Title: Harmonizing CERIF and INSPIRE metadata models to support multidisciplinary data sharing

Authors: Enrico Boldrini; Daniela Luzi; Stefano Nativi; Fabrizio Pecoraro

Addresses: Institute of Atmospheric Pollution Research, National Research Council (CNR-IIA), Via Madonna del Piano 10, 50019, Sesto Fiorentino (FI), Italy ' Institute for Research on Population and Social Policies, National Research Council (CNR-IRPPS), Via Palestro 32, 00185, Rome, Italy ' Institute of Atmospheric Pollution Research, National Research Council (CNR-IIA), Via Madonna del Piano 10, 50019, Sesto Fiorentino (FI), Italy ' Institute for Research on Population and Social Policies, National Research Council (CNR-IRPPS), Via Palestro 32, 00185, Rome, Italy

Abstract: This paper proposes a two-way crosswalk between CERIF (Common European Research Information Format) and ISO 19115 INSPIRE (Infrastructure for Spatial Information in Europe) profile metadata models. Several strategies have been investigated, depending on the quality of semantic mapping, as well as on the characteristics of the CERIF data model.

A core of corresponding metadata elements has been identified, as well as a proposal of extensions of the CERIF model to be compliant with the European directive INSPIRE.

Moreover, extensions of ISO concepts have also been described to provide contextual research information in the domain of environmental research data.

Finally, the crosswalk has been implemented in the GI-cat discovery broker framework. Successful use cases concerning both publication and discovery of CERIF result products using the extended framework demonstrated the possibility for CERIF information to be integrated in ISO compliant infrastructures and for INSPIRE information to be captured in CERIF.

1. Introduction

Research in Earth and Environmental Sciences relies on the analysis of heterogeneous data collected during both small and large-scale projects and acquired in both long and short-term observations, as well as in experiments or simulations. Discovery, acquisition and integration of enormous volumes of multidisciplinary and dispersed data are needed in order to gain knowledge of global environmental issues, such as climate change, flood and landslide risks.

Different communities and organizations are sustaining standardization efforts with the aim of reducing and managing the existing data heterogeneity by proposing and publishing different standard data models and service protocols for sharing datasets. However, their overall number as well as the number of the existing implementations remains high and therefore data integration still represents a fundamental issue.

Limiting for example the context to remote sensing, different standard data formats are currently available, such as Network Common Data Form (NetCDF), Hierarchical Data Format (HDF), Geographic Tagged Image File Format (GeoTiff). Encoded data is usually distributed through various standard access services to enable its online download, such as OGC Web Coverage Service (WCS), OGC Web Map Service (WMS), Open-source Project for a Network Data Access Protocol (OPeNDAP). Moreover, different metadata standards are available to describe data enabling its discovery, evaluation and use, such ISO 19115 (2003), Dublin Core, Geographically Encoded Objects for RSS feeds (GeoRSS), the Thematic Real-Time Environmental Distributed Data Services (THREDDS) Catalog model. Metadata records are published by standard discovery services, such as OGC Catalogue Service for the Web (CSW), THREDDS Catalogue service, OpenSearch engine. However, there are different versions and implementations of these standards that need to interoperate both at disciplinary and cross-disciplinary levels (e.g. services, clients, libraries). In particular, client applications (and in general scientists needing to access datasets) have to face the issue of implementing an additional software component in order to access a supplementary data source type. On the other hand, service applications (and in general data providers needing to publish their data) have the onus of implementing an additional software component to publish data according to further data source types.

The Brokering Approach (Nativi et al., 2013) was conceived to complement the standardization effort by moving the implementation burden from clients and services to a third "brokering" layer. In this way scientists and data providers can concentrate on their traditional goals using domain specific standards and applications, while the broker service concentrates on the technical aspects of data sharing, such harmonization of heterogeneous data models as shown in Figure 1.
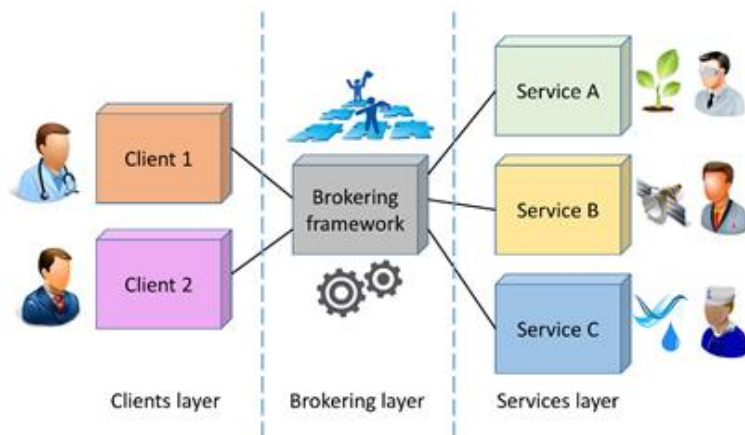
Figure 1. A brokering framework mediates between clients and services belonging to different domains. The burden required by the implementation of interoperability between components is thus shifted from client and service layers to the new brokering layer.

The Brokering architecture was first developed by the EU FP7 EuroGEOSS project and further tested at national level by the CNR to build the GIIDA (Interdepartmental Infrastructure for Environmental Data Sharing) e-infrastructure (Nativi et al., 2008). It has been further used to build up several e-infrastructures able to interoperate with heterogeneous information sources and standards, especially in global and/or multidisciplinary contexts. For instance, the Global Earth Observation System of Systems (GEOSS) is a flexible network of content providers linking together existing and planned observing systems around the world and supporting the development of new systems where gaps occur. A similar approach is also used in the US NSF EarthCube initiative. This approach makes it possible to build a system of heterogeneous systems without imposing any common standard or technology on the resource providers. Moreover, only the broker needs to be updated whenever a new standard is introduced.

The aim of this paper is to investigate the harmonization of context research information with environmental data, and its implementation in the context of a brokering framework.

CERIF (Common European Research Information Format) is a likely candidate for the discovery of research information sources (Jeffery et al., 2014) for two main reasons. On the one hand, its wide diffusion in the scenario of Research Information Systems (RISs) and its gradual integration into the data models of Institutional Repositories (OpenAIREplus) (Manghi et al., 2010) makes it an important part of the currently developing research infrastructures. On the other hand, the flexibility of the CERIF conceptual model based on rich relationships allows to test the brokering approach for the integration of research data in domain-specific environments.

This paper intends to propose different solutions to integrate CERIF in the environmental dataset domain, based on the quality of semantic mapping as well as on the characteristics of the CERIF data model. Therefore, the paper provides the results of a crosswalk of ISO 19115 INSPIRE profile to and from the CERIF model (paragraph 3 and 4) and also presents a CERIF profiler and accessor components implemented and tested in a prototype use case (paragraph 5). The paper concludes with a discussion of the achieved results.

2.   INSPIRE profile and CERIF data models: Harmonization approach

Recently different integration efforts have been directed to draft CERIF extensions to enable the exchange of research information connected with datasets in specific domains. The C4D (CERIF for Datasets) project (C4D-CERIF, 2012) proposed an extension of CERIF data model to import research data from the marine science domain. It identified a set of MEDIN elements that can be directly mapped to CERIF and in particular proposed that the entity cfResProd (CERIF result product) could represent datasets. The major result of this project was the new CERIF 1.6 version that incorporated some C4D suggestions. In projects, such as ENGAGE and EPOS CERIF is part of a multi-layer metadata architecture that links different models providing contextual information. Moreover, an interesting CERIF extension was proposed to interconnect RISs with Linked Open Data (LOD) (Joerg et al., 2012) and had an application in the Virtual Open Access Agriculture & Aquaculture (VOA3R) project. This approach generally bases the mapping on the CERIF semantic layer and for the scope of the integration takes advantage of identifiers of both CERIF and LOD data elements.

In particular, we propose a two-way data crosswalk from the CERIF to the ISO data model with a twofold aim: a) mapping from INSPIRE to CERIF to provide a CERIF guideline for the description of research datasets according to the INSPIRE profile ISO 19115; b) mapping from CERIF to ISO to provide a solution to describe CERIF research information related to the result products according to ISO 19115.

As expected when mapping two models describing different domains, the data crosswalk is likely to encounter problems to accommodate data elements of each schema. Moreover, a crosswalk has to take into account differences in the scope and structure of the two models. CERIF provides an abstract view of a platform independent database using an E-R diagram of research information, while ISO 19115 provides a conceptual representation of geographical information in a condensed and structured mode (based on UML) suitable to serve as an interoperable format in distributed systems. Therefore, the mapping between the two target models has to consider not only the semantics of attributes and concepts and relevant data types, but also the structural features, such as the CERIF semantic layer, the multiple relationships between entities and their temporal durations as well as the multi-language feature that allows to capture different translations of the same attribute. Although these features have the advantages of making the CERIF model flexible and applicable to different contexts, they make it necessary to identify specific conventions to establish a coherent integration based on both semantics and structure, in order to avoid the proliferation of dialects.

3.   Mapping from INSPIRE ISO 19115 profile to CERIF

The crosswalk analysis has been performed considering the INSPIRE profile of ISO 19115 defined in the Implementing Rules to describe the metadata for spatial data. It comprises the mandatory (core) set of ISO 19115 elements, as well as optional elements and specific constraints as required by the INSPIRE directive to describe geo-referenced datasets.

For the description of research information, we used CERIF version 1.6 (physical annotation for brevity) that accommodates metadata of research datasets based on the results of the C4D project and classifies a cfResProd instance as a dataset. This release is under a testing and review status,

so that the results of our crosswalk can be proposed to the euroCRIS community for further CERIF extensions. In the crosswalk from INSPIRE profile of ISO 19115 to CERIF, we distinguished three types of possible mappings defined as follows:

- *Straightforward*, when INSPIRE elements have semantically correspondent elements in the CERIF data model;

- *Inferential mapping*, when both INSPIRE and CERIF can refer to a data dictionary/vocabulary that contains semantically defined terms;

- *Convention*, when the CERIF metadata elements can be accommodated to express some mandatory INSPIRE elements by convention of the parties exposing their metadata.

Straightforward mapping allows an automatic discovery and interpretation of metadata elements exposed in RISs that use the CERIF model. The other two mappings require additional knowledge (about the used vocabulary or agreed convention), in order to meaningfully understand the dataset semantics content.

On the whole 22 out of 28 INSPIRE data elements could be mapped in the CERIF model. In particular 16 INSPIRE mandatory elements found a potential correspondence, of which 7 elements have a straightforward mapping, 3 elements can be integrated through an inferential mapping, and 6 elements could be accommodated by convention. Moreover, considering optional elements 6 out of 8 INSPIRE data elements could be mapped. There are cases where the mapping of data elements can be solved using different approaches: either establishing by convention the meaning of a specific data elements using CERIF semantic layer, or introducing in the CERIF model a new data element that univocally refers to an important aspect of research data management (see paragraph 3.3. and 3.4). This choice depends on the type and scope of the system that integrates both research data and context information.

3.1. *Straightforward mapping*

At the highest level the main correspondence is from the concept of dataset to the CERIF entity cfResProd (and its related elements) as reported in Table 1 focusing on ISO and CERIF paths as well as the cardinality of each element.

