Publishable summary for SeaDataNet III

Executive summary.

(The length of this part cannot exceed 1 page)

SeaDataNet is the leading network in Europe, actively operating and developing a pan-European infrastructure for managing, indexing and providing access to ocean and marine datasets and data products, acquired from research cruises and monitoring activities in European marine waters and the global oceans.

SeaDataNet is undertaken by the National Oceanographic Data Centres (NODCs), and marine information services of major research institutes from 34 coastal states bordering the European seas. It also includes experts in IT, data publishing, and modelling, as well as international organisations, namely the Intergovernmental Oceanographic Commission (IOC) of UNESCO, International Council for the Exploration of the Sea (ICES) and EU Joint Research Centre (EU-JRC). SeaDataNet is working closely together with other European RTD projects and initiatives such as EMODnet (European Marine Observation and Data Network – initiated by EU DG MARE in 2009) in the capacity of providing standards, services, data centres and infrastructure for managing marine and oceanographic data, and of providing experience and expertise for joint development of new standards and services. Technical innovations in standards and services enrich the basis of SeaDataNet and are implemented in its infrastructure, where possible.

From the start experts have been involved both in the field of operational oceanography and, later on, in the fields of chemistry, geology, geophysics and biology. This has resulted in a system that meets the users' needs. And substantial progress has been made in the SeaDataNet infrastructure and the services provided.

The SeaDataNet infrastructure provides harmonised discovery services and access to ocean and marine environmental data sets managed in distributed data centres, not only for partners in the SeaDataNet II project, but for many more data centres.

At present, SeaDataNet is an infrastructure of 102 data centres from 34 countries, giving access to 1.8 million datasets originating from more than 550 organisations in the Pan-European countries, and to a range of metadata services, tools and standards that have been widely adopted across the EU, and also more internationally: in the USA and in Australia through the cooperation with ODIP and ODIP II European projects.

Summary description of the project context and the main objectives.

(The length of this part cannot exceed 4 pages)

Oceanographic and marine data include a very wide range of measurements and variables, covering a broad, multidisciplinary spectrum of projects and programmes. Oceanographic and marine data are collected by over a thousand research institutes, governmental organisations and private companies in the countries bordering the European seas. Various heterogeneous observing sensors are used installed on research vessels, submarines, aircraft, and moorings, drifting buoys, gliders, floats and satellites. These sensors measure physical, chemical, biological, geological and geophysical parameters, with further data resulting from the analysis of water and sediment samples for a wide variety of parameters. These data are collected at a very considerable cost. Besides, they are of prime value because they are the reference for any study and, if lost, cannot be replaced. In order to make best use of these data for science and for society, a robust operational infrastructure, based on European and internationally agreed standards, is mandatory. It has to cover data quality and long term stewardship as well as technical and semantic aspects of interoperability.

The SeaDataNet infrastructure is undertaken by the National Oceanographic Data Centres (NODCs), and marine information services of major research institutes from 34 coastal states bordering the European seas. It also includes expert modelling centres and the following international organisations in its network: the Intergovernmental Oceanographic Commission (IOC) of UNESCO, International Council for the Exploration of the Sea (ICES) and EU Joint Research Centre (EU-JRC). The SeaDataNet data centres are highly skilled and have been actively engaged in data management for many decades and have the essential capabilities and facilities for data quality control, long term stewardship, retrieval and distribution.

The predecessor project, SeaDataNet (2006 – 2011), funded by the EU as an FP6 Research Infrastructures project, was actively operating and further developing a Pan-European infrastructure for managing, indexing and providing access to ocean and marine data sets and data products, acquired from research cruises and other observational activities in European marine waters and the global ocean.

The overall objective of the **SeaDataNet II project** was to upgrade the SeaDataNet infrastructure into an operationally robust and state-of-the-art Pan-European infrastructure for providing up-to-date and high quality access to ocean and marine metadata, data and data products originating from data acquisition activities by all engaged coastal states. It also aimed at setting, adopting and promoting common data management standards and at realising technical and semantic interoperability with other relevant data management systems and initiatives on behalf of science, environmental management, policy making, and the economy.

The specific **objectives** of the SeaDataNet II project are:

- Achieving an overall capacity of all connected data centres on a Pan-European scale, to provide up-to-date and high quality metadata and data access services for the data managed by their centres

- Achieving more metadata, data input and data circulation from other relevant data centres in Europe by further development of national NODC networks, thereby promoting and supporting adoption and implementation of SeaDataNet standards, tools and services.
- Achieving interoperability and exchange with other relevant data management systems in Europe, thereby promoting, fine-tuning and implementing the SeaDataNet standards, also taking into account active tuning and harmonising on an international scale.
- Establishing SeaDataNet as the core data management component of the EMODNet infrastructure and contributing on behalf of Europe to global portal initiatives, such as the IOC-IODE – Ocean Data Portal (ODP), and the Global Earth Observing System of Systems (GEOSS).
- Achieving INSPIRE compliance and contributing to the INSPIRE process for developing and implementing rules for oceanography.
- Achieving for operational oceanography, both delayed-mode and real-time data provision capacities, in close cooperation with MyOcean, EuroGOOS, its Regional groups (ROOSes), and other oceanographic monitoring agencies and systems.
- Achieving a more streamlined process from data acquisition by research fleets, *in situ* and remote sensing observation systems, and *ex-situ* analyses in laboratories, to data centres through further development and implementation of common technical and semantic standards, such as SensorML, Observation and Measurements and common vocabularies.
- Achieving an improved capability for handling also marine biological data and interoperability with the emerging biodiversity data infrastructure in close cooperation with actors of the EurOBIS, MarBEF, and LifeWatch initiatives
- Achieving active, well-structured and world class contributions from Europe to global data management and exchange projects, such as Argo, Global Ocean Surface Underway Data (GOSUD), OceanSites, and others.

The SeaDataNet consortium has grouped 58 organisations from 23 member states of the European Union and 11 other countries bordering the European seas namely:

- forty-four beneficiaries: AWI (Germany), BSH (Germany), CLS (France), CNR (Italy), ENEA (Italy), EPA (Lithuania), FMI (Finland), HCMR (Greece), IBSS (Ukraine)[1], ICES (Denmark), IEO (Spain), IFREMER (France), IHPT (Portugal), IMGW (Poland), IMR (Norway), INGV (Italy), IO-BAS (Bulgaria), IOF (Croatia), IOI-MOC (Malta), IOLR (Israel), JRC (Italy), LIAE (Latvia), MARIS (The Netherlands), METU-IMS (Turkey), MHI-DMIST (Ukraine)[1], MI (Ireland), MRI (Iceland), MSI (Estonia), NERC-BODC (United Kingdom), NERI (Denmark), NIB (Slovenia), NIMRD (Romania), NIOZ (The Netherlands), OC-UCY (Cyprus), OGS (Italy), RBINS-BMDC (Belgium), RIHMI (Russian Federation), SIO-RAS (Russian Federation), SMHI (Sweden), TSU-DNA (Georgia), TUBITAK (Turkey), ULg (Belgium), UNIHb (Germany), VLIZ (Belgium);

[1] These Crimean partners were involved for the three first year of the project and then their participation was terminated on 02/10/2014 through a decision of the European Commission

- thirteen sub-contractors: CLUBWEB (Italy), EU-CONSULT(The Netherlands), IGEWE (Albania), IMBK (Montenegro), INRH (Morocco), INSTM (Tunisia), IOC-IODE (Belgium), IOC-JCOMMOPS (France), RSHU (Russian Federation), SHODB (Turkey), SHOM (France), STFC (United Kingdom), UTM-CSIC (Spain);
- one associated partner: IOPAN (Poland).

Together these constitute a rich balanced partnership of National Oceanographic Data Centres or data focal points, international organisations, expert IT and modelling centres, representing all European Sea basins from the Baltic Sea, the North Sea, the Arctic Seas, the North Atlantic Ocean, the Mediterranean Sea up to the Black Sea.

The ambitious objectives of the project covered three activities defined as: Coordination, Support, and Research and Development.

The coordination activities included the maintenance and operation of all the SeaDataNet discovery services, on top of the metadata and data catalogues, as well as the SeaDataNet common vocabularies web services covering a broad spectrum of ocean and marine disciplines. . The capacity building and training of all consortium beneficiaries in the use of the different SeaDataNet tools, and the communication, dissemination and promotion to users forms important components of the coordinating activities.

The support activities corresponded to the operation of the SeaDataNet infrastructure by all connected beneficiaries, aiming at giving access to an increasing volume and diversity of data, metadata and data products.

The research and development activities included the specification and governance of standard metadata, data and data product formats, quality check methods, common vocabularies and new services, as well as the development and governance of software tools, services and interoperability solutions and the development and regular updating of standard data products for maritime European regions.

The overall budget of SeaDataNet II was around 7.5 million euros, 6 of which were funded by the European Commission, for a duration of four years.

Description of the main S & T results/foregrounds.

(The length of this part cannot exceed 25 pages)

Research and development has been the core activity of the SeaDataNet project representing a total of 324 person/months over the 603 person/months of the full project.

Since the central aim of SeaDataNet II project was to improve the SeaDataNet infrastructure in terms of standardisation and INSPIRE compliance, interoperability with other data management systems, machine to machine interfaces, operational updates of the SeaDataNet catalogues (metadata harvesting), robustness and preparation of data products many research and development activities took place during the 4 years of the project.

