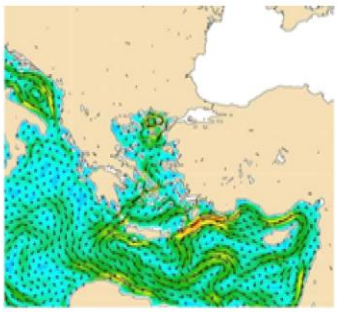
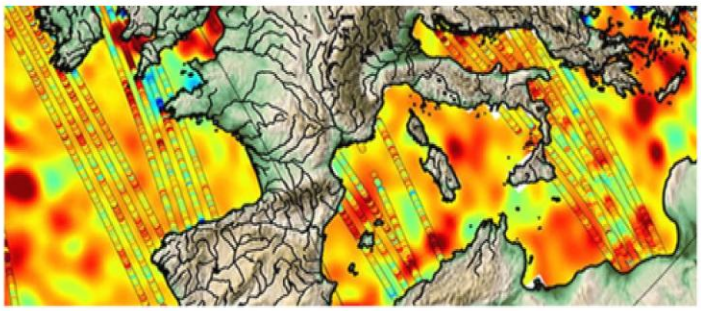
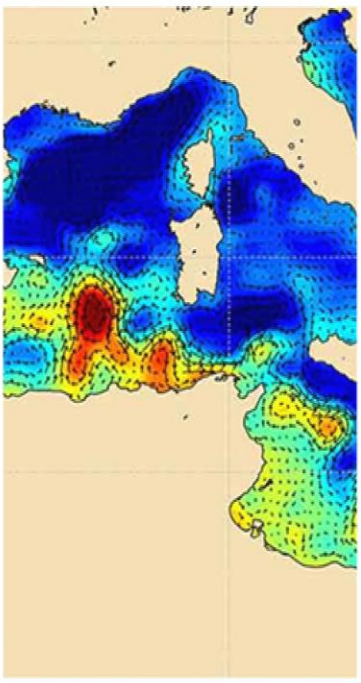
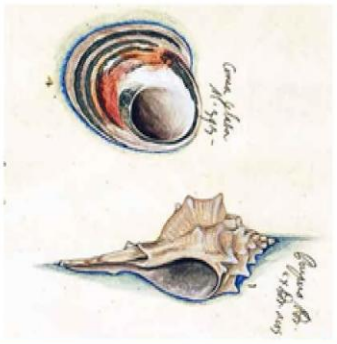








EMODnet MedSea Checkpoint
Data Adequacy Report

EMODnet
European Marine Observation and Data Network

MedSea Checkpoint

http://www.emodnet-mediterranean.eu/

GROWTH AND INNOVATION IN OCEAN ECONOMY – GAPS AND PRIORITIES IN SEA BASIN OBSERVATION AND DATA

First Data Adequacy Report
EMODNET MedSea CheckPoint

Total number of pages: 52

Workpackage	**11**	11.2 Final version of the First Data Adequacy Report	
Author(s):	G.M.R. Manzella, Nadia Pinardi, Antonio Guarnieri, Michela De Dominicis		INGV
	Eric Moussat, Erwann Quimbert, Julien Meillon		IFREMER
	Frederique Blanc, Claverie Vincent		CLS
	George Kallos		UOA
	Chara Kyriakid		HCMR
	Lluís Gómez-Pujol		SOCIB
	Giuseppe Scarcella		ISMAR
	Antonio Cruzado		OCEANS-CAT
	Federico Falcini		ISAC

A project funded by:

EUROPEAN COMMISSION, DIRECTORATE-GENERAL FOR MARITIME AFFAIRS AND FISHERIES, MARITIME POLICY ATLANTIC, OUTERMOST REGIONS AND ARCTIC

European Commission

Table of Contents

Glossary

BODC: British Oceanographic Data Centre

CFP: Common Fisheries Policy

Chl: Chlorophyll

CLS: Collecte Localisation Satellites (FR)

CLU: CLU s.r.l. (IT)

CMCC: Euro-Mediterranean Centre for Climate Change (IT)

CNR: National Research Council (IT)

Copernicus: European Programme for the establishment of a European capacity for Earth Observation

CSW: Catalogue Service for the Web

CYCOFOS: Cyprus Coastal Ocean Forecasting and Observing System

DAC: Data Assembly Center

DAR: Data Adequacy Report

DCR: Data Collection Regulation

DCF: Data Collection Framework

DG-MARE: Directorate-General for Maritime Affairs and Fisheries

EC: European Commission

ECMWF: European Centre for Medium-Range Weather Forecasts

ECV: Essential Climate Variables

EDF-EN: EDF Energies Nouvelles (FR)

EDMED: European Directory of Marine Environmental Data

EEA: European Environmental Agency

EEC: European Economic Community

EEZs: Exclusive Economic Zones

EIONet: European Environment Information and Observation Network

EMODnet: European Marine Observation and Data Network

EMSA: European Maritime Safety Agency

ESA: European Space Agency

EU: European Union

EUMETNET: European National Meteorological Services

EUNIS: European Nature Information System

EUROGOOS: European Global Ocean Observing System

FAO: Food and Agriculture Organization

FP7: Seventh Framework Programme

GEBCO: General Bathymetric Chart of the Oceans

GES: Good Environmental Status

GEO: Group on Earth Observation

Geoportal: type of web portal used to find and access geographical information

GEOSS: Global Earth Observation System of Systems

GIS: Geographic information system

GMES: Global Monitoring for Environment and Security

GNOO: National Group for Operational Oceanography

GOOS: Global Ocean Observing System

HCMR: Hellenic Centre for Marine Research (GR)

HO: Hydrostatic Office

ICES: International Council for the Exploration of the Sea

ICCAT: International Commission for the Conservation of Atlantic Tunas

ICZM: Integrated Coastal Zone Management

IEO: Instituto Español de Oceanografía

IFREMER: Institut Français de Recherche pour l'Exploitation de la Mer (FR)

IH-Cantabria: Fundación Instituto de Hidráulica Ambiental de Cantabria (ES)

IHO: International Hydrographic Organization

IMEDEA: Mediterranean Advanced Studies Institute

IMO: International Maritime Organization

INGV: National Institute of Geophysics and Volcanology (IT)

INSPIRE: Infrastructure for Spatial Information in the European Community

IOC: Intergovernmental Oceanographic Commission

IPCC: Intergovernmental Panel on Climate Change

ISAC: Institute of Atmospheric Sciences and Climate

ISCOMAR: Isleña Marítima de Contenedores

ISO: International Organization for Standardization

ISPRA: Italian National Protection Agency

JCOMM: Joint WMO-IOC Commission on Marine Meteorology

JRC: Joint Research Centre

MAP: Mediterranean Action Plan

MS: Member States

MSFD: Marine Strategy Framework Directive

MSP: Maritime Spatial Planning

MSSD: Mediterranean Strategy for Sustainable Development

NRT: Near Real Time

NKUA: National and Kapodistrian University of Athens

OSSE: Observing System Simulation Experiments

OSE: Observing System Experiment

P01: BODC Parameter Usage Vocabulary

P02: SeaDataNet Parameter Discovery Vocabulary

P03: SeaDataNet Agreed Parameter Groups

SHOM: Service hydrographique et océanographique de la marine

SOCIB: Balearic Islands Coastal Observing and Forecasting System (ES)

UCY: University of Cyprus (CY)

UN: United Nations

UNCLOS: United Nations Convention on the Law of the Sea

UNEP: United Nations Environment Programme

UNESDO: United Nations Educational, Scientific and Cultural Organization

VLIZ: Flanders Marine Institute

WISE: Water Information System for Europe

WFD: Water Framework Directive

WMO: World Meteorological Organisation

WMOP: SOCIB Western Mediterranean Sea Operational forecasting system

WWF: World Wildlife Fund

Executive Summary

EMODnet MedSea CheckPoint aims to document the quality assessment of the existing monitoring system at the sea basin level, by developing fitness-for-use indicators to show the appropriateness and availability of monitoring data for the production of Challenge targeted products. There are seven Challenges: Wind Farm Siting (CH1), Marine Protected Areas (CH2), Oil spill Platform Leaks (CH3), Climate and Coastal Protection (CH4), Fisheries (CH5), Marine Environment (CH6) and Rivers (CH7). The assessment will help identify gaps and prioritize the needs in order to optimize the system throughout the value chain (i.e. data collection, in situ and satellite data assembling, data management and networking, modelling and forecasting, geo-infrastructure) and release recommendations for future developments to better meet the application requirements.

This first Data Adequacy Report reviews the methodology used in the MedSea CheckPoint to construct the metadatabase, and develop the indicators for a selected number of assessment criteria. The indicators were defined on the basis of existing ISO standards and they are constructed directly on the metadatabase containing information on the upstream data sources (monitoring datasets). For each Challenge, CheckPoint Information on What, Why, Where, When, How data will be used to develop targeted products. The information is organised into a metadatabase which currently contains **298 data source descriptors**.

On the basis of this metadatabase, the Literature Survey results were reviewed. It was confirmed that **45 Characteristics** (variables derived from observations or models and/or the geographical representation of an object) are needed for the seven Challenges and **126 data providers** would be required by the Challenges in order to develop the targeted products. This already highlights the importance of the MedSea CheckPoint portal for collecting and organizing the information on complex and distributed data source networks that are required to derive the Challenge products. GIS technology is used to organize all this information into a metadatabase and make it available to DGMARE and the public.

The assessment criteria are subdivided into two territories: appropriateness and availability. Only availability is described in the first DAR since most of the Challenge products have not yet been developed and thus appropriateness cannot be defined properly. The availability investigates “how the input data sets are made available to the Challenge use” and **8 indicators** have been developed from the metadatabase. From this first analysis and without differentiating between the Challenges, only four indicators were assessed from which the following emerged: 1) the majority of the data

sets are accessible through an online delivery service (not always fully Inspire compliant), 2) the data policy is partially restricted; 3) most of the data are free of charge, and 4) the responsiveness in terms of data delivery time is generally high. These results will be modified when a thematic or single Challenge analysis is carried out. For example, CH1 (Wind farming) data are available but have to be purchased, and CH5 (Fisheries) data responsiveness is low (more than a week to access the data).

For CH3 (Oil Spill Leak Platform) all the eight availability indicators have been estimated. The results show that the indicators have mostly positive and very positive values except for data policy visibility and the data policy itself which is limited by moratorium and/or specific agreements.

In conclusion, this first DAR highlights that it is possible to develop objective indicators of fitness for use for the input data sets to the Challenges. The next step will be to discuss thematic indicators together with Challenge indicators and enlarge the indicator set to the appropriateness territory.

1 Introduction

CheckPoint main objectives

The EMODnet CheckPoints, a concept first conceived in the EC Marine Knowledge 2020 Communication, were set up to assess the quality and fitness for use of basin-scale monitoring systems in support of targeted applications, also called 'Challenges', i.e.: CH1- Windfarm Siting, CH2- Marine Protected Areas, CH3- Oil Platform Leak, CH4- Climate and Coastal Protection, CH5- Fisheries Management, CH6- Marine Environment, CH7- River Inputs.

The primary aims of CheckPoints are to:

1) document the appropriateness and availability of the existing monitoring system at the sea basin level;
2) develop fitness-for-use indicators to show the performance, accessibility and usability of monitoring data for the production of Challenge targeted products;
3) identify gaps and prioritize the needs in order to optimize the system throughout the value chain (i.e. data collection, in situ and satellite data assembling, data management and networking, modelling and forecasting, geo-infrastructure) and release recommendations for future developments to better meet the application requirements.

Data Adequacy Reports

Two Data Adequacy Reports (DARs) need to be produced by the MedSea CheckPoint project, containing the major assessment findings. This report documents the first DAR, building on the definitions and the methodology described in the literature survey [Ref1][1].

'Data' definition

In this report 'data' is defined as a 'reinterpretable representation of information in a formalized manner suitable for communication, interpretation or processing' (ISO 19115). The ISO 19157 standards and additional criteria were used to define the metadata specification for the CheckPoint data inventory.