INSPIRE elements	INSPIRE Section	ISO 19115 Path	ISO Card.	CERIF Path	CERIF Card.
Dataset title	B1.1	MD_Metadata > MD_DataIdentification.citation > CI_Citation.title	[1..1]	cfResProd > cfResProdName	[1..*]
Geographic Bounding Box	B4.1	MD_Metadata > MD_DataIdentification.extent > EX_Extent > EX_GeographicExtent > EX_GeographicBoundingBox	[1..*]	cfResProd > cfResProd_GeoBBox > cfGeoBBox	[0..*]
Abstract describing the dataset	B1.2	MD_Metadata > MD_DataIdentification.abstract	[1..1]	cfResProd > cfResProdDescr	[1..*]
Dataset keyword	B3	MD_Metadata > MD_DataIdentification.descriptiveKeywords > MD_Keywords	[1..*]	cfResProd > cfResProdKeyw	[1..*]
Unique resource identifier	B1.5	MD_Metadata > MD_DataIdentification.citation > CI_Citation.identifier	[1..*]	cfResProd > cfResProdID	[1..1]

Resource type	B1.3	MD_Metadata.hierarchyLevel	[1..*]	[fixed by the scope to dataset]
Metadata character set	-	MD_Metadata.characterSet	[1..1]	[fixed to UTF-8]

Table 1. Straightforward mapping describing ISO and CERIF paths and related cardinalities

Moreover, an example on how to perform this mapping considering the CERIF multi-language feature is shown in Figure 2. Elements from the ISO class diagram are mapped to correspondent elements of the CERIF class diagram, if and only if they belong to the same language. Two different types of constraint blocks are used to select the appropriate language:

1. The "conditional" blocks enable the mapping between correspondent elements only when the input condition (represented by the linked "equal" block) is satisfied. In our example we focused on mapping of three core metadata elements: (1) *title* the name by which the cited resource is known; (2) *abstract* a brief narrative summary of the content of the resource(s) and (3) *keyword* commonly used word(s) or formalized word(s) or phrase(s) used to describe the subject.

2. The "equal" blocks check two linked input elements for equality. In our example they assure that the linked conditional mappings are performed only in case languages are the same both in ISO and in CERIF (i.e. the CERIF attribute cfLangCode of the relevant class equals the ISO attribute language of the MD_Metadata entity);
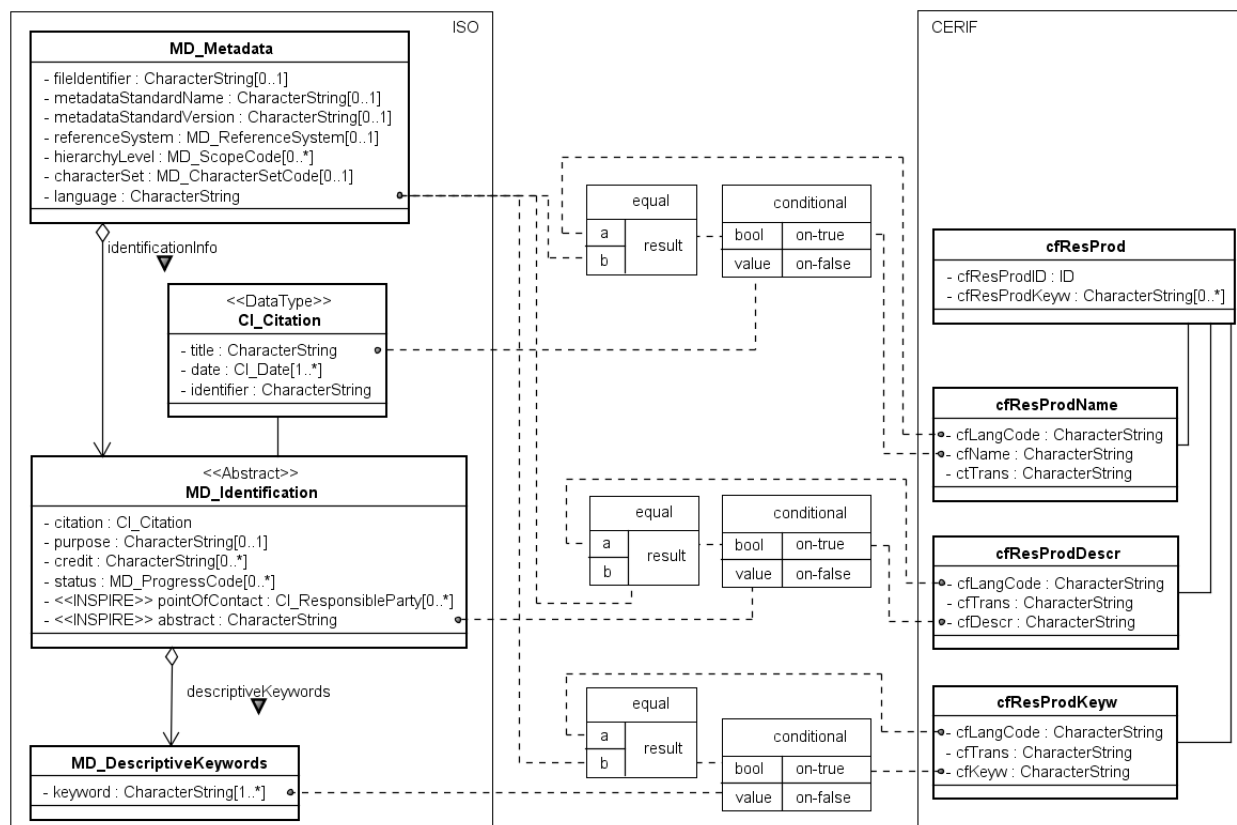
Figure 2. Straightforward mapping of ISO core metadata elements with correspondent CERIF data elements. The following constraint blocks were used: 1) "equal" blocks to check whether the value of the attribute cfLangCode matches with the specific language used by ISO for documenting metadata (attribute language of the entity MD_Metadata); 2) "conditional" blocks used to capture the correct translation stored by CERIF in case of multi-language records depending on the result of the "equal" blocks.

This results in a straightforward mapping that allows discovery brokers such as the GEO Discovery and Access Broker (GEO-DAB) to discover primary elements describing datasets exposed in RISs that use the CERIF model. Of course, in the prevision of an increasing inclusion of datasets also in RISs, it would be advisable to introduce in the CERIF model a specific entity for dataset results of a project/research activity. Such an extension was already applied when the entity cfResPub was introduced to univocally identify publications among the various research results (CERIF, 2008).

3.2. *Inferential mapping*

In the context of both geographic information systems and research infrastructures an important concept concerns the parties involved and their responsibilities. CERIF provides two entities to define this information: cfOrgUnit that represents an organizational unit, such as a university, a research institute or a private company and cfPers that represents an individual. Moreover, both entities are related with the cfPAddr and cfEAddr entities to specify, respectively, the place (i.e. address) and the electronic address (i.e. e-mail) where a person or an organization can be contacted. The relationships between the organization and the person with the research product (cfResProd)

are explicitly expressed through the so-called Link Entities (cfOrgUnit_ResProd and cfPers_ResProd, respectively). Moreover, cfClassId and cfClassSchemeId are used as primary foreign key to associate the Link Entities with the appropriate cfClass and the cfClassScheme of the Semantic Layer. This layer is used to specify the *role* played by a source object related to a target object (for instance a cfPers plays the role of Author in a cfResPubl).

Conversely, ISO 19115 captures the information related with the involved parties using the single Data Type entity CI_ResponsibleParty providing the identity of person(s) and/or organization(s) associated with the resource as well as the information required to contact them (represented by the CI_Contact Data Type) such as the position and the on-line sources from which the dataset can be obtained (represented by the CI_OnlineResource Data Type). Moreover, the function performed by the responsible party is managed using the *role* attribute collected within the Data Type. It expresses the role played by the relevant party in the management of datasets.

Given these premises, an inferential mapping can be performed to link the CERIF entities cfPers and cfOrgUnit with the corresponding ISO elements as highlighted in Figure 3. In addition, given that both standards allow to capture different responsible parties, their specific roles and responsibilities have to be taken into account by linking the cfClassId and cfClassSchemeId with the role represented by the ISO CI_RoleCode as well as the relevant dictionary (CodeList). This approach can be used to map two main INSPIRE elements, as reported in Table 2: (1) the "dataset responsible party" that identify person(s) and organization(s) associated with the resource(s) and the "metadata point of contact" that specifies the party responsible for the metadata information.
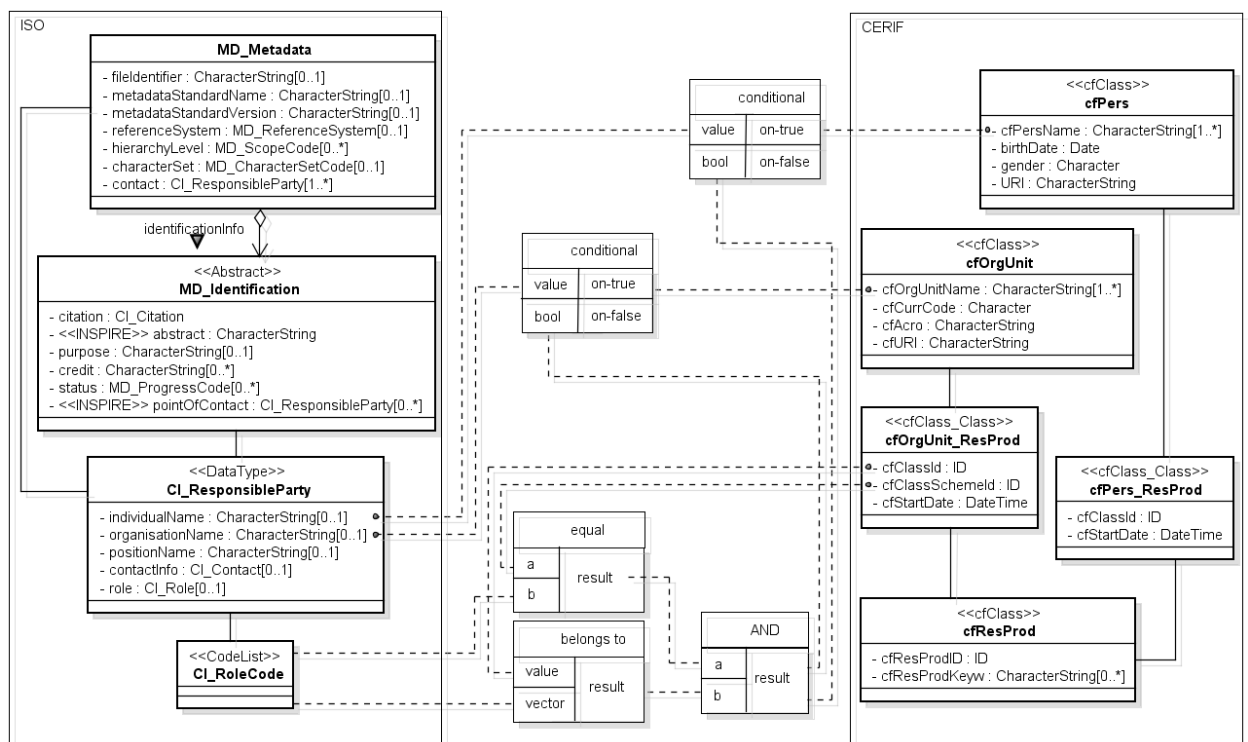

Figure 3. Inferential mapping of metadata elements related to parties involved in the management of datasets, as well as their role. To assure that mapping is applied only on entities having the same roles in both schemas, the

following constraint blocks were used: 1) "conditional" blocks map individualName and organizationName elements of the ISO CI_ResponsibleParty with the correspondent attributes in CERIF if and only the "AND" block is satisfied 2) the "AND" block is satisfied when both the linked blocks ("equal" and "belongs to") are satisfied. 3) "equal" blocks compare for equality the classification scheme of the organization unit (cfClassSchemeId attribute of the cfOrgUnit_ResProd entity) and the ISO CI_RoleCode list; 4) "belongs to" blocks verify whether the role specified by the cfClassId attribute belongs to one of the roles available in the CI_RoleCode dictionary;

A similar approach can be applied to map the ISO "dataset topic category" element using the cfResProd_Class with a specific taxonomy based on ISO MD_TopicCategoryCode terms (Table 2).

