Advances in Geosciences

CDI/THREDDS interoperability in the SeaDataNet framework

S. Nativi[1,2], P. Mazzetti[1,2], M. Santoro[1], E. Boldrini[1], G. M. R. Manzella[3], and D. M. A. Schaap[4]

[1]IMAA of the Italian National Research Council, Tito Scalo, Italy
[2]University of Florence at Prato, Prato, Italy
[3]ENEA – ENEA CLIM, La Spezia, Italy
[4]MARIS – Marine Information Service, The Netherlands

Abstract. SeaDataNet is an EU funded project aiming to create and operate a pan-European, marine data infrastructure for managing the large and diverse datasets (i.e. temperature, salinity current, sea level, chemical, physical and biological properties) collected by the oceanographic fleets and the new automatic observation systems. In order to make the SeaDataNet system compliant with the INSPIRE Implementing Rules for discovery service, an ISO 19139 encoding of the SeaDataNet Common Data Index (CDI) metadata model was defined. Moreover, the problem of heterogeneous data sources has been addressed. In fact, a widely used system of SeaDataNet partners and oceanographic-marine community is THREDDS/OPeNDAP; this raises up the problem of federating into SeaDataNet framework THREDDS/OPeNDAP systems as well. In this paper we describe an interoperability framework to access resources (i.e. data and services) that are available through CDI and THREDDS/OPeNDAP services. The proposed solution implements a common catalog interface to discover and access the two heterogeneous resources in a common way. This catalog service is fully distributed and implements international standards as far as geospatial information discovery and query are concerned. The developed solution is called GI-cat and was experimented in the framework of the SeaDataNet European project.

1 Introduction

SeaDataNet is an EU funded project aiming to create and operate a pan-European, marine data management infrastructure, accessible online through a unique portal. The primary goal of SeaDataNet is to develop a system which provides transparent access to marine data sets and data products from 36 countries in and around Europe. Hence, it must develop a standardized distributed system for managing the large and diverse datasets (i.e. temperature, salinity current, sea level, chemical, physical and biological properties) collected by the oceanographic fleets and the new automatic observation systems. By use of standards for communication and new developments in information technology, the 40 in-situ and satellite marine data platforms of the partnership provide metadata, data and products as a unique virtual data centre[1]. The core of the SeaDataNet infrastructure is composed of some discovery services (e.g. EDMED, EDMERP, EDMO, CSR, EDIOS) and the Common Data Index (CDI)[2], a metadata model for geospatial resources encoded in a proprietary XML schema.

The tradition of creating formal data centres to house and distribute ocean data and products has long been shadowed by the parallel existence of direct data downloads from research groups (usually via HTML page-links and/or FTP servers). Historically, at least half of the data available to researchers has been provided outside the IODE/World Data Center System facilities[3] (Parker, 1992). During the last decades efforts have been dedicated to create a new system of Internet communication (e.g. the NASA Data Access Protocol, 2007, OPeNDAP[4], the UNIDATA THREDDS – Thematic Realtime Environmental Distributed Data Services; Domenico et al., 2002) that would allow users to access distributed datasets directly, using a family of "middleware services" that overlay datafile caches, allowing retrieval

---

[1]http://www.seadatanet.org/

[2]SeaDataNet, Common Data Index (CDI), available at http://www.seadatanet.org/data_access/common_data_index_cdi

[3]International Oceanographic Data Exchange http://www.iode.org/.

[4]OPeNDAP organization: "Welcome to OPeNDAP", available at: www.opendap.org.

of files in desired formats, no matter what the storage format might be.

The middleware service adopted by SeaDataNet for discovery and query makes use of web technologies to exchange CDI documents. The CDI metadata model is based on ISO 19115, a standard model for geospatial resources (Organization for Standardization, 2003). Since other data management systems based on THREDDS/OPeNDAP are widely used by the oceanographic-marine community and SeaDataNet partners, there is a clear need to build a "bridge" between the CDI realm and the THREDDS/OPeNDAP realm. This would allow to discover and access data provided by both CDI and THREDDS/OPeNDAP systems. This paper presents and discusses a distributed catalog solution, based on the GI-cat framework, addressing this issue.

Moreover, since the SeaDataNet project is developed in a more general European context, it is also necessary to consider the requirements of the EU INSPIRE Directive (2007) which represents the European Legislation concerning Spatial Data Infrastructures (SDIs) and entered in force on the 15 of May 2007.

The INSPIRE Implementing Rules (IRs)[5] identify the OGC CSW ISO Application Profile (AP)[6] as the reference standard for INSPIRE compliant catalog services. They specifically describe how to use ISO 19139[7] schemas to encode INSPIRE compliant metadata in a CSW response.

This paper is structured as follows: in the next sections brief descriptions of CDI, THREDDS/OPeNDAP and INSPIRE Directive are provided; after a survey of the most relevant standard discovery and access interfaces used for geospatial information interoperability, GI-cat is presented and discussed, detailing especially the mediation aspects - that is, the solution adopted to achieve data models harmonization and access protocols adaptation; finally, in the last two sections of this work we describe the successful experimentations done, the application client we developed for demonstrating GI-cat and the conclusions.

1.1 INSPIRE

Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE) was published in the official Journal on the 25th April 2007. This Directive defines the legal framework for setting up

---

and operating a European Spatial Data Infrastructure (ESDI) based on the infrastructures for spatial information (SDIs) of the Member States of the European Union. To ensure interoperability, the Directive requires that common Implementing Rules (IRs) and Technical Guidance (TG) are adopted for:

- Metadata

  - TG based on EN ISO 19115 and EN ISO 19119

- Data Specifications

- Network Services

  - Discovery: TG for an INSPIRE Profile of CSW ISO AP

- Data and Service Sharing

- Monitoring and Reporting

Providing an INSPIRE compliant interface for discovery is thus a crucial requirement for SeaDataNet as well as describing available resources with ancillary information (Metadata) which comply with ISO 19115 and 19119 standards and their implementation ISO 19139.