The main results are described hereafter.

1 - Standardisation

SeaDataNet is a distributed infrastructure between all the connected data centres; in order to be able to deliver coherent and homogeneous data and metadata, much standardisation work has been realised since the first steps of SeaDataNet. Details about major achievements concerning INSPIRE compliance, the new version of NERC-BODC vocabulary server, the NetCDF transport format and the extension of the ODV format for handling biological data given below.

1.1 INSPIRE compliance

At the end of the first SeaDataNet project, all SeaDataNet catalogues were described through ISO-19115 format which was the standard available at that time. Since then, the ISO-19139 schema matured and was adopted by INSPIRE for the description of geographic information metadata. Thus the most frequently populated SeaDataNet catalogues that were candidate for a standard OGC-CSW harvesting (i.e. CDI - Common Data Index - and CSR - Cruise Summary Report - catalogues) had to be upgraded to that standard.

For this purpose an analysis took place of the initial CDI and CSR formats compliant with ISO 19115:2003 also taking into account specific additional constraints from the European Directive INSPIRE. This resulted in a new profile definition, which also added a number of missing elements. Furthermore a related document (SeaDataNet ISO 19115 profile – XML encoding) was prepared which defines and details the XML schema implementation for these metadata profiles, based on the XML schema defined in ISO 19139:2006 plus additional definitions and Schematron rules. The Schematron rules are defined to express XML constraints not enforceable by the schema, such as the use of SeaDataNet common vocabularies with controlled terms.

These two formats have been submitted to the Ocean Data Standards and Best Practices Project (ODSBP), the continuation of the Ocean Data Standards (ODS) Pilot Project of JCOMM-IODE for sharing the CDI and CSR formats and Schemas as European marine profile with other ocean communities worldwide (Deliverable D8.9).

1.2 Vocabulary

Use of common vocabularies in all metadatabases and data formats is an important prerequisite towards consistency and semantic interoperability. Common vocabularies consist of lists of standardised terms that cover a broad spectrum of disciplines of relevance to the

oceanographic and wider community. Using standardised sets of terms solves the problem of ambiguities associated with data mark-up and also enables records to be interpreted by computers. This opens up data sets to a whole world of possibilities for computer aided manipulation, distribution and long term reuse. Common vocabularies have to be controlled for consistency, which includes both content governance (concept population and semantic descriptions) and technical governance (content storage and distribution) services.

SeaDataNet (NERC-BODC) maintains and operates the NERC Vocabulary Server (NVS), covering a broad spectrum of ocean and marine disciplines. The common terms are used to mark-up data, metadata and data products in a consistent and coherent way. This vocabulary service provides the basis for semantic interoperability within the SeaDataNet architecture. Maintenance includes both content governance (concept population and semantic descriptions) and technical governance (content storage and distribution) services. Content governance is regulated by an international board. At present the Vocabulary Services comprise more than 60,000 terms in over 80 lists.

The NERC Vocabulary Server was originally developed in 2006. In order to support the requirements of the user community, several enhancements were required to the existing NVS, and therefore a version 2.0 (NVS2.0) was developed.

The major upgrades of NVS2.0 consist of:

- a move to the latest version of the World Wide Web Consortium's (W3C) Simple Knowledge Organisation System (SKOS) specification for encoding the data dictionaries and taxonomies served through the NVS
- the ability to serve multilingual titles and definitions for resources
- the provision for mappings to external resources enabling the results of ontology extension to be delivered.

NVS2.0 was operational in 2013 and SeaDataNet migrated to using NVS2.0 which implied the upgrade of all formats, tools, procedures and services using NVS2.0.

The Controlled Vocabularies (including content governance) have been submitted to the Ocean Data Standards and Best Practices Project (ODSBP), the continuation of the Ocean Data Standards (ODS) Pilot Project of JCOMM-IODE for sharing with other ocean communities worldwide (Deliverable D8.9).

1.3 Data transport formats

As part of the SeaDataNet services, data sets are accessible via download services. Delivery of data to users requires common data transport formats, which interact with other SeaDataNet standards (Vocabularies, Quality Flag Scale) and SeaDataNet analysis and presentation tools (ODV , DIVA). Therefore the following data transport formats have been defined:

- SeaDataNet ODV4 ASCII for profiles, time series and trajectories;
- SeaDataNet NetCDF with CF compliance for profiles, time series and trajectories, and 3D observation data such as moored or shipborne ADCP;
- SeaDataNet MedAtlas as optional extra format.

Both ODV and MedAtlas SeaDataNet formats were already defined when SeaDataNet II started, whereas the definition of SeaDataNet NetCDF format has been carried out in the frame of this project.

The **SeaDataNet NetCDF (CF) format** for profiles, time series and trajectories has been defined by bringing together a community comprising NetCDF and CF experts (such as from NCAR and UNIDATA), and as many users of CF NetCDF for oceanographic point data as possible. This included participants from MyOcean, the Australian Integrated Marine Observing System (IMOS), the Australian Navy and USNODC.

The SeaDataNet NetCDF (CF) format is based upon Version 1.6 of the CF Metadata Conventions, published by the CF community in December 2011. This publication includes a chapter on 'Discrete Sampling Geometries' to cover storage of point data in NetCDF. This was taken as starting point to formulate how basic point data - profiles, time series and trajectories - can be encoded in CF-compliant NetCDF together with the usage metadata - including parameter semantics - that SeaDataNet had included in its ASCII ODV and MedAtlas formats.

The approach taken with the development of the SeaDataNet profile based on CF 1.6 was to classify data on the basis of feature types and produce a SeaDataNet specification for storage of each of the following:

- Profile (x, y, t fixed; z variable). The specification given is for storage vertical profiles such as a CTD cast or bottle profile.
- TimeSeries (x, y, z fixed; t variable). The specification given is for storage of time series, such as a current meter record.
- Trajectory (x, y, z, t all variable). The specification given is for storage of trajectories such as underway thermosalinograph.
- TimeSeriesProfile (x, y, z fixed; t variable) but some variables can be measured at different depths at the same time var=f(t,z). The specification given is for storage of time series profiles such as moored ADCP.
- TrajectoryProfile (x, y, z, t all variable) but some variables can be measured at different depths at the same time var=f(t,z). The specification given is for storage of trajectory profiles such as shipborne ADCP.

The specification was then developed through discussions on a collaborative e-mail list involving participants in SeaDataNet, MyOcean, USNODC, NCAR and the Australian Ocean Data Network (AODN). The working objective focussed on producing profiles with the following properties:

- CF 1.6 conformant
- Have maximum interoperability with CF 1.6 implementations in use by MyOcean (OceanSITES conventions), USNODC (USNODC NetCDF templates) and two contributors to AODN (IMOS and METOC)
- Include storage for all labels, metadata and standardised semantic mark-up that were included in the SeaDataNet ODV format files for the equivalent feature type.

Significant list discussion focussed on the version of NetCDF that should be used for SeaDataNet. The conclusion was that NetCDF 4 should be used wherever possible, but that NetCDF 3, although strongly discouraged, should not be totally forbidden.

The SeaDataNet NetCDF format has been submitted to the Ocean Data Standards and Best Practices Project (ODSBP), the continuation of the Ocean Data Standards (ODS) Pilot Project of JCOMM-IODE for sharing with other ocean communities worldwide (Deliverable D8.9).

1.4 Biological data handling

In its first stage, SeaDataNet has focused on data management and access for physical oceanography, marine chemistry (to support also the EMODNet Chemistry pilot), bathymetry (to support the EMODNet Hydrography and Seabed Mapping pilots), and geology and geophysics (to support the Geo-Seas project and the EMODNet Geology pilot). Many partners in SeaDataNet are also involved in data acquisition and management for marine biology. A number are members of the Marine Biodiversity and Ecosystem Function (MarBEF) network of excellence and contribute to EurOBIS (European Ocean Biogeographic Information System) managed by the Flanders Marine Institute (VLIZ).

One of the objectives of SeaDataNet II was to undertake actions to make SeaDataNet better fit for handling marine biological data sets and establishing interoperability with biologcal infrastructure developments. Therefore an analysis was undertaken in SeaDataNet II together with actors from the initiatives mentioned above to resolve how SeaDataNet can be best adapted for also handling marine biological data sets.

Based on an analysis of the present situation, and currently existing biologcal data standards and initiatives, such as the Ocean Biogeographic Information System (OBIS), Global Biodiversity Information Facility (GBIF), Working Group on Biodiversity Standards (TDWG) and World Register of Marine Species (WoRMS) standards, a recommended format for data exchange of biological data has been developed (deliverable D8.4).

Key issues that steered the format development were:

- Requirements posed by the intended use and application of the data format (data flows, density calculations, biodiversity index calculations, community analysis, etc…)
- Availability of suitable vocabularies (World Register of Marine Species, SeaDataNet Parameter list, SeaDataNet Unit list, etc…)
- Requirements for compatibility with existing tools and software (WoRMS taxon match services, EurOBIS QC services, Lifewatch workflows, Ocean Data View, etc…)

The requirements of the extended ODV format for biological data were defined as follows:

- The format should be a general and higher level format without necessarily containing all specifics of each data type, but rather focusing on common information elements for marine biological data.
- At the same time the format needs to be sufficiently flexible/extendable to be applicable for at least part of the variety of biological data the NODC's are managing.
- It should be possible to derive OBIS or Darwin Core compatible datasets from the format.
- The format should be self-describing, in the sense that all information needed to interpret the data should be included in the file format or be available through links to vocabularies or term lists that are part of the format.

