Import Data using the OpenEHR-Flat-Loader

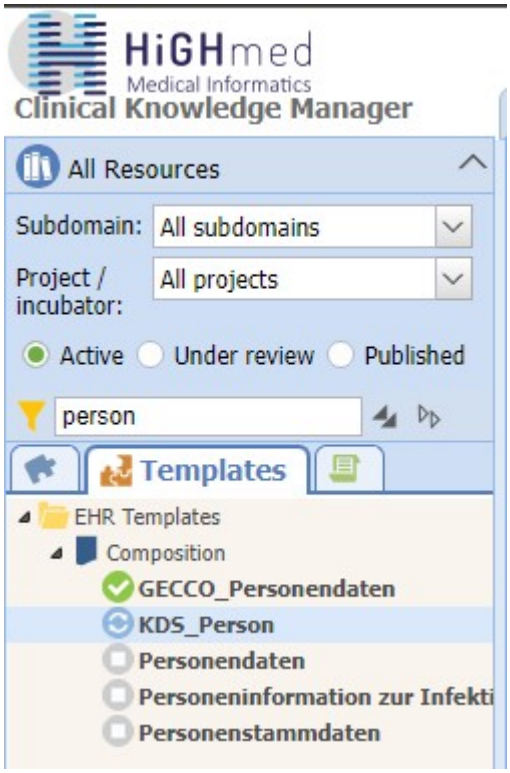
Manual and Tips by Jendrik Richter

- Question to jendrik.richter@med.uni-goettingen.de

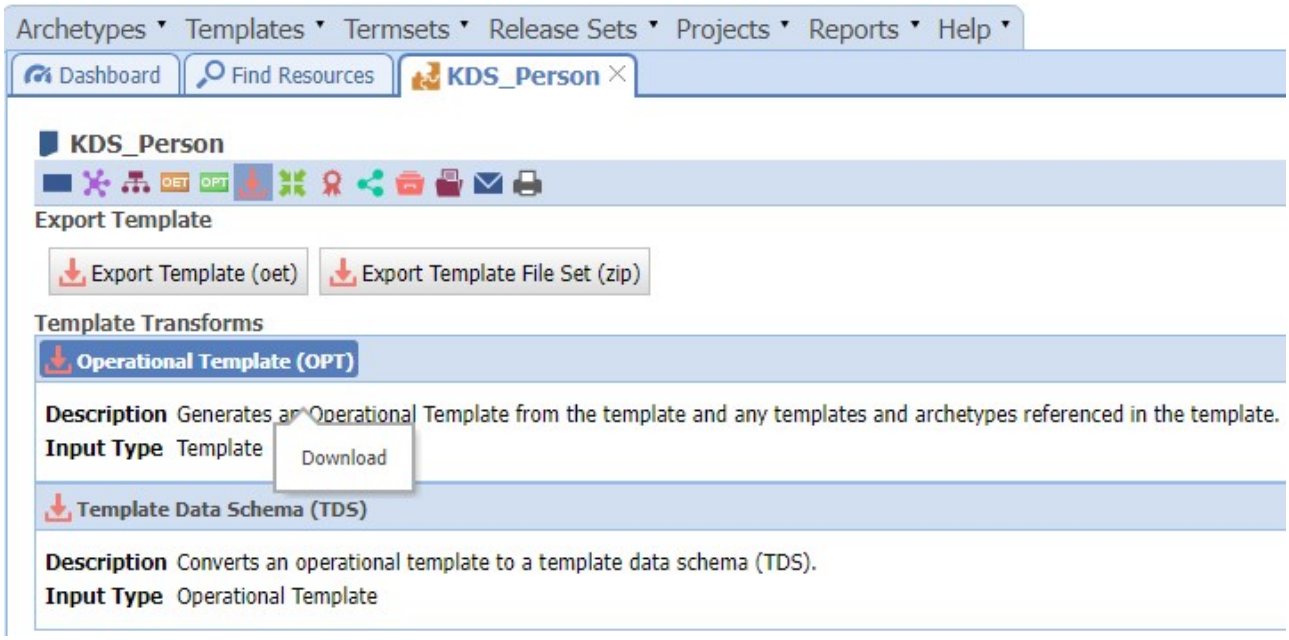
Prerequisite

OPT:

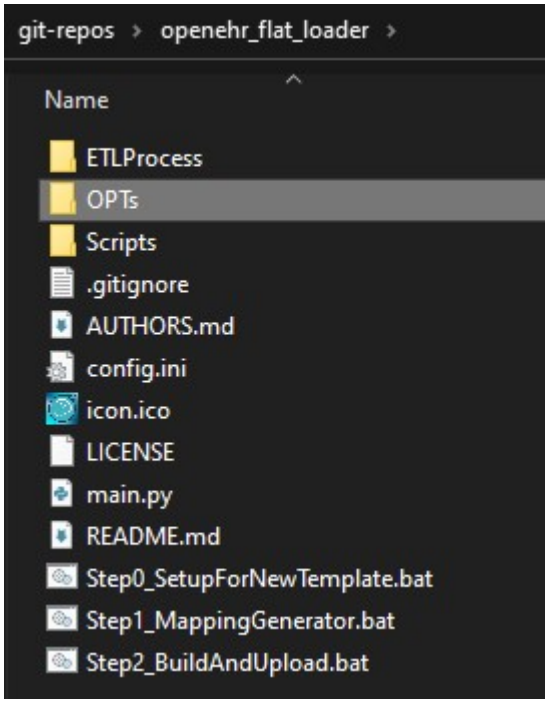
- Search the KDS_Person Template in the CKM



- Download KDS_Person-Template from CKM

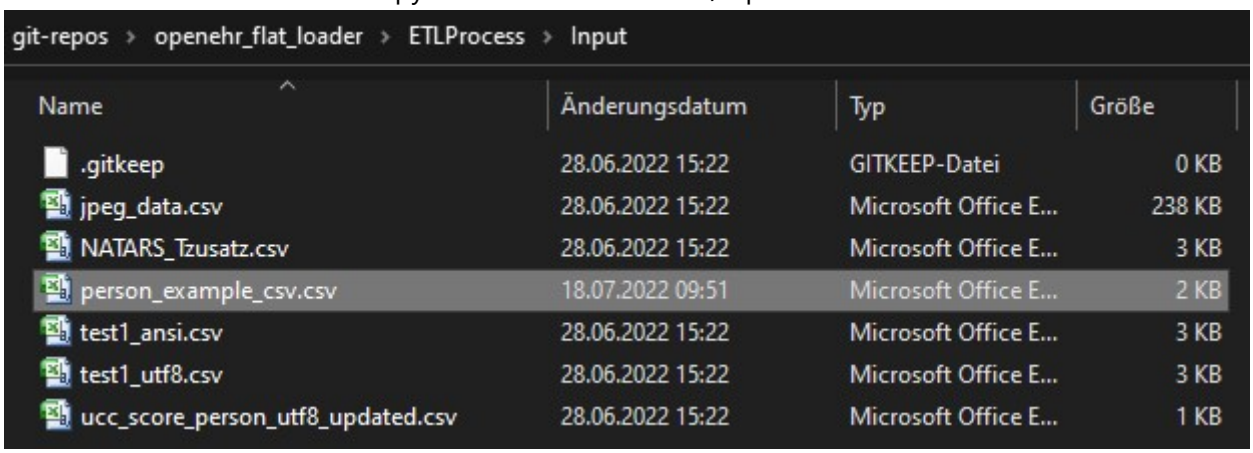


- Copy KDS_Person.opt to OPTs-Folder of the FLAT-Loader



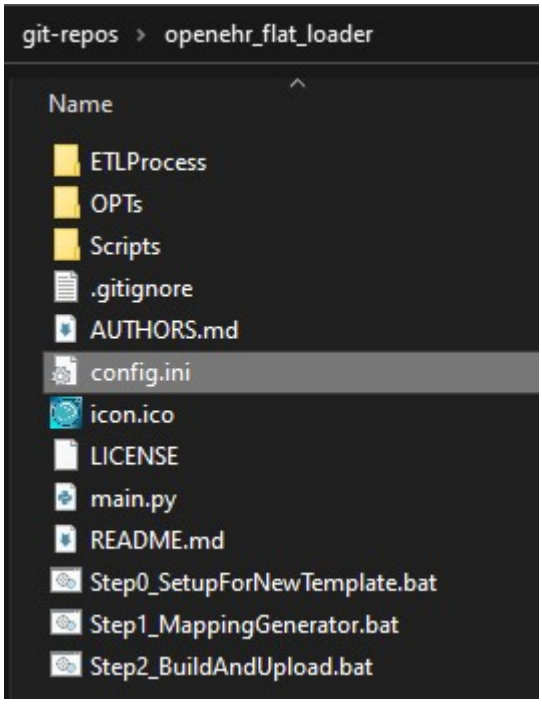
CSV:

- Provide data as CSV-File and copy CSV-File to ETLProcess/Input

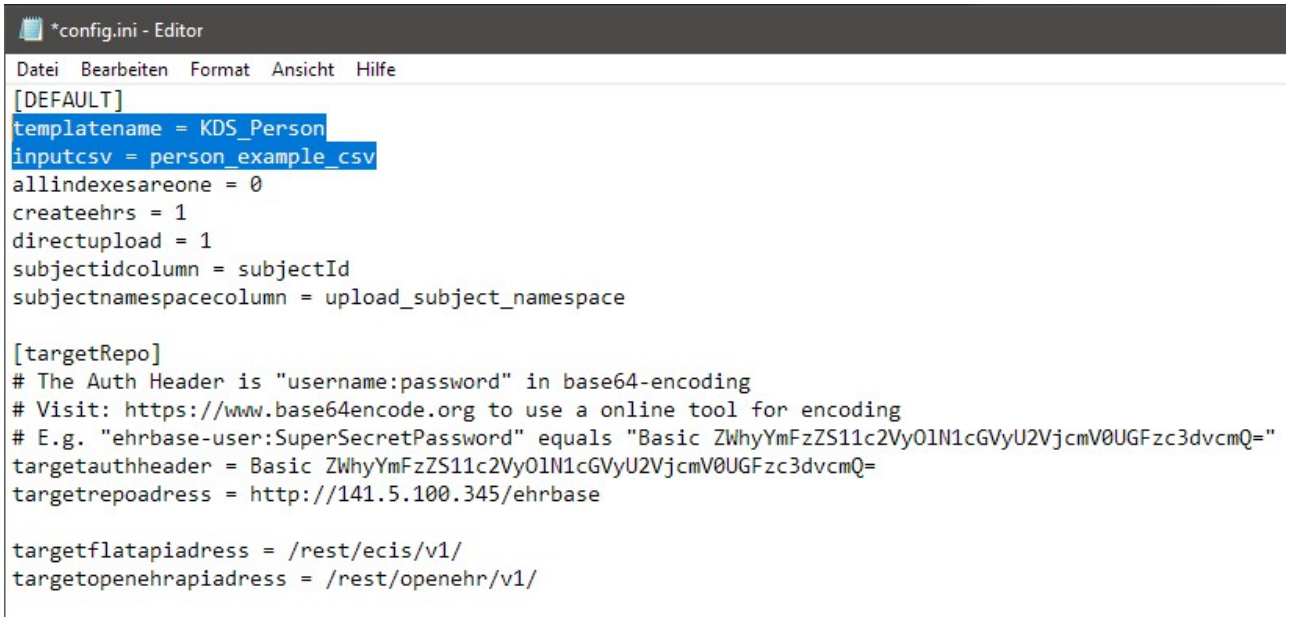


Config:

- Open config-File of the Flat-Loader



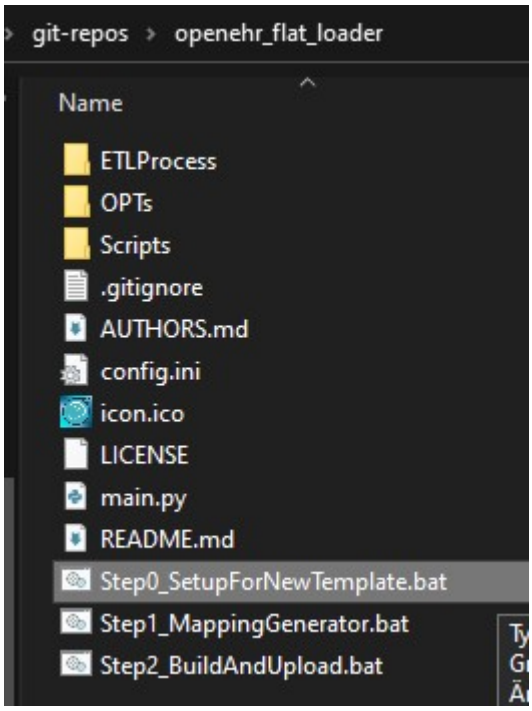
- Edit config-entries to match use case



Person Example

Setup OPT at Server:

- Run "Step0_SetupForNewTemplate.bat" to upload the OPT to your server (defined by the IP in the config)



- The Output may look like this:

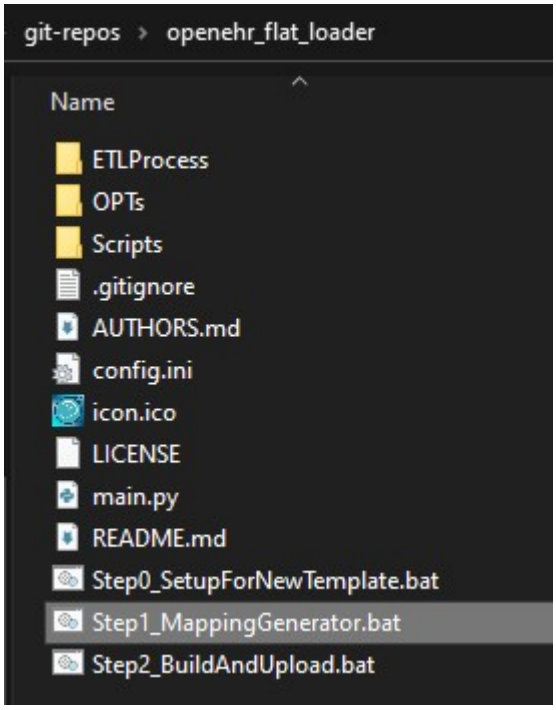
```
C:\WINDOWS\system32\cmd.exe
Used Argument: -generateExamples

HandleOPT is running:
OP Exists? 200
        OPT already exists at this server
        OPT exists at server and WebTemplate has been downloaded


OPT is uploaded to the Repository and an Example-Composition is stored in the ManualTasks-Folder.
```

Make sure necessary data columns are included in CSV:

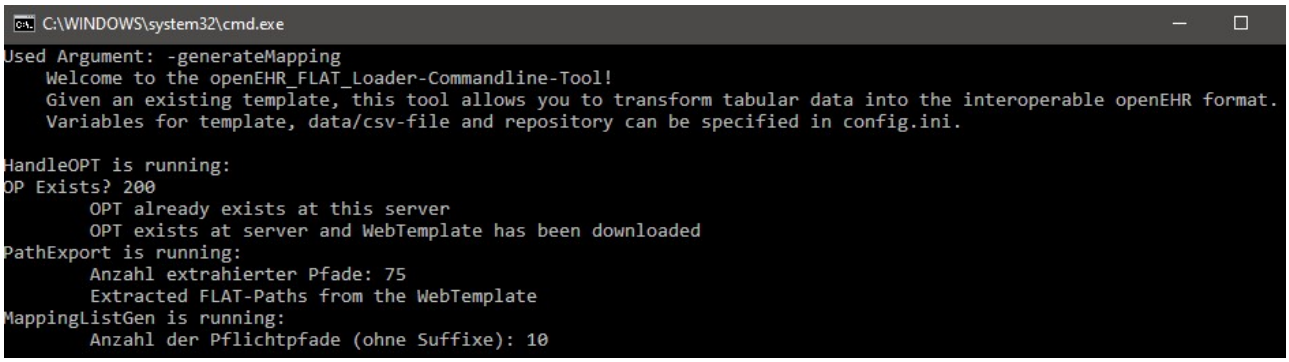
- Make sure the columns "ehrId" exists
- Make sure a subjectId-Column exists and is defined in config.
- Make sure a namespace-Column exists and is defined in config.

Generate the Excel-Mapping-Table:

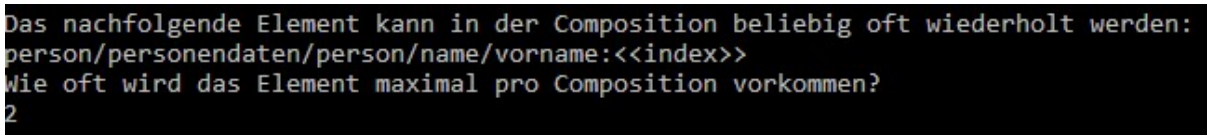
- Run "Step1_MappingGenerator.bat"



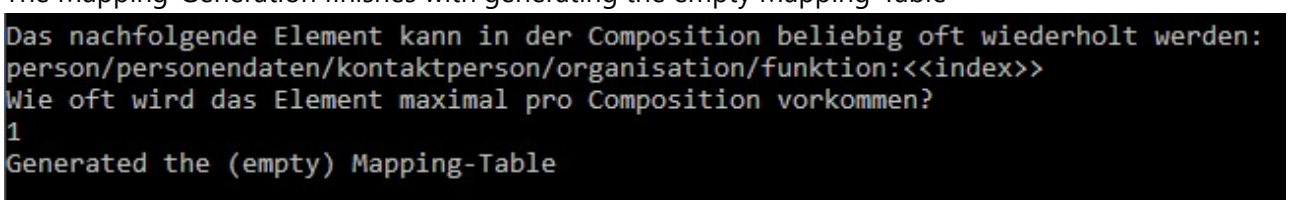
- The Output will give you some basic information about the FLAT-Paths that were extracted



- For some pathes the template takes several entries (cardinality != ([1..1] or [0..1])).
  - For this entries provide the number of maximum repetitions of the item in your data set.