1.2 CDI

The CDI provides an index (metadatabase) to individual data sets. Furthermore, it paves the way to direct online data access or direct online requests for data access or file downloads; in fact, the primary objective of CDI is to give users a highly detailed insight in the availability and geographical spreading of marine data across the different data centres and institutes across Europe[2]. CDI was initiated in the EU Sea-Search project (Sea-Serch, 2010).

Currently, it is being further developed and extended by the SeaDataNet project. All the 36 countries participating in SeaDataNet are going to utilize CDI for describing their marine data[2].

1.2.1 ISO profile for CDI

SeaDataNet started its work in 2006, basing its metadata schema upon the ISO 19115 DTD, the available schema at that time. Overtime this was replaced with the present CDI v.1 XML schema, based on ISO 19115 abstract model with specific community features and additional constraints.

In order to ensure the INSPIRE conformity, a possible solution – as explained in the next sections- can be based on a front end component publishing the available data, based on CDI v.1, through a CSW AP ISO interface. This approach requires to define an ISO profile for CDI and its ISO 19139 schema encoding.

In our work, the first step consisted in the precise definition of a CDI metadata profile of ISO19115, containing both

---

[5]Drafting Team Metadata and European commission Joint Research Centre, "INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119", V.1.1

[6]Open Geospatial Consortium, Inc. OpenGIS® Catalogue Services Specification 2.0.2 – ISO Metadata Application Profile, Ver. 1.0.0 OGC 07-045.

[7]Organization for Standardization, ISO/19139:200t(E): Geographic information – Metadata – XML schema implementation. International Organization for Standardization, Geneva, Switzerland.

INSPIRE and CDI driven constraints and extensions. This included the check of all the metadata elements present in CDI and their cardinality. A comparison was then made with respect to ISO 19115, and possible extensions were individuated. Once this community profile of ISO 19115 was established, an encoding to ISO 19139 schemas was ready to be defined.

The general approach chosen by the SeaDataNet community was to introduce additional constraints to the ISO 19139 schema (e.g. to be checked using Schematron or other validation mechanisms), rather than to define the CDI profile as an ISO 19139 Extension Package.

Several issues were faced during the definition process. Among these: (a) dynamic lists and vocabularies used in the CDI could not be easily accommodated in ISO 19139; (b) time resolution information from CDI v.1 was difficult to accommodate as well; (c) ambiguities both in the ISO 19139 specification and in the INSPIRE regulations (e.g. regarding to the bounding polygon, the language and the role of the responsible party) needed to be solved.

Another outcome of this process is the set up of conventions regarding the protocol formats to be used for a useful machine-to-machine data access. Where already registered, Uniform Resource Names (URNs) were used to identify the proper protocols (e.g. the URN "urn:ogc:serviceType:WebCoverageService:1.0:HTTP" for the OGC WCS version 1.0 with HTTP binding); in other cases, widely used identifiers were adopted as de-facto standards (e.g. the "HTTP" string for the W3C HTTP protocol).

SeaDataNet plans to adopt this new defined profile and its implementation as a replacement of the current data model implementation (CDI v.1).

1.3 THREDDS

THREDDS is a middleware to bridge the gap between data providers and data users. The goal is to simplify the discovery and use of scientific data and to allow scientific publications and educational materials to reference scientific data (Domenico et al., 2002).

THREDDS initial focus was to allow data users to find datasets that are pertinent to their specific education and research needs, access the data, and use them without necessarily downloading the entire file to their local system. To achieve this, data providers must be able to publish lists of what data are available and to describe their data enabling discovery and use: THREDDS introduced the Data Inventory Catalogs concept[8]. They are XML documents that describe on-line datasets. These catalogs can contain arbitrary metadata; besides, a standard set of metadata to bridge to discov-

ery centers –like GCMD[9], DLESE[10] and NSDL[11] – was defined. The THREDDS/OPeNDAP catalog service augments traditional MeteoOcean access services such as OPeNDAP and netCDF (Eaton et al., 2003) access via HTTP. Moreover, THREDDS/OPENDAP implements access services to bridge the MeteoOcean and GIS communities, such the OGC Web Coverage Service interfaces.

2 Distributed catalog services

According to the general definition, geospatial data catalogs are discovery and access systems that use metadata as the target for query on geospatial information. Indexed and searchable metadata provide a disciplined vocabulary against which intelligent geospatial search can be performed within or among communities (Nebert, 2004).

Due to the existence of many disciplinary application profiles, a cross-domain catalog service must be able to address heterogeneity as well as distribution. This implies the implementation of catalog components for query distribution and virtual resources aggregation by applying mediation approaches. These catalogs may be referred to as distributed catalogs; they must implement distributed discovery functionalities in a fully heterogeneous environment. In fact, we can assume that both data model and functional protocol are heterogeneous requiring mediation approaches: metadata profiles harmonization and protocol adaptation.

Indeed, the development of distributed catalog solutions is a present challenge for cross-domain geospatial data infrastructures, such as those conceived in the framework of well-known cyber/e-infrastructure initiatives, including: INSPIRE (2007), US NSDI (2007), GMES (2004) and GEOSS (2005).

Distributed catalog component may adopt a federated architecture to connect data and service components distributing client queries. This approach brings some important benefits, including: great flexibility and scalability, resources representation consistency, usability (i.e. one-stop-shop server) and performances -where caching and/or parallelism is enabled. Conversely, some issues must be addressed; they encompass: cyclic queries, data identity, and heterogeneous resources mediation.

The interoperability solution we propose consists of utilizing a standard Distributed Catalog Service for: (a) federating the THREDDS/OPeNDAP and CDI resources (i.e. services and data); (b) mediating between them; (c) exposing them through an international standard interface for geo-spatial information discovery and access (i.e. a standard catalog inter-

---

[8]UNIDATA, THREDDS Fact-sheet, available at: http://www.unidata.ucar.edu/publications/factsheets/2007sheets/threddsFactSheet-1.doc.

[9]NASA, Global Change Master Directory, available on http://gcmd.nasa.gov, last accessed on 26/01/2010/.

[10]NCAR, Digital Library for Earth System Education available at http://www.dlese.org/library/index.jsp, last accessed on 26/01/2010.

