October, 17th 2007

CERIF2006-1.1 Full Data Model (FDM)
Model Introduction and Specification

Location:

http://www.eurocris.org/

Editors:

Brigitte Jörg, DFKI GmbH, Saarbrücken, Germany
Keith Jeffery, Science and Technology Facilities Council, Didcot, UK
Anne Asserson, University of Bergen, Bergen, Norway
Geert van Grootel, Flemish Government, Economy, Science and Innovation, Brussels, Belgium
Edward Grabczewski, Science and Technology Facilities Council, Didcot, UK

Abstract:

CERIF (Common European Research Information Format) is a model format to manage Research Information. The CERIF2006-1.1 Full Data Model (FDM) specification document provides information about the CERIF2006 model; briefly introduces its history and gives insight into the model structure and range. Compared to its preceding CERIF2004 release a major change in the model has been incorporated with respect to the so called Semantic Layer. Moreover, an XML data exchange specification has been released.

CERIF (Common European Research Information Format) was developed with the support of the EC (European Commission) in two major phases: 1987-1990 and 1997-1999. It is a standard; technically it is an EU (European Union) Recommendation to member states. Since 2002 care and custody of CERIF has been handed by the EC to euroCRIS (www.eurocris.org) a not-for-profit organisation dedicated to the promotion of CRIS (Current Research Information Systems).

Status:

Model improvements are based on discussions among euroCRIS CERIF task group members. This document will be updated alongside model updates.

Table of Contents

1. Introduction

Most nation-states have publicly-supported research programmes. It is realised that public sponsorship of research and development leads to wealth creation and improvement in the quality of life. Because public funding is involved, it is necessary for there to be appropriate governance, and also for the information to be available to the public. Broadly, each nation state has a similar research process of: strategic planning; programme announcement; call for proposals; proposal evaluation and awarding; project result monitoring, project result exploitation. However, research is international. A research project in country A is likely to be based on previous research in several other countries. Many research projects are now transnational: well-known examples include the human genome and climate change but there are many others, especially where expensive infrastructure is utilised such as particle physics or space science. Furthermore, knowledge of the research activity in country A may influence the strategy towards research – including priorities and resources provided – in country B. Thus there is a need to share research information across countries, or even between different funding agencies in the same country. Research Information is used by researchers (to find partners, to track competitors, to form collaborations); research managers (to assess performance and research outputs and to find reviewers for research proposals); research strategists (to decide on priorities and resourcing compared with other countries); publication editors (to find reviewers and potential authors); intermediaries/brokers (to find research products and ideas that can be carried forward with knowledge/technology transfer to wealth creation); the media (to communicate the results of R&D in a socio-economic context) and the general public (for interest). Research Information is relevant for actors in scientific environments as well as for decision makers to support related organization, management and planning. We consider Research Information as the transmitter between Science and Society and as such as a powerful instrument for governance. Having such an impact, Research Information has to be collected carefully and preserved systematically, in order to most effectively support society and the individuals within.

1.1 CERIF2006-1.1 FDM in Brief

CERIF (Common European Research Information Format) is a model format to manage Research Information. With the latest release the model structure is highly normalized and allows for flexibility and scalability in applications by introduction of the so called Semantic Layer (Classification Entities). The strength of CERIF is with semantic relationships between research entities. Beyond that CERIF enables communication and data exchange across applications via the CERIF XML data exchange format [1]. CERIF XML has been designed for data exchange at a machine level and therefore corresponds closely with the structure of the CERIF model.

The CERIF 2006 release has been modeled with the latest version of the Toad Data Modeler[1], an easy to use tool for designing database structures, developed and maintained by Quest Software[2]. With Toad it is possible to model ERM diagrams, to generate SQL scripts for most common databases (Oracle, Microsoft, IBM, etc.), to reverse engineer from databases, to create screenshots of the model and model parts, to model at physical and logical level. SQL scripts are generated automatically from physical level [3].

1.2 Purpose of the CERIF2006-1.1 FDM Specification

This model introduction and specification document gives insight into the CERIF data model for potential use cases and applications by introduction of the involved research entities and their relationships in different kinds of research environments.

1.3 Scope of CERIF2006-1.1 FDM Specification

The latest release of the CERIF 2006 FDM comprises of the following components:

- CERIF2006-1.1 FDM – Model Introduction and Specification document
- CERIF2006-1.1 FDM – SQL scripts for most common databases

[1] Toad Data Modeler: http://www.toadsoft.com/toaddm/toad_data_modeler.htm
[2] Quest Software: http://www.quest.com/

- CERIF2006-1.1 FDM – HTML documentation of the Logical Level
- CERIF2006-1.1 FDM – HTML documentation of the Physical Level
- CERIF2006XML-1.1 FDM – Data Exchange Format Specification
  (Part of CERIF 2006-1.1 FDM, but separate documents and corresponding files.)

The CERIF 2006 related files and documents can be downloaded from the public euroCRIS website. Additional files are available for members only: http://eurocris.sharepointsite.com/default.aspx

2. The CERIF2006-1.1 FDM

This model introduction and specification document presents abstract model views to introduce the CERIF entities and their relationships and refers to screenshots from Toad Modeler for more detailed views. The abstract views provide a birds-eye perspective, whereas the model screenshots direct to the level of database and implementation with attributes, datatypes and keys. The entity names shown in the screenshots represent the physical level corresponding to table names of databases; all entity and attribute names include the CERIF prefix "cf". For consistency of SQL scripts across common databases the CERIF model has been modelled at a logical and a physical level. Where the logical level is rather meant for reading by human beings, the physical level is meant for processing by machines. Therefore, the table names at physical level have been shortened to avoid uncontrolled truncations in table names with proprietary implementations, where a length of table names is restricted to a number of characters. The table names at physical level are still understandable by human readers (see Appendix *8.2 List of CERIF Entities*).

Dynamic CERIF model documentations in HTML are available from the public euroCRIS website http://www.euroCRIS.org/ for navigation, both at logical and physical level including image extracts from model parts. The SQL scripts are available for members only. With a normalization of the CERIF structure, the CERIF 2006 model made a first step towards a Service Oriented Architecture (SOA) concept by modularizing system components and offering them separately as services.

2.1 CERIF Entities

With the latest release of the CERIF 2006 model the entity ResultPublication has become a core entity. For scalability, flexibility and simplicity reasons the model has been normalised and all role and type definitions have been moved to classification entities (the so called Semantic Layer) containing the classification schemes. The CERIF entities will be presented as five different groups as the entire model is rather complex:

(1) Core CERIF Entities
(2) 2nd Level CERIF Entities
(3) CERIF Link Entities
(4) Language-dependent CERIF Entities
(5) CERIF Classification Entities

2.2 Core CERIF Entities

The core CERIF entities are Person, OrganisationUnit, ResultPublication and Project.

Figure 1 shows the core entities and their recursive and linking relationhips in abstract view. Each core entity recursively links to itself and moreover has relationships with other core entities. The core CERIF entities represent scientific actors and their main research activities.

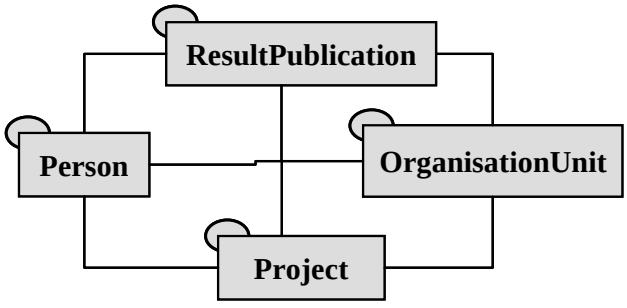

Figure 1: CERIF Core Entities in Abstract View

Figure 2 below is a screenshot keeping the core entities in the center and their associated recursive entities at corners. As indicated in figure 1 by little circles, each core entity recursively refers to itself. In this way, CERIF allows for relationships in between Projects (cfProj_Proj), Persons (cfPers_Pers), Organisations (cfOrgUnit_Org Unit), and Publications (cfResPubl_ResPubl).

The recursive entities belong to the group of link entities; the same holds for the interlinking entities in figure 3 below. Link entities will be introduced separately in section *2.4 CERIF Link Entities.*

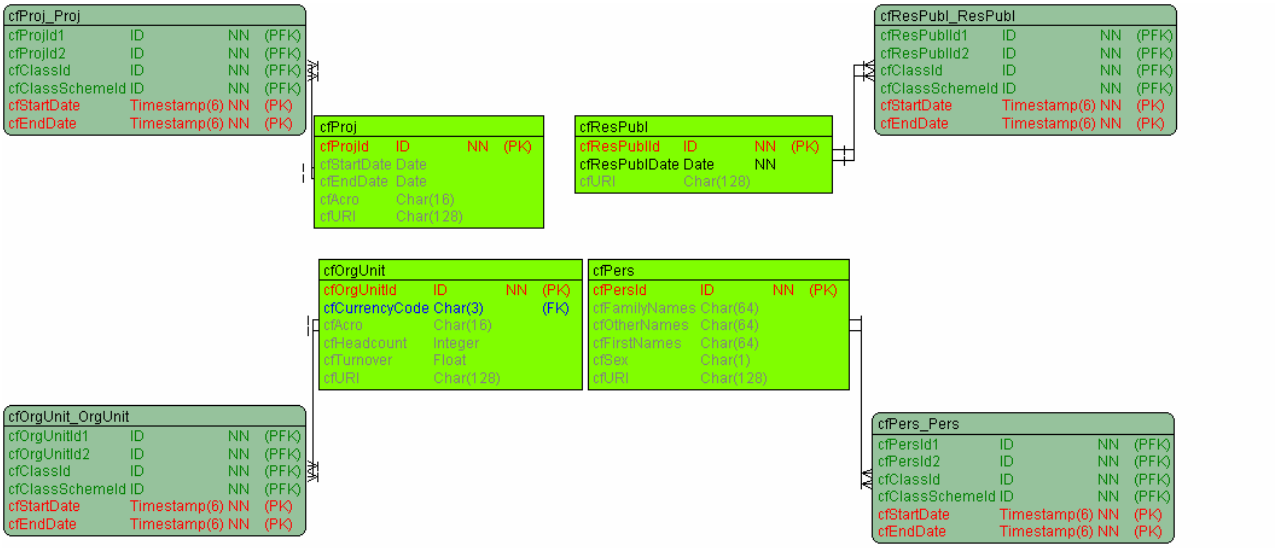

Figure 2: CERIF Core Entities and Recursive Entities in View of the Physical Level

