SEVENTH FRAMEWORK PROGRAMME

CAPACITIES

Research Infrastructures

INFRA-2009-1 Research Infrastructures

OPENAIRE

Grant Agreement 246686

"Open Access Infrastructure for Research in Europe"

OpenAIRE Data Model Specification

Deliverable Code: D5.1

Document Description

Project	
Title:	OPENAIRE, Open Access Infrastructure for Research in Europe
Start date:	1st December 2009
Call/Instrument:	INFRA-2007-1.2.1
Grant Agreement:	**246686**

Document	
Deliverable number:	D5.1
Deliverable title:	OpenAIRE Data Model Specification
Contractual Date of Delivery:	31st of March 2010
Actual Date of Delivery:	15th of May 2010
Editor(s):	Paolo Manghi
Author(s):	Paolo Manghi
Reviewer(s):	Michele Artini, Magchiel Bijsterbosch, Wolfram Horstmann, Samuele Kaplun, Natalia Manola, Salvatore Mele, Jochen Schirrwagen, Tim Smith
Participant(s):	
Workpackage:	WP5
Workpackage title:	OpenAIRE System Back-end: storage and mediation services
Workpackage leader:	CNR
Workpackage participants:	NKUA, ICM, CERN, UNIBI
Distribution:	Public
Nature:	Deliverable
Version/Revision:	v 23
Draft/Final:	Final
Total number of pages: (including cover)	
Web Resource name:	

Key words:

Disclaimer

This document contains description of the OPENAIRE project findings, work and products. Certain parts of it might be under partner Intellectual Property Right (IPR) rules so, prior to using its content please contact the consortium head for approval.

In case you believe that this document harms in any way IPR held by you as a person or as a representative of an entity, please do notify us immediately.

The authors of this document have taken any available measure in order for its content to be accurate, consistent and lawful. However, neither the project consortium as a whole nor the individual partners that implicitly or explicitly participated in the creation and publication of this document hold any sort of responsibility that might occur as a result of using its content.

This publication has been produced with the assistance of the European Union. The content of this publication is the sole responsibility of the OPENAIRE consortium and can in no way be taken to reflect the views of the European Union.

The European Union is established in accordance with the Treaty on European Union (Maastricht). There are currently 27 Member States of the Union. It is based on the European Communities and the member states cooperation in the fields of Common Foreign and Security Policy and Justice and Home Affairs. The five main institutions of the European Union are the European Parliament, the Council of Ministers, the European Commission, the Court of Justice and the Court of Auditors. (http://europa.eu.int/)

OPENAIRE is a project funded by the European Union

Table of Contents

Table of Figures

Summary

The OpenAIRE web site will offer functionalities for administrators, anonymous and registered users to manage an Information Space of FP7-funded open access publications. The aim of this document is to describe the conceived structure and semantics of this Information Space, i.e., the *Open AIRE data model*, by providing an abstract definition of its main entities and the relationships between them.

In this definitional process, the intended interaction (Task 7.2) between the OpenAIRE Information Space and the EC Participant Portal software systems (e.g., CORDIS, NEF, SESAM) plays an important role in the specification of project data, i.e., how project data should be described, stored and exported in OpenAIRE. At this stage of interaction, we shall consider as authoritative sources for project data, the SESAM system and the CORDIS web site, which contain information on publications, reports, projects and participants in FP7. As such, we shall store project information adopting a schema inspired by SESAM and CORDIS. Furthermore, with regard to data export, the results of projects such as CERIF and KE-CRIS-OAR are taken into account as inspiring guidelines.

Finally, the document will analyze the impact of the OpenAIRE data model on the repository managers willing to expose, through OAI-PMH interfaces, metadata conforming to the OpenAIRE Information Space, so as to avoid double ingestion work to their researchers.

1 Scenario

The OpenAIRE web site will offer functionalities for administrators, anonymous and registered users to manage an Information Space of FP7-funded open access publications. In particular:

- *anonymous users* will be able to search and consult the Space;
- *registered users* will be able to insert and update content in the Space;
- *administrators* will have full access and rights to such Space and in particular are in charge of validating/invalidating insertions/deletions/updates from registered users.

In our reasoning we generalize the concept of publication to that of *project result*, so as to be able of including further kinds of research outputs whose production is "economically supported" by a EC project. Examples of project results may range from traditional publications to experimental data, software products, patents, books, ORE aggregation (i.e., compound object web representations). Similarly, we extend the notion of author to that of *person*. Other typologies of person may be required in the future, so as to describe future scenarios of interest.

