

# 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

## Revised First Data Adequacy Report EMODNET MedSea CheckPoint

Total number of pages: 62

Workpackage	11	11.2 Revised 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**



## Table of Contents

<b>EXECUTIVE SUMMARY .....</b>	<b>6</b>
<b>1 INTRODUCTION .....</b>	<b>8</b>
1.1 GENERAL.....	8
1.2 KEY DEFINITIONS.....	9
1.3 ISO METHODOLOGY.....	10
1.4 CHECKPOINT TARGET AUDIENCE .....	11
1.5 STRUCTURE OF THE DOCUMENT.....	13
<b>2 CHECKPOINT ASSESSMENT METHODOLOGY.....</b>	<b>13</b>
<b>3 THE CHECKPOINT METADATABASE .....</b>	<b>14</b>
3.1 VOCABULARIES.....	17
<b>4 THE UPDATED LITERATURE SURVEY RESULTS .....</b>	<b>18</b>
4.1 DATA SET IDENTIFICATION ISSUES .....	19
<b>5 CHECKPOINT INDICATOR DEFINITIONS.....</b>	<b>20</b>
5.1 THE SMART METHODOLOGY .....	21
5.2 INDICATOR EVALUATION SCALE .....	22
<b>6 AVAILABILITY INDICATORS.....</b>	<b>22</b>
6.1 VISIBILITY INDICATORS .....	23
6.2 ACCESSIBILITY INDICATORS .....	25
6.3 PERFORMANCE INDICATORS .....	28
6.4 SUMMARY OF INDICATORS .....	29
<b>7 INDICATOR RESULTS FOR ALL CHALLENGES.....</b>	<b>30</b>
<b>8 THE INDICATORS RESULTS: OIL PLATFORM LEAK CHALLENGE.....</b>	<b>34</b>
8.1 AVAILABILITY – VISIBILITY INDICATORS .....	36
8.2 AVAILABILITY – ACCESSIBILITY INDICATORS .....	37
8.3 AVAILABILITY – PERFORMANCE INDICATOR AND SYNTHESIS .....	39
<b>9 INITIAL ASSESSMENT OF THE EMODNET PORTAL DATA SETS AND MYOCEAN SERVICE .....</b>	<b>39</b>
<b>10 SUMMARY OF THE LESSONS LEARNED BY EACH CHALLENGE.....</b>	<b>42</b>
<b>11 DISCUSSION AND CONCLUSIONS .....</b>	<b>44</b>

<b>12</b>	<b>ANNEX 1 ASSESSMENT CRITERIA.....</b>	<b>48</b>
<b>13</b>	<b>ANNEX 2 METADATA FORMAT AND MODEL .....</b>	<b>50</b>
<b>14</b>	<b>ANNEX 3 LIST OF CATEGORIES OF CHARACTERISTICS AS A FUNCTION OF CHALLENGE AND NUMBER OF PROVIDERS. ....</b>	<b>52</b>
<b>15</b>	<b>ANNEX 4 PROVIDERS FOR THE 298 UPSTREAM DATA SET DESCRIPTORS IDENTIFIED BY THE CHALLENGES .....</b>	<b>55</b>
<b>16</b>	<b>ANNEX 5 EMODNET AND MYOCEAN INDICATORS.....</b>	<b>58</b>
<b>17</b>	<b>ANNEX 6 NOMENCLATURE .....</b>	<b>61</b>

## Glossary

AIS: Automatic Identification System  
BODC: British Oceanographic Data Centre  
CFP: Common Fisheries Policy  
CH: Challenge  
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 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 Forecast  
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: Hydrographic 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 (ES)  
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 Institute for Advanced Studies (CSIC-UIB) (ES)  
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  
ISO NP: ISO New Proposal  
ISO NP TS: ISO NP Technical Specification  
ISPRA: Italian National Protection Agency  
JCOMM: Joint WMO-IOC Technical Commission for Oceanography and 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 (FR)  
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

UNESCO: United Nations Educational, Scientific and Cultural Organization

WISE: Water Information System for Europe

VLIZ: Flanders Marine Institute (BE)

WFD: Water Framework Directive

WMO: World Meteorological Organisation

WMOP: SOCIB Western Mediterranean Sea Operational forecasting system

VMS: Vessel Monitoring System

WWF: World Wildlife Fund

## Executive Summary

EMODnet MedSea CheckPoint aims to document the reliability and utility 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 or applications. There are seven Challenges: Wind Farm Siting (CH1), Marine Protected Areas (CH2), Oil 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 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 built on the basis of existing ISO standards and are constructed directly on the metadatabase that contains information on the upstream data providers. For each Challenge, CheckPoint Information has been collected 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 set descriptors**.

On the basis of this metadatabase, the Literature Survey results were reviewed. It was confirmed that **45 categories of characteristics** (a specifically defined class of variables derived from observations or models and/or the geographical representation of an object) are needed for the seven Challenges and **126 data providers** (primary provider/originator and secondary provider/distributor/curator) would be required by the Challenges in order to develop the targeted products. This highlights the importance of the MedSea CheckPoint portal for collecting and organizing the information in 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 targeted products have not yet been developed and thus appropriateness cannot be properly defined.

The availability investigates “**how the input data sets are made available to the Challenge use**” and **eight indicators** were 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 acquisition 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 need to be purchased, and CH5 (Fisheries) data responsiveness is low (more than a week to access the data).

For CH3 (Oil Platform Leaks) all eight availability indicators were estimated. All the results show highly positive indicator 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 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 area of appropriateness.

## 1 Introduction

### 1.1 General

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.

It is key to link the European ocean observations to applications identified by the Integrated Maritime Policy. The latter seeks to provide a more coherent approach to maritime issues, with increased coordination between different policy areas. It focuses on blue growth and marine knowledge. The seven Challenges can be seen as special but important study cases for blue growth and marine knowledge strategy.

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.

CheckPoint  
main  
objectives

To achieve these aims, the project produces three levels of outputs:

- *CheckPoint information* consisting of metadata on the input monitoring data sources and providers and Challenge targeted products;
- *CheckPoint indicators* based on the monitoring system assessment criteria;
- CheckPoint results consisting of a Literature Survey and two Data Adequacy Reports.

CheckPoint  
outputs

The Data Adequacy Reports (DARs) contain the major assessment findings. This report is the first DAR, building on the definitions and the methodology described in the literature survey [Ref1]<sup>1</sup> and showing how the methodology can be applied for selected indicators and a selected challenge. Naturally the results of the

Data  
Adequacy  
Reports

---

<sup>1</sup> Ref1: Literature Survey available at <https://webgate.ec.europa.eu/maritimeforum/node/3646>

assessment are valid within the specific point of view of the present Challenges and will be enlarged in scope in the future.

## 1.2 Key definitions

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'  
definition

**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.

Data  
Adequacy  
concept

In an EC Report [Ref2]<sup>2</sup> adequacy was defined as an assessment of the 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]<sup>3</sup>. The CheckPoint adequacy is close to this definition but focused on several Challenges.

In other words **adequacy** is here intended as 'sufficient to satisfy a requirement or meet a need' [Ref 4]<sup>4</sup>. Thus, '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.

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 quality of a dataset from a producer and a user point of view is shown in Figure 1.1.

Universe of  
Discourse

In the CheckPoint case, the universe of discourse provides the model to assess the effectiveness of a monitoring system or data collection strategy, quality of data and quality of services that fits the user's defined requirements, under specified operational conditions (represented by the Challenges).

---

<sup>2</sup> 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

<sup>3</sup> 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

<sup>4</sup> Ref4: Random House Unabridged Dictionary, Random House Inc, 2006

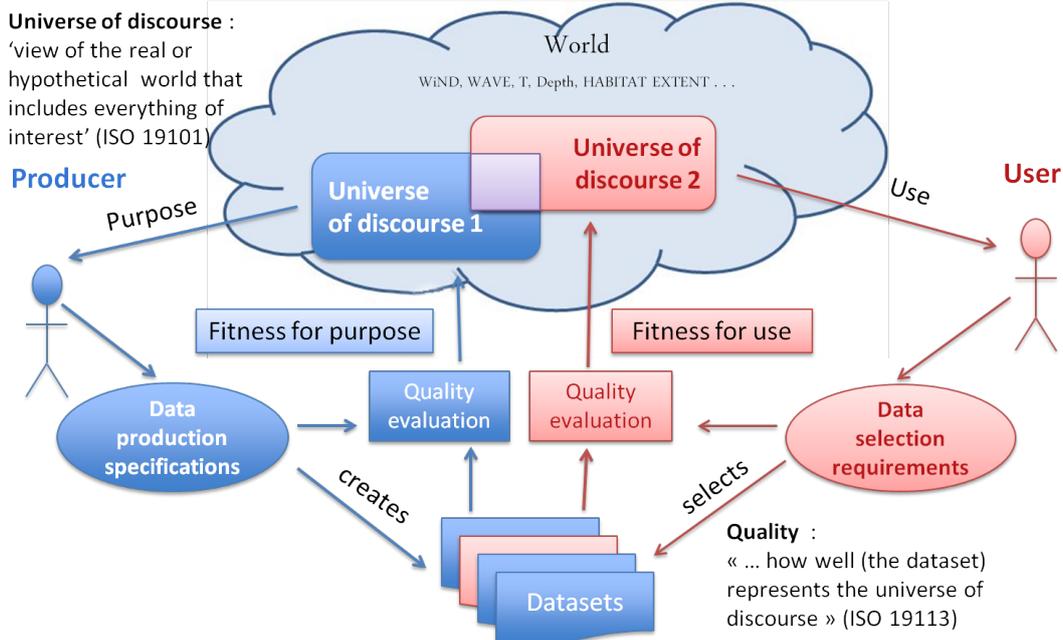


Figure 1.1. Concept behind data quality principles from the user side, as described in ISO 19113 [Ref 5] <sup>5</sup>

*Fitness for use* is different from *fitness for purpose*: *purpose* describes the rationale for creating a dataset, while *use* describes the rationale for selecting a dataset. The producer translates the purpose into production specifications, while the user translates his / her user requirements into selection specifications which may differ from the production specifications.