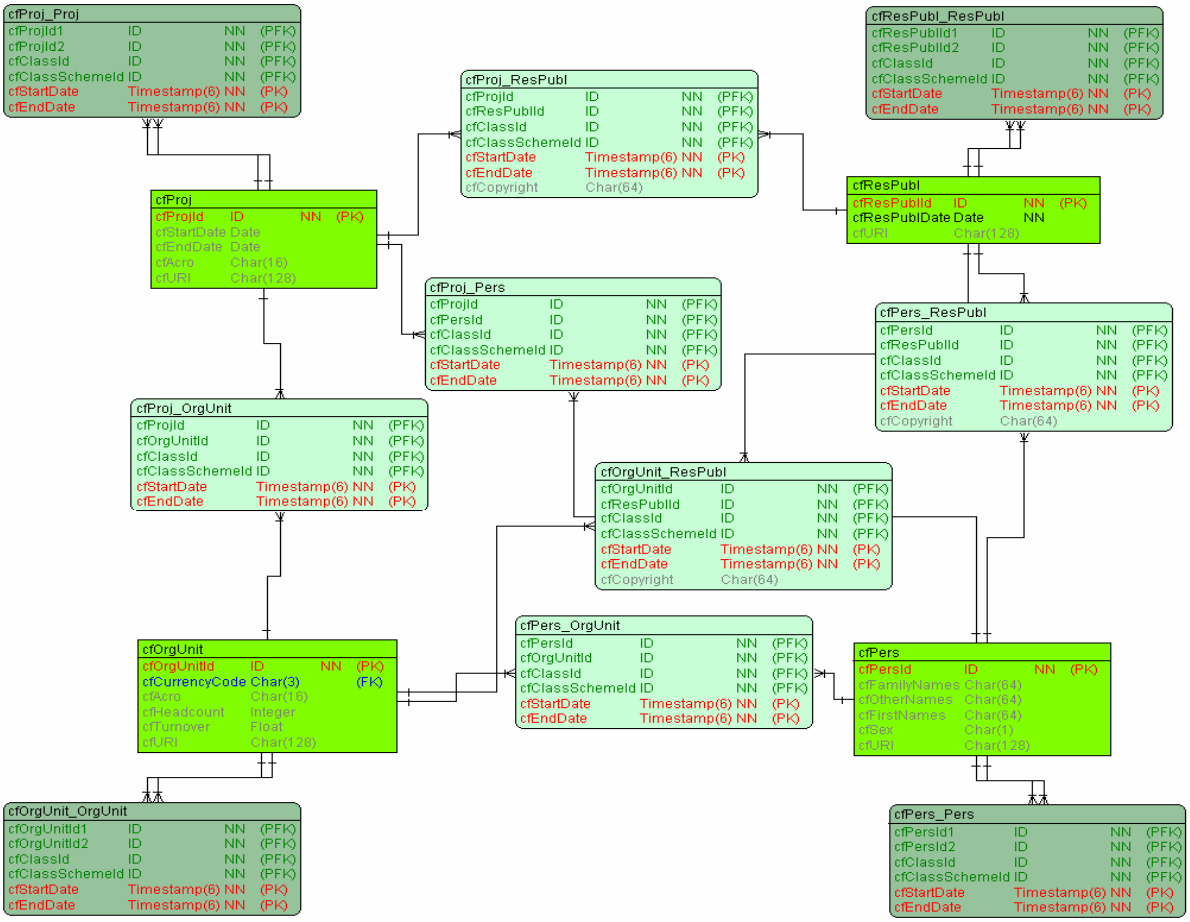

Figure 3: CERIF Core Entities and Linking Entities in View of the Physical Level

The core CERIF entities represent scientific actors (Persons and Organisations) and their main research activities (Projects and Publications): Scientists collaborate (cfPers_Pers), are involved in projects (cfProj_Pers), are affiliated with organisations (cfOrgUnit_Pers) and publish papers (cfPers_ResPubl). Projects involve people (cfProj_Pers) and organisations (cfProj_OrgUnit). Scientific publications are published by organisations (cfOrgUnit_ResPubl) and refer to projects (cfProj_ResPubl), publications involve people (cfPers_ResPubl), organisations store publications (cfResPubl), support or participate in projects (cfProj_OrgUnit), and employ people (cfPers_OrgUnit). To manage type and role definitions, references to classification schemes (cfClassId; cfClassSchemeId) are employed that will be explained separately in Section *2.6 CERIF Classification Entities.*

Each core entity contains an identifier attribute at first order; the primary key. Second there are basic attributes to represent the objects behind the entities, and finally a URI (UniformResourceIdentifier) allows for reference to (persistent) public resources. The primary key identifiers of core entities are propagated to their associated linking entities as indicated in figures 2, 3, 4. Where figure 2 shows the core entities and their associated recursive entities, figure 3 shows the relationships in between the core CERIF entities. Figure 4 demonstrates the means for representation of entity types for core CERIF entities by linkage to the Semantic Layer (cfClassId; cfClassSchemeId): projects (cfProj_Class), organisations (cfOrgUnit_Class), people (cfPers_Class), and publications (cfResPubl _Class).

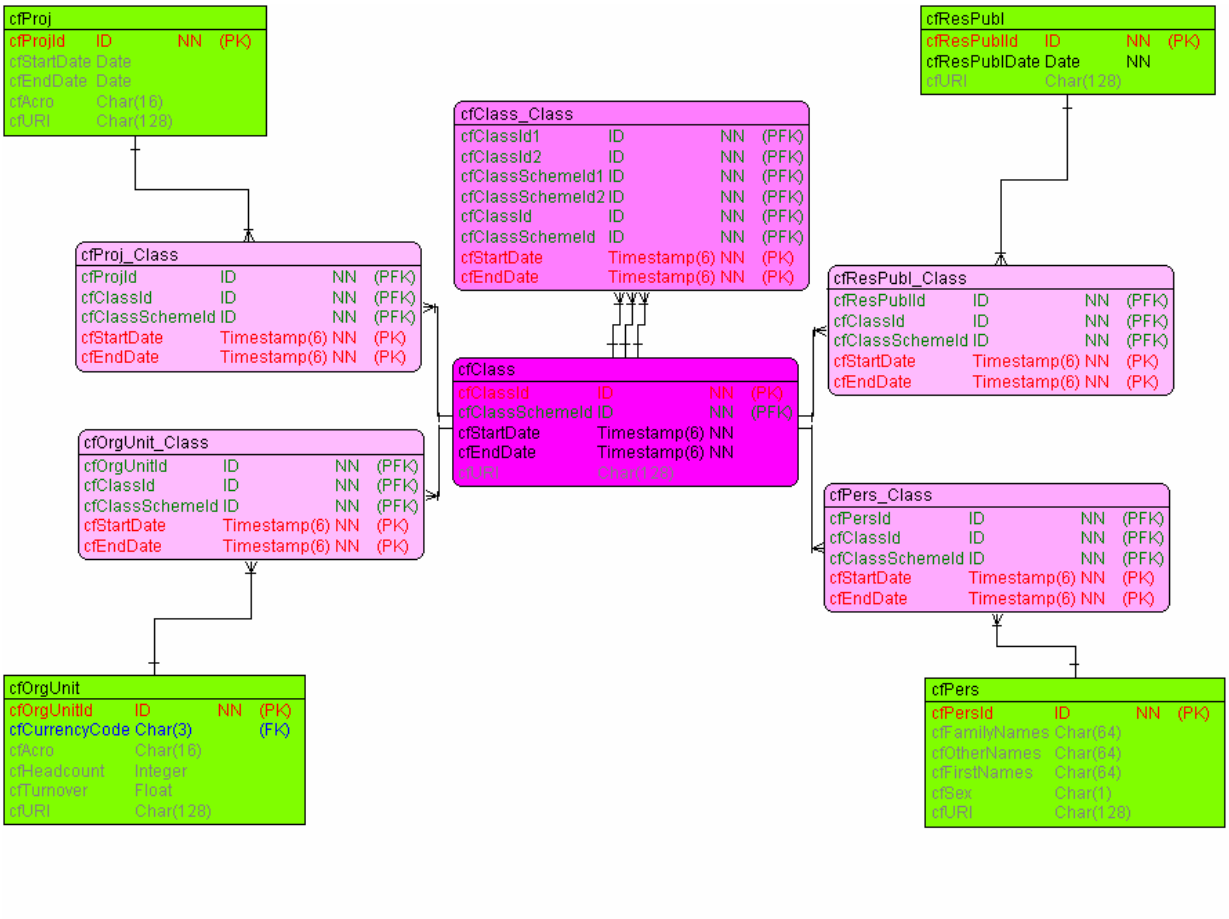

Figure 4: CERIF Core Entities and Classification Entities in View of the Physical Level

2.3 2nd Level CERIF Entities

Besides the core entities that represent players and activities, CERIF captures the context of players and their interaction in the wider range of a research environment by so called 2nd Level Entities. Figure 5 shows the core entities and some of the 2nd level entities and their linking relationships.

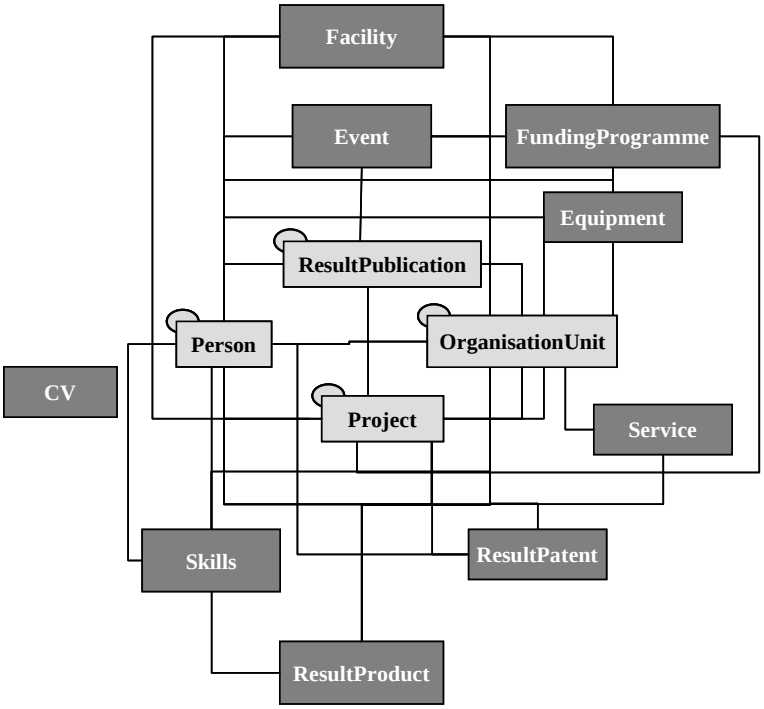

Figure 5: CERIF 2nd Level Entities in Abstract View

Figure 6 shows a screenshot of the 2nd level entities arranged around the core entities. Similar to core entities, each 2nd level entity contains an identifier attribute at first order; the primary key. Second are basic attributes to represent the objects behind the entities, and finally a Uniform Resource Identifier (cfURI) allows for reference to (persistent) public resources.

