

European Commission funding corpus.

OpenAire is populating its repository with data concerning the funding activities of the European Commission coming from the the [Open Data Portal \(https://data.europa.eu/euodp/en/home\)](https://data.europa.eu/euodp/en/home), and using the ontology [DINGO \(http://w3id.org/dingo\)](http://w3id.org/dingo) to support the modelling of the data. The data is licensed via the same license at the Open Data Portal.

The corpus includes information about a number of different type of information related to funding and research. In the particular the main classes of entities are (using the prefix dg: <https://w3id.org/dingo#> (<https://w3id.org/dingo#>)):

- Grant class dg:Grant - a disbursed fund paid to a recipient or beneficiary (a Participant) and the process for it.
- GrantPayment class dg:GrantPayment - a single payment to a recipient or beneficiary within a Grant.
- GrantShare class dg:GrantShare - the full or proper portion or part allotted or belonging to or contributed to an individual entity within a Grant.
- Project class dg:Project - an organised endeavour (collective or individual) planned to reach a particular aim or achieve a result, in the context of this corpus it indicates the funded projects.
- Role class dg:Role - the function assumed by or ascribed to an entity (typically person, group of persons or organisation) in a particular situation, which in the context of this corpus indicates the role the various mentioned entities have/had in the project or grant (for example: coordinator, participant,).
- Organisation class dg:Organisation - the social entities with a collective goal involved in the research and funding, including the ultimate funder. These are organised in a series of subclasses, using the [DINGO \(http://w3id.org/dingo\)](http://w3id.org/dingo) ontology.
- FundingAgency class dg:FundingAgency - the organisations that materially disburse and administer the Grant process.
- FundingScheme class dg:FundingScheme - the programs that determines and organizes the funding. A grant can be implementing different funding schemes or programs. The complexity of modelling the various types of funding schemes and formulas and the adopted solutions will be briefly mentioned in the next section.
- Criterion class dg:Criterion - the specification(s) of Grant coverage, Grant eligibility, Grant reimbursement rates, Grant specific criteria for funding, Grant population targets, and similar features.

All the data of the corpus is available via the OpenAire Linked Open Data SPARQL endpoint, and by downloading the OpenAire data dumps. The license of the original data can be found [here \(https://op.europa.eu/en/web/about-us/reuse-and-copyright\)](https://op.europa.eu/en/web/about-us/reuse-and-copyright).

Ontology

The data is modelled using the ontology [DINGO](http://w3id.org/dingo) (<http://w3id.org/dingo>). The ontology is purposely designed to allow modelling for a large spectrum of the funding landscape, and not only for the European Commission types of funding, as the aim would be to be able to model different funding data and perform comparative analysis across those.

Specific useful specialisations of the DINGO terminology have been encoded in the comments (rdfs:comment) associated to some of the entities. For example this has been done concerning the modelling of funding schemes. In fact, already in the case of the European Commission funding activities alone one finds programs, frameworks and actions/schemes which are all different specialisation of the general concept of funding scheme, and each funding body has its own concepts and nomenclature for funding schemes or programs. This situation which would have lead to a meaningless infinite series of narrowly specific subclasses of the general type FundingScheme, if DINGO would have attempted such a categorization. Comments have been used in order to distinguish among them in a more practically meaningful way (for example, actions and programmes are respectively commented with "Type of action in the framework work programme." and "Programme in the framework work programme.").

The dataset is also linked to the original identifiers by the European Commission: every node of type Grant is linked to the EC identifier via the property "dg:agency_identifier". (The EC terminology typically uses the word "project" to actually indicate the grant, identifying it with the research project, however DINGO allows to model the cases where a project received different grants -in sequence or also in parallel- and thus the correct mapping among types has been made in this dataset).

The modeling uses other ontologies as well, for example by employing the class MonetaryAmount from the schema.org (<http://schema.org/>) ontology. Overall the adopted ontologies are:

Prefix	Ontology	Description
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns# (http://www.w3.org/1999/02/22-rdf-syntax-ns#)	rdf-schema (https://www.w3.org/TR/rdf-schema/)
rdfs	http://www.w3.org/2000/01/rdf-schema# (http://www.w3.org/2000/01/rdf-schema#)	rdf-schema (https://www.w3.org/TR/rdf-schema/)
schema	http://schema.org/ (http://schema.org/)	schema.org (http://schema.org/)
skos	http://www.w3.org/2004/02/skos/core# (http://www.w3.org/2004/02/skos/core#)	skos-reference (https://www.w3.org/TR/skos-reference)
dg	https://w3id.org/dingo# (https://w3id.org/dingo#)	DINGO ontology (https://w3id.org/dingo)