The *fitness for purpose* is evaluated by the producer according to the specifications of the quality expected for his / her purpose. The *fitness for use* of a dataset is evaluated by the user according to the specifications of the quality expected for his / her use. The purpose of the CheckPoint is to provide an evaluation of the *fitness for use* of the datasets used by the Challenges.

The ISO 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 are presented again in Annex 1.

### 1.3 ISO Methodology

The "Methodology to assess and communicate the economic benefits of consensus-based standards" [Ref. 6]<sup>6</sup> developed by ISO proposes a general

Fitness for  
use  
and  
fitness for  
purpose

<sup>5</sup> Ref5: <http://inspire.ec.europa.eu/glossary/UniverseOfDiscourse>

framework for the objective assessment of the quality of the monitoring system and allows results to be shared, thus supporting benchmarking and identification of general trends.

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 meet requirements in order to enhance user satisfaction.

ISO  
methodology

The ISO methodology provides a set of standards that objectively measure the 'data adequacy' and was chosen by the CheckPoint as the basic methodology.

#### 1.4 CheckPoint target audience

CheckPoint users were categorized in order to better define what the CheckPoint is and the main target audience. They are:

##### - Institutional audience

Definition: Policy makers on observations and monitoring (Integrated Maritime Policy, Environmental directives, Regional convention on the sea)

Rational: The CheckPoint indicators provide support for decision making in observations and monitoring. It is a CheckPoint primary function.

Needs: Requires **the development of indicators to identify gaps and priorities (sectorial and focus status)**.

Target  
audience:  
institutional  
audience and  
upstream  
providers are  
the main  
actors

##### - Upstream providers or producers

Definition: Data producers or providers of the datasets required for the Challenges (e.g. Copernicus marine service, EMODnet in situ data, Fishery data collection framework, EU and national Research infrastructure and projects and blue growth applications)

Rational: The CheckPoint indicators depict the fitness for use of the producer or providers outputs. This is a primary function of CheckPoint.

Needs: Requires **the development of indicators for each upstream data to give indications on how to evolve them in a user-oriented manner**

##### - Intermediate users

Definition: Public or private sector developing downstream added value products

Rational: The CheckPoint targeted products could serve as reference information for building products from a complex input dataset

<sup>6</sup> Ref6: Assessing economic benefits of consensus-based standards – The ISO methodology.  
<http://www.iso.org/iso/home/standards/benefitsofstandards/benefits-detail.htm?emid=6>

landscape. This is a secondary function of CheckPoint.

Needs: Requires **the development of a GIS catalogue to present the targeted products as reference information with dependencies on upstream data**

- **End-users**

Definition: All other users interested in getting a regional status on monitoring systems and their quality.

Rational: The CheckPoint metadatabase enables different upstream data to be discovered and to understand their fitness for use. This is a secondary function of CheckPoint.

Needs: Requires **the development of CheckPoint services to search for upstream data and develop gateway to upstream catalogues.**

- **General audience**

Definition: All users interested in knowing the general monitoring system landscape and possible use for Challenges.

Rational: The Checkportal explains principles, place and rational of CheckPoint. This is a secondary function of CheckPoint.

Needs: Presenting **the project, informing regularly on deliverables and providing showcases** (demonstration of applications)

The broad categorization of users versus CheckPoint functions is overviewed in Table 1.1.

CheckPoint functions / user categories	Institutional audience	Upstream providers	Intermediate users	End users	General audience
Project deliverables and information	x	X	x	x	Main Driver
CheckPoint indicators to identify gaps and priorities	Main Driver	X	x	x	x
CheckPoint indicators to illustrate relationship between targeted products and input datasets	x	Main Driver	x	x	
GIS catalogue to present the targeted products as reference information with dependencies on upstream data	x	x	Main Driver	x	
search for upstream data to clarify the observation landscape (neutral status), gateway to upstream catalogues	x	x	x	Main Driver	x

Target audience: institutional audience and upstream providers are the main actors

Table 1.1 CheckPoint functions vs user categories for the Mediterranean CheckPoint. Blue indicates the two targeted CheckPoint users.

The conclusion is that the target audience for the CheckPoint activities is made up of the Institutional audience and the upstream providers/producers.

The main objective of CheckPoint is to continuously manage CheckPoint information in order to describe input data that are usable and used by Challenges, **to assess the quality and fitness for use** of data and data sets and provide this assessment to those targeted user categories. The CheckPoint is not aimed at providing direct access to data and data sets, as this is the role of the upstream provider catalogues.

### 1.5 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 (§7-8-9-10).

Structure of  
the first DAR

Discussion and conclusions (§11) conclude the document.

## 2 CheckPoint assessment methodology

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:

CheckPoint  
Information  
and  
CheckPoint  
Service

- 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.

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

Assessment  
Territories

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 and Annex 1 explains them.

Territory 1: Appropriateness	Territory 2: Availability
<ul style="list-style-type: none"> <li>○ <b>Spatial information</b> <ul style="list-style-type: none"> <li>▪ <b>Extent</b></li> <li>▪ <b>Resolution</b></li> </ul> </li> <li>○ <b>Time information</b> <ul style="list-style-type: none"> <li>▪ <b>Extent</b></li> <li>▪ <b>Resolution</b></li> </ul> </li> <li>○ <b>Purpose</b></li> <li>○ <b>Lineage</b></li> <li>○ <b>Usage</b></li> <li>○ <b>Completeness</b></li> <li>○ <b>(logical) Consistency</b></li> <li>○ <b>Accuracy</b> <ul style="list-style-type: none"> <li>▪ <b>Horizontal</b></li> <li>▪ <b>Vertical</b></li> <li>▪ <b>Temporal</b></li> <li>▪ <b>Thematic</b></li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>✓ <b>Visibility</b></li> <li>✓ <b>Accessibility</b> <ul style="list-style-type: none"> <li>○ <b>Service</b></li> <li>○ <b>Data policy</b></li> <li>○ <b>Pricing policy</b></li> <li>○ <b>Formats</b></li> <li>○ <b>Service interoperability</b></li> </ul> </li> <li>✓ <b>Performance</b> <ul style="list-style-type: none"> <li>○ <b>Responsiveness</b></li> <li>○ <b>Reliability</b></li> </ul> </li> </ul>

Assessment  
Territories and  
their elements

Table 2.1 Appropriateness and Availability assessment criteria

However each assessment criterion must be given a different emphasis in the various Challenges. For example, 'Accuracy' may be more relevant for some Challenges than for others. This means that once the assessment methodology has been put in place, the final assessment must be made by challenge experts.

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, information was 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 Literature Survey provided the initial list of categories of characteristics and data set descriptors which were included in a ‘CheckPoint Metadatabase’ contained in an e-infrastructure based on the Catalog Services for Web (CSW) framework. The CheckPoint metadatabase contains the metadata that link the various providers of upstream data to the selected categories of characteristics of the Challenges. The Metadata format and model is explained in Annex 2.

From  
Literature  
Survey to  
Data  
Adequacy  
Report

For the CSW and in order to store any targeted products for the Challenges, the Consortium decided to use the GIS technology called ‘Sextant’ [Ref 7]<sup>7</sup>. 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, GeoSeas.

Sextant and  
GIS  
Technology

For the purposes 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 providers, the Sextant monitoring data catalogue will use the SeaDataNet Common Vocabularies (<http://www.seadatanet.org/>), the European Directory of Marine Organisations

Sextant  
catalogue

---

<sup>7</sup> Ref7: <http://www.ifremer.fr/sextant>

(EDMO), the European Directory of Marine Environmental Research Projects (EDMERP).

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

CheckPoint  
information  
for the  
metadatabase

1. Characteristics (= What)
2. Data provider (= From/Who)
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 data provider here is intended both as primary provider, i.e. data originator, and secondary data provider, i.e. data distributor. The 2014 version of ISO19115 provides a list of the different categories of data providers:

Data provider  
definition

1. Originator : “party who created the resource “ (i.e. dataset)
2. Distributor : “party who distributes the resources”
3. Custodian “party that accepts accountability and responsibility for the resource and ensures appropriate care and maintenance of the resource”

In the present metadatabase, we have not differentiated between the two-three types of providers however in the future we will add an additional subdivision that will identify primary (originator) and secondary data providers (distributor and custodian). The lack of such subdivision makes it difficult to find overlapping data sets, which could be produced once and distributed many times through different portals.

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

The metadatabase follows 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.

CheckPoint  
Service  
components

In principle, CSW should allow the metadata to be harvested. However, harvesting is not yet possible in many cases: no catalogue for the input data set, the granularity of dataset descriptions is not appropriate, optional metadata not completed, etc. To include all the necessary information in the database, in some cases metadata were compiled manually, with an iterative process that was consolidated to a certain extent for the first DAR.