Data Adequacy concept

Data adequacy can be defined as the fitness for use of the data for a particular user or for a variety of users. Since different applications require different properties associated with the data itself, 'adequacy' should be defined objectively using standardized nomenclature and methods.

[1] Ref1: Literature Survey available at https://webgate.ec.europa.eu/maritimeforum/node/3646

In an EC Report [Ref2][2] 'adequacy' was intended as an assessment of reported information to meet the objectives of the Marine Strategy Framework Directive (MSFD) and its technical requirements listed in MSFD Articles 8, 9 and 10 [Ref3] [3].

Adequacy can be defined as 'sufficient to satisfy a requirement or meet a need' [Ref 4] [4]. From this definition, 'adequacy' relates to meeting both requirements as well as needs and is normally applied within the framework of an ISO 9001 based Quality Management System.

ISO standard used in MedSea CheckPoint

ISO/TC 211 (International Organisation for Standardisation / Technical Committee No. 211) provides a series of standards that deal with various aspects of geographical information / geomatics, which include ISO 19157 Data Quality and ISO 19115:2003 Metadata. ISO 19157 standards introduce a new element for assessing 'how' and 'how much' data meets requirements in order to enhance user satisfaction.

Universe of Discourse

In this report, both the data and literature are components of the 'Universe of Discourse' defined as a 'view of the real or hypothetical world that includes everything of interest' (ISO 19101). The concept behind the universe of discourse is shown in Figure 1.1.

The quality standard principles not only provide the concept of data quality, but also standardize the names and schemes under which all dataset differences and the corresponding universe of discourse can be categorized. The elements and sub-elements, as described in ISO 19113, are discussed in the Literature Survey, and will not be presented again here.

[2] Ref2: The first phase of implementation of the Marine Strategy Framework Directive (2008/56/EC) - The European Commission's assessment and guidance. CELEX_52014SC0049_EN_TXT

[3] Ref3: Adequacy does not necessarily mean, for instance, that if the defined data is adequate, this automatically means that the quality of the marine waters Is acceptable

[4] Ref4: Random House Unabridged Dictionary, Random House Inc, 2006

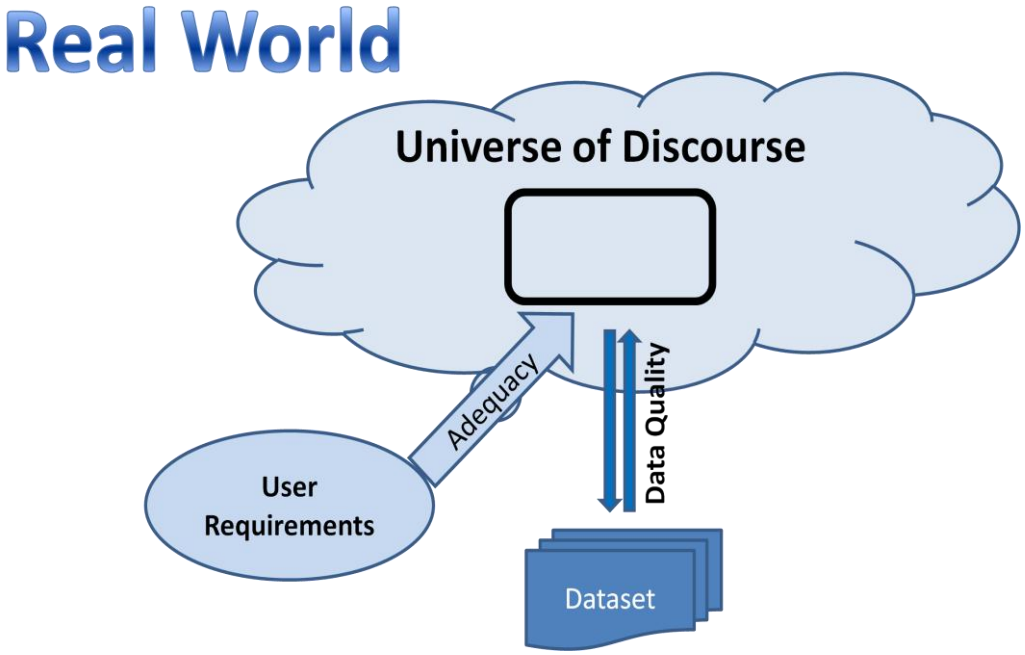


Figure 1.1. Concept behind data quality principles as described in ISO 19113 [Ref 5] [5]

1.1 Structure of the document

Structure of the document

The report is subdivided in three parts:

1) the first (§ 2 - 4) reviews the CheckPoint methodology and metadata definitions and presents the updated Literature Survey results.
2) the second part describes the indicator definitions (§5-6);
3) the third part shows the fitness for use assessment using only the availability assessment indicators (§7-8-9).

Discussion and conclusions (§10) conclude the document.

[5] Ref5: http://inspire.ec.europa.eu/glossary/UniverseOfDiscourse

2 CheckPoint assessment methodology

CheckPoint Information and CheckPoint Service

The overall aim of EMODnet CheckPoints is to assess the fitness for use or adequacy of the existing monitoring system at the sea basin level in terms of the Challenge targeted products. This involves the development of CheckPoint Information on upstream data and a CheckPoint Service to perform the assessment and make it available. In the Literature Survey, the basic assessment methodology was specified as the:

- Establishment of a framework for collection of information related to input data sets required by the Challenges. The end product of this framework is the production of a metadatabase or CheckPoint information database;
- Definition of objective assessment criteria for the production of 'adequacy' indicators;
- Analysis of the fitness for use of the input datasets with respect to specific Challenge targeted products.

Assessment Territories

The assessment criteria are subdivided into two 'Territories' that need to be evaluated in terms of Challenge requirements. They are:

Territory 1: Appropriateness
What is made available to the Challenges and what the inherent properties of such data sets are
Territory 2: Availability
How the input data sets are made available to users

The 'fitness for use' or 'adequacy' is established with indicators based on these two assessment criteria. Territory criteria provide the degree of conformity of data to the Challenge requirements and needs. Table 2.1 lists the assessment elements of the two Territories.

Territory 1: Appropriateness	Territory 2: Availability
○ Spatial information	✓ Visibility
▪ Extent	✓ Accessibility
▪ Resolution	○ Service
○ Time information	○ Data policy
▪ Extent	○ Pricing policy
▪ Resolution	○ Formats
○ Purpose	○ Service interoperability
○ Lineage	✓ Performance
○ Usage	○ Responsiveness
○ Completeness	○ Reliability
○ (logical) Consistency	
○ Accuracy	
▪ Horizontal	
▪ Vertical	
▪ Temporal	
▪ Thematic	

Table 2.1 Appropriateness and Availability assessment criteria

Data Adequacy in EMODnet MedSea CheckPoint will be assessed in two tiers. The first DAR will analyze the 'Availability' territory for each Challenge input data sets. Here each of the Challenges is considered as an 'autonomous universe of discourse', with its own specific input data. Adequacy assessments will be based on the transformation of the input metadata into an ordered set of indicators.

First version of DAR: Availability Territory

The second report will consider the links between the different Challenges and use indicators for the ‘appropriateness’ territory.

The second DAR

3 The CheckPoint metadatabase

As part of the Literature Survey, metadata were collected and analyzed for upstream data classification, an iterative process that was consolidated for the first DAR.

The Literature Survey introduced the classification of upstream data for the Challenges in terms of “characteristics” and “environmental matrices”. A “characteristic” is a distinguishing feature which refers:

Characteristic definition

- either to a variable derived from the observation, the measurement or the

numerical model output of a phenomenon or of an object property in the environment

- or to the geographical representation of an object on a map (i.e. a layer such as a protected area, a coastline or wreck) by a set of vectors (polygon, curve, point) or a raster (a spatial data model that defines space as an array of equally sized cells such as a grid or an image).

The concept of environmental matrices is introduced to avoid ambiguities when using a characteristic name such as “temperature”.

Environmental Matrix

The environment matrix is the environment to which a characteristic is defined and its elements are:

1. Air
2. Marine Waters
3. Fresh Waters
4. Biota/Biology
5. Seabed
6. Human activities

The CheckPoint database contains the metadata that link the different sources of upstream data to the characteristics required by the Challenges. The Metadata format and model is explained in Annex 1.

The Consortium decided to use the GIS technology called ‘Sextant’[Ref 6][6] to store the information on the upstream data sets and the targeted products for the Challenges, when they are available. Sextant provides access to various geographical data via web services using standards defined by the Open Geospatial Consortium (OGC) and the ISO Technical Committee ISO/TC 211, Geographic information/Geomatics. Sextant uses Geonetwork to set up the Catalogue Services for the Web and is used by several EU projects such as EMODnet, MyOcean, SeaDataNet.

Sextant catalogue and GIS Technology

For the purpose of the EMODnet Medsea CheckPoint, Sextant is used to describe the upstream data and to handle the information needed for the establishment of the fitness for use indicators. In order to classify the data sources and providers, the Sextant monitoring data catalogue will use the SeaDataNet Common Vocabularies

Sextant catalogue

[6] Ref6: http://www.ifremer.fr/sextant

(http://www.seadatanet.org/), the European Directory of Marine Organisations (EDMO), the European Directory of Marine Environmental Research Projects (EDMERP).

CheckPoint information for the metadatabase

The CheckPoint metadatabase contains information describing input data to the Challenges that are uniquely identified as a combination of characteristics, dataset and use, i.e.:

1. Characteristics (= What)
2. Data sources (= From)
3. Overview (= Why for)
4. Spatial coverage (= Where)
5. Time coverage (= When)
6. Accessibility (= How)

These descriptors make up the elements of the territory assessment criteria previously described in Section 2.

The metadatabase is enabled by a dedicated CSW technology based on an ISO XML schema embedding:

- ISO 19115/139 for contextual metadata – classical reference, for catalogue
- ISO 19157 for CheckPoint indicators – such as fitness for purpose and "used by" any application (use cases).

In the Literature Survey the following ISO standards were analyzed to construct the metadata base which helps in constructing 'adequacy' indicators:

1. ISO/NP (New Proposal) 19157 Geographic information -- Data quality
2. ISO/NP TS (Technical Specification) 19158 Geographic information - Quality assurance of data supply

CheckPoint Service

The metadatabase is built with a well-defined process to edit, validate and consolidate its content. The metadatabase will be visible and accessible through three interfaces:

- **CheckPoint Browser –**
  A public access function to search and explore the input datasets and the Challenge products;
- **CheckPoint GIS –**

A function under public and restricted access to visualize and manage CheckPoint input datasets and the Challenge products;

- **CheckPoint Dashboard –**

  A function under restricted access to compute and visualise indicators, directly built from metadatabase descriptors.

3.1 Vocabularies

Data curation and interoperability is possible only if common vocabularies are used. The same variable sometimes has different names depending on the repositories or the applications. Common vocabularies have been developed in many international initiatives, such as GEMET (promoted by INSPIRE as a multilingual thesaurus), UNIDATA, SeaDataNet, and the Marine Metadata Initiative (MMI).

SeaDataNet Vocabulary

When a vocabulary is formally managed, it becomes a controlled vocabulary. In this case, "managed" means the terms are stored and maintained using agreed-upon procedures. Procedures should exist for adding terms, modifying terms and, more rarely, rejecting terms from a controlled vocabulary.

The SeaDataNet is a controlled vocabulary containing terms that are:

1. Accepted: Each term adheres to community practices.
2. Defined: The terms are characterized precisely. Typically, this means the terms have rigorous definitions.
3. Managed: There is a body of experts that create and maintain the controlled vocabulary. The controlled vocabulary maintenance involves periodic review, addition of new terms, modification of terms, and occasionally the deprecation of terms.

SeaDataNet P01, P02, P03 vocabulary

MedSea CheckPoint decided to use the SeaDataNet Vocabulary that adopts a hierarchical approach for the classification of terms, from disciplines (P03), to parameter discovery (P02), to parameter usage (P01). This hierarchy goes from a coarser to a finer classification of a given dataset.

