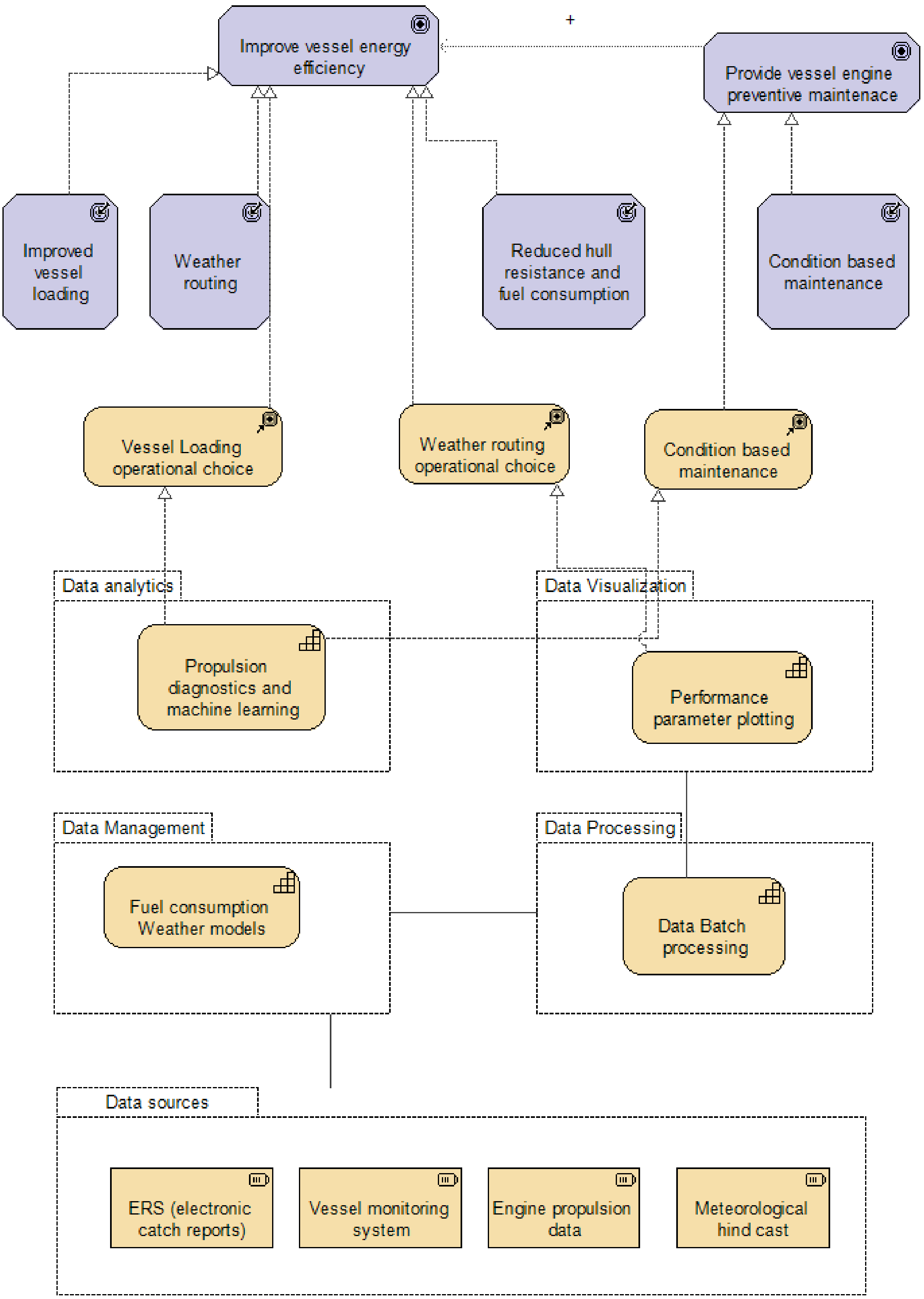


Figure 11: Fishery Pilot A1 strategy view

The immediate decision support system is built on top of a data collection and distribution system. The data collection and distribution system is used to collect sensor data from the on-board systems and makes them available in a single system. The data distribution system ensures that the decision support system only interface with a single system, instead of multiple sensors. The decision support system presents the data from the data distribution system and collect them in an internal storage system for presentation of current performance vs. historic performance.

3.4 Pilot Evaluation plan

The high-level evaluation plan for the pilot is shown in Figure 12. The evaluation plan is focused on data collection, KPI measurements and the implementation of technologies for analytics and visualization of information relevant for the operation of the vessel.

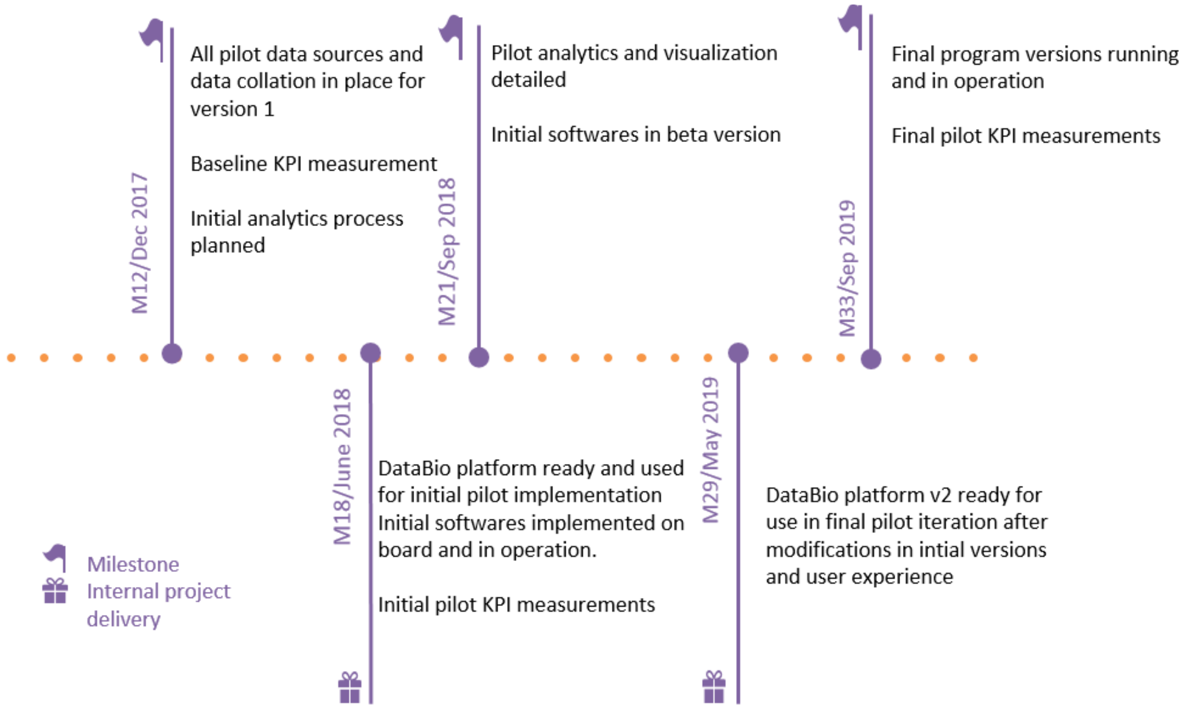


Figure 12: Fishery Pilot A1 Milestone roadmap

3.4.1 High level goals and KPIs

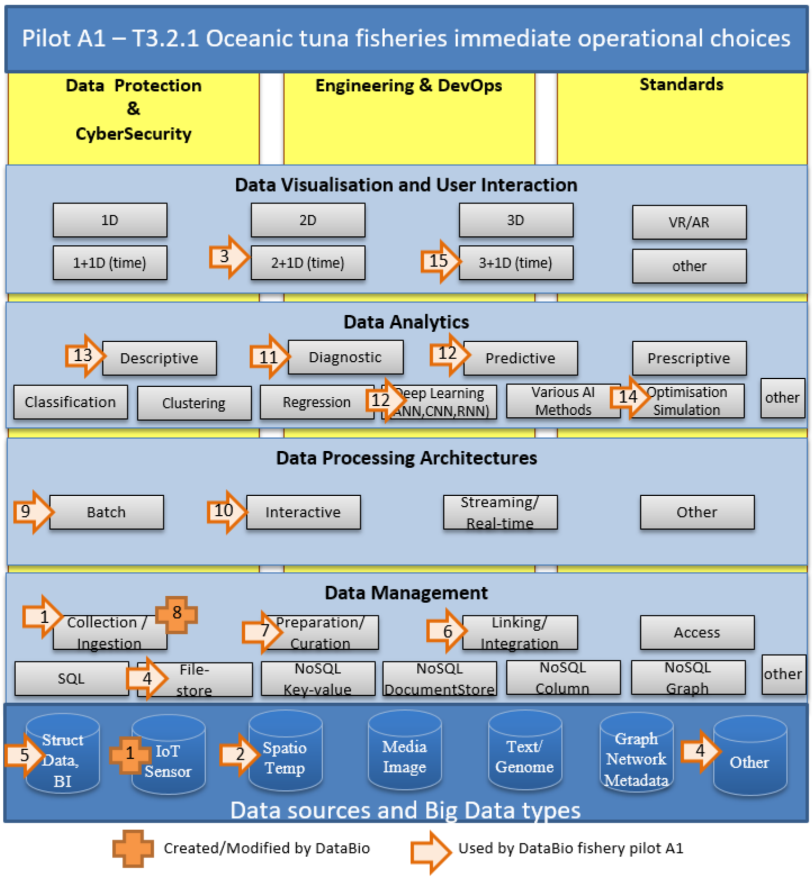
The relevant KPIs identified so far are:

- Fuel consumed per sailed nautical mile.
- Fuel consumed per catch unit of mass [kg fuel/fish ton or kg]
- Downtime hours due to main engine failure per year
- Miles sailed per catch unit of mass [Nautical mile/fish ton or kg]

The baseline values will be obtained through analysis of historical data. The baseline values will be used as reference to evaluate the proposed solutions.

3.5 Big data assets

The diagram below summarizes Big Data technology components used in this pilot using the extended BDVA reference model. Where applicable, specific partner components that are likely to be used or evaluated by this pilot are listed using the DataBio component IDs.



1. On board monitoring system for logging data from vessel
2. Ship propulsion system and auxiliary engines data
3. Plot of performance parameters models/time – R (Open VA), LabVIEW
4. Catch reports for KPI development
5. Text format data
6. Fuel consumption models with weather data
7. Data validation, and decimation/interpolation with R, LabVIEW
8. SW component gathering data from 4 and 5
9. Batch processing of historic data- R / LabVIEW?
10. Interactive and iterative development
11. Propulsion system condition diagnosis with ML methods
12. ML methods for predictive Fuel Oil Consumption models
13. Correllation/Covariance analysis
14. Optimization of operational choices
15. Interactive display propeller pitch, rpm, SOG with time for optimum sailing point

Figure 13: BDVA reference model for fishery pilot A1

4 Fishery Pilot A2: Small pelagic fisheries immediate operational choices

4.1 Pilot overview

4.1.1 Introduction

A new breed of small pelagic fishing vessels provides a greater flexibility in how they are operated. The crew must typically decide on propulsion mode (diesel-electric, diesel-mechanic and various hybrid configurations) as well as power generation (use of shaft generator and auxiliary engines). The best choice depends on desired speed, thrust and electric consumption, as well as the environmental conditions. This flexibility gives good possibilities for energy efficient operations, at the cost of extra complexity and requirements on the crew.

4.1.2 Overview

This pilot will develop technologies to assist the fishing vessel crew in making the best decisions to improve vessel energy efficiency and fish quality. The specific operational choices which will be addressed are:

- Propulsion mode (diesel electric, diesel mechanic, hybrid).
- Which auxiliary engine(s) to run.
- Using the main engine(s) for electricity production?
- Loading of the vessel to reduce wave movements and thereby increase catch quality.
- Loading of the vessel to reduce resistance.

Method

Operational data have provided valuable insight in how the operational choices affects fuel economy for fishing vessels. To increase the usefulness of such data, the considerable effect of parameters such as waves, wind, and load condition must be taken into account. This pilot will do this by combining onboard measurements with available meteorological and oceanographical data, so that the connection between more parameters can be more accurately modelled. This also makes it possible to remove the noise unmodelled effects inflict on the data, making it possible to study the effect of any operational parameter with more confidence. This includes, but is not limited to, how to operate the vessel in the most economical way in terms of propulsion mode, loading and use of auxiliary and main engines.

Another aspect which will be treated in this pilot, is how to reduce the degradation of catch quality because of vessel movements. This will be applied to vessels with RSW (refrigerated sea water) fish holding tanks. Based on large amounts of catch reports (time, species, size and position), coupled with the corresponding landing quality reports and the weather conditions along the vessel route, one wants to investigate and quantify 1) how vessel movements are affected by environmental conditions and operational choices (speed, trim, displacement) and 2) how fish quality is affected by vessel movements, species, initial catch condition and storage parameters (temperature, filling, design).

Challenges

The challenges in this pilot are common to those in the A1 Pilot above, but excluding predictive maintenance, the main focus is on fuel reduction, energy optimization and to reduce degradation of catch quality.

Relevance to and availability of Big Data and Big Data infrastructure

The connection between operational choices and its consequences are difficult to reveal, since many hard to measure parameters and effects come into play. It is therefore assumed that employing big data methods, such as machine learning, to existing datasets will prove useful. By combining large amounts of vessel measurements with meteorological and oceanographic hindcast, it will be possible to analyse how the vessel loading affects vessel movements and energy hull resistance. By combining measurements of vessel movements with reports of quality of landed fish, the relation between movements and catch quality can be analysed.

For a number of vessels:

- 10 Hz 6DOF accelerations and velocities of the vessel
- 1 Hz operational data (speed, energy consumption, machineries loads, propulsion choices)

For all Norwegian vessels:

- Catch reports for all catches
- Sales reports all landings
- Quality reports from vessels for all landings
- Quality reports from landing site for all landings
- Meteorological data (wind and waves)

SINTEF Marine Data Center employs big data tools such as Apache Hadoop, Apache SPARK and GlusterFS for storage and analysis of incoming operational data from fishing vessels. Integration against other data sources, such as oceanographic and meteorological is planned.

Benefit of pilot

This pilot will focus on the small pelagic fishing fleet, covering the North Atlantic Ocean. The Norwegian pelagic fishing fleet will be the main stakeholders. In this project they are represented by the ship owning companies Ervik & Saevik, Eros, Kings Bay, and Liegruppen. The main research partner is SINTEF Ocean.

Initially, the Norwegian pelagic fishing fleet will benefit from this pilot. But the results within energy efficiency will be general and easily transferable to both other fishing vessel types and other types of ships where the operational choices affecting energy efficiency is not trivial.

Within energy efficiency, SINTEF Ocean is working on a parallel project with much of the same aim, and a demonstrator is expected to be ready as a starting point for this pilot. For the work within the relationship between fish quality and vessel loading, the data gathering and onboard components are partly ready. The main work to be done in this pilot will be to develop Big Data methods and tools to take advantage of additional data sources. The average Technology Readiness Level of this pilot is approximately 4

4.2 Pilot case definition

Table 8: Summary of Fishery Pilot A2: Small pelagic fisheries immediate operational choices

Use case title	Small pelagic fisheries immediate operational choices	
Vertical (area)	Fishery	
Author/company/email	Karl Gunnar Aarsæther /SINTEF Ocean / karl.gunnar.aarsather@sintef.no	
Actors/stakeholders and their roles and responsibilities	Vessel masters and managing companies. The vessel master makes immediate operational choices for the vessel at sea, while the managing companies pay the operational expenses incurred by fuel usage and maintenance. The vessel master is responsible for managing the vessel at sea, while the managing companies are responsible for the economics of vessel operations.	
Goals	Provide decision support for more efficient fisheries through operational choices on-board the vessel	
Use case description		
Current solutions	**Compute(System)**	Prototype system for presentation of current situation with regards to past performance. Data collection system for vessel power system.
	Storage	Local storage in NetCDF files. Remote storage of historic data not accessible on-board
	Networking	Remote management of data collection system and uploading of historic data.
	Software	SINTEF Ocean Ratatosk signal multiplexer, logging and hardware interface system. Prototype stand-alone application on bridge with display.
Big data characteristics	**Data source (distributed/centralized)**	Most data sources located on the vessel. Distributed access points, but power system interfaced through IAS. Incorporation of EO data

		suggested as a complementary data source.
	Volume (size)	Power system logs are about 50mb each day in NetCDF binaries.
	Velocity (e.g. real time)	Real time measurements each second. Approximately 50-100 measurement channels for time series data.
	Variety (multiple datasets, mashup)	Time series measurements may be collected from a variety of sources. ● Power system ● Navigation system ● Weather sensors ● Deck machinery Additionally, the following sources which do not present straightforward time-series will be evaluated for inclusion ● EO environment ● Sonar/hydroacoustic
	Variability (rate of change)	It is expected that the rate of change will be different from onboard time-series measurements (seconds) and EO data and hydroacoustics. The time-series data can be expected to have an rate of change from 10Hz to 1hz
Big data science (collection, curation, analysis, action)	**Veracity (Robustness Issues, semantics)**	IAS data with high veracity. Exception is sensor faults which could be detected.
	Visualization	Visualization of expected results from operational changes. Correlations between measured data and historic performance.
	Data quality (syntax)	Data may contain missing values, or measurement noise. Data must be curated before storage and use in analyses as to limit errors from such errors.
	Data types	Structured time series and possible EO data.
	Data analytics	Predictive analytics for data driven decision support for the improvement of vessel operations in the immediate future
Big data specific challenges (Gaps)	There is a need, or desire, to utilize the large amount of data which is already produced, or may be collected, on-board vessel to improve their operation. The data from a variation of sources (ex: machinery, deck equipment, meteorological, loading computer) must be collected in a system which allows them to be utilized to improve vessel operations. The modern	

	vessels have numerous ways to operate the vessel, which differ in optimality for the current situation. It is therefore possible to advise the master to take corrective actions to improve vessel efficiency, and such advice can be produced by collecting statistics of vessel operations and comparing the current vessel state to the vessels operational history.
Big data specific challenges in bioeconomy	Vessel operations at sea and cost of satellite bandwidth necessitates either minimal bandwidth usage, or asynchronous operations with synchronization of data and analyses results during periods when the vessel(s) are within range for less costly communication systems (ie. vessel domestic cellular data service).
Security and privacy technical considerations	Operational data is business sensitive information and access should be restricted to authorized parties.
Highlight issues for generalizing this Use case (e.g. for ref. architecture)	Predictive analytics and analyses methods developed should be general in terms of application to other DataBio pilot applications.
More information (URLs)	http://www.databio.eu
Note: <additional comments>	

4.2.1 Stakeholders and user stories

The end users of this pilot are the masters on-board small pelagic fishing vessels. The pilot will develop tools to advise on energy consumption and effective methods of operating the vessel given the current vessel state and environment. The ship operating companies are stakeholders since the pilot attempts to reduce fuel consumption and machinery running hours during operation, and hence attempts to reduce operational costs.