Corpus content

The corpus at the moment contains the data related to the h302 and FP7 framework programs, except for Persons at the present stage. The actual content is the following:

Broken down per individual dataset file

FP7

type

entityCount

type	entityCount
https://w3id.org/dingo#GrantShare (https://w3id.org/dingo#GrantShare).	55493
schema:MonetaryAmount	55493
schema:PostalAddress	2487
https://w3id.org/dingo#OrganisationRole (https://w3id.org/dingo#OrganisationRole).	110986
https://w3id.org/dingo#Project (https://w3id.org/dingo#Project).	20289
https://w3id.org/dingo#Grant (https://w3id.org/dingo#Grant).	20289
https://w3id.org/dingo#EducationOrganisation (https://w3id.org/dingo#EducationOrganisation).	2487
type	entityCount
https://w3id.org/dingo#GrantShare (https://w3id.org/dingo#GrantShare).	35496
schema:MonetaryAmount	35496
schema:PostalAddress	3836
https://w3id.org/dingo#OrganisationRole (https://w3id.org/dingo#OrganisationRole).	70992
https://w3id.org/dingo#Project (https://w3id.org/dingo#Project).	13212
https://w3id.org/dingo#Grant (https://w3id.org/dingo#Grant).	13212
https://w3id.org/dingo#ResearchPerformingOrganisation (https://w3id.org/dingo#ResearchPerformingOrganisation).	3836
type	entityCount
https://w3id.org/dingo#GrantShare (https://w3id.org/dingo#GrantShare).	6250
schema:MonetaryAmount	6250
schema:PostalAddress	1955
https://w3id.org/dingo#OrganisationRole (https://w3id.org/dingo#OrganisationRole).	12500
https://w3id.org/dingo#Project (https://w3id.org/dingo#Project).	2535
https://w3id.org/dingo#Grant (https://w3id.org/dingo#Grant).	2535
https://w3id.org/dingo#GovernmentalOrganisation (https://w3id.org/dingo#GovernmentalOrganisation).	1955
type	entityCount
https://w3id.org/dingo#GrantShare (https://w3id.org/dingo#GrantShare).	41989
schema:MonetaryAmount	41989
schema:PostalAddress	19548
https://w3id.org/dingo#OrganisationRole (https://w3id.org/dingo#OrganisationRole).	83978
https://w3id.org/dingo#Project (https://w3id.org/dingo#Project).	9312
https://w3id.org/dingo#Grant (https://w3id.org/dingo#Grant).	9312
https://w3id.org/dingo#ForProfitOrganisation (https://w3id.org/dingo#ForProfitOrganisation).	19548
type	entityCount
https://w3id.org/dingo#GrantShare (https://w3id.org/dingo#GrantShare).	5432
schema:MonetaryAmount	5432
schema:PostalAddress	3255
https://w3id.org/dingo#OrganisationRole (https://w3id.org/dingo#OrganisationRole).	10864
https://w3id.org/dingo#Project (https://w3id.org/dingo#Project).	2672
https://w3id.org/dingo#Grant (https://w3id.org/dingo#Grant).	2672
https://w3id.org/dingo#Organisation (https://w3id.org/dingo#Organisation).	3255

And for the “prj_prog” file:

type	entityCount
------	-------------

type	entityCount
schema:MonetaryAmount	51556
https://w3id.org/dingo#Project (https://w3id.org/dingo#Project).	25778
https://w3id.org/dingo#Grant (https://w3id.org/dingo#Grant).	25778
https://w3id.org/dingo#FundingScheme (https://w3id.org/dingo#FundingScheme).	67
https://w3id.org/dingo#GovernmentalOrganisation (https://w3id.org/dingo#GovernmentalOrganisation),1	

