D5.1. FishData system specification

2019-10-29

SMARTFISH H2020

Innovation for sustainable fisheries

Project number:	7553521
Project duration:	1 Jan 2018 – 31 Dec 2021
Project coordinator:	Bent Herrmann, SINTEF Ocean
Web site:	www.smartfishh2020.eu

Deliverable ID:	Due month:	Preparation date:
D5.1	M6	2019-10-29

Title:

FishData system specification

Lead beneficiary:
SINTEF Ocean

Prepared by:
Lars T. Kyllingstad, SINTEF Ocean
Karl-Johan Reite, SINTEF Ocean
Per Gunnar Auran, SINTEF Ocean

Approved by:
Hanne Digre, Research Director, SINTEF Ocean

Abstract

The goal of SMARTFISH H2020 is to develop, test and promote a suite of high-tech systems for the EU fishing sector, thereby optimising resource efficiency, improving automatic data collection for fish stock assessment, provide evidence of compliance with fishery regulations and reduce the ecological impact of the industry.

This deliverable is an initial specification of FishData—an experimental data acquisition and analysis infrastructure to be established by and for the SMARTFISH H2020 project. We present a number of use cases that the system should support, and from those use cases we extract a set of functional requirements. Based on the requirements, we propose a high-level architecture consisting of two subsystems: An *onboard infrastructure* for data acquisition and exchange on fishing vessels, and an *onshore infrastructure* for data centres that process and store data from multiple sources, including fishing vessels.

Dissemination level

PU	Public	X
CO	Confidential, only for members of the consortium (including the Commission services)	
CI	Classified information as referred to in Commission Decision 2001/844/EC)	

Deliverable type

R	Document, report	X
DEM	Demonstrator, pilot, prototype	
DEC	Web sites, patent fillings, videos, etc.	
OTHER	Software, technical diagram, etc.	

Authorship information

Editor	Lars T. Kyllingstad, SINTEF Ocean
Contributing partners	AZTI: Guzmán Diez, Iñaki Quincoces, Gorka Gabiña DTU Aqua: Clara Ulrich, Kristian Schreiber Plet-Hansen Cefas: Thomas Catchpole, David Maxwell, Stuart Hetherington

Version history

Version number	Date	Description of changes
1	2018-06-29	Initial version. Delivered to the European Commision as D5.1.
2	2019-10-29	Updated according to feedback received after Reporting Period 1 (M18) review. Section 4: • Described in more detail the three primary use cases. • Other use cases have been kept for future reference but were moved to a separate subsection. Section 5: • Removed all previous contents of this section as they have been incorporated in or obsoleted by the changes to sections 4 and 6. • Changed title to "Data management" and added new text that discusses data access policies and security. Section 6: • Added *Implementation* subsections in section 6.1 (Onboard infrastructure) that describe in more detail how the onboard infrastructure is implemented and the specific data collected. • Added more subsections in section 6.2 (Onshore infrastructure), most with their own *Implementation* sub-subsection, to provide more details on the most important infrastructure components. • Annotated the FMC diagrams with communication protocols and data formats (to the extent that they are determined). Appendix B: • Completely replaced. Now contains a listing of data structures. Delivered to the European Commision as a revised version of D5.1.

SMARTFISH H2020 consortium

SMARTFISH H2020 (773521) is an Innovation Action within Horizon 2020, the European Union's framework programme for research and innovation, Call H2020-SFS-2017-1, Topic SFS-22-2017.

The following table contains information about the consortium members.

	SINTEF Ocean (SO) Postboks 4762 Torgard N-7465 Trondheim NORWAY www.sintef.no	Project Coordinator: Bent Herrmann bent.herrmann@sintef.no Project Manager: Rachel Tiller Rachel.tiller@sintef.no
	AZTI-Tecnalia (AZTI) 48395 Sukarrieta SPAIN www.azti.es	Guzmán Diez gdiez@azti.es +34 667 174 392
	DTU Aqua (DTU) Danmarks Tekniske Universitet Willemoesvej 2 9850 Hirtshals DENMARK www.dtu.dk	Ludvig Ahm Krag lak@aqua.dtu.dk +45 21 31 64 57
	Marine Scotland Marine Laboratory 375 Victoria Road Aberdeen, AB11 9DB UNITED KINGDOM http://www.gov.scot/Topics/marine/science/	Neil Campbell Neil.Campbell@gov.scot +44 (0)131 244 2902
	Melbu Systems AS Sminesgata 3, 8445 Melbu NORWAY http://www.melbusystems.no/	Einar Roger Pettersen einar@melbusystems.no +47 900 21 883
	Cukurova University Fisheries Faculty 01330, Saricam, Adana TURKEY http://www.cu.edu.tr/Eng/Default.aspx	Gökhan Gökce gokhan.gokce@ymail.com +90 322 330 6084 ext: 2961
	CEFAS Pakefield Road Lowestoft Suffolk NR33 0HT UNITED KINGDOM https://www.cefas.co.uk/	Thomas Catchpole thomas.catchpole@cefas.co.uk +44 (0) 7768041107
	Mersin University Fisheries Faculty Yenisehir Campus 33169 Mersin TURKEY http://www.mersin.edu.tr/	Huseyin Ozbilgin ozbilginh@gmail.com +90 3243412818

SINTEF	**Stiftelsen SINTEF** N-7465 Trondheim NORWAY www.sintef.com	Jens T. Thielmann Jens.T.Thielemann@sintef.no +47 930 59 299 Karl Henrik Haugholt Karl.H.Haugholt@sintef.no Petter Risholm Petter.Risholm@sintef.no +47 95 89 53 29
NTNU Norwegian University of Science and Technology	**NTNU** Institutt for Marin Teknikk, NTNU N-7491 Trondheim NORWAY https://www.ntnu.no/imt	Martin Ludvigsen martin.ludvigsen@ntnu.no +47 91897272
UEA University of East Anglia	**University of East Anglia** Norwich Research Park Norwich Norfolk, NR4 7TJ UNITED KINGDOM https://www.uea.ac.uk/	Michal Mackiewicz M.Mackiewicz@uea.ac.uk +44 (0)1603 59 1206
MARPORT	**Marport France SAS** 2 allée Copernic Espace Eurêka Parc Technologique de Soye 56270 Ploemeur, France	Didier Caute dcaute@marport.com +33 622 012 347
ZUNIBAL	**Zunibal** Idorsolo, 1 48160 Derio Bizkaia SPAIN http://zunibal.com/en/	Iratxe Arraibi iratxe.arraibi@zunibal.com (+34) 944 977 010 (7:30-15:30) (+34) 608 556 661
SafetyNet TECHNOLOGIES	SafetyNet Technologies LTD 29 Shand Street London, SE12ES UNITED KINGDOM http://sntech.co.uk/	Dan Watson dan@sntech.co.uk +44 779 294 9727
NERGÅRD	Nergård Havfiske Langnesveien 18 9408 HARSTAD Norway http://nergardhavfiske.no/	Tor Pedersen tor@yras.no +47 994 99 899
IFL	**Interfish** Wallsend Ind Est Cattedown Plymouth, Devon PL4 0RW, England	Andrew Pillar andrew@interfish.co.uk +44 (0) 1752 267261
LARRASMENDI, BI. S.L. BUQUES DE PESCA	Larrasmendi	Aitor Larranaga erlaxi@larrasmendibi.e.telefonica.net +34 946830617

Table of contents

Executive summary ... 11
1. SMARTFISH H2020 motivation and background ... 12
1.1. Role of this deliverable ... 12
1.2. Contributors ... 13
2. Introduction ... 14
2.1. Purpose ... 14
2.2. Scope and definition ... 15
2.3. Non-goals ... 15
2.4. This document ... 15
3. Background and motivation ... 17
3.1. Data collection and sharing ... 17
3.2. Vessel monitoring systems for fisheries management ... 17
3.2.1. SIF ... 18
3.2.2. BBV ... 19
3.2.3. AZTI Fisheries Catch Monitoring System ... 19
3.2.4. FOCUS ... 20
3.3. Fishery planning and vessel routing ... 20
3.4. Examples of machine learning in fisheries ... 21
3.5. Fishery information services and open data sets ... 21
3.6. Potential of FishData ... 23
3.7. Related work by the SMARTFISH H2020 partners ... 24
4. Use cases ... 27
4.1. Use case 1: Historical catch data analysis ... 27
4.1.1. Description ... 27
4.1.2. Method ... 27
4.2. Use case 2: Catch composition and bycatch monitoring ... 28
4.2.1. Description ... 28
4.2.2. Method ... 28
4.3. Use case 3: Improvements and updates to CatchScanner using data from commercial fisheries 29
4.3.1. Description ... 29
4.3.2. Method ... 30
4.4. Other use cases ... 31
4.4.1. Onboard fishing activity detection ... 31
4.4.2. Vessel information portal ... 31
4.4.3. Onshore fishing activity detection ... 31
4.4.4. Biomass monitoring ... 31
4.4.5. Variations in catch composition and selectivity ... 32

5. Data management ... 33
5.1. Vessel data ... 33
5.2. Publicly available data ... 33
5.3. Output from analyses of private data ... 33
5.4. Output from analyses of public data ... 33
6. Architecture ... 34
6.1. Onboard infrastructure ... 34
6.1.1. Middleware ... 35
6.1.2. Data sources ... 36
6.1.3. Logging ... 39
6.1.4. Ship–shore data transfer ... 40
6.2. Onshore infrastructure ... 41
6.2.1. Access control ... 43
6.2.2. Orchestration ... 43
6.2.3. Inbound and outbound data ... 44
6.2.4. Web portal ... 44
6.2.5. Data storage ... 44
References ... 47
A. Data collected by the SIF and BBV systems ... 51
A.1 SIF ... 51
A.1.1 Desired improvements in the current data system ... 52
A.2 BBV – Black Box Video ... 53
A.2.1 Data currently collected by the Black Box systems ... 54
A.3 Onshore grading machine data ... 54
B. Data structures ... 57
C. Fundamental Modeling Concepts ... 59

List of figures

Figure 1: Conceptual structure of SMARTFISH H2020 12

Figure 2: Conceptual figure of the difference between landings data available at haul level in the electronic logbook and the sea-packing data available in the SIF database. [4] 18

Figure 3: Example of image for video audit with grid overlay and measuring line integrated in the Black Box Analyzer set at measuring a monkfish (Lophius piscatorius) of 65.7 cm. 19

Figure 4: Touch screen of the AZTI Fisheries Catch Monitoring System. 20

Figure 5: Norwegian Sea fishing activity according to Global Fishing Watch (Jul/Aug 2017). 22

Figure 6: The FiskInfo service; example showing fishing activity with nets (blue), lines(red) and purse seiners (purple) as well as restricted (black polygons) and closed (filled polygons) fishing areas. From the fiskinfo.no web site. 22

Figure 7: FMC block diagram of overall architecture. 34

Figure 8: FMC block diagram of onboard infrastructure, with some generic examples of data sources. 35

Figure 9: FMC block diagram of the onshore infrastructure. Some connectors are left unannotated because the boxes they connect represent generic service *categories*, and their interfaces/protocols are chosen on a case-by-case basis. 42

Figure 10: Black Box Video System instruments, data flow, integration and transmission to the inland centre. Courtesy of Anchor Lab K/S. 53

Figure 11: FMC block diagram reference sheet. [43] 59

List of tables

Table 1: Examples of open data sets relevant for fisheries. 23
Table 2. Ongoing related projects. 24
Table 3. Existing technologies from SMARTFISH H2020 partners. 25
Table 4: A non-exhaustive overview of relevant data collected from fishing vessels. See Appendix B for IDL definitions of the types listed under *DDS topic type*. 38
Table 5: Properties of data volumes in the storage cluster. 45
Table 6: Overall haul and trip information in the SIF system for sea-packed landings. 51
Table 7: Haul and trip level information with sea-packing landings data. 51
Table 8: Example of packing data format from a grading machine. 55
Table 9: Example of weighing data format from a grading machine. 55

Executive summary

FishData will be an ICT infrastructure that enables systematic collection and analysis of data from fishing vessels, and which facilitates use of the results for fisheries monitoring and management purposes, for stock assessment and other marine research activities, and for decision support in the fisheries industry itself. In this document we propose a number of specific analysis use cases that the infrastructure should support:

1. Onboard fishing activity detection
2. Vessel information portal
3. Onshore fishing activity detection
4. Catch composition
5. Biomass monitoring
6. Variations in catch composition and selectivities

We then define a set of requirements based on these use cases. In brief, the infrastructure must support a great variety of onboard and onshore data sources; it must handle both high-volume and high-throughput data; it must have good security measures; and it must support a broad array of data exploration and analysis tools.

Finally, we propose a high-level architecture for the onboard and onshore parts of the infrastructure.

1. SMARTFISH H2020 motivation and background

With an increasing pressure on marine resource extraction mounting with resultant calls for sustainability in the sector, SMARTFISH H2020 will develop, test and promote a suite of high-tech systems that will optimise resource efficiency, improve automatic data collection, provide evidence of compliance with fishery regulations and reduce the ecological impact of the sector on the marine environment (Figure 1).

SMARTFISH H2020 will exploit and further develop existing technological innovations in machine vision, camera technology, data processing, machine learning, artificial intelligence, big data analysis, smartphones/tablets, LED technology, acoustics and ROV technology. The developments will assist commercial fishers throughout Europe in making informed decisions during pre-catch, catch, and post-catch phases of the harvesting process.

SMARTFISH H2020 will also provide new data for stock assessment from commercial fishing and improve the quality and quantity of data that comes from traditional assessment surveys. This provides the potential for more accurate assessment of fish stocks and allows for the assessment of stocks that are currently data-poor and therefore difficult to manage. In addition, the project will access automatically collected catch data from the fisheries which will also allow for management regulations to gain higher compliance rates.

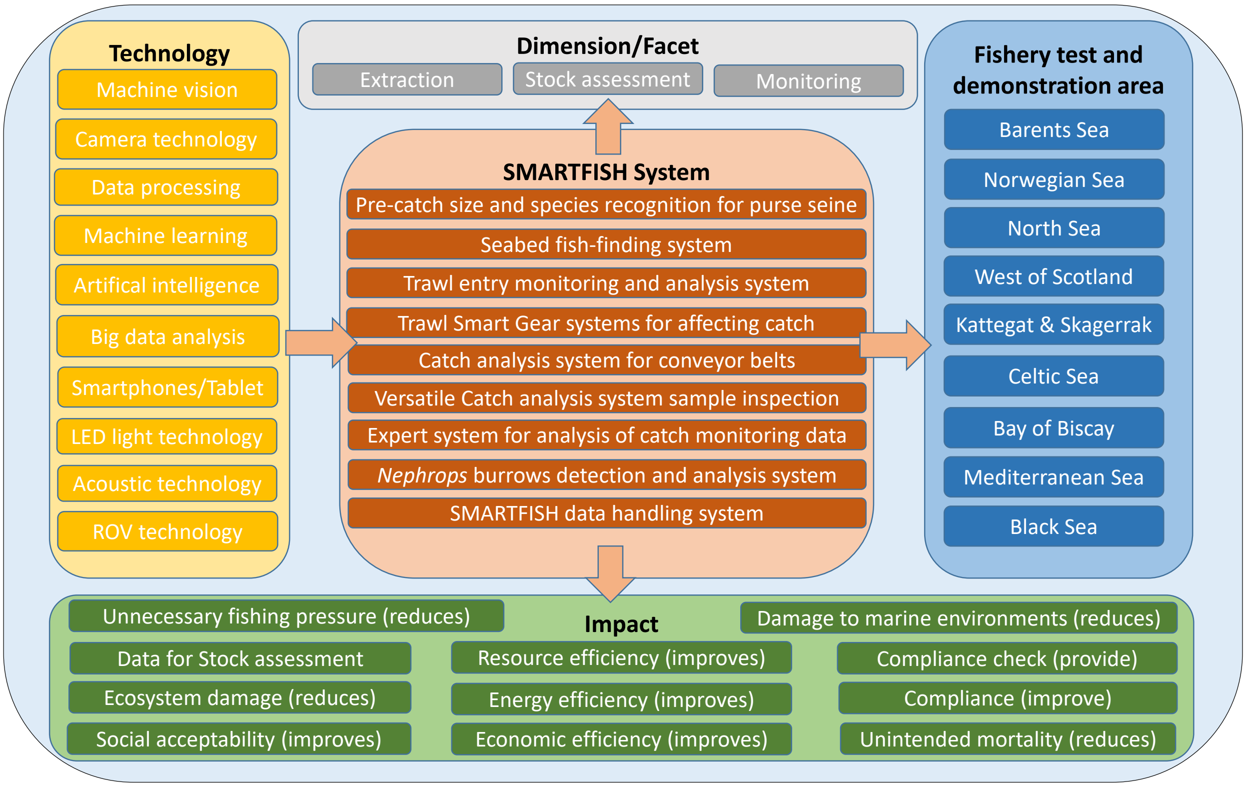

Figure 1: Conceptual structure of SMARTFISH H2020

1.1. Role of this deliverable

This document is the end result of SMARTFISH H2020 Work Package 5, Task 5.1, *Specification of overall architecture*. The Project Description defines the task as follows:

> "The architecture of the data handling system will be specified at a level of detail which enables subsequent work on Tasks 5.2–5.5 to happen mostly in parallel. This includes identifying relevant data sources, defining use cases, identifying performance and capacity requirements (storage, processing, communications, etc.), specifying security requirements and selecting the communication protocols and data formats that will be used for interoperability between different subsystems."