4 The updated literature survey results

Here we present some of the statistics obtained from the consolidated metadatabase, validated after the Literature Survey. The latter has already presented results related to input data sources for the Challenges, however the metadatabase had not been built at the time, and several iterations took place afterwards to consolidate the content and find mistakes, a process that is still on-going and will continue until the second DAR.

At present, there are **298** data sets in the MedSea CheckPoint metadatabase**.** Figure 4.1 shows the number of data sets required by each Challenge.

298 data sets required by Challenges

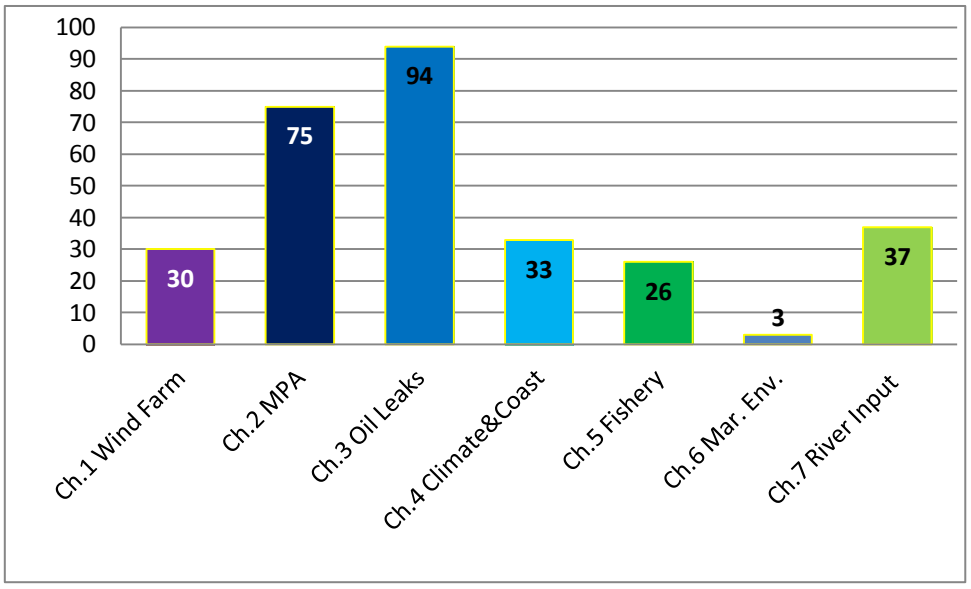


Figure 4.1 Number of data sets for each Challenge (total is 298)

The Challenge characteristics consider three kinds of upstream data:

- variables that result from the observation of a property of an object or of a phenomenon at a given place and time;
- gridded outputs from numerical models in given areas and times;
- map layers representing the geographical features (position and geometry) of natural or man-made objects such as extent of land use, protected areas, coastline or locations of wrecks defined by a set of vectors (polygon, curve, point) or by a raster.

To date there are **45** characteristics considered for all CheckPoint Challenges.

45 overall characteristics

Figure 4.2 shows the number of characteristics required by each challenge as a function of the environmental matrix .

The list of the **298** input data sets, associated with the **45** characteristics is provided in Annex 2.

Characteristics by Challenge and environmental matrix

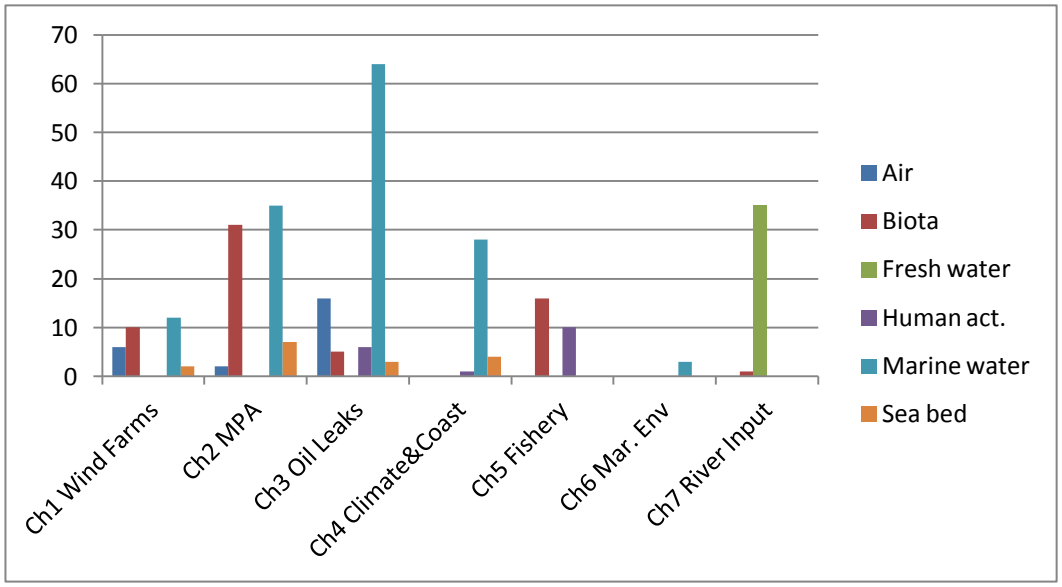


Figure 4.2 Number of characteristics by challenge and matrix

The **298** data sets are associated with **126** data providers or sources that were selected by the project experts to be the desirable upstream data inputs to the Challenges. The term data provider is used here both as a single data "producer" and a "project consortium" producer (i.e. MyOcean, Medess4MS). This creates some confusion or at least entails checking whether the accessibility evaluation is related to the producer of data or to the project providing data. The next release of Sextant will offer the possibility of adding the producer to avoid confusion.

126 data providers

It should be noted that the same characteristic is required by the different Challenges from different data providers. The list of **126** potential data sources is provided in Annex 3. Table 4.1 shows the categories of potential data sources, subdivided into six, broad categories.

Data sources broad categories

EU project, initiatives, ..	Environ. Bodies	National Data Centres	Academia, Research Centres	Space and Meteo Agencies	UN bodies and MAP RACs
21	38	7	42	10	8

Table 4.1. Number of data sources (total 126) for broad categories (details are in Annex 3)

5 CheckPoint indicator definitions

CheckPoint indicator definition

We define 'fitness for use' criteria as the weighted sum of Territory 1 and Territory 2 assessment elements which are considered the most relevant for the use by the Challenge. The degree of fitness for use of a dataset is represented by a series of indicators which are the result of the comparison of the value of the selected criteria with the user requirements (eg the horizontal accuracy of the data set either described in the producer metadata or measured by the user himself).

The present DAR concentrates only on the 'availability' indicators since as yet we do not have the Challenge products and thus it is impossible to extract the 'appropriateness' indicators. In the second DAR, Challenge targeted products will be available and final fitness for use indicators will be developed.
In order to provide an objective evaluation method, a ranking system needs to be developed for each assessment criteria. The indicator methodology is illustrated in Figure 5.1.

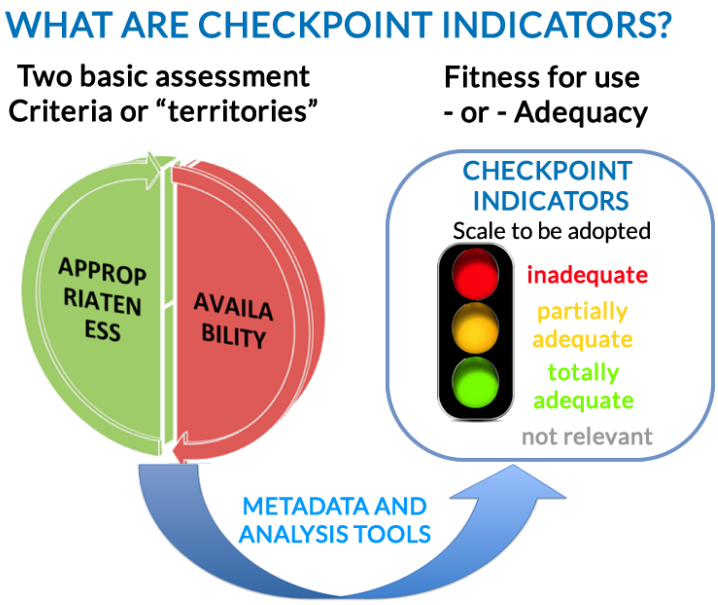


Figure 5.1 The two territory assessment criteria and the indicators

5.1 The SMART methodology

In the DAR, the availability indicators will be built using the ISO19157 methodology and according to the principles of 'SMART' methodology outlined in table 5.1.

SMART method for indicators definition

Letter	Value	Comment
S	Significant – Specific	The indicator should be significant but specific. Details for a quick understanding of the indicator should be provided so that there is no ambiguity as to what exactly the indicator stresses .
M	Measurable – Meaningful	Indicators should be measurable and quantifiable. Indicators should also measure progress, so the change can be seen as it occurs. A measurable goal for an indicator is that it can be assessed either on a sliding scale (1-10), or as a success or failure.
A	Achievable - Attainable	An achievable goal for an indicator means that it is valid given the current social, economic, or cultural resources and time available. It is action-oriented, easily attributed for the purpose of the assessment of marine data at the scale and perimeter of the MedSea CheckPoint project.
R	Realistic - Reliable	All indicators should be realistic and reliable, that is: • Feasible: • Explained (i.e. the reason for) • Reproducible: Because monitoring characteristics are changing and improving every day, it is important to envisage as much automatic computing as possible. Some CheckPoint indicator computations will be integrated or connected with the Web-GIS tools in order to dynamically update the CheckPoint information.
T	Time-Bound – Temporal aspects defined	Setting a timeframe for the indicator goals enables assessments to be tracked and adjusted over time. Temporal aspects will cover: • First date of value • Update frequency • Action list: working methods and how to improve the indicator value

Table 5.1: The principles of 'SMART' methodology

Indicators provide both an overview of the situation at a high level of aggregation as well as detailed information about trends and links. The challenge (in the indicator definition) is to find an appropriate balance between simplification and completeness. Indicators offer a unique and objective way to assess a problem

without accessing directly all the metadata.

The indicators are built from descriptors in several stages. In this DAR we build the indicators from the raw descriptors without prior aggregation of the information by the Challenges, releasing a neutral and basic status of indicators. The indicators of the first DAR will be time-bound and will be updated every six months up to the second DAR.
In the next section we describe the indicators for the availability criteria.

6 Availability indicators

In MedSea CheckPoint 'availability' measures the degree to which datasets are ready for use and obtainable. To obtain datasets, information is needed on the sources (visibility), how to access them (accessibility), and how fast the process is to take possession of them (performance).

Availability indicators:

The availability indicators (AV) provide an understanding of the readiness and service performance of the infrastructure providing access to data. The availability indicators are divided into three parts:

Visibility

1) Visibility (VI), i.e. the possibility of identifying and quickly accessing the appropriate site for the required data sets;

Accessibility

2) Accessibility (AC) i.e. the possibility, for non expert users, to understand the retrieval model status;

Performance

3) Performance (PE) i.e. the ability of a system to keep operating over time and to meet real time operational conditions. This is related to service performance.

6.1 Visibility indicators

"Visibility" is the ability to identify and quickly access the appropriate site delivering the desired data sets. In other words it is the ability for all users, including non-experts, to perform data sourcing through an EU Inspire catalogue.

Two indicators have been defined for the visibility element, i.e.:

AV-VI-1 **Easily found**	Can the data sets or series of data sets be found easily?
AV-VI-2 **EU Inspire catalogue service**	Is the dataset referenced by a EU catalogue service or other bodies (private or public, national or international non EU services[7](Ref.7))

The methodology used to compute the visibility indicators for the first DAR is explained in Table 6.1.