Currently, OpenAIRE registered users can ingest information relative to *project results of type publication*, *persons*, *organizations* (companies, research centers or institutions) in order to link them together with so-called *authorships.* An authorship represents the fact that a given *person* has (co-)authored a given *publication project result* while being affiliated with a given *organization*. Besides, project results are always associated to one or more *instances* of the results, in the sense that different "physical representations" of the same result may exist. For example, the same publication may be kept in two different repositories, both exposing the payload file (e.g., PDF) at different internet locations (URLs). Accordingly, an instance of a project result is represented as a combination of zero or more *web resources* relative to the project result and of the internet *data source* (e.g., OAI-PMH accessible repositories, FTP sites, web sites) from which such resources are accessible/available. **Note that**: the purpose of the project result-instance-web resource model is to capture a list of internet pointers relevant to the project result and not that of capturing the compound object structure that some results may have. If a project result is a compound object (e.g. ORE aggregation), its instances will likely be associated to web resources embodying its compound object nature (e.g., ORE resource maps).

Of crucial interest to the OpenAIRE Information Space is also the identification of the European *projects* which co-funded the research that has led to a given publication. In particular, data describing European *projects* will be fetched from authoritative EC databases (currently SESAM but others are being contacted, such as CORDIS and NEF), together with the *organizations* or *persons* (ERC projects) which are *participants* of projects. While project data will authoritatively originate in EC databases, information about organizations may be also ingested by users, for example to complete authorships information in the database.

Project data models Focus studies on the results of CERIF project and the KE-CRIS-OAR initiative are being carried out in WP7. In particular, with regard to project, organization and publication data, it will be considered to adopt the description schemes proposed by such projects as models of the correspondent entities in the OpenAIRE data model. In this document, however, we shall proceed taking inspiration from the choices made by the EC databases on that respect, in particular by SESAM and CORDIS web site.

Outline In the following, Section 2 provides a detailed description of the entities that come into play by providing an Entity Relationship model and a relational database representation, Section 3 provides the requirements of input to the definition of the OpenAIRE Guidelines for Repository Managers (to be produced in WP3-T3.3), Section 4 describes how the data model could be extended so as to support collaborative refinement of the data.

2 Data definition

2.1 Detailed description

Project Result (or Result) A project Result is described by a result *kind* (e.g., publication, research data, software product), a *type*, which depends on the given kind (e.g., for publications we have "article", "book", "manual", etc.), a *date of creation*, a *description*, a *publisher*, a *language*, a list of *keywords*, an *access mode* (e.g., license to access the result) and an *embargo end-date* (empty if the access kind does not imply an embargo). A Result is associated with (*i*) the set of its Authorships, (*ii*) the set of its Instances, if any (which can be located at different Data Sources), and (*iii*) the Projects which co-funded the research underlying the Result.

Note: the property *access mode* can have one of the following values: "open access", "restricted" or "embargo". If it is "embargo", the field *embargo end-date* contains the date after which the publication becomes open access.

In SESAM publication results are described as follows:

I_ID	NOT NULL	NUMBER(10)
I_PROJECT_ID		NUMBER(10)
I_NUMBER		NUMBER(10)
ST_TITLE		VARCHAR2(100)
ST_AUTHOR		VARCHAR2(100)
ST_TITLE_PERIODIC		VARCHAR2(100)
ST_PERIODIC_NUMBER		VARCHAR2(100)
ST_PUBLISHER		VARCHAR2(100)
ST_PLACE_PUB		VARCHAR2(100)
ST_RELEVANT_PAGES		VARCHAR2(30)
ST_PERMANENT_ID		VARCHAR2(20)
ST_ACCESS_OPEN		VARCHAR2(1)
ST_STATUS		VARCHAR2(50)
I_VERSION		NUMBER(5)
DA_DATE_PUB	NOT NULL	DATE
I_EMBARGO_MONTHS		NUMBER(3)
DA_EMBARGO_END		DATE

Note that the attributes that were found relevant to a resource are a subset or can be derived from those used in SESAM. If required, in the future we can devise ways to import publication data from SESAM to OpenAIRE.

Organizations An Organization is described in the SESAM database as follows:

PROJECT_REF	NOT NULL	NUMBER(10)
LEGAL_SHORT_NAME		VARCHAR2(4000)
LEGAL_NAME		VARCHAR2(4000)
LEGAL_STATUS		VARCHAR2(4000)
CECID		NUMBER
SME		VARCHAR2(4000)
CTRY_CODE		VARCHAR2(20)
CTRY_DESCR		VARCHAR2(200)
THIRDCTRY		VARCHAR2(1)
CPF_PARTICIPANT_NO		NUMBER
COST_MODEL		VARCHAR2(20)
COST_MODEL_DESCR		VARCHAR2(200)
TOTAL_COST		VARCHAR2(4000)
TOTAL_FUNDING		VARCHAR2(4000)
FUNDING_MAX_NOE		VARCHAR2(4000)
TOTAL_COST18		VARCHAR2(4000)
TOTAL_FUNDING18		VARCHAR2(4000)
SEAR_MAL		VARCHAR2(4000)

