

Void file structure for the RDF platform (November 2016)

The Void file structure is orientated on the proposals in this paper <https://www.w3.org/TR/hcls-dataset/> - however we only enforce some rules. Additional information are welcome, but not necessary.

Overall structure: Datasets have to have 3 levels: Summary, Version and Distribution Level. Some more complex datasets may also use linksets (4th level). All these levels have to be connected.

Summary Level: The summary level provides a description of a dataset that is independent of a specific version or format.

Version Level: The version level captures version-specific characteristics of a dataset.

Distribution Level: The distribution level captures metadata about a specific form and version of a dataset.

Rules that the RDF platform enforces (Void File)

Summary Level is defined through:

<<http://purl.org/dc/terms/Dataset>>
<<http://purl.org/pav/hasCurrentVersion>> (connection to Version level)

In addition it needs:

<http://purl.org/dc/terms/title>
<http://purl.org/dc/terms/publisher>
<http://purl.org/dc/terms/description>

Is **NOT** allowed to have:

<http://rdfs.org/ns/void#dataDump>
<http://purl.org/dc/terms/creator>

Version Level is defined through

<<http://purl.org/dc/terms/Dataset>>
<<http://purl.org/dc/terms/isVersionOf>> (connection to Summary level)
<<http://purl.org/dc/terms/hasDistribution>> (connections to distribution level)

It addition needs:

<http://purl.org/dc/terms/title>
<http://purl.org/dc/terms/description>
<http://purl.org/dc/terms/creator>
<http://purl.org/dc/terms/publisher>
<http://purl.org/pav/version>

Is **NOT** allowed to have:

<http://rdfs.org/ns/void#dataDump>

Distribution Level is defined through

<<http://rdfs.org/ns/void#Dataset>>
<<http://rdfs.org/ns/void#dataDump>>

Needs to have

<http://purl.org/dc/terms/title>

<http://purl.org/dc/terms/description>

Example for such a 'simple, minimal' void [file can be found here](#) .

More complex VoID structure:

Some resources (Ensembl, ChEMBL, OLS) have a more complex structure which also has to be reflected in the void file. It is possible to use hasPart, subSet and Linksets:

namedGraphs: Data is loaded into a named graph. The name is defined through the **summary level URI**. Data Sources with a simple VoID structure get pushed into one graph, which is fine in most cases.

HasPart: Can be used at the Summary level if a dct:Dataset has subParts. Each subPart again has to have a summary Level, Version level and distribution level! (Example UseCase: OLS consists of many ontologies, each ontology has its own levels and an own download link). In case hasPart is used, only the download links of the parts is loaded into the platform! (We offer a download for the whole of ols as well as each ontology separately. To load both would lead to duplication - in addition, we want every ontology to live in its own named graph!)

subSet: On Distribution Level subSets (<<http://rdfs.org/ns/void#subset>>) can be used to further split up the data. Subsets can link to other subsets, but in the end have to link to <<http://rdfs.org/ns/void#Dataset>> or to a linkSets (<<http://rdfs.org/ns/void#Linkset>>) which provide the dataDump link.

[Example hasPart](#)

[Example subSet](#)

REMINDER:

As defined in the previously mentioned [specification](#) there are a lot more things that you can(should) add to the VoID file, these rules here merely represent the minimum set of rules that we want to enforce to have a common base. Additional information is allowed unless stated otherwise!!! The table in that file gives a good overview about other fields you might want to add as well where to put that information (summary, version or distribution level).

Simple VoID Example:

#Definition of prefixes

@prefix dcterm: <<http://purl.org/dc/terms/>> .

@prefix void: <<http://rdfs.org/ns/void#>> .

@prefix dct: <<http://purl.org/dc/terms/>> .

@prefix dcat: <<http://www.w3.org/ns/dcat#>> .

@prefix idot: <<http://identifiers.org/idot/>> .

@prefix pav: <http://purl.org/pav/> .