Indicator name	Meaning (INSPIRE)	Meaningful (Symbol)	Achievable & Realistic (Choice)
AV-VI-1 **Easily found**	The indicator comments on the visibility of data in catalogues. = Part B 1.4 of INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119	Low visibility	Choice 1: Red *"Cited in peer reviewed paper or grey literature but no info on how to access"* Choice 2 : Red *"Information retrieved upon specific request to the data source "*
		Medium visibility	Choice 3: Yellow *"Use of social network, community of practices sharing information, portals of organization where no search is organized by an engine"*
		High visibility	Choice 4: Green *"Use of open search engines, searching by name either the data source or the characteristics"* Choice 5: Green *"Search via reference catalogue (e.g. MyOcean, GEOSS Geoportal…)"*

Availability – Visibility Indicators

[7] Note1: advanced services, but not following the guidelines defined in INSPIRE and its technical annexes.

AV-VI-2 EU Inspire catalogue service	The characteristics can be searched by a catalogue service (e.g. the EMODNET thematic data portals, the Copernicus core services, the INSPIRE Geoportal, EEA services, DG MARE services, MonGOOS services). Part B 1.4 of Inspire Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119	 Inadequate  Partly adequate  totally adequate	Choice 1: Red *"Data sets are not referenced in a catalogue or are referenced in a non public catalogue"* Choice 2: Yellow *"The datasets are referenced in a public national catalogue, in an international catalogue service "* Choice 3: Green *"The datasets provide a full EU Inspire catalogue service "*

Table 6.1 Definitions of Visibility Indicator Values

6.2 Accessibility indicators

Accessibility is the ability for all users, including non-experts, to understand the retrieval model status and its appropriateness.

There are five indicators devised for accessibility:

AV-AC-1 Policy visibility	Visibility on data policy adopted by data providers.
AV-AC-2 Delivery	Data delivery mechanisms, i.e. the services available to the user to access data
AV-AC-3 Data Policy	Data policy
AV-AC-4 Pricing	Cost basis / price policy
AV-AC-5 Readiness	Format for use

The methodology used to compute the accessibility indicators for the first DAR is explained in Table 6.2

Indicator name	Meaning (INSPIRE)	Meaningful (Symbol)	Achievable & Realistic (Choice)
AV-AC-1 Policy visibility	The indicator comments on the information given about the data policy adopted by data providers Part B 8.1 and 8.2 of INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119	 Low transparency  Medium transparency  High transparency	Choice 1: Red *"There is no information at all on data policy adopted by data providers"* Choice 2: Yellow *"There is information, but details are available only on request"* Choice 3: Green *"There is detailed information provided to understand data policy"*
AV-AC-2 Delivery mechanism	The indicator reports the type of services available to the user to access data Part B 2.2 of INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119	 No information  Manual  Partial Inspire function  Full Inspire function	Choice 1: Red *"No information was found on data delivery mechanisms"* Choice 2: Red *"Order form/invoice is requested"* Choice 3: Yellow *"Online downloading services "* Choice 4: Green *"Online discovery and downloading services"* Choice 5: Green *"Online discovery + downloading + viewing services*

Availability – Accessibility Indicators

			(Advanced services)"
AV-AC-3 Data policy	The indicator comments on the status of information given for data policy Part B 8.2 of INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119	Not documented Restricted Partially restricted Unrestricted	Choice 1:Red *" Not or not well documented"* Choice 2: Red *"Restricted"* Choice 3: Yellow *"Accessible under moratorium"*[8] *(Ref.8)* Choice 4: Green *"Unrestricted"*
AV-AC-4 Pricing	The indicator comments on the provision of information related to cost basis Part B 8.1 of INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119	Not documented Commercial cost Charge Free	Choice 1: Red *"Not or not well documented"* Choice 2: Red *"Commercial cost charge"* Choice 3: Yellow *"Distribution charge" "Collection charge" "Free of charge for academic institutions and uses"* Choice 4: Green *"Open and Free, No charge"*
AV-AC-5 Readiness	The indicator comments on the information given for distribution format (Optional in INSPIRE)	Not documented Not ready to be consumed	Choice 1: Red *"Not or not well documented"* Choice 2: Red *"Proprietary and not well documented "*

[8] Ref 8: Moratorium is defined in SeaDataNet: data are initially restricted, but the conditions relax for academic or unrestricted access once a specified period of time after an event has elapsed (such as collection, publication, completion of QC procedures or project cessation).

		 Not ready to be consumed  Can be processed to be consumed  Ready to be consumed	Choice 3: Red *"Not proprietary but content not clearly specified "* Choice 4: Yellow *"Proprietary but content clearly specified "* Choice 5: Green *"Not proprietary and content clearly specified (eg auto-descriptive eg ODV, NetCDF CF) or at least with appropriate document describing the content."*

Table 6.2 Definitions of Accessibility Indicator Values

6.3 Performance indicators

The performance indicators indicate the ability of a system to keep operating over time and to meet real time operational conditions. It is related to service performance. Only one indicator is defined for performance:

AV-PE-1 **Responsiveness**	How responsive is the delivery service for the available data?

The methodology used to compute the performance indicator for the first DAR is described in Table 6.3.

Indicator name	Meaning (INSPIRE)	Meaningful (Symbol)	Achievable & Realistic (Choice)
AV-PE-1 Responsiveness	The indicator comments on the information given for the timeliness or ability to process a request in a deterministic and acceptable amount of time	Low response	Choice 1: Red *"No information is found on response time"* Choice 2: Red *"More than 1 week for release"*
		Medium response	Choice 3: Yellow *"Less or equal to 1 week for release"*
		High response	Choice 4: Green *"Online downloading (i.e. a few hours or less) for release"*

Availability – Performance Indicator

Table 6.3 Definitions of Responsiveness Indicators

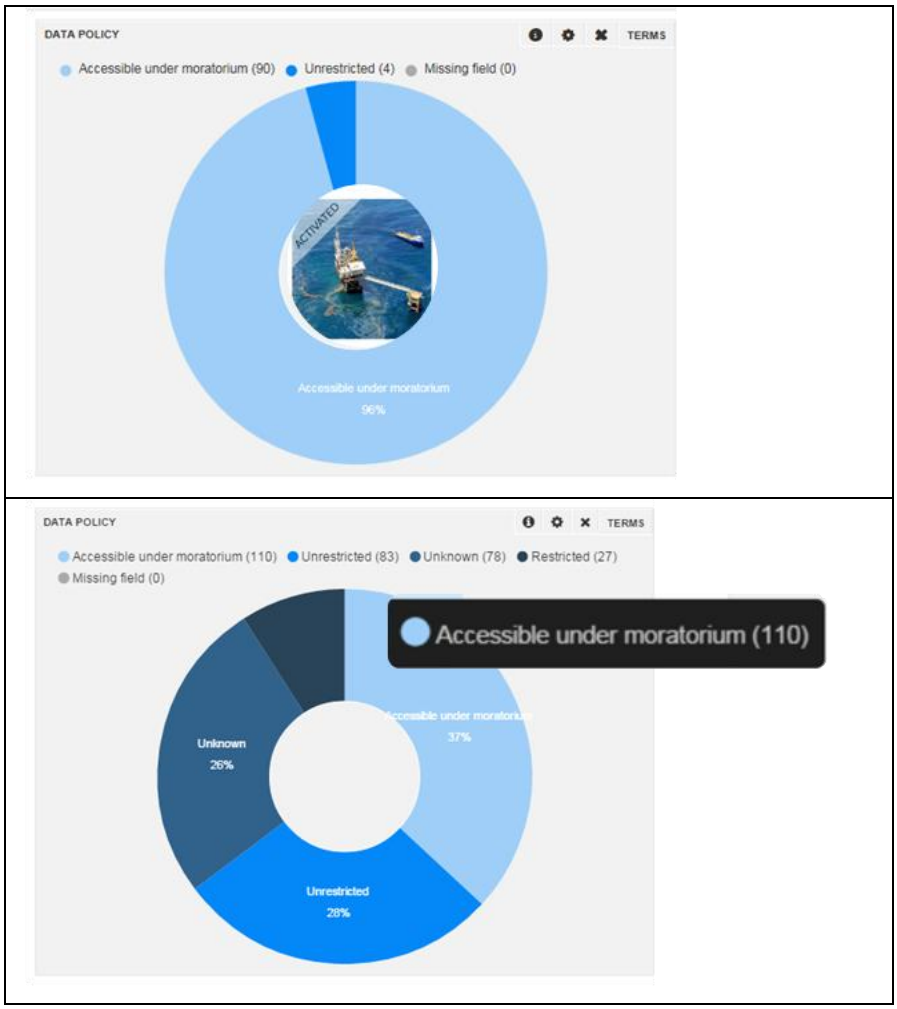
6.4 Summary of indicators

In summary, the eight availability indicators are:

AV-VI-1 Easily found	Can the data sets or series of data sets be found easily?
AV-VI-2 EU Inspire catalogue service	Is the service catalogue EU Inspire compliant?
AV-AC-1 Policy visibility	Visibility on data policy adopted by data providers.
AV-AC-2 Delivery	Data delivery mechanisms
AV-AC-3 Data Policy	Data policy
AV-AC-4 Pricing	Cost basis / price policy
AV-AC-5 Readiness	Format for use
AV-PE-1 Responsiveness	How responsive is the delivery service for the available data?

The eight final availability indicators

To facilitate the reproducibility of the indicator and other CheckPoint processes, an automated process was set-up to compute indicators from descriptors, which is directly accessible from the CheckPoint Dashboard (not yet available on-line). Indicators can be presented by challenge or for all challenges together (Fig. 6.1).



CheckPoint Dashboard: Availability – Accessibility Indicators for Oil Platform Leaks

CheckPoint Dashboard: Availability – Accessibility Indicators for all Challenges

Fig 6.1 AV-AC-3 - Variability of Data Policy as described by Challenge 3 (upper panel) or all challenges together (lower panel) as will be available from the CheckPoint Dashboard

7 Indicator results for all Challenges

The indicators are still being built, and only the indicators that have enough information in the metadatabase at this point are presented for all Challenges, i.e. AV-AC-2, AV-AC-3, AV-AC-4 and AV-PE-1. At present, no visibility indicators can be extracted from the database because of missing information. For Oil spill platform leaks, all availability indicators can be evaluated and will be presented in

the next section.

For AV-AC-2 the results are reported in Fig. 7.1 and Table 7.1

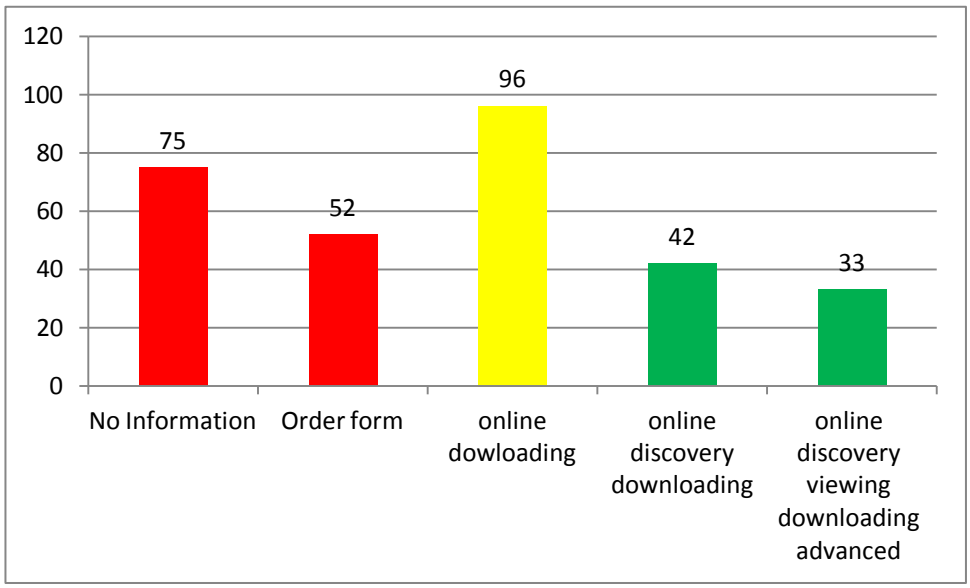


Fig. 7.1 AV-AC-2 Indicator on quality of delivery mechanisms as a function of indicator values and for all Challenges

Challenge	No information	Order form	Online downloading	Online discovery+ downloading	Online advanced sources	Total
Ch1		18			12	30
Ch2	74		1			75
Ch3		8	86			94
Ch4		1	8	6	18	33
Ch5		25	1			26
Ch6					3	3
Ch7	1			36		37
Total	**75**	**52**	**96**	**42**	**33**	**298**
%	**26**	**17**	**32**	**14**	**11**	**100**

Availability – Accessibility indicators for all challenges

Table 7.1 AV-AC-2 Indicator on quality of delivery mechanisms as a function of Challenges

Most of the datasets can be downloaded online and many are in full Inspire compliant systems (with downloading and viewing).