The specific SeaDataNet ODV extended format for biological data has been defined; templates and examples of files are described in the deliverable D8.4 for data such as:

- macrobenthos community with density and biomass values
- zooplankton community with samples from different depths
- demersal fish population with densities for different size classes and individual fish measurements

- pollutant concentrations in biota specimens

At the end of the project more than 20,000 CDIs related to biological data were available in SeaDataNet infrastructure and considerably more were in preparation (> 500,000 records).

2 - Interoperability

2.1 Interoperability with other web portals

One of the activities of SeaDataNet II has focused on establishing interoperability between the SeaDataNet CDI Data Discovery & Access service, which can be considered as a regional infrastructure, and global initiatives, such as the IOC-IODE Ocean Data Portal (ODP) and the GEOSS portal.

In discussion between MARIS, CNR and RIHMI-WDC it has been agreed that interoperability from the SeaDataNet CDI service with the 2 global portals should be arranged by using the OGC CS-W (Catalogue Services for the Web) standard protocol for exchanging ISO-19115 – 19139 XML coded metadata entries. This provides a standard interoperability which can also be applied by SeaDataNet towards other portals and is compliant with the EU INSPIRE Directive. Moreover, considering the function of the 2 global portals (ODP and GEOSS) as global guides and also taking into account performance issues, it has also been agreed to establish the interoperability at collections level and not at the granule level of a CDI metadata record. Furthermore, it has been decided to make use of the GEO-DAB brokerage service as developed and operated by CNR for establishing the SeaDataNet CS-W service on top of the SeaDataNet CDI Data Discovery and Access service as operated by MARIS. Making use of the GEO-DAB brokerage service (based upon GI-Cat) is an efficient solution and has additional advantages that optimal use can be made in the near future of the further developments that are actively taking place for GEO-DAB in the GEOSS framework and other projects. These developments guarantee maintenance and operational sustainability of the GEO-DAB service and keeping up-to-date with changing standards such as developed by ISO and OGC.

The activities undertaken for establishing the interoperability have been as follows:

- a collection has been defined by MARIS as an aggregation of CDI entries by data centre (EDMO-code), discipline (P08 terms) and by geometry (points / tracks / polygons). A program has been developed for automatic and dynamic aggregation of all CDI entries into CDI collections, each time that the CDI database is being updated (almost every day). These collections are formatted in XML following the overall CDI ISO-19139 Schema and with supporting SeaDataNet controlled vocabularies.
- an XML output REST service has been set by MARIS at the SeaDataNet CDI service to provide the CDI collections in XML level. Each collection XML file contains aggregated metadata to describe the CDI collection and this also includes a specific URL to the public CDI Data Discovery & Access service to give the CDI results within the collection and the normal options for requesting access.
- CNR has set up the GEO-DAB brokerage service to harvest the CDI collection of XML files and to convert these to the Brokerage Common Reference Model, thereby sustaining use of specific SeaDataNet vocab terms next to ISO standard terms;
- In addition, MARIS has integrated in its program a trigger URL as provided by CNR for the GI-cat server. This trigger is called each time after generating updated CDI collections. This way there is a dynamic updating of the chain of services.

- The GEO-DAB Brokerage service provides the SeaDataNet CDI collections as common XML files by 2 standard protocols **CS-W** and **OAI-PMH.** CS-W is the OGC Catalogue Service for the Web protocol, while OAI-PMH is the Open Archives Initiative Protocol for Metadata Harvesting, another protocol globally used for exchanging metadata.

2.2 Machine to Machine interoperability

The present CDI Data Discovery and Access service provides a strong and important foundation of the SeaDataNet infrastructure because it connects already more than 100 data centres and gives access to a large volume of data sets as managed by these data centres.

One of the objectives of SeaDataNet II was to extend the CDI services with a machine to machine interaction in support of regular and operational metadata and data exchanges with systems and members of specific user communities.

Projects such as EMODnet Chemistry require regular and operational metadata and data exchanges. Therefore SeaDataNet has developed a machine to machine interaction, which is now fully operational, to facilitate an efficient discovery and retrieval of data collections. It consists of a **robot harvesting system** for automatic discovery and harvesting of metadata and data sets, and an online SeaDataNet Buffer Content Management System (CMS), which allows configuration of specific group profiles (specify group, involved users, motivation, query criteria). The Buffer CMS works together with the upgraded Request Status Manager (RSM) service, an existing component of the CDI shopping mechanism, to perform and administer robot shopping transactions and to store the harvested data sets in central buffers. This also includes maintenance, whereby new and updated CDI entries are identified and used to trigger additional harvesting for the central data buffers.

Furthermore, the central buffers have been equipped with a Central Buffer CDI User Interface including the shopping mechanism, to facilitate the extraction and delivery of metadata and data sets from the central buffer databases in a regulated way. All central shopping transactions on the central buffers are administered in a new section of the RSM so that data providers can fully oversee all transactions. This buffering system is exclusive for specific applications and communities and access is secured via the SeaDataNet AAA service only for authorised users as defined in the buffer CMS. It does not replace the distributed CDI infrastructure and its shopping process for regular users. Next to the human operated Central Buffer User Interface also an Application Programming Interface (API) has been developed. This API works alongside the Central Buffer User Interface and facilitates remote and authorised machines, using the SeaDataNet AAA service.

In the last phase of the SeaDataNet II project, further use has been made of the robot harvesting system for preparing buffers of data collections for WP10 and for other projects such as EMODnet Chemistry. Thus further fine-tuning of the robot harvesting system has taken place.

2.3 Metadata harvesting

Dynamic metadata harvesting using OGC CS-W services has been implemented for the most frequently updated catalogues (CSR and CDIs) for making the transfer of new XML files between the local nodes and the central directory managers more operational.

After analysis in the SeaDataNet Technical Task Group, it has been decided to make use of GeoNetWork.

2.3.1 Implementation for CSR catalogue:

SeaDataNet nodes use MIKADO to generate their CSR XML entries in ISO 19139. The nodes have to install and configure the Open Source GeoNetwork package to set-up a CSW node to publish their metadata. For this purpose IFREMER has configured a dedicated version of GeoNetwork for making CDI and CSR XML files available for harvesting (described in Deliverable D9.2).

The central directory manager for CSR (BSH, Germany) has installed GeoNetwork at their end. BSH is now harvesting new / updated XML entries from several local SeaDataNet nodes on a weekly basis and takes into account only records with an update date beyond the last harvesting date. This ensures that only new and updated XML files are harvested and not all XML entries from the local directories every time.

The pilot test for harvesting of CSR XML files from IFREMER, HCMR, OGS and IEO (local) by BSH (central) has been successful. The test phase was concluded in February 2015. Since then the CSR harvesting is operational. Other potential harvesting candidates have been instructed how to install and configure GeoNetwork. Three more local nodes (RBINS-BMDC - Belgium, MI - Ireland and NERC-BODC – United Kingdom) are implementing this mechanism. It was agreed in the Roadmap that this will be a gradual process, establishing a mix of harvesting and sending by email of updates.

As before, the central managers of all directories will perform automatic checks on syntax and semantics of the exchanged XML entries during import, as well as visual inspections on the quality of the provided content. In the new situation CSR CS-W harvesting has been introduced to make the transfer from local data centres to the central directories more efficient. However the updating of the central public CSR directory itself will still be done via a staging process, because automatic checks on semantics and syntax only will not exclude content issues. Staging will allow for a final visual validation of information content before publishing which will guarantee the overall high quality and coherence of all information provided by the central services.

2.3.2 Implementation for CDI catalogue:

In the last year of the project also the remaining CDI Harvesting and Ingestion system has been successfully developed and implemented. MARIS uses GeoNetWork for harvesting CDI XML entries from data centres that have installed and configured GeoNetwork for CDI XML files.

MARIS has developed and implemented the CDI Ingestion Content Management System (CMS) which allows data centres to follow and control their CDI submissions from load till production (description of the processes in deliverable D5.3).

Once harvested, the CDI XML entries are taken up by the ingestion process. The system runs in steps through a batch process in which the data provider is asked to interact only a few times for the following checks:

- to identify XML errors (syntax – semantics) for possible repair in the next batch,

- to identify potential duplicates (against import and production) and undertake action to delete real duplicates from the import database,
- to control the overall phases from CDI import and submission to the production environment (after confirming up-to-date coupling table and new data availability).

MARIS monitors the ingestion processes and also keeps on processing CDI batches received from other data centres in the traditional way by email and possibly FTP. For both purposes MARIS uses the CDI Ingestion CMS which was used successfully in the last 6 months for all CDI packages received.

The CDI Harvesting and Ingestion system has successfully been tested by MARIS with 3 data centres: IFREMER (France), BSH (Germany), and IEO (Spain). All three data centres have installed the software client developed by IFREMER for GeoNetwork CSW and configured it for the distribution of CDI XML files. The pilot has helped to test thoroughly the CDI ingestion system and at the same time familiarize the persons at the partner organisations with the alternative work method for updating and adding new CDI entries to the CDI data discovery and access service. The CDI harvesting and ingestion activities accomplished have been documented and are now available also for other SeaDataNet data centres to adopt as alternative to their present submissions.