[11]US National Science Foundation, National Science Digital Library, available at http://nsdl.org/, last accessed on 26 January 2010.
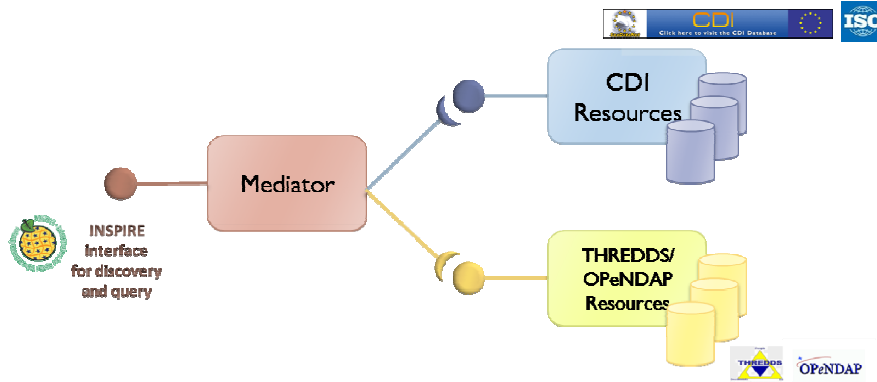
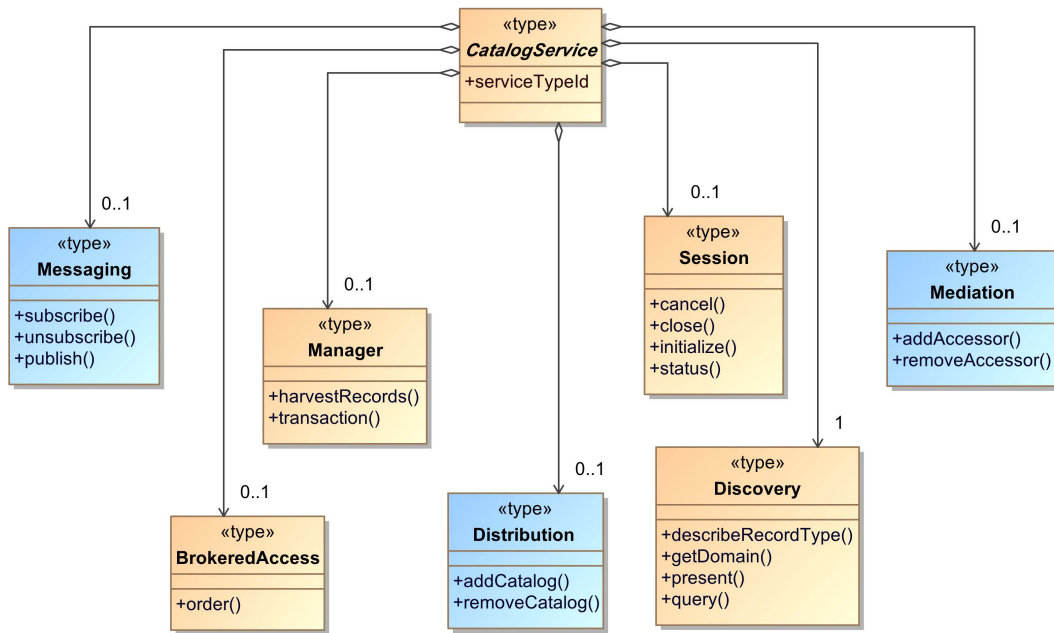**Fig. 1.** The adopted interoperability solution for CDI and THREDDS/OPENDAP.


Fig. 2. Extended interface functional model for distributed catalog service with mediation capabilities.

face). Figure 1 depicts a schema of the adopted interoperability approach. In addition to the mediation capabilities for CDI and THREDDS/OPeNDAP, this solution presents the following assets: (a) it implements and publishes the OGC CSW[12] standard catalog interface for resources discovery; (b) it is flexible allowing to federate and mediate other international standards and disciplinary interoperability specifications, such as: OGC access services (i.e. WMS, WCS), Biodiversity information systems protocols (i.e. the GBIF services), etc. These benefits will be further discussed in the final section of the paper.

2.1 Standard specifications

OGC Catalogue Service specification[12] defines a General Catalogue Interface Model for geospatial information. This specification provides a set of abstract service interfaces that support the discovery, access, maintenance and organization of catalogues of geospatial information and related resources. The specified interfaces are intended to allow users or application software to find information that exists in multiple distributed computing environments, including the World Wide Web (WWW) environment, as in the OGC CSW specification. OGC CSW functionalities include discovery as mandatory capability (i.e. Metadata Query & Presentation) and a couple of optional capabilities: session (i.e. initialize, close, status) and management (i.e. harvest, transaction).

---

[12]Open Geospatial Consortium, Inc. OpenGIS® Catalog Services Specification, Ver. 2.0.2.OGC 07-006R1.
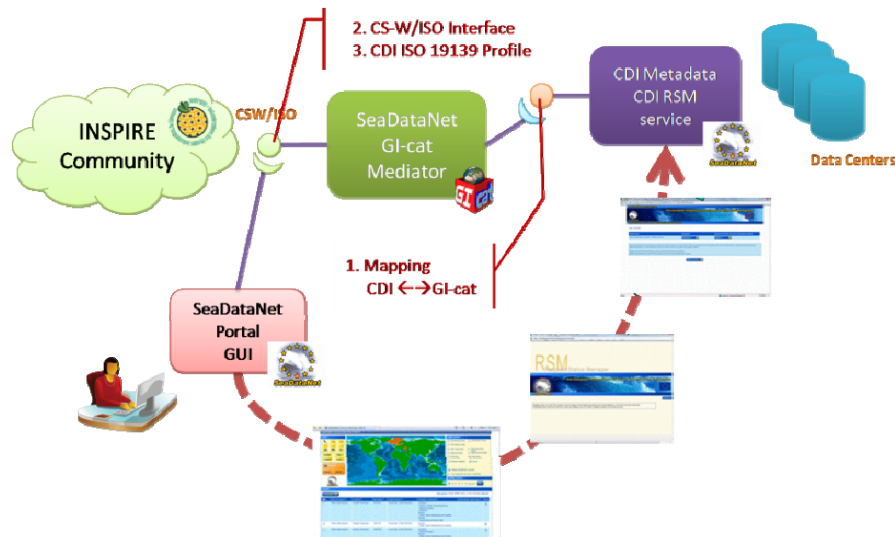
Fig. 3. GI-cat in SeaDataNet Framework.