Table 9: Stakeholders and user stories of Fishery Pilot A2: Small pelagic fisheries immediate operational choices

Stakeholders	User story	Motivation
Vessel master	As a master onboard a small pelagic fishing vessel I want to be able to quickly gauge the operational efficiency of my vessel. The feedback on operational efficiency should be based on the vessels past operation as to show what is feasible, and not "optimal, but unachievable". The feedback must be instant and simple as to allow me to take corrective actions fast without interfering with my other activities.	Optimize fuel and time expenditure during operations

Operation manager	As an operational manager for a fleet of fishing vessels I want to be able to judge the fuel consumption and operational philosophy on-board a vessel based on data. The varying environmental conditions and individual master choices makes it difficult to ascertain if my vessels are operated as optimal as I could expect. Actual data from vessel operations will help me understand the vessel operations.	Minimize fuel and time expenditure

4.2.2 Motivation and strategy

The main motivation for this pilot is to

- Improve catch efficiency through improved energy efficiency
- Improve vessel operations by use of acquired data originating on the vessel in combination with meteorological data.

The pilot motivation and strategy is summarized using ArchiMate diagrams in the next section, while goals and KPIs are addressed in the successive evaluation plan.

4.3 Pilot modelling with ArchiMate

The table below lists the optional views in this chapter, which comes in addition to the required motivational and strategy views introduced in the summary. These additional views where designed for this pilot since it was the ArchiMate modelling example for all the pilots in the DataBio project.

Table 10: Fishery A2 application view and business process view

Name	Type	Description
Fishery A2 Application view	View	This view represents the Application of the A2 Pilot. This diagram should be related to the BDVA diagram in the Big data assets section below.
Fishery A2 Business Process view	View	This view represents the A2 Pilot as seen from a business process perspective.

4.3.1 DataBio fishery pilot A2 motivation view

This chapter describes the "DataBio Fishery Pilot 3.2.2 Motivation view" view defined in the "Fishery A2 Small pelagic fisheries immediate operational choices modelling with ArchiMate" view point (Figure 14).

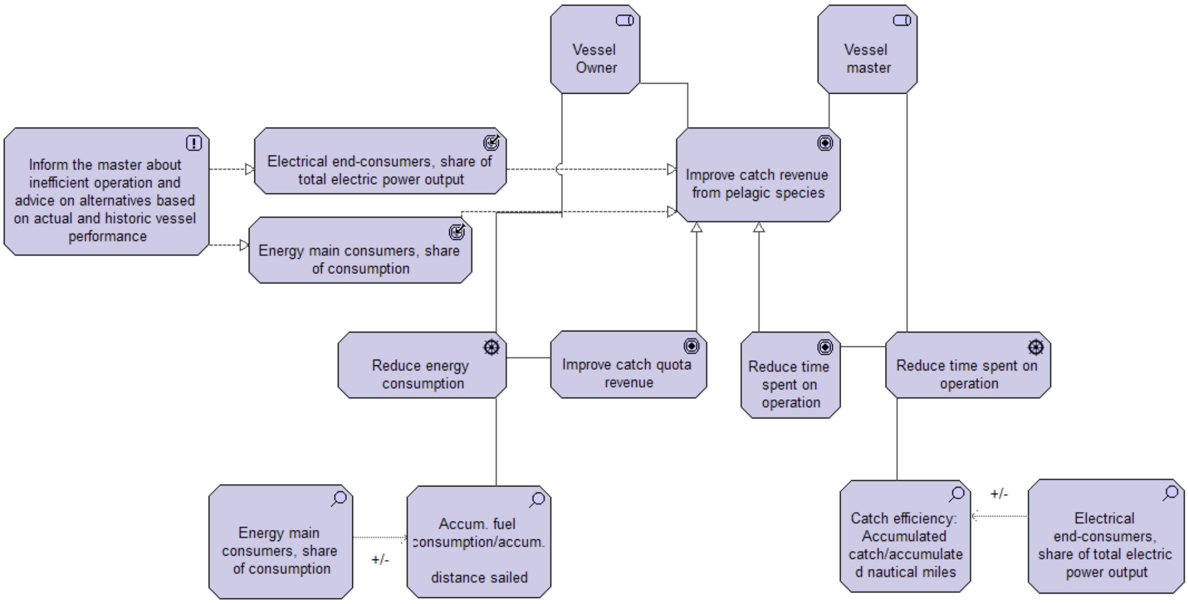


Figure 14: Fishery Pilot A2 Motivation view

The motivation of the vessel master and vessel owner is to reduce operational costs by reducing time expenditure and fuel consumption. These factors all influence the aggregate goal of increasing catch revenue from the resource base, in this case pelagic species.

4.3.2 DataBio fishery pilot A2 strategy view

The underlying strategy of the decision support system is to apply data storage and processing to augment the vessel master's perception about the operation of the vessel. The strategy model is seen in Figure 15. The immediate decision support system is built on top of a data collection and distribution system. The data collection and distribution system is used to collect sensor data from the on-board systems and makes them available in a single system. The data distribution system ensures that the decision support system only interface with a single system, instead of multiple sensors. The decision support system presents the data from the data distribution system and collect them in an internal storage system for presentation of current performance vs. historic performance.

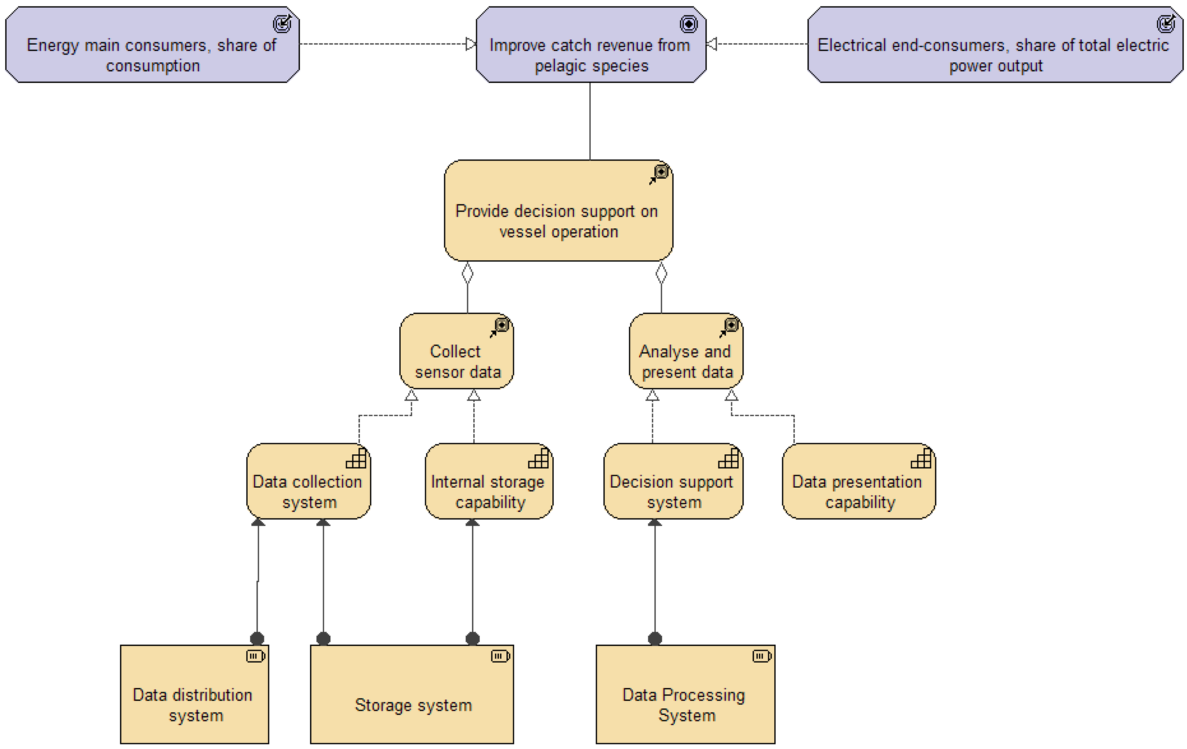


Figure 15: Fishery Pilot A2 Strategy view

The table below lists the current view's main elements.

Table 11: List of the elements defined in the Fishery A2 strategy view

Name	Description
Analyse and present data	The ability to process acquired data using a set of rules and algorithms. Present the result to the user in a preferred format on the preferred devices, e.g. map layer on a chart plotter.
Collect sensor data	The ability to collect measurements from a sensor device through a robust interface
Data Processing System	System with the ability to process data
Data collection system	System with the ability to collect data
Data distribution system	System with the ability to distribute data between nodes
Data presentation capability	System component with the ability to present data to an user.

Decision support system	System providing decision support to a process or user.

4.3.3 DataBio fishery pilot A2 business view

The business process of fisheries is inevitably about balancing revenue from catch against operational expenses incurred from crew, maintenance and fuel. Availability of fisheries resources are scarce and the fisherman is limited by the quota. Therefore, the only control the fisherman, or managing company, exerts on the immediate business of fisheries is to increase operational efficiency during location and catching of the fish. The business view of the small pelagic fishing operation is seen in Figure 16.

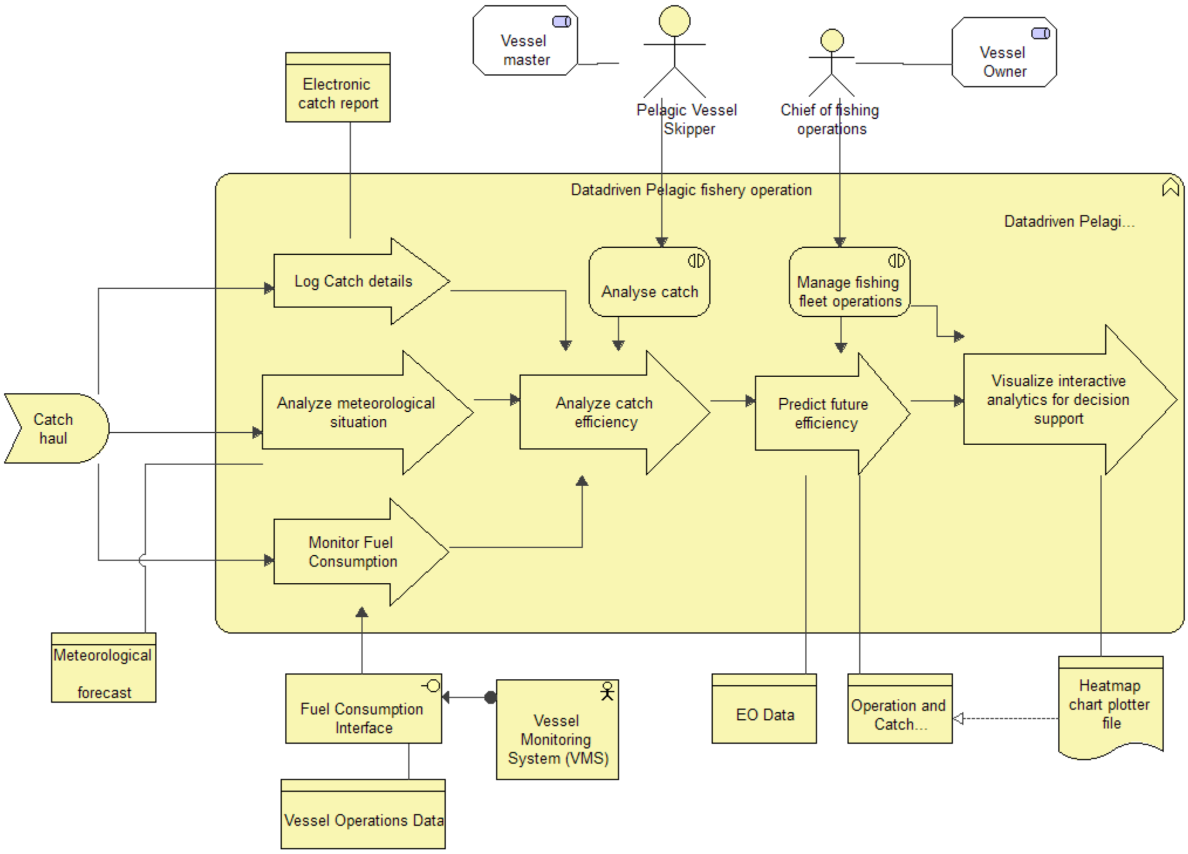


Figure 16: Fishery Pilot A2 Business process view

The table below lists the current view's key elements.

Table 12: List of the elements defined in the Fishery A2 Business Process view

Name	Description
Analyse catch	When the catch is onboard it is analysed manually or automatically using graders. Estimation of catch content (species, weight, size distribution, quality etc) is made.
Analyse catch efficiency	Divide catch by aggregated energy load in catch period
Analyse meteorological situation	Evaluate current weather conditions and weather predictions
Heatmap chart plotter file	Check plotter file for information
Log Catch details	Record the details of each hauls to comply with regulations and build knowledge base
Manage fishing fleet operations	Direct and supply the fishing vessels in ports with catch landing capability and resupply service
Meteorological forecast	Weather forecasts of wind, waves and ocean currents for use in planning of operations
Operation and Catch Efficiency prediction	Predict how much fuel and time must be spent to catch target fish
Predict future efficiency	Predict the future performance of the fisheries on the vessel
Vessel Monitoring System (VMS)	The vessel monitoring system provides energy and fuel consumption statistics for the crew and management
Vessel Operations Data	Accumulated data of the energy and fuel consumption can be stored and accessed on the vessel.
Visualize interactive analytics for decision support	Visualize the current situation from measurements and predictions. Take action from presented information

4.3.4 DataBio fishery pilot A2 application view

This chapter describes the "Fishery A2 Application view" seen in Figure 17. This figure shows the view defined in the "Fishery A2 Small pelagic fisheries immediate operational choices modelling with ArchiMate" view point.

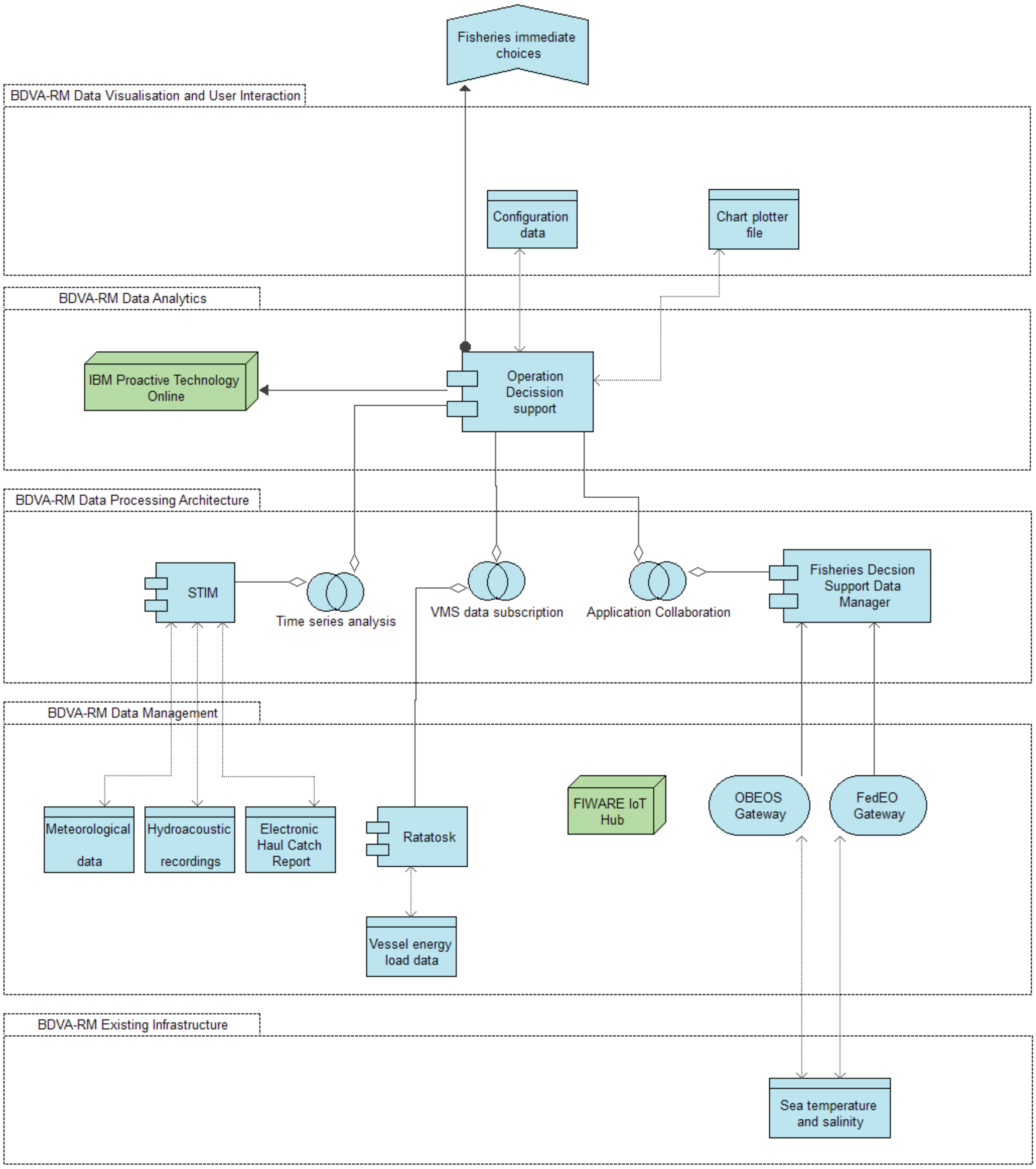


Figure 17: Fishery Pilot A2 Application view

This view represents the Application of the Pilot 3.2.2. The elements of the pilot implementation are functionally in the entire stack of the BDVA diagram from existing infrastructure through data management, processing, analytics and visualization. The on-board sensors at the existing infrastructure level are connected to a selection of possible data management components. The data management components then deliver data to the processing step before analytics is applied and results visualized. The existing components in use on board the pilot vessels today cover data management, and processing while there are possibilities for improvement in additional management, data sources from earth observation and other networked sources. The components in the analytics and visualization layers of the

BDVA diagram is expected to improve the pilot implementation form the current existing state.