Wind Farming (CH1) and Fisheries (CH5) are the Challenges with the least Inspire compliant delivery mechanism most probably due to the immediate commercial value of the datasets. For the majority of Marine Protected Area-MPA (CH2) information is not given on the delivery mechanism.

For accessibility indicator AV-AC-3, related to Data Policy, the results are shown in Fig. 7.2 and Table 7.2.

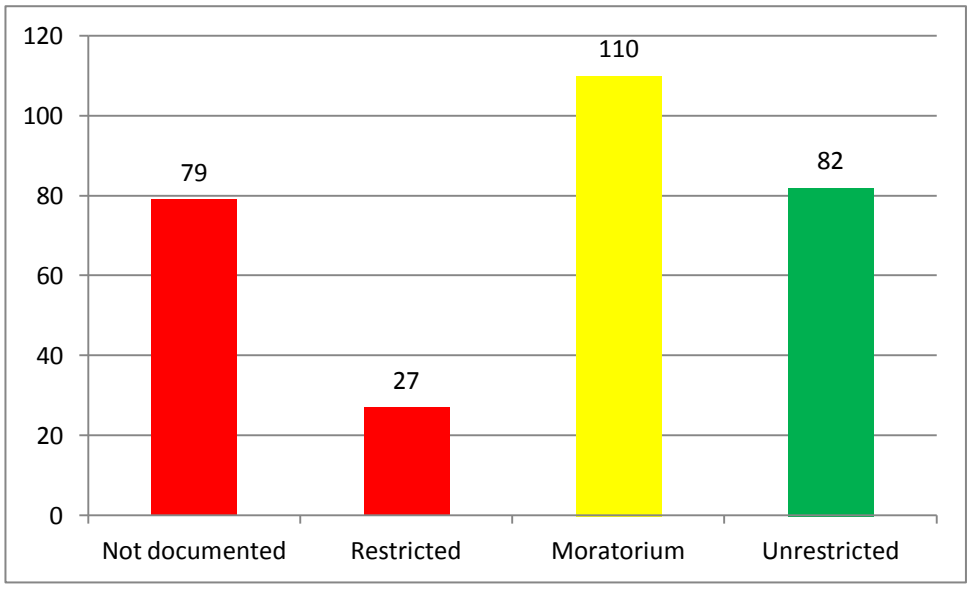


Fig. 7.2 AV-AC-3 indicator on data policy for all Challenge data sets as a function of indicator meaning

Challenge	Not documented	Restricted	Moratorium	Unrestricted	Total
Ch1			18	12	30
Ch2	74			1	75
Ch3			90	4	94
Ch4		2		31	33
Ch5		25		1	26
Ch6			2	1	3
Ch7	5			32	37
Total	**79**	**27**	**110**	**82**	**298**
%	**27**	**9**	**37**	**27**	**100**

Table 7.2 AV-AC-3 indicator on Data Policy as a function of Challenges

Availability – Accessibility Indicators for all challenges

Again Fisheries (CH5) has almost all restricted access data sets, for MPA (CH2) information is not available for data policy, and for CH3 most of the data are

available with a moratorium. All the other Challenges have unrestricted data policy upstream data sets.

For AV-AC-4 indicator, related to Pricing Policy, the results are shown in Fig. 7.3 and Table 7.3 for all Challenges.

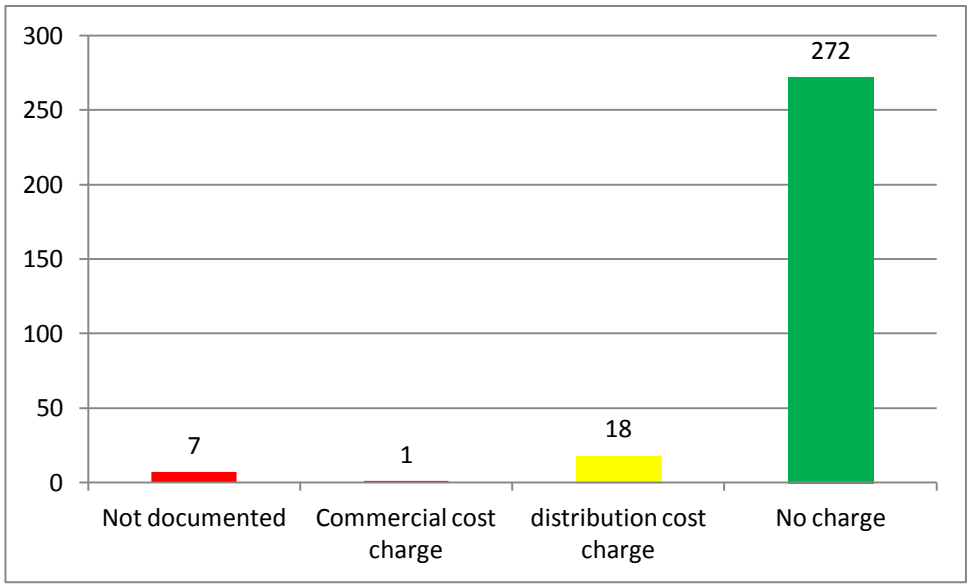


Fig. 7.3 AV-AC-4 indicator on pricing policy for all Challenges as a function of indicator meaning

Challenge	No information	Commercial cost charge	Distribution cost charge	No charge	Total
Ch1		1	18	11	30
Ch2				75	75
Ch3				94	94
Ch4				33	33
Ch5	2			24	26
Ch6				3	3
Ch7	5			32	37
Total	**7**	**1**	**18**	**272**	**298**
%	**2**	**0**	**6**	**92**	**100**

Table 7.3 AV-AC-3 indicator on Data Pricing Policy as a function of Challenges

Most data sets required by all Challenges are free of charge, with the exception of Wind Farming (CH1) where the data have a cost. This reflects the fact that Wind farming applications have relatively large costs associated with feasibility studies

regarding optimal sitings.

For AV-PE-1 indicator, related to time required to access the data, the results are shown in Fig. 7.4 and Table 7.4

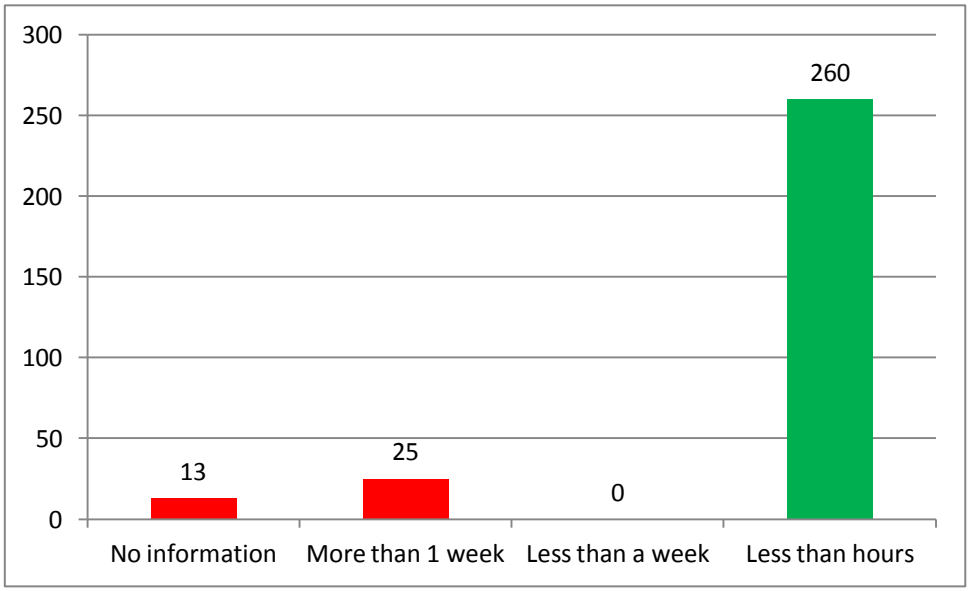


Figure 7.4 AV-PE-1 indicator on database responsiveness in terms of delivery for all Challenges as a function of indicator meaning

Challenge	No information	More than 1 week	Less than 1 week	few hours or less	Total
Ch1	12			18	30
Ch2				75	75
Ch3				94	94
Ch4				33	33
Ch5		25		1	26
Ch6				3	3
Ch7	1			36	37
Total	**13**	**25**		**260**	**298**
%	**5**	**8**		**87**	**100**

Availability – Perfomance Indicator for all Challenges

Table 7.4 AV-PE-1 Indicator on data delivery responsiveness as a function of Challenges

Most of the data delivery responsiveness is declared to be high but this will be further clarified when the data sets will be actually downloaded for the targeted

Challenge products. For Fisheries (CH5), the response time is more than a week, as part of the general problem of data availability for this Challenge.

To sum up, an overall indicator value was defined by taking the largest number of times an indicator value was chosen regardless of the challenge. This overall indicator status, for all 298 data sets, is shown in Table 7.6.

Synthesis of Indicators for all Challenges

Indicator	Meaning	Value
AV-AC-2	Delivery mechanism	 Partial Inspire function
AV-AC-3	Data policy	 Partially restricted
AV-AC-4	Pricing	 Free
AV-PE-1	Responsiveness	 High

Table 7.6 Summary of overall indicators for all Challenges

We expect this result to change slightly when the metadatabase is further checked and consolidated over the next few months and when the analysis is performed by characteristics or thematically rather than by Challenge.

8 The indicators Results: Oil Platform Leak Challenge

Oil Platform Leaks Challenge indicators

CH3, Oil spill Platform Leaks, has to provide an Oil Platform Leak Bulletin within 24 hours after a request from DG-MARE, containing information about transport, transformation and impacts of oil released from a source. The Bulletin has been in place since June 2014 and :

- the service can be accessed from:http://www.EMODnet-mediterranean.eu/portfolio/oil-platforms-leak/.
- the monitoring data are assessed here: https://webgate.ec.europa.eu/maritimeforum/en/node/3668
- and results from the service are discussed here: https://webgate.ec.europa.eu/maritimeforum/en/node/3668.

Since the service is already in place, most of the availability information has been finalized and for this Challenge we can give all eight indicators described in Section 6. There are 94 characteristics identified in the Challenge Oil Platform Leaks. They are categorized in the environmental matrices as in Figure 8.1

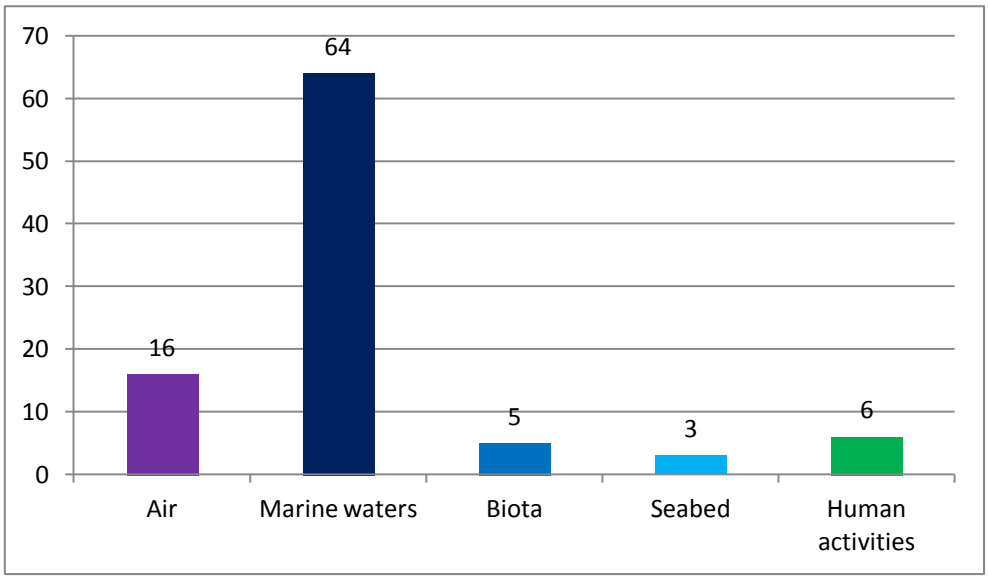


Figure 8.1 Subdivision of the 94 Oil Platform Leaks datasets in the environmental matrices

Table 8.1 shows the P01 and P02 metadata for this Challenge . The table shows that the vocabularies do not provide an internationally agreed definition for some characteristics. The lack of a code in some cases creates problems in organizing and managing the MedSea CheckPoint database.

