Joint European Research Infrastructure network for Coastal Observatories

D# 5.6 Delayed-mode Data Management Handbook V 2

Grant Agreement n° 262584
Project Acronym: JERICO
Project Title: Towards a Joint European Research Infrastructure network for Coastal Observatories

Coordination: P. Farcy, IFREMER,

jerico@ifremer.fr, www.jerico-fp7.eu:

Authors: R. Nair, L. Petit de la Villeon
Involved Institutions: OGS and IFREMER
Version and Date : V2 – 10-03-2015

TABLE OF CONTENTS

1. Document description

REFERENCES
Annex 1 to the Contract Description of Work (DoW) version of the 22 Feb. 2011

Document information	
Document Name	Delayed-mode Data Management Handbook, Version 2, D#5.6
Document ID	
Revision	
Revision Date	
Author	Rajesh Nair, Loic Petit De La Villeon
Security	

History			
Revision	Date	Modification	Author

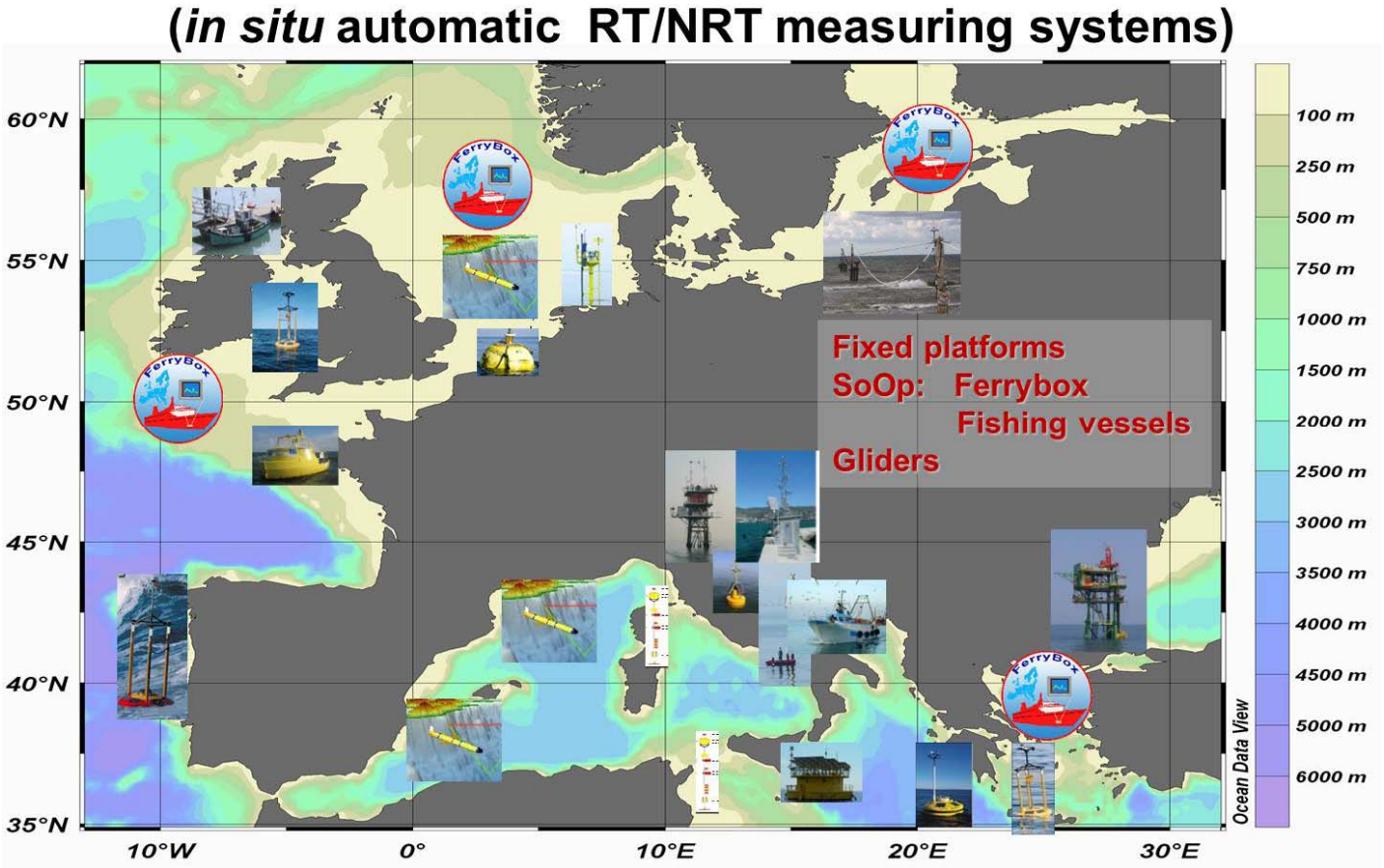
Diffusion list				
Consortium beneficiaries	X			
Third parties				
Associated Partners				
other				

