



CRIS 2014

Integrating CERIF entities in a multidisciplinary e-infrastructure for environmental research data

Enrico Boldrini^a, Daniela Luzi^{b*}, Stefano Nativi^a, Fabrizio Pecoraro^b

^aInstitute of Atmospheric Pollution Research, National Research Council (CNR-IIA), Via Madonna del Piano, 10, 50019, Sesto Fiorentino, Italy

^bInstitute for Research on Population and Social Policies, National Research Council (CNR-IRPPS), Via Palestro 32, 00185, Rome, Italy

Abstract

The paper proposes 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. A two-way crosswalk is described resulting in the identification of a core of corresponding metadata and a proposal of extensions of the CERIF model. Extensions of ISO concepts are also 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 tests demonstrated the possibility for CERIF information to be integrated in ISO compliant infrastructures and for INSPIRE information to be captured in CERIF.

© 2014 Elsevier B.V This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

Peer-review under responsibility of euroCRIS

Keywords: CERIF, ISO 19115, environmental dataset, research information, metadata crosswalk, brokering approach.

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. In order to gain knowledge of global environmental issues, such as climate change, flood and landslide risks, integration of enormous volumes of multidisciplinary and dispersed data is needed. To solve this issue it is necessary to develop e-infrastructure that integrate different sources of information at both disciplinary and cross-disciplinary levels, thus harmonizing the various data models and standards that cover specific domains with

*Daniela Luzi. Tel.: +39-06-492724-214; fax: +39-06-49383724.

E-mail address: d.luzi@irpps.cnr.it

different granularity and scopes. To facilitate the addressing of these challenges, especially in global and/or multidisciplinary contexts (see the Global Earth Observation System of Systems: GEOSS¹; the European Infrastructure for Spatial Information in the European Community: INSPIRE²; the US NSF EarthCube³; etc.), the metadata Brokering approach was introduced to complement the standardization effort⁴. The Brokering architecture was first developed by the EU FP7 EuroGEOSS project⁵ and further tested at national level by the CNR to build the GIIDA (Interdepartmental Infrastructure for Environmental Data Sharing) e-infrastructure⁶. This approach makes it possible to build a system of (heterogeneous) systems without imposing any common standard or technology on the resource Providers. It is based on three types of brokers: Access, Semantic and Discovery, each one fulfilling a specific functionality to integrate distributed resources.

The aim of this paper is to extend the current Brokering framework integrating context research information with environmental research data. CERIF (Common European Research Information Format)⁷ is a likely candidate for the discovery of information sources 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 (OpenAIREplus⁸) make 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 different solutions for the integration of research data in domain-specific environments.

Other integration efforts have been recently carried out proposing CERIF extensions to enable the exchange of research information connected with dataset in specific domains. The C4D (CERIF for Datasets) project⁹ 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* 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 ENGAGE¹⁰ and EPOS¹¹ 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)¹² and had an application in the VOA3R (Virtual Open Access Agriculture & Aquaculture) 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.

Our approach 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 19115¹³ INSPIRE profile to and from the CERIF model (paragraph 3 and 4) and also presents a CERIF profiler 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

In the context of metadata for spatial data we consider the INSPIRE profile of ISO 19115 defined in the Implementing Rules² document that 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 particular, we propose a two-way data crosswalk from the CERIF to the ISO schema with a twofold aim: a) mapping from INSPIRE to CERIF to provide a CERIF guideline for the description of research data according to the INSPIRE profile ISO 19115; b) mapping from CERIF to ISO to provide a solution to describe CERIF concepts related to *cfResProd* using the ISO 19115 data model.

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 the logical and physical 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 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, but also the structural features, such as the CERIF semantic layer, the multiple relationships between entities and their temporal duration. Although these features have the advantages of making the CERIF model flexible and applicable to different domains, they make it necessary to identify specific solutions to establish a coherent integration based on both semantics and structure.

3. Mapping from INSPIRE ISO 19115 profile to CERIF

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 shared 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.

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) (Table 1). This results in a straightforward mapping that allows discovery brokers such as GEO-DAB⁴ 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 datasets resulted in a project/research activity. This extension was already applied when the entity cfResPub was introduced to univocally identify publications among the various research results¹⁴.