Consolidation  
of the  
Literature  
Survey in the  
DAR

In the Literature Survey we defined the assessment criteria, while in this first DAR we built the metadatabase and the indicators based upon the methodological approach of the Literature Survey.

### 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).

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.

SeaDataNet  
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.

MedSea CheckPoint decided to use the SeaDataNet Vocabularies that adopts a hierarchical approach for the classification of terms, from disciplines (P03), to

SeaDataNet  
P01, P02, P03  
vocabularies

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 provider 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 set descriptions in the MedSea CheckPoint metadatabase. Figure 4.1 shows the number of data sets required by each Challenge.

298 data set  
descriptions  
required by  
Challenges

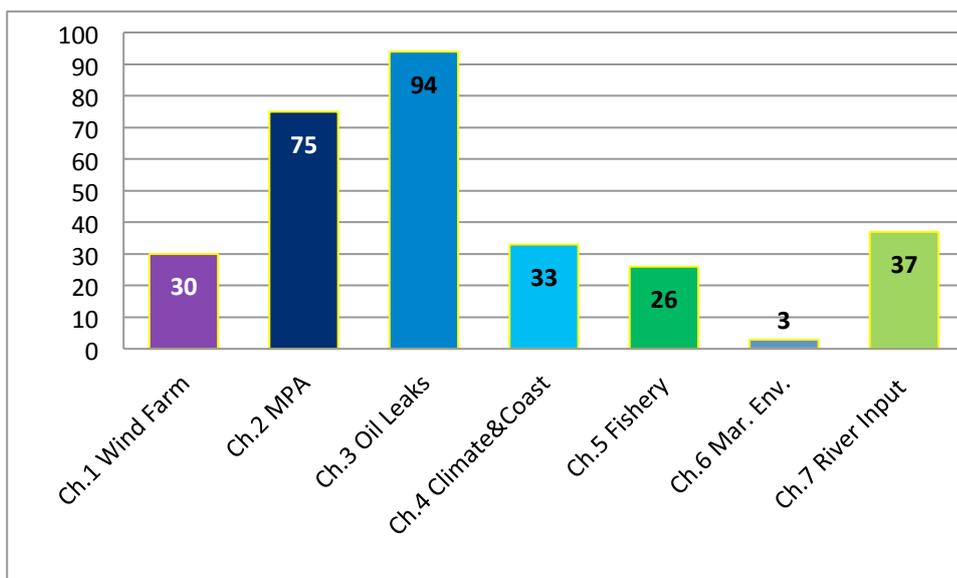


Figure 4.1 Number of descriptions of data sets for each Challenge (total is 298)

The Challenge categories of 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.

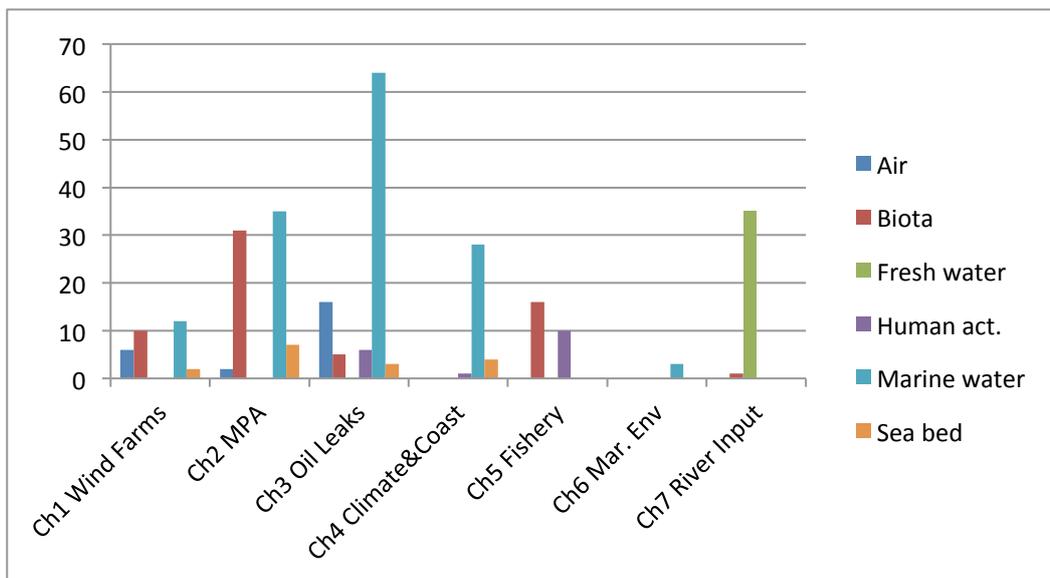
To date there are **45** categories of characteristics considered for all CheckPoint Challenges. Figure 4.2 shows the number of categories of characteristics required by each challenge as a function of the environmental matrix .

45 overall  
categories of  
characteristics

The list of the 298 data set descriptions, associated with the 45 categories of characteristics is provided in Annex 3. Note that some important data set descriptions are still missing from the list of 298, e.g. oxygen data for Challenge 6 (Marine Environment) . This depends partially on the expertise of the Challenge Team and the difficulty in finding an adequate data set. The completeness of the metadatabase will be improved in the second part of the project when Challenge targeted products will have to be produced.

The 298 data sets are associated with 126 data providers that were selected by the project experts to be the desirable upstream data inputs to the Challenges.

126 data providers



Categories of characteristics by Challenge and environmental matrix

Figure 4.2 Number of categories of characteristics by challenge and matrix

The same categories of characteristics are required by the various Challenges from different data providers. The list of **126** potential data providers (primary and secondary) is provided in Annex 4. Table 4.1 shows the categories of potential data providers, subdivided into six, 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 providers (total 126) for broad categories (details are in Annex 4)

#### 4.1 Data set identification issues

The list of categories of characteristics and data providers (primary and secondary) was defined within each challenge on the basis of the partners' knowledge and partners' communities in order to fit the user needs for each Challenge. In other

Providers, data duplications and best copies

words, the content of the CheckPoint metadatabase is strongly linked to the specific Challenges chosen by the DGMARE call for tender and expert opinion. In the future, a wider consultation will be set up to improve the alternative data providers and data sets.

Many databases are undergoing major updates while the CheckPoint metadatabase building is in progress. For example, EMODnet Geology Portal at the moment has no information on the Mediterranean Sea but it soon will have and the metadatabase should be updated. The metadatabase needs to be updated yearly to have a view of the monitoring system at the basin scale.

Different providers may make the same data sets or data set series available as different releases and without a specification. As a consequence there is a high risk of including in the CheckPoint metadatabase either duplicates or apparent duplicates with different availability conditions.

Challenges have collected information on providers by selecting the 'best copies' of data sets, i.e. the ones with the most appropriate quality for their needs and trying to avoid duplications. However, the same data sets could have been uploaded into the Med CheckPoint metadatabase by different Challenges. To reduce these problems, the metadatabase is checked initially by each challenge leader and subsequently by 'auditors' not working in the Challenges, a process that only started in this first DAR.

It is estimated that about 15% of the 298 data set descriptors currently refer to the same input data set.

## 5 CheckPoint indicator definitions

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 (e.g. the horizontal accuracy of the data set either described in the producer metadata or measured by the user himself).

CheckPoint  
indicator  
definition

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.

### What are Checkpoint indicators?

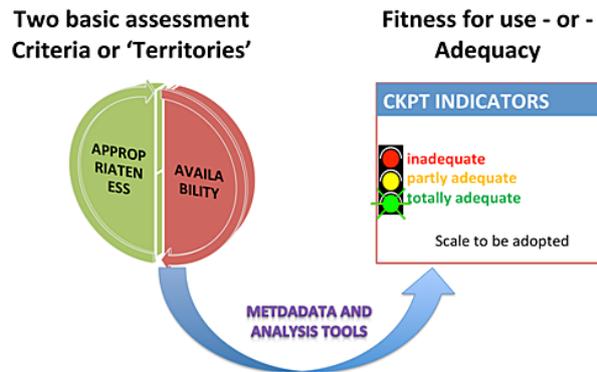


Figure 5.1 The two territory assessment criteria and the indicators

#### 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 below.

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: <ul style="list-style-type: none"> <li>• Feasible:</li> <li>• Explained (i.e. the reason for)</li> <li>• Reproducible:</li> </ul> Because monitoring categories of 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

SMART  
method for  
indicators  
definition

		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: <ul style="list-style-type: none"> <li>• First date of value</li> <li>• Update frequency</li> <li>• Action list: working methods and how to improve the indicator value</li> </ul>

## 5.2 Indicator Evaluation Scale

Indicator  
evaluation  
scale

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 is to find an appropriate balance between simplification and completeness. They offer however an 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 until the second DAR.

Visual  
representation  
of indicators

To provide an easy and straightforward visual representation of indicators, allowing a non expert to easily assess the fitness for use without looking at the metadata and reports, it was decided to use three colours:

- Red: urgent actions are required to provide datasets and services fitting for use
- Yellow: limited actions are required to provide datasets and services fitting for use
- Green: actions and services are fit for use and should be maintained

Red light  
Yellow light  
Green light

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 data provider (visibility), how to access them (accessibility), and how fast the process is to take possession of them (performance).