INSPIRE mandatory elements	INSPIRE Section	ISO 19115 Path	ISO Card.	CERIF Path	CERIF Card.	CERIF Role specification
Dataset responsible party	B9	MD_Metadata > MD_DataIdentification.pointOf Contact > CI_ResponsibleParty	[1..*]	cfResProd > cfOrgUnit_ResProd > cfOrgUnit > cfOrgUnitName AND cfOrgUnit_EAddr [AND cfResProd > cfPers_ResProd > cfPers > cfPersName]	[1..*]	cfClassId ∈ CI_RoleCode (e.g. "custodian") cfClassSchemeId=" CI_RoleCode"
Metadata point of contact	B10.1	MD_Metadata.contact > CI_ResponsibleParty	[1..*]	cfResProd > cfOrgUnit_ResProd > cfOrgUnit > cfOrgUnitName AND cfOrgUnit_EAddr [AND cfResProd > cfPers_ResProd > cfPers > cfPersName]	[1..*]	cfClassId ∈ CI_RoleCode (e.g. "pointOfContact") cfClassSchemeId=" CI_RoleCode"
Dataset topic category	B2.1	MD_Metadata > MD_DataIdentification.topicCa tegory	[1..*]	cfResProd_Class	[1..*]	cfClassId ∈ MD_TopicCategory Code (e.g. biota) cfClassSchemeId=" MD_TopicCategory Code"

Table 2. Inferential mapping describing ISO and CERIF paths and related cardinalities

3.3. Convention

Metadata standards devoted to the description of datasets usually contain some information related to the dataset quality (procedures, measurements used to collect data) and lineage (how, when and where data were collected). This information is crucial for both evaluation purposes and data re-use. To describe this information INSPIRE profile of ISO 19115 has two mandatory attributes: "conformity" and "lineage" (Table 3). Both concepts can be accommodated by a convention between the parties exposing metadata using the CERIF attribute cfValJudgeText of the cfMeas entity specifying in the cfMeasName the type of measurement used to collect data (fixed to "conformity" and/or "lineage"). cfMeas can also be used to indicate at least one of ISO temporal references chosen from the following categories: date of publication, date of last revision or date of creation ("temporal extent"). This could be accommodated by convention using cfMeasDateTime, where cfClass specifies the chosen category. Similarly, management information such as the date of creation of the dataset ("dataset reference date") and related metadata ("metadata date stamp") as well as the metadata language ("metadata language") can be accommodated given an agreed upon semantic to the relationship between dataset and organization. Considering the date element, CERIF semantic layer tracks temporal information in each relationship between two entities indicating the start and end dates (cfStartDate, cfEndDate)

and does not explicitly capture the date of creation of a research product or the date of its metadata creation. Therefore, by convention the cfStartDate element can be used to accommodate both date references. To distinguish between these two ISO elements an appropriate CERIF classification can specify the role of the organization in the creation of metadata as well as in the creation of the dataset (respectively: Publisher Institution and Author Institution).

Considering the ISO "metadata language", as there is no CERIF element for this information, it can be accommodated by convention in the cfLangCode element of one mandatory metadata field, such as cfResProdName.

Moreover, INSPIRE optional elements describing the format used by the distributor (i.e. "distribution format") as well as the on-line source from which the dataset can be obtained (i.e. "on-line resource") can be accommodated by convention using the attributes cfResProdVersInfo and cfURI of the cfResProd entity.

INSPIRE elements	INSPIRE Section	ISO 19115 Path	ISO Card.	CERIF Path	CERIF Card.	CERIF Attribute specification
INSPIRE mandatory metadata elements						
Conformity	B7	MD_Metadata > DQ_DataQuality.report	[1..*]	cfResProd > cfResProd_Meas > cfMeas > cfMeasName AND cfValJudgeText	[1..*]	cfMeasName= 'conformity'
Lineage	B6.1	MD_Metadata > DQ_DataQuality.lineage > LI_Lineage	[1..1]	union(cfResProd > cfResProd_Meas > cfMeas > cfMeasDescr)	[1..*]	cfMeasName= 'lineage'
Temporal extent	B5.1	MD_Metadata > MD_DataIdentification.extent > EX_Extent > EX_TemporalExtent	[1..*]	From minimum to maximum (cfResProd > cfResProd_Meas > cfMeas > cfDateTime)	[1..*]	
Dataset reference date	B5	MD_Metadata > MD_DataIdentification.citation > CI_Citation.date	[1..*]	cfResProd > cfOrgUnit_ResProd > cfOrgUnit > cfStartDate/cfEndDate		cfClassId='author institution'
Metadata date stamp	B10.2	MD_Metadata.dateStamp	[1..1]	cfResProd > cfOrgUnit_ResProd > cfOrgUnit > cfStartDate/cfEndDate	[1..*]	cfClassId='publis her institution'
Metadata language	B10.3	MD_Metadata.language	[1..1]	cfResProd > cfResProdName > cfLangCode	[1..*]	
INSPIRE optional metadata elements						
Distribution format	-	MD_Metadata > MD_Distribution > MD_Format.name AND MD_Format.version	[0..*]	cfResProd > cfResProdVersInfo	[1..*]	
On-line resource	B1.4	MD_Metadata > MD_Distribution > MD_DigitalTransferOption.on Line > CI_OnlineResource	[0..*]	cfURI	[0..1]	

Table 3. Convention mapping describing ISO and CERIF paths and related cardinalities

Other mandatory data elements of the INSPIRE profile of ISO 19115 that cannot be easily documented in CERIF concern the constrains related to the access and use of the dataset ("conditions for access and use" and "limitations on public access") as well as the language used within the dataset (the ISO "dataset language") and the standard used to encode it (the ISO "dataset character set"). This important information can be accommodated using the CERIF semantic layer

on the basis on an agreed-upon convention between the parties. For instance concerning the ISO "dataset languages", there are many admissible choices (i.e. ISO 639-2 standardized language codes). In this case a CERIF classification scheme with an appropriate, conventional identifier, could be built to document them. The same approach can be used considering ISO optional elements related to both metadata ("metadata file identifier", "metadata standard name", "metadata standard version" and "reference system") and dataset ("spatial representation type").

An alternative, more explicit approach would consist in extending CERIF model, as described in the following paragraph.

3.4.   A proposal of CERIF extension

Extending the CERIF model is another solution to accommodate both mandatory and optional ISO elements that don't have straightforward mappings to CERIF. This can be done by including a set of attributes in an extension of the CERIF entity cfResProd, such as cfCopyright and cfUseLimit (Fig. 4) to accommodate ISO "conditions for access and use" and "limitations on public access" information. Other attributes that could be introduced to the extended cfResProd are cfLangCode and cfCharSet to document, respectively, the ISO "dataset language" and the standard used to encode datasets (the ISO "dataset character set").

Extending the CERIF model would give more visibility to these fields that are important for INSPIRE compliancy, with respect to accommodating this information in the semantic layer using a conventional mapping.
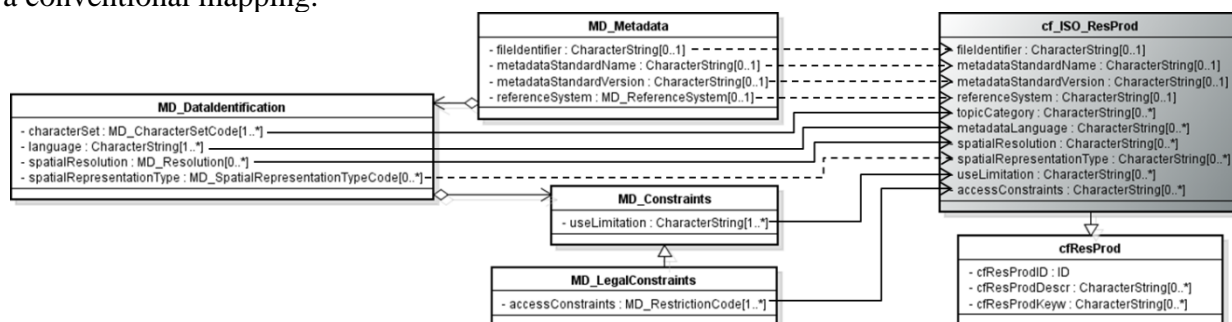

Figure 4. Proposal for CERIF extended elements (a solid line represents mandatory attributes whereas a dashed line represents optional elements)

The proposed mapping intentionally does not make use of the Dublin Core derived CERIF elements, because they are considered "deprecated" by the latest CERIF 1.6 specification. Otherwise some of the discussed mappings could have been easily resolved, with the use of elements such as cfDCLanguage, cfDCProvenance, cfDCRightsMMAccessRights.

4.   Mapping from CERIF to a ISO 19115 profile

The task of this section is to investigate the possibility of extending ISO 19115 data model with CERIF concepts, introducing research information, such as CERIF entities related to projects that provide a broader context of the research framework in which the datasets are produced. In addition this extension can improve the integration of environmental information systems with other domain-independent infrastructure such as RISs and Institutional Repositories. Therefore the

CERIF base entity Project can be introduced in ISO extending MD_Identification with an additional optional role named "initiativeInfo" that "provides information about the initiative under which the dataset was produced". This role belongs to a new object type (e.g. CERIF_InitiativeInformation) which contains the CERIF project attributes encoded when possible recurring to ISO classes (e.g. CI_Citation, CharacterString, MD_Keywords, DS_InitiativeTypeCode). Moreover, to benefit from the increasing availability of enhanced publications such as Dryad, ISO should also include metadata elements to document a publication related to a dataset. For this reason ISO 19115-1 revised version (not supported by INSPIRE at this time) has introduced an element called "additionalCitation" already adopted in the SeaDataNet profile (Boldrini and Nativi, 2013). Similarly, a CERIF profile can extend the MD_DataIdentification class with the subclass CERIF_DataIdentification, containing the optional attribute additionalDocumentation of type CERIF_Citation to collect bibliographic references to the dataset, such as articles and related publications. CERIF_Citation is also an extension of CI_Citation: it adds optional online references to the cited documentation. In addition, other concepts that are significantly represented in both schemas could be expanded in ISO to represent the information modeled in CERIF, such as person and organization. As already mentioned, ISO does not fully document a person except describing his/her belonging organization and the role played in it (e.g. originator, publisher and author). Thus additional CERIF attributes can be added, such as birthdate, gender, URI. The same stands for relevant organization attributes that are not modelled in ISO such as its research activity and the keywords used to describe it, creating two extensions to ISO concept such as CERIF_ResponsibleIndividual and CERIF_ResponsibleOrganization, each one including the correspondent attributes. The UML class diagram in Fig. 5 formalizes these metadata extensions according to ISO 19115 extension methodology.
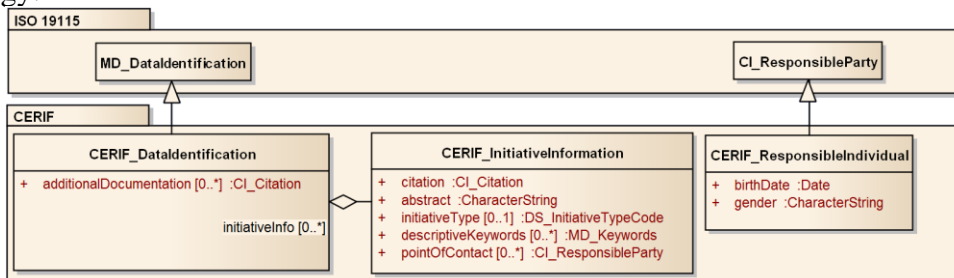

Figure 5. A subset of the possible extensions to the ISO metadata model with CERIF elements

5.   Implementation in a brokering framework

The two mappings were implemented and successfully tested by the GI-cat discovery and semantic broker (Nativi and Bigagli, 2009). GI-cat enables an harmonized discovery of datasets published by remote systems via a plethora of heterogeneous services interfaces (Bigagli et al., 2004) (e.g. CSW and OpenSearch for catalogue services, THREDDS and OAI-PMH for inventory services, and WMS, WCS, WAF, MySQL for access services). In turn, GI-cat exposes those discovery interfaces that are well adopted by existing clients and applications (e.g. CSW 2.02, OpenSearch, OAI-PMH 2.0). In GI-cat, each metadata model is mapped into a common model and vice versa

by specific components: called "accessors" and "profilers", respectively. GI-cat was extended with a CERIF "accessor" and "profiler" (see Fig. 6):
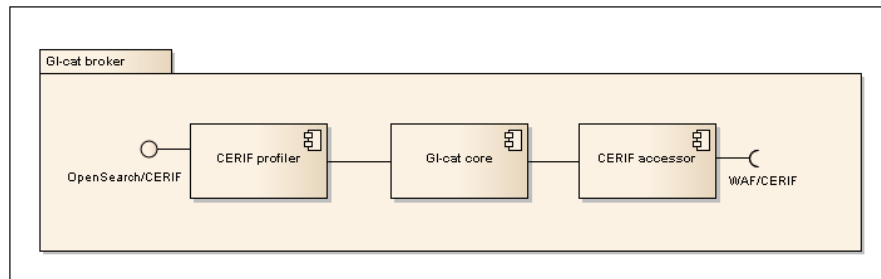

Figure 6. GI-cat component diagram, showing the new CERIF profiler and accessor, designed to support CERIF in the GI-cat brokering framework.