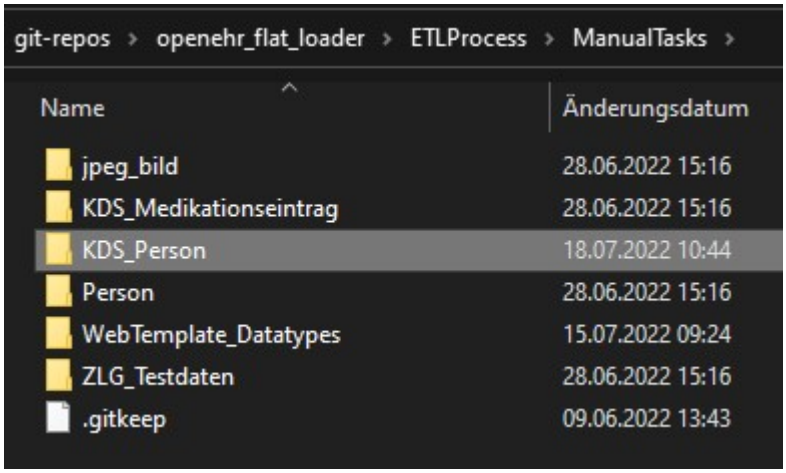


  - Setting the value to "0" will lead to the path not being included in the Mapping-Table
  - These so called indexes in the paths may be manipulated manually at a later time (setting 1 to have the path in the Mapping-Table is often fine).
- In the case of "KDS_Person"-Template all the repeatable items for "Straßenanschrift", "Postfach" and others are queried.
  - Enter either 0 or 1 for this example.
- The Mapping-Generation finishes with generating the empty Mapping-Table