Matrix	P02 Discovery Code	P01 characteristic code
Air	EWSB (Wind speed and direction)	Zonal wind component ESEWZZXX
		Meridional wind component ESNSZZXX
Marine Water	RFVL (Horizontal velocity of the water column - currents)	Water zonal velocity component LCEWZZ01
		Water meridional velocity component LCNSZZ01
	TEMP (Temperature of the water column)	Water temperature TEMPPR01
	GWDR (Wave direction)	Mean wave direction GWDRZZ01
	WVST (Wave height and period statistic)	Average zero crossing period of waves {Tz} on the water body GTZAZZ01
		Significant height of waves {Hs} on the water body GTDHZZ01
Biota	HBEX (Habitat extent)	Marine protected areas (polygon) – no P01 code
		Ecologically or Biologically significant areas (polygon) – no P01 code
		Cetacean areas (polygon) – no P01 code
		Coralligenous areas (polygon) – no P01 code
Sea bed	MBAN (Bathymetry and elevation)	Sea-floor depth BATHDPTH
	COGE (Coastal geomorphology)	Coast type – no P01 code
Human activities	MLES (Marine environment leisure usage)	Touristic area – no P01 code
	GP087 (Fishery characterisation)	Fisheries Zone (polygon) – no P01 code
	Oil - To be better defined	Oil Slick Area – no P01 code Oil Slick Thickness – no P01 code

Hierarchical organisation of vocabularies for Oil Platform Leaks

Table 8.1 The hierarchical organization of characteristics of Oil Platform Leaks

8.1 Availability – Visibility indicators

Starting with AV-VI-1 for Challenge 3, most of the data are displayed in catalogues such as EDMERP, especially for Air and Marine waters (see Fig. 8.2). For AV-VI-2, the databases are easily found using search engines or via reference catalogues.

Only three databases have no references in catalogues (Fig. 8.3).

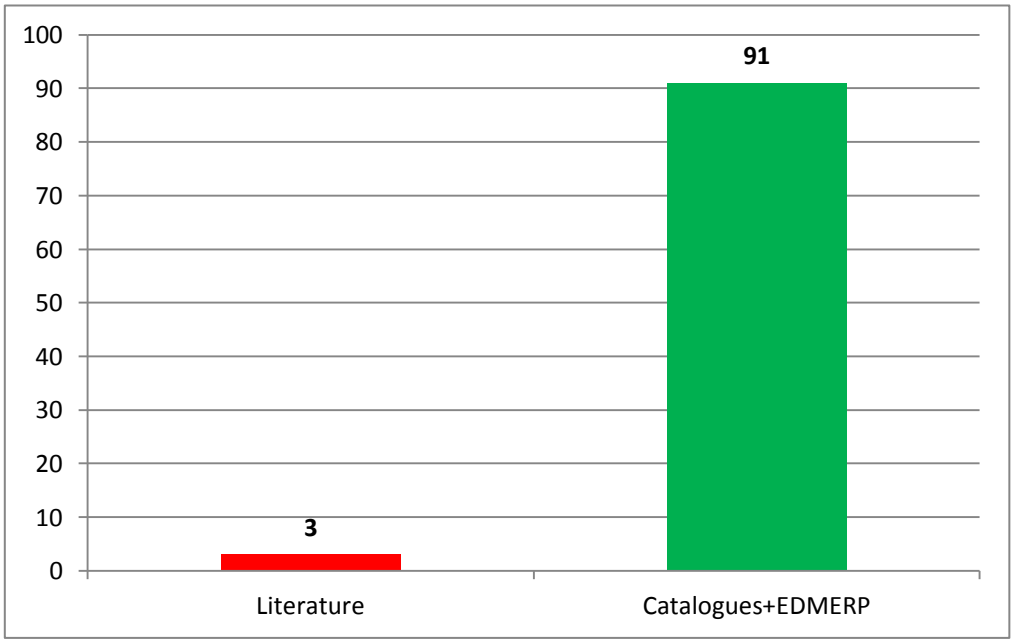


Figure 8.2 AV-VI-1 indicator on availability: Easily found (Challenge 3)

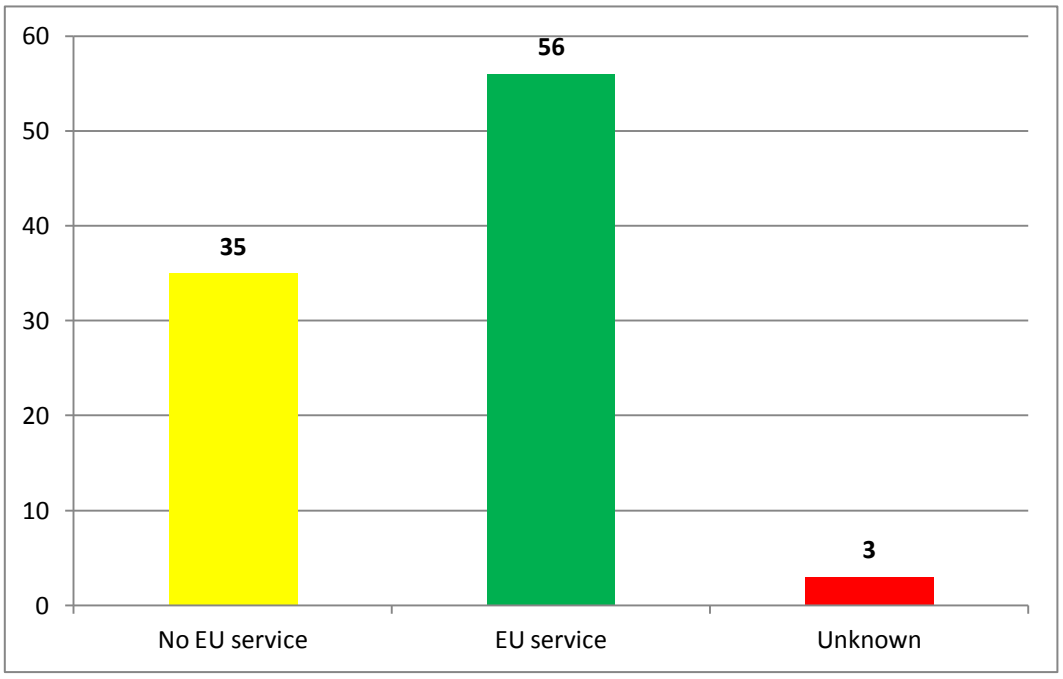


Figure 8.3 AV-VI-2 indicator on availability: datasets referenced in catalogue services (Challenge 3)

In CH3, 60% of the databases are accessible through services provided by MonGOOS related systems (http://www.mongoos.eu/), SeaDataNet (http://www.seadatanet.org/) and by EMODnet Bathymetry and Human activities. This challenge uses mainly operational forecasting system data sources both for Air and Marine matrices.

8.2 Availability – Accessibility indicators

For the AV-AC-1 indicators on "Policy visibility" for Challenge 3, the information is visible and can be extracted from internet documents. Thus this indicator is positive but details are missing and sometimes have to be requested directly from the data provider. In the report for CH3, this indicator is 'yellow'.

Figure 8.4 shows the results for the AV-AC-2 delivery mechanism: 90% of the information is accessible via online downloading. Thus the indicator is green.

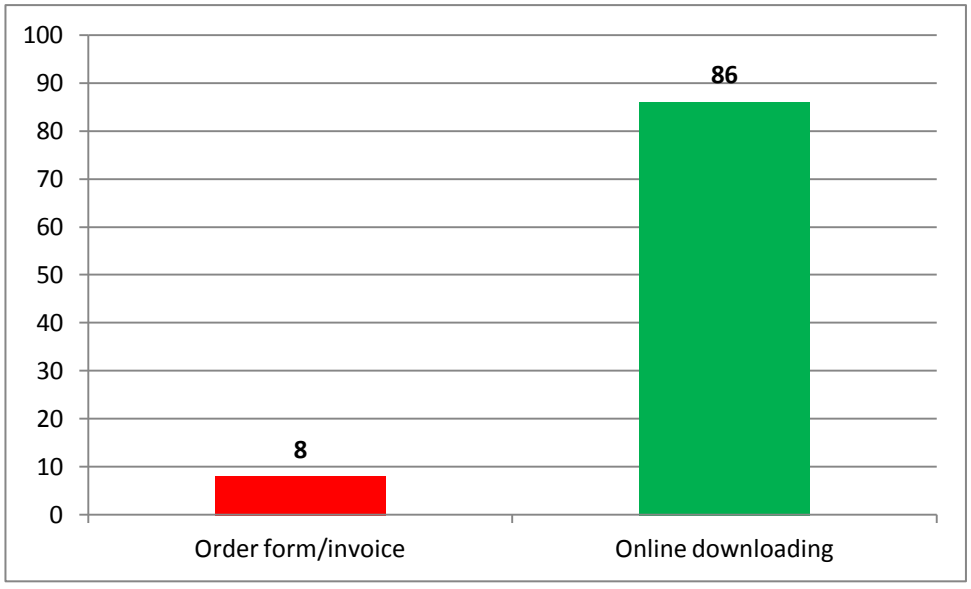


Figure 8.4 AV-AC-2 indicator on availability: delivery mechanism (Challenge 3)

Availability – Accessibility Indicators for Oil Platform Leaks

For the AV-AC-3 data Policy, the results in Fig. 8.5 show that data are accessible through a moratorium, i.e., only after agreements are put in place to access the data or after a defined period.

It is not easy for a non-expert to understand the data policy adopted by the data provider for a specific Challenge input data set. For the operational forecasting

systems in the Mediterranean Sea, the Data Exchange Agreement signed by the MonGOOS members provides the framework. In some cases the data owners also need to be contacted for information on the exact data policy adopted, as explained for the AV-VI-2 indicator.

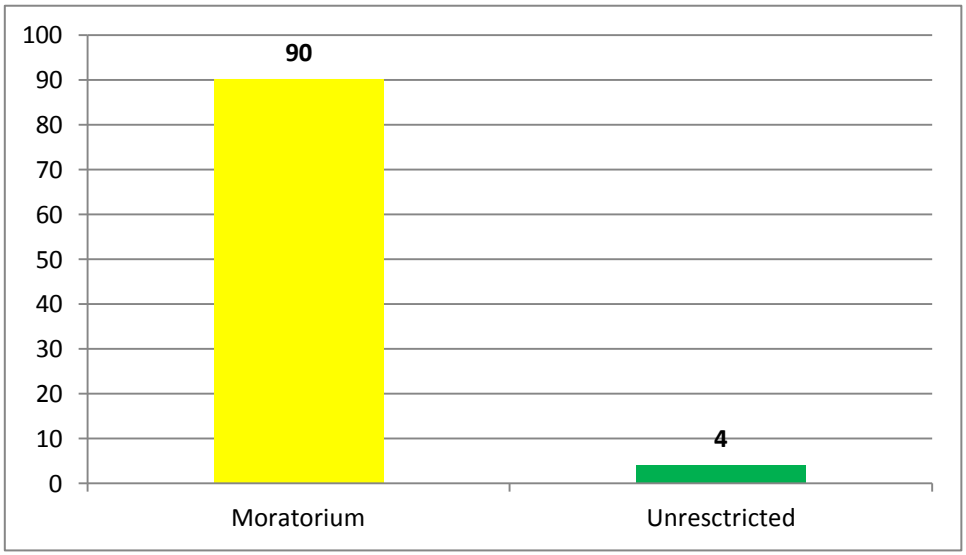


Figure 8.5 AV-AC-3 indicator on availability: data policy.