2. Executive summary

The success of the coastal observatory network created through the JERICO project depends on the reliability, accessibility and easy distribution of data coming from the participating observing systems. This necessarily implies the use of a standard platform and common procedures for handling the data produced by the network. The present document constitutes the second version of the JERICO Delayed-mode Data Management Handbook. It describes the general JERICO data management structure and policy, and provides partners with practical advice and useful information on how to manage their delayed-mode data within the framework of the project. The document also contains references and links to the basic and most important online documents needed for implementing the established procedures. The guidelines presented here are those that have been followed by partners for handling their real-time data during the JERICO Service Data Access period that began in January 2013.

3. Introduction

The FP7 JERICO project embodies the first operational network of coastal observatories - assemblages of distributed sensor systems with extensive coverage - on a truly pan-European scale. The network is strikingly heterogeneous, marked by a surprising richness in the range of parameters handled, the frequency and spatial distribution of measurements, equipment maintenance practices, and the quality assurance schemes employed for sensors and data. All the coastal observatories forming the JERICO network share the same goal: to help detect and investigate coastal processes in a timely fashion, and provide crucial operational data for planning, assessment, mitigation, and model assimilation and validation purposes. The current composition of the JERICO network of coastal observatories is summarized in Fig 1.



Source: Sparnocchia et al. (Journal of Operational Oceanography, submitted)

Fig 1: Current composition of the JERICO network of coastal observatories showing the main kinds of real-time (RT) or near real-time (NRT) observing platforms in use.

As shown, the network mainly comprises four kinds of observing platforms: fixed or station-keeping structures, Ferryboxes on ferries, fishing vessels equipped with sensors mounted on fishing gear, and gliders; in more than a few cases, operators manage complex systems incorporating more than one of these platforms. The measuring instruments and sensors on the platforms can differ widely in number, kind, scope, and technical configuration from platform to platform, and from operator to operator. The sets of targeted parameters also vary considerably across the network, though salinity, temperature and pressure (depth) are nearly always measured. Table 1 provides a breakdown of the principal parameters being handled in the JERICO project.

Table 1. The JERICO parameter list; only the chief parameters are presented.

	PARAMETER	CORE	OPTIONAL
PHYSICAL	Salinity	‼	
	Temperature	‼	
	Turbidity	‼	
	Sea level	‼	
	Surface waves	‼	
	Surface currents		‼
	Chlorophyll-a	‼	
	Turbidity	‼	
	CDOM		‼
	Noise Passive Acoustic Listener (PAL)		‼
CHEMICAL	Dissolved nutrients		‼
	Dissolved oxygen	‼	
	CO2 partial pressure	‼	
	Contaminants		‼
	pH or Alkalinity		‼