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:

- 1) Visibility (VI), i.e. the possibility of identifying and quickly accessing the appropriate site for the required data sets;
- 2) Accessibility (AC) i.e. the possibility, for non expert users, to understand the retrieval model status;
- 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.

Availability indicators:

Visibility

Accessibility

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.:

<b>AV-VI-1</b> Easily found	<b>Can the data sets or series of data sets be found easily?</b>
<b>AV-VI-2</b> 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 <sup>8</sup> (Ref.8))

The methodology used to compute the visibility indicators for the first DAR is explained in the following table

Indicator name	Meaning (INSPIRE)	Meaningful (Symbol)	Achievable & Realistic (Choice)
<b>AV-VI-1</b> <b>Easily found</b>	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	 Low visibility	Choice 1: Red <i>"Cited in peer reviewed paper or grey literature but no info on how to access"</i>  Choice 2 : Red <i>"Information retrieved upon specific request to the data source "</i>

Availability –  
Visibility  
Indicators

<sup>8</sup> Note1: advanced services, but not following the guidelines defined in INSPIRE and its technical annexes.

AV-VI-2 EU Inspire catalogue service	and EN ISO 19119	 Medium visibility	Choice 3: Yellow <i>"Use of social network, community of practices sharing information, portals of organization where no search is organized by an engine"</i>
		 High visibility	Choice 4: Green <i>"Use of open search engines, searching by name either the data provider or the characteristics"</i>
			Choice 5: Green <i>"Search via reference catalogue (e.g. MyOcean, GEOSS Geoportal...)"</i>
	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).	 Inadequate	Choice 1: Red <i>"Data sets are not referenced in a catalogue or are referenced in a non public catalogue"</i>
		 Partly adequate	Choice 2: Yellow <i>"The datasets are referenced in a public national catalogue, in an international catalogue service "</i>
	Part B 1.4 of Inspire Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119	 totally adequate	Choice 3: Green <i>"The datasets provide a full EU Inspire catalogue service "</i>

Availability –  
Visibility  
Indicators

## 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	<b>Visibility on data policy adopted by data providers.</b>
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 the following table.

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	Choice 1: Red <i>"There is no information at all on data policy adopted by data providers"</i>
		 Medium transparency	Choice 2: Yellow <i>"There is information, but details are available only on request"</i>
		 High transparency	Choice 3: Green <i>"There is detailed information provided to understand data policy"</i>
AV-AC-2 Delivery mechanism	The indicator reports the type of services available	 No information	Choice 1: Red <i>"No information was found on data delivery"</i>

Availability –  
Accessibility  
Indicators

	<p>to the user to access data</p> <p>Part B 2.2 of INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119</p>	<p> Manual</p> <p> Partial Inspire function</p> <p> Full Inspire function</p>	<p><i>mechanisms"</i></p> <p>Choice 2: Red <i>"Order form/invoice is requested"</i></p> <p>Choice 3: Yellow <i>"Online downloading services "</i></p> <p>Choice 4: Green <i>"Online discovery and downloading services"</i></p> <p>Choice 5: Green <i>"Online discovery + downloading + viewing services (Advanced services)"</i></p>
<p><b>AV-AC-3</b> <b>Data policy</b></p>	<p>The indicator comments on the status of information given for data policy</p> <p>Part B 8.2 of INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119</p>	<p> Not documented</p> <p> Restricted</p> <p> Partially restricted</p> <p> Unrestricted</p>	<p>Choice 1: Red <i>" Not or not well documented"</i></p> <p>Choice 2: Red <i>"Restricted"</i></p> <p>Choice 3: Yellow <i>"Accessible under moratorium"<sup>9</sup> (Ref.9)</i></p> <p>Choice 4: Green <i>"Unrestricted"</i></p>
<p><b>AV-AC-4</b> <b>Pricing</b></p>	<p>The indicator comments on the provision of</p>	<p> Not documented</p>	<p>Choice 1: Red <i>"Not or not well documented"</i></p>

<sup>9</sup> Ref 9: 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).

AV-AC-5 Readiness	information related to cost basis	 Commercial cost	Choice 2: Red <i>"Commercial cost charge"</i>	Availability – Accessibility Indicators
	Part B 8.1 of INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119	 Charge	Choice 3: Yellow <i>"Distribution charge"</i> <i>"Collection charge"</i> <i>"Free of charge for academic institutions and uses"</i>	
AV-AC-5 Readiness	The indicator comments on the information given for distribution format (Optional in INSPIRE)	 Free	Choice 4: Green <i>"Open and Free, No charge"</i>	
		 Not documented	Choice 1: Red <i>"Not or not well documented"</i>	
		 Not ready to be consumed	Choice 2: Red <i>"Proprietary and not well documented "</i>	
		 Can be processed to be consumed	Choice 3: Red <i>"Not proprietary but content not clearly specified "</i>	
		 Ready to be consumed	Choice 4: Yellow <i>"Proprietary but content clearly specified "</i>	
			Choice 5: Green <i>"Not proprietary and content clearly specified (eg auto-descriptive eg ODV, NetCDF CF) or at least with appropriate document describing the content."</i>	

The Expert Panel suggested to change the color scale of AV-AC-4 indicator in order to highlight the difference between the choice of “no information” and “commercial cost”. In the future we will consider commercial cost as part of the “yellow” scale instead the red one.

### 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 the table below.

Indicator name	Meaning (INSPIRE)	Meaningful (Symbol)	Achievable & Realistic (Choice)
<b>AV-PE-1 Responsive ness</b>	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 <i>"No information is found on response time"</i>
		 Medium response	Choice 2: Red <i>"More than 1 week for release"</i>
		 High response	Choice 3: Yellow <i>"Less or equal to 1 week for release"</i>
			Choice 4: Green <i>"Online downloading (i.e. a few hours or less) for release"</i>

Availability –  
Performance  
Indicator

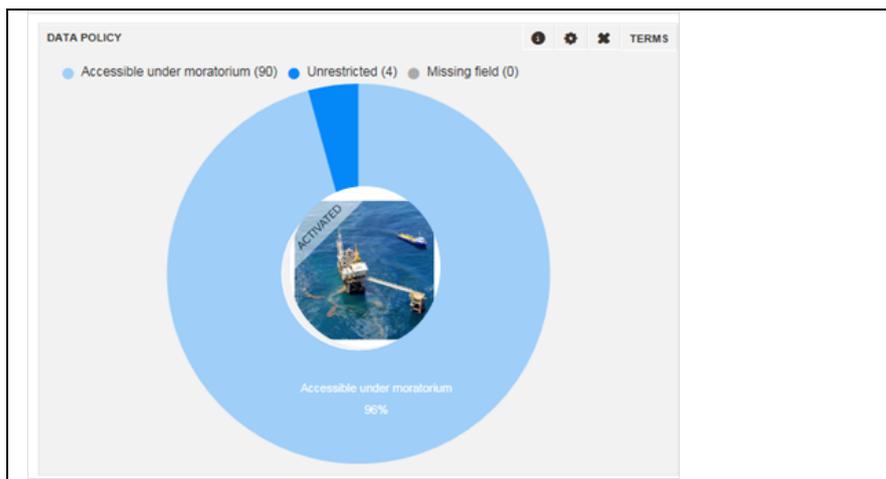
## 6.4 Summary of indicators

In summary, the eight availability indicators are:

<b>AV-VI-1</b> Easily found	Can the data sets be easily found?
<b>AV-VI-2</b> EU Inspire catalogue service	Is the service catalogue EU Inspire compliant?
<b>AV-AC-1</b> Policy visibility	Visibility on data policy adopted by data providers.
<b>AV-AC-2</b> Delivery	Data delivery mechanisms
<b>AV-AC-3</b> Data Policy	Data policy
<b>AV-AC-4</b> Pricing	Cost basis / price policy
<b>AV-AC-5</b> Readiness	Format for use
<b>AV-PE-1</b> 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 - Data Policy strategies as described by Challenge 3 (upper panel) or by all Challenges together (lower panel) as it 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

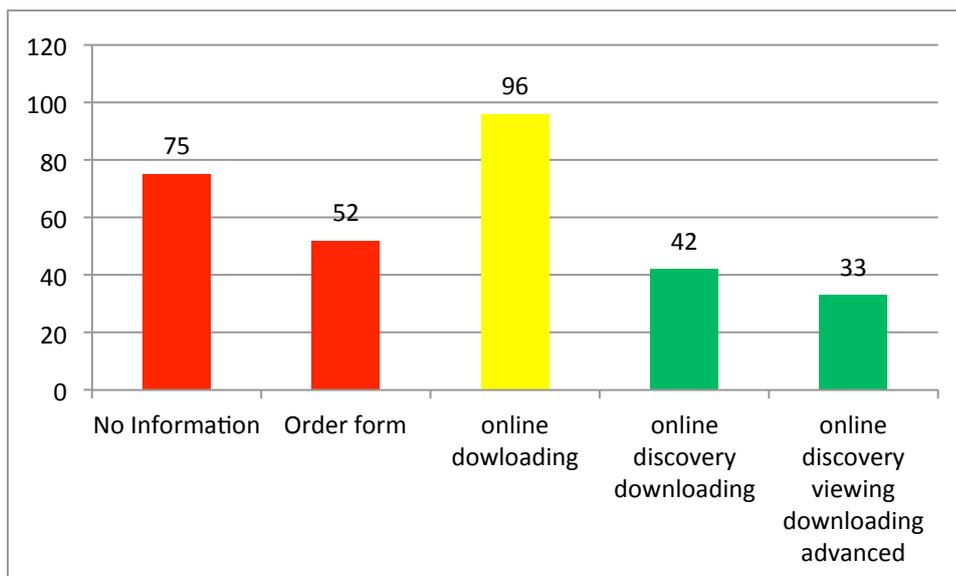


Fig. 7.1 AV-AC-2 Indicator on quality of delivery mechanisms as a function of indicator values and for all Challenges

Challenge	 No information	 Order form	 Online downloading	 Online discovery+ downloading	 Online discovery+ viewing + downloading advanced	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
<b>Total</b>	<b>75</b>	<b>52</b>	<b>96</b>	<b>42</b>	<b>33</b>	<b>298</b>
<b>%</b>	<b>26</b>	<b>17</b>	<b>32</b>	<b>14</b>	<b>11</b>	<b>100</b>

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.