- *WAF/CERIF accessor* enables brokering of CERIF XML documents published on a given Web Accessible Folder (WAF) and describing result products. This component maps the CERIF datasets to the GI-cat internal data model (based on ISO 19115) according to the mapping proposed in this paper.
- *OpenSearch/CERIF profiler* enables discovery of resources brokered by the GI-cat framework through the OpenSearch protocol; results are encoded according to the CERIF XML Schema. This component translates from the GI-cat internal data model to CERIF according to the proposed mapping.

The extended GI-cat broker framework was finally tested for a couple of simple use cases.

5.1. Use case: publishing CERIF products through standard interfaces

In this scenario a research center is managing a Current Research Information System (CRIS), where information about researchers, projects and result products are maintained according to the CERIF data model. By using the extended GI-cat broker, CERIF result products can be suddenly made available for discovery through many standard service interfaces used by the different communities and already implemented by the broker (e.g. CSW, OpenSearch, OAI-PMH, etc.). The publication of CERIF products via all these heterogeneous interfaces is transparently managed by the broker. The deployment diagram in figure 7 shows GI-cat broker publishing existing CERIF result products through a standard CSW/ISO service interface making use of the new GI-cat CERIF accessor, as recommended by INSPIRE. Data are published by CERIF CRIS using a Web Accessible Folder (WAF) containing XML documents. In addition, an INSPIRE web portal user can be used to transparently discover CERIF result products. This setup makes possible for other communities (such as the INSPIRE community) to learn about the result products managed by the given research center, as well as the related research information.
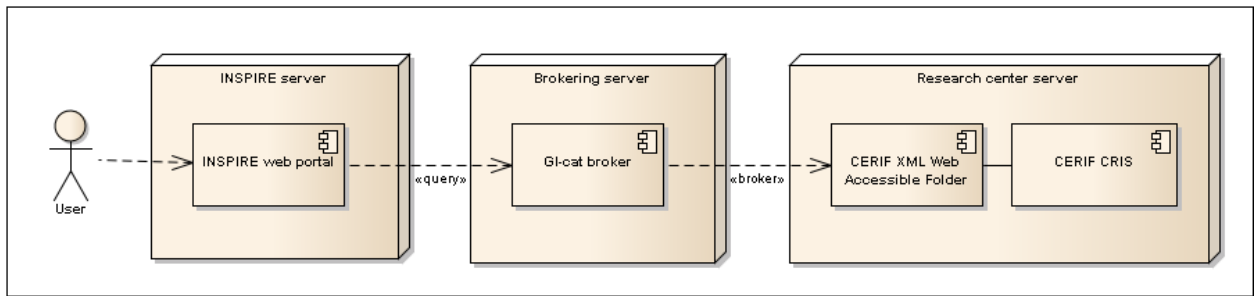
Figure 7. A CERIF CRIS is publishing a Web Accessible Folder (WAF) containing CERIF XML documents. GI-cat is used to broker it as well as to expose and publish its content for discovery through a CSW/ISO INSPIRE compliant interface, making use of the new GI-cat CERIF accessor. An INSPIRE web portal user can thus transparently discover CERIF result products.

5.2. *Use case: discovery of research products from heterogeneous data providers*

In this scenario a research team is interested in discovering result products published on heterogeneous servers maintained by different organizations all over the world. Using the extended GI-cat broker it is possible for him or her to run queries through a simple service interface, such as OpenSearch. The queries are transparently executed against the set of brokered resources. Several constraints are supported, such as keywords, spatial temporal area data provider. Matching records are then returned encoded according to the CERIF XML schema, ready to be imported into a local CRIS based on CERIF.

The deployment diagram in figure 8 shows a setup that can be used to easily discover a wide range of environmental datasets and having them encoded according to the CERIF model, based on the GI-cat server that broker environmental datasets from a set of heterogeneous data providers. The new GI-cat CERIF profiler enables their discovery through a simple OpenSearch interface returning documents encoded in conformity with the CERIF XML schema. A user is thus able to find the desired result products and import them in his/her local CERIF based CRIS.
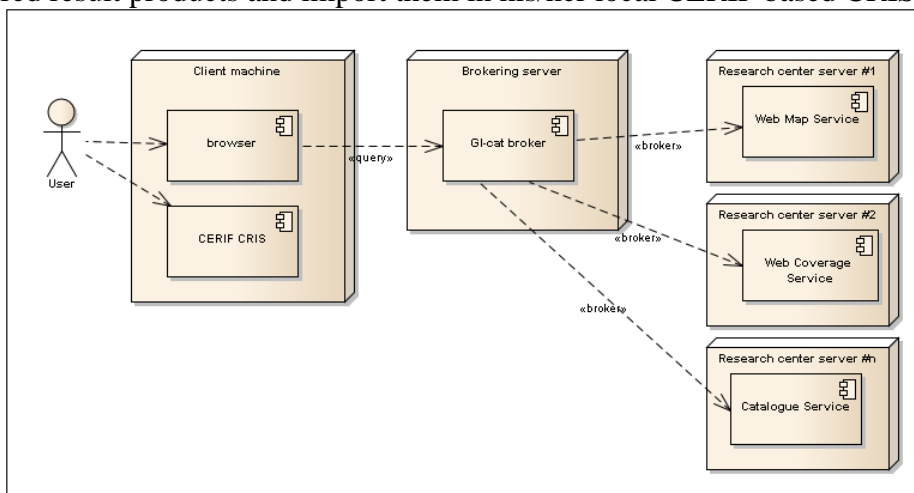

Figure 8. GI-cat being used to broker environmental datasets from a set of heterogeneous data providers. The new GI-cat CERIF profiler enables their discovery through a simple OpenSearch interface returning documents encoded in conformity with the CERIF XML schema. A user is thus able to find the desired result products and import them in a his/her local CERIF based CRIS.

Table 4 shows an excerpt XML document of a query result, encoded according to CERIF 1.6 XML schema.

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<cerif:CERIF xmlns:cerif="urn:xmlns:org:eurocris:cerif-1.6-2" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" date="2014-04-
07+02:00" sourceDatabase="GEO-DAB" xsi:schemaLocation="urn:xmlns:org:eurocris:cerif-1.6-2
http://www.eurocris.org/Uploads/Web%20pages/CERIF-1.6/CERIF_1.6_2.xsd">
  <cerif:cfResProd>
    <cerif:cfResProdId>2009d331-cefb-4fb1-8828-0e274d646a9f</cerif:cfResProdId>
    <cerif:cfURI>http://webgis1.barrowbc.gov.uk/webgis/bingis.html</cerif:cfURI>
  </cerif:cfResProd>
  <cerif:cfResProdName>
    <cerif:cfResProdId>2009d331-cefb-4fb1-8828-0e274d646a9f</cerif:cfResProdId>
    <cerif:cfName cfLangCode="eng" cfTrans="o">2009 Aerial Photography</cerif:cfName>
  </cerif:cfResProdName>
  <cerif:cfGeoBBox>
    <cerif:cfGeoBBoxId>ID-0</cerif:cfGeoBBoxId>
    <cerif:cfWBLong>-3.32482</cerif:cfWBLong>
    <cerif:cfEBLong>-3.12439</cerif:cfEBLong>
    <cerif:cfSBLat>54.03964</cerif:cfSBLat>
    <cerif:cfNBLat>54.21841</cerif:cfNBLat>
  </cerif:cfGeoBBox>
  <cerif:cfResProd_GeoBBox>
    <cerif:cfResProdId>2009d331-cefb-4fb1-8828-0e274d646a9f</cerif:cfResProdId>
    <cerif:cfGeoBBoxId>ID-0</cerif:cfGeoBBoxId>
    <cerif:cfClassId>ID-1</cerif:cfClassId>
    <cerif:cfClassSchemeId>ID-2</cerif:cfClassSchemeId>
  </cerif:cfResProd_GeoBBox>
  <cerif:cfResProdDescr>
    <cerif:cfResProdId>2009d331-cefb-4fb1-8828-0e274d646a9f</cerif:cfResProdId>
    <cerif:cfDescr cfLangCode="eng" cfTrans="o">Orthorectified aerial photography of Barrow Council administrative area flown on 31st
May 2009 at a resolution of 10cm.</cerif:cfDescr>
  </cerif:cfResProdDescr>
  <cerif:cfResProdKeyw>
    <cerif:cfResProdId>2009d331-cefb-4fb1-8828-0e274d646a9f</cerif:cfResProdId>
    <cerif:cfKeyw cfLangCode="eng" cfTrans="o">Photography</cerif:cfKeyw>
  </cerif:cfResProdKeyw>
  <cerif:cfResProdKeyw>
    <cerif:cfResProdId>2009d331-cefb-4fb1-8828-0e274d646a9f</cerif:cfResProdId>
    <cerif:cfKeyw cfLangCode="eng" cfTrans="o">Orthoimagery</cerif:cfKeyw>
  </cerif:cfResProdKeyw>
  <cerif:cfOrgUnit_ResProd>
    <cerif:cfOrgUnitId>ID-3</cerif:cfOrgUnitId>
    <cerif:cfResProdId>2009d331-cefb-4fb1-8828-0e274d646a9f</cerif:cfResProdId>
    <cerif:cfClassId>ID-8</cerif:cfClassId>
    <cerif:cfClassSchemeId>ID-9</cerif:cfClassSchemeId>
    <cerif:cfStartDate>2009-05-31T02:00:00.000+02:00</cerif:cfStartDate>
    <cerif:cfEndDate>2009-05-31T02:00:00.000+02:00</cerif:cfEndDate>
  </cerif:cfOrgUnit_ResProd>
  <cerif:cfOrgUnit>
    <cerif:cfOrgUnitId>ID-3</cerif:cfOrgUnitId>
  </cerif:cfOrgUnit>
  <cerif:cfOrgUnitName>
    <cerif:cfOrgUnitId>ID-3</cerif:cfOrgUnitId>
    <cerif:cfName cfLangCode="eng" cfTrans="o">Barrow Borough Council</cerif:cfName>
  </cerif:cfOrgUnitName>
  <cerif:cfPers_ResProd>
    <cerif:cfPersId>ID-4</cerif:cfPersId>
    <cerif:cfResProdId>2009d331-cefb-4fb1-8828-0e274d646a9f</cerif:cfResProdId>
    <cerif:cfClassId>distributor</cerif:cfClassId>
    <cerif:cfClassSchemeId>CI_RoleCode</cerif:cfClassSchemeId>
```