1.2. Contributors

The following partners have contributed to this deliverable:

- SINTEF Ocean
- AZTI
- DTU Aqua
- Cefas

2. Introduction

This document describes *FishData*—an experimental data acquisition and analysis infrastructure to be established by and for the SMARTFISH H2020 project. This document is the first of three deliverables in Work Package 5 (WP5), *Data handling systems*, and its role is to lay the ground for the work to be done in the WP.

The WP consists of the following tasks:

- Task 5.1, *Specification of overall architecture*, which concerns the writing of this document.
- Task 5.2, *Development of onboard infrastructure*, which is about developing a hardware and software infrastructure for data acquisition and transfer of data from fishing vessels.
- Task 5.3, *Development of onshore infrastructure*, which is about developing a hardware and software infrastructure for data management and analysis on shore.
- Task 5.4, *Development of analyses and presentation for fisheries decision support*, in which we will demonstrate the use of big data techniques and technologies in a case study focused on the needs of stakeholders in the fishing industry.
- Task 5.5, *Development of analyses and presentation for stock assessment and fisheries management*, in which we will do the same for a case study that focuses on stakeholders in marine research and resource management.

The results of the above tasks are formalised as three *deliverables*:

- D5.1, *FishData system specification* (this document), which is the result of Task 5.1.
- D5.2, *FishData infrastructure*, which is the result of Tasks 5.2 and 5.3, due in December 2019.
- D5.3, *FishData analysis*, which is the result of Tasks 5.4 and 5.5, due in December 2020.

2.1. Purpose

The purpose of carrying out SMARTFISH H2020 WP5 and developing the FishData technology is threefold:

Firstly, we will explore the use of *big data* techniques and technologies in fisheries. Opinions differ on the proper definition of "big data", but one that seems apt for our purposes is:

> "The ability of society to harness information in novel ways to produce useful insights or goods and services of significant value" and "…things one can do at a large scale that cannot be done at a smaller one, to extract new insights or create new forms of value." [1]

Big data in general is a field of research which has seen tremendous progress in recent decades. It is already widely used in the IT industry, and is currently finding applications across a multitude of other industries as well as in research and government. The current state of big data in fisheries, and some possible future developments, are discussed further in Section 3.

Secondly, we will explore the use of *new kinds of data* made available through recent technological advances, in particular the other "SMARTFISH H2020 technologies" and the enabling technologies upon which they are built, such as advanced 3D cameras, machine vision, hydroacoustics and more. An important element is to make essential information available in *real time*, or as close to it as possible, since certain data are only useful for certain applications for a short period of time. One example is the use of fish observations from other vessels for short-term planning of fishing operations. We will exploit the ever-increasing availability of mobile and satellite communications and computer processing power onboard fishing vessels of all sizes to achieve this.

Finally, we aim to facilitate sharing of data between the different stakeholder groups involved in fisheries. We believe that all stakeholders will benefit by sharing certain data – between fishing vessels, between companies, and between industry, research and management – and through this project we will explore and demonstrate these benefits.

2.2. Scope and definition

FishData will be based on the SMARTFISH H2020 project participants' existing systems and infrastructure. By combining these systems and augmenting them with new data sources and channels for data exchange, the project will open up opportunities to perform analyses on data of greater scale and complexity than before. The existing systems include:

- SINTEF Ocean's data infrastructure, consisting of the *Ratatosk* system for onboard data acquisition and *SINTEF Marine Data centre* (SMD) for large-scale data storage and analysis.
- AZTI's *Fisheries Catch Monitoring System*, developed for small-scale fisheries.
- Marport's *SCALA*, a real-time onboard commandant interface for Marport sensors and instruments.

FishData will have two aspects: specification and implementation. The *FishData Specification*, of which the present document is the first iteration, is a set of conventions and recommendations for enabling efficient, secure and traceable exchange of data between widely different systems such as fishing vessels and onshore data centres. The specification includes aspects such as communication protocols, file formats and metadata structure. Section **Error! Reference source not found.** deals with the requirements on such systems, and Section 6 contains the actual conventions.

Based on the specification, we will develop a proof-of-concept implementation on top of the existing systems listed above that demonstrates the usefulness of the technology for a set of relevant case studies. The case studies will be selected from a larger set of use cases discussed in Section 3. The goal is to develop the technology to a minimum technology readiness level (TRL) of 6, defined as "technology demonstrated in relevant environment (industrially relevant environment in the case of key enabling technologies)" [2].

2.3. Non-goals

To keep the work focused and limited to a reasonable scope, it is necessary to establish at this early stage what FishData is *not*. Most importantly, we do not aim to create a formal standard, nor do we seek to supplant existing official channels and standards for catch reporting. Rather, FishData will serve as a demonstration of how such reporting could benefit from new methods and technologies for data management and analysis.

Furthermore, we do not plan to develop a commercial service or a general-purpose implementation (i.e., independent of the project participants' existing infrastructure), as this would require a greater degree of technology maturity than this project can be expected to deliver. The TRL achieved in this project should not be expected to exceed 7, "system prototype demonstration in operational environment" [2].

Finally, FishData is intended to be an *extension* of the project participants' existing research infrastructure, and not a complete replacement.

2.4. This document

The present document is intended to serve multiple purposes. It is:

- An introduction to the field of big data in fisheries and a motivation for our work (Section 3).
- A non-exhaustive list of use cases that the FishData infrastructure should support (Section 4).
- A specification of the requirements on, and thus the scope of, the technology (Section **Error! Reference source not found.**).
- A definition of the FishData conventions and recommendations, to serve as an implementation guide (Section 6).

Deliverable D5.1 is the first iteration of this document, which will remain a living document throughout the project period. It will be updated as necessary during development, and at the end of the project, it will serve as a documentation of the work done.

3. Background and motivation

Modern fishing vessels are quite sophisticated, with numerous sensors for finding fish, navigating and communicating with the outside world. The temporal and spatial variability of fish stocks have led the fishermen to have an inherent need to share, restrict and seek out information from each other. Several separate groups benefit from data collection and sharing between fishing vessels: the fishermen, the managing companies, research institutes, and government bodies.

In this section we first give a short introduction to the field of big data in fisheries data. (For a more comprehensive review, see [3].) Then, in subsection 3.6, we attempt to place FishData in this context.

3.1. Data collection and sharing

The current method to obtain information about the fishing activities—where the fishing is good, which species are caught, and which vessels are active—is through utilization of communication technologies and past catch history. Sales data are routinely accessed to obtain information about which vessels deliver what quantities of different species where. VHF radio, cell phones and internet are used to contact specific vessels and companies to get an overview of how the conditions are and to obtain information for fishing trip planning. AIS (Automatic Identification System) tracking portals are used to get an overview of the regions where vessels operate. This process is manual and the access to information is limited by availability of industry contacts and the willingness to share.

Fishing vessels are operated by businesses where the shipowner controls the vessel and resource base, and the catch is landed from vessels per arrangement (auction) with buyers or by habit and location. Each fishing company report their catch diaries to the regulatory bodies and catch information is accumulated on each vessel (or company) while sales and deliveries of fish are collected into publicly available statistics. The businesses record their own data on experience of past fisheries while landing statistics are available online from various sales organizations.

3.2. Vessel monitoring systems for fisheries management

Vessel monitoring systems (VMS) are systems used in commercial fishing to allow environmental and fisheries regulatory organizations to track and monitor the activities of fishing vessels both in a country's territorial waters and the Exclusive Economic Zone outside each country's coast. VMSes are used to improve the management and sustainability of the marine environment by ensuring proper fishing practices and help prevent illegal fishing. VMS systems implementations, requirements and protocols vary between countries. EU, including Norway through EEC, requires VMS and Electronic Report Systems (ERS) aboard all fishing vessels longer than 12 meters (since 2012).

ICES, the International Council for the Exploration of the Sea, is the advisory body that determines the status of fish populations and recommends sustainable quotas for the next year through their annual cycle of scientific and stock assessment meetings of scientists and stakeholders from Europe and North America. The ICES advice helps to set quotas for both EU (including EEC) and ICES member states and carries a heavy weight for settling the quotas at the international, bilateral and national level.

The governing bodies, e.g. the EC and national EU and EEC member states, require fishermen and landing sites by law to report catch data back to them for monitoring purposes. The catch volume is first reported to the proper sales association for the particular species caught and then auctioned to determine the landing site. Fishing companies are increasingly using ERS systems integrated with each vessel in their fleet to support efficient quota management instead of the traditional logbooks. These report systems are designed according to national regulations, and examples include the eCatch system by Dualog in Norway, and the eLog system used in Denmark.

3.2.1. SIF

SIF (“sporbarhed i fiskeindustrien”) is an electronic traceability system for Danish fisheries with automated data collection that provides full tracking from vessel to consumer. It was initiated in 2011 to comply with the EU1224 – article 58 regulation and has nearly full coverage of the seafood supply chain in Denmark. The SIF system provides information on sea-packed landings at the haul level including information on species and size composition by weight, together with detailed information on date, time and position of the haul. The size classes applied are those defined by the EU Council as well as additional size classes used by the Danish fish auctions. These size classes can be linked directly and converted to a weight range. Weight for each size class of each sea-packed species are recorded automatically into the SIF system by the sea-packing equipment using a dynamic scale during the handling operations where the fish are gutted and bled. Haul spatial and temporal information is linked to the SIF system from the electronic logbook (eLog). All sea-packing vessels are part of the Danish large-scale fleet. The SIF system also contains the landing data from the sales slips; however, as sales slips become available when vessels land, the resolution of this is on a per trip basis and the current data collection in Denmark already include the sales slips information.

In essence, the sea-packing data in the SIF system resemble the eLog with the addition of size class rather than only weight of the full landing of a given species, although weights in the eLog are reported as live weight, whereas weights reported in SIF are for gutted fish (Figure 2).

Access to the SIF database is currently through the SIF website whereas the administrator grant access to the individual vessel information. As all vessels own their own data, permission for access has to be provided. To extract the data, a web scraper is used. SIF data has been extracted and investigated for 13 vessels. Investigations of accurate SIF data show that sea-packing accuracy in terms of recorded landings vary between vessels and species, whereas spatial and temporal haul information is generally accurate [4]. Of these 13 vessels, only 6 have an accurate and reliable sea-packing dataset that can be used for scientific purposes.

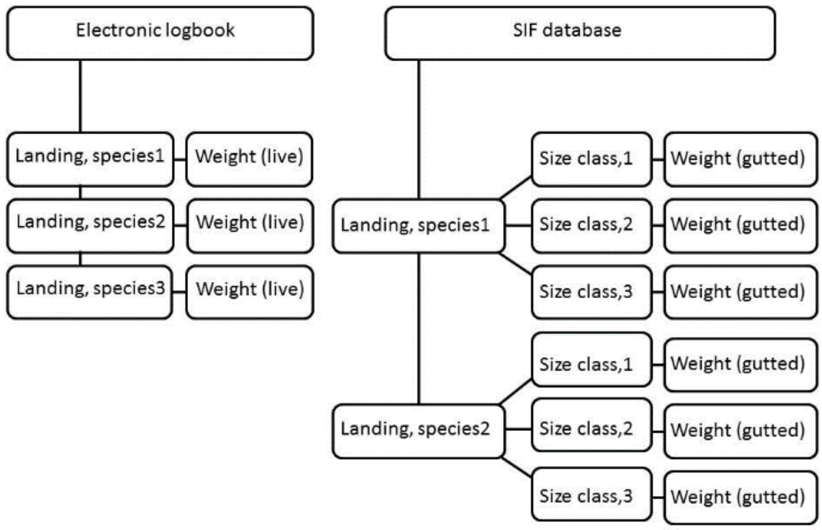

Figure 2: Conceptual figure of the difference between landings data available at haul level in the electronic logbook and the sea-packing data available in the SIF database. [4]

More details on SIF are found in appendix A, section A.1.

3.2.2. BBV

The Black Box Video (BBV) system developed by the Danish AnchorLab company has been used as part of the Danish Fully Documented Fisheries trials [5] and has proven to be a reliable system for Remote Electronic Monitoring. A pure GPS sensor version of the system is mandatory for all vessels fishing for common mussels (Mytilus edulis) in Denmark. The system has been used to collect length and weight data on discards of cod, hake, whiting, haddock and saithe with a length measurement recorded at less than 1 cm [4]. Figure 3 shows an example of measuring the length of a monkfish. The Black Box Video System has been used for several electronic monitoring trials in Denmark, mainly focusing on trawlers in the North Sea and Skagerrak.

BBV is described further in appendix A, section A.2.

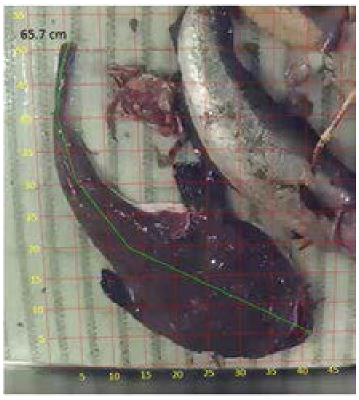

Figure 3: Example of image for video audit with grid overlay and measuring line integrated in the Black Box Analyzer set at measuring a monkfish (Lophius piscatorius) of 65.7 cm.

3.2.3. AZTI Fisheries Catch Monitoring System

AZTI, which is one of the partners in the SMARTFISH H2020 project, has developed a catch monitoring system for small scale fisheries that includes:

- User-friendly interface with touch-mode interaction
- Gear identification (bottom otter trawl and purse seine)
- GPS integration (date, time, vessel position, course and vessel speed monitoring at configurable sampling)
- Near-real-time data transferring unit (GSM) or satellite communication
- Generates 2 files: (1) catch monitoring per fishing event; (2) one day vessel track (configurable depending on fishery)
- Catch monitoring (quick and easy way): fish species selection “by touch-screen” the corresponding fish picture and catch-mass (kg)
- Likewise, “by-catch” monitoring by the same way a catch monitoring.
- Logging and sending data every catch or fishing event
- Vessel track monitoring (sampling data configurable) logging and sending data once per day (configurable)

The touch screen interface of AZTI's system is shown in Figure 4.