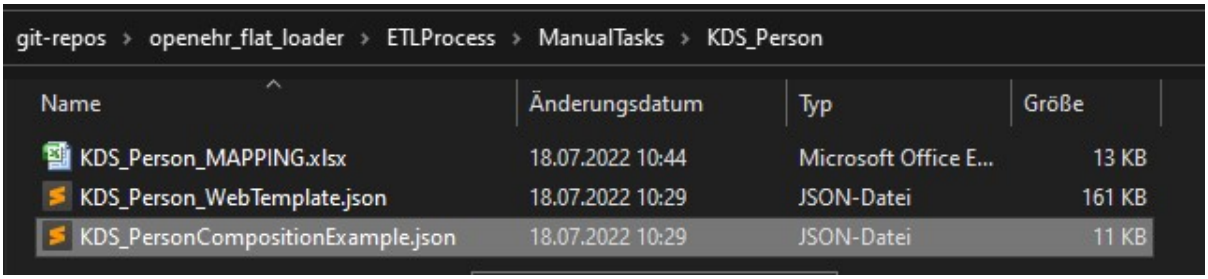


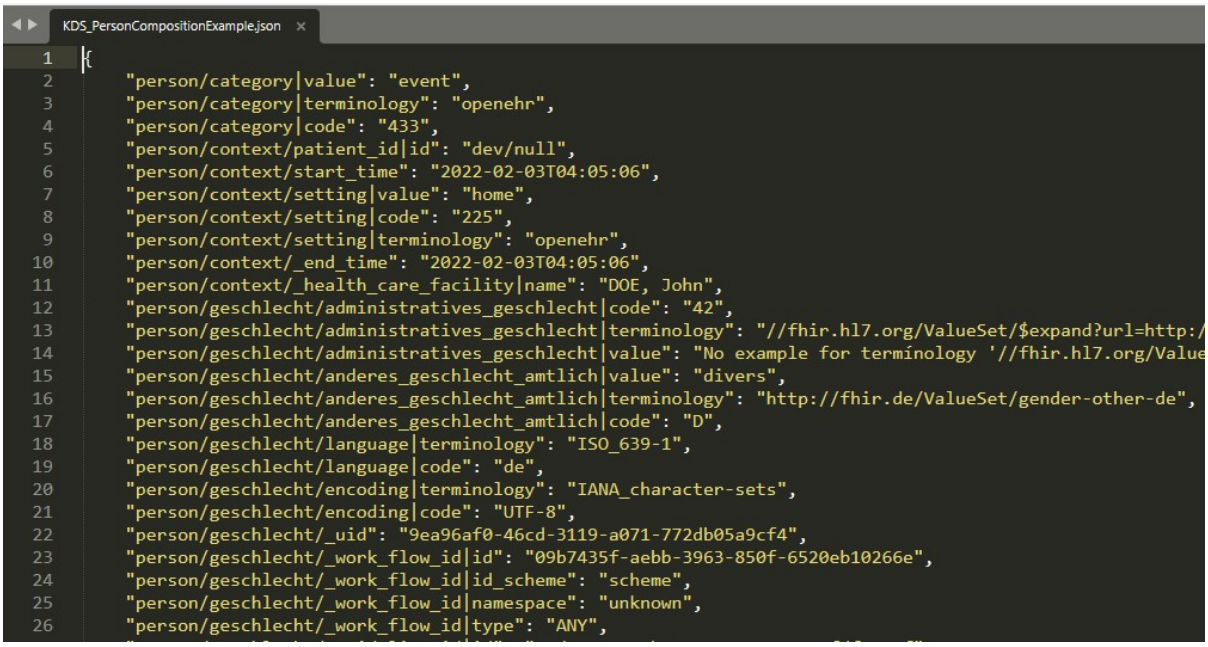
Perform the Mapping (and enrich the Data-CSV)

- You can find all related files under "openehr_flat_loader\ETLProcess\ManualTasks"



  - The example composition that has been generated by the server gives a good idea how FLAT-Paths/Items of the Template may be mapped to data items/values