In our opinion, the distribution aspects should be further specified, becoming another optional capability. Besides, we believe that distribution cannot be fully addressed, in an heterogeneous environment, without considering mediation capabilities. In fact, catalog services heterogeneity (e.g. the CSW application profiles multiplicity and the disciplinary special arrangements for resources discovery and inventory) often hinders effective distribution and discovery. Finally, there is the need to consider a resource access functionality which substantially differentiates from the "order and charge" request interface, which is provided by the present Brokered Access functionality of OGC Catalogue Service specification.

Therefore, we proposed an extended functional model for implementing distribution and mediation capabilities of catalog services. The abstract model is shown in Fig. 2.

The functional modules of the proposed model are:

– *Session* Interactive session management (e.g. close, initialize a session, return feedback).

– *Discovery* Querying/browsing of metadata, supporting asynchronous interactions.

– *Distribution* Configuration of a custom service chain to be used in subsequent queries.

– *Publication* Metadata publication.

– *Access* Resource data access and preview.

3   GI-cat mediation framework

Based on the above extended model, we developed a distributed catalog solution, called GI-cat. GI-cat supports caching and mediation capabilities and can act as a broker towards disparate resources, transforming multiple query results to a uniform and consistent interface. In fact, GI-cat implements a framework to federate standard and well-accepted catalog, inventory and access components. As far as interoperability standards are concerned, GI-cat supports the OGC CSW, WCS, WFS and WMS specifications. As to community arrangements for interoperability, GI-cat is able to federate: UNIDATA THREDDS Data Servers (TDS)/OPeNDAP, CDI resources, and GBIF (Global Biodiversity Information Facility) catalog and access components. GI-cat exposes several standard discovery interfaces, including: (a) the OGC CSW/ISO interface[6], recommended by the European Directive INSPIRE; (b) the OGC CSW/ebRIM-EO[13] interface, recommended by the GMES/ESA-HMA initiative[14]; (c) the OGC CSW Core[12], recommended by the GEO/GEOSS (2005): (d) OGC CSW/ebRIM-CIM[15]; (e) OpenSearch[16]; (f) GENESI DR[17].

---

[13]Open Geospatial Consortium Inc, Ogc™ catalogue services specification 2.0 extension package for ebrim application profile: Earth observation products, Ver. 0.2.4 OGC, 2009.

[14]European Space Agency, Heterogeneous mission accessibility – context, available at http://earth.esa.int/hma/context.html, accessed on 15 January 2010, 2007.

[15]Open Geospatial Consortium, Inc. OGC™Cataloguing of ISO Metadata (CIM) Using the ebRIM profile of CS-W, Ver. 0.1.8 OGC 07-038r1.

[16]OpenSearch, available at http://www.opensearch.org, accessed on 20/10/2010.

[17]GENESI-DR – Ground European Network for Earth Science Interoperations – Digital Repositories available at http://www.genesi-dr.eu/, accessed on 20/10/2010.
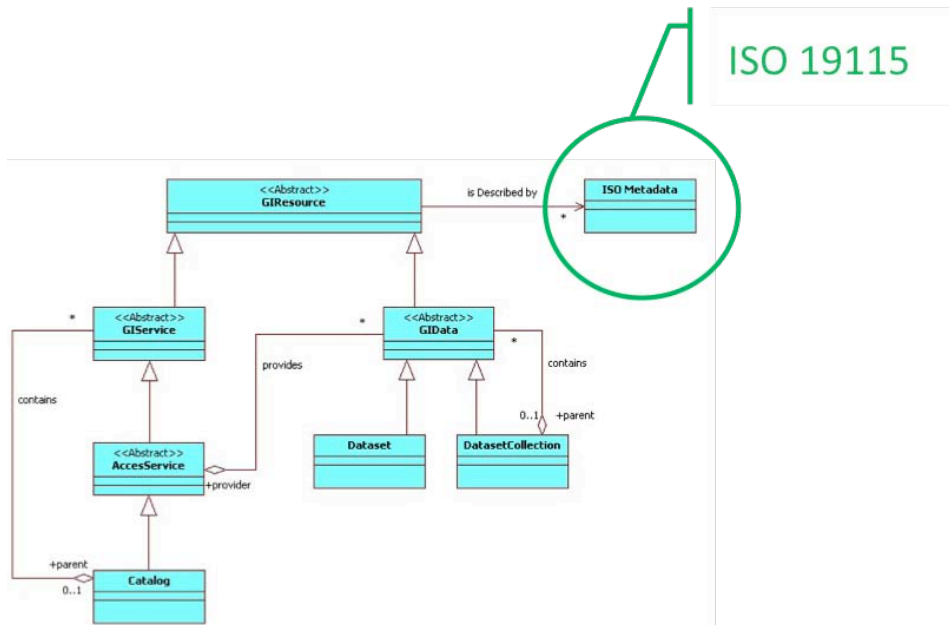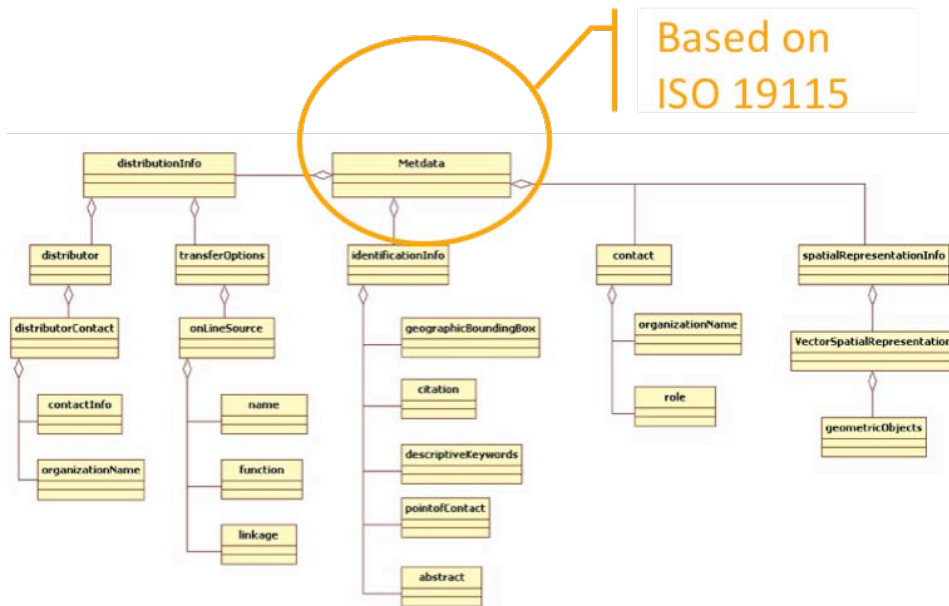
Fig. 4. Simplified GI-cat data model.