4.4 Pilot Evaluation plan

The high-level evaluation plan for the pilot is shown in Figure 18. The evaluation plan is focused on implementation of technologies and studying the effects these have on the operation of the vessel.

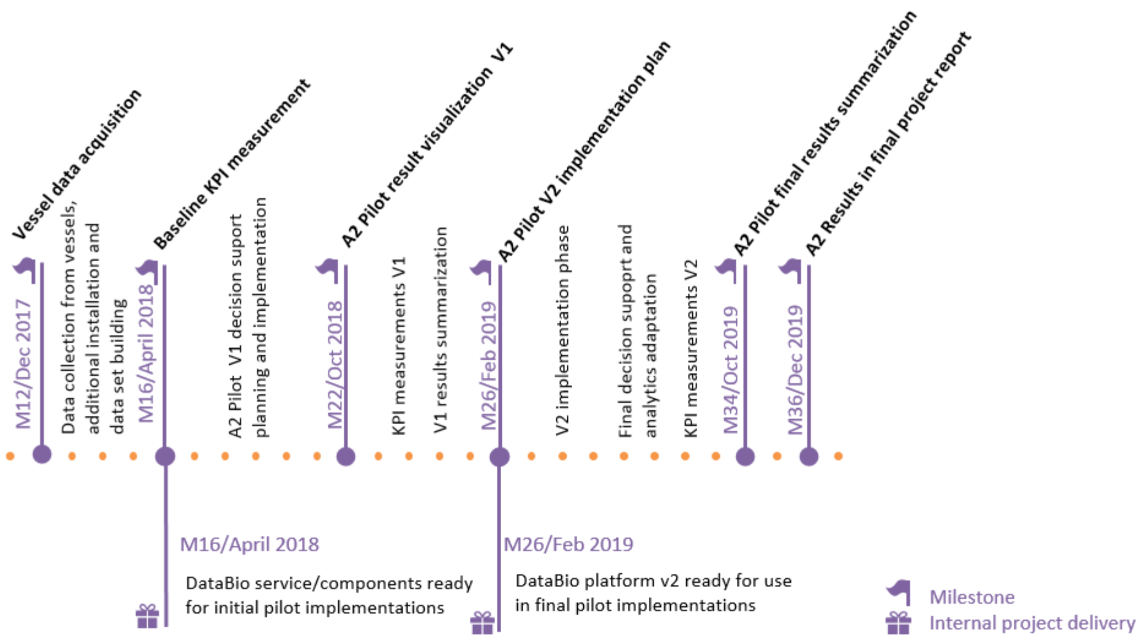


Figure 18: Fishery Pilot A2 Milestone timeline

4.4.1 High level goals and KPIs

The following KPIs have been proposed as an initial set of measurements for the pilot application. These KPIs for the master of the vessel correspond to an "Assessment" in the ArchiMate modelling framework.

- Catch efficiency: Accumulated catch/accumulated nautical miles in [kg/m]. This shows the steaming distance required in order to catch fish and crew and fuel expenditure scales with increased steaming distance.
- Accumulated fuel consumption/accumulated distance sailed [kg/m]. Fuel consumption (in tonnes or kg) pr sailed distance shows the effectiveness of the vessel and machinery system. This is both a function of machinery operation, and of nautical know how on how to save fuel with different cruising speeds.
- Electrical end-consumers, share of total electric power output [-]. This KPI is the share of the electrical consumers power, which will together with main engine load indicates if there is extra capacity on the main engine to handle electric power production.
- Energy main consumers, share of consumption [-]. This shows if a single power source

dominates consumption.

The goal for the pilot is formulated as ArchiMate goals influenced by the drivers:

- Reduce energy consumption
- Reduce time spent on operations

These combine to fulfil the overall goals of the pilot which are formulated, in the motivation view, which is economically motivated and aims to improve and rationalize vessel operations.

- Improved catch quota revenue
- Reduced time spent on operation
- Improved catch revenue from pelagic species

4.5 Big data assets

The diagram below summarizes Big Data technology components used in this pilot using the extended BDVA reference model. Where applicable, specific partner components that are likely to be used or evaluated by this pilot are listed using the DataBio component identification labels.

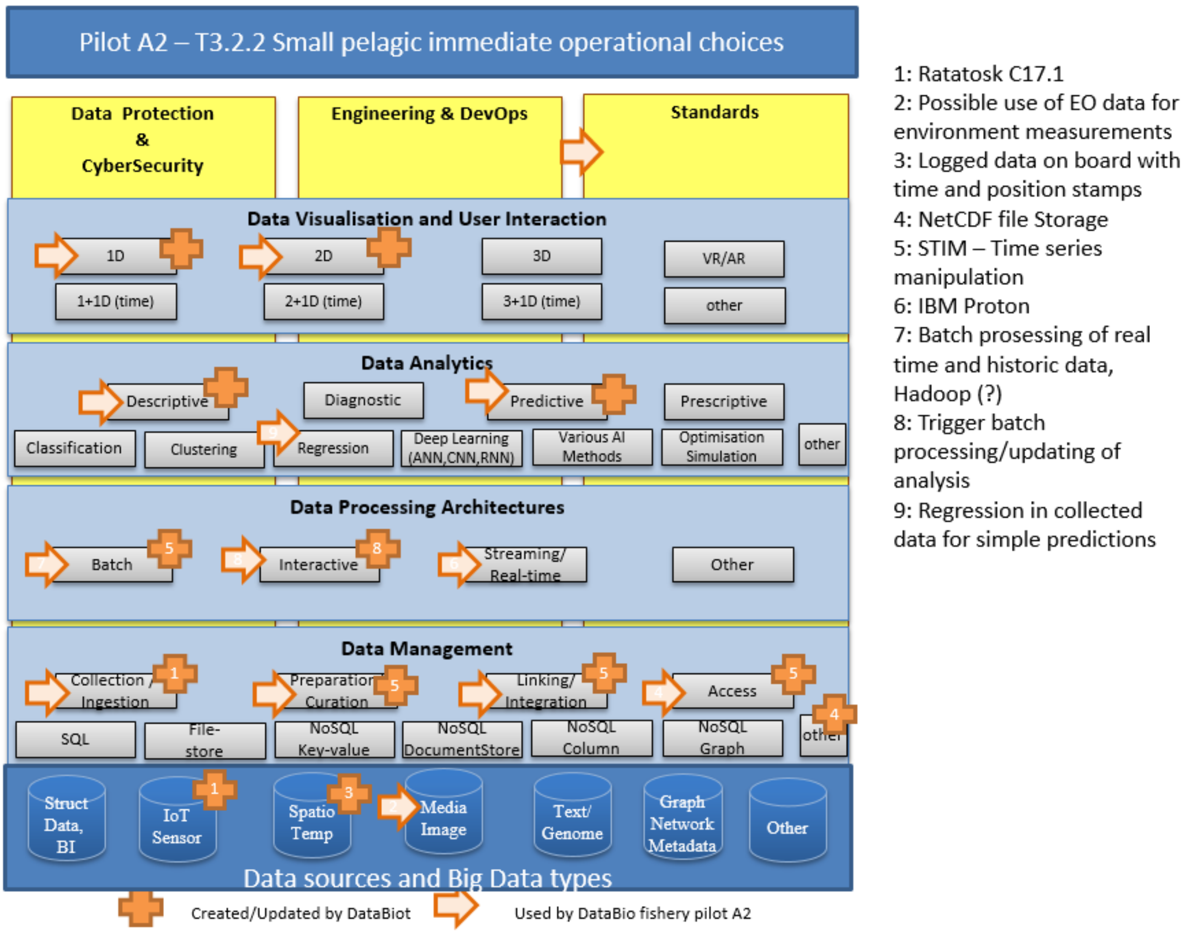


Figure 19: BDVA reference model for fishery pilot A2

5 Fishery Pilot B1: Oceanic tuna fisheries planning

5.1 Pilot overview

The purpose of this pilot is to improve profitability of oceanic tuna fisheries through savings in fuel costs through fish observation and route optimisation.

5.1.1 Introduction

Since all international bodies assessing the status of tuna populations advice not to increase the catch of tuna the approach of the pilot is instead to improve the performance of the fishery by means of saving costs thus generating more profit fishing the same amount of fish.

Purse seine is the surface gear that contributes most to the catch of yellowfin and skipjack globally. In the purse seine fishery, three main fishing strategies are used to capture tunas: (1) targeting fish swimming in free schools, (2) targeting fish swimming around drifting objects, (3) targeting fish associated with dolphins (only in the case of Eastern Pacific Ocean), and in some isolated cases associated with whales or whale sharks. In the first approach, called a free-school set, a school of fish is identified from evidence in the water's surface, and is captured by encircling it. In the second approach, a drifting object where fish are aggregated is encircled with the net. Within this second strategy, there are a subset of techniques including sets on encountered "natural" floating objects ("log sets"), and sets on fish aggregating devices (FADs). FADs are floating objects that have been modified and placed in the fishing areas by the fishers to attract fish, and to facilitate their aggregation and capture. Additionally, FADs are often outfitted with a buoy to help fishers locate them. The strategy of using FADs was developed in the 1980s, but greatly increased in use during the 1990s.

A deciding factor for the energy efficiency of the fishing vessel, is the ability to choose the FADs with more probability of having big or profitable tuna aggregations, combined with the best route available to visit and/or fish all this FADs in the most energetically efficient way. Such decisions are today based on the skills and knowledge of the skipper or not took in account at all, being one of the main areas to improve the profit of the fishing industry without harming the state of the stocks.

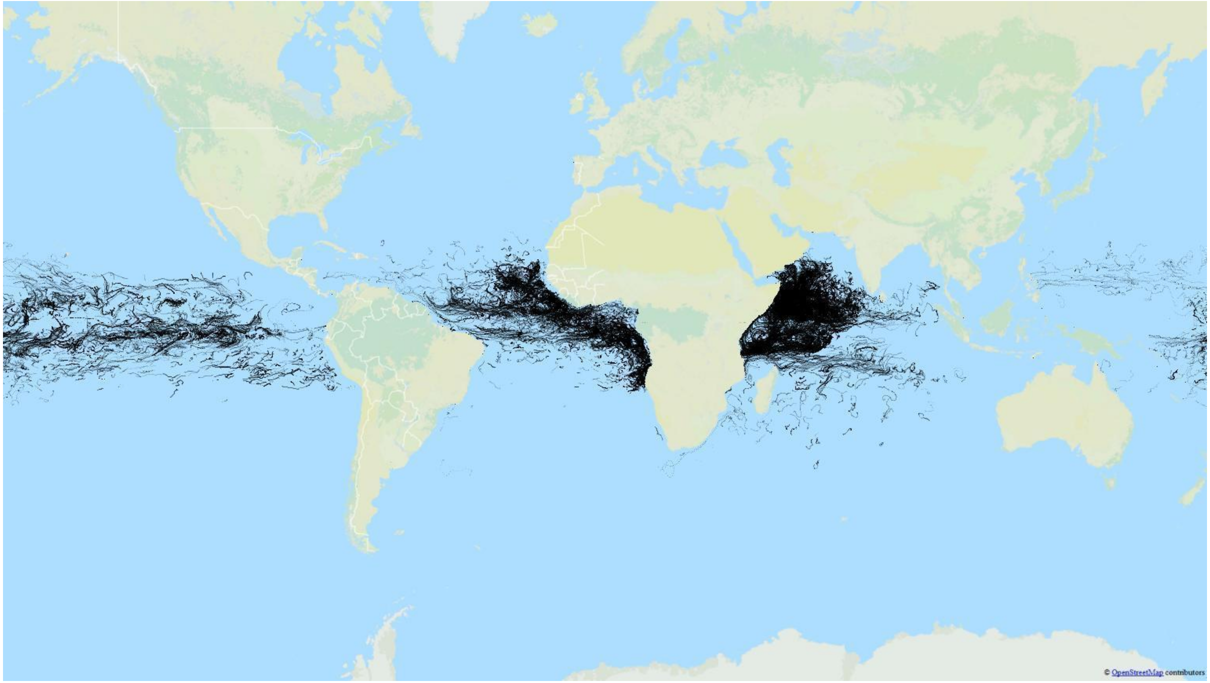


Figure 20: Plot of the active buoys deployed by the whole Basque fleet for one month of 2009, each black dot is the position sent from the buoy to the vessel (n=1.250.000).

5.1.2 Overview

Method/description

The methodology for the first objective will be the use of big data machine learning techniques in two ways

1. Unassisted machine learning, all the historical data (catch logbooks, GPS, Buoys, Observers, etc.) will be put in relation in different databases and used to learn the better way to fish with the FAD method.
2. The same databases will be used to construct a "virtual bridge" where actual skippers will be confronted against past fishing situations in a simulator-game like interface, in order to assist machine learning algorithms to extract the key variables involved in the decisions took by the skippers about which is the better FAD to go for fishing.

The second and third objectives will be addressed with the algorithms developed in the Pilot B1 using the probabilities of having tuna in the FADs from objective 1 for choosing the "waypoints" of the fishing routing of one vessel or the way a cooperative fleet has to sail for fish, saving the larger amount of fuel.

Challenges

A central challenge is to develop models to have a probability of having a given amount of tuna at a given time in a concrete FAD. In fact, this is exactly what a good skipper is calculating in his mind when taking decisions in the bridge in real-time. Developing such algorithms will allow to integrate these decisions with all the other available data in order to do the more efficient sailing for saving fuel.

The second challenge is to use this system to transmit the empiric knowledge of the skippers to the new generations since there is a problem of generational change in all European fisheries.

Finally, for optimizing the sailing the challenges are the same as in Pilot B1.

Relevance to and availability of Big Data and Big Data infrastructure

The goal of this pilot is to provide the skipper and ship owners with information which benefits fisheries planning. This information will be provided based on historical extensive data sets within fisheries activity (AIS, VMS, GPS tracking), catch statistics (logbooks), oceanographic conditions (SST, salinity, chlorophyll), meteorological aspects, data from buoys in the FADs (GPS data, echosounder data, SST), Engine data (engine parameters like RPMs, fuel consumption). Large amounts of historical data combined with machine learning and real time information may be able to predict how the fisheries best can be performed.

Historical data:

- Data from buoys (GPS, echosounder)
- Data from vessels engines sensors (fuel consumption, rpm...)
- Data from AIS, VMS and vessels own plotters (vessel routes)
- Data from observers on-board (catch and by-catch data, visits to FADs without fishing)
- Data from logbooks (catch by day)
- Oceanographic conditions (remote sensing: SST, Chlorophyll, height of sea level, currents)

Real (or near to) time data:

- Data from buoys (GPS and echosounder)
- Data from vessel engines sensors
- Position of the vessels
- Remote sensing (other partners?)
- Meteorological conditions

AZTI's Marine Datacenter. Currently has implemented different databases for AIS, Buoys, VMS, Catch statistics, Observers data and Remote sensing data.

Integration against other data sources, such as catch information, oceanographic and meteorological is planned.

Benefit of pilot

South Atlantic and Indian Oceans are the main fishing grounds for the Basque Oceanic Tuna Purse Seine fleet but Europe has vessels all around the world.

Echebastar Fleet will be the main beneficiary of the developed technology in a first step, but if successful the creation of a company for exploitation of the technology as a service to other vessels owner is expected,

The whole Tuna purse seine industry will benefit from the implementation of techniques.