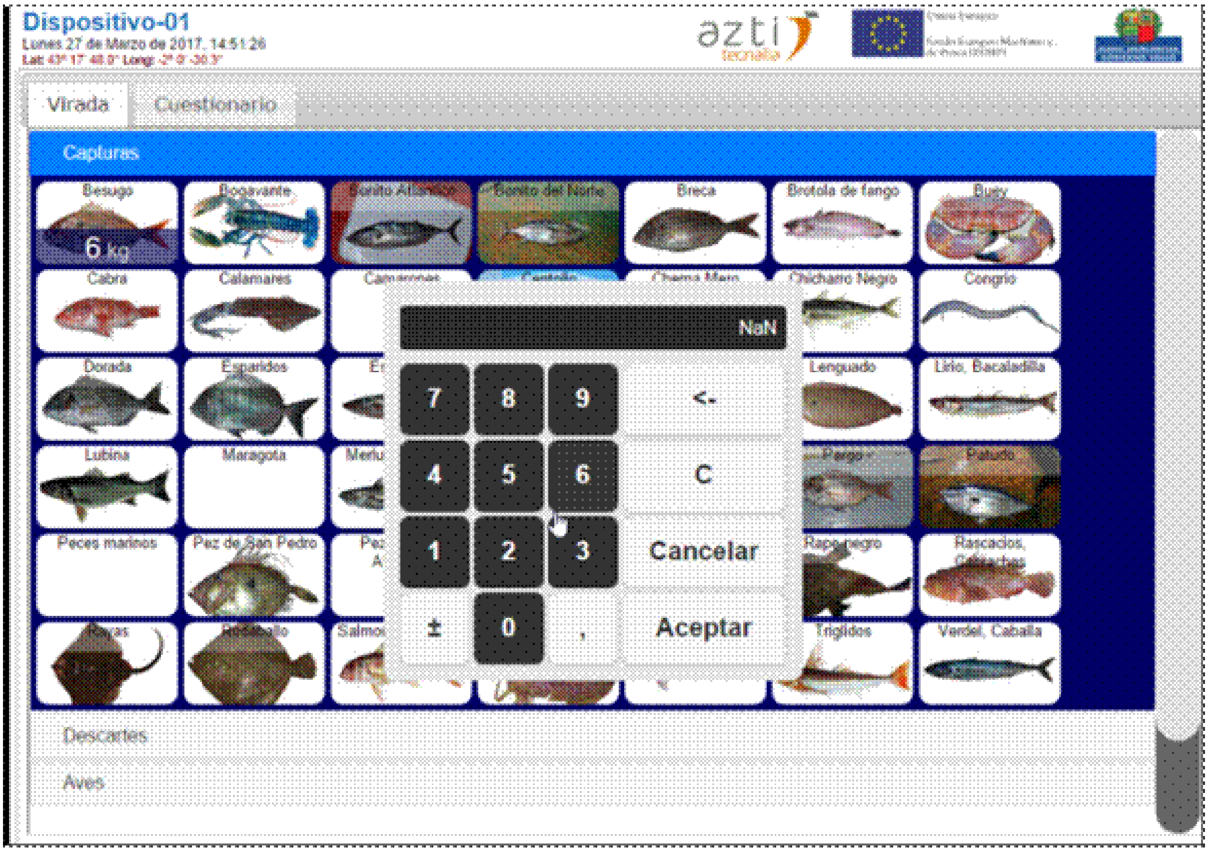

Figure 4: Touch screen of the AZTI Fisheries Catch Monitoring System.

3.2.4. FOCUS

FOCUS (Fisheries Open source CommUnity Software) is an open-source community aiming to offer a free suite of tools to support fisheries management organisations to contribute to sustainable fisheries (www.focus.fish). The project has signed a SDG partnership with the UN and is also supported by the European Commission. A key challenge is the data integration of very diverse data sets, and FOCUS support the UN/CEFACT FLUX (Fisheries Language for Universal eXchange) standards for information exchange to overcome the barrier with diverse national reporting standards for catch monitoring and fisheries management.

3.3. Fishery planning and vessel routing

A lot of time and resources are spent on searching for fish compared to the active fishing operation itself, hence, fisheries planning and vessel routing is a natural starting point for fisheries optimization.

The decision about which route a vessel should take is made by expert fishermen in a subjective way based on their own experience, technological devices (sonar, meteorological forecasts) and increasing communication with local scientists (e.g. presence/absence forecasts from habitat models). Apart from the initial planning based on best areas to fish in the past and current meteorological forecasts, the existence of fish or not in each potential fishing ground attempted (spatial correlation) as well as unforeseen events (bad weather, instrument failures, etc.) need to be considered.

This has been approached in the past using interactive optimization [6, 7], which has been also used in maritime transportation planning [8]. The definition of a fitness function that accurately represents the real world is a critical task that often require an iterative process of eliciting a fitness function from the expert [7]. This explains why so far there are only some attempts or proof of concept aiming at optimising certain elements of fishing activities [9, 10]. However, they focus on considering a single activity or destination (e.g. routing to the fishing area) or a single decision driver (e.g. meteorological conditions [9].

3.4. Examples of machine learning in fisheries

Machine learning based approaches using satellite data have been successful in the past for example in forecasting species recruitment and identifying new potential predictors [11, 12]. Time-series analysis forecasting anchovy recruitment demonstrated that climate patterns could explain a seasonal behaviour [11].

Recent advances in image analysis, segmentation and machine vision have shown promising results for automated classification of fish. Machine vision has been successfully applied for quality grading [13] and sorting [14] of Atlantic salmon, sorting of herring [15] and segmentation of blood defects in cod fillets in 3D [16] achieving 99% accuracy using Support Vector Machines(SVM) and perfect classification using Convolutional Neural Networks (CNN) classification based on GPU processing. CNN algorithms has also been applied to successfully count fish in fisheries surveillance videos to monitor discarded catch under challenging conditions with random orientation and multiple occlusions, achieving a count error varying from 2 to 16% [17]. This system has been developed further to automate the monitoring of catch quotas [18].

Sonars and echo-sounders are widely used for remote sensing of life in the marine environment and preliminary work shows the potential of automated analysis of commercial medium-range sonar signals for detecting presence/absence of tuna in fishing vessels in a cost-effective way [19]. Scientific surveys are very costly and of limited coverage [20], thus the approach in [19] uses image processing techniques to analyse sonar screenshots. Scientific data was used to classify each region into a class (tuna or no-tuna) and build a dataset to train and evaluate classification models by using supervised learning. [21] and [22] used backscatter energy levels at multiple frequencies, e.g. discrete frequency analysis, as features for classification of fish species based on echo-sounder data. The Institute of Marine Research in Norway worked with quantification and identification of fish schools by hydroacoustic data together with SIMRAD (Kongsberg Maritime) to quantify and identify fish schools through hydroacoustic data through the years [22, 23]. Furuno and Kongsberg SIMRAD are the dominating fish-finding instrument brands globally. They are both starting to apply big data technology to improve their business by providing more sophisticated analyses and better information services related to their fish-finding instruments.

3.5. Fishery information services and open data sets

Olex (www.olex.no) is a system for combining data from GPS and echo sounders to provide detailed bathymetric maps based on crowdsourcing data from their customers and sharing maps based on community data back to them. The system is highly popular, and "Olex [has] shown that a collective [of fishermen sharing their data] is capable of producing results far beyond what could be imagined by the mapping community" [24]. Their system is highly relevant as it records and shares data from echo sounders, and if it is extended to report observed biomass estimates in addition to seabed depth, the system can boost the expansion of hydroacoustic data gathering from the fishing fleet. MarineTraffic (is a very popular information service for finding location and other information about vessels, ports, stations and offshore installations, including arrivals and departure times. They have more than 6 million monthly users visiting their site.

The Global Fishing Watch (www.globalfishingwatch.org) is a large project that maps fishing activity based on machine learning of vessel motion patterns [25, 26]. The project's strength is the massive and global analysis of AIS data, and the transparent data sharing policy in this project makes it stand out as a unique global resource. The open data portal and source code website gives access to highly relevant fisheries activity that can be accessed and processed onshore as part of the planning phase. An example of one-month fishing activity (July–August 2017) in the Norwegian and Barents Sea is shown in Figure 5.

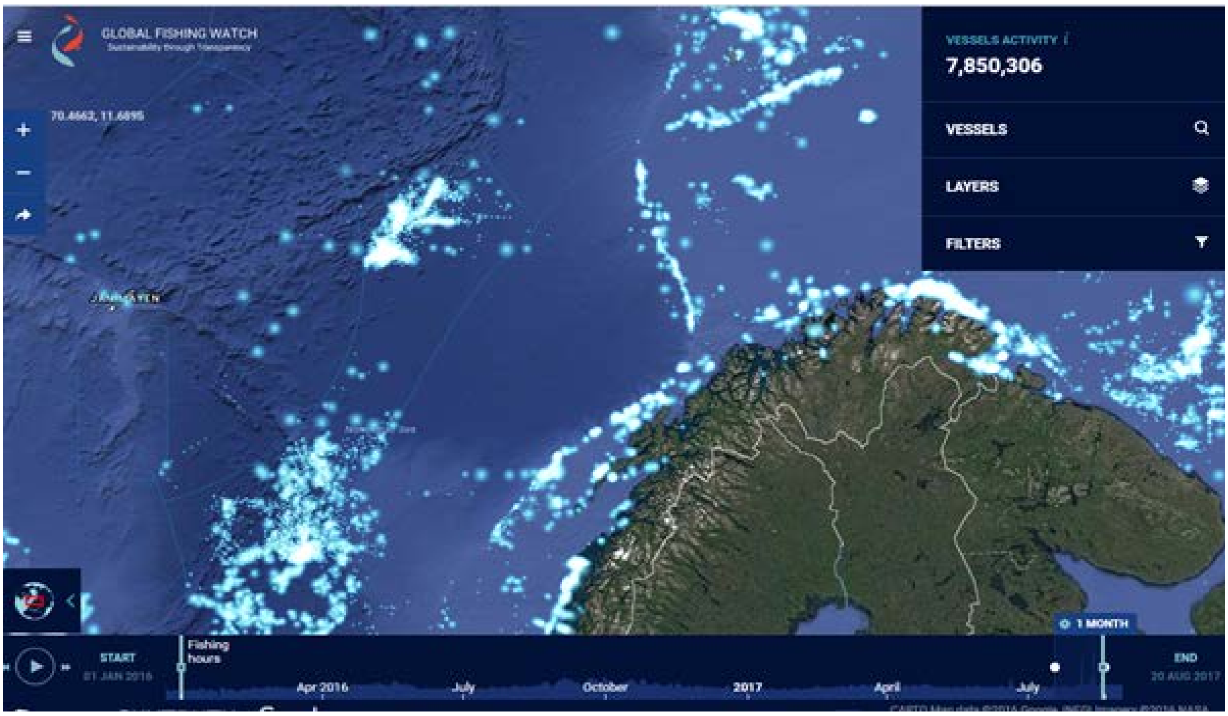

Figure 5: Norwegian Sea fishing activity according to Global Fishing Watch (Jul/Aug 2017).

BarentsWatch is a comprehensive monitoring and information system with a public portal for large parts of the northern seas focusing on the North Atlantic from Scotland to the Arctic waters (www.barentswatch.no). It includes the set of map-centric services for fishery activity, wave forecast, overview of ports, FiskInfo (translates to “FishInfo”) and polar forecasts including ice edge and concentration. The FiskInfo service is especially relevant as it shows where fishing activity is ongoing and which areas have been closed or restricted for fishing, cf. Figure 6. Hence, it is already a comprehensive portal with relevant information for fishery and more information layers are continuously being integrated. It is actively being used in Norway and maps files can be downloaded and shown using the Olex system and several chart plotters.

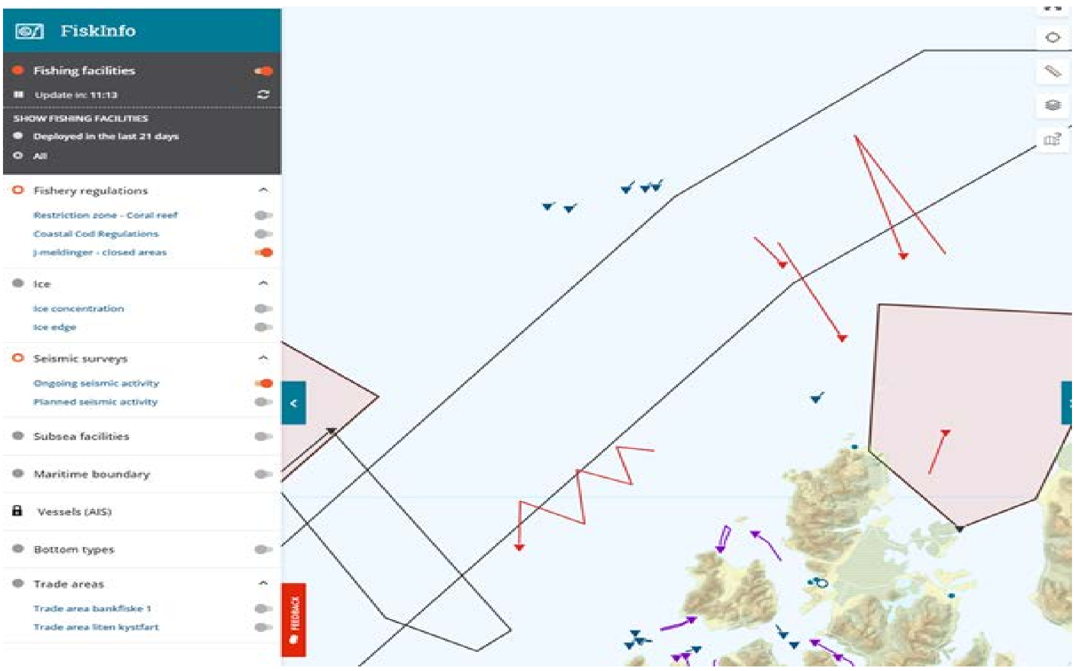

Figure 6: The FiskInfo service; example showing fishing activity with nets (blue), lines(red) and purse seiners (purple) as well as restricted (black polygons) and closed (filled polygons) fishing areas. From the fiskinfo.no web site.

EMODnet (The European Marine Observation and Data Network, www.emodnet.eu) is an organization supported by the EU's maritime policy which aims to be the gateway to marine data in Europe. A central challenge is to make available the marine data collected by many different institutions and research projects across Europe, which often have been carried out in a fragmented way for many years. Examples of data relevant for fishery includes catch statistics data per port, field observation data for many marine species, and sea surface and depth profile temperature data. EMODnet is stimulating marine innovation through open data sharing and encouraging developers to provide their marine applications as open source through GitHub.com. Refer to Table 1 for other examples of open data sets relevant for fisheries.

Table 1: Examples of open data sets relevant for fisheries.

Open Data Provider	Brief Description	Web links
ICES	Species and field observations, surveys	www.ices.dk www.ecosystemdata.ices.dk
Eurostat EU Statistics	Web API, EU + EEC export and import data	www.ec.europa.eu/eurostat
EUMOFA	Important EU/EEC fishery market data	www.eumofa.eu
EMODnet & EurOBIS	Seven themes, from physics to biology	www.emodnet.eu/portals www.eurobis.org/dataset_list
Copernicus (EU) and NOAA (US)	EO (Earth Observation), weather, climate, waves, temperatures	www.marine.copernicus.eu www.noaa.gov
Nature	Database of commercial, small-scale and illegal catch.	www.nature.com/articles/sddata201739

It is also important to mention that the SAM (State-space Assessment Model) model by Anders Nielsen and Casper W. Berg from DTU Aqua is used by ICES to estimate development in at least ten of the most economically important fisheries in Europe [27]. The model is web-based and anyone can enter data and check the intermediate results and figures used to generate a result and it is also possible to rewind all results to see the data used to reach a specific conclusion. This provides high transparency and easier insight between the researchers themselves as well as between ICES and the fishermen.

3.6. Potential of FishData

Current fishery information services and data portals go a long way in aiding the fishermen, shipping companies and authorities:

- Vessel monitoring and tracking systems show where fishing activity is ongoing, and where vessels and offshore installations are. This helps to increase transparency and making it harder for illegal, unreported and unregulated (IUU) fishing activities to avoid notice.
- Weather forecast services, Copernicus and other Earth Observation systems aid fishermen in planning routes to the fishing grounds and where to fish with consideration of weather, environment and ongoing fishing activities and regulations.
- Open data portals are doing a great job in making data available and discoverable while much work remains making it interoperable to facilitate data collation across different scientific domains. Important standardization work for data integration is ongoing, but the different data exchange services also need to implement this.