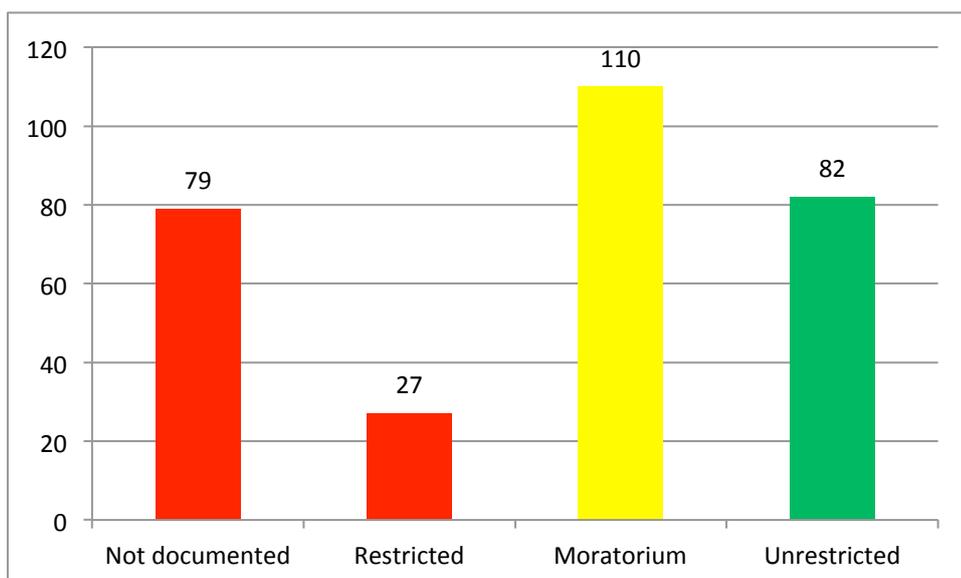


Fig. 7.2 AV-AC-3 indicator on data policy for all Challenges as a function of

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
<b>Total</b>	<b>79</b>	<b>27</b>	<b>110</b>	<b>82</b>	<b>298</b>
<b>%</b>	<b>27</b>	<b>9</b>	<b>37</b>	<b>27</b>	<b>100</b>

Availability –  
Accessibility  
Indicators for  
all Challenges

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

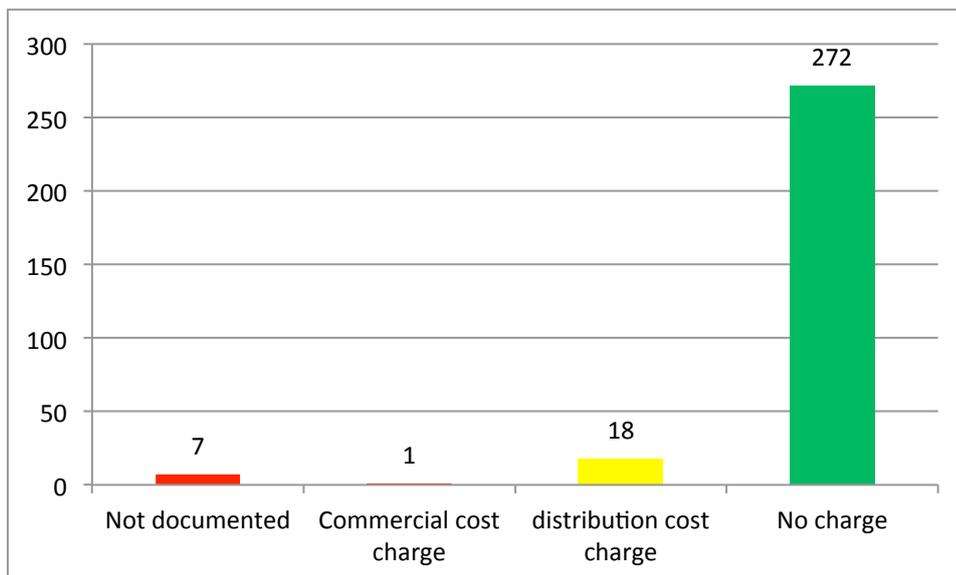


Fig. 7.3 AV-AC-4 indicator on pricing policy for all Challenges as a function of indicator meaning

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
<b>Total</b>	<b>7</b>	<b>1</b>	<b>18</b>	<b>272</b>	<b>298</b>
<b>%</b>	<b>2</b>	<b>0</b>	<b>6</b>	<b>92</b>	<b>100</b>

Availability –  
Accessibility  
Indicators for  
all Challenges

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

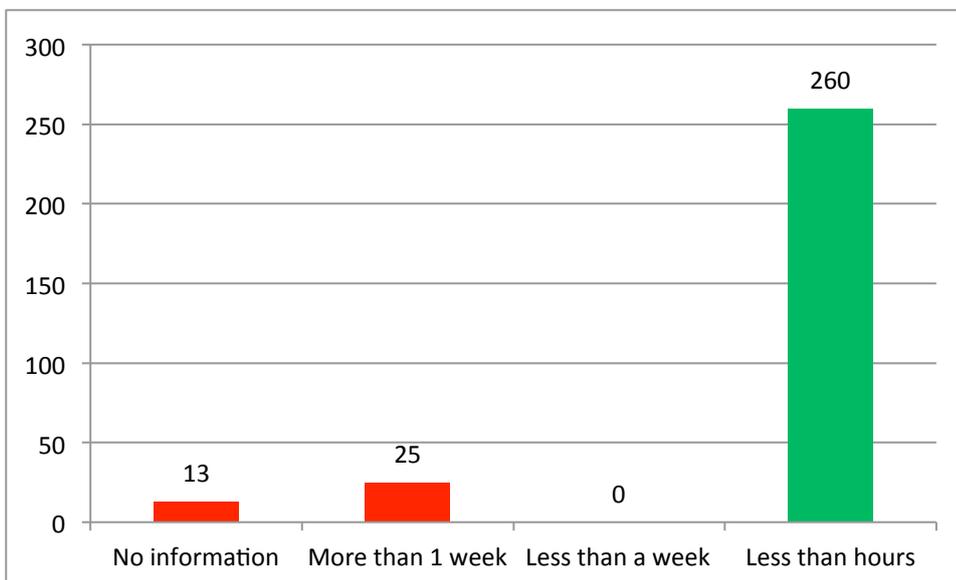


Figure 7.4 AV-PE-1 indicator on database responsiveness in terms of delivery for all Challenges as a function of indicator meaning

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
<b>Total</b>	<b>13</b>	<b>25</b>		<b>260</b>	<b>298</b>
<b>%</b>	<b>5</b>	<b>8</b>		<b>87</b>	<b>100</b>

Availability –  
Performance  
Indicator for  
all Challenges

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

Clearly, most of the data delivery is declared as being high. However, this will be clarified when the data sets are actually downloaded for the targeted Challenge products. For Fisheries (CH5), response time is longer 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.

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

Synthesis of  
Indicators for  
all Challenges

Table 7.6 Summary of overall indicators for all Challenges

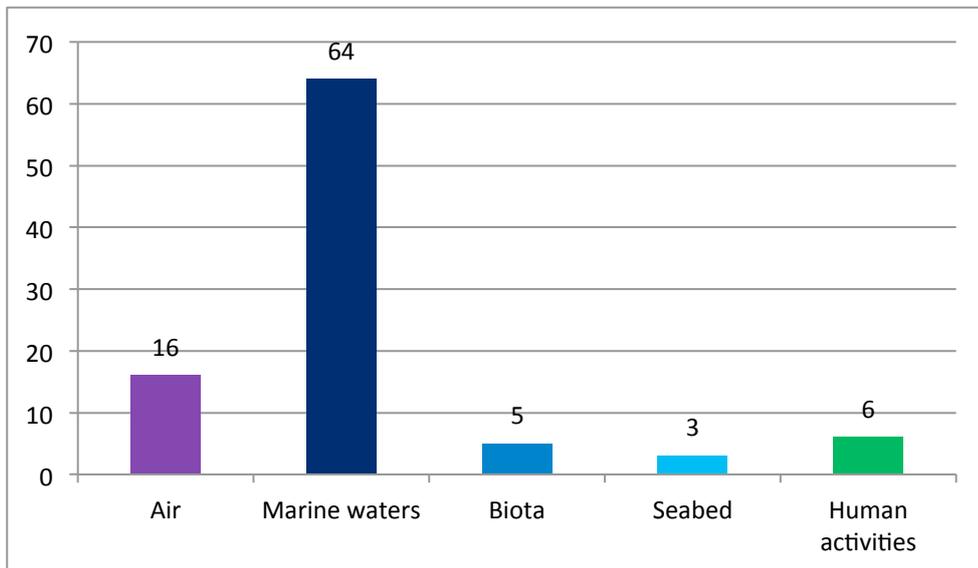
We expect this result to change slightly when the metadatabase is further checked and consolidated over the next few months and when the analysis is performed by categories of characteristics or thematically rather than by Challenge.