Current TRL: Energy efficiency: 5-6, Routing taking in account Buoys data 2-3

5.2 Pilot case definition

Table 13: Summary of Fishery Pilot B1: Oceanic tuna fisheries planning

Use case title	Oceanic tuna fisheries planning	
Vertical (area)	Fishery	
Author/company/email	Iñaki Quincoces /AZTI/ iquincoces@azti.es	
Actors/stakeholders and their roles and responsibilities	Skipper / Technical staff in the shore office / ship owner Skippers take operational decisions on board and is responsible of catching fish / Technical staff in the shore office is responsible to keep ship running and fishing with the minimum operational cost /Ship owner is the responsible for strategic decisions	
Goals	Develop tools to assist when taking fishing planning decisions regarding fishing operations and navigation routing.	
Use case description	Refer to the pilot case definition section and diagrams in the pilot modelling sections.	
Current solutions	**Compute(System)**	No existing computing system for deciding where to fish, it is a decision of the skipper from the available information
	Storage	none
	Networking	none
	Software	none
Big data characteristics	**Data source (distributed/centralized)**	Both depending on the data: distributed: earth observations, meteorological forecast and observations, fishing logbooks, observers logbooks. Centralized: buoy data, vms data.
	Volume (size)	Without earth and meteorological data, around 0.5 terabyte year. With earth and meteorological data from 1 terabyte to 10 terabyte depending on resolution.
	Velocity (e.g. real time)	Daily for earth observation, every hour for meteorological, VMS, Buoys. Not in real time due to confidentiality

	Variety (multiple datasets, mashup)	HIGH
	Variability (rate of change)	HIGH
Big data science (collection, curation, analysis, action)	**Veracity (Robustness Issues, semantics)**	The software has to give reliable advice to end user to be considered in normal operational condition
	Visualization	Geographical output of the fishing probability model and routes for optimization (Geoserver maps is a possible example)
	Data quality (syntax)	
	Data types	Excel, csv, NetCDF, GeoJSON, JSON
	Data analytics	Machine learning for prediction of probability of having X tons of fish in a given position. Complex events have to be managed in real time in order to produce the bigger fish catch with the lowest fuel consumption.
Big data specific challenges (Gaps)	The real reason for tuna aggregation and for calculating the amount of fish aggregated in a FAD is unknown, but some skippers are very successful in discovering the best locations for fishing whether other are not using the same information. Almost all the available information for a skipper onboard will be used to reproduce the fishing behavior of good skippers.	
Big data specific challenges in bioeconomy	The vessel has a reduced capability of downloading and uploading data, so the final product for the decision making must be reduced in weight. Most of the data to be used for the modelling is generated out of the vessel, the only data fully needed generated by the vessel is its position (GPS data)	
Security and privacy technical considerations	Operational areas and ship position are extremely delicate to pirate presence in Indian ocean and due to possibility to be used by competitors in their fishing operations.	
Highlight issues for generalizing this Use case (e.g. for ref. architecture)	The only issue for generalizing the use case is the type of buoy used by each fishing company since they are equipped or not with echo sounder in order to measure the amount of fish aggregated in the FAD and the accuracy of the measurements is very heterogeneous.	
More information (URLs)	www.databio.eu <other URLs to be added later if relevant>	
Note: <additional comments>		

5.2.1 Stakeholders and user stories

Table 14: Stakeholders and user stories of Fishery Pilot B1: Oceanic tuna fisheries planning

Stakeholder	User Story	Motivation
Ship owner	As owner I would like to keep the income from fish catches and reduce the operational costs with reduced fuel consumption.	Cost reduction with same or higher income will increase my profit.
Skipper	As a skipper I would like to have a tool that recommends me which buoys would provide me with maximum catches and in the same time reduce me the burden of estimating the minimum fuel consumption for each of the possible buoys with catches.	Catch the same amount of fish reducing sailed nautical miles and uncertainty on board.

5.2.2 Motivation and strategy

The main motivations for this pilot is to reduce costs from fuel consumption on board while keeping the amount of catches.

The pilot motivation and strategy is summarized using ArchiMate diagrams in the next section, while goals and KPIs are addressed in the succeeding evaluation plan.

5.3 Pilot modelling with ArchiMate

5.3.1 DataBio fishery pilot B1 motivation view

Figure 21 illustrates the "Fishery B1 Pilot Motivation view" in ArchiMate 3.0 modelling. The main goal and its breakdown in subgoals mirrors that of the previous A1 and A2 pilots, improving catch revenue through reduced energy consumption and time spent on the operation and improving catch efficiency. The three elements to the upper right summarizes the special focus in this pilot in addition to these goals: Fish species distribution forecast and optimized routing.

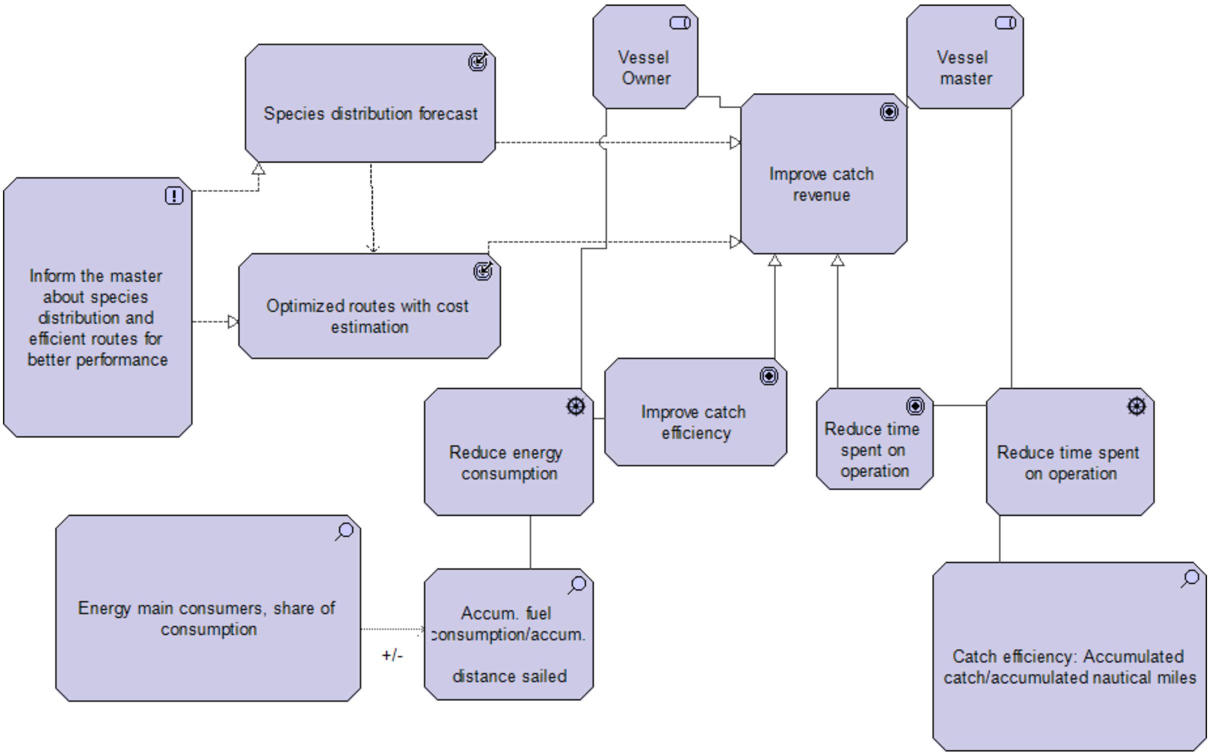


Figure 21: Fishery Pilot B1 Motivation view

5.3.2 DataBio fishery pilot B1 strategy view

The "Fishery B1 Pilot Strategy view" ArchiMate diagram illustrates how the decision support is realized through the collection and analysis of sensor data and presentation of relevant fisheries information for oceanic tuna fisheries planning (Figure 22). The FADs mentioned in the pilot overview section above are the main sensor input for the species distribution forecast, and they are special to this pilot.

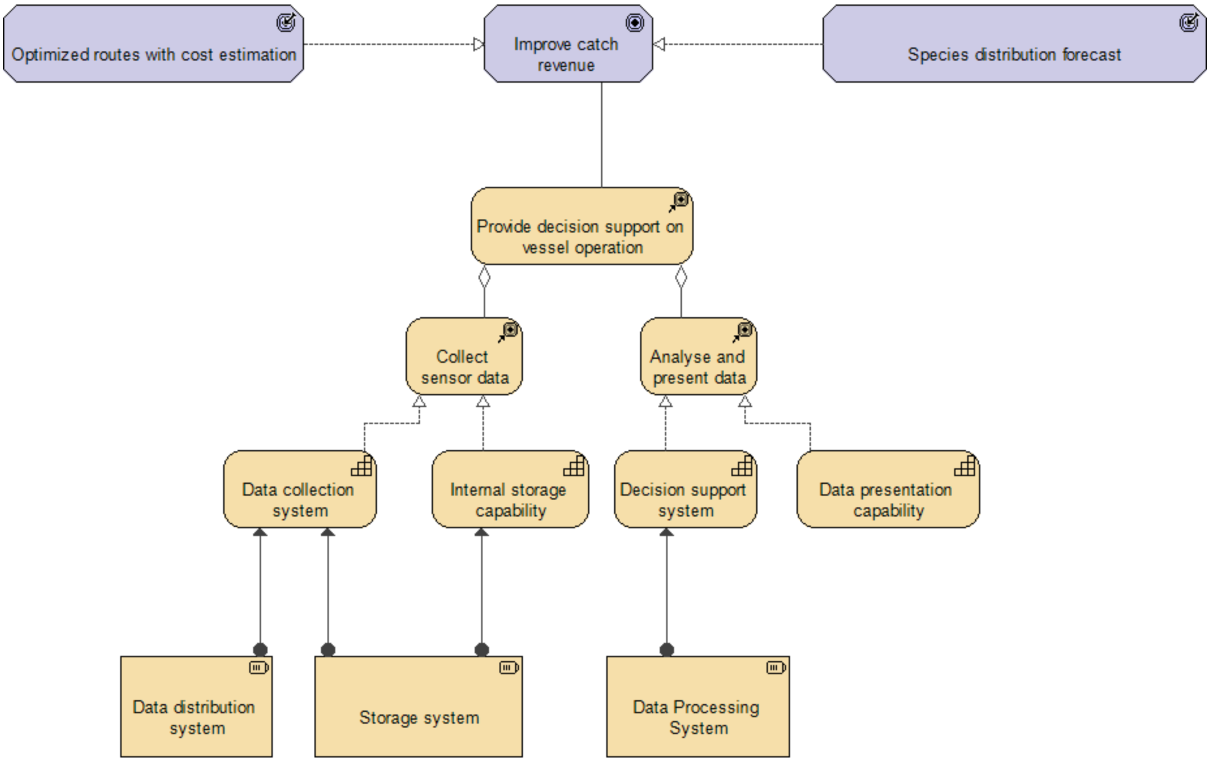


Figure 22: Fishery Pilot B1 Strategy view

Table 15 lists key elements of the Strategy view.

Table 15: List of the elements defined in the Fishery B1 Strategy view.

Name	Description
Analyse and present data	The ability to process acquired data using a set of rules and algorithms. Present the result to the user in a preferred format on the preferred devices, e.g. map layer on a chart plotter.
Collect sensor data	The ability to collect measurements from a sensor device through a robust interface
Data Processing System	System with the ability to process data
Data collection system	System with the ability to collect data
Data distribution system	System with the ability to distribute data between nodes

Data presentation capability	System component with the ability to present data to an user.
Decision support system	System providing decision support to a process or user.
Improve catch revenue	The link between strategy and motivation goes through the Goal element

5.4 Pilot Evaluation plan

The high-level evaluation plan for the pilot is shown in Figure 23. The evaluation plan is focused on data gathering from satellite, oceanographic models and FAD buoys and catch data and the subsequent pilot implementation stages.

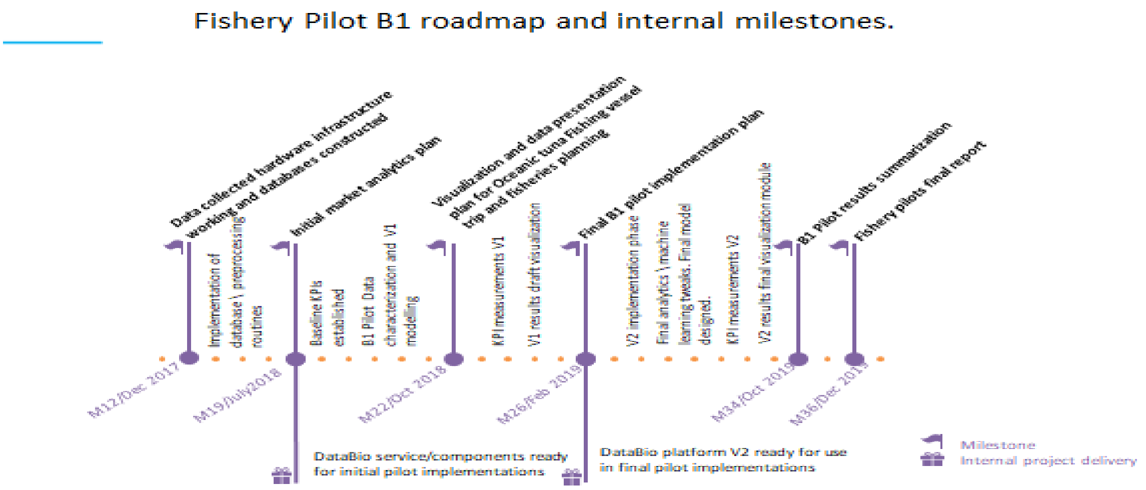


Figure 23: Fishery Pilot B1 Roadmap and internal milestones

5.4.1 High level goals and KPIs

The high-level goal is to better understand the geographical aggregation patterns of the fish, to better plan trips, and hence save fuel and time.

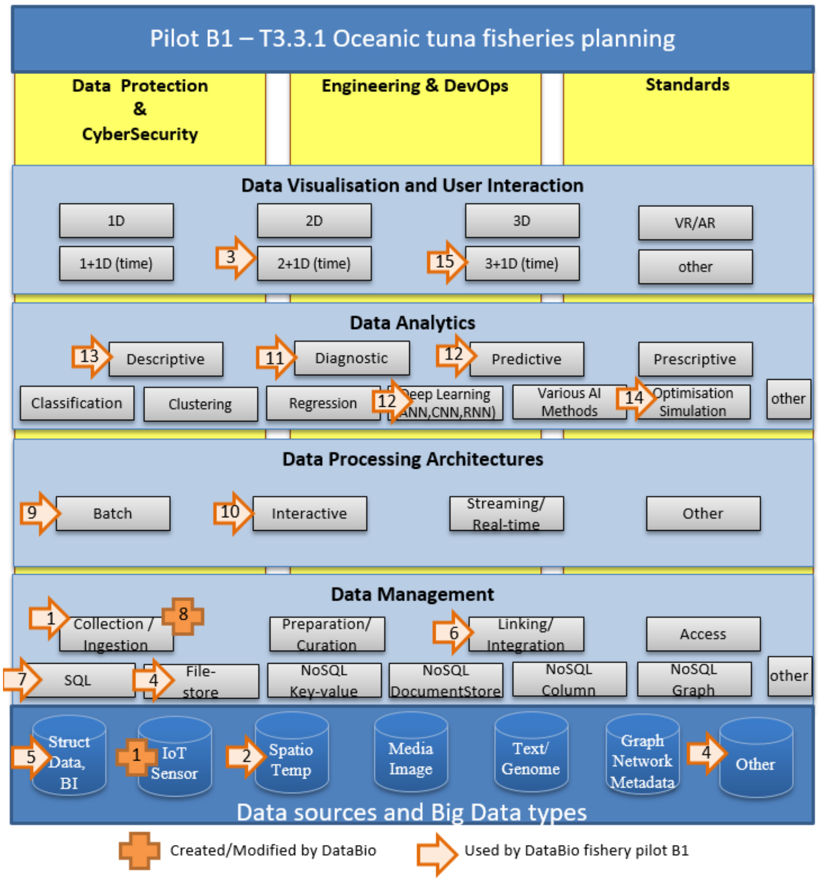
At the moment the relevant KPIs identified are:

- Fuel consumed per sailed nautical mile.
- Fuel consumed per catch unit of mass [kg fuel/fish ton or kg]
- Miles sailed per catch unit of mass [Nautical mile/fish ton or kg]

The baseline values will be obtained through analysis of historical data. The baseline values will be used as reference to evaluate the proposed solutions.

5.5 Big data assets

The diagram below summarizes Big Data technology components used in this pilot using the extended BDVA reference model.



1. VMS, FAD Buoys.
2. Logbooks, Observers database, Earth Observation, Metereologic.
3. Plot of performance parameters models/time – R (Shiny)
4. Catch reports for KPI development
5. Text format data
6. Fuel consumption models with weather data
7. Interrelation of spatiotemporal data from diferent sources
8. SW component gathering data from 4 and 5
9. Batch processing of historic data Bash, SQL, Bigdata programs
10. Interactive and iterative development
11. Amount of catch and all the other parameters as hindcast of ML methods
12. ML methods for predictive catch probability linked to fuel consumption models for better route forecasting.
13. Correllation/Covariance analysis
14. Optimization of operational choices
15. Maps of catch probability with best sailing route (Geoserver for serving maps generated by R or other ML software)

Figure 24: BDVA reference model for fishery pilot B1

6 Fishery Pilot B2: Small pelagic fisheries planning

6.1 Pilot overview

6.1.1 Introduction

A deciding factor for the energy efficiency of fishing vessels, is the ability to find the best suitable fishing grounds, combined with the best suitable fishing methods and tools. Such decisions are today based on the skills and knowledge of the crew.

6.1.2 Overview

The goal of this pilot is to provide the crew and ship owners with information which benefits fisheries planning. It will run in two phases. The first phase will focus on providing an interactive end user tool for browsing compilations of past history, while the second level will provide the end user with a tool which will do predictions of where and when fish can be caught in an energy efficient manner.

Method/description

This pilot will collect and combine as many data sets as possible which may be of interest for predicting/planning of future fisheries effort. Extensive data sets within fisheries activity and catch statistics will be combined with information from that time and history of the same, such as meteorological and oceanographic data, moon phase, time of day, time of year. Machine learning will be used to form a predictive model, which will then be used for predictions of fish availability.

This information will provide insight of how the fisheries best can be performed, and will give valuable input for the planning process.

Relevance to and availability of Big Data and Big Data infrastructure

Large and varied data sets are needed for this pilot. Information about all pelagic catches that have been landed in Norway the last decades will be provided by Norges Sildesalgslag, as well as current data on the same. Meteorological and oceanographic hindcasts and forecasts will be provided by e.g. the Norwegian Meteorological Institute. Operational data will be provided by the participating fishing vessels.

The analyses and data aggregation will be performed at the SINTEF Marine Data Center, using Big Data tools such as Apache Hadoop, Apache SPARK and GlusterFS for storage and analysis.

Benefit of pilot

This pilot will benefit the small pelagic fishing fleet, typically covering the North Atlantic Ocean, with the Norwegian pelagic fishing fleet as the main stakeholders. In this project they are represented by the ship owning companies Ervik & Saevik, Eros, Kings Bay and Liegruppen. The main research partner will be SINTEF Ocean.

The developed methods will in general be easily transferable to other fisheries.

There is an ongoing Norwegian project looking into some of the same issues within bottom trawl fisheries, and synergies are expected between these projects. Current TRL: 3.