2.4 Marine-ID

Since the first stage of SeaDataNet, the IFREMER partners' directory has been used to manage registration and authentication of the data users. This solution was not sustainable for security reasons as SeaDataNet users can automatically register in the directory unlike IFREMER partners whose registration is manually validated.

For SeaDataNet II and other data service providers it has been decided to set up a new system for registration and authentication of marine datasets users. This system is called Marine-ID (http://www.marine-id.org). As for the previous system and eduGain federation of identity, a Central Authentication System CAS (http://jasig.github.io/cas/4.1.x/index.html) is used for authentication management. For user registration the open-source software PWM (https://github.com/pwm-project/pwm) is used. It provides the required security features for such function: CAPTCHA, email confirmation. The full marine-id implementation for user registration and authentication has been deployed in production in April, 2014.

This Marine-ID system manages user registration and authentication. Besides, projects might require specific data license agreement or authorisation role assignment. Marine-ID has been designed to cooperate with the related project dedicated databases.

For SeaDataNet, the full network of around one hundred data centres and a number of central services (in particular the RSM for request status management) had to be migrated to the new user registration and authentication system. As the technologies for authentication remain the same, CAS, the migration was simple at data centre level but still challenging because of the large number of nodes. The authorisation and role management system has also been upgraded to cooperate with Marine-ID functions.

2.5 Sensor Web Enablement (SWE) for supporting operational oceanography

Users of oceanographic data resulting from monitoring stations are generally interested in both the long-term archived time series, as well as the near-real time or real time data. This is also requested in dedicated projects such as the new AtlantOS project, which is aiming for an integrated Atlantic Ocean Observing System, and EMODnet Physics which aims at providing users an easy portal for discovery and access to physics data from all operational stations, such as buoys, fixed moorings, gliders, floats (ARGO) and other instrumented platforms. In practice, provision of real-time data and delayed-mode data is arranged by different infrastructures. Real-time and near real-time data are managed in a cooperation between the EuroGOOS ROOSs (Regional Operational Oceanography Systems) and the INSTAC (In Situ Thematic Assembly Centre) of the Copernicus Marine Environment Monitoring Service (CMEMS). The present Copernicus INSTAC.CMEMS has been developed in coherence with other components of the pan-European data management landscape following the EuroGOOS recommendation adopted in 2008 by the EuroGOOS Assembly. The INSTAC.CMEMS makes use of standards and vocabularies developed within SeaDataNet and is interoperable both upstream with the observation provider networks and downstream with initiatives including EMODnet Physics.

The physical data sets are also gathered by data centres (in particular NODCs) from organisations operating monitoring stations in their country (both from research and monitoring activities) and after additional QA-QC and processing stored in their databases for long-term archiving. The additional scientific validation and documentation of the data sets is required to make these data sets also fit for purpose of scientific analyses such as generating climatologies, studying climate changes, sea level rise and other topics which require high quality data sets. The data centres populate metadata and data also into the European SeaDataNet infrastructure, which is managed by the NODCs in Europe, and which then provides the basis for making archived physical data available for users and portals such as EMODnet Physics.

EuroGOOS, CMEMS and SeaDataNet work closely together to provide discovery and access to physics data sets in a harmonised way. SeaDataNet also plays an important role in developing standards and protocols that can be adopted for wider implementation. For that purpose SeaDataNet has undertaken additional activities to extend its traditional data access services as provided for long term archive datasets. This extension aims at facilitating access and especially visualisation of the time-series datasets provided in near real time by operational oceanography data providers. In doing this, SeaDataNet has taken advantage of the Sensor Web Enablement (SWE) technologies which are designed and increasingly implemented to provide standard services close to the observatories. Discovery and access to long-term time series is provided by SeaDataNet through its CDI Data Discovery and Access service. To demonstrate support for operational oceanography data sets, it was decided to develop an extension to the CDI service in such a way that users of the SeaDataNet CDI service will have additional functionality to access and visualise near real-time data from the same monitoring station that also provided the long-term data sets. These long term data sets are normally described in the CDI entries.

This extension has been demonstrated for selected monitoring stations. The demonstrator has been set-up to with data sets resulting from 2 monitoring stations: E2M3A mooring as

operated by OGS; EMSO Azores deep sea observatory as operated by IFREMER. The technical architecture is composed of 3 new components and one upgraded and comprises:

- (new) 2 SOS servers deployed at data provider level;
- (new) 1 SOS client for time-series visualisation;
- (upgraded) the CDI data access portal and catalogue providing observation description and federating the links to the pre-existing download access points and the new services for operational oceanography.

The client allows visualisation of the location of the monitoring station on a geographical map and also visualisation of the data time series of the station per parameter. The CDI Data Discovery and Access portal manages CDI entries which describe in metadata the details of the long term data sets as collected at the monitoring stations. Normally it gives users the functionality for discovering relevant data sets, to browse the related metadata in ISO 19115 – 19139 format, and to request access to the long-term data sets that are stored at the SeaDataNet data centres. For the Demonstrator, use has been made of the test server of the CDI service: http://seadatanet.maris2.nl/v_cdi_v3_test/search.asp. A number of CDIs for the demonstrator monitoring stations have been amended to include description and links to the extra data services. The new restFUL 52°North API services have been added in the CDI observation description records (in ISO-19139 format), as on-line resources with protocol type "52n-sos-restful-ts-api". This protocol has also been added to the L07 (SeaDataNet data access mechanisms) vocabulary. The resulting dialogue is that a user can go to the CDI test service and look for CDI entries from OGS in Italy for the monitoring station. This results in several CDI entries.

The demonstrator has successfully shown a possible way to make use of Sensor Web Enablement (SWE) for giving access by visualisation of NRT time series from operational oceanography monitoring stations. This facility can be provided to users of the SeaDataNet CDI service as an extension to the regular functionality for browsing metadata and requesting access to the archived data sets. A wider implementation will be evaluated in the near future.

3 - Robustness

3.1 Monitoring of the infrastructure:

The SeaDataNet II infrastructure consists of interconnected data centres that provide harmonized discovery, access and viewing services to metadata, data, aggregated data sets and data products.

Since the aim of SeaDataNet II has been to provide services highly available and accessible to users, several actions have been planned to minimize unavailability and failure of the services provided. Monitoring is crucial when implementing widely distributed systems over the Internet. SeaDataNet II had to permanently monitor the system, using appropriate IT tools, to detect failures of systems components and to warn persons in charge.

The benefit of monitoring is not only on the short term to be able to alert the right people when incidents are detected in order to correct them as soon as possible, but on the longer term, to identify critical components within the widely distributed system in order to update them to improve their robustness.

At the beginning of the project a monitoring plan for system monitoring and metrics was defined (Deliverable D2.4), it planned the implementation of the monitoring service in nine successive phases which have all been completed successfully at the end of SeaDataNet II.

The SeaDataNet Monitoring System, developed by HCMR, is a service that monitors through Internet the SeaDataNet services (web pages, Download Managers, etc.) to confirm that they are accessible and available, notify persons in charge to take action in case of service failure and keep statistics. It is based on Nagios (TM) software. The system constantly checks the network state detecting for slow speed of connectivity or failure components. The monitoring system, automatically, notifies the network administrator in case of issues.

The SeaDataNet components that are monitored are divided into two groups of services:

- The **Core services**, which are centrally-based provided services:
  - Common Data Index (CDI) portal
  - European Directory of Marine Organisations (EDMO) portal
  - European Directory of the initial Ocean-observing Systems (EDIOS) portal
  - European Directory of Marine Environmental Research Projects (EDMERP) portal
  - European Directory of Marine Environmental Data Sets (EDMED) portal
  - Cruise Summary Reports (CSR) portal
  - SeaDataNet homepage
  - SeaDataNet Central Authentication Service (CAS)
  - Common Vocabularies Web Services
  - Request Status Manager (RSM).
- The **Local services**, which are services that are provided by the partners' infrastructures.
  - 49 Download Managers installed locally in 49 data centres (for SeaDataNet II project).

Nowadays the delivered operational monitoring system is monitoring in total 13 core services and 86 Download Managers supporting not only SeaDataNet projects but also other associated projects such as Geo-Seas, UBSS (Upgrade Black Sea Scene), EMODNet-Chemistry and EMODNet-Bathymetry. The monitoring results are also used in regular reports to oversee the overall performance of the SeaDataNet infrastructure and to analyse 'weak links'.

On-top of the monitoring services, an online web interface has been developed in order to give a tool, to all connected data centres, to search and access a variety of detailed information about their own servers and services status. It gives to the responsible of local data centres, an excellent instrument to improve the knowledge of quality of the services provided. To correct their services or server malfunctions local administrators need more than a notification email; they need to have access to log files to get more detailed information about the source of the problem. That is the role of this user friendly SeaDataNet Monitoring Portal which is addressed to local administrators who are responsible to maintain their services up and running.

Using the portal they can:

- search and access a variety of detailed information about their own servers and services status and logs,
- view all SeaDataNet services status of operation (availability),

- record and show the history of their servers/services status of operation,
- calculate an availability indicator for each of their services that reports as percentage their uptime,
- calculate a global availability indicator which represents the total uptime percentage of all services related to SeaDataNet.

In the latest stage of SeaDataNet II, the monitoring services have been duplicated. They are now installed in HCMR (Greece) and in OGS (Italy) in order to crosscheck monitoring data and to avoid false alarms.