SEAR_FEM	VARCHAR2(4000)
DOCT_MAL	VARCHAR2(4000)
DOCT_FEM	VARCHAR2(4000)

Of such descriptive fields, we consider in OpenAIRE the following: *legal short name*, *legal name*, *legal status*.[1] To these properties, we add further ones, relevant to search, statistics and visualization: an *URL of its web site*, an *URL of the logo*, a *country of origin*, and a *location* (longitude, latitude, and time zone). An Organization is associated with (*i*) the set of Data Sources, if any, of which the Organization is responsible for and (*ii*) if any, with the Persons authoring a Result while begin affiliated with such Organization. An Organization may be the Participant of one or more Projects.

Persons A Person is described by a *name*, a *surname*, a *nationality*, and a *gender*. A Person is associated to zero (if not an author of Results in OpenAIRE) or more Authorships and may be the Participant of one or more Projects.

Participants A Participant is either a Participant Person or a Participant Organization which is benefiting from the funding of a project. It is described by an *EC participant number*, provided the first time the participant is granted a project by the EC, and to be used in future participation to projects. A Participant is always associated to one or more Projects.

Authorships An Authorship describes the fact that one Person has authored one Result while being affiliated with one Organization, if available. As such, an Authorship is associated to the Person, to the Result and, possibly, to the Organization.

Instances An Instance represents the combination of the Web Resources associated with a Result and the Data Source where such Web Resources are stored. As such, an Instance is described by the *original unique identifier* (e.g. DOI if the Result is a publication) of the Result at the original Data Source. Furthermore, an Instance is associated with the Result, to the relative Data Source and, if any, to a list of Web Resources. Constraint: if the Result is of kind "publication", the Instance must be associated to at least one Web Resource, relative to the payloads of the publication.

Data Sources A Data Source is described by an *official name*, an *English name*, a *URL of its web site*, an *URL of the logo*, an email of the technical contact, a *typology of source* (e.g., OAI-PMH, OAI-ORE, FTP, Web Site), *location* (longitude, latitude and time zone), and an *access information package*, i.e., an XML text bearing the information needed to access the specific source: for example, an OAI-PMH URL, if source typology is OAI-PMH, or an FTP address with login and password, if source typology is FTP. A Data Source is associated with (*i*) the set of Instances relative to Results available to OpenAIRE that are located at the Data Source and (*ii*) with the set of Organizations responsible for the Data Source.

Web Resources A Web Resource is described by *its unique URL*. A Web Resource is associated to the Instance of the Result of which it contains relevant content.

Projects A Project is described in the SESAM database as follows:

ST_TEL	VARCHAR2(255)
ST_FAX	VARCHAR2(255)
ST_EMAIL	VARCHAR2(255)

[1] We are still waiting for a document from SESAM describing the meaning of all table attributes. So for example, we can only assume that CPF_participant is the unique identifier assigned by the EC to organizations.

```
ST_WEB_SITE                                          VARCHAR2(255)
DA_PERIOD_END                                        DATE
DA_PERIOD_START                                      DATE
ST_FUNDING_SCHEME                NOT NULL            VARCHAR2(255)
ST_GRANT_AGREEMENT_NUMBER                            VARCHAR2(255)
ST_PROJECT_ACRONYM                                   VARCHAR2(255)
ST_PROJECT_TITLE                                     VARCHAR2(255)
DA_PROJECT_START_DATE                                DATE
DA_PROJECT_END_DATE                                  DATE
ST_SCI_REP                                           VARCHAR2(255)
ST_COORDINATOR_ORGANISATION                          VARCHAR2(255)
ST_COORDINATOR_NAME                                  VARCHAR2(255)
```

Of such descriptive fields, we consider in OpenAIRE the following: *web site*, *EC project web site* (e.g., the project page at CORDIS), *grant_agreement_number*, *acronym*, *title*, *start_date* and *end_date*. To these, we add a *project call identifier* and a list of *Keywords*. A Project is associated with (*i*) the set of Participants participating to the Project, (*ii*) the set of Results whose research was co-funded by the Project, (*iii*) the Contract Types (e.g., Coordination and Supporting Action, I3) it conforms to, and (*iv*) the Subdivisions (e.g., FP7-ENERGY, FP7-INFRASTRUCTURES) under which it is funded.

EC Projects classification

The entities described below mirror the CORDIS classification scheme (see Figure 1), accessible at the address http://cordis.europa.eu/fp7/info-programmes_en.html.

