CDI/THREDDS Interoperability: the SeaDataNet developments

P. Mazzetti\textsuperscript{1,2}, S. Nativi\textsuperscript{1,2},
1. CNR-IMAA; 2. PIN-UNIFI
Outline

- Interoperability Issues in SeaDataNet

- A broker solution for CDI/THREDDS interoperability
  - Mediation for CDI/THREDDS interoperability with the GI-cat middleware

- INSPIRE compliance
  - Query adaption with GI-cat middleware
  - Definition of an ISO 19139 Profile for CDI
SeaDataNet project overview

- 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.

- SeaDataNet infrastructure addresses the development of infrastructure blocks that **allow to manage legacy problems** stemming from the multiplicity of information systems that must be federated.
The SeaDataNet context: CDI and THREDDS

CDI

- The Common Data Index (CDI) is part of the SeaDataNet specifications for discovery and query

THREDDS/OPeNDAP

- Many data management systems are based on the Unidata THREDDS (*Thematic Realtime Environmental Distributed Data Services*) and on the OPeNDAP (*Open-source Project for a Network Data Access Protocol*) middlewares for data discovery and access

There is a need to bridge CDI and THREDDS
The SeaDataNet context: INSPIRE Directive


- To ensure interoperability, the Directive requires that common Implementing Rules (IR) 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

There is a need to be INSPIRE compliant
Data discovery and access in SeaDataNet: Requirements

- SeaDataNet has expressed two interoperability requirements:
  - To discover/access resources published by CDI and THREDDS/OPeNDAP systems
  - To provide an INSPIRE compliant interface for discovery
A broker-based solution

- A specific component, the Mediator, acts as a broker...
  - ...mediating between heterogeneous resources
  - ...exposing a common interface (e.g. INSPIRE)
The broker’s role in a SOA

- Service-Oriented Architectures (SOAs) include three roles:
  - Service Providers
  - Service Consumers
  - Service Registries

- Clients need to implement all interfaces adopted by the Servers

- The Consumers address Service Providers heterogeneity
The broker’s role in a SOA

- A fourth role can be added in a SOA:
  - Service Broker
- A Mediator request services to the Servers on behalf of the Client.
  - Clients need to implement only the Mediator interface
- Only the Broker needs to address the Service Providers heterogeneity
In SeaDataNet, a data discovery and access broker performs the following tasks:

- **federates** the THREDDS/OPeNDAP and CDI resources (i.e. services and data)
  - providing a common (federation) data model
- **mediates** between THREDDS/OPeNDAP and CDI resources
  - mapping their data models to the common data model
- **adaption** of THREDDS/OPeNDAP and CDI discovery/access
  - mapping query interfaces for discovery/access
- exposes THREDDS/OPeNDAP and CDI resources through a **common interface** for discovery and access and CDI resources
  - providing resources through an INSPIRE compliant interface
GI-cat is an Open Source implementation of a discovery and access broker designed and developed by the ESSI Lab of CNR-IMAA and PIN-University of Florence.

- It allows clients to discover and evaluate geoinformation resources over a federation of data sources;
- It allows different clients to use the service, as it publishes different catalog interfaces.
GI-cat components: Profilers

Each profiler publishes a specific catalog interface. Its task is to mediate between this interface and the GI-cat internal interface (for queries) and vice-versa (for query results).
GI-cat components: Distributor component

It implements the query distribution functionalities:

a) query distribution amongst the federated resources

b) results aggregation
GI-cat components: Accessors

They implement mediation services for interfacing heterogeneous data providers which expose interfaces based on different specifications.

These mediating components map the heterogeneous providers metadata models into a uniform data model (implementing the ISO 19115 specification).
GI-cat v6.1

Exposed Interfaces (Profilers)
- OGC CSW 2.0.2 AP core
- OGC CSW 2.0.2 AP ISO 1.0
- OGC CSW 2.0.2 ebRIM EO
- OGC CSW 2.0.2 ebRIM CIM
- OpenSearch
- GENESI DR
- GI-cat extended interface

Data sources (Accessors)
- OGC WCS 1.0, 1.1
- OGC WMS 1.3.0, 1.1.1
- OGC WFS 1.0.0
- OGC CSW 2.0.2 Core, AP ISO 1.0, ebRIM/CIM, ebRIM/EO
- THREDDS 1.0.1, 1.0.2
- CDI 1.04, 1.3
- GI-cat 5.x, 6.x
- GBIF
- GeoNetwork (versions 2.2.0 and 2.4.1) catalog service
- Deegree (tested with version 2.2) catalog service
- OpenSearch
- GENESI DR OpenSearch

http://zeus.pin.unifi.it/cgi-bin/twiki/view/GIcat
GI-cat in CDI V1 service chain

1. Mapping CDI ↔ GI-cat
2. CS-W/ISO Interface
3. CDI ISO 19139 Profile
GI-cat: the SeaDataNet configuration
The mediation and adaption process: the GI-cat Accessors

