**Scope**

We want to push the design time metadata for a basic ETL process to EDC via metadata registry interface.

**ETL scenario (see attached excel sheet Analysis\_Lineage\_InnerSource.xlsx)**

In our test setup we have an ETL process (implemented in Optimus).

The ETL process copies a dataset pc\_contract\_header (Acquisition layer) to

dataset har\_pc\_cont\_header (Harmonization layer).

**Metadata registry interface EDC**

We need to use the metadata registry interface for publishing the metadata to EDC. We deliver json files that comply to the interface specification to a ADLS folder and EDC will pick this files up. The interface specification (with samples) are described here : <https://dev.azure.com/cbsp-abnamro/GRD0001030/_git/metadata-registry-interface-specifications>

**Implementation POC**

For the ETL scenario we have created 6 json files.

|  |  |
| --- | --- |
| Name | Explanation |
| acq\_physical\_entity.json | Physical entity : pc\_contract\_header |
| acq\_physical\_attribute.json | Physical attributes for table : pc\_contract\_header |
| harmo\_physical\_entity.json | Physical entity : har\_pc\_cont\_header |
| harmo\_physical\_attribute.json | Physical attributes for table : har\_pc\_cont\_header |
| harmo\_physical\_entity\_association.json | The association between entity pc\_contract\_header and har\_pc\_cont\_header. |
| harmo\_physical\_attribute\_association.json | The association of two (or more) physical attributes between entity pc\_contract\_header and har\_pc\_cont\_header. |
| 2c43088b-0ae3-411f-bb0f-f33e40b21f9c.json | Implementation Reference to the (isc-optimus) job property defined In the entity association. |
| 24231a7a-acd0-4f5d-86a6-77fa61028c1a.json | Implementation Reference to the (isc-orchestrator) DAG property defined In the entity association. |
| dddf8c61-24f3-420b-8be8-8edcc834c7a4.json | Implementation Reference to the (isc-orchestrator) Task property defined In the entity association. It’s contains simplified content. |

**Question Json file content**

Regarding to the physical\_entity schema:

1. Can we merge acq\_physical\_entity.json and harmo\_physical\_entity.json into one json. Is it allowed to deliver both (or more) entities in one physical json file?

Regarding to the physical\_entity\_association schema:

1. In the example file the DAG ,Job and Task property is referenced with an UID. Should the actual implementation for the DAG , Job and Task property be provided in a separate file name , named conform the UID value?

Regarding to the physical\_attribute\_association schema:

1. The physicalEntityAssocation attribute in the test json provided in the repo, it is pointing to the UID of the physical entity not to the physical entity association. We think it needs to point to the physical\_entity\_association UID. Is this assumption correct?
2. how the formula property needs to be configured in the case of a one to one mapping of column from source to target (without any transformation logic). Our assumption is that the formula looks like this {{d042b292-7c81-408a-ad43-49833ca804c1}} where the UID value it the reference to the physical source attribute column. Is this assumption correct?
3. When we have a transformation that adds an additional column to the target dataset can we exclude the from property or we leave the from property blank ? (see example json file provided).