The current catch technology is quite efficient and the commercial trawlers and purse seiners rarely have trouble filling their catch quotas during the fishing season. The key question is rather how to leverage Big Data technology to optimise operation, planning and management of the fisheries to secure a best possible profit with minimal environmental impact on oceans and climate. A key challenge for the fishermen and shipping companies is to locate the fish as efficient as possible to reduce the time and energy needed to fill the quota at a time when prices are good. This is becoming harder as pelagic species have an increasingly changing migration pattern, and this observation is especially noticeable in Arctic waters.

How can data collection through FishData help in this perspective?

Taking the FishData platform to the next level for fisheries management than current systems could be achieved by creating a synergy between the commercial fisheries and the fisheries management and fisheries science, by utilising the fishing vessels presence in the marine environment.

Additionally, a large field of research dealing with the impact of fisheries on the ecosystem is limited by insufficient measurements of the impact of fishing gears on the sea bottom. A prerequisite for meaningful and reliable calculations of actual area swept and for assessing the scale and nature of the contact between the fishing gears and the benthic habitats would require a better information on the type and dimension of the gear actually used by fishers [28] and improvements could be brought.

Expanding the data collection could thus include:

- Integration of acoustic data collected by fisheries
- Measurement of sediment resuspension and seabed impact of towed fishing gears
- New technology for analysing the gear performance, fishing effort, energy use/fuel consumption and GreenHouseGas (GHG) emissions in fisheries
- Use of underwater cameras for in-situ monitoring
- Detailed bathymetry and seaweed habitat mapping
- Detailed gear specifications/dimensions (e.g. door spread or beam width),

Fishing locations can be produced by using the haul level spatial information in landing reports. An initial comparison of two sea-packing vessels, where Fully Documented Fisheries (FDF) data including GPS sensor data was also available, show good correspondence between SIF positional data and the positional data recorded in the FDF at a 1 minute temporal interval. Outputs for stakeholders would be maps and graphical representations of the data. Initially these could be static plots, with development the next stage would be interactive selection and customization of the outputs.

3.7. Related work by the SMARTFISH H2020 partners

Table 2. Ongoing related projects.

Project	Description	Relation to SMARTFISH H2020
DataBio (2017–2019)	Focus on datadriven bioeconomy in agriculture, forestry and fishery through pilot demonstrators utilising Big Data technology.	Provides initial work and concepts to develop further within SMARTFISH H2020.
Catch control in Norwegian purse seine fishery: Phase 2 (2017–2021)	Improve catch control in purse seine fisheries by developing instruments and analysis methods which provide a better basis for decision making during the fishing process.	May provide access to new data sources that SMARTFISH H2020 can also benefit from, especially echo sounder and ASDIC.

Cefas DP389: Automatic fish weights (2016–2018)	Investigating the availability and accessibility of data from fish grader machines in England. (Cefas funded project.)	Provided initial work to build on within SMARTFISH H2020.
DiscardLess WP4 subproject: Sea-packing data in the SIF database	Investigating the accessibility, coverage, consistency and reliability of data from sea-packaging machines and fish traceability system in Denmark, in order to assess whether it can be used for scientific studies and in management advice.	Provide example of existing system in use in the industry from which experience relevant to SMARTFISH H2020 can be gained
EBARTESA	The objective is to increase the current knowledge of activity of the artisanal fleet and their fishing grounds.	Provide a computer tool for the fleet to both reduce fuel consumption and record the daily catches which can be remotely sent to the ground stations.

Table 3. Existing technologies from SMARTFISH H2020 partners.

Technology	Description	Relevance for SMARTFISH H2020
Ratatosk	Onboard data acquisition, exchange and storage. Scalable, modular architecture.	Basis for the FishData onboard infrastructure.
STIM	Efficient analysis of time series.	Can be useful as part of a data analysis pipeline.
SINTEF Marine Data centre (SMD)	Robust, replicated storage. Centralised vessel configuration. Data processing and analytics. Communication channel with vessels.	Basis for the FishData onshore infrastructure.
PurSense	Decision support system for energy efficient operation of purse seiners.	Good example of how data acquisition and analysis can be used to improve the efficiency of fishing vessels.
Cefas C7361: Spurdog By-catch Avoidance Programme 2	An industry-managed scheme to monitor, avoid and reduce spurdog by-catch, preventing a ‘choke’ to UK fisheries.	Provided an initial low-tech real-time reporting tool to help avoid catches of spurdog, to be developed within the SMARTFISH project.
GESTOIL	Device and software to calculate and model the fuel consumption of the main engine in fishing vessels.	The system can monitor the fuel consumption, and thus to improve the energy-saving and the knowledge of the carbon-footprint in fishing vessels

Black Box Video System	Onboard data acquisition, exchange and storage system for haul level information including CCTV	System which has been used with success in several fisheries in Denmark. A current project in Australia is ongoing for a low power version of the system, enabling the system to be configured to different vessel types. Part of the project include the development of machine learning for automated species identification from images. Currently, monkfish (*Lophius piscatorius*) is used as initial example with promising results.
Fisheries Catch Monitoring System	System for monitoring the catch onboard especially developed for small scale fisheries (artisanal fisheries).	Provide a system that can be installed in small fishing vessels and has been already tested.

4. Use cases

In this section, we propose a number of use cases that the FishData infrastructure should support. Each of them is intended to highlight an area where systematic collection and analysis of available data can provide new insights and/or create value for one or more of the SMARTFISH H2020 project's target stakeholder groups. To recapitulate, the stakeholder groups are:

- The fishing industry
- Marine researchers who perform stock assessments
- Public fisheries management

One central challenge in fulfilling the objectives of the SMARTFISH H2020 project is to achieve the necessary collaboration and compliance from the fishing industry, most importantly the shipowners. It is therefore paramount that they benefit from participating, and several of the use cases are defined with this in mind. This will be achieved by providing both short- and long-term decision support for the shipowners and fishing vessel crew. In addition to fostering collaboration with the fisheries sector, this is expected to result in decreased fuel consumption and more resource-friendly fisheries.

Of the many use cases identified, three use cases are selected for initial development. These are described in enough detail to form the main requirements to the FishData architecture. The remaining use cases are still relevant for future development, either in this project or later. They are therefore also relevant for the infrastructure specification and are briefly described to support this.

As part of Task 5.4 and Task 5.5, the use cases will be further defined and developed into *demonstrators*. The extent to which they will be implemented will be decided at a later stage, when the cost/benefit trade-off can be determined more accurately. This will be done based on discussions with the relevant stakeholders.

4.1. Use case 1: Historical catch data analysis

4.1.1. Description

The amounts of time, fuel and emissions a fishing vessel spends per unit catch are directly related to the short- and long-term decisions made by the shipowner and the crew. These decisions are based on previous experiences and knowledge acquired from detecting patterns in past history. A use case for FishData is to enable shipowners and crew to more easily investigate historical data. As the fisheries show strong tendencies of cyclic behaviour, fishermen often express the importance of knowing and understanding how the fisheries unfolded in former catch seasons. By looking at for example how catch composition historically has varied between locations, the fishermen will be better able to plan when and where to go fishing for specific species. Similarly, historical catch efficiencies and historical prices can be of great value when planning the fisheries. Better tools for understanding the historical fisheries will therefore provide the means to improve their decision-making basis.

4.1.2. Method

To facilitate easier access to and improved analyses of historical data, we will create a web portal for presenting historical catches and trades. The portal will be regularly updated with new catches, and it will have functionality to enable the user to interactively analyse the data.

The primary data source will be publicly available catch data. For the sake of demonstration, we will use data from the Norwegian Directorate of Fisheries, available in the form of CSV files for every year since 2000. In principle, they contain all catches of Norwegian vessels and catches from foreign vessels landed in Norway, but the earlier years are not complete. The data for the current year are regularly updated. We will download the data for the current year at a regular basis, following the updates from the Directorate, while the data from previous years will be downloaded only once or when corrections to the official data are detected. The data which are newer than e.g. the last week are error checked and possibly corrected or augmented with other data (such as weather, tide etc.), before they are grouped according to e.g. time, species and location. The resulting dataset is then appended to, and partially overwriting, the operational dataset. This dataset contains 132 variables for each catch. The most relevant information which can be extracted is:

- Buyer: Name and location
- Fishing vessel: Name, ID, geographic affiliation, length, tonnage, building year, engine power
- Catch: Date, quota type, fishing gear, location and zone, landing time, species, quality, preservation method, size
- Transaction: Application, weight, price, value

We will create a web portal which show the most relevant of this data using visual elements such as charts and maps. The web portal will further make the user able to apply various filters and analyses to the variables. As he changes the filters, he can see the results as dynamically updated charts and maps. This way, we will make both historical and up-to-date information about catches and transactions available for end users. The user will be able to investigate for example how for a given species, price has varied by time, location and size. Or one can see how the species, sizes and quality of the latest catches have varied by location.

4.2. Use case 2: Catch composition and bycatch monitoring

4.2.1. Description

Information about the species and size composition of fish catches is of great importance for fishermen, fisheries management and fish stock assessment.

Fishermen can use knowledge about how the catch composition varies with fishing location to find the fishing areas which gives them the best combination of species and sizes with regard to their fishing quotas and the market situation. For example, they can avoid areas with too much bycatch or an undesirable size distribution. In itself, this could lead to less bycatch and discards.

Catch composition is also one of the factors which may dictate closing or opening of fishing areas by national fisheries authorities. The time it takes to establish a sufficient basis for such decisions will often determine the time it takes for the authorities to react to changes in catch composition. By making the basis for such decisions available with less delay, decisions on opening and closing of fishing areas can be made faster and more precise. This can lead to less undesired catches. Conversely, it can also help to avoid unnecessary closing of fishing areas.

Finally, the information is also highly relevant for stock assessment purposes, as the catch compositions will be correlated with the age and size composition and spatial distribution of fish stocks.

A use case for FishData is therefore to provide close-to-real-time information to vessel crew, fisheries management and stock assessors about the composition of catches made in different fishing areas. This could make the fisheries management able to respond faster when catch compositions dictates closing of fishing fields, and the fishing vessel crew would be able to find areas where their allotted quota can be fished to provide maximum value, with less fuel consumption. The idea is to combine available data – catch reports, landings reports and onboard catch analyses – and to provide catch composition as a continuously updated service. This could be either a web portal and/or push messages to subscribers when certain criteria are met.

4.2.2. Method

The method employed for this use case builds further on that described for the first use case in section 4.1.2. Specifically, we use the data downloaded from the Norwegian Directorate of Fisheries and the same methods for analysing it, and we use the previously-described web services as the basis for development of new ones.

In addition to the catch data from the Directorate of Fisheries, we now take advantage of detailed, real-time catch composition information collected directly from the individual vessels while they are at sea.

Measurements of catch composition onboard the contributing fishing vessels are performed continuously using the CatchScanner technology developed in SMARTFISH H2020 WP4 [29]. For development and demonstration purposes in this project (WP6), this is limited to the one fishing vessel with a CatchScanner unit installed.

The data collected from the CatchScanner comprise estimates of the following *for each individual fish*:

- Weight
- Length
- Species

In addition, the following data will be collected from vessel systems:

- The vessel's position, acquired from the onboard GPS
- The vessel's water speed, acquired from the speed log
- Onboard energy consumption, acquired from the power generators
- Propeller RPM and torque/thrust, acquired from the propulsion system (if possible)

These data are needed to be able to associate the catch composition data with the time and geographical location at which the fish was caught. Substantial time may pass from capture to processing of the fish, so simply sampling the position at processing time is insufficient. Instead, data collected from vessel systems, such as those listed above, can be used to determine with reasonable certainty which type of operation the vessel was performing at any given time in the past using a method developed by SINTEF, outlined in a report from the DANTEQ project [30]. (As an example, if the vessel is moving slowly but with high propulsion power, this could be an indication that it is trawling.) Each CatchScanner sample will then be associated with the midpoint of the last trawling operation.

Catch composition reports are sent to shore, containing information per individual fish or in the form of high-fidelity distributions for shorter time periods. For the onshore services, it is important to have the data available without much delay. At the same time, these data are quite small. The transfer of these data is therefore given a high priority. Communication channels with high cost and low bandwidth, such as satellite, are therefore used for transferring these data when other communication channels are not available.

On shore, the incoming data are aggregated and combined with the data from the Norwegian Directorate of Fisheries. This dataset forms the basis for both visualisation in the web portal and for pushing messages to subscribers. In the web portal, the data will be shown in maps, charts and numbers, dynamically updated according to the user's selection of filters and analyses. As an example, the user can choose to show in a map the average weight of caught fish or the bycatch percentage in each catch location, only considering catches exceeding a given quantity of a given species in a given time period.

4.3. Use case 3: Improvements and updates to CatchScanner using data from commercial fisheries

4.3.1. Description

The CatchScanner technology being developed in SMARTFISH H2020 WP4 is based on machine learning techniques [29]. It uses special cameras to acquire 3D and 2D colour images of fish that pass on a conveyor belt, and a neural network analyses the images to assess the species, length and weight of each individual fish. This neural network needs to be trained on pre-labeled images of fish, and the more samples it is trained on, the better its estimates become.

A use case for FishData is to:

1. Collect images acquired by CatchScanner units deployed on commercial fishing vessels and transfer them to shore, where they can be used for further training of the neural networks and thus improving the product.
2. Transfer updated neural networks – or more precisely, updated numerical *weights* associated with the neurons and connections in the network – from shore back to the CatchScanner units.

This has a number of benefits:

- Continuous improvement of the CatchScanner product so its estimates become more accurate over time.
- Automatic updates of the neural network weights on deployed CatchScanner units so that users benefit from the improvements.
- Tailoring of the neural networks to specific fisheries, geographical areas and/or seasons, to potentially improve their accuracy even more.

For example, if a vessel plans to fish in a specific region, it could download the most up-to-date weights for that region before it sets out, and it could contribute with new data for the same region when it returns.

4.3.2. Method

From an onboard CatchScanner, the following data will be collected *per scanned fish*:

- The image data, which constitute the *input* to the neural network
- The *feature vector*, which is the last layer in the neural network
- Species, size and weight estimates, which are the *output* from the neural network

In addition, the following data will be collected from vessel systems:

- The vessel's position, acquired from the onboard GPS
- The vessel's water speed, acquired from the speed log
- Onboard energy consumption, acquired from the power generators
- Propeller RPM and torque/thrust, acquired from the propulsion system (if possible)

The data will be accumulated for several hauls in onboard storage and transferred to shore when the vessel is in range of mobile communication or Wi-Fi networks. The inclusion of images makes these data relatively high-volume, so satellite communication will not be used. Satellite bandwidth is typically limited and expensive, and this particular use case does not require real-time data; a delay of one or more weeks from collection to transfer is just fine.

The position, speed and energy consumption of the vessel will be used to estimate approximately when and where each fish was caught, in the manner mentioned in section 4.2.2 and described in the DANTEQ report [30].

Once the data have been received and stored in the onshore data centre, the CatchScanner developers will have access to them through a download service. This will enable them to improve the image recognition software, for example through manual inspection and correction of the results, followed by re-training of the neural networks. This is discussed in SMARTFISH H2020 deliverable D4.1 [29], and we will not go into further details in the present document.

Once the neural networks have been re-trained, the new weights will be uploaded to the data centre, from which they can be downloaded by the vessels and applied in the onboard CatchScanner units. There may be several sets of weights, each associated with a particular fishery, geographic region and/or time of year, and the organisation of the weight files must reflect this and allow selection of specific sets.

4.4. Other use cases

While the three use cases described above have been selected as the primary use cases for demonstration of the FishData technology, several others have been identified. These are briefly described in the following. They may be considered for development depending on the progress of the primary use case demonstrators.

4.4.1. Onboard fishing activity detection

By using only information and systems available onboard a fishing vessel, it should be possible to detect fishing activity around the vessel in real time. The main idea is to use machine learning to analyse the data available from the vessel AIS receiver and find patterns in the movements and behaviour of other fishing vessels which indicate that they are engaged in fishing activity. This would both increase the range where such activity can be detected from approximately 10 nautical miles to 30 nautical miles (increasing the detection area 9 times), as well as remove the need for the vessel crew to continuously monitor the radar for fishing activity.