By combining the AV-AC-2 (Fig. 8.4) and AV-AC-3 (Fig. 8.5) indicators, it is possible to infer that only a few databases are accessible through partially Inspire-compliant services.

For the indicator AV-AC-4 (not shown), there is no cost for all 94 datasets after the delivery agreements have been put in place.

For AV-AC-5, the datasets for this Challenge adopt formats that have been suggested by SeaDataNet (mainly NetCDF for 83 databases). However, the databases are better organized in the cases of 'physical data' (e.g. wind, wave, current). The Biota databases provide 'shape files' and Human Activity databases are in 'text' format.

8.3 Availability – Performance indicator and synthesis

For the AV-PE-1 indicator, all datasets can be delivered within a few hours.

To sum up, for Oil Platform Leaks targeted products, availability indicators are

shown in the following table

AV-VI-1	Easily found	 High visibility
AV-VI-2	EU Inspire catalogue service	 Totally adequate
AV-AC-1	Policy visibility	 Medium transparency
AV-AC-2	Delivery mechanism	 Full Inspire function
AV-AC-3	Data policy	 Partially restricted
AV-AC-4	Pricing	 Free
AV-AC-5	Readiness	 Ready to be consumed
AV-PE-1	Responsiveness	 High response

Synthesis of Indicators For Oil Spill Platform Leaks

9 Summary of the lessons learned by each Challenge

Synthesis of Challenge Reports on Lessons Learned

A summary of the lessons learned was explicitly requested in the Call for Tender. The 'analysis' for the first DAR refers here to the search for the input datasets and their cataloguing. Some specific questions were posed to Challenges on the accuracy of data sources, gaps in the input data, most useful data sources, trends in availability (improving, worsening), usefulness of secondary data sources (e.g. MyOcean, EMODnet), availability-accessibility issues, priorities in data collection or assembling.

The overall questions posed to the Challenges were:

Q1: Are your input data sources accurate enough for the Challenge products?

Q2: Where are the main gaps in the input data?

Q3: (If there are different sources for similar parameters) Which data source is the most useful?

Q4: Is the availability of data improving or worsening?

Q5: Are there more data available that could not be used for this Challenge because they were too expensive or complicated to access?

Q6: What would the main priority be for data collection or assembly in order to improve the accuracy of the results or help overcome the difficulty in producing them?

The answers were given in specific Challenges reports, which are summarized here.

Q1: Are your input data sources accurate enough for the Challenge products?

The data deriving from modeling outputs are normally considered as accurate. It has been suggested that EMODnet portals should be used as independent data sources of observations to quantify the model output accuracy and suitability. However, it has been observed that accuracy is sufficient if multi-model and multi-forcing forecast systems are used (as is the case of Oil Platform Leaks Challenge).
In general, the accuracy of in situ data is considered higher than the satellite data for the selected applications, although the temporal and spatial coverage is better assured by satellite data. The accuracy of in situ data is not yet considered suitable for some assessment criteria (e.g. connectivity) and information on priority species and habitats is also limited. The final consideration is that data accuracy can be assessed correctly if the 'availability indicators' are 'green' or in the worst case 'yellow'.

Q2: Where are the main gaps in the input data?

The spatial and temporal resolution of all data is an issue for all Challenges, but the main gaps are related to biological and biodiversity data sources. The availability of VMS maps related to vessel coverage is critical and there are also gaps in the accuracy, resolution and availability of the socio-economic and environmental data.
In some cases the reason for gaps is related to the lack of human and financial resources, as well as to the non availability of data sources, especially at a national level.

Q3: Which data source is the most useful?

The most useful data sources are for example the most available and high resolution sources for meteo-oceanographic condition datasets.
The use of multiple data sources often enables gaps to be filled such as between tide gauges and sea level altimetry to meet sea level rises. Multiple data sources also enable uncertainties to be estimated in the products derived from the input datasets.

Q4: Is the availability of data improving or worsening?

For Challenges that need to access model data, data availability is improving due to the development of 'Marine Services' and 'EMODnet Thematic portals'. The availability is also improving for Fisheries, due to the implementation of R-routines (a recruit model for stock assessment) for the analysis of VMS and AIS datasets.
In general there is the idea that there is no need to access data from original

sources. Their availability in federated systems is considered to be good, since they offer the 'best copies' of data.

The follow up of indicators every six months along with iteration with upstream providers should also help to improve the process.

Q5: Are there more data available that could not be used for this Challenge because they were too expensive or complicated to access?

All Challenges have underlined that much data can be available under specific agreements. The main issues related to data availability and access are:

- need to access bathymetric data at high resolutions
- biological data are restricted in some cases (e.g. MEDISEH project)
- AIS data are not available
- Information provided by data sources are, in many cases, not sufficient to assess their usefulness.

There is also a well-known difficulty in retrieving wildlife tracking, because proprietary issues and data tracking are centralised and metadata are managed elsewhere or not managed. This is a key issue for PMA (CH2).

Q6: What would be the main priority for data collection or assembly in order to improve the accuracy of the results or help overcome the difficulty in producing them?

In terms of numerical model data, the state-of-the-art is to use well-established modelling systems, such as the one used to derive the 10-year dataset used in Challenge 1, which also assimilates all the available measurements. In addition, a combination of model outputs with observations and statistical methods for local value-adding purposes could improve the accuracy of the results. The need to assess data assembly activities on the Mediterranean biodiversity has also been highlighted.

It is important to promote agreements with different authorities in order to be able to provide higher resolution, better accuracy, as well as environmental coastal and human activities data layers.

Bearing in mind the MedSea CheckPoint experience in the compilation of the metadatabase, incorporating descriptive metadata would be useful, following the example of MyOcean targeted products. They are easy to consult and visualize. The acquisition of information on most of the MedSea CheckPoint datasets requires a 'document archaeology work'. The use of a metadata model, such as the one used

in MyOcean or SeaDataNet, would simplify the implementation of the MedSea CheckPoint database.

10 Discussion and conclusions

The first Data Adequacy Report contains results related to:

Three main DAR results

1. the building of the Metadatabase which began during the Literature Survey;
2. the definition of the methodology for the objective assessment of the Data Adequacy;
3. Initial results for the assessment of input datasets for the Challenges.

The adoption of well-defined standards and controlled vocabularies has guided the selection of metadata models and formats for upstream data sets. This enabled a MedSea CheckPoint Inspire compliant information system to be built with a metadatabase for developing and computing data adequacy indicators.

Vocabulary is not large enough to contain all characteristics

It soon became clear that the hierarchical organisation of the SeaDataNet vocabularies is helping the development of the information system. At the same time existing SeaDataNet controlled vocabularies were found not to contain all the terms needed for implementing and managing a metadatabase for the seven MedSea CheckPoint Challenges. This is a basic problem that should be considered in future projects.

298 data sets and 126 data providers for the Challenge characteristics

Data originates from different providers (research centers, private companies, public agencies, etc.) and is accessed by different users (public administrations, decision-makers, citizens, etc.) for different purposes (research, government, civil protection, etc.). A total of 298 data sets are now listed as potential contributors to the Challenge targeted products. They are provided by 126 different data sources.

45 Characteristics needed for Challenges

The 45 different characteristics selected for the Challenges are extremely heterogeneous. They originate from different sensors, with different acquisition geometries and sampling strategies, different numerical and statistical models, different spatial resolutions, etc. In addition the same characteristics originate from different data providers depending on the expert practice. In this first DAR each challenge was considered as an 'isolated universe of discourse'. This means that

many problems on the multiple use of the same datasets for different challenges have not been considered here.

The availability indicators

This first Data Adequacy Report defined the first set of indicators for the 'fitness for use' assessment. Adequacy is defined on the basis of two territories: 'appropriateness' and 'availability'. For 'availability' eight DAR indicators were defined as follows:

Territory	Indicator name	Meaning
Visibility	AV-VI-1	Easily found
	AV-VI-2	EU Inspire Catalogue Service
Accessibility	AV-AC-1	Policy visibility
	AV-AC-2	Delivery mechanism
	AV-AC-3	Data policy
	AV-AC-4	Pricing
	AV-AC-5	Readiness
Performance	AV-PE-1	Responsiveness

The indicators were extracted from the metadatabase, however not all of them could be evaluated at this time. There are still gaps in the database that prevent this first assessment from being completed. This is true for visibility indicators and for part of the accessibility indicators.

At the level of all the Challenges, the overall results for the indicators are:

First DAR results : indicators for all Challenges

Indicator	Meaning	Value
AV-AC-2	Delivery mechanism	Partial Inspire function
AV-AC-3	Data policy	Partially restricted
AV-AC-4	Pricing	Free
AV-PE-1	Responsiveness	High

This initial analysis highlights that in terms of accessibility and performance, the input data sets for the Challenges are medium to high, although the situation varies greatly from Challenge to Challenge.

For the Oil spill Leaks Challenge, we were able to assess all the availability

indicators and the result is

AV-VI-1	Easily found	 High visibility
AV-VI-2	EU Inspire catalogue service	 Totally adequate
AV-AC-1	Policy visibility	 Medium transparency
AV-AC-2	Delivery mechanism	 Full Inspire function
AV-AC-3	Data policy	 Partially restricted
AV-AC-4	Pricing	 Free
AV-AC-5	Readiness	 Ready to be consumed
AV-PE-1	Responsiveness	 High response

For this Challenge, data inputs mainly come from numerical ocean, wave and atmospheric models, as well as human activities. Data are generally available but the data policy is difficult to understand and the delivery mechanism is only partially compliant with an advanced Inspire service.

In conclusion, this first DAR highlights that it is possible to develop objective indicators of fitness for use for the input data sets to the Challenges. The availability of input data sets for all Challenges seems to run between medium to high with the only exception of CH1 (Wind Farm siting) and CH5 (Fisheries). CH2 shows the least Inspire catalogues and data policy transparency.

The next step will be to discuss thematic indicators together with Challenge indicators, i.e. considering the average weighted indicator among Challenges that require the same characteristics. The addition of appropriateness indicators will also probably be combined with the availability indicators in order to change the assessment of this first DAR for the fitness for use of monitoring data sets for the Mediterranean Sea.

11 Annex 1 Metadata format and model

Metadata describe data elements or attributes (name, size, data type, etc) and data structures (length, fields, columns, etc). Metadata also document pieces of software associated with the data (platform, language, input parameters, etc). Metadata should include descriptive information on the context, quality and condition, or characteristics of the data.

The metadatabase implemented for the MedSea CheckPoint is based on the metadata formats and models presented in Table A1

Metadata format	Description
Dublin Core Metadata Element Set (DCMES): ANSI/NISO Z39.85 or EN ISO 15836 = Dublin Core	This standard provides a list of descriptive metadata to support resource discovery for geographical and non-geographical communities. It is used considerably by libraries for their repositories. There are two implementations with a set of associated tools: -HTML -Resource Description Framework (RDF)/XML a format for metadata, which is now used the most. Here is the list of associated metadata: -language-title-subject-description-publisher-rights-identifier-creator-contributor-date created-date modified-references-replaces.
EN ISO 19119 (taxonomy of services)	This is a European standard for documenting services. It belongs to the INSPIRE Recommendation.
Geographical information – Metadata – Implementation specification: EN ISO 19139	This is the unique standard interpretation (implementation) of the ISO 19115 based upon the use of XML schema. It belongs to the INSPIRE Recommendation.
Geographical information: Metadata EN ISO 19115	This is a GIS European standard for documenting data sets. It is the basis of INSPIRE Recommendation for the description of metadata. Three levels of resource discovery metadata: - discovery - browsing - exploitation There are different categories of metadata: - identification -spatial representation -quality information -contents description -presentation modalities -dissemination modalities -maintenance modalities -security level -restriction level with a model for the Medadata itself: -identifier -language -character set -metadatastandardname -metadatastandardversion -hierarchy level
ISO 19108	This standard defines concepts for describing temporal characteristics of geographic information. It belongs to the INSPIRE Recommendation.
ISO 8601	Standard for numeric representation of date and time. It belongs to the INSPIRE Recommendation.