6.2 Pilot case definition

Table 16: Summary of Fishery Pilot B2: Small pelagic fisheries planning

Use case title	Small pelagic fisheries planning	
Vertical (area)	Fishery	
Author/company/email	Jørgen Haavind Jensen / SINTEF Ocean / jorgen.jensen@sintef.no	
Actors/stakeholders and their roles and responsibilities	Vessel captain/Shipping company	
Goals	Use historical data to forecast geographical fish stock distribution for the purpose of planning trips.	
Use case description		
Current solutions	**Compute(System)**	NA
	Storage	NA
	Networking	NA
	Software	NA
Big data characteristics	**Data source (distributed/centralized)**	Both. Distributed data from ships, and centralized data from satellites, catch reports, etc.
	Volume (size)	Unknown. Probably order of TB
	Velocity (e.g. real time)	Varying. Some data feeds are on the order of 1Hz, other can be much slower (days, months).
	Variety (multiple datasets, mashup)	Fairly heterogeneous data. Vessel time series, discrete geographical events, satellite map data, etc.
	Variability (rate of change)	Similar to data velocity
Big data science (collection, curation, analysis, action)	**Veracity (Robustness Issues, semantics)**	
	Visualization	2D map forecast/hindcast
	Data quality (syntax)	
	Data types	Time series from vessels. Geo-located catch events 2D time series from satellites
	Data analytics	
Big data specific challenges (Gaps)	Establish automated dataflow from all relevant sources	
Big data specific challenges in bioeconomy	A system intended for use at open sea faces severe network connectivity challenges, both for on-shore data logging and the use of web based services. The vessel system must be able to operate for days without Internet access based on the latest information that was available, and on-shore systems as data	

	loggers and catch reporting must resume operation when network connections are re-established.
Security and privacy technical considerations	Data from vessels are potentially sensitive
Highlight issues for generalizing this Use case (e.g. for ref. architecture)	
More information (URLs)	www.databio.eu
Note: <additional comments>	

6.2.1 Stakeholders and user stories

Table 17: Stakeholders and user stories of Fishery Pilot B2: Small pelagic fisheries planning

Stakeholders	User Story	Motivation
Vessel captain Shipping company	As a captain or ship owner I want to know the likely future distribution of fish stocks in order to better plan my route	Reduction of fuel- and resource-consumption, and time spent at sea

6.2.2 Motivation and strategy

The main motivation for this pilot is to improve catch revenue through improved fisheries planning, that is improving catch efficiency through reducing the time spent looking for fish and the fuel consumption. The strategy to achieve this goal is through providing an information service with historical catch data and estimated species distribution coupled to a model for zooplankton distribution to reduce the time for deciding on the fishing grounds.

The pilot motivation and strategy is summarized using ArchiMate diagrams in the next section, while goals and KPIs are addressed in the succeeding evaluation plan.

6.3 Pilot modelling with ArchiMate

6.3.1 DataBio fishery pilot B2 motivation view

The ArchiMate 3.0 "Fishery B2 Motivation view" diagram is shown in Figure 25. The main goal is to improve catch revenue through improved efficiency and reduced energy consumption and time spent on the fishing operation, similar to was seen for the A2 pilot. The special motivation elements for the B2 pilot can be seen to the left of the figure: Fishery decision support through species distribution information (catch reports and forecast) coupled with the SINMOD oceanographic model with plankton distribution, and routing optimization to fishing grounds.

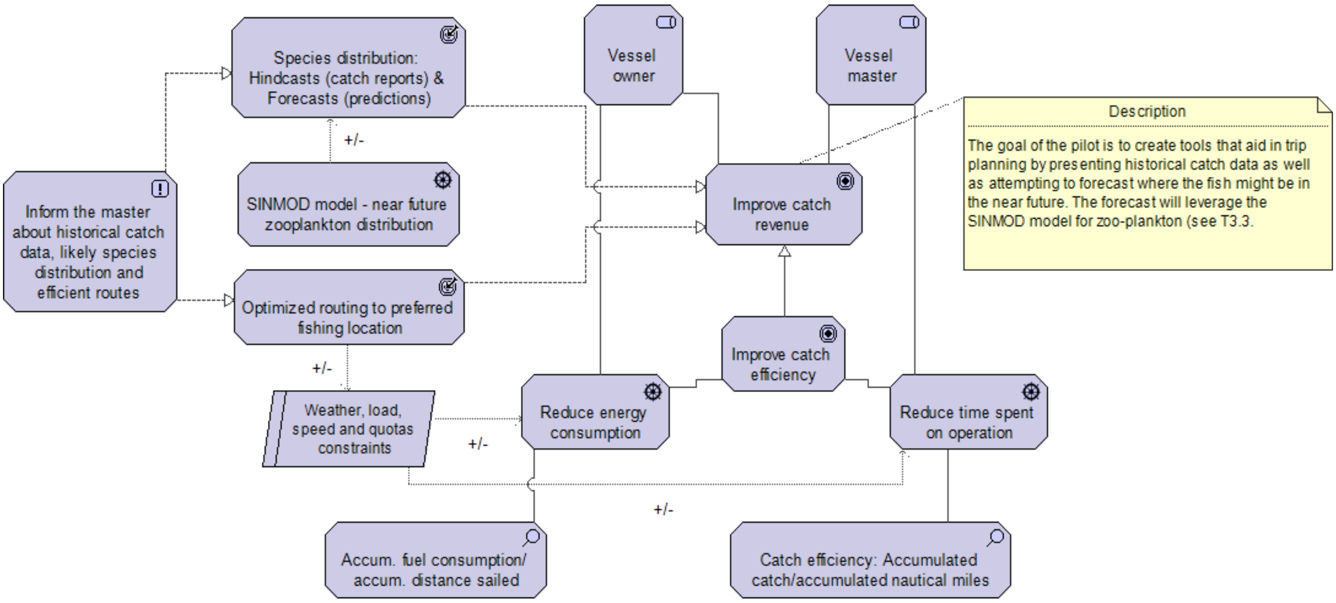


Figure 25: Fishery Pilot B2 Motivation view

6.3.2 DataBio fishery pilot B2 strategy view

The "Fishery B2 Strategy view" diagram in Figure 26 outlines how the decision support is composed: The bottom layer shows the different resources utilized to create the capabilities needed to support the main courses of action, i.e. to collect and analyse data and present the information needed for pelagic fisheries planning.

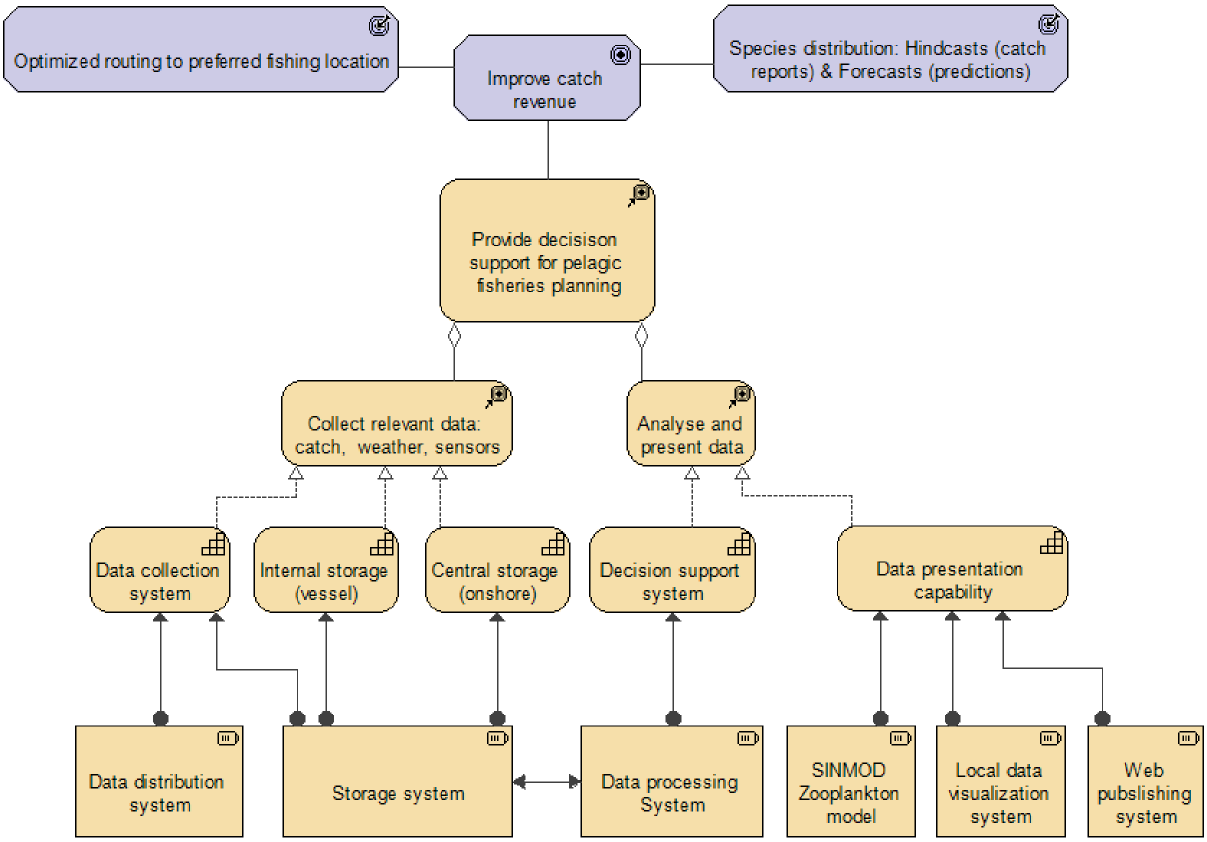


Figure 26: Fishery Pilot B2 Strategy view

6.4 Pilot Evaluation plan

The high-level evaluation plan for the pilot is shown in Figure 27. The evaluation plan is focused on data gathering from satellite, oceanographic models and hydroacoustic and catch data and the subsequent pilot implementation stages.

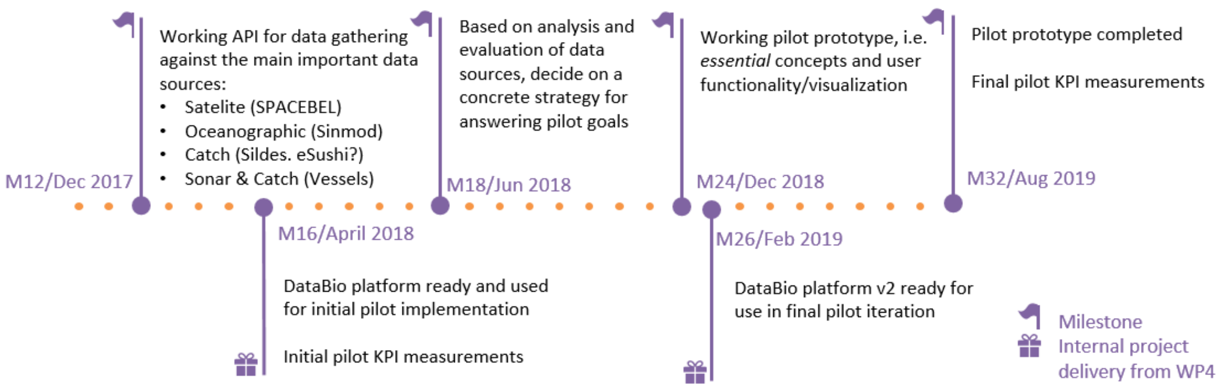


Figure 27: Fishery Pilot B2 Milestone timeline

6.4.1 High level goals and KPIs

The high-level goal is to better understand the geographical movement patterns of the fish, to better plan a trip, and hence save fuel and time.

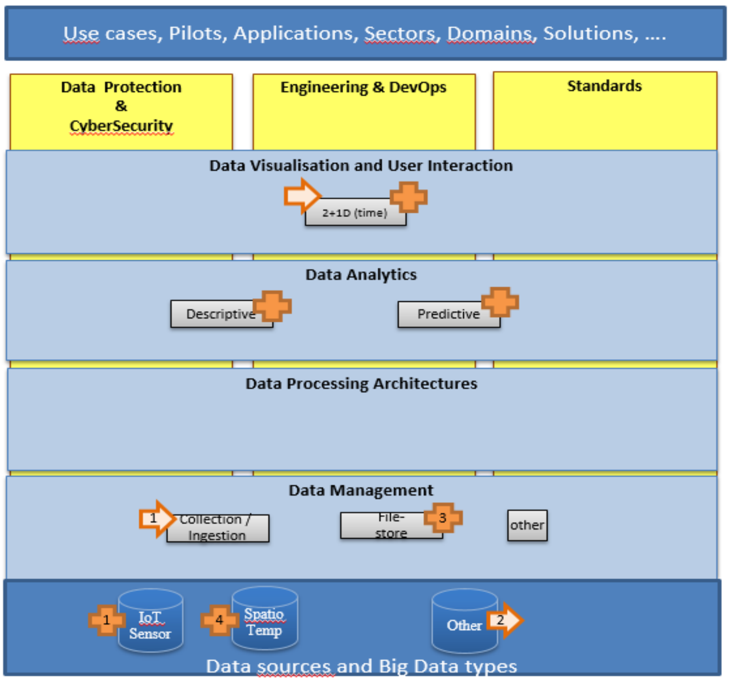
Direct quantitative KPIs are difficult to define. The fuel/time baseline varies significantly based on other factors, and the available sensor data makes it difficult to establish a "ground truth" about fish distribution.

A more indirect KPI can be to map whether the stakeholders experience the tool as "useful".

Statistical analysis of whether the forecast agrees with the sonar data might also be possible.

6.5 Big data assets

The diagram below summarizes Big Data technology components used in this pilot using the extended BDVA reference model. Where applicable, specific partner components that are likely to be used or evaluated by this pilot are listed using the DataBio component ids.



1. Ratatosk (C17.01) for logging data from vessel, Hydroacoustic/sonar, position, etc
2. Pelagic catch reports (Vessels, Sildes)
3. Data collation for different data sources of high variety in sampling frequency (spatial/time) and formats – from DataBio platform
4. Satellite data (Spacebel, etc.)

Figure 28: BDVA reference model for fishery pilot A1

7 Fishery Pilot C1: Pelagic fish stock assessments

7.1 Pilot overview

7.1.1 Introduction

Fish stock assessments are based on scientific surveys conducted at dedicated research vessels, biological sampling of the catch at the ports, and from a randomly selected small number of trips in commercial fishing vessels. These results are in good data quality, but the associated costs prevent the spatial and temporal coverage necessary for reducing the large uncertainty in the population estimates with satisfactory accuracy. In parallel to this activity, fishing vessels are covering interesting ocean areas to a considerable degree, either during fishing activities or during transit to or from fishing areas. Figure 29 below shows the coverage of AIS equipped fishing vessels in the northeast Atlantic during 2011 as observed by satellite-based receivers. The data produced by sensory on board of all these vessels are presently not being utilized for stock assessments. Initiatives have been set in motion to systematically collect large amounts of useful data from these vessels, such as hydroacoustics, oceanographic and meteorological data.

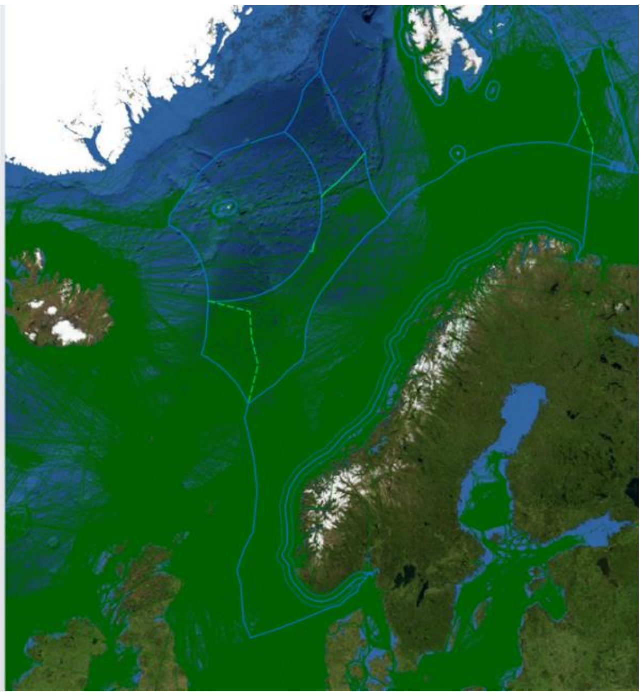


Figure 29: Green regions indicate tracks run by AIS equipped fishing vessels in 2011 as observed from satellite based receivers.

7.1.2 Overview

The objective of this pilot is to demonstrate that the combination of information from many various assets can be used to produce better population dynamics estimates. It is anticipated that a crowdsourced data collection effort from fishing vessels combined with public/private data assets and data analytics can increase both the accuracy and precision of stock assessments.

Method

The general approach for this pilot is to acquire data assets and utilize them together with other sources of insight. Some of the data assets can be supplied by remote sensing, including ocean surface currents and temperatures, and meteorological information. Other information must be measured locally, such as estimates of fish species and densities. These data assets will be provided by fishing vessels equipped with the necessary sensors and communications abilities. Other assets include catch reports, oceanographic simulations, and stock simulations.

Considerable effort in refining the data assets is needed to create value from them. The data will need to pass through several levels of processing and analyses to cultivate information relevant to fish stock assessments. It is expected that value creation is achieved by fusion of curated big data sets with pre-existing knowledge. The knowledge includes simulation models of both physical and biological processes. The hybrid analytics to be used is expected to entail several analytic types, including descriptive and predictive analyses, with activities within model creation, (re)training, and data assimilation.

Relevance to and availability of Big Data and Big Data infrastructure

This pilot will use hybrid analytics and the availability of big data is essential. Big data is needed for training fish stock models, and for assimilating/validating combined biological and oceanographic models. The data needs include satellite data (meteorological and oceanographic), model data (descriptive and predictive), local measurements (buoys and shipborne instruments), and reports on fish catches.

Benefit of pilot

The fishing fleet will benefit from better stock estimates. This may lead to less variable fishing quota from year to year, and it may also lead to more sustainable fish stocks. Reliable stock estimates will provide the governing institutes a better basis for determining fishing quota, both in relation to stock migration and distribution of the stocks outside of conventional research cruises. In total, this pilot will give the means to accomplish better management of the fish resources.

This pilot will depend upon both existing data sources and data sources expected to become available during the course of the project. The Technology Readiness Level varies between the different parts of the pilot, but an estimated average value is TRL 3.

