Definition of a CDI metadata profile and its ISO 19139 based encoding

Enrico Boldrini (1), Arjen De Korte (2), Mattia Santoro (1), Dick M. A. Schaap (2), Stefano Nativi (1,3), and Giuseppe Manzella (4)

(1) Italian National Research Council (CNR), Institute of Methodologies for Environmental Analysis (IMAA), Prato, Italy {boldrini, santoro, nativi}@imaa.cnr.it, (2) Marine Information Service (MARIS), Voorburg, Netherlands {arjen, dick}@maris.nl, (3) University of Florence, Prato, Italy stefano.nativi@pin.unifi.it, (4) Italian National Agency for New Technologies, Energy and Sustainable Economic Development (ENEA) - ENEA CLIM, La Spezia, Italy giuseppe.manzella@santateresa.enea.it

The Common Data Index (CDI) is the middleware service adopted by SeaDataNet for discovery and query. The primary goal of the EU funded project SeaDataNet is to develop a system which provides transparent access to marine data sets and data products from 36 countries in and around Europe. The European context of SeaDataNet requires that the developed system complies with European Directive INSPIRE.

In order to assure the required conformity a GI-cat based solution is proposed.

GI-cat is a broker service able to mediate from different metadata sources and publish them through a consistent and unified interface. In this case GI-cat is used as a front end to the SeaDataNet portal publishing the original data, based on CDI v.1 XML schema, through an ISO 19139 application profile catalog interface (OGC CSW AP ISO).

The choice of ISO 19139 is supported and driven by INSPIRE Implementing Rules, that have been used as a reference through the whole development process.

A mapping from the CDI data model to the ISO 19139 was hence to be implemented in GI-cat and a first draft quickly developed, as both CDI v.1 and ISO 19139 happen to be XML implementations based on the same abstract data model (standard ISO 19115 - metadata about geographic information).

This first draft mapping pointed out the CDI metadata model differences with respect to ISO 19115, as it was not possible to accommodate all the information contained in CDI v.1 into ISO 19139. Moreover some modifications were needed in order to reach INSPIRE compliance.

The consequent work consisted in the definition of the CDI metadata model as a profile of ISO 19115. This included checking of all the metadata elements present in CDI and their cardinality. A comparison was made with respect to ISO 19115 and possible extensions were individuated.

ISO 19139 was then chosen as a natural XML implementation of this new CDI metadata profile.

The mapping and the profile definition processes were iteratively refined leading up to a complete mapping from the CDI data model to ISO 19139.

Several issues were faced during the definition process. Among these: dynamic lists and vocabularies used by SeaDataNet could not be easily accommodated in ISO 19139, time resolution information from CDI v.1 was also difficult to accommodate, 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).

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.

Changes to the original ISO 19139 schema were at the maximum extent avoided because of practical reasons within SeaDataNet: additional constraint required by the profile have been defined and will be checked by the use of Schematron or other validation mechanisms.

The achieved mapping was finally ready to be integrated in GI-cat by implementation of a new accessor component for CDI. These type of components play the role of data model mediators within GI-cat framework.

The new defined profile and its implementation will also be used within SeaDataNet as a replacement of the current data model implementation (CDI v.1).