3.2 Duplicate checks

Duplicate data sometimes occur and represent a problem for data management, for example different data providers may submit the same data sets, because data were collected in the same oceanographic cruise. Also it can happen that CDI entries are produced for duplicate data sets by the same data centre, possibly due to small differences in metadata. Therefore, to ensure integrity of the CDI Data Discovery and Access service and to improve the overall quality of data deliveries by SeaDataNet infrastructure, a global potential duplicate check has been done on the central CDI database. This was through a cooperation between the SeaDataNet and the MyOcean (now Copernicus) projects for preparing a European Temperature and Salinity (T&S) climatology. A procedure and algorithm have been formulated and applied using ODV software. This has identified a range of potential duplicates, of which their providers have been informed. Consequently actions to solve the issues have been undertaken. In many cases, potential duplicates were not real duplicates, e.g. in case of repeated measurements at the same time and location, but at different water depths. In other cases, CDI entries had to be de-activated.

In order to avoid in the future further duplicates in new entries to the CDI service, the duplicate detection algorithm has been integrated into the CDI validation and import procedures as operated by MARIS as central CDI manager. Every new CDI submission is validated this way for potential duplicates and reports are forwarded back to data providers for checking and possible correction before giving the green light for moving new and updated CDI entries from the import service to the production service.

4 – Products

The regional products of SeaDataNet are made on the 6 following sea areas: Arctic Ocean, North Atlantic Ocean, Baltic Sea, North Sea, Mediterranean Sea and Black Sea. They are described in the SeaDataNet product catalogue (SEXTANT) and are freely available for identified Marine-ID users.

Data access services made available by SeaDataNet have been used to prepare these SeaDataNet regional products which are of 2 different types: (1) the aggregated datasets of temperature and salinity which are ODV collections of all SeaDataNet measurements of temperature and salinity by sea basins and (2) the temperature and salinity climatologies which are regional gridded field products based on the aggregated datasets.

The preparation of these products is a collaboration between the following:

- MARIS for the data harvesting,
- AWI for the preparation of the temperature and salinity collection,
- ULg for the training workshop and the assistance on DIVA software
- the 6 regional coordinators (IMR – for the Arctic Ocean, IFREMER for the North Atlantic Ocean, SMHI for the Baltic Sea, RBINS-BMDC for the North Sea, INGV for the Mediterranean Sea and METU-IMS for the Black Sea) for the quality check of the data and the compilation of the aggregated datasets and the climatologies.
- All CDI partners for the correction of the data anomalies or errors that are detected during the quality check step.

In addition, a collaboration with the MyOcean 2 project has been undertaken in order to share the SeaDataNet recent data (1990-2012) with them and to include these data in a common MyOcean–SeaDataNet product.

4.1 Aggregated datasets

The aggregated datasets are regional ODV collections of all temperature and salinity measurements contained within SeaDataNet database covering all the European sea basins (Arctic Sea, Baltic Sea, Black Sea, North Sea, North Atlantic Ocean and Mediterranean Sea).

Two versions of the aggregated datasets have been published during SeaDataNet II (version 1.1 and version 2). Each of them is the result of the Quality Check Strategy (QCS) implemented during SeaDataNet II that contributed to greatly improve the quality of temperature and salinity data. The QCS was carried out in four main phases:

1. data harvesting from the central CDI
2. files and parameters aggregation
3. quality check analysis at regional level
4. analysis and correction of data anomalies.

These aggregated datasets have been prepared using ODV software for the quality checks of the data. They are distributed as ODV collection files. They have been divided into 2 parts: (1) all the SeaDataNet license/unrestricted data on one hand, (2) all the restricted data on the other hand. Both parts have been used for the climatology computation, but only part (1) is distributed through the SeaDataNet SEXTANT product catalogue and available from SeaDataNet web site.

4.2 Climatologies

SeaDataNet gridded climatologies are based on the aggregated datasets v1.1. The preparation of these products has also improved the quality, the consistency and the overall coherence of the data made available by SeaDataNet.

They have been computed using **DIVA** software tool (Data-Interpolating Variational Analysis), developed by the University of Liege. **DIVA** allows the user to spatially interpolate (or analyse) observations on a regular grid in an optimal way. The analysis is performed on a finite element grid allowing for a variable spatial resolution and a good representation of the coastline and isobaths. As some areas covered in the European seas have complex coastlines, the finite-element grid of DIVA will be able to adequately resolve those.

4.3 Product catalogue, viewing and download services

The regional climatologies and the aggregated data collections have been described with metadata and can be discovered in the SeaDataNet Sextant products catalogue, operated by IFREMER, which can be found at: http://sextant.ifremer.fr/en/web/seadatanet/

The additional **OceanBrowser Viewing** and **Oceanotron** services, which interact with the Sextant catalogue and the products, are described further in the SeaDataNet tools section.

5 - SeaDataNet tools

SeaDataNet provides documentation and common software tools for metadata and data formatting, for Quality Control (QC)/Quality Assurance (QA), and for statistical analysis and also versatile software packages for data analysis and presentation. These tools can be downloaded from the SeaDataNet portal without any restriction.
Several categories of tools have been developed and maintained to support data management and data processing in the SeaDataNet context:

5.1 Tool for metadata formatting

MIKADO is an ISO-19115 and ISO-19139 XML catalogue description generator used to create XML files for metadata exchange of the SeaDataNet catalogues: CSR, EDMED, CDI, EDMERP and EDIOS. MIKADO can be used in manual mode and in automatic mode, to generate XML files automatically if information is catalogued in a relational database or in an Excel file.

New versions of MIKADO are released via the SeaDataNet portal following project requirements and user feedback, and can be downloaded from the SeaDataNet web site by any user without any constraints.

MIKADO has been available for SeaDataNet users since 2007, developed during the first stage of SeaDataNet. During SeaDataNet II, twelve releases have been delivered to take into account the new major functionalities such as the migration to NVS 2.0, the implementation of ISO-19139 schemas for CDI and CSR and for other smaller changes or bug corrections.

5.2 Tools for data files formatting

NEMO is a reformatting software used for data exchange between SeaDataNet data centres and the portal to users. Its objective is to reformat ASCII files of vertical profiles (like CTD, Bottle, XBT) or time-series (like current meters, sea level data) or trajectories (like thermosalinograph data) to a SeaDataNet format (ODV, NetCDF or MEDATLAS). As the entry file can be all kinds of ASCII format, NEMO must be able to read all these formats to translate them to one of the SeaDataNet data transport formats.

Regularly new versions of NEMO are released via the SeaDataNet portal following project requirements and user feedback; they can be downloaded from SeaDataNet web site by any user without any restrictions.

NEMO has been available for SeaDataNet users since 2007, developed during the first stage of SeaDataNet. During SeaDataNet II, ten releases have been delivered to take into account the new major functionalities such as the conversion to SeaDataNet NetCDF format, the migration to NVS 2.0, the migration to CSR ISO-19139, the integration of format extensions for more interoperability and other smaller changes or bug corrections.

Two format converters have been delivered in order to be able to convert existing formats to the new SeaDataNet NetCDF format; these tools can also be downloaded from SeaDataNet web site without any constraints:

- The **MedSDN2CFPOINT** software, developed by IFREMER, reformats **MEDATLAS SeaDataNet** files of vertical profiles, time series or trajectories to **NetCDF SeaDataNet** files.
- The **OdvSDN2CFPOINT** software, developed by IFREMER, reformats **ODV SeaDataNet** files of vertical profiles, time series or trajectories to **NetCDF SeaDataNet** files.

5.3 Tool for data downloading

The DOWNLOAD MANAGER (DM) is the software used to download data files from the SeaDataNet portal.

The DM has to be installed and configured at each SeaDataNet node to enable receiving and processing data requests from users submitted to the SeaDataNet portal.
The first operational version of the DM was delivered at the beginning of 2010, during the first phase of SeaDataNet.

The DM consists of the following components:

- **RSM-DM request dialogue interface** provides interaction with RSM for receiving user requests.
- **DOWNLOAD MANAGER Request processor** (with Read-method and local application) provides interaction with the local data at the SeaDataNet node and prepares the zipped data file (Datasets pre-processed in SeaDataNet formats or via a read-method and preparation of the datasets from a database system) or enable advanced services (e.g. viewing service on seismic datasets).
- **RSM-DM response dialogue interface** provides interaction with the RSM for reporting about transport data status. The RSM interface gives users an overview of the status of their individual requests for datasets at each SeaDataNet node. This status is updated continuously by communication from the DM at each data centre.
- **DOWNLOAD MANAGER User page for downloading** contains a download section and a visualisation section and allows the user to get files or view datasets.
- **DOWNLOAD MANAGER checker and cleaner batches** allow SeaDataNet nodes to perform checks on the data and DM system and to maintain users' directories.
- **DOWNLOAD MANAGER monitor and status pages** allow Nagios monitoring by HCMR and gives an overview of the DM status to RSM.
- **DOWNLOAD MANAGER splitter** to extract one CDI from a multiCDI data file

Regularly new versions of the DM are released via the SeaDataNet portal following project requirements and user feedback; this can be by data centres that show interest in getting connected to the SeaDataNet infrastructure and making their data discoverable and accessible through the SeaDataNet portal. Interested data centres beyond the consortium are advised to contact the SeaDataNet support desk (sdn-userdesk@seadatanet.org). Installation and configuration requires professional guidance from the SeaDataNet overall support desk and the CDI support desk.