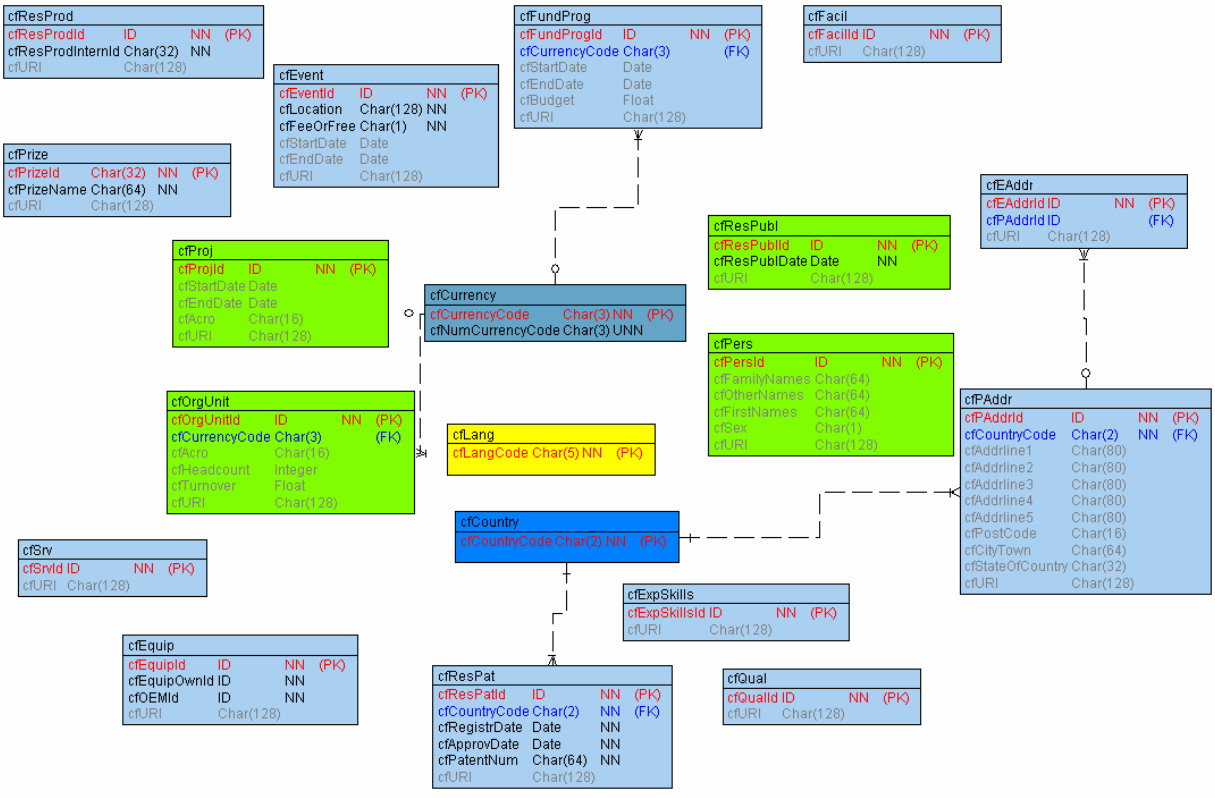

Figure 6: CERIF Core Entities and 2nd Level Entities in View of the Physical Level

Outside the core entities in the center of figure 6 are the 2nd level entities: Curriculum Vitaes (cfCV) play a major role in research environments; projects closely refer to funding programmes (cfFundProg); researchers attend events (cfEvent); use facilities (cfFacil) and services (cfSrv); their work depends on equipment (cfEquip); they apply for patents (cfResPat) and their work results in products (cfResProd). In figure 6 the linking entities have been omitted for a better visibility of the 2nd level entities. Some link entities associated with 2nd level entities are therefore presented separately in figure 7.

From figure 6 it is clear, that the entities cfCurrency, cfCountry and cfLang are explicitly kept in the model and have not been moved to the Semantic Layer. The three do not perfectly fit into the group of 2nd level entities, however are kept there for convenience. A transfer of the three entities will be further discussed within the CERIF task group because it implies further model changes. The structure of the three entities is different from other 2nd level entities. Instead of identifiers they apply codes for primary keys, as for currency, languages and country names there are public standard codes available which can be re-used [7, 8, 9]. Ecqual to core entities the 2nd level entities employ semantic relationships (cfClassId; cfClassSchemeId) as indicated in figure 7.

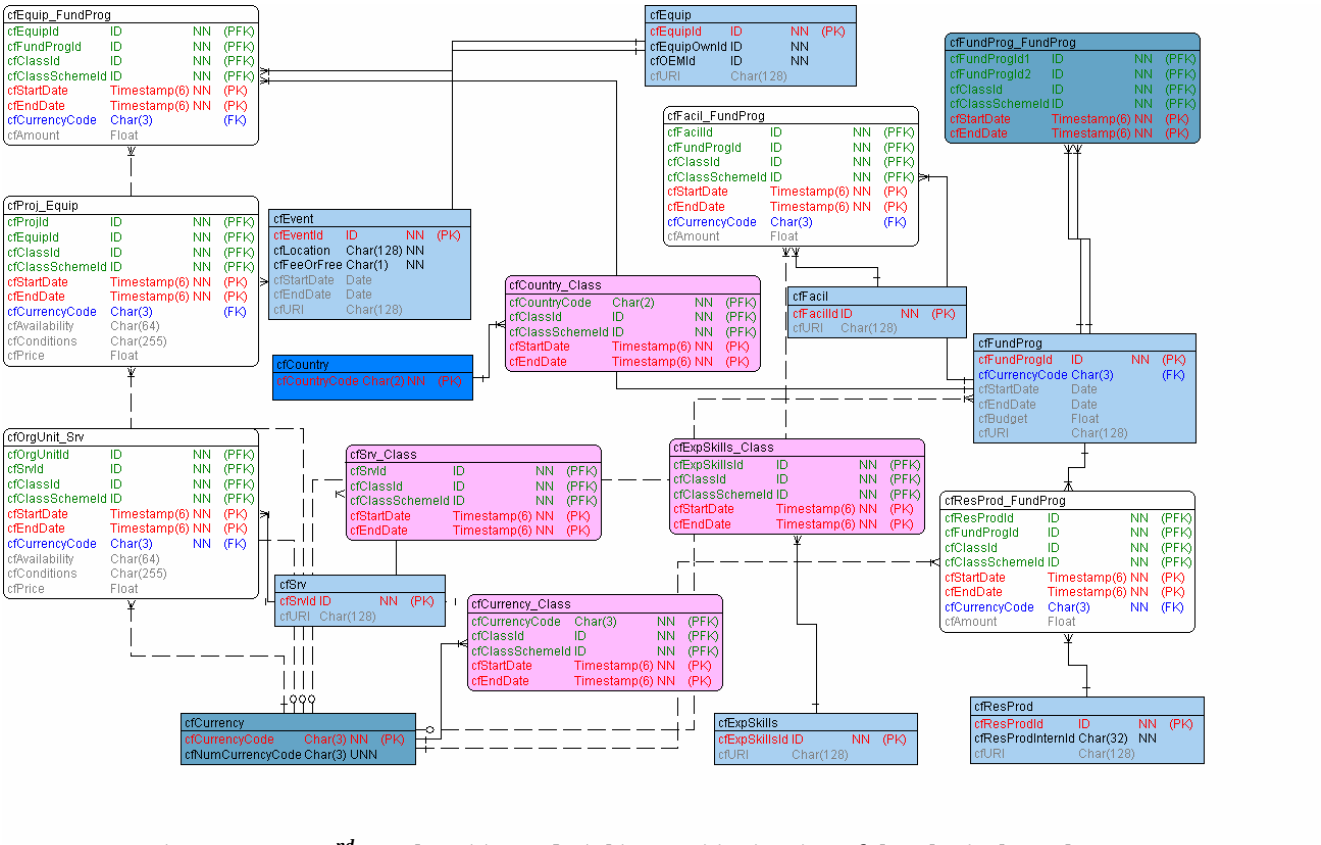

Figure 7: Some 2nd Level Entities and Linking Entities in View of the Physical Level

In figure 6 and 7 the entity cfCurrency propagates its primary key to the core entity cfOrgUnit and to the 2nd level entiy cfFundProg. The entity cfCountry propagates its identifier *cfCountryCode* to the entity cfResPat and cfPAddr. The entity cfLang in figure 6 is not associated with any 2nd level entity at this stage and will be explained separately in section *2.5 Language-dependent CERIF Entities*.

2.4 CERIF Link Entities

For relationships between core and 2nd level entities, CERIF uses so called Link Entities. In CERIF, a link entity in the current release always connects two, either core entities or 2nd level entities. Figure 8 shows the abstract view of the link entities that combine the core CERIF entities with black background color: cfPers_ResPubl, cfProj_Pers, Pers_OrgUnit, Proj_ResPubl, OrgUnit_ResPubl, Proj_OrgUnit.