```
   </cerif:cfPers_ResProd>
   <cerif:cfPers>
     <cerif:cfPersId>ID-4</cerif:cfPersId>
   </cerif:cfPers>
   <cerif:cfOrgUnit_EAddr>
     <cerif:cfOrgUnitId>ID-3</cerif:cfOrgUnitId>
     <cerif:cfEAddrId>ID-5</cerif:cfEAddrId>
     <cerif:cfClassId>ID-6</cerif:cfClassId>
     <cerif:cfClassSchemeId>ID-7</cerif:cfClassSchemeId>
   </cerif:cfOrgUnit_EAddr>
   <cerif:cfEAddr>
     <cerif:cfEAddrId>ID-5</cerif:cfEAddrId>
     <cerif:cfPAddrId>gis@barrowbc.gov.uk</cerif:cfPAddrId>
   </cerif:cfEAddr>
   <cerif:cfOrgUnit_ResProd>
     <cerif:cfOrgUnitId>ID-10</cerif:cfOrgUnitId>
     <cerif:cfResProdId>2009d331-cefb-4fb1-8828-0e274d646a9f</cerif:cfResProdId>
     <cerif:cfClassId>author</cerif:cfClassId>
     <cerif:cfClassSchemeId>CI_RoleCode</cerif:cfClassSchemeId>
     <cerif:cfStartDate>2014-04-07T02:00:00.000+02:00</cerif:cfStartDate>
     <cerif:cfEndDate>2014-04-07T02:00:00.000+02:00</cerif:cfEndDate>
   </cerif:cfOrgUnit_ResProd>
   <cerif:cfOrgUnit>
     <cerif:cfOrgUnitId>ID-10</cerif:cfOrgUnitId>
   </cerif:cfOrgUnit>
   <cerif:cfOrgUnitName>
     <cerif:cfOrgUnitId>ID-10</cerif:cfOrgUnitId>
     <cerif:cfName cfLangCode="eng" cfTrans="o">Barrow Borough Council</cerif:cfName>
   </cerif:cfOrgUnitName>
   <cerif:cfPers_ResProd>
     <cerif:cfPersId>ID-11</cerif:cfPersId>
     <cerif:cfResProdId>2009d331-cefb-4fb1-8828-0e274d646a9f</cerif:cfResProdId>
     <cerif:cfClassId>author</cerif:cfClassId>
     <cerif:cfClassSchemeId>CI_RoleCode</cerif:cfClassSchemeId>
   </cerif:cfPers_ResProd>
   <cerif:cfPers>
     <cerif:cfPersId>ID-11</cerif:cfPersId>
   </cerif:cfPers>
   <cerif:cfOrgUnit_EAddr>
     <cerif:cfOrgUnitId>ID-10</cerif:cfOrgUnitId>
     <cerif:cfEAddrId>ID-12</cerif:cfEAddrId>
     <cerif:cfClassId>ID-13</cerif:cfClassId>
     <cerif:cfClassSchemeId>ID-14</cerif:cfClassSchemeId>
   </cerif:cfOrgUnit_EAddr>
   <cerif:cfEAddr>
     <cerif:cfEAddrId>ID-12</cerif:cfEAddrId>
     <cerif:cfPAddrId>gis@barrowbc.gov.uk</cerif:cfPAddrId>
   </cerif:cfEAddr>
   <cerif:cfResProd_Class>
     <cerif:cfResProdId>2009d331-cefb-4fb1-8828-0e274d646a9f</cerif:cfResProdId>
     <cerif:cfClassId>imageryBaseMapsEarthCover</cerif:cfClassId>
     <cerif:cfClassSchemeId>MD_TopicCategoryCode</cerif:cfClassSchemeId>
   </cerif:cfResProd_Class>
   <cerif:cfResProd_Meas>
     <cerif:cfResProdId>2009d331-cefb-4fb1-8828-0e274d646a9f</cerif:cfResProdId>
     <cerif:cfMeasId>ID-15</cerif:cfMeasId>
     <cerif:cfClassId>ID-16</cerif:cfClassId>
     <cerif:cfClassSchemeId>ID-17</cerif:cfClassSchemeId>
   </cerif:cfResProd_Meas>
   <cerif:cfMeas>
     <cerif:cfMeasId>ID-15</cerif:cfMeasId>
     <cerif:cfValJudgeText>conformity to INSPIRE specification</cerif:cfValJudgeText>
   </cerif:cfMeas>
   <cerif:cfMeasName>
```