- In the GI-cat framework, the geospatial resources heterogeneity is faced applying mediation technologies.
- GI-cat Accessors are in charge of the mediation process:
  - Accessors **map the providers data models to a common data model**: the ISO 19115/19135 based GI-cat model
  - Accessors map query protocols **translating query requests** expressed according to the interface protocols exposed by GI-cat (i.e. CS-W) into the multiple query dialects spoken by the resource providers.
The mediation process: CDI resources

- Mapping between:
  - GI-cat v5.0 ↔ CDI v1.04
  - GI-cat v5.0 ↔ CDI v1.3
GI-cat Model vs CDI Model

GI-cat v5.0
- Extends ISO 19115
- Hierarchical model
  - Dataset collection

CDI v1.04
- Based on ISO 19115
- Non-hierarchical model
  - No dataset collection concept
GI-cat / CDI Mapping

- Metadata mapping is straightforward (both based on ISO 19115)
- E.g. Spatial envelope mapping:
GI-cat / CDI Mapping

- The CDI provides an index (metadatabase) to individual data sets.

- A CDI data model extension to support dataset collection, has been experimented.
The mediation process: THREDDS resources

- Mapping between:
  - GI-cat v5.0 ↔ THREDDS v1.0.1
  - GI-cat v5.0 ↔ THREDDS v1.0.2
GI-cat Model vs THREDDS Model

GI-cat v5.0

- Extends ISO 19115
- Hierarchical model
  - Dataset collection

THREDDS 1.0.1/1.0.2

- Hierarchical model
- Catalog and dataset concepts
GI-cat / THREDDS Mapping

- Metadata mapping is straightforward
- THREDDS metadata inheritance must be addressed
The adaption process: CDI resources

- Adaption between:
  - CSW query \(\rightarrow\) query on CDI document
  - GI-cat harvests the CDI resource
  - Queries are performed on the internal data model
The adaption process: THREDDS resources

- Adaption between:
  - CSW query ➔ query on THREDDS catalog

- GI-cat harvests the THREDDS Dataset Inventory Catalog
- Queries are performed on the internal data model
ISO 19139 Profile for CDI

- SeaDataNet aims to assure that the output from the SeaDataNet portal is INSPIRE compliant

- This require exploring how to upgrade the SeaDataNet metadata XML files to ISO 19139

- Draft 0.3.2 developed in cooperation between SeaDataNet partners at IFREMER, CLS and MARIS, together with ESSI lab
ISO 19139 Profile for CDI

- Main requirements:
  - To accommodate all the data model information from CDI v.1.3.
  - In general this will be easily accomplished, as both CDI v.1.3 and ISO 19139 are based on ISO 19115

- Possible drawbacks:
  - Need to modify -by extension- the original ISO 19139 schemas in case of discrepancies
ISO 19139 Profile for CDI: Issues

1. How to accommodate information about SDN lists and vocabularies
   a) using gmx:Anchor where possible, or the ISO 19139 code lists

2. How to accommodate dataset time resolution
   a) Include the information under the temporal extent, under gml:TimePeriod/gml:timeInterval (with mapping from Vocab list L031 and allowed values)

3. How to map geographic extensions based on GML polygons
   a) using the element gmd:geographicElement/gmd:EX_BoundingPolygon

4. Determine the correct way to express language metadata

5. To put up an agreement to express dataset access information (this is vaguely defined in ISO 19115).
Conclusions

Some interoperability issues have been addressed in the SeaDataNet context:

- How to federate CDI and THREDDS/OPeNDAP systems
- How to provide an INSPIRE compliant interface

A solution has been successfully experimented:

- Use of the Gl-cat mediator, acting as a discovery/access broker to federate CDI/THREDDS resources, exposing an INSPIRE compliant interface
- Definition of an “ISO 19139 Profile for CDI” (draft).
Thank you for your attention
Extended Catalog Service for distribution & mediation

- Extension of the **CS-W model** for Distributed Catalog service (federated architecture)
  - Extended functionalities: **Messaging, Mediation, Distribution**
ISO 19139 Profile for CDI: timeline

- July 2009
  - GI-cat interoperability tests with mapping of CDI v.1.3 to ISO 19139
- August 2009
  - Draft proposal of ISO 19139 profile for CDI
- October-November 2009
  - Updates to the proposal
- January 2010
  - Profile definition?