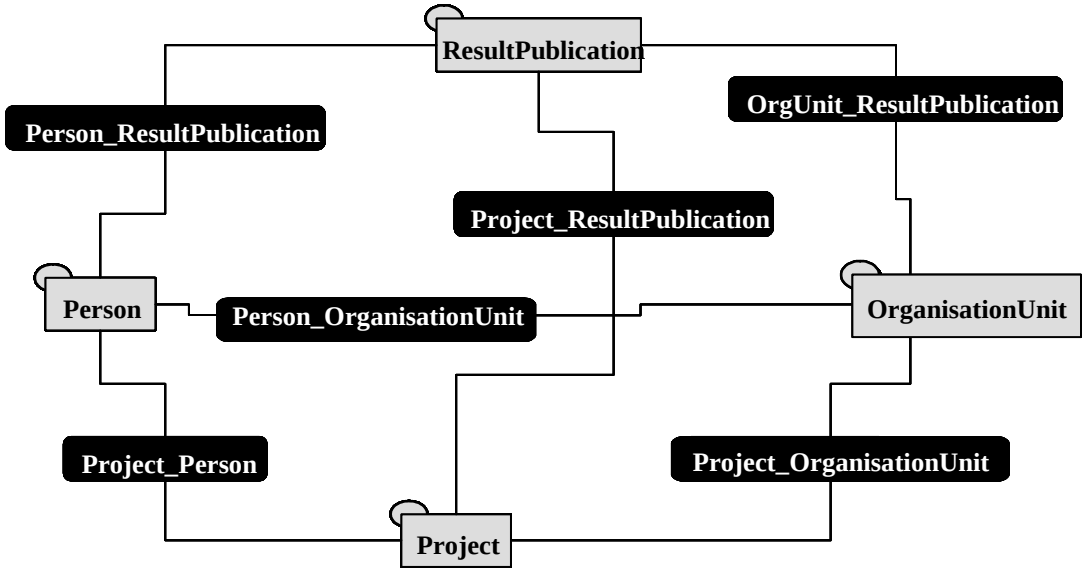

Figure 8: CERIF Link Entities connecting Core Entities in Abstract View

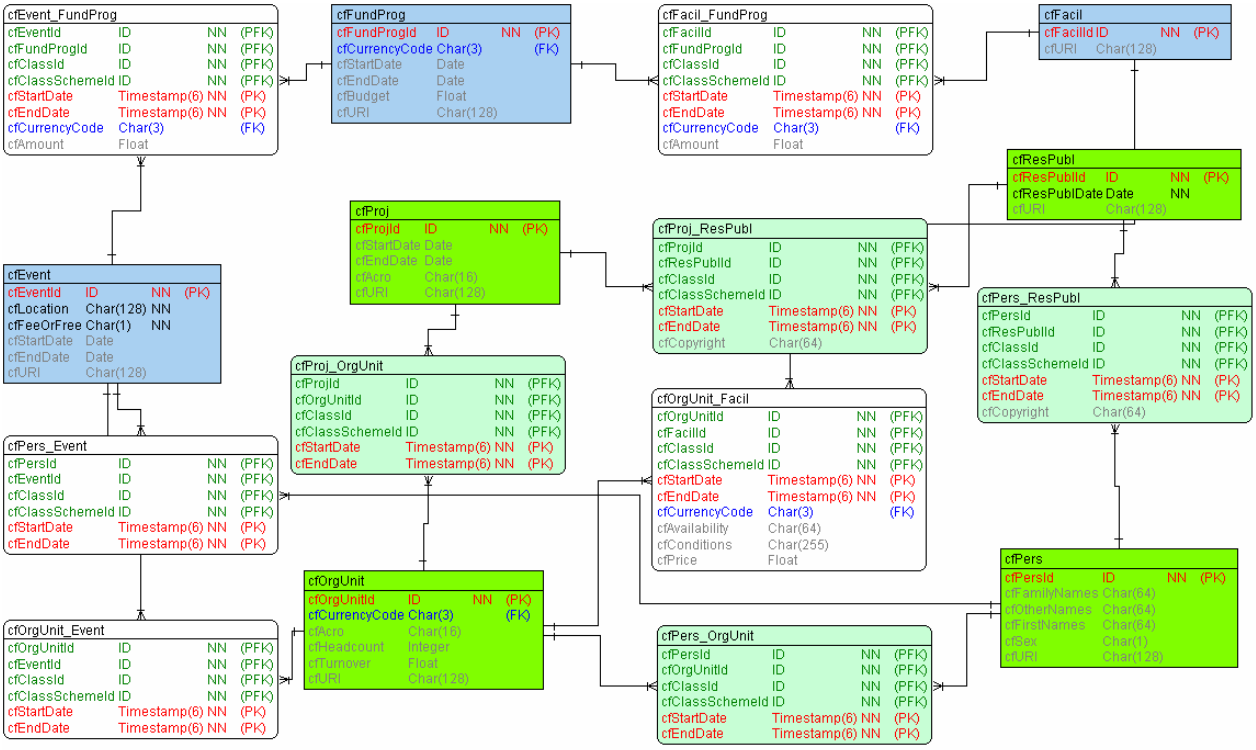

Figure 9: CERIF Link Entities in between Core Entities and some 2nd Level Entities in View of the Physical Level

Figure 9 shows some linking entities associated with core and 2nd level entities. The name of such composite entities mirrors the two entities involved and implies the reading direction for the two identifiers in the primary key. For example, the link entity cfPers_Facil inherits the identifiers (cfPersId from cfPers; cfFacilId from cfFacil). The identifiers are applied in this explicit order followed by semantic identifiers (cfClassId; cfClassSchemeId) and timestamps. Identifiers, semantic references and timestamp together build the primary key of link entities like in figure 10.

cfEntity1Name_Entity2Name	
cfInheritedEntity1Identifier	**(PFK)**
cfInheritedEntity2Identifier	**(PFK)**
cfInheritedClassificationIdentifier	**(PFK)**
cfInheritedClassificationSchemeIdentifier	**(PFK)**
cfStartDate	**(PK)**
cfEndDate	**(PK)**

Figure 10: Composition of the Primary Key for Link Entities

The composition of the primary key is obvious from preceding screenshot figures that specify the primary keys of entities within brackets behind the attribute names like in figure 10. Following the primary key, at some entities there are additional attributes to further describe the relationships: cfCurrencyCode for those inheriting from cfFundProg; cfCopyright for those inheriting from cfResPubl as captured in figure 9.

2.5 Language-dependent CERIF Entities

Much information in research environments needs representation in more than one language. As indicated in figure 11, CERIF contains many language-dependent entities, such as Keywords, Abstract, ResearchInterest, Name, ResearchActivity, Title and more. The semantic layer also allows for multiple language representations.

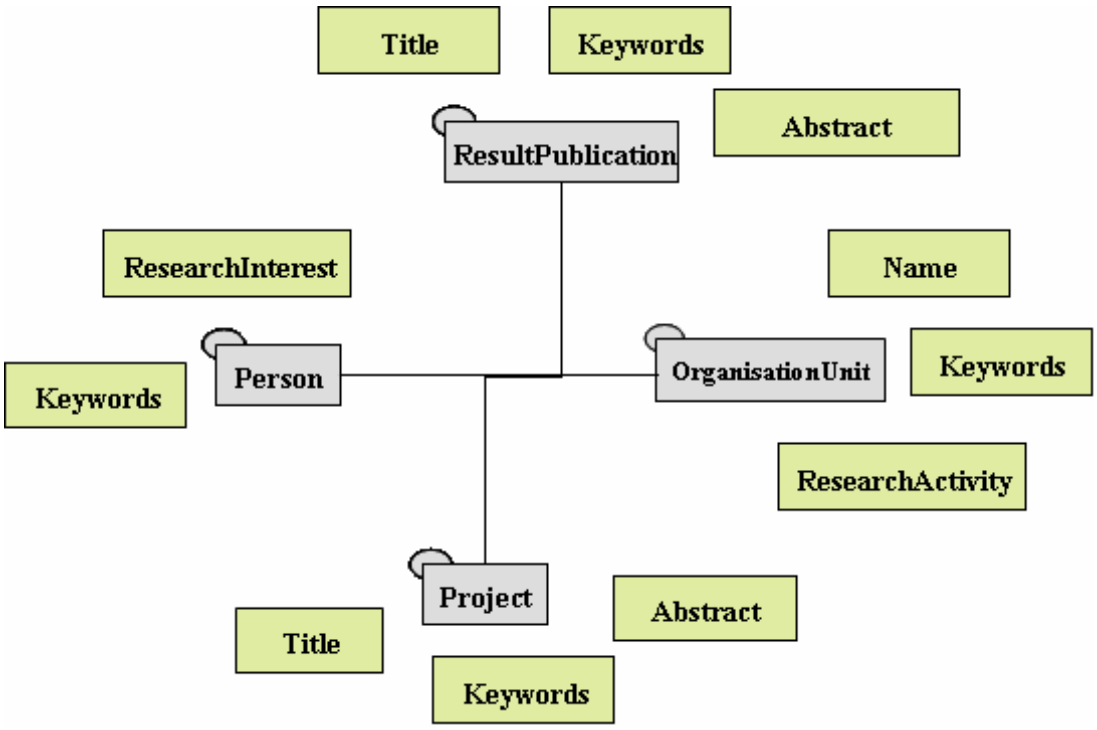

Figure 11: CERIF Core Entities supporting Multilingual Entries in Abstract View

Figure 12 shows that each language-dependent entity contains an identifier at first order which is inherited from its associated entity. Associated entities can either be core, 2nd level, or belong to the semantic layer. Second, each language-dependent entity requires a reference to language code (cfLangCode). Third, to the translation types (cfTranslation); both are part of their primary key. Subsequent to the primary key are the keyword, abstract or description attributes that capture the text in various languages (see figure 13).

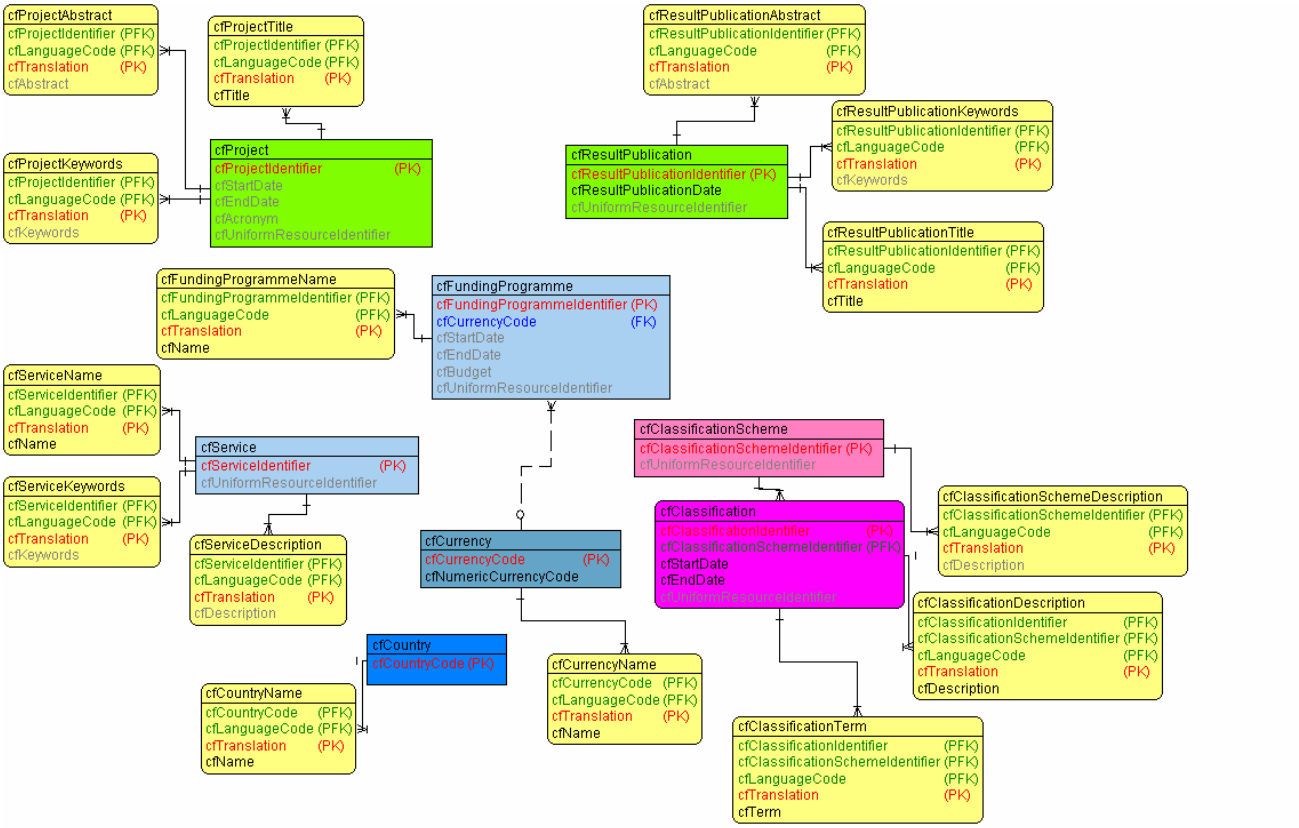

Figure 12: CERIF Language-dependent Entities supporting Multilingual Entries and some associated Core and 2nd Level Entities in Logical View

cfEntityName	
cfInheritedEntityIdentifier **cfLanguageCode** **cfTranslation** **cfKeyword; cfAbstract; cfDescription**	**(PFK)** **(PFK)** **(PK)**