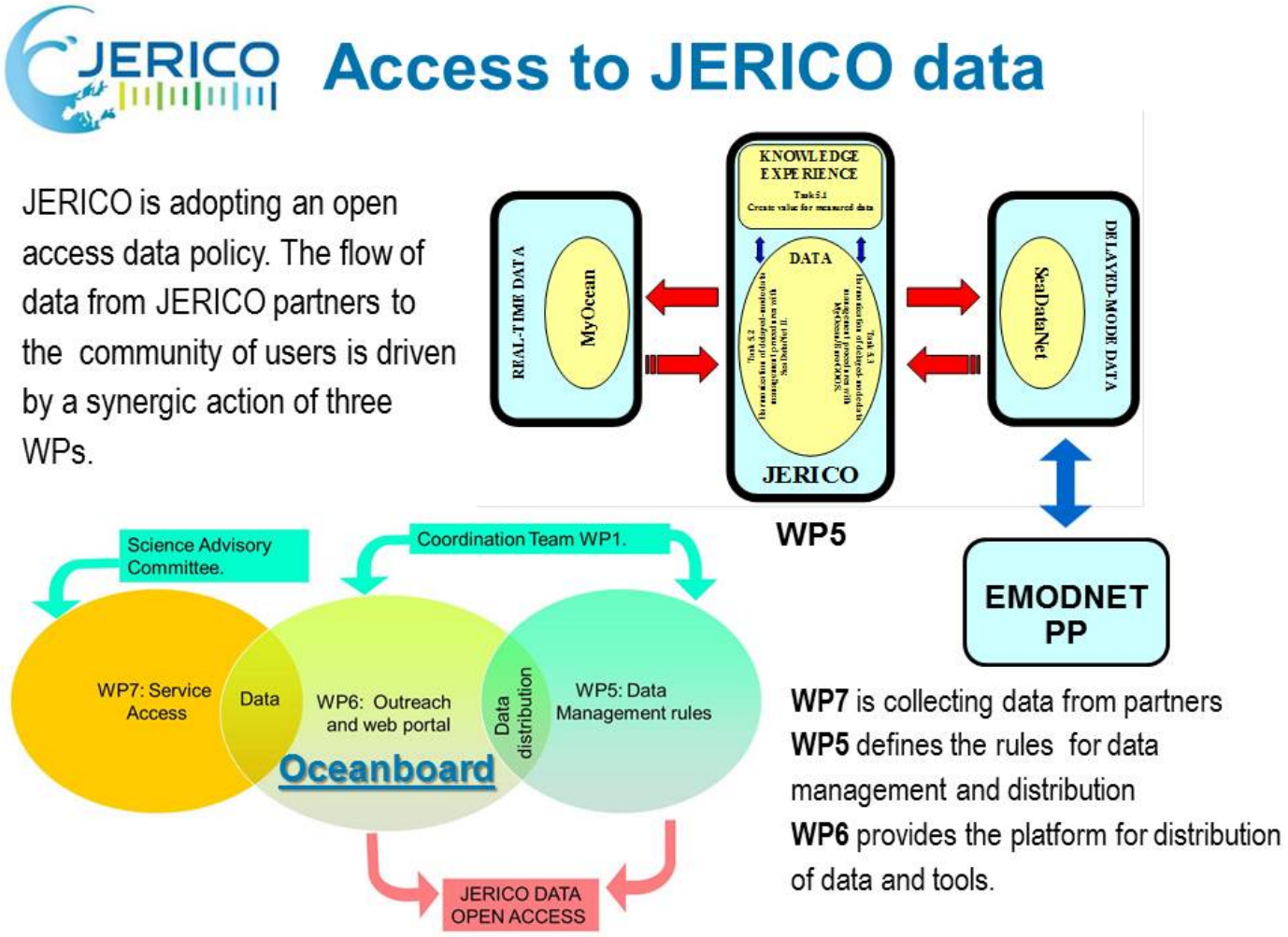
Source: Sparnocchia et al. (Journal of Operational Oceanography, submitted)

The JERICO network is geared towards ensuring continuous, timely access to coastal observations. The emphasis is on assuring valid data streams for coastal data at the transnational level in near real-time to ensure prompt availability and in delayed-mode for more deferred use.