During SeaDataNet II, seven releases of the DM have been delivered to take into account new important functionalities such as the integration of Nagios monitoring, compliance with the

upgraded vocabulary server (NVS 2.0), support of the SeaDataNet NetCDF format, improvement of the DM_Checker component and other smaller changes or bug corrections.

5.4 Tool for data analysis and presentation

Ocean Data View (ODV) is a freely available software package for the interactive exploration, analysis and visualization of oceanographic and other geo-referenced profile, time-series and trajectory data. The ODV data format allows dense storage and very fast data access. Large data collections with millions of stations can easily be maintained and explored on inexpensive desktop and notebook computers. Data downloaded from the SeaDataNet infrastructure can be directly imported into ODV.

SeaDataNet has adopted the ODV software package as its fundamental data analysis and visualisation software. ODV provides interactive exploration, analysis and visualisation of oceanographic and other geo-referenced profile or sequence data. It is available for all major computer platforms and currently has more than 40,000 registered users. ODV has a very rich set of interactive capabilities and supports a very wide range of plot types. This makes ODV ideal for visual and automated quality control. The latest release overcomes many limitations of previous versions and now supports more flexible metadata models, an unlimited number of variables and custom quality flag schemes, and is fit for loading and aggregating data sets in the SeaDataNet ODV ASCII format.

The ODV software is also being used in SeaDataNet for producing generic data products for each of the regional seas for various variables. As a part of SeaDataNet, the DIVA method (see below) has been integrated into ODV, and the integration greatly facilitates the usage of DIVA. Features supported by the ODV/DIVA integration include proper treatment of domain separation due to land masses and undersea ridges or seamounts and the realistic estimation of water mass properties on both sides of the divides. This is important in areas, such as the Kattegat, with many islands separated by narrow channels.

The Ocean Data View (ODV) software has been used for three crucial steps in the creation of the SeaDataNet aggregated and quality controlled dataset: (1) aggregation of the more than 1.3 Million individual SeaDataNet data files for Temperature and Salinity into a single integrated dataset; (2) aggregation and harmonization of parameters using the new P35 vocabulary; and (3) duplicate station detection and quality control of the data using automatic and visual procedures.

Most of the functionality needed for file and parameter aggregation in steps 1 and 2 above were developed and optimized by partner AWI during the SeaDataNet II project phase. This also includes implementation of efficient duplicate station detection algorithms. Many quality control (QC) techniques were refined and generalised during the project phase. The QC procedures were taught to SeaDataNet partners during training courses at the IOC Project Office for IODE in Oostende.

The SeaDataNet data importer in ODV now imports all CDI metadata and extracts instrument metadata as well as additional data references from the data files. Many metadata items, such as CDI reference, EDMO code, data references, as well as instrument and parameter codes are clickable in ODV thereby displaying the respective SeaDataNet webpages in the web browser.

In addition to the work on the ODV software itself, partner AWI also developed and released an ODV Application Programming Interface (ODVAPI). The ODVAPI allows other developers to write their own standalone or web-based applications to access the metadata and data held in ODV collections and process the information for their specific needs. This API is successfully employed by partner IFREMER to serve SeaDataNet data in aggregated ODV collections over the Internet.

A total number of 18 new ODV versions and 3 ODVAPI versions were released during the SeaDataNet II project phase.

5.5 Tool for statistical analysis of data

In practice, in-situ measurements can be sparse and heterogeneously distributed. The freely available DIVA software tool (Data-Interpolating Variational Analysis) allows the user to spatially interpolate (or analyse) those observations on a regular grid in an optimal way. The analysis is performed on a finite element grid allowing for a spatial variable resolution and a good representation of the coastline and isobaths. As some areas covered in the European seas have complex coastlines, the finite-element grid of DIVA will be able to adequately resolve those. It is also possible to compute error maps for the gridded fields which reflect the accuracy of the observations and their distribution. This allows assessment of the reliability of the gridded fields and to objectively identify areas with poor coverage.

Within the time frame of the SeaDataNet II project, 12 new releases of DIVA have been made available. The improvements with particular relevance for SeaDataNet include a new efficient algorithm to derive the error fields, parallel and iterative solver, optimized grid generation, better handling of time series and the inclusion of the metadata of the observations used in the final analysis. The full list of changes is available at http://modb.oce.ulg.ac.be/mediawiki/index.php/New_Diva_Features.

The XML metadata of the DIVA analysis are generated with the tool divadoxml (8 releases in the frame of the project).

The web interface of DIVA (diva-on-web) now also allows use of an advection constraint and addition of NetCDF attributes to enrich the analysis produced with relevant metadata. DIVA analyses are visualised with the tool OceanBrowser which allows extraction of horizontal sections and arbitrary vertical sections.

The **OceanBrowser Viewing service**, developed and maintained by University of Liege (Ulg), provides access to the DIVA interpolated maps. Output images are available as horizontal sections and vertical sections. The latter can be selected by drawing an appropriate transect. ULg together with IFREMER have integrated the Sextant data products catalogue with the OceanBrowser. This now gives users the option to select SeaDataNet data products in the Sextant catalogue and then to visualise and browse them in OceanBrowser. Alternatively users can visualise data products in OceanBrowser and then to look up their description from the Sextant Catalogue.

In the frame of the SeaDataNet II project, the responsiveness of this web-application has been improved, the support for animation has been enhanced and the in situ visualisation service

(Oceanotron) has been integrated. In total 15 releases of OceanBrowser have been distributed on docker hub.

In addition a new service has been added by adaption of **Oceanotron** for giving access and visualization services to the SeaDataNet data collections. Oceanotron has been initiated by IFREMER in the framework of the MyOcean project and it has been further developed as part of SeaDataNet II and the MyOcean projects. It gives advanced services on top of in-situ data collections.

The visualisation and interactivity of the regional T&S data collections is arranged by using Oceanotron. In practice it works as follows:

- The regional T&S data collections have been prepared with the ODV (Ocean Data View) software and are managed as ODV binary collections;
- The ODV software has been equipped by AWI with an API to allow reading the ODV binary collections;
- A plug-in has been developed by IFREMER to enable Oceanotron to make use of the ODV API and in this way to interact with the data collections;
- Furthermore WMS services have been implemented as part of Oceanotron using ncWMS from the University of Reading to enable interacting and sharing results of interactions of Oceanotron with the data collections ;
- The Oceanotron WMS service has been integrated into the OceanBrowser.

6 - SeaDataNet metadata directory services

The SeaDataNet infrastructure comprises a network of interconnected data centres and a central SeaDataNet portal. The portal provides users not only background information about SeaDataNet and the various SeaDataNet standards and tools, but also a unified and transparent overview of the metadata and controlled access to the large collections of data sets, managed by the interconnected data centres.

SeaDataNet maintains and operates at its portal several discovery services with overviews of marine organisations in Europe and their engagement in marine research projects, managing large datasets, and data acquisition by research vessels and monitoring programmes for the:

- **EDMO**: European Directory of Marine Organisations,
- **EDMED**: European Directory of Marine Environmental Data,
- **EDMERP**: European Directory of Marine Environmental Research Projects,
- **CSR**: Cruise Summary Reports,
- **EDIOS:** European Directory of the Ocean Observing Systems,
- **CDI:** Common Data Index, the data discovery and data access service.

The SeaDataNet architecture has been designed to provide a coherent system of services. For implementation, a range of technical components have been defined and developed. These make use of recent web technologies, and also comprise Java components, to provide multi-platform support.

Each of these central directory services has been upgraded during SeaDataNet II as part of the innovation cycle. The technical upgrades have been focused on:

- Adopting the CDI and CSR services to the new INSPIRE compliant ISO-19139 based XML schemas
- Adopting the Common Vocabularies Version 2 (NVS 2.0) in all metadata directories (CDI, CSR, EDMED, EDMERP, EDIOS) and data formats (SeaDataNet ODV, SeaDataNet NetCDF)
- Adopting a more effective transfer of updated and new metadata entries for CDI and CSR by applying OGC CS-W harvesting between central portals and distributed data centres
- Installing and configuring new visualisation and data delivery services for the CDI service consisting of OGC WMS, OGC WFS, OGC CS-W, OAI-PMH, OGC OpenSearch services and the robot harvesting and buffer system
- Optimising the CDI discovery and shopping mechanism e.g. by adding more search criteria, tuning performance, enlarging the shopping basket from 500 to max 10,000 entries, adding extra functionalities such as the search result summary with drilling down capability and extra reporting and analysis options in the Request Status Manager service.

Each directory service includes components for import, validation, storage, discovery, retrieval and publishing with a considerable interaction and interrelationship between them. Therefore a roadmap has been developed with an analysis of which tools and services had to be upgraded and in what planning phase. Thereafter this plan has been followed and successfully resulted in the upgraded directory services.