Figure 13: Composition of the Primary Key for Language-dependent Entities

Besides the core and 2nd level entities, classification entities allow for language references (cfLanguageCode): cfClassificationDescription, cfClassificationSchemeDescription, cfClassificationTerm (see figure 12). It is thus possible to maintain classification schemes in multiple languages. This feature is an important issue in countries where several official languages are spoken. Country names for example can now be maintained in multiple languages: België (du), Belgien (ge), Belgique (fr), Belgium (en).

2.6 CERIF Classification Entities [Semantic Layer]

By normalization of the model with the current release, a flexible and consistent way for semantic relationahips has been introduced. It is therefore possible to include roles, types and other semantic references or predefined classifiers (cfClassId) from multiple classification schemes (cfClassSchemeId) or to define mappings between classifiers from different classification schemes. CERIF allows for a classification of all core and 2nd level entities according to pre-defined type schemes; link entities include references to the semantic layer (figure 14). Many examples of link entities with references to classification entities have been presented within this document.

This section will concentrate on the interior of the semantic layer. Examples will be provided in XML format.

We start by explaining basic role schemes for i.e. relationships between the core CERIF entities as indicated in figure 14 in abstract view. The black squares represent relationships as simplified link entities between core entities where the roles stand for classification references.

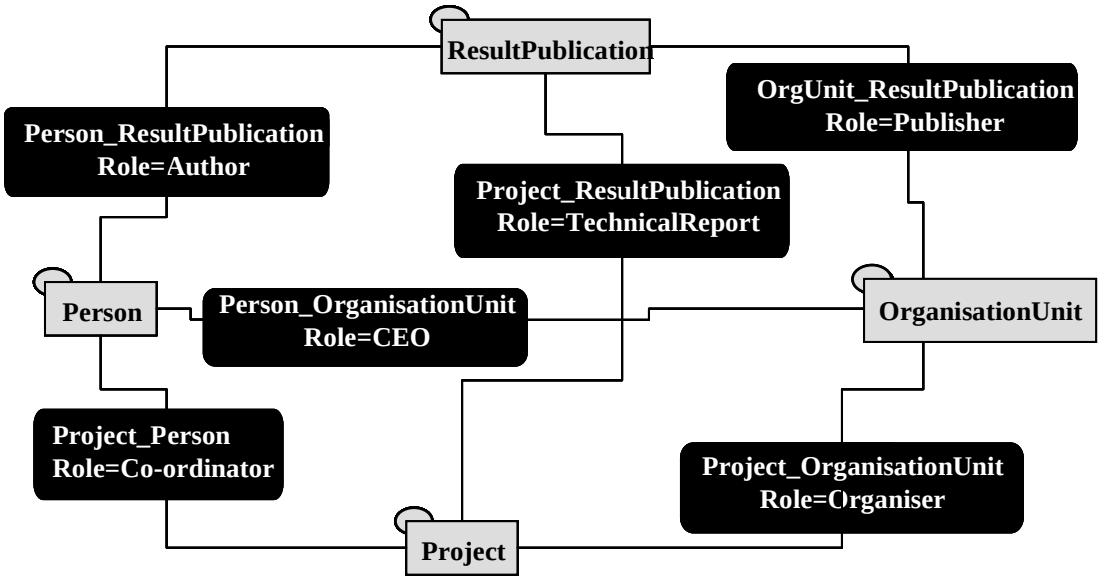

Figure 14: CERIF Core and Link Entities with Role Examples in Abstract View

The following examples show CERIFXML representations of some linking entity records with above roles. The roles corresponding to a role scheme are pre-defined and maintained within the semantic layer (figure 15).

```
<cfPers_ResPubl>
    <cfPersId>BrigitteJoerg</cfPersId>
    <cfResPublId>AnalyticInformationServicesForTheERA</cfResPublId>
    <cfClassId>isAuthor</cfClassId>
    <cfClassSchemeId>Pers_ResPublRolesDefaultScheme</cfClassSchemeId>
    <cfStartDate>2006-05-12 T00:00:00-00:00</cfStartDate>
    <cfEndDate>2099-12-31T00:00:00-00:00</cfEndDate>
</cfPers_ResPubl>
```

Example 1: Relationship between Person and Publication according to a Default Role Scheme

```
<cfPers_ResPubl>
    <cfPersId>BrigitteJoerg</cfPersId>
    <cfResPublId>JoergEtAl_2006</cfResPublId>
    <cfClassId>Author1</cfClassId>
    <cfClassSchemeId>Pers_ResPublRolesOrderScheme</cfClassSchemeId>
    <cfStartDate>2007-09-28T00:00:00-00:00</cfStartDate>
    <cfEndDate>2099-12-31T00:00:00-00:00</cfEndDate>
</cfPers_ResPubl>
```

Example 2: Relationship between Person and Publication according to a Role Scheme taking into account theOrder of Authors

```
<cfPers_OrgUnit>
    <cfPersId>BrigitteJoerg</cfPersId>
    <cfOrgUnitId>DFKI_GMBH</cfOrgUnitId>
    <cfClassId>isAffiliatedWith</cfClassId>
    <cfClassSchemeId>Pers_OrgUnitRolesDefaultScheme</cfClassSchemeId>
    <cfStartDate>2000-08-01T00:00:00-00:00</cfStartDate>
    <cfEndDate>2099-12-31T00:00:00-00:00</cfEndDate>
</cfPers_OrgUnit>
```

Example 3: Relationship between Person and Organisation according to a Default Role Scheme

The so called Semantic Layer consists of the entities displayed in figure 15. The entity classification (cfClass) introduces the identifier attribute *cfClassId* and inherits the *cfClassSchemeId* attribute from the classification scheme entity. Both attributes are consequently propagated to their associated entities: cfClassDescr, cfClassTerm, and Link Entities (cfClass_Class, cfClassScheme_ClassScheme and all other link entities not shown in figure 15). Where the entity classification manages the roles, types, terms, mappings, and other relationship definitions, the cfClassScheme entity assigns to those roles, terms, and mappings a particular classification scheme. It is therefore possible to capture the semantics for multiple application environments as a term, role and type semantics is consistently supported by schema references (Example 1, 2, 3).

Within the entity cfClass_Class a whole range of schema structures can be represented; be it simple **isA** relationships in taxonomies, vocabulary relationships like in thesauri **synonymOf**, **broaderTerm**, **narrower Term** or open relationships in ontologies. Some examples are given below.

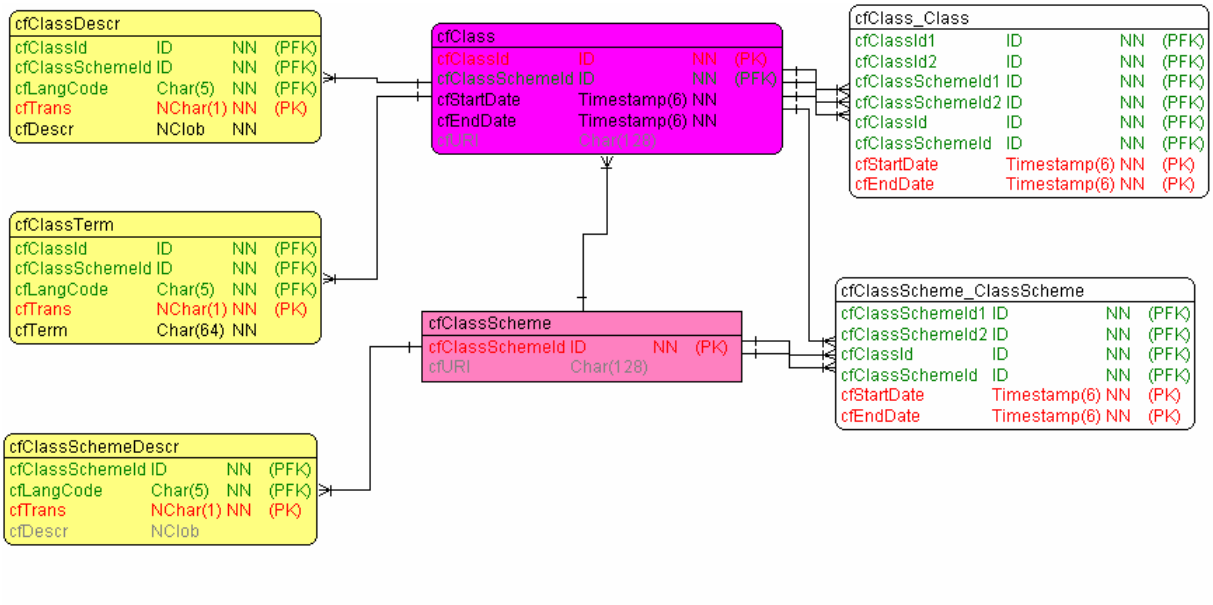

Figure 15: CERIF Classification Entities – Semantic Layer in View of thePhysical Level

```
<cfClass>
    <cfClassId>isA</cfClassId>
    <cfClassSchemeId>TaxonomicStructure</cfClassSchemeId>
    <cfStartDate>2007-09-28T00:00:00-00:00</cfStartDate>
    <cfEndDate>2099-12-31T00:00:00-00:00</cfEndDate>
</cfClass>

<cfClass_Class>
    <cfClassId1>InformationScience</cfClassId1>
    <cfClassId2>Science</cfClassId2>
    <cfClassSchemeId1>TaxonomyOfScience</cfClassSchemeId1>
    <cfClassSchemeId2>TaxonomyOfScience</cfClassSchemeId2>
    <cfClassId>isA</cfClassId>
    <cfClassSchemeId>TaxonomicStructure</cfClassSchemeId>
    <cfStartDate>2007-09-28T00:00:00-00:00</cfStartDate>
    <cfEndDate>2099-12-31T00:00:00-00:00</cfEndDate>
</cfClass_Class>
```

Example 4: Example of aTaxonomic isARelationship

In example 4, first the **isA** relationship of taxonomy is defned by cfClassId=isA; cfClassSchemeId=Taxonomic Structure. This definition is subsequently applied in the process of classifiying within the cfClass_Class entity. The concept “InformationScience” isA “Science” in the range of the classification scheme called “TaxonomyOf Science”. As a result we have a representation as depicted in figure 16.