4. Main Report

4.1. The current JERICO data management structure and principal data flows

The general approach to data handling that has been adopted by the JERICO project is outlined in Figure 2. It is based on an “open” data access policy, following the conventional European outlook in these matters. The flow of data from partners to the community of users concerns three Work Packages: WP7 is collecting data from partners, WP5 defines the rules for data management and distribution, and WP6 provides the platform for the distribution of data and related tools.



Source: Sparnocchia et al. (submitted)

Fig 2: Outline of the JERICO data management scheme, featuring the driving actions and main information flows.

The data management structure is built on the principle of “using what exists”. Suitable partnerships have been created with ongoing European data management initiatives so as to minimize possible duplication of efforts. Thus, there has been no dedicated development of a specific data management structure for JERICO. Instead, the use of, and integration with, already available data management infrastructure has been pursued, a strategy consistent with the policy behind SeaDataNet and MyOcean, the major ongoing European initiatives for the establishment and coordination of infrastructures for the management and distribution of marine data and products.

The JERICO data management framework for delayed-mode data uses the SeaDataNet infrastructure while the near real-time data are being routed through MyOcean. There is continuous interaction with SeaDataNet, MyOcean, EuroGOOS and EMODnet to enable the seamless integration of the JERICO data stream into these two established European marine data management infrastructures. Many difficulties remain to be overcome, though, especially in relation to parameters and data types that are not actually being handled in SeaDataNet and/or in MyOcean.

Some specific dataset indexing procedures were developed to help ensure compatibility with MyOcean Thematic Assembly Center requirements and assist non-NODC data providers in using the SeaDataNet infrastructure. An appropriate indexing and data distribution scheme was created. A distinct tag, an explicit JERICO index, was designed to easily recognize and select JERICO data from larger archives. This measure also allows to clearly define the JERICO contribution to the global marine observing system. Furthermore, procedures using the OGC’s SWE suite and SensorML format have been developed to help implement standardized descriptions of the different elements of the JERICO observing infrastructure. The descriptions can contain technical specifications of platforms and sensors, details of instrument settings, calibrations and performances, and some information on data processing procedures.

4.2. Delayed-mode data

Delayed-Mode (DM) data are data that have been submitted to a data management infrastructure more than 30 days after collection. It is implicitly assumed that this kind of data has undergone rigorous quality evaluations and calibration checks. If this is not the case, then data should be submitted as near real-time data. Most of the time, DM datasets can be considered as reference datasets, especially when applied quality control and calibration procedures are openly acknowledged and described. Data providers are recommended to obtain a DOI (Digital Object Identifier) for any submitted data to aid easy referencing in publications.

4.2.1 Description of the MyOcean delayed-mode data management infrastructure

The SeaDataNet data management infrastructure (Figure 4) is built around the National Oceanographic Data Centres (NODCs) of the different European countries involved in the initiative.