OpenSearch has been added to the CDI service. It is an interesting protocol supported by major portals such as Google, Yahoo, Twitter. OpenSearch provides metadata about the contents along with a set of URL Templates which illustrate the parameters accepted by the service and the variety of output formats in which results can be obtained. The OpenSearch request interface is simple, consisting of a description of an HTTP GET request with a series of optional key-value parameters that can be used to constrain the search: Free search; Geospatial (area or point + radius); Temporal (from to). A SeaDataNet OpenSearch implementation has been developed that considers the full scope of SeaDataNet CDIs. The OpenSearch principle is that so-called OpenSearch entry points are prepared in JSON XML or RDF. Each file describes an access point as an aggregation of the marine data as available in SeaDataNet. To make the entry points not too large in number of CDIs and to provide distinctive attributes to each aggregation of CDIs it was decided to prepare SeaDataNet entry points as collections just as defined, applied and earlier described for the CDI OGC CS-W service. Querying is made possible on spatial and temporal criteria PLUS free search on the following CDI metadata attributes: parameters (from vocabularies P02), station name (free text), alternative station name (free text), cruise name (free text), and alternative cruise name (free text).

Another way to achieve interoperability and viewing has been achieved by making use of the Open Geospatial Consortium (OGC) **Web Map Service (WMS) and Web Feature Service (WFS)** protocols to exchange metadata including URLs to further metadata and data. WMS makes it possible to share and exchange map layers between internet portals with map viewing services and WFS makes it possible to interrogate the objects on these maps by clicking on

those thereby retrieving feature information. The feature info then can include URLs to the associated portal to retrieve further metadata and options to request access and delivery of the data sets. The SeaDataNet WMS GetCapabilities specifies how the WMS and also the WFS can be called and integrated in another portal. However implementing WFS is dependent on the client and needs programming.

The SeaDataNet infrastructure has been well positioned in many other projects funded by the EU as part of the RTD programme and as part of EMODnet. The SeaDataNet standards and CDI infrastructure has been adopted in the following projects: Geo-Seas, Eurofleets 2, JERICO, JERICONEXT, Upgrade Black Sea SCENE, CaspInfo, Micro-B3, CLIPC, CitClops, AtlantOS, CoCoNet, EMODnet Chemistry, EMODnet Bathymetry, EMODnet Geology, EMODnet Physics, and EMODnet Biology.

This has resulted in a steady increase of the entries in each of the directory services over the lifetime of the SeaDataNet II project:

- **EDMO**: European Directory of Marine Organisations, from ca 2100 to ca 3600 entries
- **EDMED**: European Directory of Marine Environmental Data, from ca 3800 to ca 4050 entries
- **EDMERP**: European Directory of Marine Environmental Research Projects, from ca 2300 to ca 2900 entries
- **CSR**: Cruise Summary Reports, from ca 32000 to ca 46500 entries
- **EDIOS:** European Directory of the Ocean Observing Systems, from ca 294 to ca 365 observing programmes
- **CDI:** Common Data Index, the data discovery and data access service, from ca 1.05 M to ca 1.8 M entries.

The CDI Data Discovery and Access service gives users a highly detailed insight in the availability and geographical spreading of a large variety of marine and ocean data sets, that are managed by the data centres connected to the SeaDataNet infrastructure. Moreover it provides a unique interface for requesting access, and if granted, for downloading data sets from these distributed data centres across Europe. All users can freely query and browse in the CDI directory; however submitting requests for data access via the shopping basket requires that users are registered in the SeaDataNet central user register, thereby agreeing with the overall SeaDataNet User Licence. This registration also allows users to submit a request for multiple data sets from multiple data centres in one go and to follow-up its processing via their personal user account. All data requests are forwarded automatically from the SeaDataNet portal to the relevant data centres. This process is controlled via the Request Status Manager (RSM) service at the portal, that communicates with the data centres. Users receive a confirmation e-mail of their data set requests and a link to the RSM service. By log-in to the RSM service users can check regularly the status of their requests and download data sets from the associated data centres, after access has been granted. In turn, data centres can follow via the RSM service all transactions for their data sets online and can handle requests which might require their mediation.

The number of connected data centres has increased over the SeaDataNet II project from 65 to now 102 data centres and the number of CDI entries has increased from circa 1.05 M to 1.8 M. The data

centres are based in 34 countries, riparian to European seas, and the data sets cover physical oceanography, chemistry, geology, geophysics, bathymetry and biology, originating from more than 550 organisations. All the data sets are collected in research projects and monitoring programmes and after a process of quality control and formatting, archived in the local databases of the data centres that are connected to the SeaDataNet CDI Data Discovery and Access service.

Potential impact (including the socio-economic impact and the wider societal implications of the project so far) and the main dissemination activities and the exploitation of results.
(The length of this part cannot exceed 10 pages).

SeaDataNet and its partners cooperate with many other EU projects and EU initiatives in order to expand its infrastructure with more data centres and more data sets, and to develop further their standards and services.

The European Commission states in its communication to the European Parliament and the Council entitled "MARINE KNOWLEDGE 2020 - marine data and observation for smart and sustainable growth" that the creation of marine knowledge begins with observation of the seas and oceans. An impact assessment estimates that the collection of marine data by public institutions in EU Member States costs more than €1 billion annually, with a further €0.4 billion for marine-related satellite data. A related public consultation concluded that users continue to find it hard to discover what data already exist. This can be due to restrictions on access or the pricing policy of some providers. Moreover, fragmented standards, formats and nomenclature, lack of information on precision and accuracy, and insufficient temporal or spatial resolution are further barriers.

SeaDataNet provides an operational infrastructure and is also further developing its services and products to meet the challenges of the Marine Knowledge communication, and also the EU INSPIRE and Marine Strategy Framework Directives. In 2008 the EU initiated the development of an overarching European Marine Observation and Data Network (EMODnet). This must facilitate long-term and sustainable access to the high-quality data necessary to understand the biological, chemical and physical behaviour of seas and oceans.

Partnerships from the SeaDataNet consortium successfully bid to develop and manage a number of the EMODnet prototypes and follow-on projects, such as for example the EMODnet Bathymetry and Chemistry portals, which began in 2009. These portals make use of the SeaDataNet standards and basic infrastructure, and have provided an excellent opportunity to engage and reach out to other data centres. This has resulted in more data centres connecting to the SeaDataNet infrastructure and populating their metadata and data to the SeaDataNet services. It has also resulted in the development of a number of interesting products and services. In this way the SeaDataNet consortium and infrastructure contribute to play a major role in the development and actual implementation of EMODnet.

SeaDataNet has focused, with success, on establishing common standards and on applying those standards for interconnecting the data centres enabling the provision of integrated online access to comprehensive sets of multi-disciplinary, in situ marine data, metadata and products. It is able to support a wide variety of data types and to serve several sector communities. SeaDataNet is also actively sharing its technologies and expertise, spreading and expanding its approach, and building bridges to other well established infrastructures in the marine domain. This has resulted in adoption and an active role for a number of SeaDataNet partners in many related data management projects, such as Geo-Seas, Upgrade Black Sea Scene, CaspInfo, EuroFleets 1 and 2, Jerico, Jerico Next, ClipC, Micro-B3, AtlantOS, ENVRIplus and others.

SeaDataNet cooperates with EuroGOOS, the association of national governmental agencies and research organisations committed to European-scale operational oceanography within the context of the intergovernmental Global Ocean Observing System, and the consortium, that is implementing and operating the Copernicus Marine Environment Monitoring Service (CMEMS). The cooperation focuses on improving the availability of high quality and harmonised physical oceanography data sets in real-time and delayed mode, through long term archives, in support of operational oceanography.

In addition, SeaDataNet partners cooperate with the Partnership for Observation of the Global Oceans (POGO), EurOcean, International association of Oil and Gas Producers (IOGP) (SIMORC service), and leading oceanographic institutes from USA and Australia in the frame of the Ocean Data Interoperability Platform (ODIP) project.

Training courses, capacity building

Innovation for managing data, training and career development have been addressed by the SeaDataNet II project. By organising training courses, SeaDataNet II has contributed to the development of practices of the emerging field of marine data scientist activities and built an overall capacity for marine data management in all of the countries participating in the project. Linkage with the IOC-IODE OceanTeacher programme ensures that materials developed for training are preserved and are also available for international dissemination and use by other countries outside Europe, thus contributing to international standardisation.

Two training courses covering several domains (data quality, data analysis, data management, standards adoption and implementation) have been organised during the project. Participants were data manager or technicians at the data centres and scientists.

The objectives of these training courses were:

- Training of data managers and technicians of data centres in the operation of SeaDataNet services;
- Training in the uptake of upgraded standards and data management procedures and in the implementation and configuration of new infrastructure components;
- Bringing all data centres onto an equal footing.

The aim for the participants was:

- To learn about the infrastructure and the general data and metadata management rules to be applied to realise a SeaDataNet node
- To learn how to connect the node to the SeaDataNet infrastructure
- To learn how to use the SeaDataNet tools
- To become acquainted with the new developments

42 people attended the first training course (2-6 July 2012), and 38 attended the second training course (20-22 May 2014).

Data Management conference

The cycle of conferences focusing on marine data management (International Marine Data and Information System Conference - IMDIS -) started ten years ago with the first conference

organised in Brest, France in 2005. Then in 2008 and 2010 two of these conferences were organised in the frame of the first stage of SeaDataNet.

The IMDIS cycle of conferences has the aim of providing an overview of the existing information systems to serve different users in ocean science. It also shows the progress of development of efficient infrastructures for managing large and diverse data sets, standards, interoperable information systems, services and tools for education.