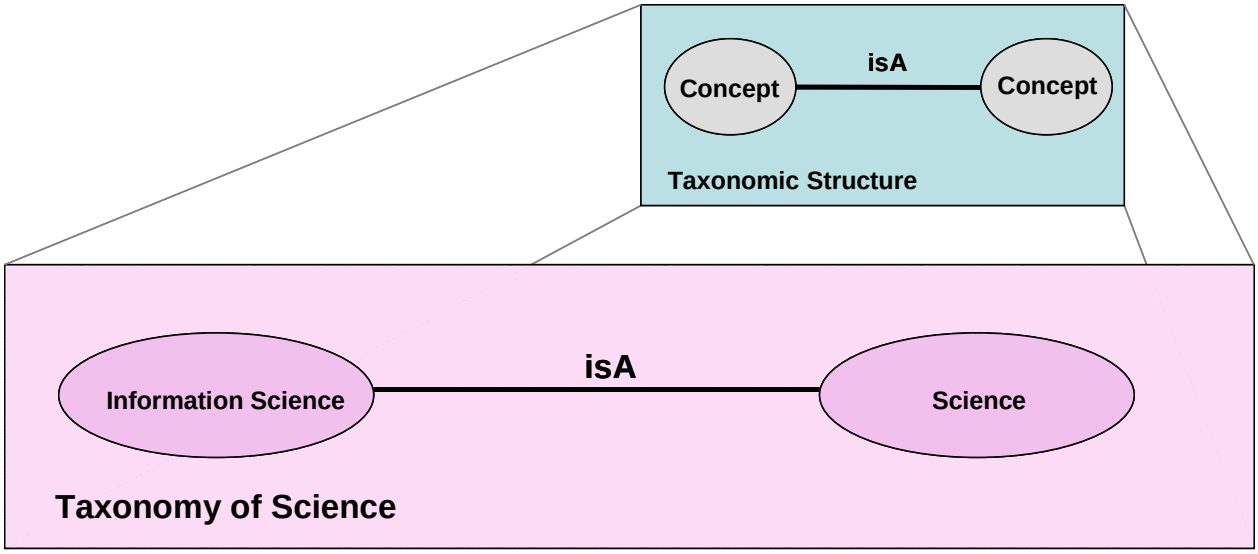

Figure 16: Explicit Graph representation of the Taxonomic isARelationship from Example 4

Figure 16 represents the graph of the statements from example 4. Beyond such taxonomic structures, thesauri structures or ontology structures can be represented in a similar way by assigning relationship labels like synonymOf, narrowerTerm, broaderTerm to a particular scheme as in example 5.

```
<cfClass>
    <cfClassId>synonymOf</cfClassId>
    <cfClassSchemeId>ThesaurusStructure</cfClassSchemeId>
    <cfStartDate>2007-09-28T00:00:00-00:00</cfStartDate>
    <cfEndDate>2099-12-31T00:00:00-00:00</cfEndDate>
</cfClass>
<cfClass>
    <cfClassId>narrowerTerm</cfClassId>
    <cfClassSchemeId>ThesaurusStructure</cfClassSchemeId>
    <cfStartDate>2007-09-28T00:00:00-00:00</cfStartDate>
    <cfEndDate>2099-12-31T00:00:00-00:00</cfEndDate>
</cfClass>

<cfClass_Class>
    <cfClassId1>Nature</cfClassId1>
    <cfClassId2>Character</cfClassId2>
    <cfClassSchemeId1>ThesaurusTerm</cfClassSchemeId1>
    <cfClassSchemeId2>ThesaurusTerm</cfClassSchemeId2>
    <cfClassId>synonymOf</cfClassId>
    <cfClassSchemeId>ThesaurusStructure</cfClassSchemeId>
    <cfStartDate>2007-09-28T00:00:00-00:00</cfStartDate>
    <cfEndDate>2099-12-31T00:00:00-00:00</cfEndDate>
</cfClass_Class>
```

Example 5: Example of Thesauri Relationships

Multiple classification terms and structures can be maintained in parallel and easily identified as semantically different due to their scheme assignments. Furthermore, it is possible to map terms across classification schemes like in example 6.

The cfClass_Class entity is a recursive entity and from a technical viewpoint it is treated like a link entity. Link Entities have been introduced and described in Section *2.4 CERIF Link Entities.* The structure of a link entity has been explained in figure 10. Each link entity inherits the two identifiers from the two entities involved. In this particular case, where the cfClass_Class entity is a recursive entity, the two identifiers could not be distinguished. Therefore, the first identifier is defined as cfClassId1, the second is defined as cfClassId2.

According to the structure of link entities as explained in figure 10, each link entity refers to the Semantic Layer with cfClassId and cfClassSchemeId; the same holds for the cfClass_Class entity. Given the two identifiers cfClassId1; cfClassId2 and their scheme references cfClassSchemeId1; cfClassScheme2 however reveals nothing about the kind of relationship between the two. Only another reference provides the type of relationship between the both as with example 6. Only the **mappesTo** assigned to a CERIF_DublinCore scheme supports an understanding of the semantics.

```
<cfClass>
    <cfClassId>mappesTo</cfClassId>
    <cfClassSchemeId>CERIF_DublinCoreMappings</cfClassSchemeId>
    <cfStartDate>2007-09-28T00:00:00-00:00</cfStartDate>
    <cfEndDate>2099-12-31T00:00:00-00:00</cfEndDate>
</cfClass>

<cfClass_Class>
    <cfClassId1>Author</cfClassId1>
    <cfClassId2>Creator</cfClassId2>
    <cfClassSchemeId1>cfPers_ResPublRolesDefaultScheme</cfClassSchemeId1>
    <cfClassSchemeId2>DublinCoreScheme</cfClassSchemeId2>
    <cfClassId>mappesTo</cfClassId>
    <cfClassSchemeId>CERIF_DublinCoreMappings</cfClassSchemeId>
    <cfStartDate>2007-09-28T00:00:00-00:00</cfStartDate>
    <cfEndDate>2099-12-31T00:00:00-00:00</cfEndDate>
</cfClass_Class>
```

Example 6: Example of a Mapping Relationship across Schemes

As explained in figure 10 as part of the introduction to link entities, each link entity finally contains timestamps. By default, we set timestamps for cfEndDate to the year 2099.

To allow for relationships between different classification schemes the entity cfClassScheme_ClassScheme is available (see figure 15 above). The entities cfClassDecr, cfClassTerm, and cfClassSchemeDescr belong to the group of language-dependent entities and allow for multiple languages.

The Semantic Layer of CERIF resulted from model improvements with the current release and reveals the major strength of the model. The content of the semantic layer has to be conceptually pre-defined so that semantic references from linking entities can be set via cfClassId and cfClassSchemeId in a subsequent step. For more about the content of the Semantic Layer see section *5 CERIF Semantic Layer Content.*

Each CERIF entity can be viewed as a member of one of the presented groups:

- CERIF Core Entities
- 2nd Level CERIF Entities
- CERIF Link Entities
- Language-dependent CERIF Entities
- CERIF Classification Entities

A list showing all CERIF entities is provided within the Appendix.

3. CERIF-based SQL scripts

From the ERM model in Toad Data Modeler the SQL scripts are generated automatically for most common databases. See some extract examples in figures 17, 18, 19, 20 [6].

```
Create table [cfPers] (
        [cfPersId] Nchar(32) NOT NULL,
        [cfFamilyNames] Nchar(64) NOT NULL,
        [cfFirstNames] Nchar(32) NULL,
        [cfOtherNames] Nchar(32) NULL,
        [cfSex] Nchar(1) Default u NOT NULL Check (f, m, u ),
        [cfURI] Nchar(128) NULL UNIQUE,
Primary Key  ([cfPersId])
)
```

Figure 17: SQL Extract for MS SQL7 database

```
Create table "cfPers" (
        "cfPersId" NChar(32) NOT NULL ,
        "cfFamilyNames" NChar(64) NOT NULL,
        "cfFirstNames" NChar(32),
        "cfOtherNames" NChar(32),
        "cfSex" NChar(1) Default u NOT NULL  Check (f, m, u ) ,
        "cfURI" NChar(128) NOT NULL  UNIQUE ,
primary key ("cfPersId")
)
```

Figure 18: SQL Extract for Oracle9i database

```
Create table "cfPers" (
        "cfPersId" Char(32) NOT NULL,
        "cfFamilyNames" Char(64) NOT NULL,
        "cfFirstNames" Char(32),
        "cfOtherNames" Char(32),
        "cfSex" Char(1) Default u NOT NULL Check (f, m, u ),
        "cfURI" Char(128) UNIQUE)
```

Figure 19: SQL Extract for DB2 UDB v.8

```
Create table `cfPers` (
        `cfPersId` Char(32) NOT NULL,
        `cfFamilyNames` Char(64) NOT NULL,
        `cfFirstNames` Char(32),
        `cfOtherNames` Char(32),
        `cfSex` Char(1) NOT NULL DEFAULT u,
        `cfURI` Char(128),
        UNIQUE (`cfURI`),
 Primary Key (`cfPersId`)) ENGINE = MyISAM/
```

Figure 20: SQL Extract for mySQL

4. CERIF XML

With the current release an XML format for data exchange has been provided. Related documents and schema files for validation can be viewed and downloaded from the public website: http://www.euroCRIS.org/. The XML specification corresponds to the physical level of the CERIF2006-1.1 FDM release and will be updated accordingly.

5. CERIF Semantic Layer Content

The structure and strength of the semantic layer of the CERIF model has been presented. In close cooperation with the CERIF Best Practice task group, some classification schemes will be prepared in CERIF XML format for default usage and import: http://eurocris.sharepointsite.com/default.aspx

5.1 XML Files from Best Practice Cases

- **Default Schemes for Type Definitions of Core and 2nd Level Entities**
  - Schemas for Project Types, Project Status, …
  - Schemas for Academic Titles, Honorific Titles, …
  - Schemas for Organisational Types
  - Schemas for Publication Types

- **Default Schemes for Role Definitions of Relationships between Core and 2nd Level Entities**
  - Roles, defining relationships between Projects and OrgUnits
  - Roles, defining relationships between OrgUnits
  - Roles, defining relationships between Person and OrgUnits
  - Roles, defining relationships between Project and Person
  - Roles, defining relationships between Project and OrgUnit
  - …

5.2 Supported Standard Codes

- (ISO639-1) ISO Two-Letter Language Codes
- (ISO3166-1) ISO Country Codes
- (ISO4217) ISO Currency Codes

6. CERIF Extensions

Contributions, thoughts, error reports or bug reports are very welcome. For submission we refer to the formal change procedure: http://www.eurocris.org/

Incoming feedback will first be discussed within the CERIF task group and subsequently presented to members. A decision towards extension will finally be taken and the CERIF model will be accordingly adapted.