4.4.2. Vessel information portal

A vessel's own past experiences contains important information for planning future fisheries. By looking into what the vessel did previous years, one can get an idea of what would be sensible to do this year. But the documented history of a given vessel is often either too detailed (e.g. time series of vessel speed, machinery operation, etc), too cumbersome to access (e.g. written logbooks) or too coarse (e.g. catch reports) for detailed manual analysis of how the fisheries have developed over several years. As an answer to this, one can create a searchable database of the past, based on available data sources. A central task will be to analyse the individual vessel time series of operational data to extract information about fisheries activity. This will cover not only catching operations, but also information from the search phase.

4.4.3. Onshore fishing activity detection

An up-to-date overview of ongoing fishing activity would be useful as a decision support tool for both shipowners and fisheries management. Multiple available data sources can be integrated to automatically detect such activity and provide this information to both shipowners, fishing vessel crew and fisheries management. As different data sources have different delay, the accuracy of the detection will depend upon how close to real-time one wants to operate. The most promising data source is probably AIS data relayed by satellites. This is available in real time, and can enable detection far from shore. But the quality is often poor and the coverage is variable, and this will present additional challenges for the analysis. Catch reports will give a larger time lag but can give more accurate information about the outcome of the fisheries effort. Access to the necessary data sources must be secured before such a service can be developed.

4.4.4. Biomass monitoring

The marine biomass is to a large extent unknown, despite the considerable resources spent every year to chart it. This is not only related to the fish stock sizes, but also to their time varying spatial distributions and fishery-relevant behaviour. This use case will seek to provide information about the spatial distribution of fish stocks. The shipowners will be able to improve their planning, and thus catch their set quota spending less time and fuel, and also optimise their profit by being able to serve their market better. To make this possible, this use case should analyse data sources which can tell something about biomass distribution, such as echo sounder data collected from vessels in addition to the data sources mentioned in Use case 4. In addition, this use case may include simulations and predictions of physical and biological processes in the ocean, such as nutrients and plankton. The results would be distributed to fishing vessels contributing with data through either a web service or directly to their chart plotter.

4.4.5. Variations in catch composition and selectivity

One important aspect within fisheries is to maximise the yield of the sea in general and of the fish stocks in particular. Optimising the biomass production is one important aspect, but which species and sizes are taken out and at what time is also significant. This use case will aim to provide a better understanding of how catch composition and selectivity changes with vessel, area and time. To this end, data from grading machines, catch reports and onboard packing will be collected and analysed.

5. Data management

The use cases defined in section 4 can be seen as processes which take input data and deliver output data. The data which flow through the systems will have different requirements with regard to security and access policies, and the developed architecture must support these requirements. This section describes some of the considerations which need to be taken into account.

5.1. Vessel data

Data which are collected directly from fishing vessels are considered confidential unless the shipowner instructs otherwise. In addition, these data are impossible to reconstruct if they are lost or if their correctness is in doubt. When handling these data, the following requirements must be met:

- Onboard the vessel, access to the data must be controlled.
- The data must be transferred over an encrypted channel.
- The data must be stored on shore protected against hardware failure and bit rot.
- The data must be stored on shore with strong access control.
- If the data are used in public or private services, permissions from the shipowner must be obtained and care must be taken to not disclose confidential information.

5.2. Publicly available data

Public data, such as catch data and meteorological data do not have any privacy issues. These data can be divided in two categories, according to how severe their loss would be. In many cases lost data can be very simply be replaced; in other cases the most recent data are easily obtained but historical data could e.g. be either expensive or impossible to obtain. The following requirements must be met:

- The data must be stored onshore protected against hardware failure and bit rot. If the data can easily be replaced, this requirement is reduced to being able to detect any loss or corruption of data.

5.3. Output from analyses of private data

When private data are run through analyses, the access to the data will to a large extent depend upon to which extent the analyses hide the identity of the data source and/or the details of the data. Another important criterion is how time and cost consuming the recreation of the data will be. The requirements are:

- The data must be stored onshore protected against hardware failure and bitrot, or at least it must be possible to detect corruption or loss of data.
- The data must be stored onshore with strong access control, unless the data are deemed to have lost their status of confidentiality.
- If the data are used in public or private services, permissions from the shipowner must be obtained and care must be taken to not disclose confidential information.

5.4. Output from analyses of public data

When public data are run through analyses, one possible reason to secure the resulting data is to hide the analysis method and/or implementation. This is not regarded as relevant for the SMARTFISH H2020 use cases. Another important criterion is how time and cost consuming the recreation of the data will be. The requirements are:

- The data must be stored onshore protected against hardware failure and bit rot, or at least it must be possible to detect corruption or loss of data.

6. Architecture

Figure 7 shows a bird's-eye view of the FishData infrastructure and how it relates to external agents, including stakeholders (which will be the end users), other data centres, and data sources such as fishing vessels and third-party services. It uses the *Fundamental Modeling Concepts* (FMC) notation, about which we've provided more information in Appendix C. The main entities in the diagram are:

- *Data centre*: Servers and systems responsible for storage and analysis of data.
- *Stakeholders*: The users of the system; beneficiaries of data analysis results.
- *Trusted partners*: Other data centres or services with which data should be exchanged.
- *Fishing vessels*: Vessels which contribute and/or use data.
- *Third-party data sources*: Data sources not under our control, e.g. meteorological services.

The data centre should be secured behind a firewall that restricts inbound connections (though it may allow outbound connections). Most of the communication with the outside world should therefore happen via intermediaries—services that are exposed to the Internet and can publish or receive data:

- *Web portal*: Web services that provide data and analysis results and are available for access by stakeholders. It is fed with data from the data centre.
- *Outbound*: An intermediate storage that can be accessed by trusted partners, typically through secure channels that require authentication. The data centre periodically pushes selected data and results to it.
- *Inbound*: An intermediate storage to which data gets pushed by trusted sources, such as fishing vessels and trusted partners. For most purposes, this should happen through secure channels that require authentication.

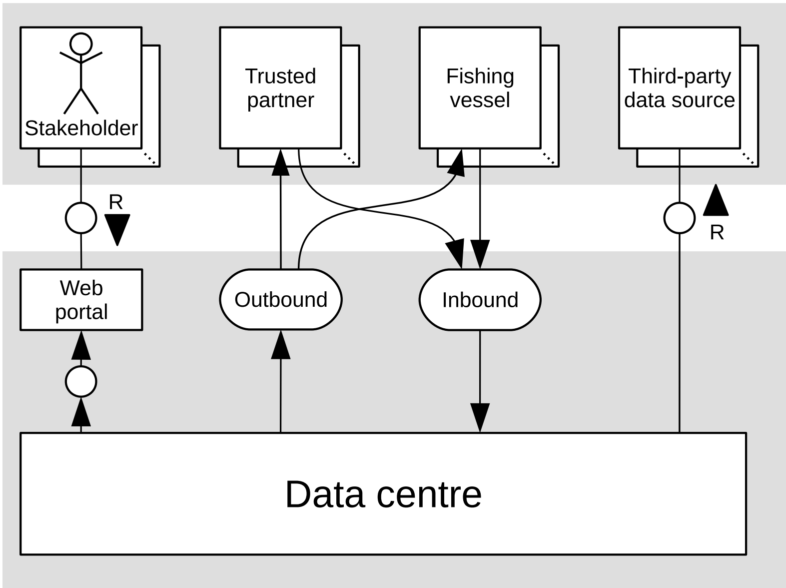

Figure 7: FMC block diagram of overall architecture.

The elements inside the lower-left gray area of Figure 7 are part of the FishData *onshore infrastructure*, described in Section 6.2. The *onboard infrastructure*, which we discuss in Section 6.1, is the recommended way to collect, manage and transfer data within each fishing vessel.

6.1. Onboard infrastructure

The primary responsibilities of the FishData *onboard infrastructure* are:

- Acquisition of data from onboard instruments and sensors.
- Making the acquired data available on board over a common protocol.
- Time synchronisation of data samples.
- Storage of data in a common file format.
- Transmission of data between ship and shore.

A schematic illustration of the infrastructure is shown in Figure 8. Details on the constituent components are provided in the following subsections.

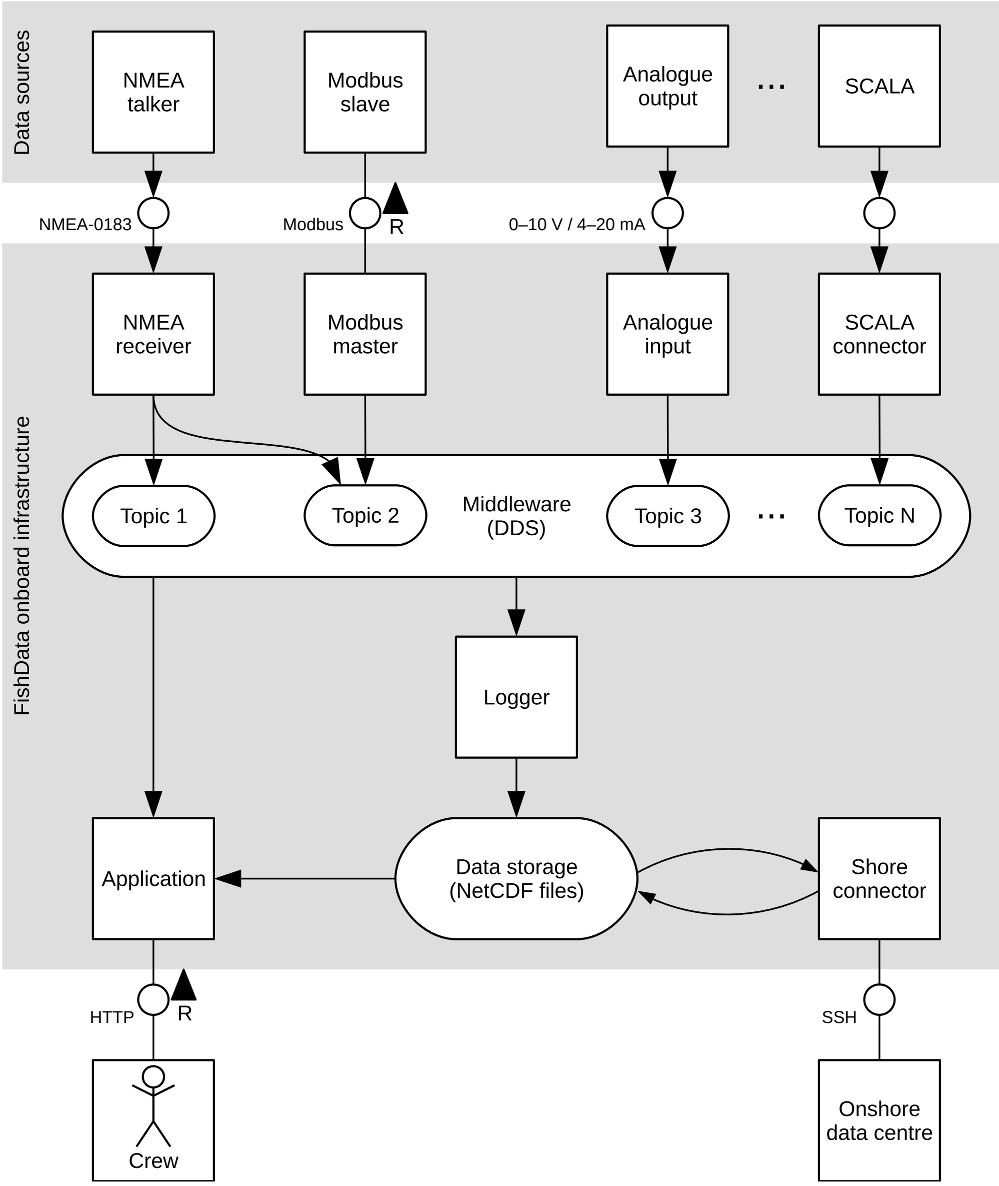

Figure 8: FMC block diagram of onboard infrastructure, with some generic examples of data sources.

6.1.1. Middleware

The middleware layer provides a common protocol and transport mechanism for the collected data and ensures that the time stamps associated with data samples all follow the same clock. It should enable a highly modular architecture where data "producers" and data "consumers" are decoupled from each other to the extent possible. This has two major advantages:

- New components, including both data sources and applications, can be added without reconfiguration of the rest of the system.
- Data consumers do not need to know the details of the underlying data sources, such as the original protocols, formats, IP addresses, and so on. They only deal with the middleware.

These properties will ensure scalability and robustness in the face of future changes and extensions to the system. They also enable the development of functionality (applications) without regard to the specifics of a particular vessel. (This is analogous to how developers of a smartphone app usually do not need to worry about the specific makes and models of a phone's hardware components; it's all abstracted away by the device operating system APIs.)

Implementation

For the FishData middleware layer we use the *Data Distribution Service for real-time systems* (DDS), an open standard for communication middleware [31]. DDS is used for real-time communication in a variety of industries, including transportation, smart energy systems, medical devices, automation and more [32].

DDS is based on a *publish-subscribe* model of communication. Participating entities can publish or subscribe to *topics*, which are channels for communicating structured data. Each topic may have multiple publishers and multiple subscribers. Conceptually, the topic exists independently of these entities. The entities only have to agree on the topic definition, and do not need any other information about each other. Thus, the participating entities are decoupled, making the system extremely scalable.

Formally, a topic consists of a *type* and a *name*. The type defines the structure of the data, while the name distinguishes the topic from others. A type is defined using a subset of the *OMG Interface Definition Language* (IDL) [33], from which an *IDL compiler* can generate software code to perform the actual communication.

As a simple example[1], a topic that represents a ship's position (from GPS) might be called `vesselPosition` and have a structure defined by the following IDL type declaration:

```
struct GlobalPosition
{
    string sensorID;
    double latitude;
    double longitude;
};
```

Here, the type name is `GlobalPosition`. (Note that the type and the topic are named independently, because the same type may be used for multiple topics.) If the ship has several redundant GPSes, they can all publish to the same topic, only distinguished by separate `sensorID` values.

A single data source will often be associated with several topics. To use the GPS as an example again, it would typically publish the vessel's position, course, and speed over ground to three separate topics.

6.1.2. Data sources

On a modern fishing vessel a multitude of instruments and sensors produce data in a variety of formats, communicating those data over many different protocols. Some of the most common protocols are illustrated in the topmost gray area of Figure 8:

- *NMEA*: A combined specification of an electrical interface and a data protocol for communications between marine instrumentation [34], commonly used for GPS receivers, gyrocompasses, speed logs, anemometers and more. A "talker" is an instrument which sends data over NMEA. In our experience, the version most commonly used in commercial fishing vessels is NMEA-0183.
- *Modbus*: A communications protocol, originally designed for use with programmable logic controllers, which has now become a *de facto* standard protocol for communications between industrial electronic devices. For the purposes of on-board data acquisition, Modbus is a common means to expose data from automation and control systems. It is used over both serial connections and ethernet.
- *Analogue I/O*: Many low-level sensor measurements are available as analogue signals only, either as a voltage signal or a current signal. The industry standard is 0–10 V for the former and 4–20 mA for the latter.
- *SCALA:* Marport's bridge software system, which enables real-time communication with a wide range of sensors and instruments mounted on the fishing gear, such as catch sensors, depth sensors, temperature sensors, motion sensors, echo sounders and more.

[1] While we do actually specify a data type named `GlobalPosition`, this example is deliberately simplified. See appendix B for a listing of the proper data structures.

Implementation

As we have chosen DDS as the common data exchange platform, it is necessary to create bridge modules that sample the signals from the other communication methods and publish them over DDS. This is illustrated by the upper line of boxes in the large gray area of Figure 8. Table 4 provides an overview of the specific data for which such bridge modules are needed to support the use cases described in Section 4.

Table 4: A non-exhaustive overview of relevant data collected from fishing vessels. See Appendix B for IDL definitions of the types listed under *DDS topic type*.