The fourth edition of IMDIS conference was organised as part of WP6 activity of SeaDataNet II: IMDIS 2013, was held on 26 – 27 September 2013 in Lucca (Italy). It was organised by: ENEA, IFREMER, OGS (abstracts) and INGV (Side Event). IMDIS 2013 was a great success with 180 participants from 37 Countries, including some representatives from the IOC UNESCO, FAO, GOOS and EC-JRC and many participants from non-European countries, mainly the USA and Australia. There were 50 oral presentations and 68 posters. The proceedings have been published in Vol. 54 – supplement of "*Bollettino of Geofisica"*. An electronic version of the book of abstracts is also downloadable at this web address: http://imdis2013.seadatanet.org/Proceedings. All the presentations have been video recorded by IOC-IODE and are available, together with the pdfs of the oral presentations and posters: http://imdis2013.seadatanet.org/Presentations.

The Conference presented different systems for on-line access to data, meta-data and products, communication standards and adapted technology to ensure platform interoperability. Sessions have focused on infrastructures, technologies and services for different users: environmental authorities, research, schools, universities…

The IMDIS 2013 Conference was organised in four sessions:

1. **Marine information and data management**
   - Exchange, processing and interactive work with marine data sets from highly heterogeneous sources
   - Federation and integration
   - Network services and technologies
2. **Marine environmental data bases: infrastructures and data access systems**
   - Coastal and deep-sea operational oceanography metadata/data systems
   - Physical and bio-chemical databases for climate studies
   - Geophysical and geological metadata/data systems
3. **Data Services in ocean science**
   - Standards and quality-assurance issues
   - Services and Visualisation tools
   - User oriented services
4. **Services for Users and Education**
   - Historical evolution in data collection and management
   - Tools for dissemination
   - Test bed development for educational purposes

The **dissemination activities** implemented within the project were defined in the "Promotion and Dissemination plan" developed early in SeaDataNet II (Deliverable D6.1) and must realise the 3 following main tasks:

- reinforce the SeaDataNet base of data holders; increase the number of holding centres in the project as well as the amount of data to be accessed freely and openly;
- maintain/reinforce the collaboration with other initiatives on informatics;
- organise a feedback from the user community in order to improve the services of the infrastructure.

With the following objectives:

1. Increase the number of data centres participating to the infrastructure
2. Increase the amount of data
3. Stimulate the open and free access to data centres
4. Collaborate with other initiatives in informatics
5. Open collaboration with users and receive feedback
6. Increase the capacity of each partner and reach a major internal homogenisation
7. Suggest the adoption of SeaDataNet standards at the international level
8. Present SeaDataNet as the Pan-European infrastructure capable of providing data to major EU organisations and programs (EEA, GMES, etc.)
9. Enhance the link between SeaDataNet and EMODnet

Each of these objectives have been addressed in the frame of the project as described hereafter:

1. SeaDataNet interconnects most of the marine data centres from countries bordering the European seas, at present more than one hundred data centres are connected to the infrastructure, the number of connected data centres has increased from 65 to 102 since the beginning of SeaDataNet II.

2. An enormous amount of data is also available; 1.8 million datasets can be discovered and downloaded from the SeaDataNet web-site. The number of data sets has also almost doubled since the beginning of the project (from 1.05 M to 1.8M)

3. All metadata services are public domain. The major part of data (85%) is freely available. The remaining 15% data have access restrictions and the user has to make arrangements with the data provider to negotiate access for downloading.

4. The collaboration with other initiatives in informatics was implemented in the frame of the SeaDataNet Technical Task Group (TTG). TTG participants are all highly skilled in IT and also involved in many other projects. In addition, through participation in the ODIP project, the USA and Australian partners shared their technical experience with European participants (many of them partners of SeaDataNet II). Specific sessions organised at EGU conferences in 2013, 2014 and 2015, chaired by the SeaDataNet II technical coordinator, further increased this collaboration: the theme of these specific sessions was 'Informatics in Oceanography and Ocean Science'.

5. User feedback has been addressed during the project in order to improve the access to data. A user panel was established aiming to involve as many different types of users

as possible, from the scientific community to the policy makers, industry or private companies, education and data managers. Two questionnaires have been sent to this user panel during the project (one of them, jointly with the EMODnet chemistry project). Suggestions made by the users have been taken into account as much as possible.

6. The capacity of each partner of the project, and more particularly of the small data centres which do not have the same means as the larger ones, has been greatly increased by the development of the SeaDataNet tools demonstrated during the training courses and distributed to all partners.

7. SeaDataNet has developed and maintains a set of common standards for the marine domain, collaborating with European and international experts within the framework of IOC-IODE, ICES, and ODIP, adopting and adapting ISO and OGC standards, and achieving INSPIRE compliance. These comprise: the adoption of **the ISO 19115 – 19139 standard for describing metadata** on data sets, the **controlled vocabularies** for the marine domain with international governance, the user interfaces and web services and the standard **data exchange formats.** These standards have been submitted have been submitted to the Ocean Data Standards and Best Practices Project (ODSBP), the continuation of the Ocean Data Standards (ODS) Pilot Project of JCOMM-IODE, for sharing with other ocean communities worldwide (Deliverable D8.9).

8. The SeaDataNet Infrastructure has been adopted by the following European projects: Black Sea Scene, CASPINFO (Caspian environmental and industrial data & information service), EMODnet (The European Marine Observation and Data Network) for the Bathymetry, Chemistry, Geology and Biology themes. Interoperability with other data portal like GEOSS (Global Earth Observation System of Systems) and IODE-ODP (International Oceanographic Data and Information Exchange-Ocean Data Portal) has also been developed.

9. EMODnet Chemistry started in August 2013 as a follow up to the earlier pilot project. The project is well underway in achieving its goals of assembling measurements for chemicals related to eutrophication and contaminants and establishing a dedicated infrastructure that provides access to all assembled data and derived regional data products for all sea basins in the European waters in a uniform way. EMODnet Chemistry has adopted and contributed to upgrading of various SeaDataNet services and tools like the SeaDataNet CDI Data Discovery and Access service, ODV and DIVA software with specific developments and the data visualisation services. Another very interesting and illustrative example of what one can achieve with the SeaDataNet approach can be found at the EMODnet Bathymetry portal. Its development started in June 2009 as a pilot followed by 2 consecutive projects. At present almost 14,000 bathymetric survey datasets, managed by 27 data centres from 14 countries and originating from 167 institutes, have been gathered and populated in the EMODnet Bathymetry Data Discovery and Access service, adopting SeaDataNet standards.

   The EMODNet Physics project started early 2011 in cooperation between **EuroGOOS, MyOcean** and **SeaDataNet**. Many EuroGOOS members are also partners in SeaDataNet; and EuroGOOS and SeaDataNet have an agreement for maintaining and

operating the European Directory of Ocean Observing Systems (EDIOS). MyOcean and SeaDataNet have agreed a MoU with the aim of making available a comprehensive dataset of *in situ* observations from both operational oceanography programmes and scientific surveys to serve both the Operational Oceanography and research communities as well as other users. MyOcean and SeaDataNet also strive for common standards. The EMODnet Physics portal aims at improving for users the overview and access to in-situ physical data sets, both in near real-time and delayed mode. The EMODnet Physics portal also encourages other physical data providers outside the present communities to come forward, to contribute and to become engaged. A comparable effect can be seen in the other EMODnet projects. The pillars under EMODnet Physics comprise:

- The EuroGOOS regional operational observing systems (ROOSes), most of them collaborating with regional Conventions (HELCOM, OSPARCOM, MAP)
- The MyOcean in situ TAC providing access to physical near real time data acquired by continuous, automatic and permanent observation networks integrated in 6 regional portals and a global one operated by Coriolis
- SeaDataNet infrastructure, providing access to long term physical time series residing in a distributed system of data centres and giving information about the monitoring stations and their networks / programmes via the European Directory of Ocean Observing Systems (EDIOS).

In the portal the station and track locations contain direct links to view and browse near real-time time series of selected parameters in charts and tables, that are derived from the EuroGOOS ROOSes / MyOcean in-situ TAC system, and to submit requests for downloading near real-time data sets from the in-situ TAC system. The charts and tables are freely available for a sliding time window of 60 days. Stations might also contain direct links to retrieve and browse metadata records from the SeaDataNet CDI service for requesting access to quality controlled and long term time series of selected parameters. This facilitates to download data sets in common formats for agreed requests.

Dissemination activities implemented during the project were also the result of a collective effort as many SeaDataNet II beneficiaries conducted opportunistic dissemination actions through articles published on their own web sites or press releases on project's result and promoted the project in national, European or international conferences and workshops. Some university courses demonstrating SeaDataNet were also delivered.

The SeaDataNet Internet web-site was launched on November 30th 2011, 2 months after the start date of SeaDataNet II, at the URL: http://www.sedatanet.org,
It is a platform for presenting the project and for accessing the metadata, data, products and software made available by the distributed infrastructure.
Since its launch, the SeaDataNet web-site has been regularly updated with publications, news, events (meetings, training courses…), new software versions, newsletters…

A dedicated identity set (Logo, web-site and documents banner) and several printed materials have been developed to assist the project dissemination, such as SeaDataNet posters (4 versions produced during the project), SeaDataNet roll-up and SeaDataNet leaflets (4 versions produced during the project). These printed materials aim at dissemination of the project as

a whole, all material is available on the web site and can be downloaded and printed by all partners for presentations or distribution at conferences or workshops.