```xml
      <cerif:cfMeasId>ID-15</cerif:cfMeasId>
      <cerif:cfName cfLangCode="eng" cfTrans="o">Conformity</cerif:cfName>
   </cerif:cfMeasName>
   <cerif:cfResProd_Meas>
      <cerif:cfResProdId>2009d331-cefb-4fb1-8828-0e274d646a9f</cerif:cfResProdId>
      <cerif:cfMeasId>ID-18</cerif:cfMeasId>
      <cerif:cfClassId>ID-19</cerif:cfClassId>
      <cerif:cfClassSchemeId>ID-20</cerif:cfClassSchemeId>
   </cerif:cfResProd_Meas>
   <cerif:cfMeas>
      <cerif:cfMeasId>ID-18</cerif:cfMeasId>
      <cerif:cfValJudgeText>process steps for this dataset</cerif:cfValJudgeText>
   </cerif:cfMeas>
   <cerif:cfMeasName>
      <cerif:cfMeasId>ID-18</cerif:cfMeasId>
      <cerif:cfName cfLangCode="eng" cfTrans="o">Lineage</cerif:cfName>
   </cerif:cfMeasName>
   <cerif:cfResProd_Meas>
      <cerif:cfResProdId>2009d331-cefb-4fb1-8828-0e274d646a9f</cerif:cfResProdId>
      <cerif:cfMeasId>ID-21</cerif:cfMeasId>
      <cerif:cfClassId>ID-22</cerif:cfClassId>
      <cerif:cfClassSchemeId>ID-23</cerif:cfClassSchemeId>
   </cerif:cfResProd_Meas>
   <cerif:cfMeas>
      <cerif:cfMeasId>ID-15</cerif:cfMeasId>
      <cerif:cfValJudgeText>conformity to INSPIRE specification</cerif:cfValJudgeText>
   </cerif:cfMeas>
   <cerif:cfResProd_Meas>
      <cerif:cfResProdId>2009d331-cefb-4fb1-8828-0e274d646a9f</cerif:cfResProdId>
      <cerif:cfMeasId>ID-26</cerif:cfMeasId>
      <cerif:cfClassId>ID-24</cerif:cfClassId>
      <cerif:cfClassSchemeId>ID-25</cerif:cfClassSchemeId>
   </cerif:cfResProd_Meas>
   <cerif:cfMeas>
      <cerif:cfMeasId>ID-18</cerif:cfMeasId>
      <cerif:cfValJudgeText>process steps for this dataset</cerif:cfValJudgeText>
   </cerif:cfMeas>
 </cerif:CERIF>
```

Table 4. XML excerpt document describing of the CERIF encoding of a sample INSPIRE metadata document, as encoded by the OpenSearch/CERIF profiler

6.   Conclusions

The paper proposes a two-way mapping to integrate CERIF research information model with ISO 19115 environmental datasets model.

A first result of this crosswalk is the identification of a minimum set of elements based on the main correspondence between the CERIF "result product" and the ISO 19115 "dataset" concept.

In this perspective, a possible proposal to be examined by the euroCRIS community (in charge of maintaining the CERIF model) concerns the introduction of a specific entity to identify research products as datasets.

We have therefore shown the possibility to accommodate additional information required by INSPIRE in CERIF, by extending the CERIF result product entity including other relevant attributes to express significant information such as those related to the accessibility and the re-use of datasets.

This approach can also be a starting point for the CERIF community to establish a set of rules/procedures with the aim of standardizing a new CERIF metadata extension in different contexts. The success of this integration takes advantage of the CERIF semantic layer that facilitates a flexible application of the model in heterogeneous environments. Of course, the use of the semantic layer as well as of the multiple relationships between CERIF entities makes it necessary to define and document specific constraints and rules to establish a consistent semantic integration. For instance the different translations provided by the CERIF multi-language feature have to be taken into account (as highlighted in the straightforward mapping section). This requires not only a semantic correspondence of an INSPIRE attribute with a CERIF element but also to check the consistency of the element translation with the specific language used by ISO for documenting metadata. The same mechanism has demonstrated to be valid also to align the different dictionaries provided by INSPIRE and CERIF to specify, for instance, roles and responsibilities of the parties involved in the management of datasets, as described in paragraph 3.2.

The other part of the crosswalk concerns the integration of CERIF research information in the ISO geographic datasets model. We have investigated the possible development of a CERIF profile of ISO 19115 data model including extended elements to describe contextual research information related to a given dataset. In this case we have leveraged the already existent ISO extension methodology to sketch the extended elements that can be part of a CERIF ISO metadata profile.

Finally, the drafted crosswalk has been successfully implemented in the GI-cat discovery broker: a web service able to search amongst heterogeneous data sources and to publish heterogeneous discovery interfaces. Two components have been implemented and added to the extendible software framework: the CERIF profiler, responsible for publishing a discovery interface returning records according to the CERIF data model; and the CERIF accessor able to broker result products described according to the CERIF schema.

Successful prototype use cases demonstrated both the possibility for CERIF repositories to be integrated in ISO compliant infrastructures and the possibility to ingest INSPIRE compliant records coming from heterogeneous sources in a CERIF repository. Indeed, thanks to the extended brokering framework CERIF information can be published and discovered by scientists through

standard interfaces, without any additional implementation effort. Future work will extend tests on existing CERIF repositories.

References

Bigagli, L., Nativi, S., Mazzetti, P. and Villoresi, G. (2004) 'GI-Cat: a Web service for dataset cataloguing based on ISO 19115', *Proceedings of the 15th International Workshop on Database and Expert Systems Applications,* pp 846-850

Boldrini, E. and Nativi, S. (2013) *SeaDataNet metadata profile of ISO 19115*, [online] FP7 Technical Report of the SeaDataNet II Project. http://essi-lab.eu/schemas/SeaDataNet/documents/CDI-profile-XML-encoding-V10.0.pdf (Accessed 10 October 2014).

C4D – CERIF for Datasets. D2.1 Metadata Ontology. *Jisc* 01/2012. [online] http://cerif4datasets.wordpress.com/c4d-deliverables/ (Accessed 10 October 2014).

CERIF – Common European Research Information Format v. 1.2 (2008). [online] www.eurocris.org/Index.php?page=CERIF2008&t=1 (Accessed 10 October 2014).

CERIF – Common European Research Information Format v. 1.6 (2013). [online] www.eurocris.org/Index.php?page=CERIF-1.6&t=1 (Accessed 10 October 2014).

Codelists for description of metadata datasets compliant with ISO 19115 (2008). [online] www.isotc211.org/2005/resources/Codelist/gmxCodelists.xml (Accessed 10 October 2014).

CSW – OGC Catalogue Service for the Web (2007). [online] http://www.opengeospatial.org/standards/cat (Accessed 10 October 2014).

EarthCube – Community Inventory of EarthCube Resources for Geosciences Interoperability. [online] www.nsf.gov/geo/earthcube/ (Accessed 10 October 2014).

ENGAGE – An Infrastructure for Open, Linked Governmental Data Provision towards Research Communities and Citizens. [online] engagedata.eu (Accessed 10 October 2014).

EPOS – The European Plate Observing System. [online] www.epos-eu.org (Accessed 10 October 2014).

EuroGEOSS – European contribution to a Global Earth Observation System of Systems (GEOSS). [online] www.eurogeoss.eu/ (Accessed 10 October 2014).

GEO-DAB – GEO Discovery and Access Broker. [online] http://api.eurogeoss-broker.eu/docs/index.html (Accessed 10 October 2014).

GeoRSS – Geographically Encoded Objects for RSS feeds. [online] http://www.georss.org/ (Accessed 10 October 2014).

GEOSS – Global Earth Observation System of Systems. [online] www.earthobservations.org/geoss.shtml (Accessed 10 October 2014).

GeoTiff – Geostationary Earth Orbit Tagged Image File Format. [online] http://geotiff.osgeo.org/ (Accessed 10 October 2014).

HDF – Hierarchical Data Format. [online] www.hdfgroup.org/HDF5/ (Accessed 10 October 2014).

INSPIRE (2009): *Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119.* EC JRC 02/2009.

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

Jeffery, K., Houssos, N., Jörg, B. and Asserson, A. (2014), 'Research information management: the CERIF approach', *International Journal of Metadata, Semantics and Ontologies*, Vol. 9 No.1, pp. 5-14

Joerg, B., Ruiz-Rube, I., Sicilia, M.A., Dvořák, J., Jeffery, K., Hoellrigl, T., Rasmussen, H.S., Engfer, A., Vestdam, T., Barriocanal, E.G. (2012), 'Connecting closed world research information systems through the linked open data web', *International Journal of Software Engineering and Knowledge Engineering* Vol. 22, pp. 345-364

Manghi, P., Bolikowski, L., Manold, N., Schirrwagen, J., Smith, T. (2010) 'OpenAIREplus: the European Scholarly Communication Data Infrastructure', *D-Lib Magazine* Vol. 18, pp. 9-10

Nativi, S., Mazzetti, P., Bigagli, L. (2008) *Gestione Integrata ed Interoperativa dei Dati Ambientali (GIIDA): Architettura per l'Interoperabilità v.1.1*, Technical report at the Department of Earth and Environment of the Italian National Research Council.

Nativi, S. and Bigagli, L. (2009), 'Discovery, Mediation, and Access Services for Earth Observation Data', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* Vol. 2, pp. 233-240

Nativi, S., Craglia, M. and Pearlman, J. (2013), 'Earth Science Infrastructures Interoperability: The Brokering Approach', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* Vol. 6, pp. 1118-1129

NetCDF – Network Common Data Form. [online] http://www.unidata.ucar.edu/software/netcdf/ (Accessed 10 October 2014).

OPeNDAP – Open-source Project for a Network Data Access Protocol. [online] www.opendap.org/ (Accessed 10 October 2014).

THREDDS – Thematic Real-Time Environmental Distributed Data Services. [online] http://www.unidata.ucar.edu/software/thredds/current/tds/TDS.html (Accessed 10 October 2014).

VOA3R – Virtual Open Access Agriculture & Aquaculture project. [online] www.voa3r.eu (Accessed 10 October 2014).

WCS – OGC Web Coverage Service. [online] http://www.opengeospatial.org/standards/wcs (Accessed 10 October 2014).

WMS – OGC Web Map Service. [online] www.opengeospatial.org/standards/wms (Accessed 10 October 2014).
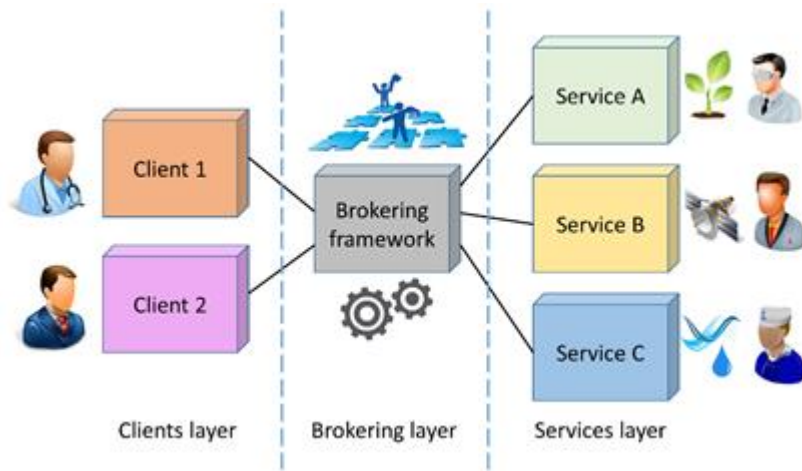
Figures.

Figure 1. A brokering framework mediates between clients and services belonging to different domains. The interoperability burden is thus shifted from clients and services layers to the new brokering layer.
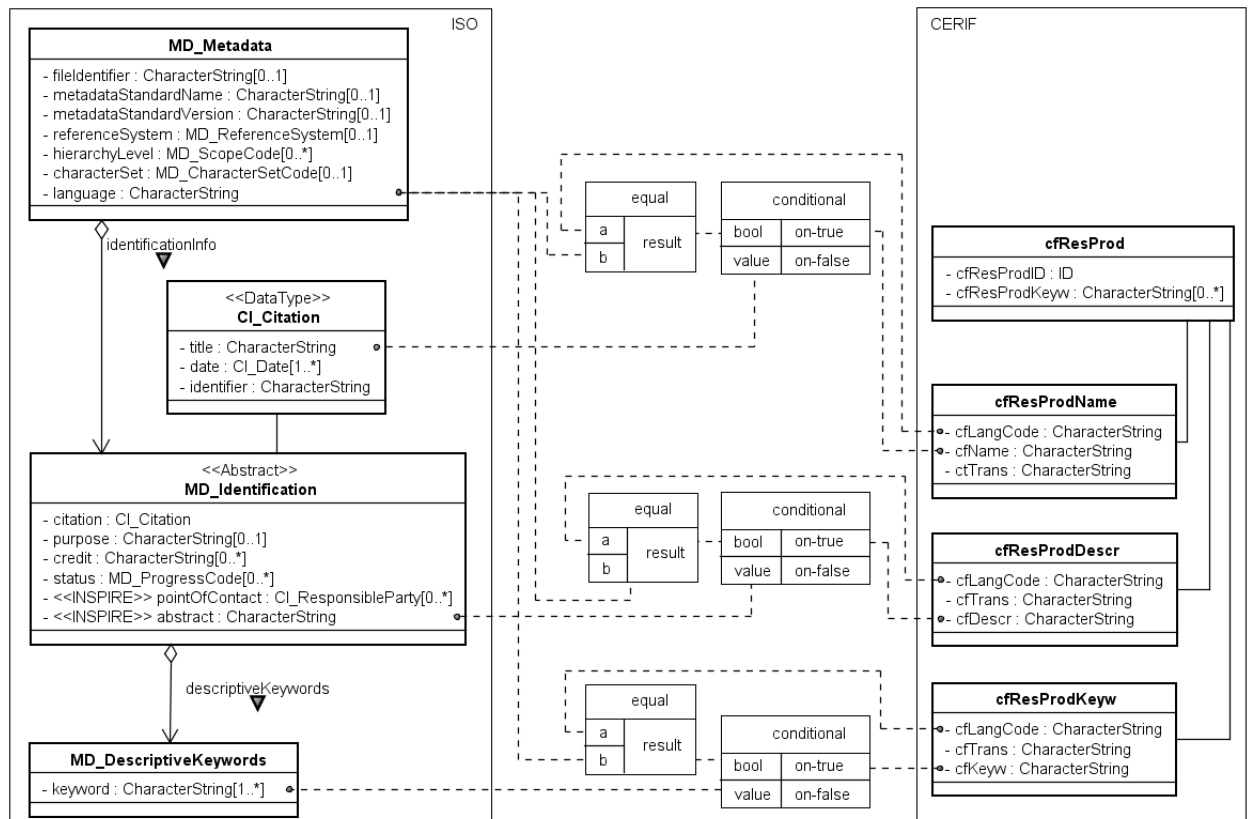
Figure 2. Straightforward mapping of ISO core metadata with CERIF data elements. The following functionalities have been considered: 1) EQUAL to check wheatear the value of the attribute cfLangCode matches with the specific language used by ISO for documenting metadata (attribute language of the entity MD_Metadata); 2) CONDITIONAL to capture the correct translation stored by CERIF using multi-language records based on the result of the EQUAL function.
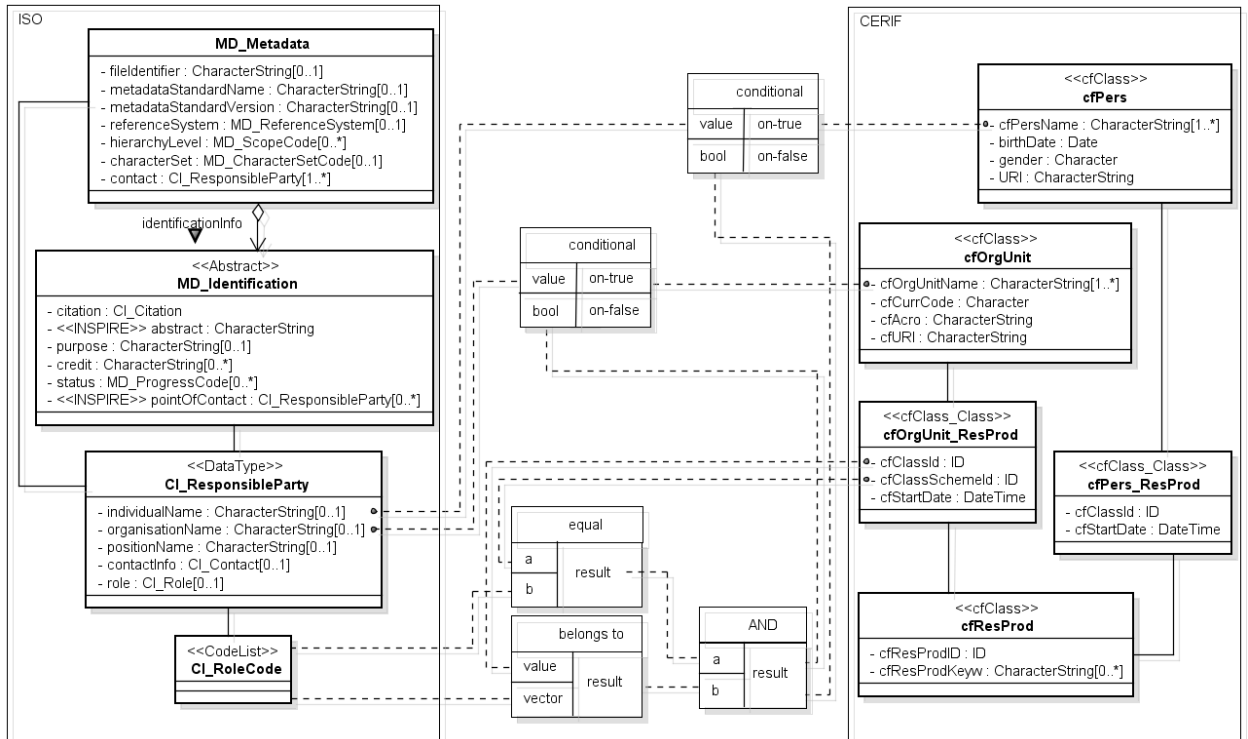

Figure 3. Inferential mapping of the parties involved as well as their role in the management of datasets. The following functionalities have been considered: 1) EQUAL to compare the classification scheme of the organization unit (cfClassSchemeId attribute of the cfOrgUnit_ResProd entity) with the ISO CI_RoleCode list; 2) BELONGS TO to verify whether the role specified in the attribute role of the CI_ResponsibleParty matches with the classification identifier captured in the cfClassId attribute of the cfOrgUnit_ResProd entity; 3) CONDITIONAL to map the individualName and the organizationName of the ISO CI_ResponsibleParty with the relevant attribute in CERIF only if both EQUAL and BELONGS TO functionalities are positive (AND).
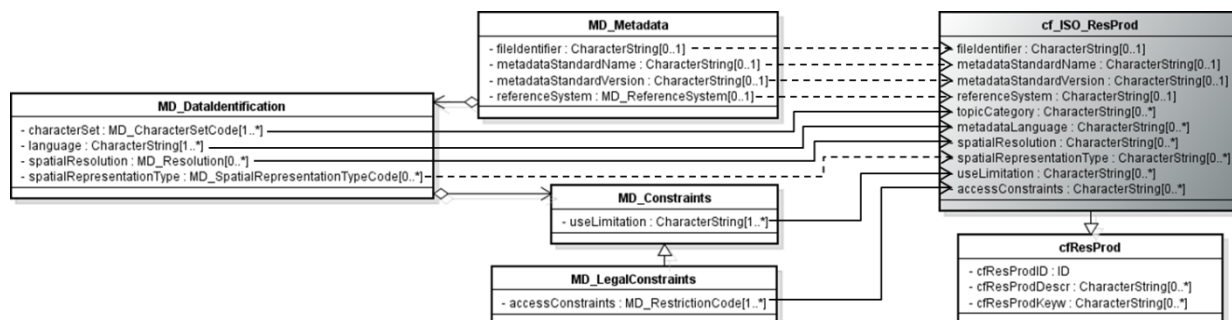
Figure 4. Missing elements (Solid line represents mandatory attributes whereas dashed line represents optional elements)
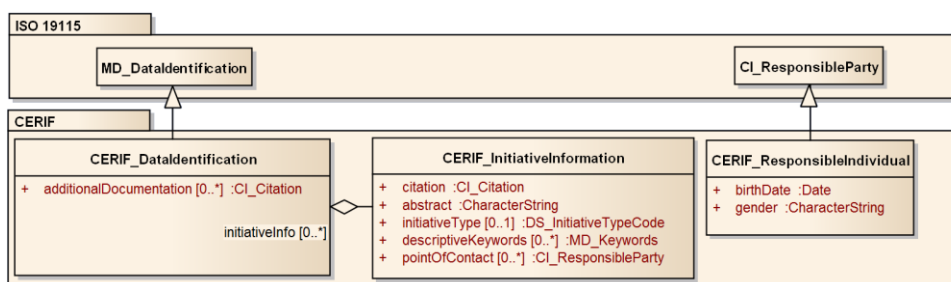

Figure 5. Possible extensions of ISO metadata model with CERIF elements
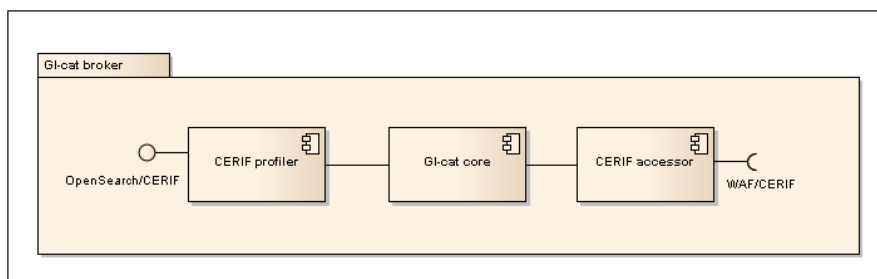

Figure 6. GI-cat component diagram, showing the new CERIF profiler and accessor, designed to support CERIF in the GI-cat brokering framework.
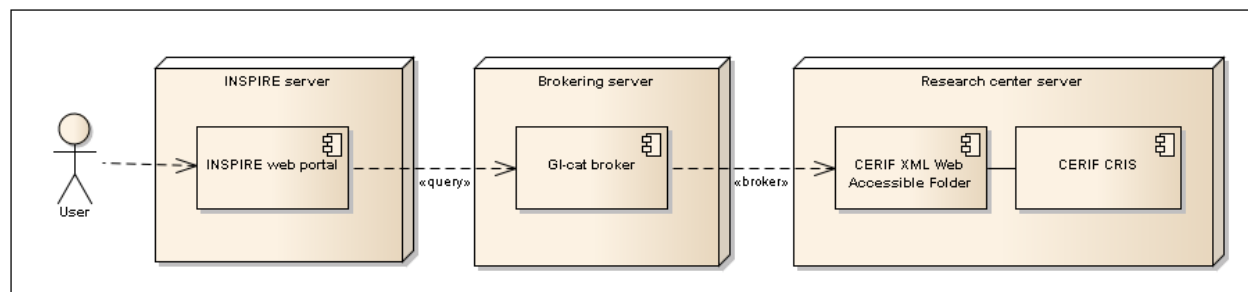

Figure 7. A CERIF CRIS is publishing a Web Accessible Folder (WAF) containing CERIF XML documents. GI-cat is used to broker it and expose publish its content for discovery through a

CSW/ISO INSPIRE compliant interface, making use of the new GI-cat CERIF accessor. An INSPIRE web portal user can thus transparently discover CERIF result products.
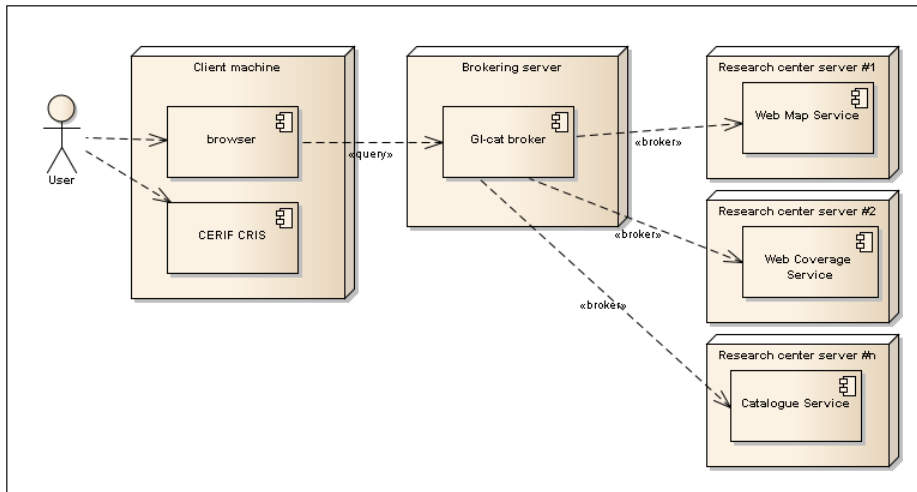

Figure 8. GI-cat being used to broker environmental datasets from a set of heterogeneous data providers. The new GI-cat CERIF profiler enables their discovery through a simple OpenSearch interface returning documents encoded in conformity with the CERIF XML schema. A user is thus able to find the desired result products and import them in a his/her local CERIF based CRIS.

Tables.

INSPIRE elements	INSPIRE Section	ISO 19115 Path	ISO Card.	CERIF Path	CERIF Card.
Dataset title	B1.1	MD_Metadata > MD_DataIdentification.citation > CI_Citation.title	[1..1]	cfResProd > cfResProdName	[1..*]
Geographic Bounding Box	B4.1	MD_Metadata > MD_DataIdentification.extent > EX_Extent > EX_GeographicExtent > EX_GeographicBoundingBox	[1..*]	cfResProd > cfResProd_GeoBBox > cfGeoBBox	[0..*]
Abstract describing the dataset	B1.2	MD_Metadata > MD_DataIdentification.abstract	[1..1]	cfResProd > cfResProdDescr	[1..*]
Dataset keyword	B3	MD_Metadata > MD_DataIdentification.descriptiveKeywords > MD_Keywords	[1..*]	cfResProd > cfResProdKeyw	[1..*]
Unique resource identifier	B1.5	MD_Metadata > MD_DataIdentification.citation > CI_Citation.identifier	[1..*]	cfResProd > cfResProdID	[1..1]
Resource type	B1.3	MD_Metadata.hierarchyLevel	[1..*]	[fixed by the scope to dataset]	
Metadata character set	-	MD_Metadata.characterSet	[1..1]	[fixed to UTF-8]	

Table 1. Straightforward mapping describing ISO and CERIF paths and related cardinalities

INSPIRE mandatory elements	INSPIRE Section	ISO 19115 Path	ISO Card.	CERIF Path	CERIF Card.	CERIF Role specification
Dataset responsible party	B9	MD_Metadata > MD_DataIdentification.pointOf Contact > CI_ResponsibleParty	[1..*]	cfResProd > cfOrgUnit_ResProd > cfOrgUnit > cfOrgUnitName AND cfOrgUnit_EAddr [AND cfResProd > cfPers_ResProd > cfPers > cfPersName]	[1..*]	cfClassId ∈ CI_RoleCode (e.g. "custodian") cfClassSchemeId=" CI_RoleCode"
Metadata point of contact	B10.1	MD_Metadata.contact > CI_ResponsibleParty	[1..*]	cfResProd > cfOrgUnit_ResProd > cfOrgUnit > cfOrgUnitName AND cfOrgUnit_EAddr [AND cfResProd > cfPers_ResProd > cfPers > cfPersName]	[1..*]	cfClassId ∈ CI_RoleCode (e.g. "pointOfContact") cfClassSchemeId=" CI_RoleCode"
Dataset topic category	B2.1	MD_Metadata > MD_DataIdentification.topicCa tegory	[1..*]	cfResProd_Class	[1..*]	cfClassId ∈ MD_TopicCategory Code (e.g. biota) cfClassSchemeId=" MD_TopicCategory Code"

Table 2. Inferential mapping describing ISO and CERIF paths and related cardinalities

INSPIRE elements	INSPIRE Section	ISO 19115 Path	ISO Card.	CERIF Path	CERIF Card.	CERIF Attribute specification
INSPIRE mandatory metadata elements						
Conformity	B7	MD_Metadata > DQ_DataQuality.report	[1..*]	cfResProd > cfResProd_Meas > cfMeas > cfMeasName AND cfValJudgeText	[1..*]	cfMeasName= 'conformity'
Lineage	B6.1	MD_Metadata > DQ_DataQuality.lineage > LI_Lineage	[1..1]	union(cfResProd > cfResProd_Meas > cfMeas > cfMeasDescr)	[1..*]	cfMeasName= 'lineage'
Temporal	B5.1	MD_Metadata >	[1..*]	From minimum to maximum	[1..*]	

extent		MD_DataIdentification.extent > EX_Extent > EX_TemporalExtent		(cfResProd > cfResProd_Meas > cfMeas > cfDateTime)		
Dataset reference date	B5	MD_Metadata > MD_DataIdentification.citation > CI_Citation.date	[1..*]	cfResProd > cfOrgUnit_ResProd > cfOrgUnit > cfStartDate/cfEndDate		cfClassId='author institution'
Metadata date stamp	B10.2	MD_Metadata.dateStamp	[1..1]	cfResProd > cfOrgUnit_ResProd > cfOrgUnit > cfStartDate/cfEndDate	[1..*]	cfClassId='publis her institution'
Metadata language	B10.3	MD_Metadata.language	[1..1]	cfResProd > cfResProdName > cfLangCode	[1..*]	
INSPIRE optional metadata elements						
Distribution format	-	MD_Metadata > MD_Distribution > MD_Format.name AND MD_Format.version	[0..*]	cfResProd > cfResProdVersInfo	[1..*]	
On-line resource	B1.4	MD_Metadata > MD_Distribution > MD_DigitalTransferOption.on Line > CI_OnlineResource	[0..*]	cfURI	[0..1]	

Table 3. Convention mapping describing ISO and CERIF paths and related cardinalities