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

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_GeoBoundingBox > cfGeoBoundingBox	[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]	

3.2. Inferential mapping

CERIF makes it possible to explicitly express the relationship between Research entities (Person, Project, Organization, Research products, etc.) through the so-called Link Entities. Moreover, *cfClassId* and *cfClassSchemeId* are used as primary foreign key to associate the Link Entities with the *cfClass* and the *cfClassScheme* of the Semantic Layer that defines the *role* played by a source object related to a target object (for instance a *cfPers* plays the role of Author in a *cfResPubl*).

Also ISO 19115 uses the attribute *role* to express the function performed by the responsible party in the management of datasets. Moreover, ISO maintains dictionaries of the so-called *CodeList* that contains a set of possible roles and related definitions used in the research data domain¹⁵. Given these premises, an inferential mapping can be used to link *cfClass* with roles listed in the relevant ISO *CodeList* dictionaries (Table 2).

In particular, the ISO “dataset responsible party” package describes both responsible organization and responsible person modeling them in a single class (*CI_ResponsibleParty*), while CERIF provides two entities to define this information (*cfOrgUnit* and *cfPers*). Thus a mapping from ISO to CERIF can be expressed with *cfOrgUnitName* or *cfPersName* associated with the *cfResProd* with its relevant role. In this way important context information can be mapped, such as the organization responsible for the creation, distribution, and/or custodian, or the person who is the author of the dataset, considering the classification reported on ISO *CI_RoleCode*.

A similar approach can be applied to map the ISO “dataset topic category” using the *cfResProd_Class* with a specific taxonomy based on ISO *MD_TopicCategoryCode* terms.

Table 2. Inferential 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.pointOfContact > CI_ResponsibleParty	[1..*]	<i>cfResProd</i> > <i>cfOrgUnit_ResProd</i> > <i>cfOrgUnit</i> > <i>cfOrgUnitName</i> AND <i>cfOrgUnit_EAddr</i> [AND <i>cfResProd</i> > <i>cfPers_ResProd</i> > <i>cfPers</i> > <i>cfPersName</i>]	[1..*]	<i>cfClassId</i> ∈ <i>CI_RoleCode</i> (e.g. “custodian”) <i>cfClassSchemeId</i> =” <i>CI_RoleCode</i> ”
Metadata point of contact	B10.1	MD_Metadata.contact > CI_ResponsibleParty	[1..*]	<i>cfResProd</i> > <i>cfOrgUnit_ResProd</i> > <i>cfOrgUnit</i> > <i>cfOrgUnitName</i> AND <i>cfOrgUnit_EAddr</i> [AND <i>cfResProd</i> > <i>cfPers_ResProd</i> > <i>cfPers</i> > <i>cfPersName</i>]	[1..*]	<i>cfClassId</i> ∈ <i>CI_RoleCode</i> (e.g. “pointOfContact”) <i>cfClassSchemeId</i> =” <i>CI_RoleCode</i> ”
Dataset topic category	B2.1	MD_Metadata > MD_DataIdentification.topicCategory	[1..*]	<i>cfResProd_Class</i>	[1..*]	<i>cfClassId</i> ∈ <i>MD_TopicCategoryCode</i> (e.g. biota) <i>cfClassSchemeId</i> =” <i>MD_TopicCategoryCode</i> ”

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 not only for evaluation purposes, but also because it is fundamental for 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 convention of the parties exposing metadata using the attribute *cfValJudgeText* of the *cfMeas* specifying in the *cfMeasName* the type of measurement used to collect data (fixed to “conformity” and/or “lineage”). The CERIF entity *cfMeas* can be also 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 research product. Therefore, by convention the cfStartDate element can be used to accommodate both date references. To distinguish between these two ISO elements the cfClass 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.

Table 3 Convention 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
<i>INSPIRE mandatory metadata elements</i>						
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= ‘publisher institution’
Metadata language	B10.3	MD_Metadata.language	[1..1]	cfResProd > cfResProdName > cfLangCode	[1..*]	
<i>INSPIRE optional metadata elements</i>						
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.onLine > CI_OnlineResource	[0..*]	cfURI	[0..1]	

3.4. Missing elements: A proposal of CERIF extension

Other mandatory data elements of the INSPIRE profile of ISO 19115 that cannot be easily documented in CERIF concern the constraints related to the access and use of the dataset (“conditions for access and use” and “limitations on public access”). To express this important information for the re-use and accessibility of datasets, the CERIF entity cfResProd should be extended including respectively cfCopyright and cfUseLimit (Fig. 1).