7.2 Pilot case definition

Table 18: Summary of Fishery Pilot C1: Pelagic fish stock assessments

<table>
<tr><td>Use case title</td><td colspan="2">Pelagic fish stock assessments</td></tr>
<tr><td>Vertical (area)</td><td colspan="2">Fishery</td></tr>
<tr><td>Author/company/email</td><td colspan="2">Joakim Haugen / SINTEF Ocean / joakim.haugen@sintef.no</td></tr>
<tr><td>Actors/stakeholders and their roles and responsibilities</td><td colspan="2">Pelagic ship-owning companies: Share collected data assets, comply with a given catch share, which depends on fish stock assessments.
Managing authorities: Provide management decisions for utilization of resources based on recommendations from experts.
Pelagic fish research groups: Experts that produce and make use of available data assets and new knowledge.</td></tr>
<tr><td>Goals</td><td colspan="2">● Create a data pipeline that takes a variety of data assets and makes them ready for hybrid analytics in fish stock assessments – improving data readiness level.
● Investigate whether and how we can leverage big data technologies in making better fish stock assessments.</td></tr>
<tr><td>Use case description</td><td colspan="2">Refer to the pilot case definition section and diagrams in the pilot modelling sections.</td></tr>
<tr><td rowspan="4">Current solutions</td><td>Compute(System)</td><td>No known big data effort exists. The Institute of Marine Research (IMR) executes research cruises and uses a reference fleet.</td></tr>
<tr><td>Storage</td><td>Norwegian Marine Data Centre, but it does not combine all available relevant assets.</td></tr>
<tr><td>Networking</td><td></td></tr>
<tr><td>Software</td><td>LSSS, IMR's in-house analytics.</td></tr>
<tr><td rowspan="2">Big data characteristics</td><td>Data source (distributed/centralized)</td><td>Both distributed and centralized. Distributed data producers are vessels, satellites, buoys, and other sensor networks. Centralized are aggregators and cultivators: service providers, research institutes, directorate of fisheries, the Norwegian Fishermen's' sales organization.</td></tr>
<tr><td>Volume (size)</td><td>Terabytes per year. Proportional to number of contributors (vessels and satellites). Including hydroacoustics, each vessel can produce ~100 gigabytes of raw data per day depending on resolution. Satellites are also highly productive, but only subsets are relevant.</td></tr>
</table>

	Velocity (e.g. real time)	Depends on data producer. Vessel and satellite create and store data in real time, but not continuously for vessels. ~1 Hz for vessel and satellites. Other data producers are burst-like (batches). Service providers: hourly (meteorological), daily (earth observation, oceanographic), catch log (daily), other: sporadic/rarely.
	Variety (multiple datasets, mashup)	Huge. Many different asset originators using various data structures and formats.
	Variability (rate of change)	High. Depends on observed phenomena and sensing device. Physical processes and dynamics can be both fast and slow. There is a hierarchy of processes at different time scales that are of interest to observe, and the output from fish stocks estimates can be orders of magnitude slower than some of the incoming data assets' variability.
Big data science (collection, curation, analysis, action)	**Veracity (Robustness Issues, semantics)**	Important so that analysis output can provide valid representation and prediction of observed processes. This will in turn influence fish quotas. The challenge is that each sensor samples different phenomena with various levels of uncertainty and precision associated with it. Data processing must maintain/improve veracity.
	Visualization	Important for some of the data assets to make them vivid and for understanding the limitations and possibilities of the datasets. Collated views may uncover connections that spawn ideas for further analytics actions. Also, needed for annotation of data by human experts to be used in supervised learning. Visualization is an essential tool in the analytics development.
	Data quality (syntax)	Data providers should have quality assurance mechanisms in place, but additional data cleaning are needed to detect and remove corrupt or inaccurate records. Standard methods such as range checks, outlier detection, principal

		component analysis-based fault detection, and data fusion are all relevant as part of the data management pipeline (curation). Mitigation, such as inclusion of interpolated values for drop-outs, may also be relevant.
	Data types	Highly structured numerical data: time series, spatio-temporal observations. Graph: network topology specific to time and location. (Formats include CSV, NetCDF, GeoJSON, JSON, SQL, proprietary formats and various APIs (specific to data provider).)
	Data analytics	● Descriptive and predictive analytics. ● Hybrid analytics (data-driven + first principles + simulations) ● Supervised learning (classification, deep learning, recurrent neural network (long short-term memory) ● Unsupervised learning (clustering, PCA)
Big data specific challenges (Gaps)	There are gaps in almost all aspects of the data processing pipeline: Sources ● Uncover data providers, establish data access agreements, implementing application interfaces for access points / databases, picking relevant data types. Management ● Improve data readiness level o Understanding the data and what it can offer o Annotation of data by human experts o Anonymization and legal constraints ● Collation and unified access, automated preparation Processing architectures ● How to collate unstructured and highly structured data ● Graph connection of jobs/tasks Analytics ● How to apply hybrid analytics to the problem at hand o Deciding which machine learning methods. o How to use big data with simulation models. Develop estimation models for fish stocks based	

	on existing models and to make use of data assets. o Visualization and user interaction ● Purposeful visualization for o Improving data readiness level o Analytics development Data protection and cybersecurity ● Anonymization procedures ● Security of web portals and communication back-ends ● Access control web portals and databases. Standards ● Unified data format for vessel originating data Engineering & DevOps ● In-house architecture for running the data processing pipeline. Best practices and workflows for big data components development.
Big data specific challenges in bioeconomy	A vessel acts as a mobile sensor platform that produces data at a high velocity. Vessels operate in remote regions with very limited communication infrastructure. This pose a huge challenge in transferring the large data volumes during the short time periods vessels are at shore (using an affordable high bandwidth connection). For the biggest data producers, such as hydroacoustics, an unprocessed data transfer is simple not viable with the current communication infrastructure.
Security and privacy technical considerations	Catch data per vessel is highly sensitive and should be anonymized and subject to access restrictions. User authentication mechanisms for various access levels must be in place. Application of standards and legislation (non-disclosure agreements) should be considered and formalized when applicable.
Highlight issues for generalizing this Use case (e.g. for ref. architecture)	Knowledge sharing for in scientific community to bring solution even further is challenging because of: ● Data format standardization for data assets. ● Proprietary data acquisition platforms and maintenance costs. ● Collation and storage of data from multiple sources can lead to issues regarding access control, long-term management, ownership, access rights. ● Looking beyond the data itself, intellectual property rights for analytics solutions that are essential parts of the processing pipeline.
More information (URLs)	www.databio.eu http://cmr.no/projects/10396/lsss/ <other URLs to be added later if relevant>
Note:	

7.2.1 Stakeholders and user stories

Table 19: Stakeholders and user stories of Fishery Pilot C1: Pelagic fish stock assessments

Stakeholders	User story	Motivation
Pelagic ship-owning companies	As a manager for a pelagic fisheries company, I want to plan the necessary size of the required workforce. I want to reduce the risk of unnecessary expenses by increasing predictability of the future market condition. Indications of next year's fishing quotas is useful.	Optimize revenue
Managing authorities: Sales organization for pelagic fish	As a sales organization for pelagic fish, I want to distribute predictions and estimates on stocks, so that I can help the fishermen in maximizing profit and ensuring predictability of the market.	A good foundation for decision making.
Managing authorities: ICES working group member	As an ICES working group member, I want the best possible basis for decision making when producing scientific advice. A good foundation of information may help both ensuring sustainability of the stocks, but also predictability in estimates from year to year.	A good foundation for decision making.
Pelagic fish research groups: Experts that produce and make use of available data assets and new knowledge.	As a pelagic fish researcher, I need data and means to perform analyses in a systematic manner. This includes making use of available data assets and existing knowledge.	Tools and architecture to improve fish stock assessments.

7.2.2 Motivation and strategy

The motivation for improved stock assessments is primarily better management of pelagic fish stocks for sustainability, but could also facilitate more predictable fishing quotas from year to year. The latter would benefit business management of ship-owning companies. The strategy to impact these motivations are to

- Ensure a crowdsourced data collection initiative from fishing vessels.
- Perform collation of many data assets relevant for enabling big data analytics.
- Create a purposeful data pipeline that improves the data readiness level of assets.
- Explore big data analytics methods for fish stock assessments.

The pilot motivation and strategy is summarized using ArchiMate diagrams in the next section, while goals and key performance indicators (KPIs) are indicated in the succeeding evaluation plan.

7.3 Pilot modelling with ArchiMate

The current chapter presents the Fishery C1 Pilot using the ArchiMate standard with motivation and strategy views.

7.3.1 DataBio fishery pilot C1 motivation view

This chapter describes the "Fishery C1 Motivation view" view defined in the "Fishery C1 Pelagic fish stock assessment modelling with ArchiMate" view point. Figure 30 provides a motivation view for the pilot. This diagram might be best interpreted if read from bottom up.

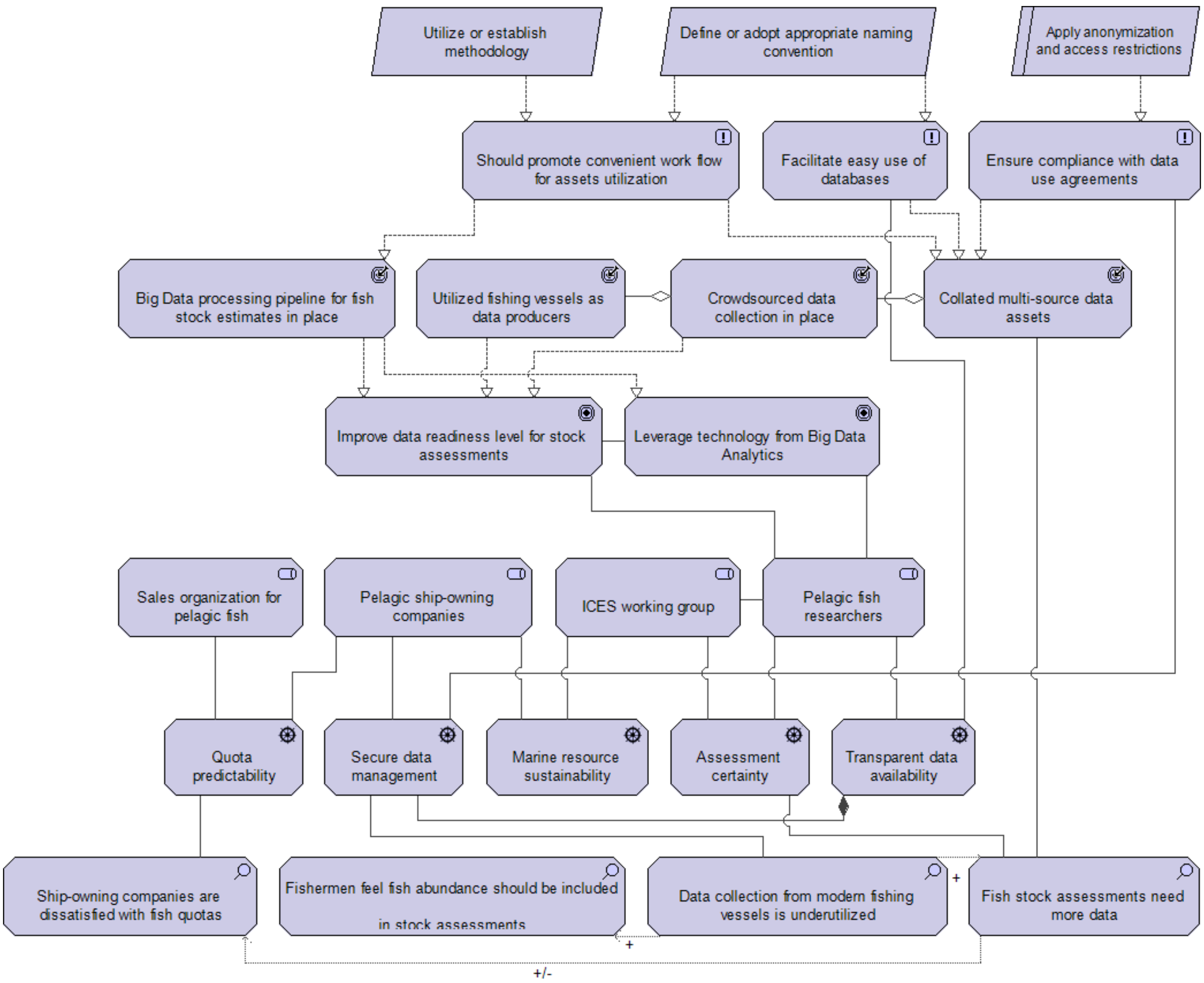


Figure 30: Fishery Pilot C1 Motivation view

7.3.2 DataBio fishery pilot C1 strategy view

This chapter describes the "Fishery C1 Strategy view" view defined in the "Fishery C1 Pelagic fish stock assessment modelling with ArchiMate" view point. Figure 31 illustrates the strategy for the pilot implementation. The resources/assets are located in the leftmost column, with new capabilities and courses of action to the right, respectively. Hence, the diagram is best understood if read left to right.

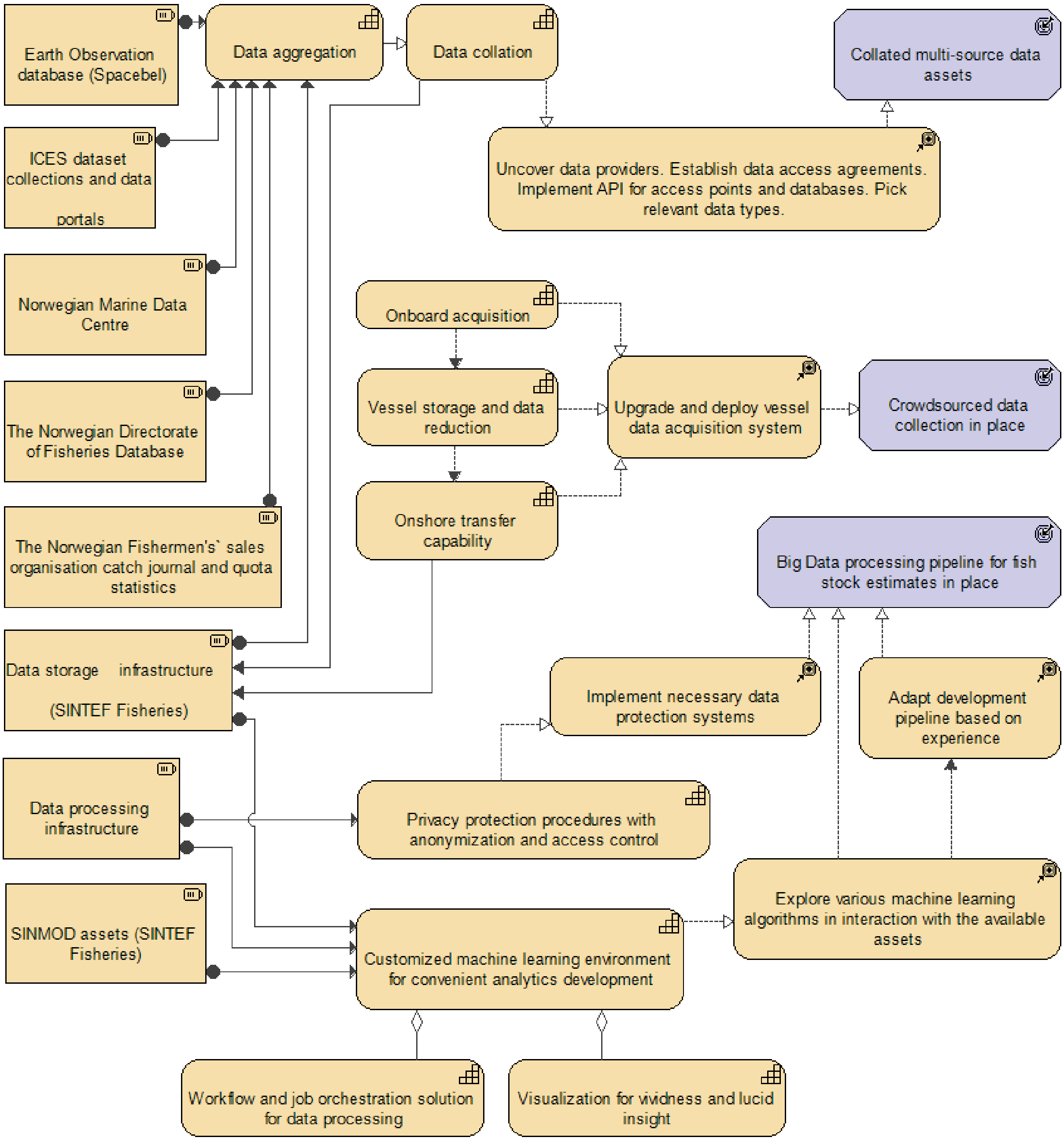


Figure 31: Fishery Pilot C1 Strategy view

7.4 Pilot Evaluation plan

Figure 32 shows a pilot-specific roadmap for the C1 pilot. Several of the milestones in the roadmap are linked to “the course actions” items in the strategy view of this pilot. The activities may be related to various levels of the BDVA reference model, which is indicated in a figure below. During the course of the project, a bottom-up approach is planned – if compared against the BDVA reference model levels. Three evaluations of KPI are planned, initial, intermediate, and final. The final results of the pilot must be summarized in due time for the final project report at month 36. The pilot starts with gathering and collating the market data, as it is not based on previous work, but initiated as part of the DataBio project.

