## 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. To test your void file you can use this online validator: https://www.w3.org/2015/03/ShExValidata/ - it also indicates the SHOULD, MAY, MAY not fields …

In addition there is a python script in this folder, which is the quality control that is used for the platform and which the void files have to pass.

**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 in 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>
<http://www.w3.org/ns/dcat#Distribution>

Needs to have
http://purl.org/dc/terms/title
http://purl.org/dc/terms/description
http://purl.org/dc/terms/creator
http://purl.org/dc/terms/format
http://purl.org/dc/terms/license
http://purl.org/dc/terms/publisher

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 cna 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)