Moreover, other extensions of the cfResProd could be introduced to represent the language used within the dataset (the ISO “dataset language”) as well as the standard used to encode datasets (the ISO “dataset character set”). This could be achieved extending the cfResProd with the attributes cfLangCode and cfCharSet. A similar approach can be used to extend the CERIF model in order to include optional ISO elements related to both metadata (“metadata file identifier”, “metadata standard name”, “metadata standard version” and “reference system”) and dataset (“spatial representation type”).

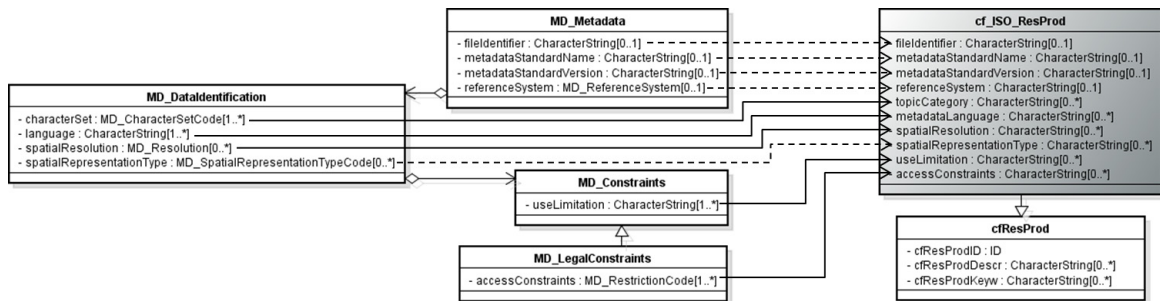


Fig. 1. Missing elements (Solid line represents mandatory attributes whereas dashed line represents optional elements)

The proposed mapping intentionally does not use of the Dublin Core derived CERIF elements, because they are considered “deprecated” by the latest CERIF 1.6 specification. Otherwise some 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 concepts with CERIF data model, introducing research information, such as CERIF data 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 a profile 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 SeaDataNet profile¹⁶. Similarly, a CERIF profile can extend the MD_DataIdentification class with the subclass SDN_DataIdentification, containing the optional attribute additionalDocumentation of type SDN_Citation to collect bibliographic references to the dataset, such as articles and related publications. SDN_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, author). Thus additional CERIF attributes can be added, such as birthdate, gender, URI. The same stand 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. 2 formalizes these metadata extensions according to ISO 19115 extension methodology. Of course other CERIF concepts can be mapped to ISO that are not considered here for space limits.

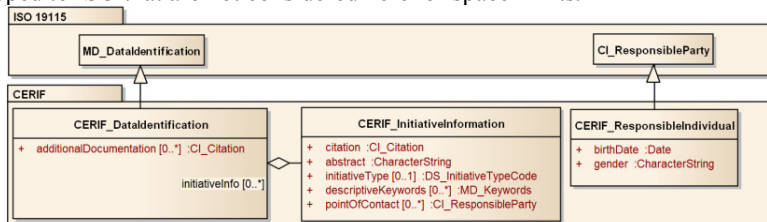


Fig. 2 Possible extensions of ISO metadata model with CERIF elements

5. Implementation in a brokering framework

The two mappings were implemented and successfully tested by the GI-cat discovery and semantic broker¹⁷. GI-cat enables an harmonized (semantic) discovery of datasets published by remote systems via a plethora of heterogeneous services interfaces¹⁸ (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 modules: called “accessors” and “profilers”, respectively. GI-cat was extended with a CERIF “accessor” and “profiler” (see Fig. 3):

- **WAF/CERIF accessor** enables brokering of CERIF datasets published on a given Web Accessible Folder (WAF) according to the CERIF XML Schema, without requiring any additional development by the information sources. 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. Thus, CERIF datasets can suddenly be discovered through one of the many interfaces already published by the broker;
- **OpenSearch/CERIF profiler** enables discovery of resources being brokered by the GI-cat framework through the OpenSearch protocols returning documents which are conform to the CERIF XML Schema. This component translates from the GI-cat internal data model to the CERIF model according to the proposed mapping. Hence, it can be used to discover a wide range of environmental datasets formatting them according to the CERIF model.