7. Next Steps

In a next step, some content for the Semantic Layer will be prepared according to Best Practice feedback. We first aim to support default application scenarios and requirements and aim for more elaborated schemes in the long term.

More work on proper namespaces is being considered for the CERIF XML specifications in the longer term.

8. Appendix

8.1 List of CERIF Entities

Following is a full list of the CERIF entities in alphabetic order, grouped by entity type, giving the Logical and Physical Name of entities in brackets.

8.1.1 Core CERIF Entities (Logical (PhysicalName))

cfProject (cfProj)
cfPerson (cfPers)
cfOrgUnit (cfOrgUnit)
cfResultPublication (cfResPubl)

8.1.2 2nd Level CERIF Entities (Logical (PhysicalName))

cfCountry (cfCountry)
cfCurrency (cfCurrency)
cfCurriculumVitae (cfCV)
cfElectronicAddress (cfEAddr)
cfEquipment (cfEquip)
cfEvent (cfEvent)
cfExpertiseAndSkills (cfExpSkills)
cfFacility (cfFacil)
cfFundingProgramme (cfFundProg)
cfLanguage (cfLanguage)
cfPostalAddress (cfPAddr)
cfPrizeAward (cfPrize)
cfPublicationReference (cfPublRef)
cfQualification (cfQqual)
cfResultPatent (cfResPat)
cfResultProduct (cfResProd)
cfService (cfSrv)
cfDCAudience (cfDCAudience)
cfDCContributor (cfDCContributor)
cfDCCoverage (cfDCCoverage)
cfDCCoverageSpatial (cfDCCoverageSpatial)
cfDCCoverateTemporal (cfDCCoverageTemporal)
cfDCCreator (cfDCCreator)
cfDCDate (cfDCDate)
cfDCDescription (cfDCDescription)
cfDCFormat (cfDCFormat)
cfDCLanguage (cfDCLanguage)
cfDCProvenance (cfDCProvenance)
cfDCPublisher (cfDCPublisher)
cfDCRelation (cfDCRelation)
cfDCResourceIdentifier (cfDCResourceIdentifier)
cfDCResourceType (cfDCResourceType)
cfDCRightsHolder (cfDCRighsHolder)
cfDCRightsManagement (cfDCRightsMM)
cfDCRightsManagementAccessRights (cfDCRightsMMAccessRight)
cfDCRightsManagementLicense (cfDCRightsMMLicence)
cfDCSource (cfDCSource)
cfDCSubject (cfDCSubject)
cfDCTitle (cfDCTitle)
cfDublinCore (cfDC)

8.1.3 CERIF Link Entities (Logical (PhysicalName))

cfClassification_Classification (cfClass_Class)
cfClassScheme_ClassScheme (cfClassScheme_ClassScheme)
cfCountry_Classification (cfCountry_Class)
cfCurrency_Classification (cfCurrency_Class)
cfCV_Classification (cfCV_Class)
cfElectronicAddress_Classification (cfEAddr_Class)
cfEquipment_Classification (cfEquip_Class)
cfEquipment_FundingProgramme (cfEquip_FundProg)
cfEvent_Classification (cfEvent_Class)
cfEvent_FundingProgramme (cfEvent_FundProg)
cfEvent_ResultPublication (cfEvent_ResPubl)
cfExpertiseAndSkills_Classification (cfExpSkills_Class)
cfFacility_Classification (cfFacil_Class)
cfFacility_FundingProgramme (cfFacil_FundProg)
cfFundingProgramme_Classification (cfFundProg_Class)
cfFundingProgramme_FundingProgramme (cfFundProg_FundProg)
cfLanguage_Classification (cfLanguage_Class)
cfOrganisationUnit_Classification (cfOrgUnit_Class)
cfOrganisationUnit_DublinCore (cfOrgUnit_DC)
cfOrganisationUnit_ElectronicAddress (cfOrgUnit_EAddr)
cfOrganisationUnit_Equipment (cfOrgUnit_Equip)
cfOrganisationUnit_Event (cfOrgUnit_Event)
cfOrganisationUnit_ExpertiseAndSkills (cfOrgUnit_ExpSkills)
cfOrganisationUnit_Facility (cfOrgUnit_Facil)
cfOrganisaitonUnit_FundingProgramme (cfOrgUnit_FundProg)
cfOrganisationUnit_OrgUnit (cfOrgUnit_OrgUnit)
cfOrganisationUnit_PostalAddress (cfOrgUnit_PAddr)
cfOrganisationUnit_PrizeAward (cfOrgUnit_Prize)
cfOrganisationUnit_ResultPatent (cfOrgUnit_ResPat)
cfOrganisationUnit_ResultProduct (cfOrgUnit_ResProd)
cfOrganisationUnit_ResultPublication (cfOrgUnit_ResPubl)
cfOrganisationUnit_Service (cfOrgUnit_Srv)
cfPerson_Classification (cfPers_Class)
cfPerson_CV (cfPers_CV)
cfPerson_DublinCore (cfPers_DC)
cfPerson_ElectronicAddress (cfPers_EAddr)
cfPerson_Equipment (cfPers_Equip)
cfPerson_Event (cfPers_Event)
cfPerson_ExpertiseAndSkills (cfPers_ExpSkills)
cfPerson_Facility (cfPers_Facil)
cfPerson_FundingProgramme (cfPers_FundProg)
cfPerson_Language (cfPers_Language)
cfPerson_Country (cfPers_Country)
cfPerson_OrganisationUnit (cfPers_OrgUnit)
cfPerson_Person (cfPers_Pers)
cfPerson_PostAddress (cfPers_PAddr)
cfPerson_PrizeAward (cfPers_Prize)
cfPerson_Qualification (cfPers_Qual)
cfPerson_ResultPatent (cfPers_ResPat)
cfPerson_ResultProduct (cfPers_ResProd)
cfPerson_ResultPublication (cfPers_ResPubl)
cfPerson_Service (cfPers_Srv)
cfPostAddress_Classification (cfPAddr_Class)
cfProject_Classification (cfProj_Class)
cfProject_DublinCore (cfProj_DC)
cfProject_Equipment (cfProj_Equip)

cfProject_Event (cfProj_Event)
cfProject_Facility (cfProj_Facil)
cfProject_FundingProgramme (cfProj_FundProg)
cfProject_OrganisationUnit (cfProj_Orgunit)
cfProject_Person (cfProj_Pers)
cfProject_PrizeAward (cfProj_Prize)
cfProject_Project (cfProj_Proj)
cfProject_Service (cfProj_Srv)
cfProject_ResultPatent (cfProj_ResPat)
cfProject_ResultProduct (cfProj_ResProd)
cfProject_ResultPublication (cfProj_ResPubl)
cfResultPatent_Classification (cfResPat_Class)
cfResultPatent_FundingProgramme (cfResPat_FundProg)
cfResultProduct_Classification (cfResProd_Class)
cfResultProduct_FundingProgramme (cfResProd_FundProg)
cfResultPublication_Classification (cfResPubl_Class)
cfResultPublication_DublinCore (cfResPubl_DC)
cfResultPublication_FundingProgramme (cfResPubl_FundProg)
cfResultPublication_ResultPublication (cfResPubl_ResPubl)
cfService_Classification (cfSrv_Class)

8.1.4 Language-dependent CERIF Entities (Logical (PhysicalName))

cfClassificationDescription (cfClassDescr)
cfClassificationTerm (cfClassTerm)
cfClassificationSchemeDescription (cfClassSchemeDescr)
cfCountryName (cfCountryName)
cfCurrencyEntityName (cfCurrencyEntityName)
cfCurrencyName (cfCurrencyName)
cfEquipmentDescription (cfEquipPDescr)
cfEquipmentKeywords (cfEquipKeyw)
cfEquipmentName (cfEquipName)
cfEventDescription (cfEventDescr)
cfEventKeywords (cfEventKeyw)
cfEventName (cfEventName)
cfExpertiseAndSkillsDescription (cfExpSkillsDescr)
cfExpertiseAndSkillsKeywords (cfExpSillsKeyw)
cfExpertiseAndSkillsName (cfExpSkillsName)
cfFacilityDescription (cfFacilDescr)
cfFacilityKeywords (cfFacilKeyw)
cfFacilityName (cfFacilName)
cfFundingProgrammeDescription (cfFundProgDescr)
cfFundingProgrammeKeywords (cfFundProgKeyw)
cfFundingProgrammeName (cfFundProgName)
cfLanguageName (cfLanguageName)
cfOrganisationUnitKeywords (cfOrgUnitKeyw)
cfOrganisationUnitName (cfOrgUnitName)
cfOrganisationUnitResearchActivity (cfOrgUnitResAct)
cfPersonResearchInterest (cfPersResInt)
cfPersonKeywords (cfPersKeyw)
cfProjectAbstract (cfProjAbstr)
cfProjectKeywords (cfProjKeyw)
cfProjectTitle (cfProjTitle)
cfResultPatentAbstract (cfResPatAbstr)
cfResultPatentKeywords (cfResPatKeyw)
cfResultPatentTitle (cfResPatTitle)
cfResultProductDescription (cfResProdDescr)
cfResultProductKeywords (cfResProdKeyw)

cfResultProductName (cfResProdName)
cfResultPublicationAbstract (cfResPublAbst)
cfResultPublicationKeywords (cfResPublKeyw)
cfResultPublicationTitle (cfResPublTitle)
cfServiceDescription (cfSrvDescr)
cfServiceKeywords (cfSrvKeyw)
cfServiceName (cfSrvName)

8.1.4.1 Language-dependent attributes including cflangCode and cfTrans

cfAbstr
cfDescr
cfKeyw
cfName
cfQualDescr
cfPublName
cfResAct
cfResInt
cfTerm
cfTitle

8.1.4.2 Currency-dependent attributes

cfBudget
cfAmount
cfPrice
cfTurnover

8.1.5 CERIF Classification Entities (Logical (PhysicalName))

cfClassification (cfClass)
cfClassificationScheme (cfClassScheme)

8.2 Logical / Physical CERIF Entity Names

The following table 1 gives an overview of all CERIF 2006 entities, their corresponding attributes with logical and physical names (including cf prefixes).