```xml
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<cerif:CERIF xmlns:cerif="urn:xmlns:org:eurocris:cerif-1.6-2" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" date="2014-04-
07+02:00" sourceDatabase="GEO-DAB" xsi:schemaLocation="urn:xmlns:org:eurocris:cerif-1.6-2
http://www.eurocris.org/Uploads/Web%20pages/CERIF-1.6/CERIF_1.6_2.xsd">
  <cerif:cfResProd>
    <cerif:cfResProdId>2009d331-cefb-4fb1-8828-0e274d646a9f</cerif:cfResProdId>
    <cerif:cfURI>http://webgis1.barrowbc.gov.uk/webgis/bingis.html</cerif:cfURI>
  </cerif:cfResProd>
  <cerif:cfResProdName>
    <cerif:cfResProdId>2009d331-cefb-4fb1-8828-0e274d646a9f</cerif:cfResProdId>
    <cerif:cfName cfLangCode="eng" cfTrans="o">2009 Aerial Photography</cerif:cfName>
  </cerif:cfResProdName>
  <cerif:cfGeoBBox>
    <cerif:cfGeoBBoxId>ID-0</cerif:cfGeoBBoxId>
    <cerif:cfWBLong>-3.32482</cerif:cfWBLong>
    <cerif:cfEBLong>-3.12439</cerif:cfEBLong>
    <cerif:cfSBLat>54.03964</cerif:cfSBLat>
    <cerif:cfNBLat>54.21841</cerif:cfNBLat>
  </cerif:cfGeoBBox>
  <cerif:cfResProd_GeoBBox>
    <cerif:cfResProdId>2009d331-cefb-4fb1-8828-0e274d646a9f</cerif:cfResProdId>
    <cerif:cfGeoBBoxId>ID-0</cerif:cfGeoBBoxId>
    <cerif:cfClassId>ID-1</cerif:cfClassId>
    <cerif:cfClassSchemeId>ID-2</cerif:cfClassSchemeId>
  </cerif:cfResProd_GeoBBox>
  <cerif:cfResProdDescr>
    <cerif:cfResProdId>2009d331-cefb-4fb1-8828-0e274d646a9f</cerif:cfResProdId>
    <cerif:cfDescr cfLangCode="eng" cfTrans="o">Orthorectified aerial photography of Barrow Council administrative area flown on 31st
May 2009 at a resolution of 10cm.</cerif:cfDescr>
  </cerif:cfResProdDescr>
  <cerif:cfResProdKeyw>
    <cerif:cfResProdId>2009d331-cefb-4fb1-8828-0e274d646a9f</cerif:cfResProdId>
    <cerif:cfKeyw cfLangCode="eng" cfTrans="o">Photography</cerif:cfKeyw>
  </cerif:cfResProdKeyw>
  <cerif:cfResProdKeyw>
    <cerif:cfResProdId>2009d331-cefb-4fb1-8828-0e274d646a9f</cerif:cfResProdId>
    <cerif:cfKeyw cfLangCode="eng" cfTrans="o">Orthoimagery</cerif:cfKeyw>
  </cerif:cfResProdKeyw>
  <cerif:cfOrgUnit_ResProd>
```