Fig. 5. Simplified CDI Data Model for a dataset element.

GI-cat supports distributed queries over a hierarchical data model, allowing incremental collection of queries results. A simplified schema of GI-cat data model is depicted in Fig. 4. A *catalog* element is modeled as an access service sub-type; in fact, access services provide access to resources (i.e. *GIResources*) which can be: datasets, dataset collections, and services (i.e. *Datasets, DatasetCollections*, and *GIServices*). *GIResource* elements are described by one or more *ISOMetadata* element.

Our catalog solution implements translucent and opaque service chaining, supporting asynchronous queries distribution and incremental presentation of results. Opaque service chaining occurs anonymously behind one service interface without any implication for the functionality of that interface; while, translucent service chaining is invoked to chain to another service behind it –in this case, the service interface must support the chain invocation.
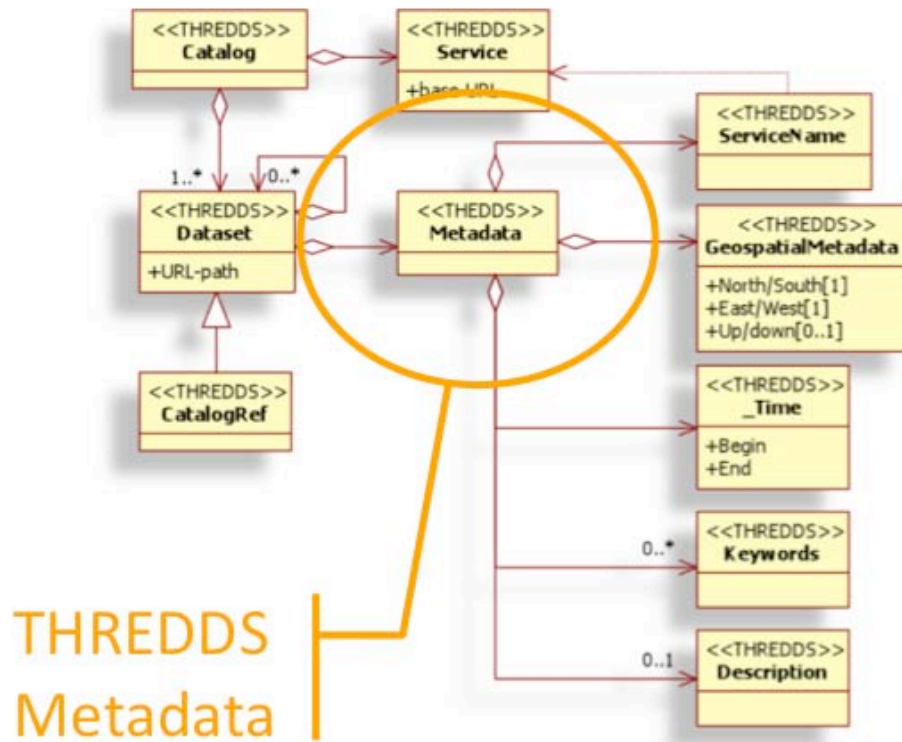
Fig. 6. THREDDS data model.

GI-cat is based on a service-oriented framework of modular components which implement the different capabilities characterizing the extended catalog model previously described. GI-cat can be customized and tailored to support different deployment scenarios. Figure 3 depicts how GI-cat is used, in the framework of the SeaDataNet project, in order to federate THREDDS/OPeNDAP and CDI resources, mediating and exposing them through the INSPIRE compliant interface OGC CSW/ISO[6] (as well as the other supported interfaces).

3.1 Accessors

In the GI-cat framework, the geospatial resources heterogeneity is faced by applying a mediation approach. Specific components implement mediation services for interfacing heterogeneous resource providers; such components are called *Accessors* and all realize the general functional interface described in the Catalog extended model. Accessors solve providers data models multiplicity by mapping them onto a more general data model: the ISO 19115 Core profile (Organization for Standardization, 2003); that is the GI-cat data model. The accessors also implement query protocols mediation by translating query requests expressed according to the interface protocols exposed by GI-cat (i.e. CSW) into the multiple query dialects spoken by the resource providers. Therefore, in the SeaDataNet context,

the CDI and THREDDS/OPeNDAP accessors play a key role for the interoperability implementation.

The described interoperability approach results quite general and flexible. In fact, it enables to implement the interoperability not only between CDI and THREDDS/OPeNDAP, but with any other supported protocol as well.

3.1.1 CDI Accessor

Figure 5 shows a UML class diagram representing a simplified version of the CDI data model. This is not a hierarchical data model. Actually, the data model only deals with low level granularity not defining high level granularity elements, such as "catalog" or dataset "aggregations". Hence, a CDI document (i.e. an XML file) describes only one dataset formalizing its content and access metadata. To carry out CDI data model interoperability with GI-cat, the correct solution would be to map a CDI resource element (i.e. a CDI dataset) onto a GI-cat "catalog" element. Clearly, this approach entails some important scalability and management issues.