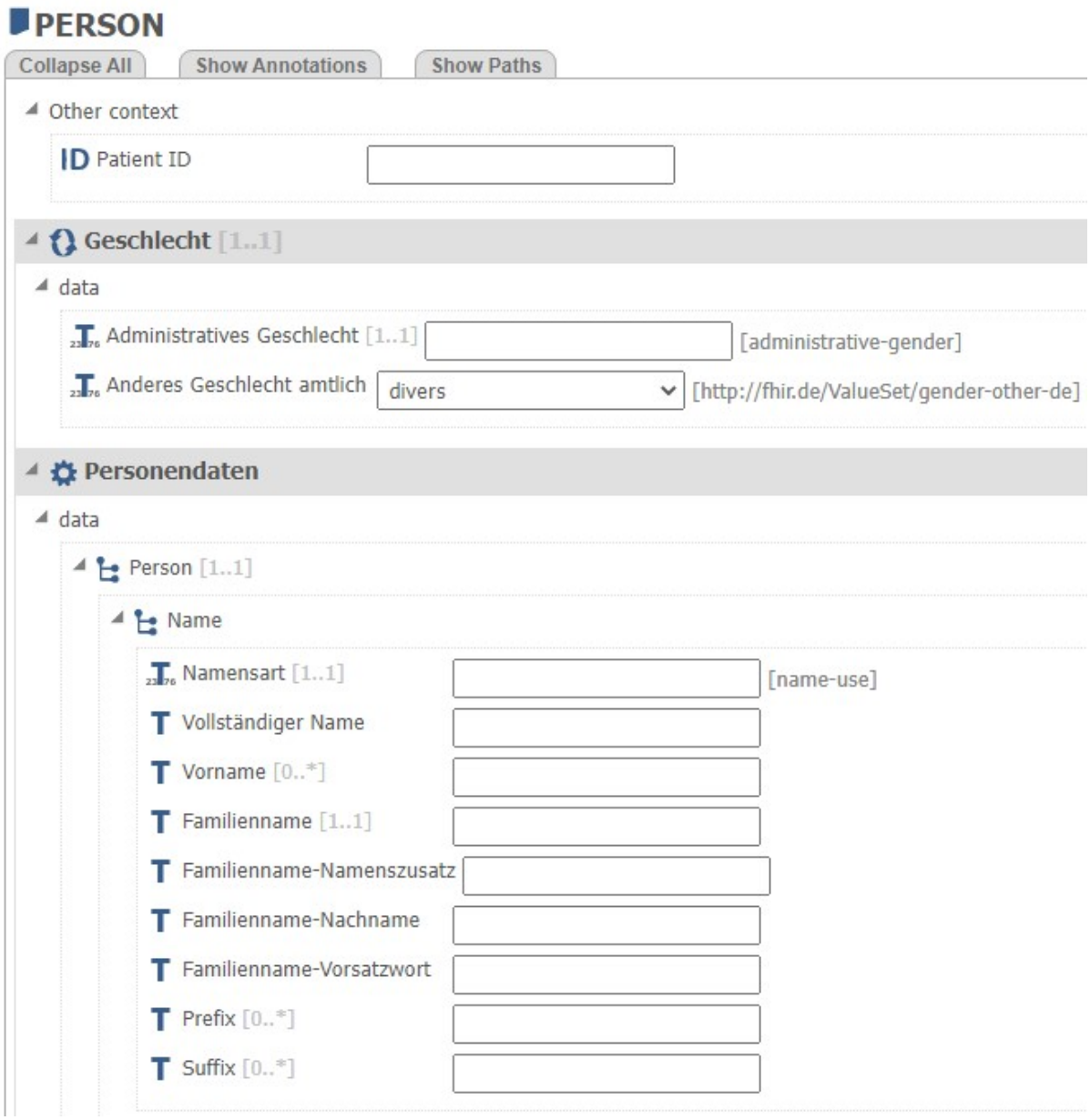


  - You can use the KDS_Person_WebTemplate.json to lookup paths and cardinalities as well as value sets.

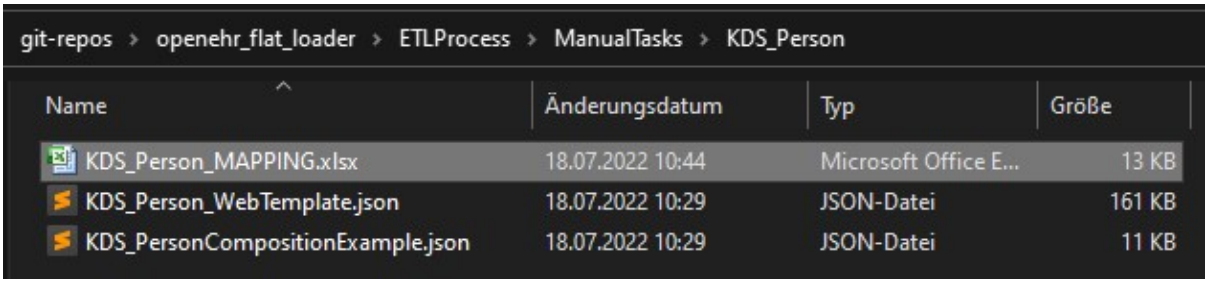
- The FLAT-Paths result from the tree-structure of the elements identified by their "id"

```
{
    "templateId": "KDS_Person",
    "version": "2.3",
    "defaultLanguage": "de",
    "languages": [
        "de"
    ],
    "tree": {
        "id": "person",
        "name": "Person",
        "localizedName": "Person",
        "rmType": "COMPOSITION",
        "nodeId": "openEHR-EHR-COMPOSITION.person.v0",
        "min": 1,
        "max": 1,
        "localizedNames": {
            "de": "Person"
        },
        "localizedDescriptions": {
            "de": "Dokument zur Übermittlung von Informationen über Personen"
        },
        "children": [
            {
                "id": "category",
                "name": "category",
```

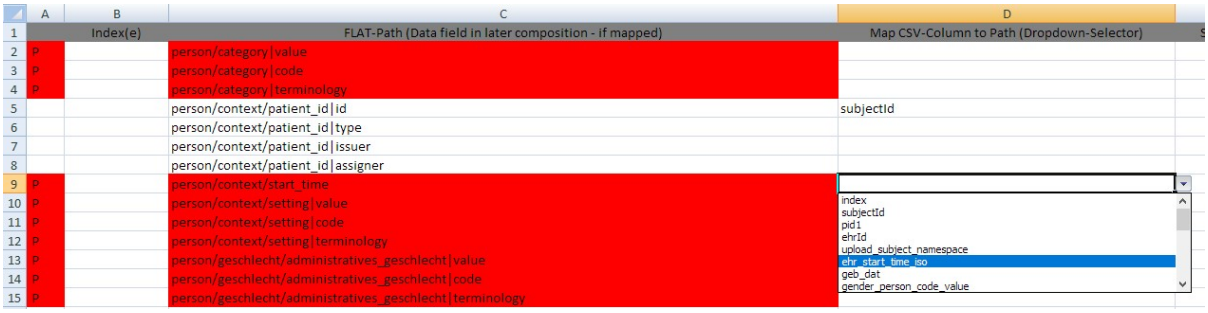
- You can lookup the structure, cardinalities and value sets online in the CKM