#Summary Level

```
<http://rdf.ebi.ac.uk/dataset/reactome> a dct:Dataset;  
dcterms:title "Reactome" ;  
pav:hasCurrentVersion <http://rdf.ebi.ac.uk/dataset/reactome/16.0>;  
dct:publisher "EBI" ;  
dcterms:description "The Description of this dataSet" .
```

#Version Level

```
<http://rdf.ebi.ac.uk/dataset/reactome/16.0> a dct:Dataset ;  
<http://purl.org/dc/terms/isVersionOf> <http://rdf.ebi.ac.uk/dataset/reactome/16.0>;  
dcterms:description "The Description of the Version level of Reactome";  
dcterms:title "Reactome Version" ;  
dct:publisher "EBI" ;  
dct:creator "TestCreator in that case" ;  
dct:hasDistribution <http://rdf.ebi.ac.uk/dataset/reactome/16.0.void>.
```

#Distribution Level

```
<http://rdf.ebi.ac.uk/dataset/reactome/16.0.void> a void:Dataset ,  
dcat:Distribution ;  
void:dataDump  
<ftp://ftp.ebi.ac.uk/pub/databases/spot/rdf/glue/test_reactome.tar.bz2> ;  
dct:creator "TestCreator in that case" ;  
dcterms:description "The Description of the Version level of Reactome";  
dct:publisher "EBI" ;  
dct:license "http://www.ebi.ac.uk/about/terms-of-use" ;  
idot:preferredPrefix "reactome" ;  
dcterms:title "Test Distribution level of reactome" .
```

HasPart Example

#Summary Level OLS

```
ols:ols a dct:Dataset ;  
dct:description "This is just a test to create a OLS VoID File" ;  
dct:hasPart ols:aeo,  
ols:aero,  
ols:agro,  
ols:ancestro, (....and many more .....).  
dct:publisher "http://www.ebi.ac.uk/" ;  
dct:title "Ontology Lookup Service" ;  
pav:hasCurrentVersion <http://rdf.ebi.ac.uk/dataset/ols/2016-10-31> ;  
foaf:page "http://www.ebi.ac.uk/rdf/services/" .
```

#Version Level OLS

```
<http://rdf.ebi.ac.uk/dataset/ols/2016-10-31> a dct:Dataset ;  
dct:creator ebi:SPOT ;  
dct:description "OLS description on version level" ;  
dct:hasDistribution <http://rdf.ebi.ac.uk/dataset/ols/2016-10-31.rdf> ;  
dct:isVersionOf ols:ols ;  
dct:issued "2016-10-31"^^xsd:date ;
```

dct:license "http://www.ebi.ac.uk/about/terms-of-use" ;
dct:publisher "http://www.ebi.ac.uk/" ;
dct:rights "Apache version 2.0" ;
dct:title "Title on Version level? " .

#Distribution Level OLS

<http://rdf.ebi.ac.uk/dataset/ols/2016-10-31.rdf> a void:Dataset,
dcat:Distribution ;
dct:creator ebi:SPOT ;
dct:description "This is just a test to create a OLS VoID File" ;
dct:license "http://www.ebi.ac.uk/about/terms-of-use" ;
dct:publisher "http://www.ebi.ac.uk/" ;
dct:title "Add title for ols on Distribution level" ;
pav:version "2016-10-31"^^xsd:date ;
void:dataDump "ftp://ftp.ebi.ac.uk/pub/databases/spot/ols/ols_rdf/ols_rdf.tar.gz" ;
foaf:page <http://www.ebi.ac.uk/rdf/service/ols> .

#Summary Level for aero

ols:aero a dct:Dataset ;
dct:description "The Adverse Event Reporting Ontology (AERO) is an ontology aimed at supporting clinicians at the time of data entry, increasing quality and accuracy of reported adverse events" ;
dct:publisher "Who do we put in here" ;
dct:title "Adverse Event Reporting Ontology" ;
pav:hasCurrentVersion <http://rdf.ebi.ac.uk/dataset/aero/17-05-2016> .

#Version Level for aero

<http://rdf.ebi.ac.uk/dataset/aero/17-05-2016> a dct:Dataset ;
dct:creator "Who is the creator/Versionlevel" ;
dct:description "Description of ..." ;
dct:hasDistribution <http://rdf.ebi.ac.uk/dataset/aero/17-05-2016.rdf> ;
dct:isVersionOf ols:aero ;
dct:issued "2016-10-31"^^xsd:date ;
dct:publisher "Who is the publisher/Versionlevel" ;
dct:title "What is the Version title?" .