H2020

type	entityCount
https://w3id.org/dingo#GrantShare (https://w3id.org/dingo#GrantShare).	33270
schema:MonetaryAmount	33270
schema:PostalAddress	2240
https://w3id.org/dingo#OrganisationRole (https://w3id.org/dingo#OrganisationRole).	66540
https://w3id.org/dingo#Project (https://w3id.org/dingo#Project).	13873
https://w3id.org/dingo#EducationOrganisation (https://w3id.org/dingo#EducationOrganisation),2240	
https://w3id.org/dingo#Grant (https://w3id.org/dingo#Grant).	13873

type	entityCount
https://w3id.org/dingo#GrantShare (https://w3id.org/dingo#GrantShare).	33668
schema:MonetaryAmount	33668
schema:PostalAddress	18820
https://w3id.org/dingo#OrganisationRole (https://w3id.org/dingo#OrganisationRole).	67336
https://w3id.org/dingo#Project (https://w3id.org/dingo#Project).	10603
https://w3id.org/dingo#Grant (https://w3id.org/dingo#Grant).	10603
https://w3id.org/dingo#ForProfitOrganisation (https://w3id.org/dingo#ForProfitOrganisation),18820	

type	entityCount
https://w3id.org/dingo#GrantShare (https://w3id.org/dingo#GrantShare).	21135
schema:MonetaryAmount	21135
schema:PostalAddress	2805
https://w3id.org/dingo#OrganisationRole (https://w3id.org/dingo#OrganisationRole).	42270
https://w3id.org/dingo#Project (https://w3id.org/dingo#Project).	8273
https://w3id.org/dingo#Grant (https://w3id.org/dingo#Grant).	8273
https://w3id.org/dingo#ResearchPerformingOrganisation (https://w3id.org/dingo#ResearchPerformingOrganisation).	2805

type	entityCount
https://w3id.org/dingo#GrantShare (https://w3id.org/dingo#GrantShare).	5824
schema:MonetaryAmount	5824
schema:PostalAddress	2065
https://w3id.org/dingo#OrganisationRole (https://w3id.org/dingo#OrganisationRole).	11648
https://w3id.org/dingo#Project (https://w3id.org/dingo#Project).	2190
https://w3id.org/dingo#GovernmentalOrganisation (https://w3id.org/dingo#GovernmentalOrganisation),2065	
https://w3id.org/dingo#Grant (https://w3id.org/dingo#Grant).	2190

type	entityCount
https://w3id.org/dingo#GrantShare (https://w3id.org/dingo#GrantShare).	5422

type	entityCount
schema:MonetaryAmount	5422
schema:PostalAddress	2792
https://w3id.org/dingo#OrganisationRole (https://w3id.org/dingo#OrganisationRole),	10844
https://w3id.org/dingo#Project (https://w3id.org/dingo#Project),	2621
https://w3id.org/dingo#Grant (https://w3id.org/dingo#Grant),	2621
https://w3id.org/dingo#Organisation (https://w3id.org/dingo#Organisation),	2792

And for the “prj_prog” file:

type	entityCount
schema:MonetaryAmount	44304
https://w3id.org/dingo#Project (https://w3id.org/dingo#Project),	22152
https://w3id.org/dingo#Grant (https://w3id.org/dingo#Grant),	22152
https://w3id.org/dingo#FundingScheme (https://w3id.org/dingo#FundingScheme),	256
https://w3id.org/dingo#GovernmentalOrganisation (https://w3id.org/dingo#GovernmentalOrganisation),	1

Combined datasets

type	entityCount
https://w3id.org/dingo#GrantShare (https://w3id.org/dingo#GrantShare),	243979
schema:MonetaryAmount	339839
schema:PostalAddress	49160
https://w3id.org/dingo#OrganisationRole (https://w3id.org/dingo#OrganisationRole),	487958
https://w3id.org/dingo#Project (https://w3id.org/dingo#Project),	47930
https://w3id.org/dingo#EducationOrganisation (https://w3id.org/dingo#EducationOrganisation),	3123
https://w3id.org/dingo#Grant (https://w3id.org/dingo#Grant),	47930
https://w3id.org/dingo#ForProfitOrganisation (https://w3id.org/dingo#ForProfitOrganisation),	32926
https://w3id.org/dingo#Organisation (https://w3id.org/dingo#Organisation),	5329
https://w3id.org/dingo#FundingScheme (https://w3id.org/dingo#FundingScheme),	323
https://w3id.org/dingo#GovernmentalOrganisation (https://w3id.org/dingo#GovernmentalOrganisation),	3203
https://w3id.org/dingo#ResearchPerformingOrganisation (https://w3id.org/dingo#ResearchPerformingOrganisation),	4674

Data quality

The corpus is the union of data concerning the H2020 and the FP7 framework programs of the European Commission. The data quality (correctness and validity of the identifiers, absence of nulls, ...) is higher for the H2020 part than for the FP7. The dataset suffers from a certain percentage of standard issues for data of this kind: for example concerning the participant organisations' names, a well-known problem due to the fact that there is no authoritative identifier for those presently (some initiatives are developing in that sense but it is not clear if the Commission data will be aligned with them). The modelling of the corpus has tried to cope with those issue by doing conservative data cleaning, and marking or unclear cases with special identifiers (containing the fragment “unkn”).