A **Funding Programme** is an EC funding programme (e.g., FP7), here characterized by an *identifier*, a *name* and an *acronym* as provided by the EC. A Funding Programme is associated with one or more **Specific Programmes** (e.g., Capacities, Cooperation, Ideas and People, JRC), here characterized by an *identifier*, a *name* and an *acronym*. In turn, a Specific Programme is associated with one or more **Subdivisions** (e.g., research infrastructures), here characterized by an *identifier*, a *name* and an *acronym*. Finally, **Contract Types** (e.g., supporting and coordination actions, eContentPlus, I3, IP), characterized by an *identifier*, a *name* and an *acronym*, are associated to zero, one or more Projects.

- *Funding Programmes*: e.g., FP7
    - *Specific Programmes*: e.g., Capacities, Cooperation, Ideas, People, Euratom direct, Euratom indirect
        - *Subdivisions*: e.g., research infrastructures, Energy, Initial Training

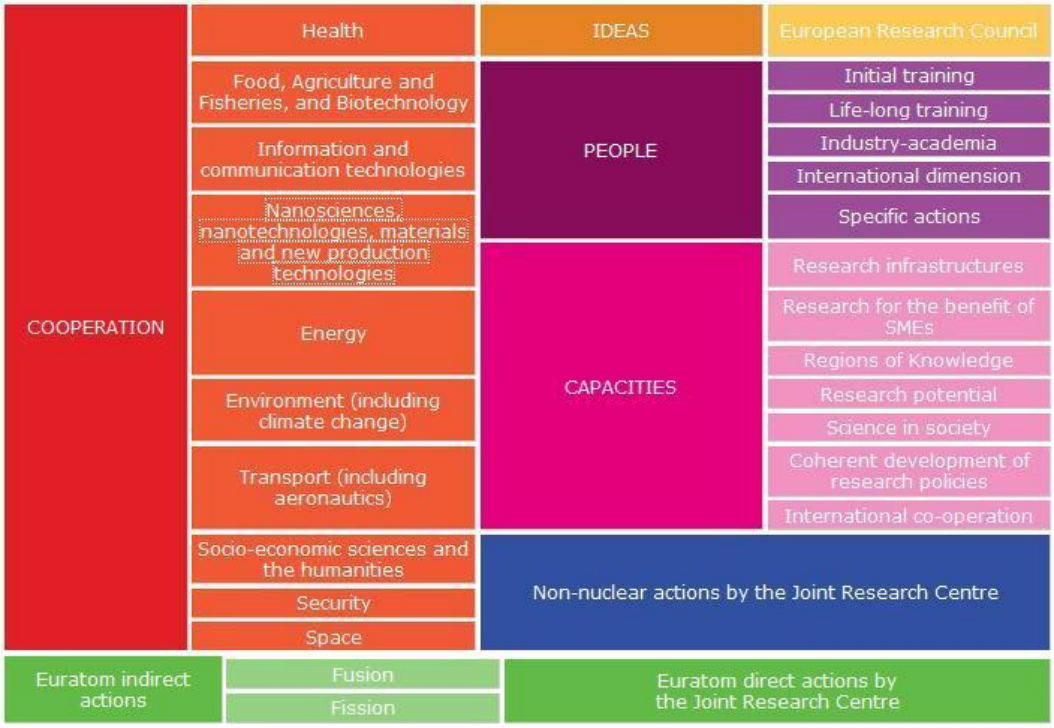

Figure 1 - CORDIS Classification Scheme: Specific programmes

2.2 Entity-Relationship model

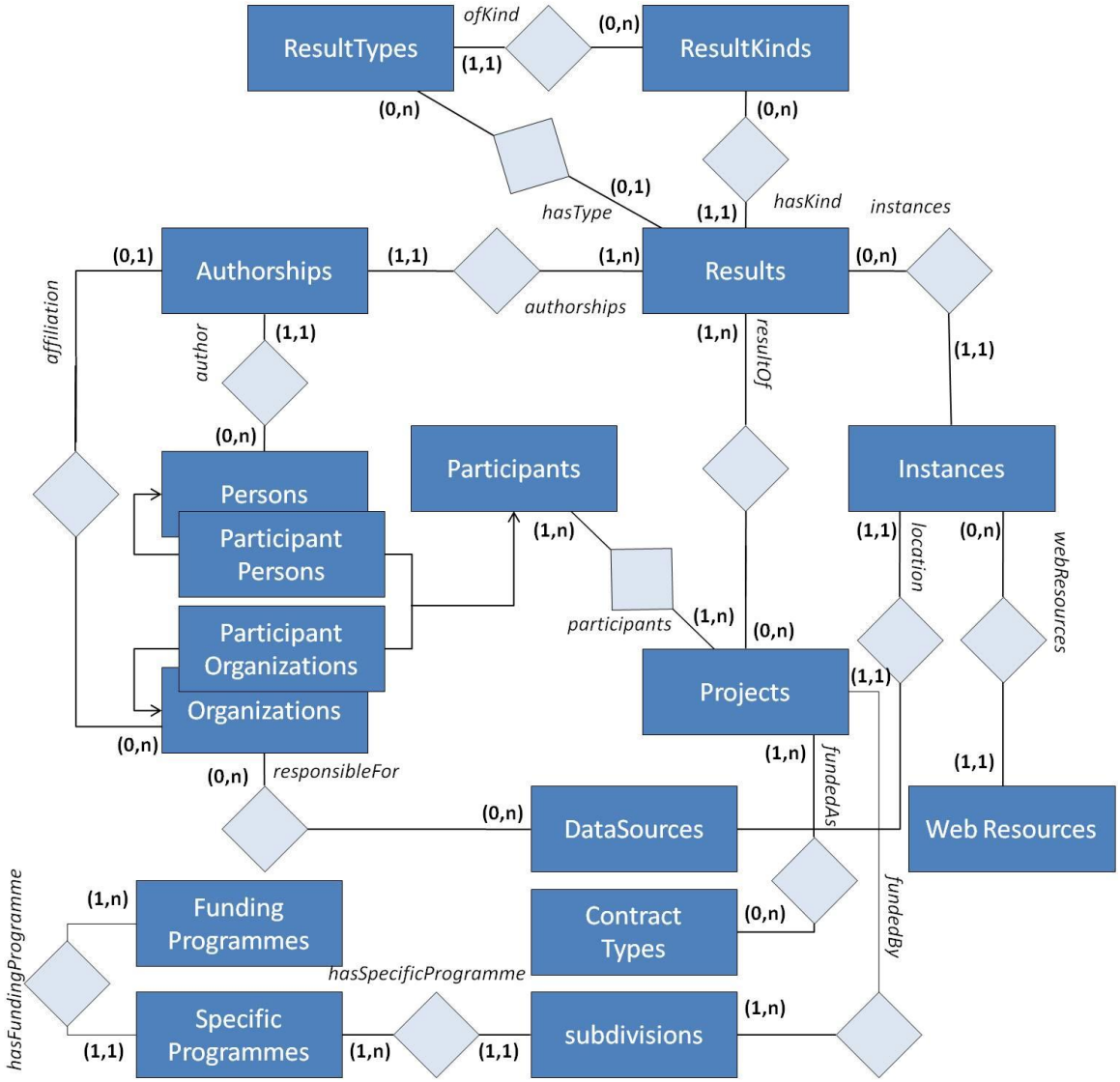

Figure 2 – OpenAIRE Entity-Relationship model

Table 1 reports the attributes and the named associations for the classes in the E-R schema in Figure 2.

Table 1 – E-R Schema: class properties

Persons • Name • surname • Nationality ➔ author^{-1} (0 or N Authorships) *Has specializations*: • **Participant Persons**	**Authorships** ➔ authorships^{-1} (1 Results) ➔ author (1 Persons) ➔ affiliation (0 or 1 Organizations)	**Instances** • Unique_identifier (URI) ➔ instances^{-1} (1 Results) ➔ location (1 Data Source) ➔ webResources (0 or N Web Resources)
Organizations • Legal short name • Legal name • Legal status • Web site URL • Logo URL	**Results** • Title • Date of publication (optional) • Description • Publisher (optional) • Language	**Projects** • Web site • Grant_agreement_number • Call_identifier (optional) • Acronym • Title