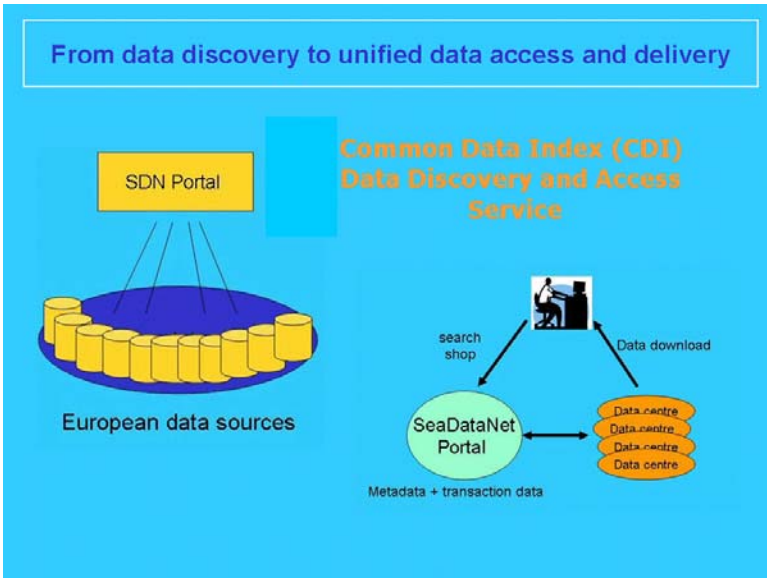


Figure 1. The main elements of the SeaDataNet data management infrastructure.

Each NODC is in charge of making relevant data from its specific country available to SeaDataNet. A NODC is normally not allowed to share the data from a country other than its own through SeaDataNet, although.some exceptions may be considered when the whole dataset from a single project is accessible from a single repository. In such cases, prior contact must be made with the SeaDataNet Project Coordinator (sdn-userdesk@seadatanet.org) to seek the necessary authorization.

The SeaDataNet data access mechanism (Figure 5) relies on a set of common metadata catalogues that allow a user to discover which datasets are available through the

SeaDataNet portal. In a second step, the user's query is forwarded to the NODC concerned and, in the final step, the whole data set that is required is made available.

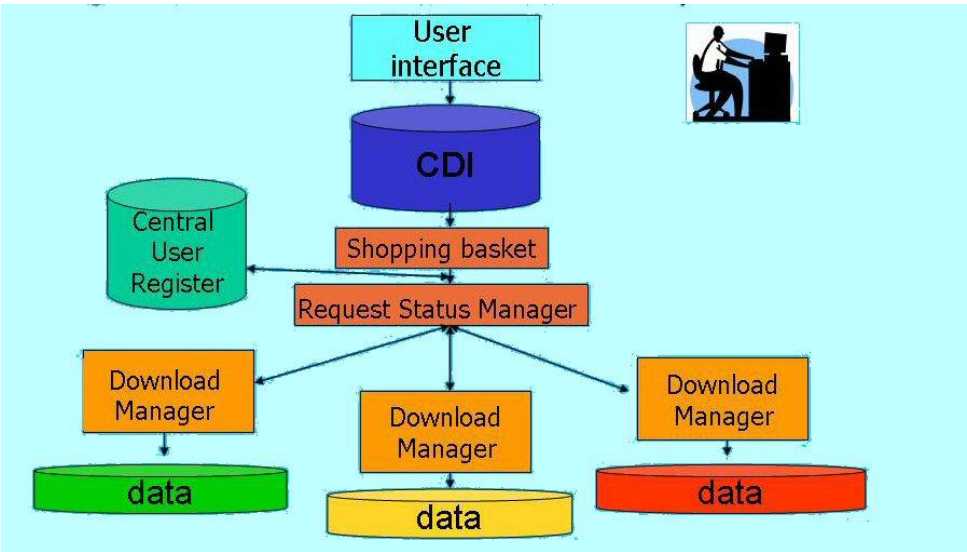


Figure 2. The main elements of the SeaDataNet data access system.

4.2.2 Uploading the main JERICO delayed-mode data types

The data generated with the JERICO network mainly come from four kinds of observing platforms: fixed or station-keeping structures, ferryboxes on ferries, fishing vessels equipped with sensors mounted on fishing gear (Fishery Observing Systems), and gliders.

SeaDataNet can handle most of the data types involved. As a first step, it is recommended that new data providers contact their respective NODCs or a local Data Center which is able to implement the different SeaDataNet components. A list of contacts can be found at the web address, "http://www.seadatanet.org/Overview/Partners".