```xml
    <cerif:cfOrgUnitId>ID-3</cerif:cfOrgUnitId>
    <cerif:cfResProdId>2009d331-cefb-4fb1-8828-0e274d646a9f</cerif:cfResProdId>
    <cerif:cfClassId>ID-8</cerif:cfClassId>
    <cerif:cfClassSchemeId>ID-9</cerif:cfClassSchemeId>
    <cerif:cfStartDate>2009-05-31T02:00:00.000+02:00</cerif:cfStartDate>
    <cerif:cfEndDate>2009-05-31T02:00:00.000+02:00</cerif:cfEndDate>
  </cerif:cfOrgUnit_ResProd>
  <cerif:cfOrgUnit>
    <cerif:cfOrgUnitId>ID-3</cerif:cfOrgUnitId>
  </cerif:cfOrgUnit>
  <cerif:cfOrgUnitName>
    <cerif:cfOrgUnitId>ID-3</cerif:cfOrgUnitId>
    <cerif:cfName cfLangCode="eng" cfTrans="o">Barrow Borough Council</cerif:cfName>
  </cerif:cfOrgUnitName>
  <cerif:cfPers_ResProd>
    <cerif:cfPersId>ID-4</cerif:cfPersId>
    <cerif:cfResProdId>2009d331-cefb-4fb1-8828-0e274d646a9f</cerif:cfResProdId>
    <cerif:cfClassId>distributor</cerif:cfClassId>
    <cerif:cfClassSchemeId>CI_RoleCode</cerif:cfClassSchemeId>
  </cerif:cfPers_ResProd>
  <cerif:cfPers>
    <cerif:cfPersId>ID-4</cerif:cfPersId>
  </cerif:cfPers>
  <cerif:cfOrgUnit_EAddr>
    <cerif:cfOrgUnitId>ID-3</cerif:cfOrgUnitId>
    <cerif:cfEAddrId>ID-5</cerif:cfEAddrId>
    <cerif:cfClassId>ID-6</cerif:cfClassId>
    <cerif:cfClassSchemeId>ID-7</cerif:cfClassSchemeId>
  </cerif:cfOrgUnit_EAddr>
  <cerif:cfEAddr>
    <cerif:cfEAddrId>ID-5</cerif:cfEAddrId>
    <cerif:cfPAddrId>gis@barrowbc.gov.uk</cerif:cfPAddrId>
  </cerif:cfEAddr>
  <cerif:cfOrgUnit_ResProd>
    <cerif:cfOrgUnitId>ID-10</cerif:cfOrgUnitId>
    <cerif:cfResProdId>2009d331-cefb-4fb1-8828-0e274d646a9f</cerif:cfResProdId>
    <cerif:cfClassId>author</cerif:cfClassId>
    <cerif:cfClassSchemeId>CI_RoleCode</cerif:cfClassSchemeId>
    <cerif:cfStartDate>2014-04-07T02:00:00.000+02:00</cerif:cfStartDate>
    <cerif:cfEndDate>2014-04-07T02:00:00.000+02:00</cerif:cfEndDate>
  </cerif:cfOrgUnit_ResProd>
  <cerif:cfOrgUnit>
    <cerif:cfOrgUnitId>ID-10</cerif:cfOrgUnitId>
  </cerif:cfOrgUnit>
  <cerif:cfOrgUnitName>
    <cerif:cfOrgUnitId>ID-10</cerif:cfOrgUnitId>
    <cerif:cfName cfLangCode="eng" cfTrans="o">Barrow Borough Council</cerif:cfName>
  </cerif:cfOrgUnitName>
  <cerif:cfPers_ResProd>
    <cerif:cfPersId>ID-11</cerif:cfPersId>
    <cerif:cfResProdId>2009d331-cefb-4fb1-8828-0e274d646a9f</cerif:cfResProdId>
    <cerif:cfClassId>author</cerif:cfClassId>
    <cerif:cfClassSchemeId>CI_RoleCode</cerif:cfClassSchemeId>
  </cerif:cfPers_ResProd>
  <cerif:cfPers>
    <cerif:cfPersId>ID-11</cerif:cfPersId>
  </cerif:cfPers>
  <cerif:cfOrgUnit_EAddr>
    <cerif:cfOrgUnitId>ID-10</cerif:cfOrgUnitId>
    <cerif:cfEAddrId>ID-12</cerif:cfEAddrId>
    <cerif:cfClassId>ID-13</cerif:cfClassId>
    <cerif:cfClassSchemeId>ID-14</cerif:cfClassSchemeId>
  </cerif:cfOrgUnit_EAddr>
  <cerif:cfEAddr>
```

```
      <cerif:cfEAddrId>ID-12</cerif:cfEAddrId>
      <cerif:cfPAddrId>gis@barrowbc.gov.uk</cerif:cfPAddrId>
   </cerif:cfEAddr>
   <cerif:cfResProd_Class>
      <cerif:cfResProdId>2009d331-cefb-4fb1-8828-0e274d646a9f</cerif:cfResProdId>
      <cerif:cfClassId>imageryBaseMapsEarthCover</cerif:cfClassId>
      <cerif:cfClassSchemeId>MD_TopicCategoryCode</cerif:cfClassSchemeId>
   </cerif:cfResProd_Class>
   <cerif:cfResProd_Meas>
      <cerif:cfResProdId>2009d331-cefb-4fb1-8828-0e274d646a9f</cerif:cfResProdId>
      <cerif:cfMeasId>ID-15</cerif:cfMeasId>
      <cerif:cfClassId>ID-16</cerif:cfClassId>
      <cerif:cfClassSchemeId>ID-17</cerif:cfClassSchemeId>
   </cerif:cfResProd_Meas>
   <cerif:cfMeas>
      <cerif:cfMeasId>ID-15</cerif:cfMeasId>
      <cerif:cfValJudgeText>conformity to INSPIRE specification</cerif:cfValJudgeText>
   </cerif:cfMeas>
   <cerif:cfMeasName>
      <cerif:cfMeasId>ID-15</cerif:cfMeasId>
      <cerif:cfName cfLangCode="eng" cfTrans="o">Conformity</cerif:cfName>
   </cerif:cfMeasName>
   <cerif:cfResProd_Meas>
      <cerif:cfResProdId>2009d331-cefb-4fb1-8828-0e274d646a9f</cerif:cfResProdId>
      <cerif:cfMeasId>ID-18</cerif:cfMeasId>
      <cerif:cfClassId>ID-19</cerif:cfClassId>
      <cerif:cfClassSchemeId>ID-20</cerif:cfClassSchemeId>
   </cerif:cfResProd_Meas>
   <cerif:cfMeas>
      <cerif:cfMeasId>ID-18</cerif:cfMeasId>
      <cerif:cfValJudgeText>process steps for this dataset</cerif:cfValJudgeText>
   </cerif:cfMeas>
   <cerif:cfMeasName>
      <cerif:cfMeasId>ID-18</cerif:cfMeasId>
      <cerif:cfName cfLangCode="eng" cfTrans="o">Lineage</cerif:cfName>
   </cerif:cfMeasName>
   <cerif:cfResProd_Meas>
      <cerif:cfResProdId>2009d331-cefb-4fb1-8828-0e274d646a9f</cerif:cfResProdId>
      <cerif:cfMeasId>ID-21</cerif:cfMeasId>
      <cerif:cfClassId>ID-22</cerif:cfClassId>
      <cerif:cfClassSchemeId>ID-23</cerif:cfClassSchemeId>
   </cerif:cfResProd_Meas>
   <cerif:cfMeas>
      <cerif:cfMeasId>ID-15</cerif:cfMeasId>
      <cerif:cfValJudgeText>conformity to INSPIRE specification</cerif:cfValJudgeText>
   </cerif:cfMeas>
   <cerif:cfResProd_Meas>
      <cerif:cfResProdId>2009d331-cefb-4fb1-8828-0e274d646a9f</cerif:cfResProdId>
      <cerif:cfMeasId>ID-26</cerif:cfMeasId>
      <cerif:cfClassId>ID-24</cerif:cfClassId>
      <cerif:cfClassSchemeId>ID-25</cerif:cfClassSchemeId>
   </cerif:cfResProd_Meas>
   <cerif:cfMeas>
      <cerif:cfMeasId>ID-18</cerif:cfMeasId>
      <cerif:cfValJudgeText>process steps for this dataset</cerif:cfValJudgeText>
   </cerif:cfMeas>
 </cerif:CERIF>
```

Table 4. XML excerpt document describingof the CERIF encoding of a sample INSPIRE
metadata document, as encoded by the OpenSearch/CERIF profiler