• Country of origin • Longitude, Latitude, Time zone ➔ affiliation^{-1} (0 or N Authorships) ➔ responsibleFor (0 or N Data Sources) *Has specializations*: • **Participant Organizations**	• Access mode • Embargo end-date (optional) • Keywords ➔ hasKind (1 ResultKinds) ➔ hasType (1 ResultType) (optional) ➔ authorships (1 or N Authorships) ➔ instances (0 or N Instances) ➔ resultOf (1 or N Projects) *Constraints*: • The *hasType* entity, if specified, must be in the association *relatedTypes* of the entity *hasKind*	• Start_date • End_date • Keywords ➔ participants (1 or N Participants) ➔ resultOf^{-1} (0 or N Results) ➔ fundedBy (1 Subdivisions) ➔ fundedAs (1 or N Contract Types) *Derived properties*: • Duration (from start_date and end_date) • EC_Project_Website(from grant_agreement_number)
Participants • EC_participant_number ➔ participants^{-1} (1 or N Projects) *Has specializations*: • **Participant Organizations** • **Participant Persons**	**Data Sources** • Official name • English name (optional) • Web Site URL • Logo URL • Contact email • Longitude, Latitude, Time zone • Typology (e.g., OAI-PMH, OAI-ORE, FTP) • Access Info Package (XML) ➔ location^{-1} (0 or N Instances) ➔ responsibleFor^{-1} (0 or N Organizations)	**Web Resources** • Web Resource URL ➔ webResources^{-1} (1 Instances)
Funding Programmes • Identifier • Name • Acronym ➔ hasFundingProgramme^{-1} (1 or N Funding Programmes)	**Specific Programmes** • Identifier • Name • Acronym ➔ hasSpecificProgramme^{-1} (1 or N Subdivisions) ➔ hasFundingProgramme (1 Funding Programme)	**Subdivisions** • Identifier • Name • Acronym ➔ hasSpecificProgramme (1 Specific Programmes) ➔ fundedBy^{-1} (1 or N Projects)
Contract Types • Identifier • Name ➔ fundedAs^{-1} (0 or N Projects)	**Result Kinds** • Name ➔ ofKind^{-1} (0 or N ResultTypes)	**Result Types** • Name ➔ ofKind (1 ResultKinds)