#Distribution level for aero

<http://rdf.ebi.ac.uk/dataset/aero/17-05-2016.rdf> a void:Dataset,
dcat:Distribution ;
idot:preferredPrefix "aero" ;
dct:creator "Who is the creator/DistLevel" ;
dct:description "Description of ...",
"The Adverse Event Reporting Ontology (AERO) is an ontology aimed at supporting clinicians at the time of data entry, increasing quality and accuracy of reported adverse events" ;
dct:license "http://www.ebi.ac.uk/about/terms-of-use" ;
dct:publisher "Who is the publisher/DistLevel" ;
dct:title "Distribution title" ;
pav:version "17-05-2016" ;
void:dataDump "ftp://ftp.ebi.ac.uk/pub/databases/spot/ols/ols_rdf/aero.tar.gz" ;
foaf:page "http://ebi.ac.uk/ols/ontologies/aero" .

subSet Example (Ensembl) - using hasPart and subsets

#Definition of prefixes

```
@prefix dcterm: <http://purl.org/dc/terms/> .
@prefix void: <http://rdfs.org/ns/void#> .
@prefix dct: <http://purl.org/dc/terms/> .
@prefix dcat: <http://www.w3.org/ns/dcat#> .
@prefix idot: <http://identifiers.org/idot/> .
@prefix pav: <http://purl.org/pav/> .
@prefix dcterm: <http://purl.org/dc/terms/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
```

#Summary level

```
<http://rdf.ebi.ac.uk/dataset/ensembl> a dct:Dataset ;
  dcterm:title "VoID Description"@en ;
  dcterm:description "The VoID description for the RDF representation of this
dataset."@en ;
  dct:publisher "EBI" ;
  pav:hasCurrentVersion <http://rdf.ebi.ac.uk/dataset/ensembl/86> ;
  dct:hasPart <http://rdf.ebi.ac.uk/dataset/ensembl/86/homo_sapiens>,
    <http://rdf.ebi.ac.uk/dataset/ensembl/86/7719> .
```

#Version level

```
<http://rdf.ebi.ac.uk/dataset/ensembl/86> a dct:Dataset ;
  dct:isVersionOf <http://rdf.ebi.ac.uk/dataset/ensembl> ;
  dcterm:description "Ensembl is a joint project between EMBL - EBI and the
Wellcome Trust Sanger Institute to develop a software system which produces and
maintains automatic annotation on selected eukaryotic genomes."@en ;
  dcterm:title "Ensembl RDF"@en ;
  dcterm:publisher <http://www.ebi.ac.uk> ;
  dct:creator "EBI - Ensembl" ;
  pav:version "86" ;
  dct:hasDistribution <http://rdf.ebi.ac.uk/dataset/ensembl/86.rdf> .
```

#Distribution Level

```
<http://rdf.ebi.ac.uk/dataset/ensembl/86.rdf> a void:Dataset ;
  dcterm:description "Ensembl is a joint project between EMBL - EBI and the
Wellcome Trust Sanger Institute to develop a software system which produces and
maintains automatic annotation on selected eukaryotic genomes."@en ;
  dcterm:license <http://www.ebi.ac.uk/Information/termsofuse.html> ;
  dcterm:publisher <http://www.ebi.ac.uk> ;
  dcterm:title "Ensembl 86 RDF"@en ;
  void:dataDump <ftp link to the whole of ensembl e.g -
ftp://ftp.ensembl.org/pub/current_rdf/ensembl.gz> .
```

#Summary home sapiens

```
<http://rdf.ebi.ac.uk/dataset/ensembl/86/homo_sapiens> a dct:Dataset ;
```

```
dct:description " ..... " ;
dct:publisher "Publisher" ;
dct:title "Ensembl - Homo sapiens" ;
pav:hasCurrentVersion
<http://rdf.ebi.ac.uk/dataset/ensembl/86/homos_sapiens/14_11_2016> .
```

#Version Level of homo sapiens

```
<http://rdf.ebi.ac.uk/dataset/ensembl/86/homos_sapiens/14_11_2016> a dct:Dataset
;
dct:creator "Who is the creator/Versionlevel" ;
dct:description "Description of ..." ;
dct:hasDistribution <http://rdf.ebi.ac.uk/dataset/ensembl/86.rdf> ;
dct:isVersionOf <http://rdf.ebi.ac.uk/dataset/ensembl/86/homo_sapiens>;
dct:issued "2016-10-31"^^xsd:date ;
pav:version "86 - 14_11_2016" ;
dct:publisher "Who is the publisher/Versionlevel" ;
dct:title "What is the Version title?" .
```