Exploring the corpus: example queries

The corpus can be explored using SPARQL queries via the [sparql_endpoint\(\)](#). Some useful and simple example query that can give an immediate feeling of the available data will be presented here. This will also allow us to illustrate some particular characteristics of the data.

A simple query to count all entities in the corpus is:

```
PREFIX dg:<https://w3id.org/dingo#>
      PREFIX schema:<http://schema.org/>

select ?type (COUNT(distinct ?entity) as ?entityCount)
where {

    ?entity rdf:type ?type.
}
group by ?type
```

The data of the funding activities of the European Commission as available in the Open Data Portal has some uncertainty, for example concerning the participant organisations' names. Indeed, in different moments, the registrars of the participant organisations have sometimes indicated different organisation names, which have not yet been fully harmonised. This is a standard well-known problem with organisations' identification in datasets, in absence, at this moment, of an accepted working identifier. To cope with this feature, a query capturing, say, the amounts of funding to each participant organisation for a given project (say OpenAire, edition FP7, with grant identifier 246686, and a followup grant "643410" in H2020), may be:

```
PREFIX dg:<https://w3id.org/dingo#>
```

```
PREFIX schema: <http://schema.org/>
```

```
select (sample(?projectname) as ?PROJECTNAME) (group_concat(distinct ?orgnName;separator="; ") as ?organisationNames) (sample(?organisationType) as ?ORGANISATIONTYPE) (sample(?grantShareValue) as ?GRANTSHARE) (sample(?currency) as ?CURRENCY) ?title
```

```
where {
```

```
  ?proj rdf:type dg:Project.
```

```
  ?proj dg:funded_by ?grant.
```

```
  ?proj dg:short_name ?projectname.
```

```
  ?proj dg:title ?title.
```

```
  ?grant dg:hasPart ?grantshare.
```

```
  ?grant dg:agency_identifier "246686". #"643410".
```

```
  ?grantshare dg:economic_value ?amount.
```

```
  ?amount schema:value ?grantShareValue.
```

```
  ?amount schema:currency ?currency.
```

```
  ?grantshare dg:recipient ?orgn.
```

```
  ?orgn rdf:type ?organisationType.
```

```
  ?orgn dg:legalName ?orgnName.
```

```
}
```

```
group by ?orgn
```

This query also allows us to make an important remark: the OpenAire project has existed and been funded both during the FP7 and H2020 frameworks. The operational databases of the European Commission typically identify project with grant, and thus have to fictitiously create two OpenAire projects, while actually OpenAire had two consecutive grants. In this corpus

A query enabling to obtain some of the data concerning funding schemes and actions is instead:

```
PREFIX dg: <https://w3id.org/dingo#>
PREFIX schema: <http://schema.org/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

select *
where {
  ?proj rdf:type dg:Project.
  ?proj dg:funded_by ?grant.
  ?proj dg:title ?title.
  ?proj dg:short_name ?acronym.
  ?grant dg:official_website ?grant_webs.
  ?grant dg:start_time ?grantStartTime.
  ?grant dg:end_time ?grantEndTime.
  ?grant dg:implementation_of ?fundingProgram.
  ?fundingProgram dg:isPartOf+ ?fundingProgram2.
  ?fundingProgram dg:funder ?funder.
  ?fundingProgram dg:short_name ?fundingProgramName.
  ?fundingProgram rdfs:comment ?comment.
}
```



(<http://creativecommons.org/licenses/by/4.0/>) UNLESS OTHERWISE INDICATED, ALL MATERIALS CREATED BY THE OPENAIRE CONSORTIUM ARE LICENSED UNDER A CREATIVE COMMONS ATTRIBUTION 4.0 INTERNATIONAL LICENSE (<http://creativecommons.org/licenses/by/4.0/>).

OPENAIRE IS POWERED BY D-NET (<http://www.d-net.research-infrastructures.eu/>).