Data providers can interact with SeaDataNet in two ways:

i. by asking the NODCs of their parent countries to manage, qualify, archive and provide long-term stewardship of presented data sets, and include these in the SeaDataNet data discovery and access service;

ii. by interfacing and connecting their in-house data systems with SeaDataNet via the Common Data Index (CDI) V3 system.

The procedure to enable the sharing of data is the same whether a data provider is the producer of the data or a downstream data management facility like a data center.

Integrating with SeaDataNet entails certain commitments:

i. data providers have to establish interfaces between their data repositories and the SeaDataNet web portal - or more correctly, the SeaDataNet Request Status Manager (RSM) service - by installing the SeaDataNet Download Manager (available at the web address, "http://www.seadatanet.org/Standards-Software/Software/Download-Manager").
ii. data must be managed and formatted in accordance with SeaDataNet formats.
iii. data sets have to be described by metadata using the SeaDataNet CDI system, which is based on the ISO 19139 geographic information standard.
iv. the metadata should be encoded as XML files utilizing the SeaDataNet CDI XML Schema (which is fully INSPIRE-compliant) and its controlled vocabularies.

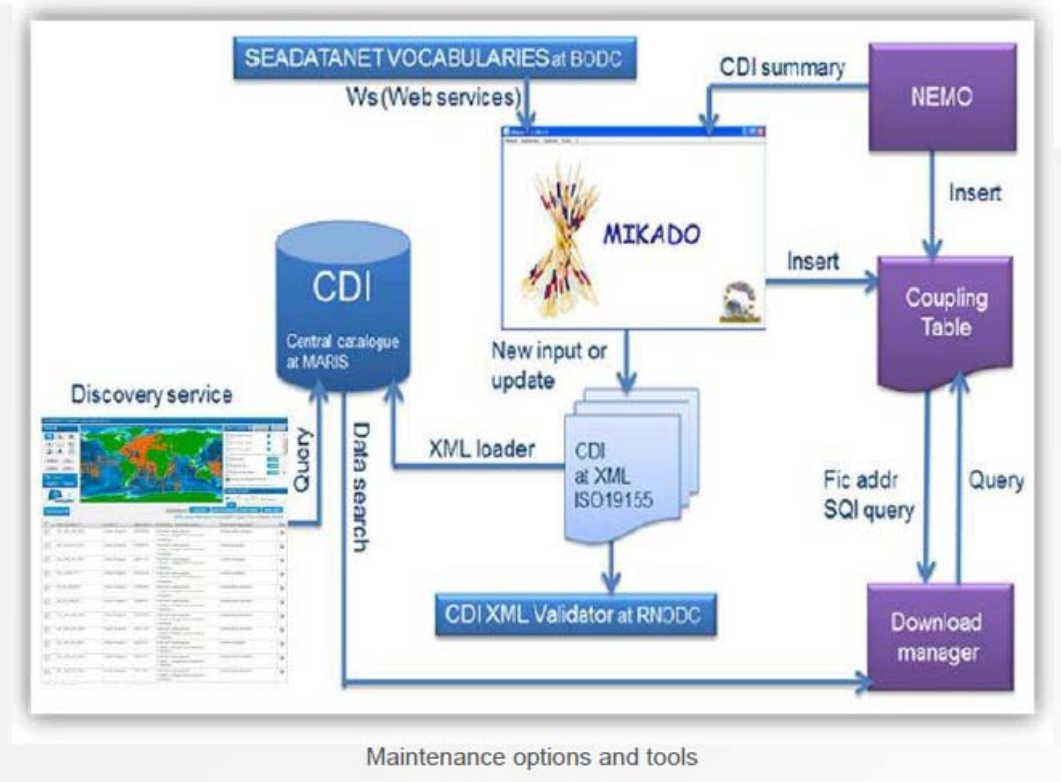


Figure 3. Overview of the Common Data Index (CDI) mechanism.