## 8 The indicators Results: Oil Platform Leak Challenge

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 categories of characteristics identified in the Challenge Oil Platform Leaks. They are categorized in the environmental matrices as in Figure 8.1



Oil Platform  
Leaks  
Challenge  
datasets

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 categories of 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) Water temperature TEMPPR01

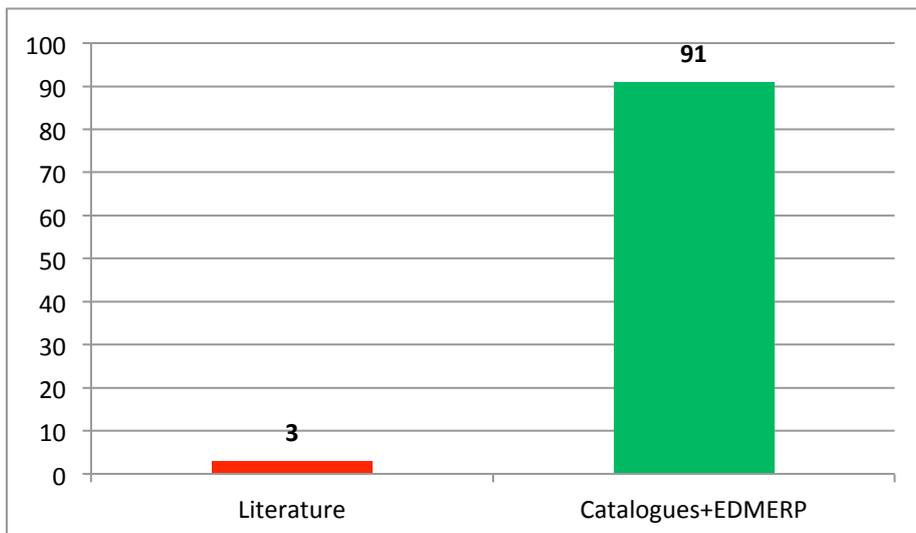
	column)	
	GWDR (Wave direction)	Mean wave direction GWDRZZ01
	WVST (Wave height and period statistic)	Average zero crossing period of waves {Tz} on the water body GTZAZZ01
		Significant height of waves {Hs} on the water body GTDHZZ01
Biota	HBEX (Habitat extent)	Marine protected areas (polygon) – no P01 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 - Pollution event category	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 categories of characteristics of Oil Platform Leaks

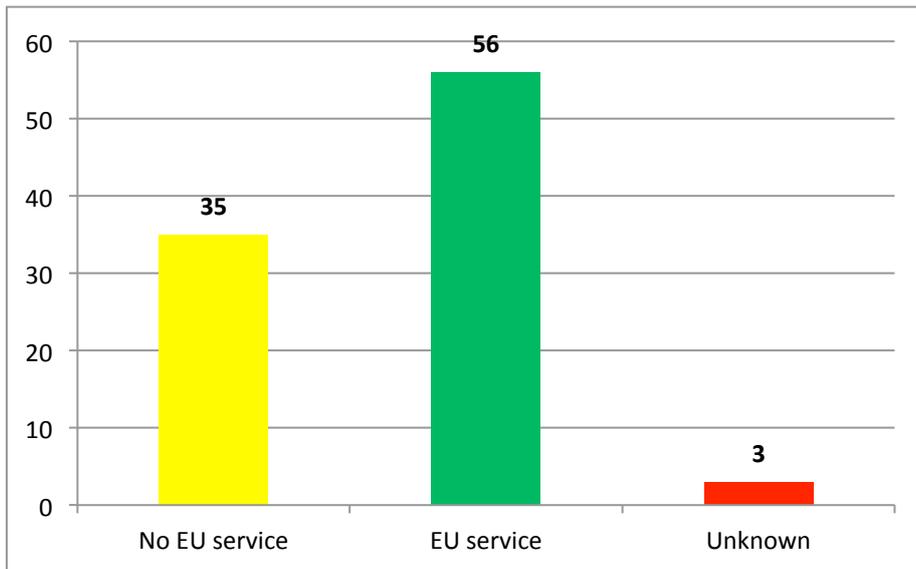
### 8.1 Availability – Visibility indicators

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).



Availability – Visibility Indicators for Oil Platform leaks

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



Availability –  
Visibility  
Indicators for  
Oil Platform  
leaks

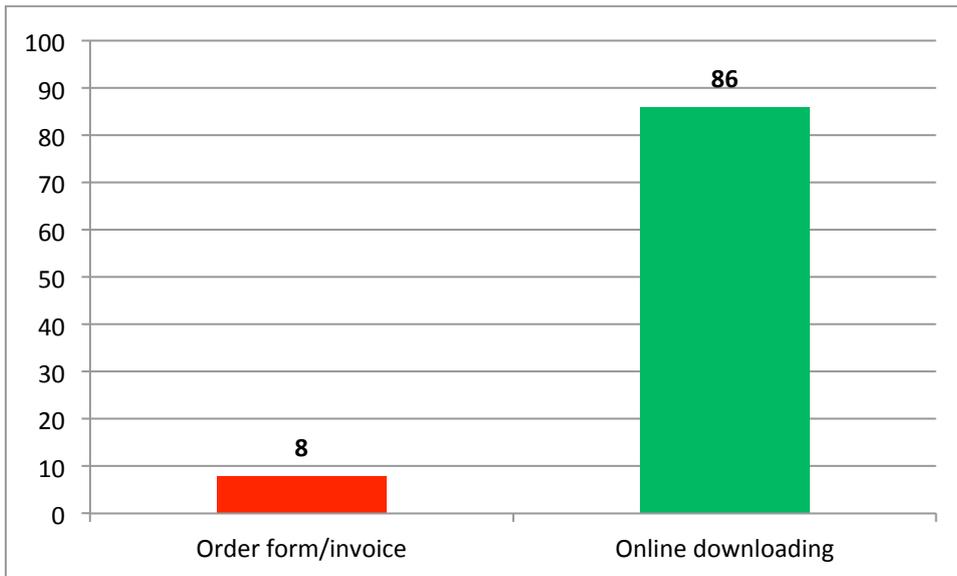
Figure 8.3 AV-VI-2 indicator on availability: datasets referenced in catalogue services (Challenge 3)

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 providers 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.

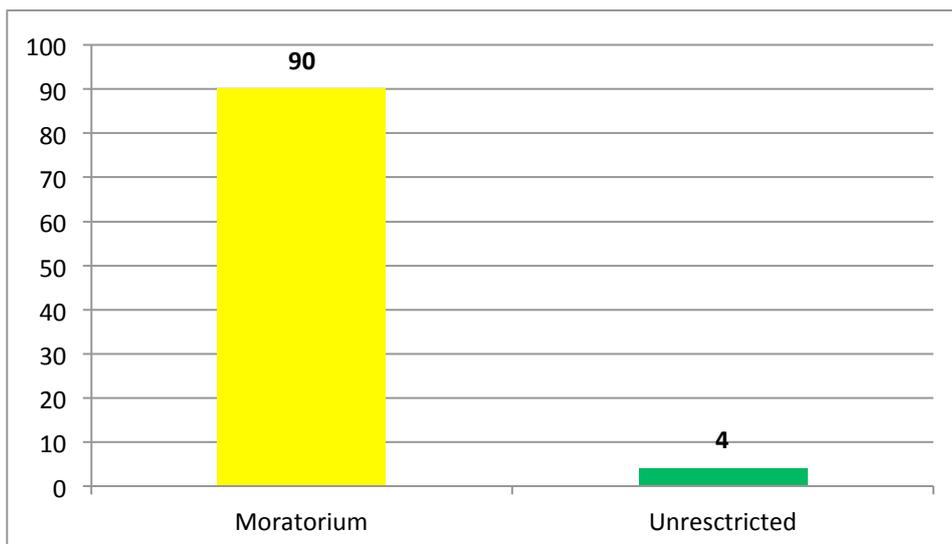


Availability –  
Accessibility  
Indicators for  
Oil Platform  
Leaks

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

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.



Availability –  
Accessibility  
Indicators for  
Oil Platform  
Leaks

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

Availability –  
Performance  
Indicator for  
Oil Platform  
Leaks

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	 Partial 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 Initial assessment of the EMODnet Portal data sets and MyOcean service

The main principles on which the EMODnet portals are built are:

1. Collect data once and use it many times;
2. Develop standards across disciplines as well as within them;
3. Process and validate data at different levels. Structures are already developing at national level but an infrastructure at sea-basin and European level is needed;
4. Develop a user-driven decision-making process for priorities ;
5. Accompany data with statements on ownership, accuracy and precision, and
6. Recognize that marine data is a public good, and discourage cost-recovery pricing from public bodies.

Initial  
assessment of  
EMODnet  
Portals and  
MyOcean  
Service

MyOcean has been designed to prepare and to lead the demonstration phases of the future Copernicus Marine Environment Monitoring Service. MyOcean products and services are meant to serve all marine applications: Marine resources, Maritime safety, Coastal and Marine Environment, Seasonal Forecast & Climate with products in Real Time, in particular value added products from earth monitoring satellites and marine oceanographic forecasts.