Measurement	Data source(s)	Source protocol	Type	Components	DDS topic type	DDS topic name
Vessel position	GPS	NMEA-0183	Sampled continuous signal	latitude (deg.), longitude (deg.)	`GlobalPositionMeasurement`	`vesselPosition`
Vessel speed over ground	GPS	NMEA-0183	Sampled continuous signal	velocity (m/s)	`VelocityMeasurement`	`vesselGroundSpeed`
Vessel speed through water	Speed log	NMEA-0183	Sampled continuous signal	velocity (m/s)	`VelocityMeasurement`	`vesselWaterSpeed`
Heading	Gyrocompass GPS	NMEA-0183	Sampled continuous signal	direction relative to north (deg.)	`HeadingMeasurement`	`vesselHeading`
Generated power	Generators (main and aux.)	Modbus	Sampled continuous signal	power, voltage	`ElectricalPowerMeasurement`	`generatedElectricalPower`
Caught fish	CatchScanner	DDS (no translation needed)	Discrete messages	species, length (m), weight (kg), thumbnail image	`IndividualFishStats-Measurement`	`individualFishStats`

6.1.3. Logging

Most use cases require collected data to be stored onboard, even if only temporarily. This will be handled by a dedicated module ("Logger" in Figure 8). Typically, the logger will subscribe to topics that represent raw data, sample those data at a given frequency and store them to file along with the sample time. The fact that all signals are sampled at the same time, in a single component, is crucial for maintaining the coherence of the data and avoiding the difficult and time-consuming work of merging different data sources later.

The logger component will primarily handle structures of "primitive" data types such as numbers, boolean values, text strings, etc. Dedicated modules should probably be developed to handle special data like images and video streams.

Data should primarily be stored in a file format that allows storing metadata, so that the files are as self-documenting as possible. Most importantly, it should always be possible to trace the data back to the source from which it was acquired.

Implementation

The chosen storage format for time series data is Unidata's *Network Common Data Form* (NetCDF) [35]. Quoting the NetCDF fact sheet [36],

> "[NetCDF] is a set of software libraries and machine-independent data formats that support the creation, access, and sharing of array-oriented scientific data. It is also a community standard for sharing scientific data. Data in netCDF format is:
>
> - *Self-Describing.* A netCDF file includes information about the data it contains.
> - *Portable.* A netCDF file can be accessed by computers with different ways of storing integers, characters, and floating-point numbers.
> - *Scalable.* Small subsets of large datasets in various formats may be accessed efficiently through netCDF interfaces, even from remote servers.
> - […]
> - *Archivable.* Access to all earlier forms of netCDF data will be supported by current and future versions of the software."

The combination of a binary format and built-in compression makes NetCDF very efficient in terms of storage space and network bandwidth.

Metadata

The following file-level metadata will be stored in each data file:

- `OriginType`: An overall class for the data source at which the file's contents originated. Valid values are:
  - `ship` for data collected on board vessels.
  - `web` for data retrieved from web services.
  - `undefined` for other sources.
- `OriginId`: A machine-readable name which uniquely identifies the origin of the data.
  - For `ship`: The vessel's IMO number on the form `IMO<number>`.
  - For `web`: The service's URL.
  - For `undefined`: Unspecified.
- `Origin`: A human-readable name which identifies the origin of the data.
  - For `ship`: The vessel's name.
  - For `web`: The service name.
  - For `undefined`: Unspecified, but an appropriate value should always be provided.
- `Title`: A human-readable, descriptive name for the data set stored in the file.
- `Summary`: A human-readable description of the data set stored in the file.
- `CreationTime`: The time at which the file was created.

- `ModificationTime`: The time at which the file was last modified.

In addition, the following metadata should be stored *per variable* (i.e., column) in the file:

- `Name`: A machine-readable name for the variable, consisting only of alphanumeric characters and underscores (like the variable naming rules in most programming languages).
- `LongName`: A human-readable, descriptive name for the variable.
- `Units`: The physical units in which the variable values are stored (if applicable).
- `Offset` and `Scale`: (If applicable.) The values of a variable v may be stored in the file in a scaled form u so that $v = o + su$, where o is `Offset` and s is `Scale`. This is useful if:
    - The magnitudes of the values are typically much larger than zero, in which case scaled storage can help to preserve precision.
    - The raw values coming from the data source are already scaled in this way, and one wants to store them in their original form.

Chunking

While NetCDF is able – indeed, designed – to store very large amounts of data in a single file, we limit the number of samples in each file for the purpose of onboard storage. This has to do with the challenging conditions under which the system will operate, in particular:

- Unreliable satellite/mobile communication means that file transfers to shore will frequently be interrupted. Smaller files allow transfers to take place in smaller chunks.
- Power outages are relatively frequent on a ship, at least compared to onshore installations. If the logging computer's power is cut during a write operation, the file being written could become corrupt. Smaller files ensure that smaller chunks of data are at risk of corruption at any given time. (It should also be noted that the NetCDF format has built-in resilience mechanisms that further lower the risk of data corruption.)

Typically, we store one hour's worth of data in each NetCDF file, but this is configurable. The files may be combined into larger units once they have been transferred to shore, if this is desirable.

Layout

Each file can be seen as a matrix where the columns represent data series and the rows represent samples. For sampled continuous signals (see Table 4 for examples), the first column represents the time axis, i.e., it contains a common time stamp for all samples in each row. This means that signals that are sampled with different frequencies need to be stored in different file sets.

For data that consist of discrete messages, possibly received at irregular time points, the primary axis, or *dimension* in NetCDF terminology, needs to be defined on a case-by-case basis. Typically, it will be some value that uniquely identifies a particular message. (A system like the CatchScanner will probably just generate unique, sequential ID numbers.)

6.1.4. Ship–shore data transfer

Communication of data between the vessel and a shore-based data centre is handled by a dedicated module labelled "Shore connector" in Figure 8. It is responsible for monitoring the vessel's connectivity status—that is, whether it has access to the Internet, and through which channels—and for transferring the appropriate data at the appropriate time. For example, certain data may be small enough and/or volatile enough to transfer in near-real time via satellite, while the bulk of the collected data should most certainly only be transferred with cheaper and faster methods like mobile broadband or Wi-Fi whenever such is available.

Implementation

The *rsync* software utility [37] is used to perform the actual file transfers. It uses delta encoding and compression to minimise network usage and can recommence broken transfers, which makes it very suitable for use onboard ships where connectivity is often variable. For security (authentication and encryption), rsync will always connect to the data centre using the SSH protocol. Each vessel will be equipped with its own, unique SSH key that gives it access to a vessel-specific account on the onshore server (section 6.2.3).

Data may be transferred in both directions: Data collected onboard the vessel is transferred to shore, while data which have been aggregated and processed in the onshore data centre, may be transferred to the vessel for use by onboard applications. Such outbound data can originate from *other* vessels, onshore data sources or data center simulations. Rsync supports bidirectional synchronisation out of the box and is the preferred tool for data transfer in both directions.

6.2. Onshore infrastructure

The architecture of the FishData onshore infrastructure is illustrated in Figure 9. Starting from the bottom, it consists of the following three main areas:

- *Private network*: Services running in this part of the data centre cannot be accessed from the outside, as they are protected by a firewall which blocks incoming connections.
- *DMZ* (demilitarised zone): Services running in this network zone may be accessed from the outside, but in a controlled manner. It is protected by a firewall which only allows certain types of inbound traffic, such as HTTP requests and SSH connections.
- The outside world (typically the Internet).

Each zone may consist of any number of computers, sharing the workload and network traffic between them. For the chosen use cases, a major part of the workload will be analysis and handling of non-public data. For this reason, most servers are placed in the private network zone. Another point which supports this choice, is that the structure shown in Figure 9 is conceptual and does not necessarily reflect the actual topology. For example, one possibility is to let the DMZ contain only one or more reverse proxies that relay communications from the outside to the private network. In that case, the actual nodes that run the *Web portal*, *Outbound data* and *Inbound data* services could be on the private network. This is especially useful if the private network consists of a large number of machines. Then, the various services, and even multiple instances of each service, can be distributed dynamically across the cluster, while the reverse proxies route and load-balance the incoming traffic to them.

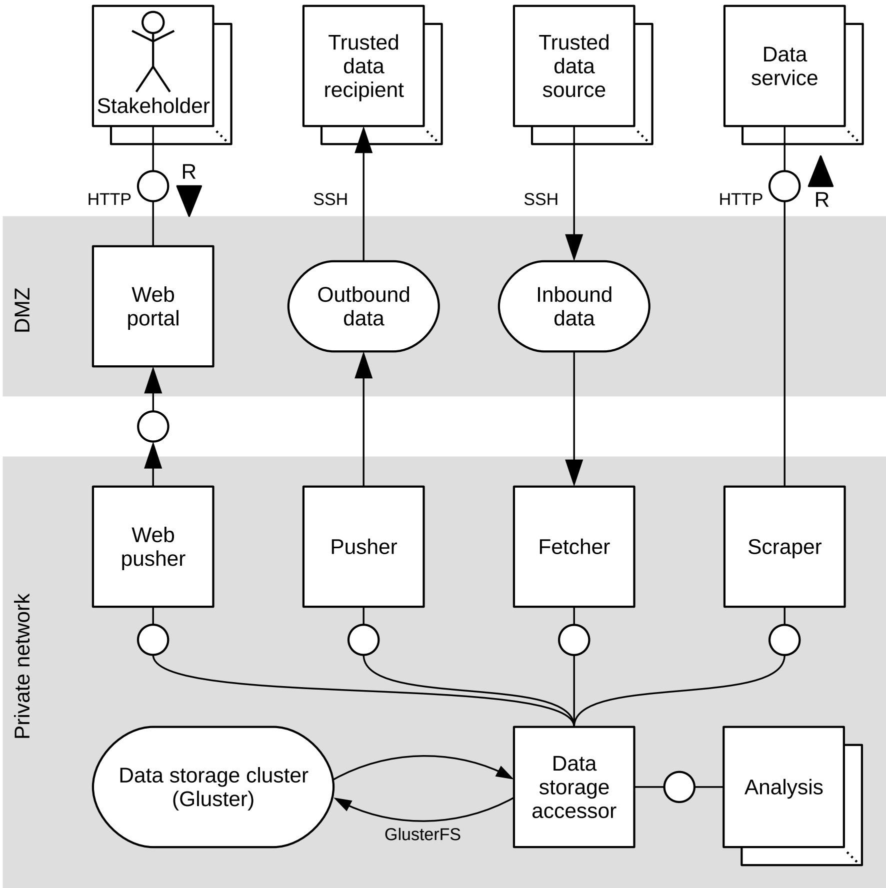

Figure 9: FMC block diagram of the onshore infrastructure. Some connectors are left unannotated because the boxes they connect represent generic service *categories*, and their interfaces/protocols are chosen on a case-by-case basis.

The services shown in the DMZ in Figure 9 are:

- *Web portal*: A web server that hosts a web portal where stakeholders can access data services.
- *Outbound data*: A file storage to which trusted data recipients can connect and acquire data.
- *Inbound data*: A file storage to which trusted data sources can connect and push data.

Trusted data recipients and *trusted data sources* include fishing vessels and trusted data centres. One entity, such as a fishing vessel, may be both a recipient and a source of data.

Connections between trusted entities and the inbound and outbound data storages should require authentication and be encrypted. The recommended method is to use SSH, for example with *rsync* as described in section 6.1.4.

The *private* services shown in Figure 9 are:

- *Data storage cluster*: A secure, redundant data storage.
- *Data storage accessor*: A central component which handles authentication and access control for the data storage cluster. All access to the storage must go through this service.
- *Analysis*: Services that process data to produce results.
- *Web pusher*: Periodically pushes data to the web portal.
- *Pusher*: Periodically reads specific data from the main data storage and pushes them to the outbound data storage.
- *Fetcher*: Periodically retrieves data from the inbound data storage and writes them to the main data storage.

- *Scraper*: Periodically queries one or more third-party data sources for data (e.g. meteorological services).

Note that there may be multiple instances of all these services. There are two main reasons one would want this. The first is separation of concerns, as it may improve security and scalability to have specific instances only handle particular data aimed at specific outside entities. The second is performance, as it makes it possible to share the computational load or traffic between multiple machines.

In the following sections, we provide more details on the various services outlined above.

6.2.1. Access control

User authentication for external web services is based on the OAuth2 protocol, with Okta as an external authentication service. To provide single sign on (SSO) capabilities for arbitrary services, the authentication and authorisation is handled by a dedicated service. This service maps user email addresses to authorisation groups and returns for specific combinations of user email, service and path either "Access granted", "Access denied" or "Login required". This service is implemented as a Flask web service, running on the same physical server as the external load balancer and interrogated by the load balancer for the access rights for each access request.

User authentication for server access and direct data access is handled by SSH keys, which means that the public keys of each client which is to have access must be sent to the servers for which he is to have access. The public keys are primarily exchanged by a dedicated key exchange service, but they can also be exchanged via other channels. The dedicated key exchange service publishes the public keys in a dedicated web server over https. The keys are published in a structured way which makes it possible for servers to extract and update their list of authorized keys for the user accounts which are handled.

6.2.2. Orchestration

Since the services are distributed across several servers in a cluster, we need an orchestration mechanism. By orchestration we mean the following tasks:

- Starting and monitoring long-running services
- Running short-lived jobs (periodically or event triggered)
- Allocating cluster-wide resources such as CPU, memory and disk space to services and jobs

Implementation

For orchestration, we use *DC/OS* (the Distributed Cloud Operating System), an “open-source distributed operating system” [38]. DC/OS consists of a large number of components that provide different aspects of its functionality. Here, we will only mention the ones that directly address our requirements:

- *Apache Mesos*: A distributed systems kernel that “abstracts CPU, memory, storage, and other compute resources away from machines (physical or virtual), enabling fault-tolerant and elastic distributed systems to easily be built and run effectively” [39].
- *Marathon*: A service orchestration platform for Mesos [40] that takes care of starting long-running services and ensuring that they remain up and running. Examples of components in Figure 9 that are implemented as Marathon services include the web portal, the inbound and outbound data buffers, the storage accessor, and possibly some analysis services.
- *Chronos*: A job scheduler for Mesos [41] that runs the “short-lived” jobs, either according to some time schedule or triggered by the completion of other jobs. Examples of components in Figure 9 that are managed by Chronos include the pusher, fetcher and scraper jobs, as well as certain analysis jobs.

(By way of analogy, if DC/OS were a GNU/Linux distribution, then Mesos would be the Linux kernel, Marathon the *init* process, and Chronos would be *cron*.)

The services themselves are run as *containers*, i.e., operating-system-level virtual machines. This means that, from the perspective of the software components that together constitute a service, the service appears to have a whole computer to itself.

Notably, the data storage sub-cluster is the one part of the data centre which is not managed by DC/OS. We discuss the reasons for this in Section 6.2.5.

6.2.3. Inbound and outbound data

Communication of data between the data centre and some external entity, whether it is a fishing vessel or another data centre, is performed in a two-stage process. For incoming data, the external entity connects to an Internet-facing service (*Inbound data* in Figure 9), and uploads its data to it. This service is merely a buffer; it is not intended for permanent storage. An internal service (*Fetcher* in Figure 9), regularly pulls data from the buffer and transfers it to the internal data storage, removing it from the buffer. For outbound data, the process is the exact opposite. This reduces the level of access needed by external entities to an absolute minimum, and it limits the consequences of a potential security breach since only a modest amount of data will be present in the buffer at any given time.

Implementation

The inbound and outbound data services are simple containers that mainly run an SSH server instance, and which have access to a sufficiently large amount of temporary storage space. Each external entity has its own user account with an isolated directory in which to store files. User authentication is performed with SSH keys, as described earlier.

6.2.4. Web portal

The *web portal* service is a web server that allows stakeholders to access data, especially the results of the analyses performed by the analysis services. Like the inbound and outbound data buffers, it does not have direct access to the internal data storage; instead, any relevant data updates get pushed to it by an internal service. The reasoning is the same as for the inbound and outbound data buffers.

Implementation

[Note: This is the least-specified part of the infrastructure, since its complete specification and development is to be carried out in the final stage of WP5 – tasks 5.4 and 5.5. The following should therefore be read as notes of a work in progress.]