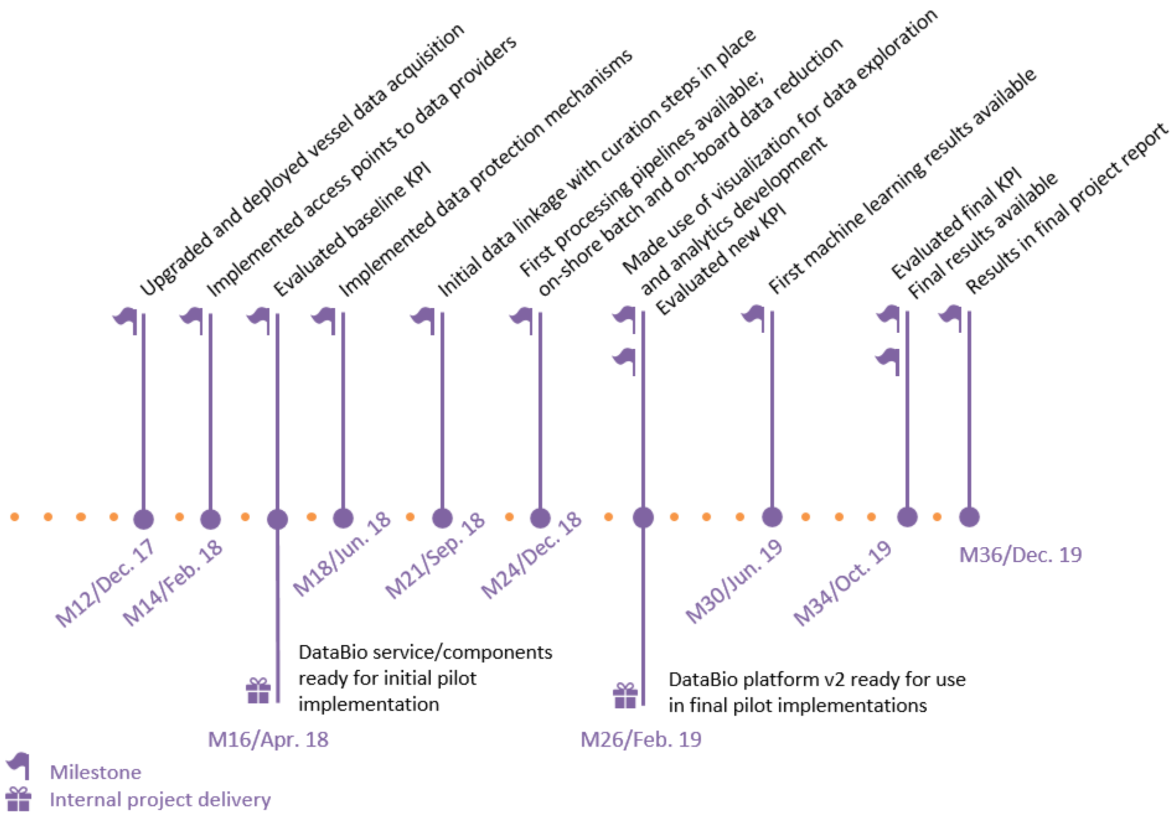


Figure 32: Fishery Pilot C1 Roadmap and internal milestones.

7.4.1 High level goals and KPIs

High level goals are to

- Create a data pipeline that takes a variety of data assets and makes them ready for hybrid analytics in fish stock assessments – improving data readiness level.
- Investigate whether and how we can leverage big data technologies in making better fish stock assessments.

Proposed key performance indicators are

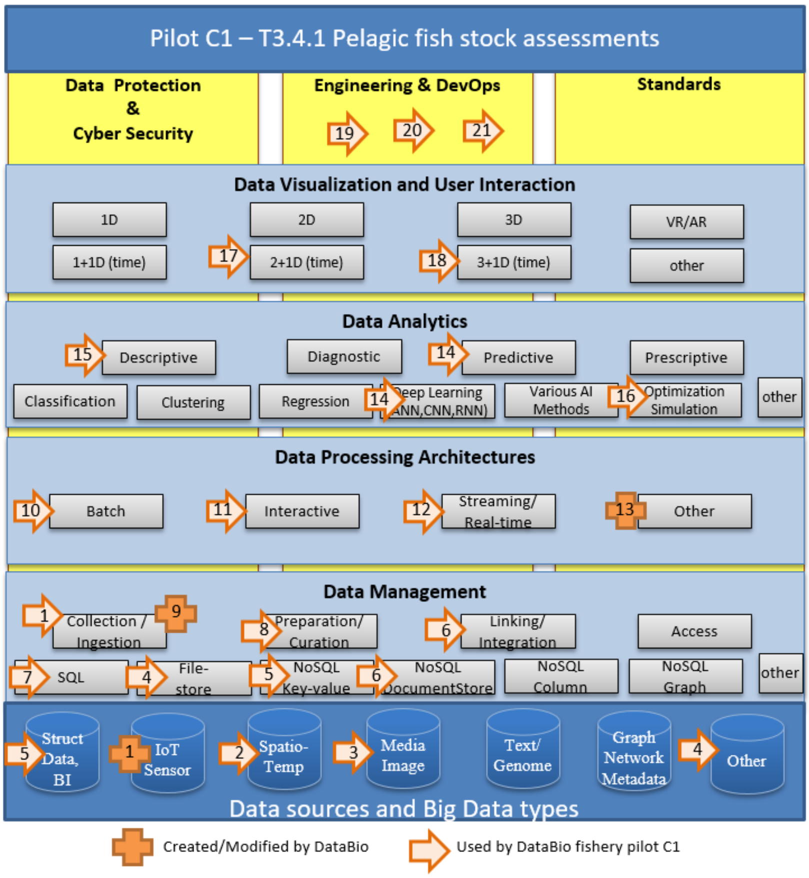
- Data readiness level: A quantification of the availability and the degree of maturity of data assets.
    - Measure of which data assets are being collected and how well they are understood in the context of use for pelagic fish stock assessments. Completeness with respect to management of desired/known sources.
    - Number of vessels that collects useful data, perhaps with secondary

completeness indication for each vessel. Collected duration is also part of the indicator.

- Data pipeline maturity: Each component in the data processing pipeline is quantified by a percentage readiness for use in this pilot. The combined readiness indicates pipeline maturity.
- Data analytics performance: Establish or find in literature a fish stock assessment performance indicator that can compare analytics against other assessment methods.

7.5 Big data assets

The diagram below summarizes Big Data technology components used in this pilot using the extended BDVA reference model. Where applicable, specific partner components that are likely to be used or evaluated by this pilot are listed using the DataBio component ids.



1. Ratatosk (C17.01) for logging data from vessel
2. Hydroacoustics/sonar and weather data
3. Sonar, radar and EO images (C07.01-4)
4. Pelagic auction data and catch reports
5. Databases
    a) ICES data portal
    b) MET.no
6. GeoJSON data store – CouchDB?
7. Map feature server – WMS – GeoServer, GeoRocket or other (C04.01-4)?
8. Stim (C17.02) for data validation, and decimation/interpolation
9. SW component gathering data from 4 and 5
10. Batch processing of historic data- Hadoop?
11. Interactive and iterative development
12. On vessel processing, ref 1,2,3, data filtering for fish observations. Image classification (C29.01, C31.01 or other?)
13. Data collation for different data sources of high variety in sampling frequency (spatial/time) and formats – From DataBio platform
14. ML methods for predictive fish stock estimates (C34.01, C16.02, C34.01 or other)
15. Correlation/Covariance analysis
16. ML optimization of factors affecting fish stocks
17. Plot of data assets and model outputs – R, matlab or octave?
18. Interactive display spatio-temporal stock information- CESIUM) and/or Rasdaman (C04.01, C05.01)?
19. GlusterFS
20. Docker for setup/job/service run
21. DCOS w/MESOS

Figure 33: BDVA reference model for Fishery Pilot C1 (shared with C2).

8 Fishery Pilot C2: Small pelagic market predictions and traceability

8.1 Pilot overview

8.1.1 Introduction

Adapting the fisheries to the future market needs is one of the major challenges for the small pelagic fisheries. One tries to do the fishing when the fisheries are good, the prices are high and the quality is good. As the covariation of these aspects are not simple to predict, there is often room for improvements with regards to both income, quality and energy efficiency.

The consumers in general, and also fish consumers, are to an increasing degree occupied by questions about sustainability, quality and environmental friendliness of their food. There are certification labels for sustainable seafood products, but the availability of certified products is relatively low in Norway. There is a lack of more objective and quantitative measures.

8.1.2 Overview

The goal of this pilot is to provide information for predicting the development of various market segments, so that the fisheries may be targeted against the most beneficial fisheries. This pilot will also act as a basis for providing the consumers with information about the products, so that he/she can take into consideration quantitative aspects such as sustainability, environmental impact, energy consumption per kg of fish, etc. This could be presented either directly or through sustainable seafood certification labels such as those provided by Marine Stewardship Council[1], which today is found on more than 20,000 seafood products. More than half of the Norwegian fisheries are already MSC-certified, but there are relatively few MSC-certified seafood products found in Norway as the neighbouring countries Denmark and Sweden has roughly 5 times more.

Method

The basis for the market predictions will be to combine as many data sources as possible which may be related to the market situation. For each market segment this could be such information as:

- The predicted volume of species in question.
- Financial factors (both international and for single country).
- The state of competing fish stocks and predicted fished volume of these.
- The local weather forecasts.
- The prices of alternative food.

Using historical data, machine learning will be employed to model consumption habits in certain areas, as well as the relation between market development and other factors. These

[1] https://www.msc.org/

models will then be used for providing predictions for how various market segments will develop in the future.

To provide information which can describe the fish products with respect to e.g. quality and environmental footprint, local vessel measurements will be combined with data such as catch reports including quality data and fish stock sustainability level.

Relevance to and availability of Big Data and Big Data infrastructure

Norges Sildesalgslag has a database with information about all pelagic catches that is landed in Norway the last decades which will be accessible for the pilot. Additionally, the associated partner, the Norwegian Seafood Council, which works together with the fisheries and aquaculture industry to develop markets for Norwegian seafood, will provide the pilot with market insight data, statistics, trade information, consumption and consumer insight.

Benefit of pilot

This pilot will focus on the small pelagic fisheries in the North Atlantic Ocean, and the main stakeholders will be Norges Sildesalgslag (Sildes), which is the world's largest marketplace for pelagic fish, and all the market players in the fisheries sector. The main research partner will be SINTEF Ocean. All the players along the value chain from catch to consumer will benefit from this pilot, from the fishing vessels companies, processing companies, markets players and finally the consumer. Current TRL: 3-4.

8.2 Pilot case definition

Table 20: Summary of Fishery Pilot C2: Small pelagic market predictions and traceability

<table>
<tr><td>Use case title</td><td colspan="2">Small pelagic market predictions and traceability</td></tr>
<tr><td>Vertical (area)</td><td colspan="2">Fishery</td></tr>
<tr><td>Author/company/email</td><td colspan="2">Per Gunnar Auran / SINTEF Ocean / per.gunnar.auran@sintef.no</td></tr>
<tr><td>Actors/stakeholders and their roles and responsibilities</td><td colspan="2">Vessel Master (captain): responsible for vessel/quota/catch/quality and cost of vessel fishing activities.
Vessel Owner (Shipping company): Company and fleet responsible
End Consumer, key market driver for consumer behavior, seafood quality, safety, pricing and sustainability labelling are important aspects of this.</td></tr>
<tr><td>Goals</td><td colspan="2">Improve revenue from pelagic fishery by targeting optimal species.
Investigate whether price trends for pelagic species and markets can be predicted by using BigData technologies.
Refer to the evaluation section for specific goals and KPIs.</td></tr>
<tr><td>Use case description</td><td colspan="2">Refer to the pilot case definition section and the diagrams in the ArchiMate modelling section below.</td></tr>
<tr><td>Current solutions</td><td>Compute(System)</td><td>Non-existing system today, fishery planning and catch species based on</td></tr>
</table>

	Storage	Local system + online information systems for catch reports and auctions
	Networking	Manual assessment of Web, AIS, VHF, Facebook pages for vessels
	Software	Multiple individual systems, not integrated processing and display.
Big data characteristics	**Data source (distributed/centralized)**	Combination of both types: Centralized: Market trends by the World bank and Norwegian Seafood Council, pelagic auction data by Sildesalgslaget. Distributed/local: Fish stock observations (hydroacoustic and sonar instruments), quality measurements, vessel operations data (motion and cost of operation)
	Volume (size)	A few terabytes per year when considering operational, motion and EO/weather data, but can increase significantly based on inclusion of hydro-acoustics/sonar data (raw data of 200GB/day when continuous logging).
	Velocity (e.g. real time)	Highly varying, from 1-10Hz for vessel position and orientation to quarterly and yearly for market trends.
	Variety (multiple datasets, mashup)	Great variety as the intention is to include as many relevant data sources as possible.
	Variability (rate of change)	Same as above, rate of change depends very much on data source/type.
Big data science (collection, curation, analysis, action)	**Veracity (Robustness Issues, semantics)**	Need to have a system for fisheries planning that can be used online and
	Visualization	Visualization of correlation and covariation of different data types affecting price. Projection curves for price development for the near future (next weeks/months/quarter/ year)
	Data quality (syntax)	Data validity filtering w.r.t. completeness, general interpolation and decimation support needed for comparisons, alternatively could the data be projected into a selected timeframe, but relevance and validity for the timeframe must be quantified and filtered on minimum

		confidence levels. Data fusion is central here.
	Data types	DDS data from vessel, weather forecast and EO, JSON Rest api for World Bank, pelagic auction and seafood council data type TBD, but JSON preferred. Time series data, both structured and unstructured with varying frequency, image data (sonar, radar and EO/weather).
	Data analytics	Predictive analytics for the development of data-driven projection models for price trends in pelagic markets.
Big data specific challenges (Gaps)	There is a need for closing the gap between fisheries planning and maximizing profit within sustainable quotas and catch by utilization of big data methods, e.g. combining varying sources of relevant information and into an integrated view showing impact and price projection results.	
Big data specific challenges in bioeconomy	A system intended for use at open sea faces severe network connectivity challenges, both for on-shore data logging and the use of web based services. The vessel system must be able to operate for days without Internet access based on the latest information that was available, and on-shore systems as data loggers and catch reporting must resume operation when network connections are re-established.	
Security and privacy technical considerations	Catch data per vessel highly sensitive, should be anonymized and treated by fleet / species group. Access to specific/detailed results by login for partners providing data?	
Highlight issues for generalizing this Use case (e.g. for ref. architecture)	Predictive analytics using machine learning for price forecasting based on big data is a general problem that extends to all the three sectors in DataBio and beyond.	
More information (URLs)	www.databio.eu <other URLs to be added later if relevant>	
Note:		

8.2.1 Stakeholders and user stories

The end users are pelagic fisheries companies and consumers of pelagic Seafood products. Sildesalgslaget will present market predictions (pelagic price predictions in various markets) on their website based on catch statistics, quality, consumer data and market demand (economic growth & fluctuations). This information will be used by the pelagic fleet to optimize revenue by choosing which species to catch when within the sustainable quotas. Stakeholders and user stories are summarized in the table below.

Table 21: Stakeholders and user stories of Fishery Pilot C2: Small pelagic market predictions and traceability

Stakeholders	User story	Motivation
Vessel Master and Vessel Owner (Captain and Shipping company)	As a captain on a pelagic vessel, I would like to take into account market predictions for price development when planning the fishing activity on the different species so that the revenue can be maximized better throughout the catch season. A price difference of a few kroner has a huge impact as the catch size often can reach several hundred metric tons.	Optimize revenue from small pelagic fishery
End Consumer	As a consumer of seafood I would like to know product quality, price, data and location of the catch, vessel of origin and environmental impact of the fishery, represented in an easily understandable way, for example as energy (fuel) consumption per kilo fish.	Raise awareness of sustainable fisheries and seafood price-quality relations

8.2.2 Motivation and strategy

The main motivation and goals for this pilot is

- Primarily to provide a service where market information and results from predictive analytic results from market data is presented for pelagic fishery stakeholders to optimize catch revenue for the near future. There is a large potential for using machine learning in small pelagic fisheries to optimize catch revenue through estimating price prediction models by using all the relevant data available which will be the approach pursued in this pilot.
- Secondarily to raise the awareness of the consumer on sustainable fisheries and seafood products by making relevant catch and stock information publicly available, and informing about sustainable labels such as MSC.

The pilot motivation and strategy is summarized using ArchiMate diagrams in the next section, while goals and KPIs are addressed in the successive evaluation plan.

8.3 Pilot modelling with ArchiMate

This chapter presents C2 pilot described using diagram according to the ArchiMate 3.0 standard, and includes the additional business view introduced in the A2 pilot together with the motivational and strategy views. Together this diagram gives a good visual overview of why the pilot is important and how the primary objectives are addressed.

8.3.1 DataBio fishery pilot C2 motivation view

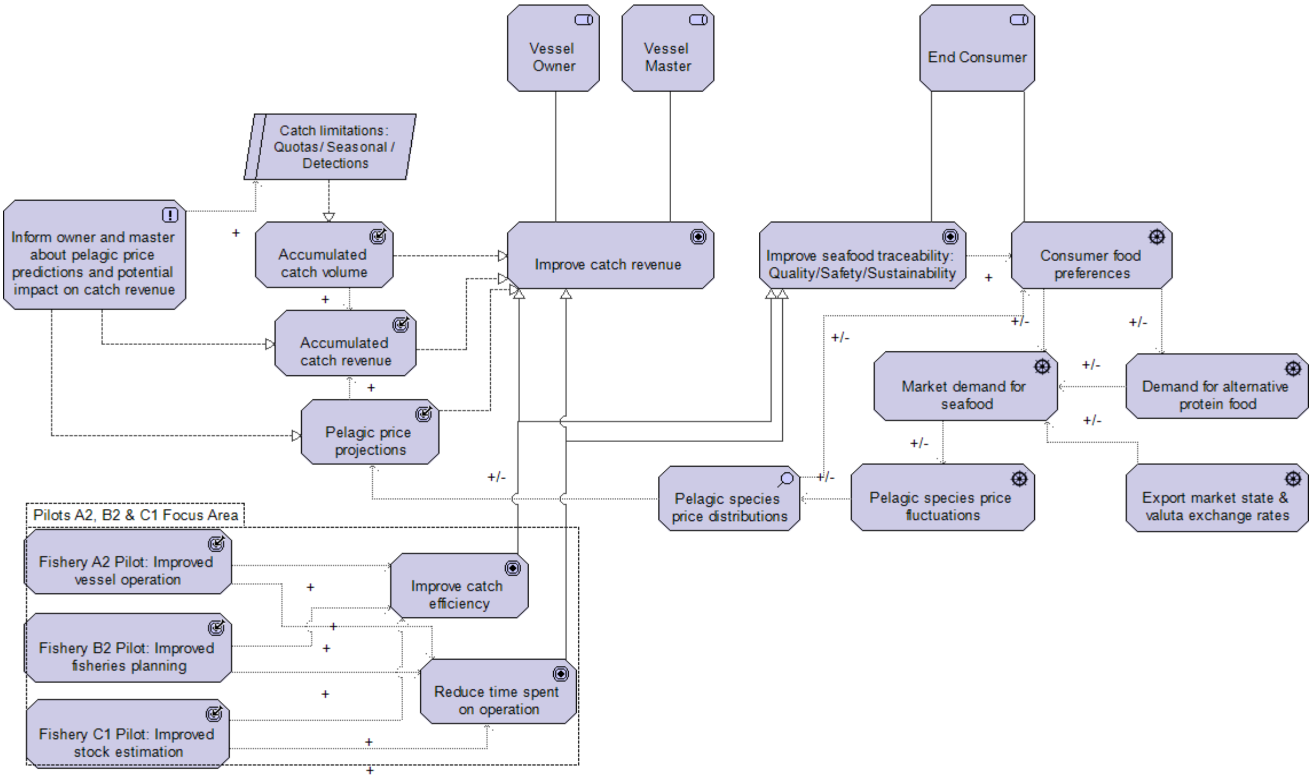


Figure 34: Fishery Pilot C2 Motivation view

Figure 34 shows the motivation view for the C2 pilot. The key motivation element is to improve catch revenue through price predictions for pelagic species (upper left) and a secondary objective is to improve seafood traceability and indirectly increase market demand for seafood products by providing more detailed information about them to end consumers (right side). In the lower left side, it is indicated how the A2, B2 and C1 pilot also provide positive influence catch revenue through better catch efficiency and reduced time on operation, but as indicated this pilot will rely on the related pilots for that and concentrate on the market aspects of pelagic catch.