We experimented a CDI data model extension introducing a dataset aggregation level. Thus, data provider can generate and publish an aggregation document (encoded as an XML file) which aggregates existing CDI documents -CDI XML files. For the aggregation, the CDI datasets are referenced by pointing the XML file they are encoded by. This is an example of CDI datasets aggregation document.
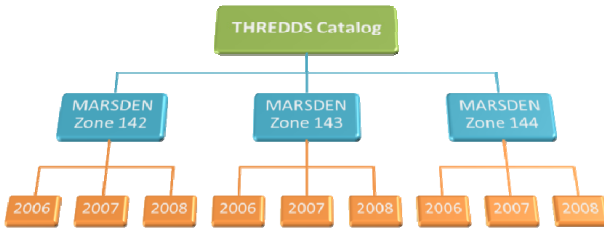
Fig. 7. THREDDS/OPENDAP Catalog tree strategy for improving query performance.

Table 1. CDI elements mapping onto GI-cat ones.

CDI element	GI-cat element
CDI document	Catalog containing one dataset
CDI document (with datasets aggregation)	Catalog containing one or more datasets

```
<cdiGroup>
        <cdiUrl title="rad1B0BB">
                http://moon.santateresa.enea.it/SDN_CDI/rad1B0BB.xml
        </cdiUrl>
        <cdiUrl title="rad4EF28">
                http://moon.santateresa.enea.it/SDN_CDI/rad4EF28.xml
        </cdiUrl>
        <cdiUrl title="rad8E3EC">
                http://moon.santateresa.enea.it/SDN_CDI/rad8E3EC.xml
        </cdiUrl>
</cdiGroup>
```

In this way there exists only one catalog for all the resources published by the aggregation document. GI-cat supports the CDI dataset aggregation extension. Table 1 describes the two mapping possibilities.

When a remote resource publishes a query interface, the Accessor component translates the query request received by GI-cat and passes it to the resource component. Of course, this is not the case of CDI resources. To perform queries over CDI resources, GI-cat harvests the CDI resource (i.e. retrieves and parses the CDI XML document), maps the content onto the GI-cat data model, and caches it for subsequent queries.

3.1.2 THREDDS/OPeNDAP Accessor

A simplified schema of the THREDDS/OPeNDAP data model is shown in Fig. 6. TDS implements a catalog service; hence, its data model is hierarchical defining a *catalog* element that may contain zero or more *dataset* – in turn, a *dataset* may contain zero or more sub-*dataset*. Besides, there

Table 2. THREDDS elements mapping onto GI-cat ones.

THREDDS element	GI-cat element
Catalog	Catalog
CatalogRef	DatasetCollection
Dataset (aggregated)	DatasetCollection
Dataset (simple)	Dataset

is a *catalogref* element that allows a catalog to refer to other remote catalogs.

It is noteworthy that THREDDS/OPeNDAP data model supports metadata inheritance. Therefore, when an aggregated *dataset* (or a *catalogref*) has some inheritable metadata, they propagate to the nested (or referred) datasets as well. In this way it is possible to factor out common metadata avoiding to duplicate the same information and encouraging to publish datasets characterized by high level granularity. This feature, along with the use of *catalogref* element, allows to query large catalogs in an efficient way.

THREDDS/OPeNDAP data model mapping on the GI-cat model is quite straightforward (see Table 2).

THREDDS/OPeNDAP publishes "inventory catalogs" (i.e. XML documents) over the Web. Catalogs can reference other catalogs and can also provide links for accessing datasets. THREDDS/OPeNDAP services include the Dataset Query Capabilities (DQC) which allows users to request a subset of a dataset collection. However, we preferred to harvest THREDDS/OPeNDAP inventory catalogs, following the approach already described for CDI documents. Differently from the CDI accessor, the THREDDS/OPeNDAP accessor must deal with nested structures which make use of the *CatalogRef* element. The accessor implements an incremental approach to avoid unnecessary network overhead and improve performances. Once a THREDDS/OPeNDAP catalog content is requested, referenced catalogs are not accessed until the content of those dataset collections is explicitly requested.

Since THREDDS data model is hierarchical, a THREDDS/OPeNDAP catalog can be represented as a tree where each internal node is either a dataset collection or a referenced catalog, and each leaf is a dataset. Each node, and leaf, has its own metadata (which may be inheritable) plus inherited metadata. A query is performed visiting the tree with a "depth first" strategy, reading the metadata of each node and checking if the query clause is satisfied. When a leaf is reached and its metadata satisfy the query constraints too, the leaf (i.e. dataset) is added to the query response.

The presence of inheritable metadata allows to discard a node along with its children if the node metadata do not sat-
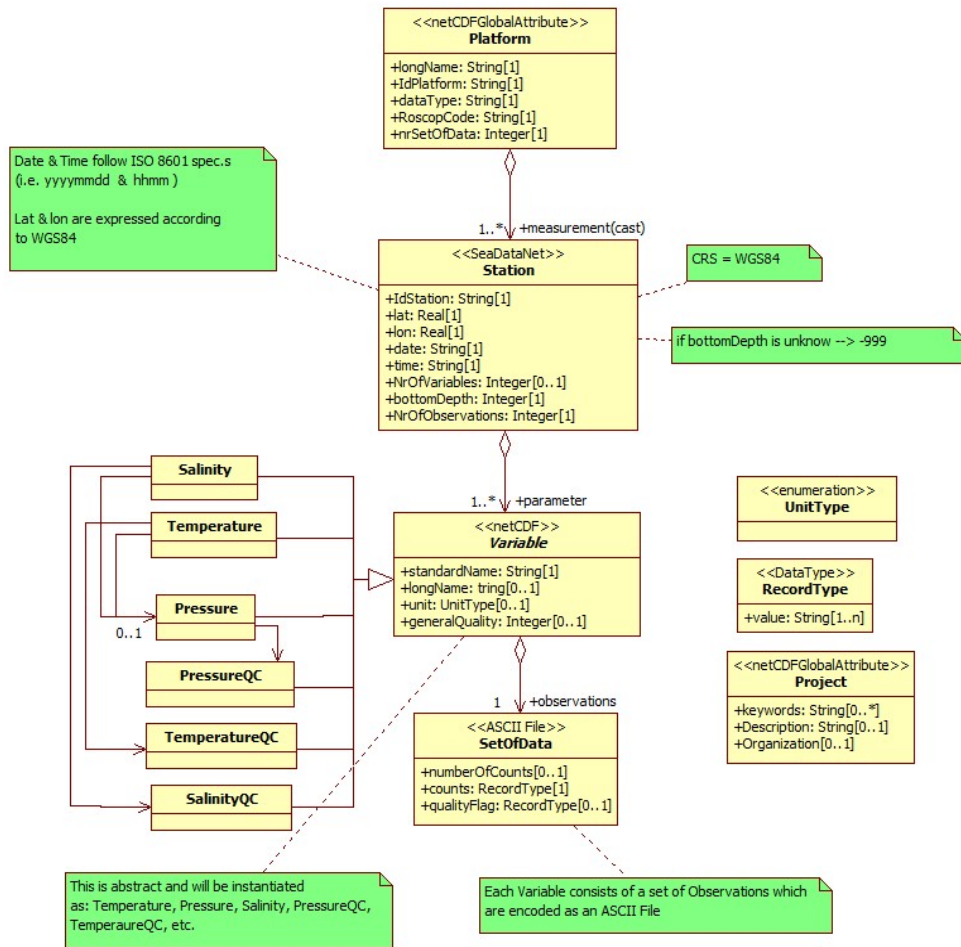
Fig. 8. CF-netCDF Data Model for XBT datasets.