The application of these principles is substantiated by the initial assessment.

**AV-VI-1**  
**Easily found**

Can the data sets or series of data sets be found easily?



The EMODnet and MyOcean portals have been presented in many meetings and conferences (e.g. European Geosciences Union – Earth and Space Science Informatics – Informatics in Oceanography and Ocean Science) and are part of a community of practices sharing information. The existing reference catalogues make the data sets search easy. However no unique identifier such as a DOI exists and this reduces the possibility of citing and searching datasets using a common way (Emodnet bathymetry is creating a DOI for the European seas DTM).

Availability –  
Visibility  
indicators for  
EMODnet  
Portals and  
MyOcean  
Service

**AV-VI-2**  
**EU Inspire  
Catalogue Service**

Is the dataset referenced in a EU or non-EU catalogue service?



The portals provide Inspire Services.

**AV-AC-1**  
**Policy Visibility**

Visibility on data policy adopted by data providers



Bathymetry does not require registration to access the EMODNET DTM product. It is downloadable directly from the Emodnet bathymetry site. Only background data used to create the DTM need a registration using the SeaDataNet registration rules.

Geology uses a web delivery mechanism with open source standards to ensure the long-term sustainability and optimal delivery of the best available and up-to-date marine geological information.

SeaBed Habitats is providing a page with terms and conditions and privacy policy.

Chemistry is providing this information: *The CDI data discovery including its services for authentication accessing, and downloading is driven by the underlying central SeaDataNet database and infrastructure of connected data centres. Users can freely search and browse the CDI discovery services and identify relevant data sets.*

Biology does not expressly state the data policy in pages, although it is clear that the data policy is for an open and sea access.

Availability –  
Accessibility  
indicators for  
EMODnet  
Portals and  
MyOcean  
Service

Physics is stating: *latest 60 days are freely viewable and downloadable while the access to older data (monthly archives) request credentials. Archived data series and metadata are organized according and in collaboration with NODCs network.*

Human Activities is declaring that they are *seeking as much as possible to provide web-services that extract data from their databases, rather than develop one large database.* The easiest interpretation is that the policy visibility is relying on information contained in metadata.

MyOcean is indicating in the home page the need to register before accessing products.

**AV-AC-2**  
**Delivery mechanisms**

Data delivery mechanisms, i.e. the services available to the user to access data



Availability –  
Accessibility  
indicators for  
EMODnet  
Portals and  
MyOcean  
Service

All the portals have an on-line delivery mechanism and most of them have an advanced system ISO, OGC and INSPIRE compliant.

**AV-AC-3**  
**Data policy**

Data policy



All the portals declare that the data are unrestricted. However, the users that want to access SeaDataNet data must agree on SeaDataNet licence, which means that users must agree on citing the providers. This entails the user being authorised and authenticated before being allowed to download the data.

The data policy is found in pages that are different from one EMODnet portal to another, e.g. in **Metadata and Data** for Bathymetry, in **About** for Geology, in **Welcome** for SeaBed Habitats, in **Data Access** for Chemistry, in **Home** for Physics, in **About** for Human Activities.

**AV-AC-4**  
**Pricing**

Cost basis/price policy



Availability –  
Accessibility  
indicators for  
EMODnet  
Portals and  
MyOcean  
Service

All data are free of charge. For those portals that follow SeaDataNet data policy, the data are open after registration and access is through authorisation and authentication.

**AV-AC-5**  
**Readiness**

Format for use



All the portals provide data in well-known formats that can be selected by the user.

**AV-PE-1**  
**Responsiveness**

How responsive is the delivery service for the available data?



Availability –  
Performance  
indicator for  
EMODnet  
Portals and  
MyOceans  
Service

All the portals deliver data within 24 hours.

The overall list of indicator values as a function of Challenge and EMODnet Portal and MyOcean Service is given in Annex 5.

## 10 Summary of the lessons learned by each Challenge

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 providers, gaps in the input data, most useful data providers, trends in availability (improving, worsening), usefulness of secondary data providers (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 providers accurate enough for the Challenge products?
- Q2: Where are the main gaps in the input data?
- Q3: (If there are different providers 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 providers 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 providers 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 providers. 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 providers, especially at a national level.

### **Q3: Which data source is the most useful?**

The most useful data providers are for example the most available and high resolution providers for meteo-oceanographic condition datasets.

The use of multiple data providers often enables gaps to be filled such as between tide gauges and sea level altimetry to meet sea level rises. Multiple data providers 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 specific R-software (VMSbase and VMStools) for the analysis of VMS and AIS datasets.

In general there is the idea that there is no need to access data from original providers. 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 for free
- Information provided by data providers 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.

Given the MedSea CheckPoint experience in the compilation of the metadatabase, incorporating descriptive metadata would be useful, following the example of MyOcean targeted products. Such products 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 ISO standards for geographic information for the description of datasets as used by MyOcean and Seadatanet combined with the use of common vocabularies for the classification of the characteristics and the discovery of the corresponding datasets would simplify the data collection for the CheckPoints.

## 11 Discussion and conclusions

The first Data Adequacy Report contains results related to:

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.

Synthesis:  
three main  
DAR results

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.

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. The extension of the vocabulary is a basic

Vocabulary is  
not large  
enough to  
contain all  
characteristics

problem that should be addressed in future projects.

Data originate from different providers (research centers, private companies, public agencies, etc.) and are accessed by different users (public administrations, decision-makers, citizens, etc.) for different uses (research, government, civil protection, etc.). A total of 298 data set descriptors are now in the metadatabase that refer to input data sets that will be used to produce Challenge targeted products. They have been provided by 126 different data providers.

298 data set  
descriptors  
and 126 data  
providers for  
the Challenge  
characteristics

The 45 different categories of 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 categories of 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.

45  
Characteristics  
needed for  
Challenges

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 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  
availability  
indicators

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	 <b>Partial Inspire function</b>
AV-AC-3	Data policy	 <b>Partially restricted</b>
AV-AC-4	Pricing	 <b>Free</b>
AV-PE-1	Responsiveness	 <b>High</b>

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

First DAR  
results :  
indicators for  
Challenge 3

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

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.

The assessment of the EMODnet and MyOcean services is very positive. However the analysis of the EMODnet Thematic Portals is not complete and will be updated in the near future because data availability will change rapidly. This is the case of EMODnet Geology portal that is not chosen as provider by the Challenges because it does not currently provide any data in Mediterranean.

First DAR  
results :  
indicators for  
EMODnet and  
MyOcean

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 smallest number of Inspire catalogues and least data policy transparency.

General  
conclusions

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 categories of 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.

## 12 Annex 1 Assessment criteria

The criteria are defined in the Literature Survey and are reviewed here. They focus on two questions : “what” is made available to the Challenges and “how”? They are derived from the quality principles of the ISO19113 standard for geographical information which concerns 99% of the data needed by the Challenges and additional criteria related to services recognized in the literature as key elements for the establishment of the fitness for use.

“What” is also defined as ‘**appropriateness**’ and includes :

- ✓ **Spatial extent** covered by the dataset or the collection  
Box or geographic region bounding the datasets
- ✓ **Spatial resolution** :  
Size of the smallest object that can be resolved on the ground. In a raster dataset, the resolution is limited by the cell size.
- ✓ **Time extent** covered by the dataset or the collection:  
Time interval represented by the dataset or by the collection.
- ✓ **Time resolution**  
Size of the smallest interval of time that can be resolved.
- ✓ **Purpose** (ISO19113 overview element of quality)  
Describes the rationale for creating a dataset (product specification)
- ✓ **Lineage** (ISO19113 overview element of quality)  
Describes the history of a dataset from collection and acquisition (source information) through compilation and derivation to its current form (process steps)
- ✓ **Usage**  
Describes the application(s) for which a dataset has been used by the data producer or by other, distinct, data users.
- ✓ **Completeness** (ISO19113 quality element)  
Degree of absence or of excess of data in a dataset
- ✓ **Consistency** (ISO19113 quality element)  
Degree of adherence to rules (concept, value domain, format...)
- ✓ **Accuracy** (ISO19113 quality element) including
  - absolute or relative **positional accuracy** :
  - **temporal accuracy** including the correctness of the temporal references of an item and the correctness of ordered events and temporal validity
  - **thematic accuracy** including classification correctness

“How” is also referred to as ‘**availability**’ and refers to :



#### *Visibility*

Ability to identify and quickly access the appropriate site delivering the desired datasets from existing EU catalogues.



#### *Accessibility*

Conditions in which users can obtain data:

- *services* : manual ordering, discovery, downloading, advanced services
- *data policy*: restricted, accessible under moratorium, unrestricted
- *pricing policy* : no charge, at a cost, cost charge depends on intended use and category of users
- *formats* (including semantic conventions)
- *interoperability* of on-line services: (OGC standards...)