The process for generating CDI entries and linking these to local data sets is illustrated in Figure 3, above. It is warmly recommended to use the MIKADO software tool (which can

be downloaded at "http://www.seadatanet.org/Standards-Software/Software/MIKADO") for the automatic encoding of CDI XML files. The tool permits the processing of multiple files in batch mode after an initial mapping analysis.

A data provider can work with files that have already been converted into one of the allowed SeaDataNet formats, ODV, MEDATLAS or SeaDataNet NetCDF. Alternatively, compatible data sets can be generated on-the-fly with appropriate database queries. For historical data sets, it is best to organise files as validated yearly time series.

Detailed guidelines for utilizing the CDI service can be found at the web address, "http://www.seadatanet.org/Data-Access/How-to-contribute". Data providers are strongly advised to consult the various SeaDataNet manuals posted there to familiarize themselves with the components of the CDI system and its workings in a stepwise manner. Moreover, it is very important that potential data providers contact the CDI Support Desk at MARIS (write to "cdi-support@maris.nl") when they are beginning the process of establishing their connections with the SeaDataNet system. The Support Desk will help the data providers to set the secure configurations by which the SeaDataNet portal can exchange messages with the SeaDataNet Download Manager utility they have installed on their systems. The Support Desk is also the contact point for matters concerning updates and new releases of the CDI software.

4.2.3 Useful contacts for delayed-mode data submission

The main contacts for data submission are the NODCs of each data provider's home country. A full list of contacts is available at "http://www.seadatanet.org/Overview/Partners".

It is also possible to approach SeaDataNet through the SeaDatanet User desk at "sdn-userdesk@seadatanet.org".

4.3 Downloading the main JERICO delayed-mode data types

Users who wish to download data from SeaDataNet can freely browse, search for, and submit requests for accessing data sets after completing a one-time registration process. Data sets can be downloaded via the CDI user interface on the SeaDataNet portal at the web address, "http://www.seadatanet.eu.org". In practice, the data are transferred directly from relevant data holders, but the transmission is facilitated through the SeaDataNet portal.

4.4 Some useful documents

a) SeaDataNet user manuals and instructions concerning metadata (EDMED, EDMERP, EDIOS, EDMO and CSR formats) can be downloaded using the link given below.

"http://www.seadatanet.org/content/download/11284/75105/file/SDN2_D46_NC_UserManualforUpdatingMetadataDirectories.pdf".

b) SeaDataNet user manuals and instructions for compiling CDI metadata, coupling tables and associated data can be downloaded using the link given below.

"http://www.seadatanet.org/content/download/11663/77627/file/SDN2_D57_TC_UserManualforUpdatingCDI-Data-Access-Service.pdf".

5. Conclusions

This document gives an overview of the general data management structure and policy of the JERICO project, and furnishes instructions and information that partners involved in the project will need to join and contribute to its delayed-mode data flow. As most of the pertinent resources, including contacts, references and useful documentation, are available online, they have been referred to by their web addresses to ensure that partners will always have access only to the most up-to-date information. The present document constitutes the second version of the JERICO Delayed-mode Data Management Handbook.

6. List of acronyms

CDI: Common Data Index.

CSR: Cruise Summary Reports.

DOI: Digital Object Identifier.

EDIOS: European Directory of the Ocean Observing Systems.

EDMED: European Directory of Marine Environmental Data.

EDMERP: European Directory of Marine Environmental Research Projects.

EDMO: European Directory of Marine Organisations.

EMODnet: European Marine Observation and Data Network.

EuroGOOS: European Global Ocean Observing System.

INSPIRE: Infrastructure for Spatial Information in the European Community.

ISO: International Organization for Standardization.

MARIS: Marine Information Service (Mariene Informatie Service).

NetCDF: Network Common Data Form.

NODC: National Oceanographic Data Centre.

ODV: Ocean Data View.

OGC: Open Geospatial Consortium.

WMO: World Meteorological Organization.

XML: Extensible Markup Language.