isfy the query clause. This avoids network overhead when the node to be discarded is a *CatalogRef* element. This strategy results also efficient in order to deal with very large catalogs; in fact, it allows to discard many branches skipping to read their metadata.

3.1.3 Improving query performance with THREDDS/OPENDAP

Query performances depend on the matching pattern existing between the THREDDS hierarchical structure and the clause types characterizing most common queries – i.e. users common query strategy.

Considering most of queries have a space-time clause, datasets should be structured according to their spatial and temporal extension, respectively. To optimize query performances with respect to spatial domain, a possible strategy consists of dividing Earth spatial domain in finite areas. A valuable example in point are the MARSDEN zones (Sweers, 1970) or WMO zones. Following this solution, each node at the first level of the THREDDS/OPeNDAP catalog tree con-

tains datasets belonging to the same MARSDEN area. Thus, three nodes at the first hierarchical level are characterized by spatial information as inheritable metadata; while second-level nodes inherit spatial metadata and have temporal interval metadata as inheritable one. Figure 7 shows the tree structure strategy stemming from this approach. Datasets are first divided by space (i.e. MARSDEN areas), then by year, subsequently by month, and so on in accordance with the level of temporal granularity required by user queries.

3.2 Access protocol adaptation

Neither THREDDS nor CDI define any special access protocol; they both use standard protocols – e.g. WMS, WCS, OPeNDAP and HTTP download. When GI-cat receives a request to download a specific resource discovered through a CDI or THREDDS catalog document, it performs the following actions:

a) retrieves the available access service(s) as specified in the resource metadata; b) in case of more than one available service, it selects the most flexible service – e.g. allowing
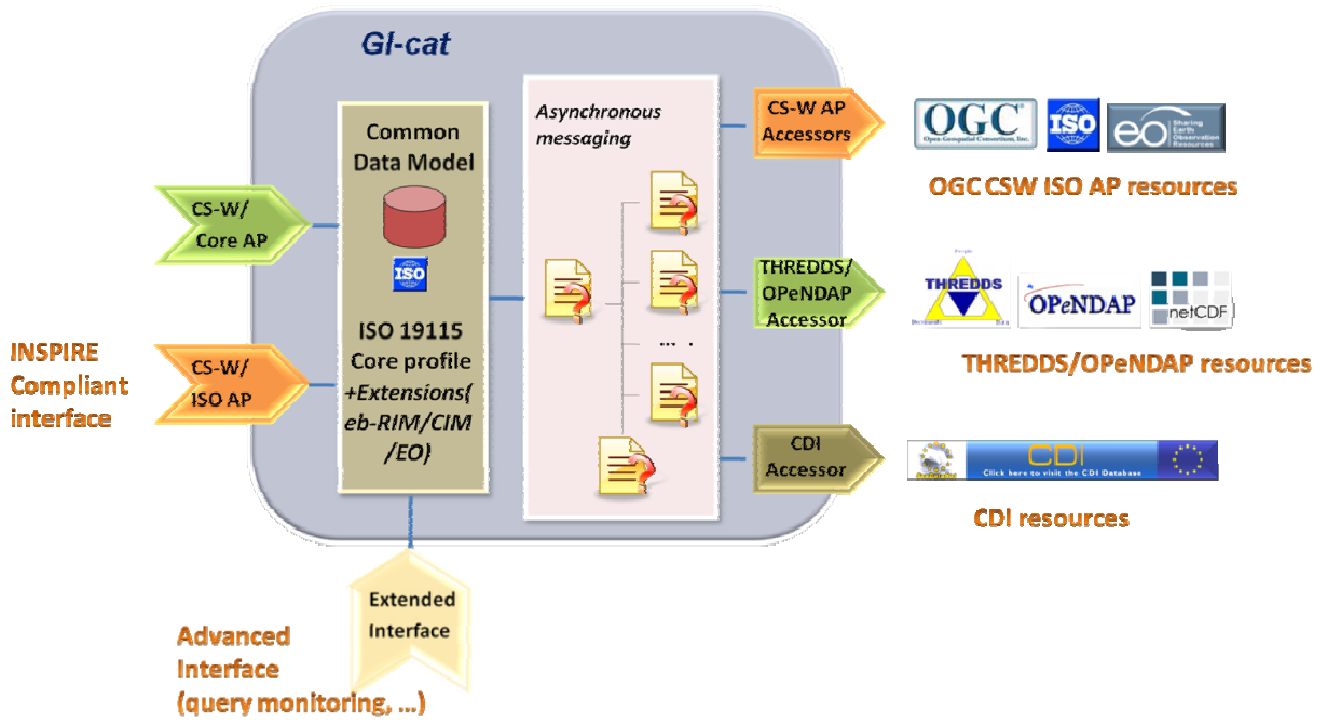
Fig. 9. The system developed to run interoperability tests.

domain and co-domain sub-settings c) instantiates a correspondent accessor component –i.e. a service client; c) delegates the accessor to formulate the appropriate access request, including sub-setting parameters, according to the received download clause, if this is supported by the resource download server.