2.3 Relational Schema

The relational schema in Figure 3 is obtained as a normalized transformation from the E-R class schema in Figure 2.

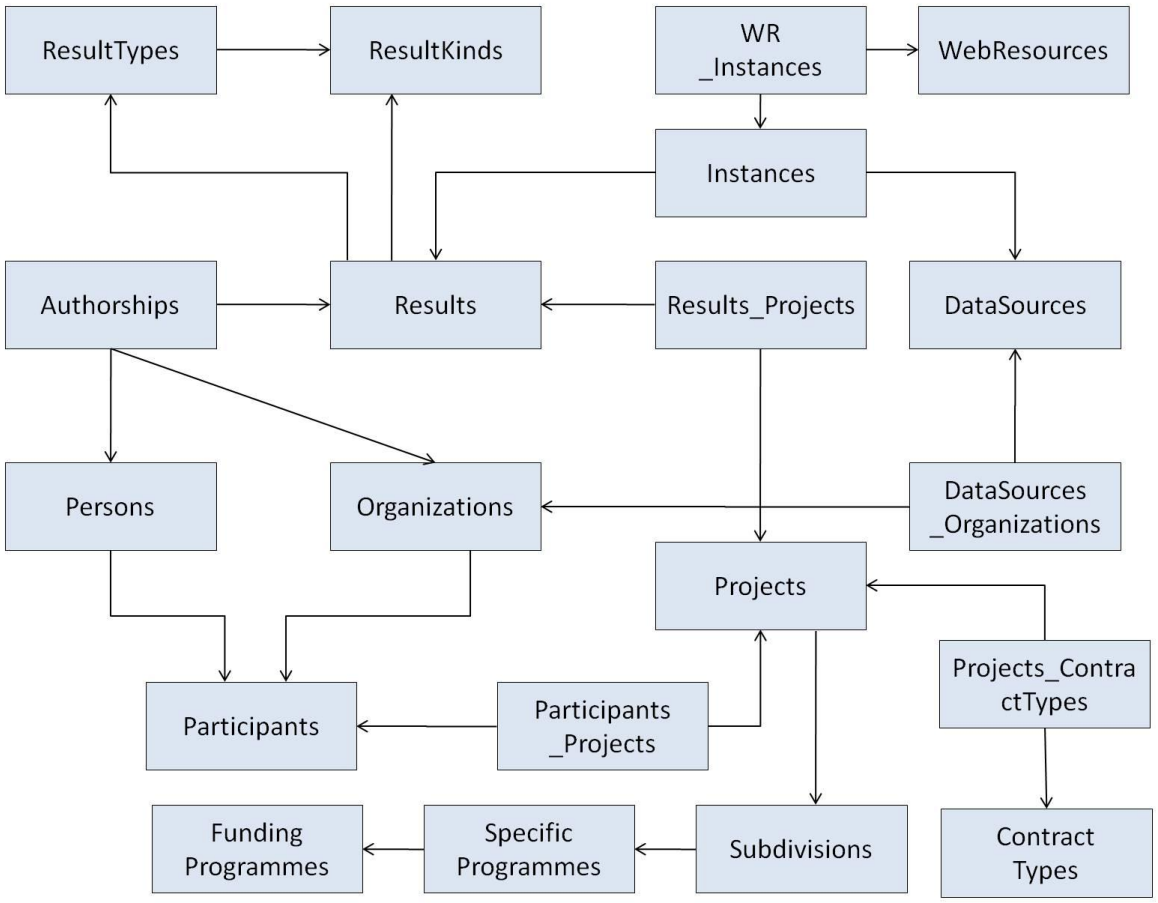

Figure 3 – OpenAIRE Relational Schema

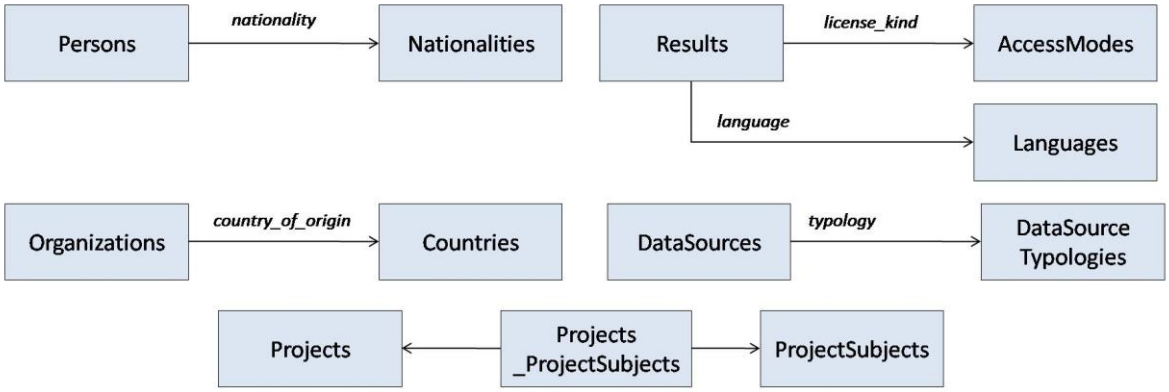

Figure 4 – Controlled Vocabularies Relational Schema: "vocabulary tables"

Figure 4 introduces the tables representing controlled vocabularies in the E-R schema, namely *vocabulary tables*. A vocabulary table has a structure of the form (key, value) is introduced whenever a controlled vocabulary needs to be managed. Currently, vocabularies are: Nationalities (of people), Result Kinds (categories of results, e.g., publication, research data, software product), Result Types (dependent on ResultKinds, for publication for example: article, book, manual, etc), Countries and Languages (possibly taken from ISO standard), Access Modes (described in the introduction), Data Source typologies (e.g., institutional repository, web site) and Project Subjects (taken from an authoritative list of subjects, to be provided by the EU: for example the seven research areas of the Pilot).

It is to be considered whether "vocabulary tables" should referred to from other tables by only using the primary key or by using the pair (key,value) in order to optimize searches (*tech*: no join query needed) and enable full-text search on values.

Finally, the "vocabulary tables" defined in this document may be subject to changes in the future: other attributes may be added to such tables, thereby extending the notion of vocabulary to that of an authority file, or new vocabularies be added (e.g., publisher). The trade-off stands in the balance between having clean and precise data and the effort required to keep a consistent and meaningful authoritative vocabulary of terms.

Table 2 – Relational schema: tables structure