Table A1- MedSea CheckPoint metadata formats

12 Annex 2 List of Characteristics as a function of Challenge and number of data sources.

P02 Parameter list	Ch.1 Wind Farm	Ch.2 MPA	Ch.3 Oil leaks	Ch. 4 Climate & coast	Ch. 5 Fish.	Ch. 6 Mar. Env.	Ch. 7 River input	Total by P02
1. Administrative units		24						24
2. Air pressure	1							1
3. Air temperature	1							1
4. Atmospheric humidity	1							1
5. Bathymetry and Elevation	1	1	2	1				5
6. Bird behaviour	1	2						3
7. Bird reproduction	1							1
8. Bird taxonomy-related counts	2	2						4
9.Cetacean abundance		1						1
10. Chlorophyll pigment concentrations in the water column		1				1		2
11. Coastal geomorphology		1	1	2				4
12. Concentration of suspended particulate material in the water column							9	9
13. Depositional Eenvironment		1						1
14. Dissolved oxygen parameters in the water column		2						2
15. Dissolved total and organic nitrogen concentrations in the water column							4	4
16. Dissolved total or organic phosphorus concentrations in the water column							4	4

17. Fauna abundance per unit area of the bed	4	1						5
18. Fish abundance in water bodies	1	2					1	4
19. Fish and shellfish catch statistics					10			10
20. Fish reproduction	1							1
21. Fishery characterisation			2					2
22. Fishing by-catch					6			6
23. Habitat extent		20	5	1				26
24. Horizontal platform movement					10			10
25. Horizontal spatial co-ordinates		2						2
26. Horizontal velocity of the water column (currents)	2	2	28					32
27. Lithology		1						1
28. Man-made structures				1				1
29. Marine environment leisure usage			1					1
30. Nitrate concentration parameters in the water column						1	4	5
31. Nutrient fluxes between the bed and the water column		1						1
32. Phosphate concentration parameters in the water column						1	4	5
33. River flow and discharge							8	8
34. Salinity of the water column	1	3						4
35. Sea level	1			8				9
36. Seal abundance		1						1
37. Sediment resuspension		1						1
38. Sedimentary structure	1							1

39. Spectral wave data parameters	1							1
40. Temperature of the water column	1	3	14	20			3	41
41. Terrestrial mapping			1					1
42. Unspecified		1	2					3
43. Wave direction	1		7					8
44. Wave height and period statistics	5		14					19
45. Wind speed and direction	3	2	17					22
Total #	30	75	94	33	26	3	37	298
P02 Parameter list (45 in total)	Ch.1 Wind Farm	Ch.2 MPA	Ch.3 Oil leaks	Ch. 4 Climate & coast	Ch. 5 Fish.	Ch. 6 Mar. Env.	Ch. 7 River input	Total by P02

13 Annex 3 Data sources for the 298 upstream data identified by the Challenges

1	Aarhus University Department of Bioscience Marine Ecology Roskilde	64	International Sediment Initiative (ISI)
2	Accobams	65	ISAC - Institute of Atmospheric Sciences and Climate
3	Argos CLS	66	ISAC - Institute of Atmospheric Sciences and Climate (Rome)
4	Balearic Islands Coastal Observing and Forecasting System (SOCIB)	67	ISPRA-Institute For Environmental Protection And Research
5	Biodiversity Structure in the Romanian Coastal Zone	68	Istituto Nazionale di Geofisica e Vulcanologia – INGV Sede Di Bologna
6	British Oceanographic Data Centre (BODC)	69	Italian Fisheries Control Agencies
7	BSH	70	IUCN
8	Centre d'Etudes Techniques Maritimes et Fluviales (CETMEF)	71	Joint Research Center
9	CNES	72	JRC - Institute for Environment and Sustainability (IES)
10	CNR Institute for the Marine and Coastal Environment (IAMC) - Oristano	73	Laboratory of Oceanography of Villefranche (Lov)
11	CNR-ISAC	74	Maltese Fisheries Control Agencies
12	CNR-ISAC (MONGOOS)	75	MarBEF - Marine Biodiversity and Ecosystem Functioning
13	CNRM - National Center for Meteorological Research - Toulouse	76	Marine Renewable Integrated Application Platform
14	CoCoNet Towards Coast To Coast Networks of Marine Protected Areas (From The Shore To The High And Deep Sea) Coupled With Sea-Based Wind Energy Potential	77	Marine Traffic
15	Collecte Localisation Satellite	78	Mediseh
16	Complex Systems Research Center (CSRC) University of New Hampshire	79	Mediterranean Decision Support System for Marine Safety
17	Croatian Fisheries Control Agencies	80	Mediterranean Sensitive Habitats
18	Cyprus Oceanography Center (OC-UCY)	81	MedPan
19	Data Support Section of the Computational and Information Systems Laboratory at the National Center for Atmospheric Research - Toulouse	82	Mercator Ocean
20	Development of the Med Pol Phase III Data Base UNEPMAP	83	Met Office

21	DG Environment,	84	Meteo France
22	ELNAIS	85	Meteo France DpServFDP
23	EMODnet	86	MyOcean
24	EMODnet Bathymetry	87	MyOcean (CLS)
25	EMODnet Hydrography	88	MyOcean (CNR-ISAC)
26	EMODnet SeaBed Habitats	89	MyOcean Med Mfc (INGV)
27	EMODnet Physics	90	National and Kapodistrian University of Athens Department of Physics Atmospheric Modeling and Weather Forecasting Group
28	EMODnet Chemistry	91	National Council of Research - ISMAR
29	ENSTA	92	National Institute for Marine Research and Development "Grigore Antipa"
30	ENSTA Bretagne	93	National Oceanic And Atmospheric Administration (Noaa)
31	Ente Nazionale Idrocarburi (ENI)	94	National Oceanographic Data Committee (The Netherlands?)
32	Estuarine Quality Classes for Water Framework Directive Indicators	95	Natura
33	European Centre for Medium-Range Weather Forecasts (ECMWF)	96	Natura2000
34	European Commission	97	Netherlands Institute of Ecology Centre for Estuarine and Marine Ecology (NIOO-CEME)
35	European Global Ocean Observing System (EUROGOOS)	98	Network of Marine Protected Area Managers in the Mediterranean
36	European Environment Agency	99	Ocean Color Tac - MyOcean
37	European Space Agency (ESA)	100	Oceana
38	EuroStat	101	Oceanographic Data Center ???
39	Evaluation of the Demersal Resources in the GFCM Area Using Direct Methods	102	Ogs (Istituto Nazionale di Oceanografia e di Geofisica Sperimentale) Department of Biological Oceanography
40	FAO - Food and Agriculture Organization of the United Nations	103	Ogs (Istituto Nazionale di Oceanografia e di Geofisica Sperimentale) Division of Oceanography
41	Federal Institute of Hydrology (BFG)	104	Permanent Service For Mean Sea Level (PSMSL)
42	Flanders Marine Institute (VLIZ)	105	Policy Oriented Marine Environmental Research in the Southern European Seas
43	Food and Agriculture Organization of the United Nations Fisheries and Aquaculture Department	106	Puertos del Estado
44	French Fisheries Control Agencies	107	RAC-SPA
45	French Marine Protected Areas Agency	108	RAMSAR
46	GEBCO	109	Regional Activity Centre for Specially Protected Areas (RACSPA)
47	German Oceanographic Datacentre (DODS)	110	Rempec Medslik-Ii

48	GFCM Fisheries Restricted Areas	111	Seadatanet-Pan-European Infrastructure for Marine Data 2
49	Greek Fisheries Control Agencies	112	Service Contract Concerning Coastal Erosion Evaluation of the Needs for Action
50	Hellenic Centre for Marine Research (HCMR)	113	SHOM
51	Hellenic Centre for Marine Research Hellenic National Oceanographic Data Centre (HCMR HNODC)	114	Slovenian Fisheries Control Agencies
52	Hydrological Cycle in Mediterranean Experiment (HYMEX)	115	Spanish Fisheries Control Agencies
53	ICCAT	116	Systeme d'Observation du Niveau des Eaux Littorales
54	IEO Spanish Oceanographic Institute	117	The Pelagos Sanctuary
55	Ifremer	118	Tulane University Department of Earth and Environmental Sciences
56	Ifremer Centre De Brest	119	TWReferencenet - Management And Sustainable Development Of Protected Transitional Waters
57	Ifremer Idm Sismer	120	UNEP MAP
58	Institute of Accelerating Systems and Applications (IASA-UAT)	121	UNESCO
59	Institute of Marine Sciences Middle East Technical University	122	United Nations Environment Programme Global Environment Monitoring System (Unepgems)
60	Institute of Research for Development (IRD)	123	University of Hawaii Sea Level Center
61	INSU	124	University of New Hampshire
62	INSU (Institut National Sciences de L'univers) Serv. d'Obs. en Milieu Littoral - Somlit	125	World Database on Protected Areas
63	International Ocean Institute - Malta Operational Centre (University Of Malta) Physical Oceanography Unit (UMTIOIPOU)	126	WWF

14 Annex 4 Nomenclature

Adequacy	Sufficient to satisfy a requirement or meet a need.
Appropriateness	What is made available to the challenge? What motivated/decided them to select one observation rather than another one.
Assessment criteria	The criteria aim to characterize/depict the inputs in terms of 3 territories capable of showing performance and gaps in the present monitoring system, appropriateness, availability and fitness for purpose.
Availability	How a characteristic is made available to the user.
Comparability	To examine things to assess how they are alike and how they are different; to judge one thing and measure it against another thing.
Compatibility	Things that can exist together without problems and conflicts.
Data	Reinterpretable representation of information in a formalised manner suitable for communication, interpretation or processing (ISO 19115).
Dataset	A “dataset” is an identifiable collection of data (ISO 19115). It can be a time series, a lithological description of a marine sample, a gridded dataset such as a DTM, an hydrodynamic model output, a GIS dataset or a feature layer of a GIS dataset, a database or a table of values in a publication. A dataset can consist of several files
Dataset series	A dataset series is a collection of datasets sharing the same specifications of production (INSPIRE).
Environmental matrices	This concept is introduced to avoid ambiguities when using the name of a characteristic such as “temperature”. The environment matrix is the environment to which a characteristic is related and defined as: Air, Fresh water, Marine water, Riverbed, Seabed, Biota/Biology, Human activities.
External adequacy	External adequacy is defined in terms of the formal specification of question relevance, which expresses natural data dependencies among the external facts used in the ontology of the domain of expertise.
Fit for purpose	Defining the usefulness of data for assessment purposes. Ability to appreciate the data exploitability (Challenge feedback on efficiency & reliability of marine data)
Harmonisation	Changes/restructuration of the reference version of the data set to make it compatible
Input dataset	This is the collection of existing data to be input to the Challenges. They are uniquely identified as a combination of (variable, dataset, intended use) or of (geographical feature, dataset, intended use) depending on their nature. They can be shared between challenges.
Relevance	Covering the extent to which data are appropriate for objectives of Challenges.
Unique identification	Combination of (characteristic, dataset, intended use) or of (geo. feature, dataset, intended use).
Universe of Discourse	View of the real or hypothetical world that includes everything of interest (ISO 19101).
Validation	Confirmation by examination and provision of objective evidence that the particular requirements for a specific intended use are fulfilled.