4 Experimentation

Some tests were performed using XBT (Expendable Bathythermograph) data encoded in netCDF (network Common Data Form) format[18] following the CF1 conventions (Eaton et al., 2003). The utilized CF-netCDF data model is shown in Fig. 8. The CDI documents describing the XBT resources are available at: http://moon.santateresa.enea.it/ SDN_CDI/

For testing the proposed CDI catalog extension supporting files aggregation, we worked out an aggregated catalog document, available at: http://zeus.pin.unifi.it/projects/ cdi_registry/registry.xml

The same XBT resources encoded in CF-netCDF were also utilized to test the interoperability of THREDDS/OPENDAP inventory catalogs. The THREDDS/OPENDAP catalog server was configured

for improving query performances applying spatial and temporal domain clauses –using the MARSDEN convention for dividing Earth in indexed areas. This catalog is available at: http://apollo.pin.unifi.it:8080/thredds/XBTdemocatalog.xml

The set up system for interoperability testing is composed of GI-cat catalog (http://zeus.pin.unifi.it/gi-cat) deployed in front of a CDI and a THREDDS/OPENDAP catalog server. Besides, standard OGC Web access services were federated in order to extend the interoperability tests. To demonstrate test results, a graphical client application for GI-cat was utilized: the GI-go geo-browser. This is a CSW client being to interact with the corresponding interface published by GI-cat. GI-go can be downloaded at: http://zeus.pin.unifi.it/ gi-go.

Figure 9 depicts the developed system used for the interoperability experimentations.

Tests were performed querying CDI and THREDDS/OPENDAP catalogs with combinations of spatial, temporal and keywords clauses.

5 Conclusions

In the work presented in this paper we addressed some interoperability issues concerning SeaDataNet project. To achieve compliance of the SeaDataNet discovery service with INSPIRE IRs an ISO 19139 encoding for CDI was

---

[18]http://www.unidata.ucar.edu/software/netcdf/docs/

defined. Specific constraints were introduced in the ISO 19139 schema to accommodate an ISO profile of the CDI metadata model.

Moreover, to make CDI and THREDDS/OPeNDAP catalog components interoperable we propose a solution based on a distributed catalog service implementing mediation functionalities, namely GI-cat. The carried out interoperability tests were successful, demonstrating the effectiveness of discovering and accessing CDI and THREDDS/OPeNDAP resources through a common and standard catalog interface. In fact, GI-cat publishes a standard CSW interface. Besides, this solution allows, with few additional efforts, to implement interoperability also with other components, which implement international standards, such as OGC access services. In fact, the GI-cat framework architecture permits a great flexibility for implementing interoperability with well-described catalog, inventory and access services. Extensibility and flexibility are important features which brought us to prefer this approach rather than a customized one.

Flexibility was pursued by introducing the accessor component which should deal with all the interoperability issues regarding a specific service –either standard or implementing a community specification. Scalability is achieved by queries supporting asynchronous and incremental responses.

As expected, the experimentation pointed out that data models characterizing remote services influence GI-cat query and access performances. The valuable case of THREDDS/OPeNDAP catalog model was presented and a possible approach for SeaDataNet was discussed. The case of CDI data should be further investigated in order to improve GI-cat query responses time.

A dataset aggregation element was proposed in order to extend the present CDI data model and improve query and access performances. This extension is very simple and might be refined. A valuable example in point concerns the definition of a specific hierarchy for the dataset aggregation; this would improve performances, as demonstrated for THREDDS/OPENDAP.

The test was also extended to include access to CF-NetCDF files. In order to perform this test a data model was implemented, as shown in Fig. 8. The implementation of a common data model could improve data aggregation and interoperability. This work describes a CF-netCDF based data model utilized for describing and encoding XBT acquisitions. This model is under revision to extend and refine it in collaboration with the international Community – e.g. a collaboration liaison is under definition with UNIDATA on this subject.

Edited by: G. M. R. Manzella and S. Nativi
Reviewed by: two anonymous referees

References

Parker, B. B.: Oceanographic Data Archeology, finding historical data for climate and global change research", Oceanography, 5, 2, http://www.tos.org/oceanography/issues/issue_archive/issue_pdfs/5_2/5.2_parker.pdf, 1992.

NASA Earth Science Data Systems Standards Process Group, The Data Access Protocol – DAP 2.0, ESE-RFC-004.1.1., 2007.

Domenico, B., Caron, J., Davis, E., Kambic, R., and Nativi, S.: Thematic Real-time Environmental Distributed Data Services (THREDDS): Incorporating Interactive Analysis Tools into NSDL, Journal of Digital Information, 2, 4, May 2002.

Nebert, D.: The SDI Cookbook, ver. 2.0, 2004.

DIRECTIVE 2007/2/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE), available at http://inspire.jrc.ec.europa.eu/, 2007.

Federal Geographic Data Committee: The National Spatial Data Infrastructure, http://www.fgdc.gov/nsdi/nsdi.html, accessed 18 January 2010, 2007.

ESA: Final Report for the GMES Initial Period (2001–2003), http://www.gmes.info/107.0.html, February 2004.

Group on Earth Observations: in: Global Earth Observation System of Systems (GEOSS) 10-Year Implementation Plan, ESA Publications Division, The Netherlands, edited by: Battrick, B., ISSN No.: 0250-1589, ISBN No.: 92-9092-495-0, February 2005.

Organization for Standardization, ISO/19115:2003(E): Geographic information – Metadata. International Organization for Standardization, Geneva, Switzerland, 2003.

Sweers, H. E.: An improved code to classify the location of marine and terrestrial data, Limnol. Oceanogr., 15, 5, 827–832, http://www.aslo.org/lo/toc/vol_15/issue_5/0827.pdf, 1970.

Eaton, B., Grogory, J., Drach, B., Taylor, K., and Hankin, S.: NetCDF Climate and Forecast (CF) Metadat Conventions, Ver. 1.0, available at http://cf-pcmdi.llnl.gov/documents, 28 October 2003.

Sea-Search: Your gateway to oceanographic and marine data & information in Europe, available at http://www.sea-search.net/, accessed on 18 January 2010, 2010.