Persons • personID: PK • name: string • surname: string • nationality: FK Nationality(nationalityID) (optional) • participantID: FK Participants(participantID) (optional)	**Authorships** • authorshipID: PK • result: FK Results(resourceID) • author: FK Persons(personID) • affiliation: FK Organizations(organizationID) (optional)	**Instances** • instanceID: PK • unique_identifier: String • datasource: FK dataSources(DataSourceID) • resource: FK Results(ResultID)
Web Resources • webResourceID: PK • file_URL: "URL" String	**Web Resources_Instances** • webResourceID: PK FK Web Resources(fileID) • instanceID: PK FK Instances(manifestationID)	**Organizations** • organizationID: PK • legal_short_name: String • legal_name: String • legal_status: String • web_site_URL: String • logo_URL: String (optional) • country_of_origin: FK Countries(countryID) • participantID: FK Participants(participantID) (optional)
Results • resultID: PK • title: String • publication_date: Date (optional) • description: String • publisher: String (optional) • hasKind: FK ResultKinds(resourceKindID) • hasType: FK ResultTypes(resourceTypeID) (optional) • language: FK Languages(languageID) • access_mode: FK Access_Modes(accessModeID) • embargo_end_date: Date • keywords: String	**Projects** • projectID: PK • web_site: "URL" String • EC_project_website: "URL" String • grant_agreement_number: String • call_identifier: String (optional) • acronym: String • title: String • start_date: Date • end_date: Date • fundedBy: FK Subdivisions(subdivisionID)	**Participants** • beneficiaryID: PK • EC_participant_number: String (unique)
DataSources • datasourceID: PK • official_name: String • English_name: String (optional) • web_site_URL: String • logo_URL: String • contact_email: "email" String	**Participants_Projects** • participant: PK FK Participants(participantID) • project: PK FK Projects(projectID)	**DataSources_Organizations** • datasource: PK FK DataSources(datasourceID) • organization: PK FK Organizations(organizationID)