The web service is based on the Flask framework. This service exposes a web portal with endpoints for both data and visual web pages. The visual web pages are created using a combination of the javascript libraries, such as React, Crossfilter. D3, DC, Bootstrap and Leaflet. The data are visualised in a map and on various DC and D3 plots. The user can apply various filters (implemented using Crossfilter) on variables such as species, time, location and size. The results are seen as dynamically updated charts (implemented using e.g. DC). User authentication is performed using the OAuth protocol, as described earlier.

To support rapid iterative development, both a development and a production version of the web portal are created. Each version is based on their separate branch in the Git version control system. The Docker containers running the web services will regularly check for updates in the version control system and build the new version if updates are available.

6.2.5. Data storage

The data storage is perhaps the most crucial part of the entire infrastructure, as the data it contains can be of high value and are often irreplaceable. To protect against data loss and service downtime, the following criteria must be met:

- The data storage consists of two or more machines in a cluster.
- Each machine contains multiple disks which are set up in a RAID configuration with redundancy, so that at least one disk may fail without loss of data.

- Important data is replicated across multiple machines, so that at least one entire machine may fail without loss of data.
- Backups should be made regularly to an off-site storage, at least once every 24 hours.

The above criteria ensure that the storage adheres to the “rule of three” of data backups, also called the “3-2-1 rule”:

- 3 copies of important data
- 2 different formats
- 1 off-site backup

By “important data” we mean *data which are difficult or impossible to reobtain or reproduce*. Raw data collected from fishing vessels certainly fall into this category, as do the results of especially demanding analyses.

Implementation

Data storage cluster

The data storage cluster consists of 3 dedicated servers. By “dedicated”, we mean that these machines are not part of the shared resource pool managed by DC/OS or have any other role than storage. This is for security reasons, as we want to isolate the storage cluster to the extent possible at the network infrastructure level, and also for practical reasons: We want to ensure that the data storage service operates in a highly reliable manner with as little interference from other computing tasks as possible. (This could have been achieved with DC/OS too, but there would be little to gain from doing so.)

Each storage server is equipped with several hard disks, pooled together in a software RAID configuration (ZFS) so that they appear as one. The storage capacities of the individual machines is then combined into an even larger pool using Gluster, a free and open source network file system [42].

We have organised the Gluster filesystem into three volumes with different levels of replication: In the lowest-level volume, each file is stored on only one server; at level 2, each file is duplicated on two servers; and in the level-3 volume, there is a copy of each file on all 3 servers. We call these volumes *replica1*, *replica2* and *replica3*, respectively. Table 5 summarises the properties and intended uses of the three volumes.

The *replica3* volume is regularly backed up to an offsite data centre in encrypted form.

Table 5: Properties of data volumes in the storage cluster.

Volume	File replication	Redundancy	Space efficiency	Intended use
replica1	1 copy	Low[2]	High	Temporary data and data which can be easily reobtained or reproduced. Workspaces of various services in the DC/OS cluster.
replica2	2 copies	Medium	Medium	Data which can be reobtained or reproduced, but at a relatively high cost, such as the results of time-consuming analyses and large data sets collected from external databases/services.
replica3	3 copies	High	Low	Data which are impossible to reobtain or reproduce, such as raw data from vessels.

[2] While each file only exists on one server, there is some redundancy in the RAID setup of the disks on each individual server, so that a disk may fail without causing any data loss.

Data storage accessor

The *data storage accessor* shown at the bottom-centre of Figure 9 is a special service that runs on the DC/OS cluster and acts as the single gateway through which other services can access the data storage. The accessor mounts the data volumes using the *GlusterFS* protocol. Other users and services access the data through the storage accessor using SSH or protocols based on SSH, such as SSHFS and rsync. This facilitates:

- A simpler setup on the storage servers, because they do not need to concern themselves with any advanced form of user authentication.
- A simpler setup for data-accessing services, because they do not need to be tailored to the Gluster file system.
- Easier user/service management and authorisation by having a single point of access.

References

[1] V. Mayer-Schönberger and K. Cukier, Big Data: A revolution that will transform how we live, work and think, London: John Murray, 2013.

[2] European Commission, "Horizon 2020 – Work Programme 2018–2020," Brussels, 2017.

[3] G. Granholm, K. Charvat, S. Huurinainen, P. G. Auran, J. Pukacki, C. Södergård, R. Tergujeff, J. H. Simarro, M. Angel and F. Fournier, "DataBio deliverable D6.3 – State of the Art," 2017.

[4] C. S. Plet-Hansen, E. Larsen, L. O. Mortensen, J. R. Nielsen and C. Ulrich, "Unravelling the scientific potential of high resolution fishery data," *(under review),* 2018.

[5] C. Ulrich, H. J. Olesen, H. Bergsson, J. Egekvist and K. B. Håkansson, "Discarding of cod in the Danish Fully Documented Fisheries trials," *ICES Journal of Marine Science,* 2015.

[6] G. W. Klau, N. Lesh, J. Marks and M. Mitzenmacher, "Human-guided search," *Journal of Heuristics,* vol. 16, no. 3, pp. 289-310, 2010.

[7] I. Ibarbia, A. Mendiburu, M. Santos and J. A. Lozano, "An interactive optimization approach to a real-world oceanographic campaign planning problem," *Applied Intelligence,* vol. 36, no. 3, pp. 721-734, 2012.

[8] M. H. Kang, H. R. Choi, H. S. Kim and B. J. Park, "Development of a maritime transportation planning support system for car carriers based on genetic algorithm," *Applied Intelligence,* vol. 36, no. 3, pp. 585-604, 2012.

[9] J. M. T. Palenzuela, L. G. Vilas, E. Spyrakos, L. R. Dominguez and F. Cetmar, "Routing optimization using neural networks and oceanographic models from remote sensing data," in *Proceedings of the 1st International Symposium on Fishing Vessel Energy Efficiency E-Fishing*, Vigo, Spain, 2010.

[10] R. Vettor, M. Tadros, M. Ventura and C. G. Soares, "Route planning of a fishing vessel in coastal waters with fuel consumption restraint," *Maritime Technology and Engineering,* vol. 3, pp. 167-173, 2016.

[11] J. A. Fernandes, X. Irigoien, N. Goikoetxea, J. A. Lozano, I. Inza, A. Pérez and A. Bode, "Fish recruitment prediction, using robust supervised classification methods," *Ecological Modelling,* vol. 221, no. 2, pp. 338-352, 2010.

[12] N. Trifonova, A. Kenny, D. Maxwell, D. Duplisea, J. Fernandes and A. Tucker, "Spatio-temporal Bayesian network models with latent variables for revealing trophic dynamics and functional networks in fisheries ecology," *Ecological Informatics,* vol. 30, pp. 142-158, 2015.

[13] Ø. Sture, Ø. E. Ruud, A. Skavhaug and J. R. B. Mathiassen, "A 3D machine vision system for quality grading of Atlantic salmon," *Computers and Electronics in Agriculture,* vol. 123, pp. 142-148, 2016.

[14] H. Westavik, E. R. Øye, M. S. Bondø, A. Eilertsen and M. J. R. B., "Automatisk singulering og kvalitetssortering i produksjonslinje for hel laks - Sluttrapport FHF-prosjekt 900847 (In Norwegian)," SINTEF Fiskeri og Havbruk, Trondheim, 2016.

[15] E. Guttormsen, B. Toldnes, M. S. Bondø, A. Eilertsen, J. T. Gravdahl and M. J. R. B., "A Machine Vision System for Robust Sorting of Herring Fractions," *Food and Bioprocess Technology,* vol. 9, no. 11, pp. 1893-1900, 2016.

[16] E. Misimi, E. R. Øye, Ø. Sture and M. J. R. B., "Robust classification approach for segmentation of blood defects in cod fillets based on deep convolutional neural networks and support vector machines and calculation of gripper vectors for robotic processing," *Computers and Electronics in Agriculture,* vol. 139, pp. 138-152, 2017.

[17] G. French, M. Fisher, M. Mackiewicz and C. Needle, "Convolutional Neural Networks for Counting Fish in Fisheries Surveillance Video," in *Proceedings of the Machine Vision of Animals and their Behaviour (MVAB).*, 2015.

[18] G. French, M. Fisher, M. Mackiewicz and C. Needle, "A deep learning system for analysis of fisheries surveillance video and automated monitoring of catch quota," in *17th Biennial Conference of the Challenger Society for Marine Science*, University of Liverpool, 2016.

[19] J. Uranga, H. Arrizabalaga, G. Boyra, Hernandez, G. N. M. C., I. Arregui, J. Fernandes, Y. Yurramendi and J. Santiago, "Detecting the presence-absence of bluefin tuna by automated analysis of medium-range sonars on fishing vessels," *PloS one,* vol. 12, no. 2, p. e0171382, 2017.

[20] L. Mayer, Y. Li and G. Melvin, "3D visualization for pelagic fisheries research and assessment," *ICES Journal of Marine Science,* vol. 59, no. 1, pp. 216-225, 2002.

[21] N. Gorska, R. J. Korneliussen and E. Ona, "Acoustic backscatter by schools of adult Atlantic mackerel," *ICES Journal of Marine Science,* vol. 64, p. 1145–1151, 2007.

[22] R. J. Korneliussen, Y. Heggelund, G. Macaulay, D. Patel, E. Johnsen and I. Eliassen, "Acoustic identification of marine species using a feature library," *Methods in Oceanography,* vol. 17, pp. 187-205, 2016.

[23] K. G. Foote, H. K. Knudsen, R. J. Korneliussen, N. P. E. and K. Røang, "Postprocessing system for echosounder data," *Journal of Acoustic Society of America,* vol. 90, no. 1, pp. 37-47, 1991.

[24] M. Jakobsson, L. Mayer, B. Coakley, J. A. Dowdeswell, S. Forbes, B. Fridman, H. Hodnesdal, R. Noormets, R. Pedersen, M. Rebesco, H. W. Schenke, Y. Zarayskaya, D. Accettella, A. Armstrong, R. M. Anderson, P. Bienhoff, A. Camerlenghi, I. Church, M. Edwards, J. V. Gardner, J. K. Hall, B. Hell, O. Hestvik, Y. Kristoffersen, C. Marcussen, R. Mohammad, D. Mosher, S. V. Nghiem, M. T. Pedrosa, P. G. Tavaglini and P. Weatherall, "The International Bathymetric Chart of the Arctic Ocean (IBCAO) Version 3.0," *Geophysical Research Letters,* vol. 39, no. 12, 29 June 2012.

[25] D. Hess and J. Savitz, "OCEANA Global Fishing Watch report," Global Fishing Watch, 2016.

[26] E. N. de Souza, K. Boerder, S. Matwin and B. Worm, "Improving Fishing Pattern Detection from Satellite AIS Using Data Mining and Machine Learning," *PLoS ONE,* vol. 11, no. 7, p. e0158248, 2016.

[27] A. Nielsen and C. W. Berg, "Open source fisheries management," 13 March 2014. [Online]. Available: http://www.aqua.dtu.dk/english/News/2014/03/140313_Fisheries_management_as_open_source. [Accessed 18 June 2018].

[28] O. R. Eigaard, F. Bastardie, M. Breen, G. E. Dinesen, N. T. Hintzen, P. Laffargue, L. O. Mortensen, J. R. Nielsen, H. C. Nilsson, F. G. O'Neill, H. Polet, D. G. Reid, A. Sala, M. Sköld, C.

Smith, T. K. Sørensen, O. Tully, M. Zengin and A. D. Rijnsdorp, “Estimating seabed pressure from demersal trawls, seines, and dredges based on gear design and dimensions,” *ICES Journal of Marine Science,* vol. 73, no. Supplement 1, p. i27–i43, 2016.

[29] C. Salomonsen and J. R. Mathiassen, “SMARTFISH H2020 deliverable D4.1: Report with specification of system requirements,” 2019.

[30] I. G. Aursand, H. Digre, J. Ladstein, L. T. Kyllingstad, U. G. Erikson, G. M. Tveit, C. J. Backi and K.-J. Reite, “Development and assessment of novel technologies improving the fishing operation and on board processing with respect to environmental impact and fish quality (DANTEQ),” SINTEF, Trondheim, 2015.

[31] Object Management Group, “Data Distribution Service specification version 1.4,” 2015.

[32] Object Management Group, “Who's Using DDS?,” 2017. [Online]. Available: http://portals.omg.org/dds/who-is-using-dds-2/. [Accessed 14 06 2018].

[33] Object Management Group, “Interface Definition Language specification version 4.2,” 2018.

[34] P. Bennett, “The NMEA FAQ,” 15 September 1997. [Online]. Available: https://web.archive.org/web/20140215150802/http://www.kh-gps.de/nmea.faq. [Accessed 14 06 2018].

[35] Unidata, “Network Common Data Form (NetCDF),” 2018. [Online]. Available: https://doi.org/10.5065/D6H70CW6. [Accessed 18 June 2018].

[36] Unidata, “NetCDF fact sheet,” December 2016. [Online]. Available: https://www.unidata.ucar.edu/publications/factsheets/current/factsheet_netcdf.pdf. [Accessed 18 June 2018].

[37] “rsync,” [Online]. Available: https://rsync.samba.org/. [Accessed 22 October 2019].

[38] D2iQ, Inc., “DC/OS – The definitive platform for modern apps,” [Online]. Available: https://dcos.io/. [Accessed 22 October 2019].

[39] Apache Software Foundation, “Apache Mesos,” [Online]. Available: http://mesos.apache.org/. [Accessed 22 October 2019].

[40] Mesosphere, Inc., “Marathon: A container orchestration platform for Mesos and DC/OS,” [Online]. Available: https://mesosphere.github.io/marathon/. [Accessed 22 October 2019].

[41] Apache Software Foundation, “Chronos: Fault tolerant job scheduler for Mesos,” [Online]. Available: https://mesos.github.io/chronos/. [Accessed 22 October 2019].

[42] Red Hat, Inc., “Gluster: Storage for your cloud,” [Online]. Available: https://www.gluster.org/. [Accessed 23 October 2019].

[43] R. Apfelbacher and A. Rozinat, “FMC Notation Reference,” December 2003. [Online]. Available: http://www.fmc-modeling.org/notation_reference. [Accessed 13 June 2018].

[44] E. Bachiller, J. A. Fernandes and X. Irigoien, “Improving semiautomated zooplankton classification using an internal control and different imaging devices,” *Limnology and Oceanography: Methods,* vol. 10, no. 1, pp. 1-9, 2012.

[45] A. Gislason and T. Silva, "Comparison between automated analysis of zooplankton using ZooImage and traditional methodology," *Journal of Plankton Research,* vol. 31, no. 12, pp. 1505-1516, 2009.

[46] X. Irigoien, J. A. Fernandes, P. Grosjean, K. Denis, A. Albaina and M. Santos, "Spring zooplankton distribution in the Bay of Biscay from 1998 to 2006 in relation with anchovy recruitment," *Journal of plankton research,* vol. 31, no. 1, pp. 1-17, 2009.

[47] R. Di Mauro, G. Cepeda, F. Capitanio and M. D. Viñas, "Using ZooImage automated system for the estimation of biovolume of copepods from the northern Argentine Sea," *Journal of sea research,* vol. 66, no. 2, pp. 69-75, 2011.

[48] P. Grosjean, M. Picheral, C. Waremboug and G. Gorsky, "Enumeration, measurement, and identification of net zooplankton samples using the ZOOSCAN digital imaging system," *ICES Journal of Marine Science,* vol. 61, no. 4, pp. 518-525, 2004.

A. Data collected by the SIF and BBV systems

A.1 SIF

The SIF system is divided into two levels of haul information: Overall information of haul and trip operations by the vessels and high resolution haul information with sea-packing landings data.

Table 6 and Table 7 present the data available in the SIF system.

Table 6: Overall haul and trip information in the SIF system for sea-packed landings.