8.3.2 DataBio fishery pilot C2 strategy view

This chapter describes the "Fishery C2 Strategy view" view defined in the "Fishery C2 Small pelagic market predictions and traceability modelling with ArchiMate" view point.

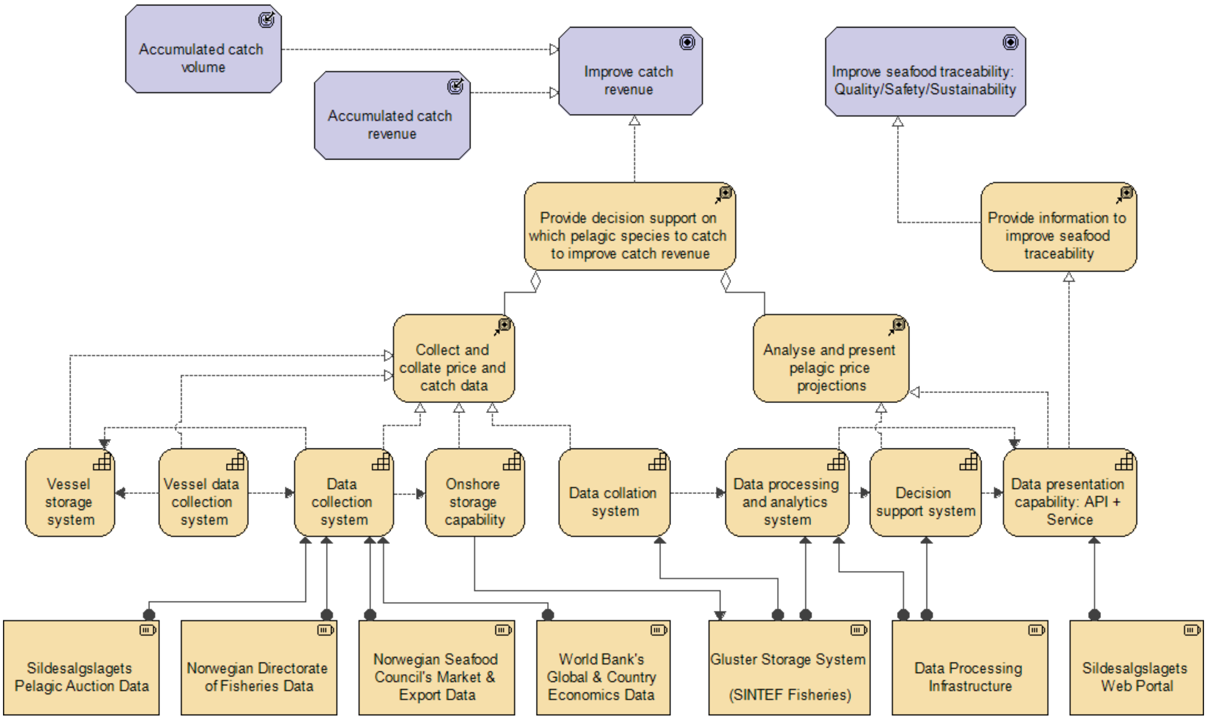


Figure 35: Fishery Pilot C2 Strategy view

Figure 35 shows the strategy view with the key elements: In the bottom row we find the main resources, i.e., the market data sources (left to middle) and storage, processing and visualization systems (right). These support the capabilities in the next layer, primarily data collection, analytics and visualization, which again provide the desired course of actions.

8.3.3 DataBio fishery pilot C2 business view

This chapter describes the "Fishery C2 Business Process view" view defined in the "Fishery C2 Small pelagic market predictions and traceability modelling with ArchiMate" view point. This view has been included to illustrate how the market data and models interact with the relevant pelagic fishery stakeholders and processes.

The Business Process view in Figure 36 is, in essence, quite similar to the strategy view, but here the main focus is on the processes in the main block at the centre of the diagram. At this stage the refinement has some increased detail, note that that the market estimates has been split into two processes, as decision support is needed based on the current/ near future (1-2W) catch process, but also how this current decision and historical quota fulfilment affect longer term target species recommendations.

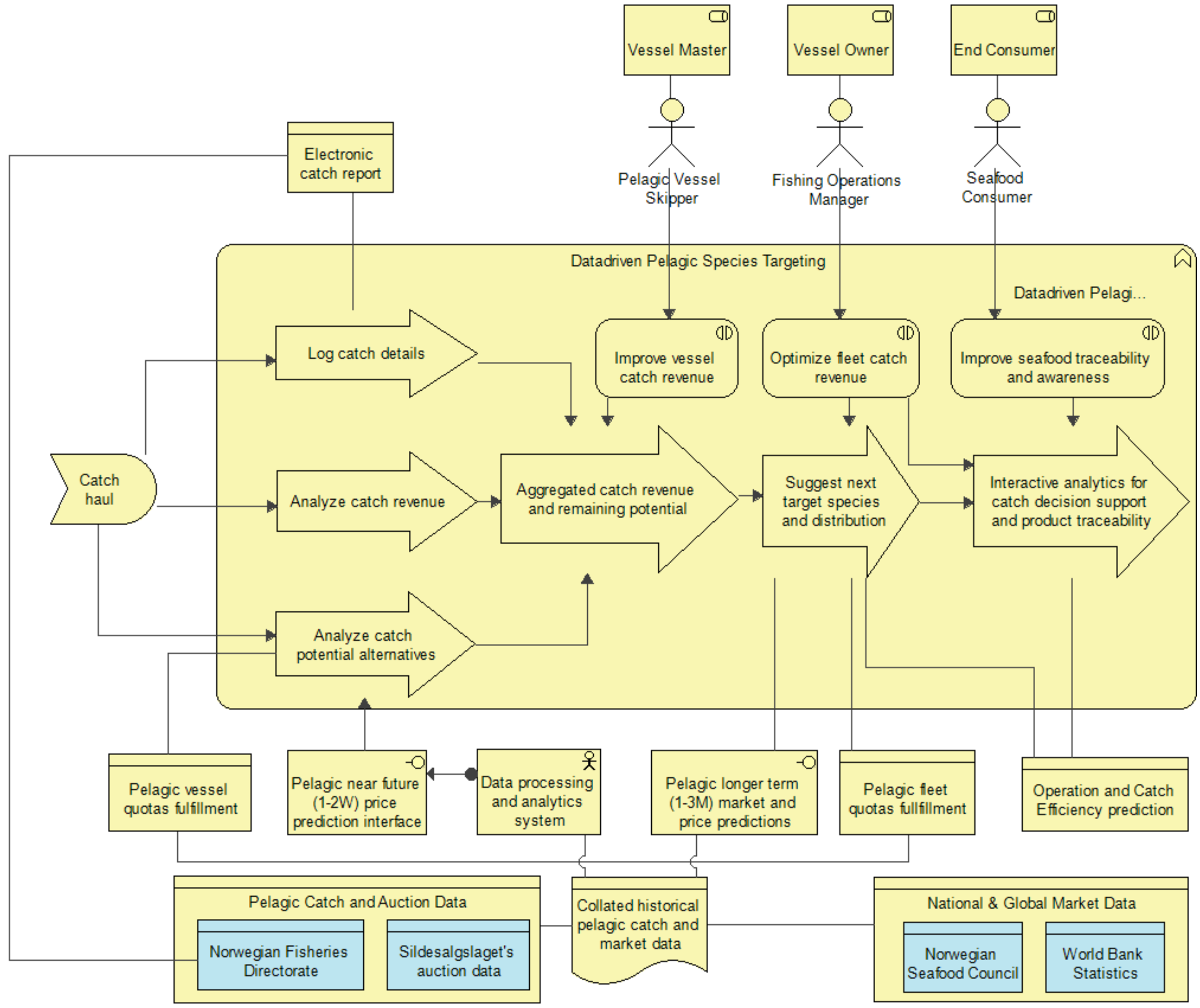


Figure 36: Fishery Pilot C2 Business process view

8.4 Pilot Evaluation plan

Figure 37 shows a pilot specific roadmap for the C2 pilot based on the general one in the introduction. There will be two iterative implementation runs adapted to the DataBio platform cycles and measurements of the pilot specific KPIs at the beginning and finalization of each implementation. The final results of the pilot must be summarized in due time for the final project report at month 36. The pilot starts with gathering and collating the market data, as it is not based on previous work, but initiated as part of the DataBio project.

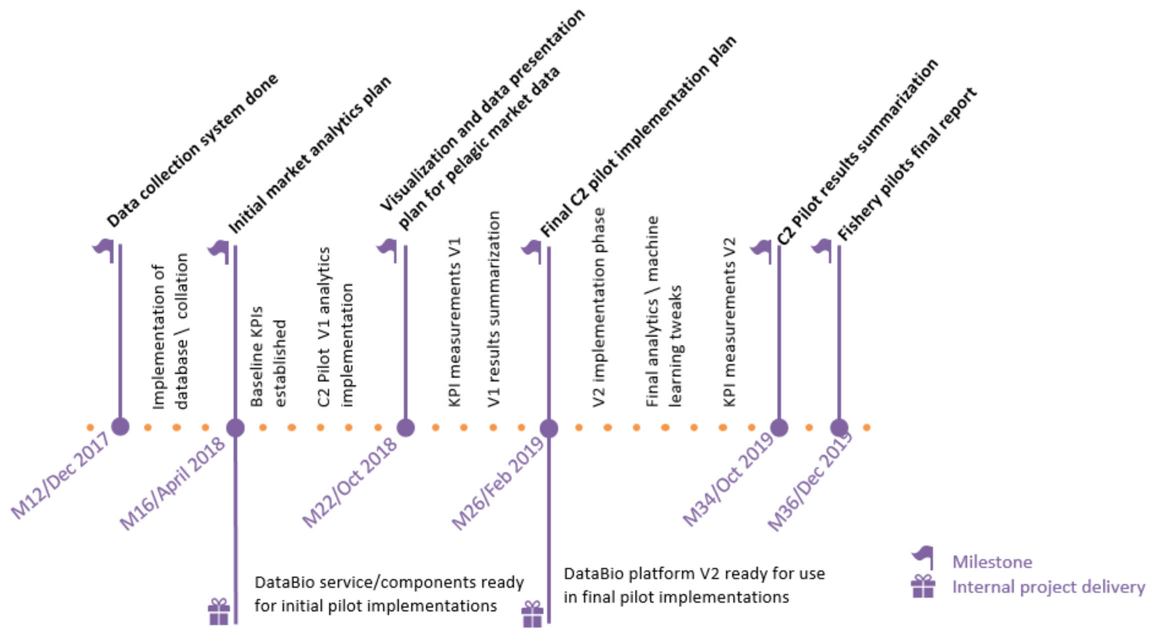


Figure 37: Roadmap for the fishery pilot C2

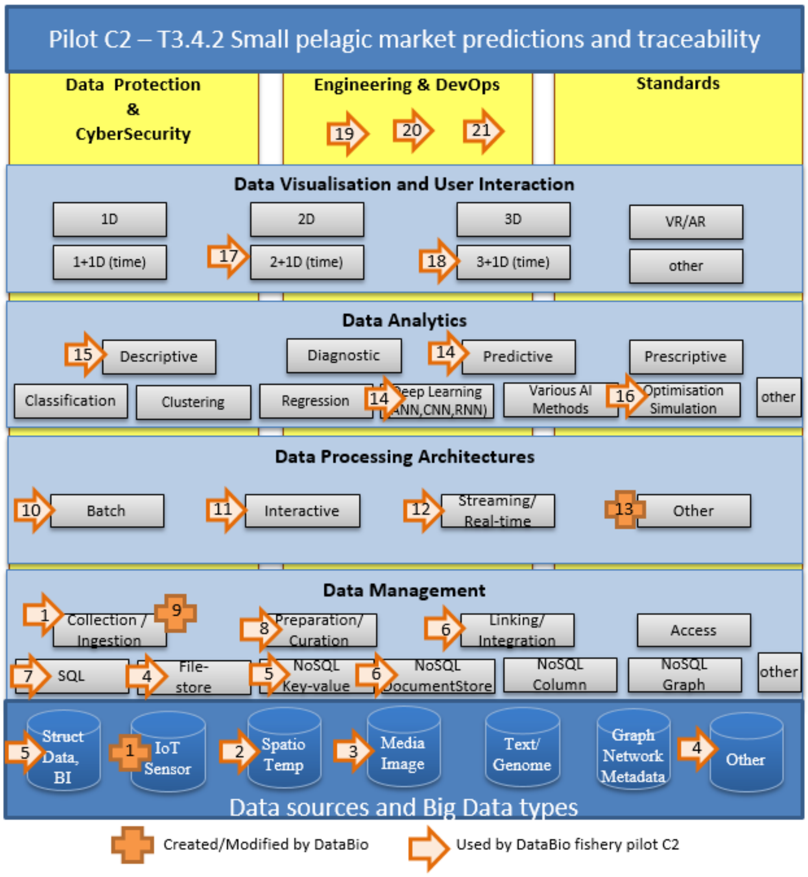
8.4.1 High level goals and KPIs

Three relevant KPIs that has been identified so far:

- Model correctness: **price prediction success ratio** having an acceptable error rate when tested on historic data that it was not trained on.
- **Revenue potential with alternative catch strategy vs. what happened**: Quantify increased revenue potential on historical data, .e.g. what was the accumulated value, vs what could have been achieved on catching alternative species.
- **System usage**: Number of users of the market information service in pelagic fisheries, and number of users, including growth rate, visiting the website with information about. Note that this takes time to establish, and for this project, a baseline can be measured first, then followed by monitoring usage after system deployment.

8.5 Big data assets

The diagram below summarizes Big Data technology components used in this pilot using the extended BDVA reference model. Where applicable, specific partner components that are likely to be used or evaluated by this pilot are listed using the DataBio component ids.



1. Ratatosk (C17.01) for logging data from vessel
2. Hydroacoustic/sonar and weather data
3. Sonar, radar and EO images (C07.01-4)
4. Pelagic auction data and catch reports
5. World bank market data (JSON)
6. GeoJSON data store – CouchDB?
7. Map feature server – WMS – GeoServer, GeoRocket or other (C04.01-4)?
8. Stim (C17.02) for data validation, and decimation/interpolation
9. SW component gathering data from 4 and 5
10. Batch processing of historic data- Hadoop?
11. Interactive and iterative development
12. On vessel processing, ref 1,2,3, data filtering for fish observations. Image classification (C29.01, C31.01 or other?)
13. Data collation for different data sources of high variety in sampling frequency (spatial/time) and formats – From DataBio platform
14. ML methods for predictive price models (C34.01, C16.02, C34.01 or other)
15. Correllation/Covariance analysis
16. ML optimization of factors affecting fishery revenue
17. Plot of price models/time – R, matlab or octave?
18. Interactive display spatio-temporal catch and market information- CESIUM) and/or Rasdaman (C04.01, C05.01)?
19. Gluster
20. Docker for setup/job/service run
21. DCOS w/MESOS

Figure 38: BDVA reference model for fishery pilot C2 (shared with C.1)

9 Conclusions

D3.1 provides the background for the fisheries pilots, and presents initial pilot descriptions and models. These descriptions are the basis for further development and implementation of the pilots within WP3 but are also important for communicating pilot plans and needs towards WP4 and WP5 for software components and EO data, respectively. This document is also used for matchmaking activities among pilots and technology providers, for modelling and analysis of synergies among pilots and for the definition of the DataBio reference architecture. The work in WP3 continues with selection and testing of components and implementing the first version of the pilot applications. After the initial testing of components from technology partners the pilot experience will be summarized and used as input to the DataBio Reference Big Data Architecture.

10 References

Reference	Name of document (include authors, version, date etc. where applicable)
[REF-01]	DataBio website. www.databio.eu. Retrieved 2017-07-07.
[REF-02]	ArchiMate® 3.0 Specification: http://pubs.opengroup.org/architecture/ArchiMate3-doc/toc.html
[REF-03]	Global Fishing Watch: http://globalfishingwatch.org/
[REF-04]	The State of World Fisheries and Aquaculture, Sofia 2016, http://www.fao.org/fishery/sofia/en
[REF-05]	FAO 2009 How to feed the World: http://www.fao.org/fileadmin/templates/wsfs/docs/expert_paper/How_to_Feed_the_World_in_2050.pdf