• longitude: Number • latitude: Number • time zone: String • typology: FK DataSourceTypologies(datasourceTypologyID) • access_info_package: "XML" String		
Results_Projects • result: PK FK Results(resultID) • project: PK FK Projects(projectID)	**FundingProgrammes** • fundingProgrammeID: PK String • programme_name: String • programme_acronym: String	**SpecificProgrammes** • specificProgrammeID: PK String • specificProgramme_name: string • specificProgramme_acronym: String • hasFundingProgramme: FK FundingProgrammes(fundingProgrammeID)
Subdivisions • subdivisionID: PK String • subdivision_name: string • subdivision_acronym: String • hasSpecificProgramme: FK SpecificProgrammes(specificProgrammeID)	**ContractTypes** • contractTypeID: PK String • contractType_name: string	**Projects_ContractTypes** • project: PK FK Projects(projectID) • contractType: PK FK ContractTypes(contractTypeID)
ResultTypes • resultTypeID: PK String • name: String • ofKind: FK ResultKinds(resultKindID)	**ResultKinds** • resultKindID: PK String • name: String	**DataSourceTypologies** • datasourceTypologyID: PK String • name: String
ProjectSubjects • projectSubjectID: PK String • name: String	**Projects_ProjectSubjects** • project: PK FK Projects(projectID) • project_subject: PK FK Project_Categories(projectCategoryID)	**Nationalities** • nationalityID: PK String • name: String
Countries • countryID: PK String • name: String	**AccessModes** • accessModeID: PK String • name: String	**Languages** • languageID: PK String • name: String

3 Implications for the definition of OpenAIRE Guidelines

This section analyzes the OpenAIRE data model and identifies some of the requirements the model may impose to OpenAIRE repository managers. As such, it will serve as input to the definition of OpenAIRE Guidelines for Repository Managers.

Based on the OpenAIRE data model and on the OpenAIRE system requirements, in order to facilitate the automatic ingestion process of metadata records relative to publications from the repositories to OpenAIRE, the following requirements were identified:

1) The repository should implement an OAI-PMH interface (OpenAIRE provides a suite for managing harvesting and aggregation of a federation of OAI-PMH conformant repositories);
2) The repository should export an OAI-PMH set (possibly with an OpenAIRE-flavored name, such as "EC_Project_Results") whose metadata records correspond to the publication results of interest to OpenAIRE; if such a set is not available, an alternative way of selecting records of interest to OpenAIRE should be provided (e.g., a special metadata field allowing for direct or indirect identification of such records)
3) Information about funding projects and licenses of the publications should come along with the bibliographic metadata: repository managers should find ways to "enrich" their Dublin Core records with new fields, in particular:
    - *access_mode*, mandatory (possibly values from the OpenAIRE controlled vocabulary "AccessKinds")
    - *embargo_end_date*, mandatory when *access_mode* = "embargo"
    - *project*, mandatory sequence (values from the OpenAIRE vocabulary obtained by projection of the attribute *projectID* of the table Projects)
    - *unique identifier*, optional (the unique identifier of the publication at its original repository, different from the value expected in *dc:identifier*, see below)
4) Some of the Dublin Core fields should be mandatory or optional based on their correspondence with the OpenAIRE properties for Results of typology "publication". Similarly, values of the DC fields should possibly, not mandatorily, adhere to the OpenAIRE vocabularies to which the corresponding properties adhere. Alternatively, if values of such fields are provided according to different vocabularies, a mapping onto the matching OpenAIRE vocabulary values should be provided. Some examples are:
    - *dc:title*, mandatory
    - *dc:creator*, mandatory sequence
    - *dc:description, mandatory*
    - *dc:identifier, mandatory sequence* (the URL of the file, in the style of DRIVER Guidelines)
    - *dc:subject, optional sequence*
    - *dc:publisher, optional*

- *dc:language, optional (possibly values from the OpenAIRE controlled vocabulary "Languages")*
- *dc:type, optional (possibly values from the OpenAIRE controlled vocabulary "ResourceTypes" of ResourceKind="Publication")*

4 Collaborative refinement of the Information Space

The OpenAIRE Web Site enables registered users to access and modify the Information Space in order to add, remove clean or enrich its content. The Information Space structure is conceived to enable a growing process of data quality improvement, so as to construct and grow authority files for persons, results, organizations, data sources and projects. This process is based on trust of registered users, whose corrective actions are double-checked by administrators, and supported by automatic tools for identification of duplicates.

In this scenario, actions from registered users should be recorded and limited to those that are reversible by administrators; actions from administrators should also be recorded, but can be irreversible. Different technical solutions are possible and will be discussed at the time of implementing the Information Space.