#Distribution Level of Homo sapiens

```
<http://rdf.ebi.ac.uk/dataset/ensembl/86/homos_sapiens/14_11_2016.rdf> a
void:Dataset,
dcat:Distribution ;
dct:creator "Who is the creator/DistLevel" ;
dct:description "Description of ...";
dct:publisher "Who is the publisher/DistLevel" ;
dct:title "Distribution title" ;
pav:version "17-05-2016" ;
void:subset <http://rdf.ebi.ac.uk/dataset/ensembl/86/homo_sapiens_subset>,
<http://rdf.ebi.ac.uk/dataset/ensembl/86/homo_sapiens_xrefs_subset> .
```

#Subset of Distribution level

```
<http://rdf.ebi.ac.uk/dataset/ensembl/86/homo_sapiens_subset> a void:Dataset ;
dcterms:title "Ensembl 86 homo sapiens" ;
dcterms:description "Ensembl 86 homo sapiens" ;
void:dataDump
<ftp://ftp.ensembl.org/pub/current_rdf//homo_sapiens/homo_sapiens.ttl.gz> .
```

#Subset of xref for homo_sapiens

```
<http://rdf.ebi.ac.uk/dataset/ensembl/86/homo_sapiens_xrefs_subset> a
void:Dataset ;
dcterms:title "xrefs for homo_sapiens" ;
dcterms:description "xrefs for homo_sapiens" ;
void:dataDump
<ftp://ftp.ensembl.org/pub/current_rdf//homo_sapiens/homo_sapiens_xrefs.ttl.gz>.
```

#Summary Level of 7719

```
<http://rdf.ebi.ac.uk/dataset/ensembl/86/7719> a dct:Dataset ;
dct:description "The Adverse Event Reporting Ontology (AERO) is an ontology aimed
at supporting clinicians at the time of data entry, increasing quality and accuracy of
reported adverse events" ;
dct:publisher "Who do we put in here" ;
dct:title "Adverse Event Reporting Ontology" ;
pav:hasCurrentVersion <http://rdf.ebi.ac.uk/dataset/ensembl/86/7719/1> .
```

#Version Level of 7719

```
<http://rdf.ebi.ac.uk/dataset/ensembl/86/7719/1> a dct:Dataset ;
  dct:creator "Who is the creator/Versionlevel" ;
  dct:description "Description of ..." ;
  dct:hasDistribution <http://rdf.ebi.ac.uk/dataset/ensembl/86/7719/1.rdf> ;
  dct:isVersionOf <http://rdf.ebi.ac.uk/dataset/ensembl/83/7719> ;
  pav:version "1" ;
  dct:issued "2016-10-31"^^xsd:date ;
  dct:publisher "Who is the publisher/Versionlevel" ;
  dct:title "What is the Version title?" .
```

#Distribution level of 7719

```
<http://rdf.ebi.ac.uk/dataset/ensembl/86/7719/1.rdf> a void:Dataset,
  dcat:Distribution ;
  dct:creator "Who is the creator/DistLevel" ;
  dct:description "Description of ...",
    "The Adverse Event Reporting Ontology (AERO) is an ontology aimed at
    supporting clinicians at the time of data entry, increasing quality and accuracy of
    reported adverse events" ;
  dct:license "http://www.ebi.ac.uk/about/terms-of-use" ;
  dct:publisher "Who is the publisher/DistLevel" ;
  dct:title "Distribution title" ;
  pav:version "17-05-2016" ;
  void:subset <http://rdf.ebi.ac.uk/dataset/ensembl/83/7719/1/subset>,
    <http://rdf.ebi.ac.uk/dataset/ensembl/83/1/30538_xref> .
```

#Subset of Distribution level

```
<http://rdf.ebi.ac.uk/dataset/ensembl/83/7719/1/subset> a void:Dataset ;
  dcterms:title "7719 - human readable" ;
  dcterms:description "7719 stands for ... " ;
  void:dataDump
  <ftp://ftp.ensembl.org/pub/current_rdf//gorilla_gorilla/gorilla_gorilla.ttl.gz> .
```

#Subset of xref for 7719

```
<http://rdf.ebi.ac.uk/dataset/ensembl/83/1/30538_xref> a void:Dataset ;
  dcterms:title "x_ref for 30538" ;
  dcterms:description "xref for 30538 - whatever it is, most likely not homo sapiens
  but this is just an example" ;
  void:dataDump
  <ftp://ftp.ensembl.org/pub/current_rdf//gorilla_gorilla/gorilla_gorilla_xrefs.ttl.gz> .
```