Table 1: List of Entities with Logical and Physical Names

Logical CERIF2006 Entities	Physical CERIF2006 Entities
cfClassification	cfClass
cfClassification_Classification	cfClass_Class
cfClassificationDescription	cfClassDescr
cfClassificationScheme	cfClassScheme
cfClassificationScheme_ClassificationScheme	cfClassScheme_ClassScheme
cfClassificationSchemeDescription	cfClassSchemeDescr
cfClassificationTerm	cfClassTerm
cfCountry	cfCountry
cfCountry_Classification	cfCountry_Class
cfCountryName	cfCountryName
cfCurrency	cfCurrency
cfCurrency_Classification	cfCurrency_Class
cfCurrencyEntityName	cfCurrencyEntName
cfCurrencyName	cfCurrencyName
cfCurriculumVitae	cfCV
cfCurriculumVitae_Classification	cfCV_Class
cfDublinCore	cfDC
cfDublinCoreAudience	cfDCAudience
cfDublinCoreContributor	cfDCContributor
cfDublinCoreCoverage	cfDCCoverage
cfDublinCoreCoverageSpatial	cfDCCoverageSpatial
cfDublinCoreCoverageTemporal	cfDCCoverageTemporal
cfDublinCoreCreator	cfDCCreator
cfDublinCoreDate	cfDCDate
cfDublinCoreDescription	cfDCDescription
cfDublinCoreFormat	cfDCFormat
cfDublinCoreLanguage	cfDCLanguage
cfDublinCoreProvenance	cfDCProvenance
cfDublinCorePublisher	cfDCPublisher
cfDublinCoreRelation	cfDCRelation
cfDublinCoreResourceIdentifier	cfDCResourceIdentifier
cfDublinCoreResourceType	cfDCResourceType
cfDublinCoreRightsHolder	cfDCRightsHolder
cfDublinCoreRightsManagement	cfDCRightsMM
cfDublinCoreRightsManagementAccessRights	cfDCRightsMMAccessRights
cfDublinCoreRightsManagementLicense	cfDCRightsMMLicense
cfDublinCoreSource	cfDCSource
cfDublinCoreSubject	cfDCSubject
cfDublinCoreTitle	cfDCTitle
cfElectronicAddress	cfEAddr
cfElectronicAddress_Classification	cfEAddr_Class
cfEquipment	cfEquip
cfEquipment_Classification	cfEquip_Class
cfEquipment_FundingProgramme	cfEquip_FundProg
cfEquipmentDescription	cfEquipDescr
cfEquipmentKeywords	cfEquipKeyw
cfEquipmentName	cfEquipName
cfEvent	cfEvent
cfEvent_Classification	cfEvent_Class

cfEvent_FundingProgramme	**cfEvent_FundProg**
cfEvent_ResultPublication	**cfEvent_ResPubl**
cfEventDescription	**cfEventDescr**
cfEventKeywords	**cfEventKeyw**
cfEventName	**cfEventName**
cfExpertiseAndSkills	**cfExpSkills**
cfExpertiseAndSkills_Classification	**cfExpSkills_Class**
cfExpertiseAndSkillsDescription	**cfExpSkillsDescr**
cfExpertiseAndSkillsKeywords	**cfExpSkillsKeyw**
cfExpertiseAndSkillsName	**cfExpSkillsName**
cfFacility	**cfFacil**
cfFacility_Classification	**cfFacil_Class**
cfFacility_FundingProgramme	**cfFacil_FundProg**
cfFacilityDescription	**cfFacilDescr**
cfFacilityKewords	**cfFacilKeyw**
cfFacilityName	**cfFacilName**
cfFormalisedDublinCoreRightsManagementPricing	**cfFDCRightsMMPricing**
cfFormalisedDublinCoreRightsManagementPrivacy	**cfFDCRightsMMPrivacy**
cfFormalisedDublinCoreRightsManagementRights	**cfFDCRightsMMRights**
cfFormalisedDublinCoreRightsManagementSecurity	**cfFDCRightsMMSecurity**
cfFundingProgramme	**cfFundProg**
cfFundingProgramme_Classification	**cfFundProg_Class**
cfFundingProgramme_FundingProgramme	**cfFundProg_FundProg**
cfFundingProgrammeDescription	**cfFundProgDescr**
cfFundingProgrammeKeywords	**cfFundProgKeyw**
cfFundingProgrammeName	**cfFundProgName**
cfLanguage	**cfLang**
cfLanguage_Classification	**cfLang_Class**
cfLanguageName	**cfLangName**
cfOrganisationUnit	**cfOrgUnit**
cfOrganisationUnit_Classification	**cfOrgUnit_Class**
cfOrganisationUnit_DublinCore	**cfOrgUnit_DC**
cfOrganisationUnit_ElectronicAddress	**cfOrgUnit_EAddr**
cfOrganisationUnit_Equipment	**cfOrgUnit_Equip**
cfOrganisationUnit_Event	**cfOrgUnit_Event**
cfOrganisationUnit_ExpertiseAndSkills	**cfOrgUnit_ExpSkills**
cfOrganisationUnit_Facility	**cfOrgUnit_Facil**
cfOrganisationUnit_FundingProgramme	**cfOrgUnit_FundProg**
cfOrganisationUnit_OrganisationUnit	**cfOrgUnit_OrgUnit**
cfOrganisationUnit_PostAddress	**cfOrgUnit_PAddr**
cfOrganisationUnit_PrizeAward	**cfOrgUnit_Prize**
cfOrganisationUnit_ResultPatent	**cfOrgUnit_ResPat**
cfOrganisationUnit_ResultProduct	**cfOrgUnit_ResProd**
cfOrganisationUnit_ResultPublication	**cfOrgUnit_ResPubl**
cfOrganisationUnit_Service	**cfOrgUnit_Srv**
cfOrganisationUnitKeywords	**cfOrgUnitKeyw**
cfOrganisationUnitName	**cfOrgUnitName**
cfOrganisationUnitResearchActivity	**cfOrgUnitResAct**
cfPerson	**cfPers**
cfPerson_Classification	**cfPers_Class**
cfPerson_Country	**cfPers_Country**
cfPerson_CurriculumVitae	**cfPers_CV**
cfPerson_DublinCore	**cfPers_DC**
cfPerson_ElectronicAddress	**cfPers_EAddr**
cfPerson_Equipment	**cfPers_Equip**
cfPerson_Event	**cfPers_Event**
cfPerson_ExpertiseAndSkills	**cfPers_ExpSkills**
cfPerson_Facility	**cfPers_Facil**

cfPerson_FundingProgramme	**cfPers_FundProg**
cfPerson_Language	**cfPers_Language**
cfPerson_OrganisationUnit	**cfPers_OrgUnit**
cfPerson_Person	**cfPers_Pers**
cfPerson_PostAddress	**cfPers_PAddr**
cfPerson_PrizeAward	**cfPers_Prize**
cfPerson_Qualification	**cfPers_Qual**
cfPerson_ResultPatent	**cfPers_ResPat**
cfPerson_ResultProduct	**cfPers_ResProd**
cfPerson_ResultPublication	**cfPers_ResPubl**
cfPerson_Service	**cfPers_Serv**
cfPersonKeywords	**cfPersKeyw**
cfPersonResearchInterest	**cfPersResInt**
cfPostAddress	**cfPAddr**
cfPostAddress_Classification	**cfPAddr_Class**
cfPrizeAward	**cfPrize**
cfPrizeAward_Classification	**cfPrize_Class**
cfProject	**cfProj**
cfProject_Classification	**cfProj_Class**
cfProject_DublinCore	**cfProj_DC**
cfProject_Equipment	**cfProj_Equip**
cfProject_Event	**cfProj_Event**
cfProject_Facility	**cfProj_Facil**
cfProject_FundingProgramme	**cfProj_FundProg**
cfProject_OrganisationUnit	**cfProj_OrgUnit**
cfProject_Person	**cfProj_Pers**
cfProject_PrizeAward	**cfProj_Prize**
cfProject_Project	**cfProj_Proj**
cfProject_ResultPatent	**cfProj_ResPat**
cfProject_ResultProduct	**cfProj_ResProd**
cfProject_ResultPublication	**cfProj_ResPubl**
cfProject_Service	**cfProj_Srv**
cfProjectAbstract	**cfProjAbstr**
cfProjectKeywords	**cfProjKeyw**
cfProjectTitle	**cfProjTitle**
cfPublicationReference	**cfPublRef**
cfQualification	**cfQual**
cfQualification_Classification	**cfQual_Class**
cfQualificationDescription	**cfQualDescr**
cfQualificationKeywords	**cfQualKeyw**
cfResultPatent	**cfResPat**
cfResultPatent_Classification	**cfResPat_Class**
cfResultPatent_FundingProgramme	**cfResPat_FundProg**
cfResultPatentAbstract	**cfResPatAbstr**
cfResultPatentKeywords	**cfResPatKeyw**
cfResultPatentTitle	**cfResPatTitle**
cfResultProduct	**cfResProd**
cfResultProduct_Classification	**cfResProd_Class**
cfResultProduct_FundingProgramme	**cfResProd_FundProg**
cfResultProductDescription	**cfResProdDescr**
cfResultProductKeywords	**cfResProdKeyw**
cfResultProductName	**cfResProdName**
cfResultPublication	**cfResPubl**
cfResultPublication_Classification	**cfResPubl_Class**
cfResultPublication_FundingProgramme	**cfResPubl_FundProg**
cfResultPublication_ResultPublication	**cfResPubl_ResPubl**
cfResultPublicationAbstract	**cfResPublAbstr**
cfResultPublicationKeywords	**cfResPublKeyw**

cfResultPublicationTitle	**cfResPublTitle**
cfService	**cfSrv**
cfService_Classification	**cfSrv_Class**
cfServiceDescription	**cfSrvDescr cfSrvKeyw**
cfServiceKeywords	**cfSrvName**
cfServiceName	

9. References

[1] CERIF2006XML-1.1: http://www.eurocris.org/

[2] CERIF2006-1.1 Full Data Model (FDM): http://www.eurocris.org/

[3] CERIF2006-1.1 Release in Toad Format: http://www.eurocris.org/

[4] CERIF2006-1.1 Full Data Model (FDM) – Physical Level: http://www.eurocris.org/

[5] W3C XML Schema: http://www.w3.org/XML/Schema

[6] CERIF2006-1.1 SQL Scripts for common databases: http://eurocris.sharepointsite.com/default.aspx

[7] ISO Country Codes: http://www.iso.org/iso/country_codes.htm

[8] ISO Language Codes: http://www.iso.org/iso/language_codes.htm

[9] ISO Currency Codes: http://www.iso.org/iso/currency_codes.htm