Haul or trip information (Column header)	Description of provided data
Start	Start of haul or trip provided as a datetime in the format yyyy-mm-dd HH:MM
Slut (End)	End of haul or trip provided as a datetime in the format yyyy-mm-dd HH:MM
Fartøj (Vessel)	Harbour number and name of the inquired vessel
Virksomhed (Company)	Company name to which the vessel belong
Fangstområde (Catch area)	Overall catch area. Areas are not ICES standard and are very overall (e.g. Northern North Sea, Baltic Sea)
Fangst ID (Catch ID)	Either a SIF or a pas:catch code which link to the next level of information. SIF codes relate to haul level information whereas pas:catch codes relate to sales slips information in the SIF system, that is, trip level information

Table 7: Haul and trip level information with sea-packing landings data.

Haul or trip information (Column header)	Description of provided data
Fangst ID (Catch ID)	Either a SIF or a pas:catch code which link to the next level of information. SIF codes relate to haul level information whereas pas:catch codes relate to sales slips information in the SIF system, that is, trip level information
Virksomhed (Company)	Company name to which the vessel belong
Fartøj (Vessel)	Harbour number and name of the inquired vessel
eLog ID	ID link to the eLog. However this ID column has not been fully incorporated yet and no link ID to the eLog is therefore available yet.
Fangststarttid (Catch start time)	Start of haul provided as a datetime in the format yyyy-mm-dd HH:MM This is only filled in for SIF code data (haul level information)
Fangstsluttid (Catch end time)	End of haul provided as a datetime in the format yyyy-mm-dd HH:MM This is only filled in for SIF code data (haul level information)

Startposition (Start position)	GPS position of the start of the haul in the format dd° mm' ss" N/S and dd° mm' ss" E/W This is only filled in for SIF code data (haul level information)
Slutposition (End position)	GPS position of the end of the haul in the format dd° mm' ss" N/S and dd° mm' ss" E/W This is only filled in for SIF code data (haul level information)
Fangstområde (Catch area)	Overall catch area. Areas are not ICES standard and are very overall (e.g. Northern North Sea, Baltic Sea)
Redskab (Gear)	Overall gear type. Gear is not ICES standard and very overall (e.g. gillnet, bottom trawl)
Art (Species)	Danish name of the landed species as well as 3ALPHA FAO code for the species.
Videnskabeligt navn (Scientific name)	The scientific name of the landed species
Størrelse (Size)	Size class/sorting of the landed species using the size classes used by the Danish fish auctions and defined by the EU. The size class is written as characters in Danish (e.g. size class 1 will be written as First sort)
Kvalitet (Quality)	The quality of the fish. This is only filled in for pas:catch information, that is sales slips information. Quality is written as characters in Danish
Vægt (Weight)	Weight of the sea-packed size class of the species at the haul. In kg with 2-digits.

The SIF code is a unique identifier for the individual haul operation. Together with the pas:catch codes and end date and times of operations, trips can be identified and linked to the DTU AQUA DFAD (Danish Fisheries Analyses Database) dataset in order to link the SIF information to existing datasets. As the lower and upper weight limit for each size class by species is the standard from the EU regulation "COUNCIL REGULATION (EC) No 2406/96 of 26 November 1996 laying down common marketing standards for certain fishery products", weight distributions can be created directly by landing, vessel or time period. Accurate conversion factors to estimate whole live weights from gutted weights could improve interpretation and is needed for uses related to stock assessment.

A.1.1 Desired improvements in the current data system

While SIF records do provide additional data than current data acquisition systems for fishery dependent data in Denmark, extension data information for several factors would potentially benefit the fisheries management. System update to include specific trip ID and implementation of ID to directly link SIF and eLog would greatly benefit the usability of SIF.

In addition to the already collected data in SIF, an extension on data collection of fish condition but rather than limiting the fish information to landings, be it sea-packed or not, full catch accountability from either full data collection of all directly catches or full data collection of discards coupled with the data collection of landings is paramount. In an EU context, the landing obligation has made this particularly important, but as a discard ban has been in place in Norway since the 1980'es such data acquisition on the system should be highly relevant.

A.2 BBV – Black Box Video

The Black Box Video (BBV) System is designed to capture sensor, GPS and video data for the use within fisheries compliance monitoring. The system functions in an "always on" capacity, meaning that it is ready to start the capture of video when a fishing trip commences until it is completed. The system can operate and record video footage from up to 12 CCTV cameras simultaneously in addition to recording sensor data for data analysis and positional data from a GPS antenna. Acquired data is stored on an internal and external hard drive when online transmission is no available. Upon online connection acquired data is transmitted to data receiver inland. The data connection can be via GSM, Wi-Fi, 3G, 4G/LTE, LTE-A or satellite (Figure 10).

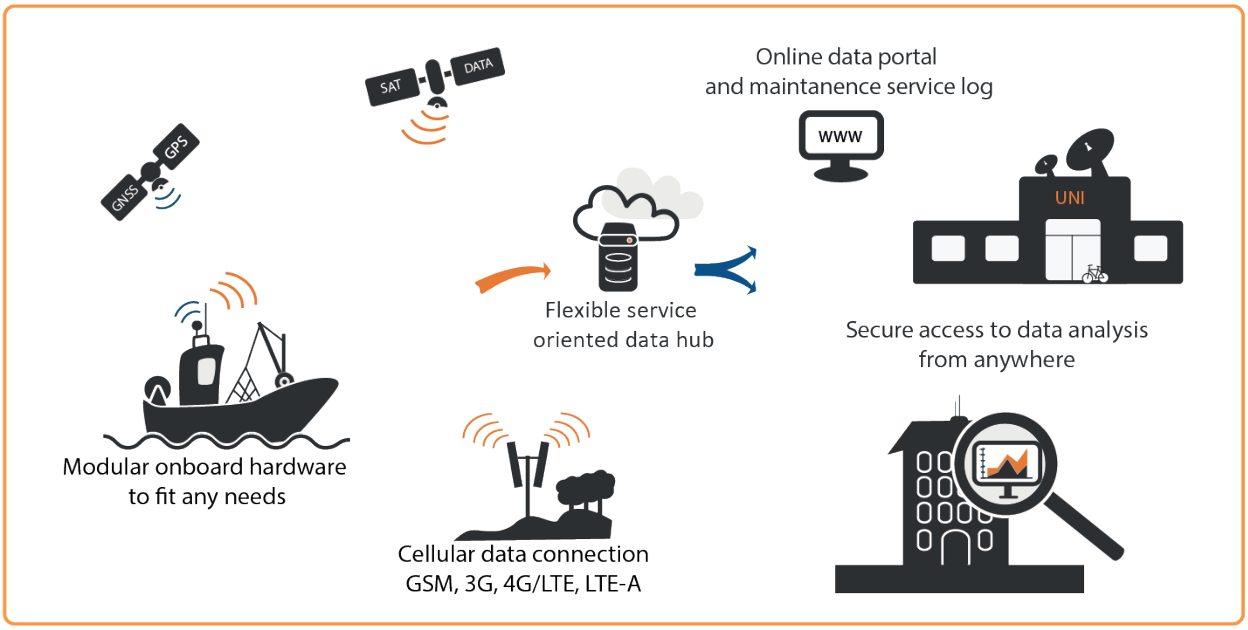

Figure 10: Black Box Video System instruments, data flow, integration and transmission to the inland centre. Courtesy of Anchor Lab K/S.

Collected sensor, GPS and video data is analysed using the Black Box Analyzer. The Black Box Analyzer provide a fleet overview and GIS layer for spatial tracking and allow for live view, geo-fencing alarms and synchronised video playback. Additionally, using photogrammetry, the length of objects on a plane of known size can be extracted. Within the square lens camera video footage the photogrammetry technique then allow for adding a grid overlay to the analysed video footage. The grid overlay can be set at custom length intervals and a measuring line can be added with an accuracy of ±0.5 cm (Figure 3). Measurements can be recorded with a specific length-weight relationship per species, allowing for video audit recorded the species, length and estimated weight of each assessed species.

A current project in Australia is ongoing for a low power version of the system, enabling the system to be configured to different vessel types. The low power version is a modular system design which enables a tailoring of the BBV system to vessel specifics. Examples: Larger vessels like trawlers could have a system consisting of the control unit mounted in the wheelhouse connect to mast mounted antennas. The system could carry cameras overlooking catch processing and discard areas and include winch mounted sensors to identify fishing activities in the Black Box Analyzer. Smaller vessels like gillnetters could have a system with a control unit using a battery pack mounted on-deck. The system could use one camera mounted on a orientable mounting arm could view the setting and hauling of nets in order to identify fishing activities and another camera viewing the catch in the hauled nets.

A.2.1 Data currently collected by the Black Box systems

The Black Box Video system collect sensor, GPS and video data which is analyzed using the Black Box Analyzer. Identification of hauls and catch processing's is performed by the software using the gear sensor records with human verification. Photogrammetry enable a grid overlay and a measuring line with an accuracy of ±0.5 cm to be added to the analyzed video footage. A species-specific length-weight relationship allow for video audit records of the species, length and estimated weight. Video audit is thus performed with human species recognition, computer assisted grid overlay, and measuring line to assess discards lengths and weights by species.

A.3 Onshore grading machine data

Fish grading machines are conveyor belt systems that weigh every fish fed into them, assign each one to a weight class and pass the fish into boxes for auction. Grading machine data are provided in two files for each time period. One file provides weight class data on fish (referred to as weighing or piece data) and the other file provides information about the packed boxes (referred to as packing data). The format of these two files is shown in Table 8 and Table 9.

Table 8: Example of packing data format from a grading machine.

Packing data variable (Column header)	Description of provided data
Number	Cumulative label number
Registration time	Date and time label was printed
Pallet Id	Empty unassigned column
Record type	Type of measurement (always ‘Pack’)
Device	Grader machine responsible for weighing
PO	Vessel name abbreviation with concatenated date and occurrence
Lot	Description of capture
Shift	Description of fish (always ‘Gutted’)
Product	Species name and a number to identify packing occurrence for this species
Record status	Empty unassigned column
Weight (kg)	Weight after correction
Production day	Date label was printed
Nominal weight (kg)	Weight after correction
Tare weight (kg)	Unladen weight
Layout	Record description (always ‘Pack Label’)
Gross weight (kg)	Scale weight not subtracting Tare weight
Target weight (kg)	Pack target weight
Nominal unit	Weight measurement units (always ‘Kilos’)
Status	Box status (always ‘closed’)

Table 9: Example of weighing data format from a grading machine.

Weighing data variable (Column header)	Description of provided data
Material	Species name and a number to identify weighing occurrence for this species
Class	Class, a weight category
Pieces	Pieces, number of fish in weight category
Value	Weight of this particular species/category combination
Value type	Description of measurement value (always ‘Weight’)
Value Sqr	Square of value field

PO	Empty unassigned column
Batch	Empty unassigned column
Shift	Empty unassigned column
Log time	Time and date of weighing
Record type	Description of record (always 'Physical')

B. Data structures

The following are examples of IDL data structures that represent various data collected on fishing vessels, corresponding to the entries in Table 4. Here, we define a *measurement* type as a combination of a *sensor* identifier and a *value*. The former is a universally unique identifier (UUID) that identifies a particular sensor. This is to distinguish it from other sensors of the same type, and to be able to relate different measurements from the same sensor (possibly in different topics).

```
// Sensor ID
typedef sequence<octet, 16> UUID;


// Value types
struct GlobalPosition
{
    double latitude;
    double longitude;
};
enum HeadingReference
{
    TRUE,
    MAGNETIC
};
struct Heading
{
    double degrees;
    HeadingReference reference;
};
struct ElectricalPower
{
    double power;
    double voltage;
};
enum FishSpecies
{
    COD,
    HADDOCK,
    SAITHE,
    // etc.
};
struct IndividualFishStats
{
    FishSpecies species;
    double length;
    double weight;
    sequence<octet> thumbnail;
};


// Measurements
struct GlobalPositionMeasurement
{
    UUID sensor;
    GlobalPosition value;
};
```

```
struct VelocityMeasurement
{
    UUID sensor;
    double value;
};
struct HeadingMeasurement
{
    UUID sensor;
    Heading value;
};
struct ElectricalPowerMeasurement
{
    UUID sensor;
    ElectricalPower value;
};
struct IndividualFishStatsMeasurement
{
    UUID sensor;
    IndividualFishStats value;
};
```

C. Fundamental Modeling Concepts

In this document we have used Fundamental Modeling Concepts (FMC) to model system structures as block diagrams. Figure 11 shows a summary of the notation for quick reference. For more information about FMC, see the FMC web site (f-m-c.org).

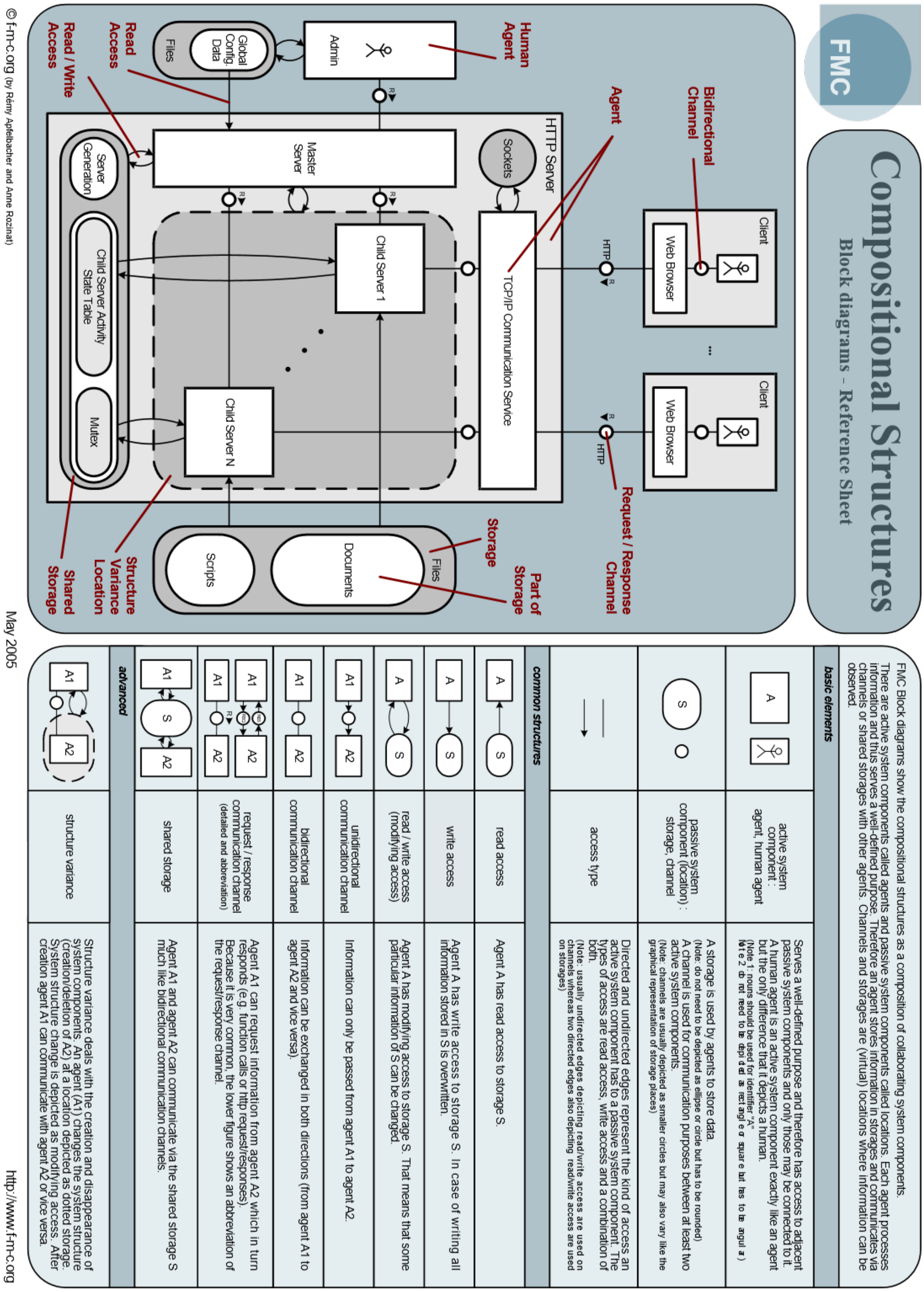

Figure 11: FMC block diagram reference sheet. [43]