#### *Performance*

- *responsiveness* is the timeliness or ability to process a request in a deterministic and acceptable amount of time
- *reliability* i.e. the ability of the services (to request data) to keep operating over time and to operate correctly, and does not fail or reports any failure to the service user for compensation

For more details on the Methodology see Annex 1 of the Literature Survey:  
[http://www.emodnet-mediterranean.eu/reports\\_news/](http://www.emodnet-mediterranean.eu/reports_news/)

## 13 Annex 2 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 categories of 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	<p>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.</p> <p>There are two implementations with a set of associated tools:</p> <ul style="list-style-type: none"> <li>-HTML</li> <li>-Resource Description Framework (RDF)/XML a format for metadata, which is now used the most.</li> </ul> <p>Here is the list of associated metadata:</p> <ul style="list-style-type: none"> <li>-language-title-subject-description-publisher-rights-identifier-creator-contributor-date created-date modified-references-replaces.</li> </ul>
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	<p>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:</p> <ul style="list-style-type: none"> <li>- discovery</li> <li>- browsing</li> <li>- exploitation</li> </ul> <p>There are different categories of metadata:</p> <ul style="list-style-type: none"> <li>- identification</li> <li>-spatial representation</li> <li>-quality information</li> <li>-contents description</li> <li>-presentation modalities</li> <li>-dissemination modalities</li> <li>-maintenance modalities</li> <li>-security level</li> <li>-restriction level</li> </ul>

Metadata format	Description
	with a model for the Metadata 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**

## 14 Annex 3 List of Categories of Characteristics as a function of Challenge and number of Providers.

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
Administrative units		24						24
Air pressure	1							1
Air temperature	1							1
Atmospheric humidity	1							1
Bathymetry and Elevation	1	1	2	1				5
Bird behaviour	1	2						3
Bird reproduction	1							1
Bird taxonomy-related counts	2	2						4
Cetacean abundance		1						1
Chlorophyll pigment concentrations in the water column		1				1		2
Coastal geomorphology		1	1	2				4
Concentration of suspended particulate material in the water column							9	9
Depositional environment		1						1
Dissolved oxygen parameters in the water column		2						2
Dissolved total and organic nitrogen concentrations in the water column							4	4
Dissolved total or organic phosphorus concentrations in the water column							4	4

Fauna abundance per unit area of the bed	4	1						5
Fish abundance in water bodies	1	2					1	4
Fish and shellfish catch statistics					10			10
Fish reproduction	1							1
Fishery characterisation			2					2
Fishing by-catch					6			6
Habitat extent		20	5	1				26
Horizontal platform movement					10			10
Horizontal spatial coordinates		2						2
Horizontal velocity of the water column (currents)	2	2	28					32
Lithology		1						1
Man-made structures				1				1
Marine environment leisure usage			1					1
Nitrate concentration parameters in the water column						1	4	5
Nutrient fluxes between the bed and the water column		1						1
Phosphate concentration parameters in the water column						1	4	5
River flow and discharge							8	8
Salinity of the water column	1	3						4
Sea level	1			8				9
Seal abundance		1						1
Sediment resuspension		1						1
Sedimentary	1							1

structure								
Spectral wave data parameters	1							1
Temperature of the water column	1	3	14	20			3	41
Terrestrial mapping			1					1
Unspecified		1	2					3
Wave direction	1		7					8
Wave height and period statistics	5		14					19
Wind speed and direction	3	2	17					22
<b>Total #</b>	<b>30</b>	<b>75</b>	<b>94</b>	<b>33</b>	<b>26</b>	<b>3</b>	<b>37</b>	<b>298</b>
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

## 15 Annex 4 Providers for the 298 upstream data set descriptors identified by the Challenges

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

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

45	<u>French Marine Protected Areas Agency</u>	108	<u>RAMSAR</u>
46	<u>GEBICO</u>	109	<u>Regional Activity Centre for Specially Protected Areas (RACSPA)</u>
47	<u>German Oceanographic Datacentre (DODS)</u>	110	<u>Rempec Medslik-li</u>
48	<u>GFCM Fisheries Restricted Areas</u>	111	<u>Seadatanet-Pan-European Infrastructure for Marine Data 2</u>
49	<u>Greek Fisheries Control Agencies</u>	112	<u>Service Contract Concerning Coastal Erosion Evaluation of the Needs for Action</u>
50	<u>Hellenic Centre for Marine Research (HCMR)</u>	113	<u>SHOM</u>
51	<u>Hellenic Centre for Marine Research Hellenic National Oceanographic Data Centre (HCMR HNODC)</u>	114	<u>Slovenian Fisheries Control Agencies</u>
52	<u>Hydrological Cycle in Mediterranean Experiment (HYMEX)</u>	115	<u>Spanish Fisheries Control Agencies</u>
53	<u>ICCAT</u>	116	<u>Systeme d'Observation du Niveau des Eaux Littorales</u>
54	<u>IEO Spanish Oceanographic Institute</u>	117	<u>The Pelagos Sanctuary</u>
55	<u>Ifremer</u>	118	<u>Tulane University Department of Earth and Environmental Sciences</u>
56	<u>Ifremer Centre De Brest</u>	119	<u>TWReferencenet - Management And Sustainable Development Of Protected Transitional Waters</u>
57	<u>Ifremer Idm Sismer</u>	120	<u>UNEP MAP</u>
58	<u>Institute of Accelerating Systems and Applications (IASA-UAT)</u>	121	<u>UNESCO</u>
59	<u>Institute of Marine Sciences Middle East Technical University</u>	122	<u>United Nations Environment Programme Global Environment Monitoring System (Unepgems)</u>
60	<u>Institute of Research for Development (IRD)</u>	123	<u>University of Hawaii Sea Level Center</u>
61	<u>INSU</u>	124	<u>University of New Hampshire</u>
62	<u>INSU (Institut National Sciences de L'univers) Serv. d'Obs. en Milieu Littoral - Somlit</u>	125	<u>World Database on Protected Areas</u>
63	<u>International Ocean Institute - Malta Operational Centre (University Of Malta) Physical Oceanography Unit (UMTIOIPOU)</u>	126	<u>WWF</u>

## 16 Annex 5 EMODnet and MyOcean Indicators

A total of 34 data sets are indicated by the Challenges from EMODnet Portals and 15 from Myocean.

Emodnet Portal	Challenge	# Data Sets	AV-AC-3 (Data Policy)	AC-PE-1 (responsiveness)	AC-AV-5 (Format)	AC-AV-4 (Cost basis)	AC-AV-2 (Delivery mech.)
<b>Chemistry</b>	Climate&Coast						
	Fishery						
	Marine Env						
	MPA						
	Oil Leaks						
	River Input	8	Unrestricted	< 24 H	Ascii	No Charge	On-Line Downloading
	Wind Farm						
Total 8 Data Sets							
<b>Bathymetry</b>	Climate&Cost	1	Unrestricted	< 24 H	Various	No Charge	On-Line Downloading
	Fishery						
	Marine Env						
	MPA	3	Unrestricted	< 24 H	Various	No Charge	On-Line Downloading
	Oil Leaks	1	Unrestricted	< 24 H	Various	No Charge	On-Line Downloading
	River Input						
	Wind Farm						
Total 5 Data Sets							
<b>Sea Habitats</b>	Climate&Cost	2	Unrestricted		Ascii, ESRI	No Charge	Unknown
	Marine Env						
	MPA	1	Unrestricted		Various	No Charge	On-Line Downloading

	Oil Leaks						
	River Input						
	Wind Farm						
<b>Total 3 Data Sets</b>							
<b>Physics</b>	Climate&Coast	12	Unrestricted	< 15 Min	Various	No Charge	On-Line Downloading + Advanced
	Fishery						
	Marine Env						
	MPA						
	Oil Leaks						
	River Inputs	2	Unrestricted	< 15 Min	Various	No Charge	On-Line Downloading + Advanced
	Windfarm						
<b>Total 14 Data Sets</b>							
<b>Geology</b>	Climate&Coast						
	Fishery						
	Marine Env						
	MPA						
	Oil Leaks						
	River Input						
	Wind Farm						
<b>Total 0 Data Sets</b>							
<b>Biology</b>	Climate&Coast						
	Fishery						
	Marine Env						
	MPA	1	Unrestricted		Various	No Charge	On-Line Downloading
	Oil Leaks						

	River Input	1	Unrestricted		Various	No Charge	On-Line Downloading
	Wind Farm						
Total 2 Data Sets							
<b>Human Activities</b>	Climate&Coast						
	Fishery						
	Marine Env.						
	MPA						
	Oil Leaks	2	Unrestricted	15 Min	ESRI	No Charge	On-Line Downloading
	Wind Farm						
Total 2 Data Sets							
<b>Myocean</b>	Climate&Coast	5	Unrestricted	< 15	Netcdf	No Charge	On-Line Advanced
	Fishery						
	Marine Env.	1	Unrestricted	<15	Netcdf	No Charge	On-Line Downloading
	MPA	9	Unrestricted	< 15	Various?	No Charge	On-Line Downloading
	Oil Leaks						
	Wind Farm						
Total 15 Data Sets							

## 17 Annex 6 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.
Categories of characteristics	Specifically defined class of variables derived from observations or models and/or the geographical representation of an object.
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	Degree of conformance of datasets to the product specifications of the producer.
Fit for use	Degree of conformance of dataset to selection specifications of the user. Ability to appreciate the data’s exploitability (Challenge feedback on appropriateness and availability of datasets)
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.
Provider	Identify both Primary Providers (who are the producers) and Secondary Providers (who are the distributors only)

Relevance	Covering the extent to which data are appropriate for objectives of Challenges.
Territory	Domain of assessment that expresses the subdivision of the assessment criteria into the two categories of "appropriateness" and "availability".
Unique identification	Combination of (characteristic, dataset, intended use) or of (geographical 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.