Fig. 3 The two new GI-cat components: CERIF profiler and accessor, designed to support CERIF data model in the GI-cat brokering framework.

The extended GI-cat broker framework was finally tested for a couple of simple use cases: 1) a WAF publishing CERIF XML documents is exposed through a CSW INSPIRE compliant interface, making use of the new GI-cat CERIF *accessor*; 2) a WAF publishing INSPIRE compliant documents (encoded in ISO 19139) is exposed as a CERIF catalog using the OpenSearch protocol; this is achieved making use of the new GI-cat CERIF *profiler*.

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

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

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<cerif:CERIF xmlns:cerif="urn:xmlns.org:eurocris:cerif-1.6-2" 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:CERIF>

```

6. Conclusions

The paper proposes different solutions to integrate CERIF model with ISO 19115 environmental datasets model based on a two-way crosswalk. A first result of this mapping is the identification of a minimum set of elements based on the correspondence of the CERIF “result product” concept with the ISO 19115 “dataset” concept.

To accommodate additional information required by INSPIRE in CERIF, we also propose to extend the cfResProd entity including other relevant attributes that can be examined by the euroCRIS community in charge of maintaining the CERIF model. 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 establish specific constraints and rules to establish consistent semantic integration.

Another proposed solution concerns the development of a CERIF profile to extend ISO 19115 concepts with contextual research information. Moreover, the proposed crosswalk has been implemented in the GI-cat discovery framework and successful tests demonstrated the possibility for CERIF repositories to be integrated in ISO compliant infrastructures. Thanks to the brokering framework CERIF information can thus be published and discovered by scientists through standard interfaces, without any additional implementation effort. Future work will include service discoverability, extending the mapping to ISO 19119.

References

- 1 GEOSS – Global Earth Observation System of Systems, www.earthobservations.org/geoss.shtml
- 2 INSPIRE – Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119, *EC JRC* 02/2009.
- 3 EarthCube – Community Inventory of EarthCube Resources for Geosciences Interoperability, www.nsf.gov/geo/earthcube/
- 4 Nativi S, Craglia M, Pearlman J. Earth Science Infrastructures Interoperability: The Brokering Approach. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of* 2013;6: 1118-29.
- 5 EuroGEOSS – European contribution to a Global Earth Observation System of Systems (GEOSS), www.eurogeoss.eu/
- 6 Nativi S, Mazzetti P, Bigagli L. Gestione Integrata ed Interoperativa dei Dati Ambientali (GIIDA): Architettura per l’Interoperabilità v.1.1., 2008.
- 7 CERIF – Common European Research Information Format v. 1.6, www.eurocris.org/Index.php?page=CERIF-1.6&t=1
- 8 Manghi P, Bolikowski L, Manold N, Schirrwagen J, Smith T. OpenAIREplus: the European Scholarly Communication Data Infrastructure. *D-Lib Magazine* 2010;18.
- 9 C4D – CERIF for Datasets. D2.1 Metadata Ontology. *Jisc* 01/2012.
- 10 ENGAGE – An Infrastructure for Open, Linked Governmental Data Provision towards Research Communities and Citizens, engagedata.eu
- 11 EPOS – The European Plate Observing System, www.epos-eu.org
- 12 Joerg B, Ruiz-Rube I, Sicilia MA, Dvořák J, Jeffery K, Hoellrigl T, Rasmussen HS, et al. Connecting closed world research information systems through the linked open data web. *International Journal of Software Engineering and Knowledge Engineering* 2012;22: 345-364.
- 13 ISO 19115:2003 Geographic information – Metadata, 2003.
- 14 CERIF – Common European Research Information Format v. 1.2 (2008), www.eurocris.org/Index.php?page=CERIF2008&t=1
- 15 Codelists for description of metadata datasets compliant with ISO 19115, www.isotc211.org/2005/resources/Codelist/gmxCodelists.xml
- 16 Boldrini E, Nativi S. SeaDataNet metadata profile of ISO 19115. 09/2013
- 17 Nativi S, Bigagli L. Discovery, Mediation, and Access Services for Earth Observation Data. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of* 2009;2:233-40.
- 18 Bigagli L, Nativi S, Mazzetti P, Villorosi G. GI-Cat: a Web service for dataset cataloguing based on ISO 19115. *Database and Expert Systems Applications, Proceedings of 15th International Workshop on* 2004:846-50.