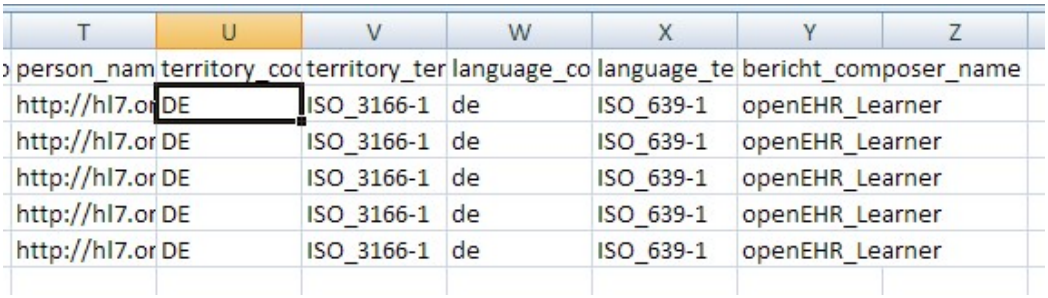
- With this documents as support you may be ready to perform the mapping task
  - Open the Mapping-Table in Excel



  - Select each CSV-Element that shall be mapped to a Path

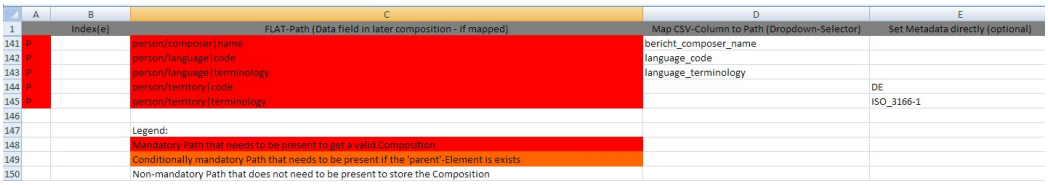


- Cells with RED-Colour are mandatory for the composition. (Sometimes they are filled by the server if left blank.)
- Cells with ORANGE-BROWN-Colour are mandatory if the element which they belong to is present.
- If you are missing a CSV-Column (e.g. you have a language but you are missing the termninology-name you have to add it in your CSV-File (per subject) or you have to set it in the Mapping via Colum E - "Set Metadata directly (optional)".
  Note that values given in Colum E are set for every subject/composition)
  - Metadata set per subject in the CSV:



	T	U	V	W	X	Y	Z
	person_nam	territory_coc	territory_ter	language_co	language_te	bericht_composer_name	
	http://hl7.or	DE	ISO_3166-1	de	ISO_639-1	openEHR_Learner	
	http://hl7.or	DE	ISO_3166-1	de	ISO_639-1	openEHR_Learner	
	http://hl7.or	DE	ISO_3166-1	de	ISO_639-1	openEHR_Learner	
	http://hl7.or	DE	ISO_3166-1	de	ISO_639-1	openEHR_Learner	
	http://hl7.or	DE	ISO_3166-1	de	ISO_639-1	openEHR_Learner	

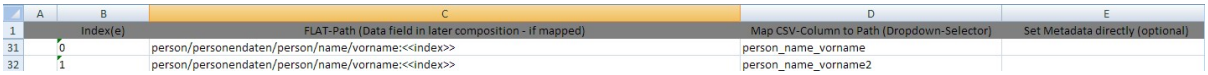
  - Metadata for territory set directly in the mapping for every subject but metadata for language taken from respective columns in the CSV:



	A	B	C	D	E
1		Index(e)	FLAT-Path (Data field in later composition - if mapped)	Map CSV-Column to Path (Dropdown-Selector)	Set Metadata directly (optional)
141	P		person/composer|name	bericht_composer_name	
142	P		person/language|code	language_code	
143	P		person/language|terminology	language_terminology	
144	P		person/territory|code		DE
145	P		person/territory|terminology		ISO_3166-1
146					
147			Legend:		
148			Mandatory Path that needs to be present to get a valid Composition		
149			Conditionally mandatory Path that needs to be present if the 'parent'-Element is exists		
150			Non-mandatory Path that does not need to be present to store the Composition		

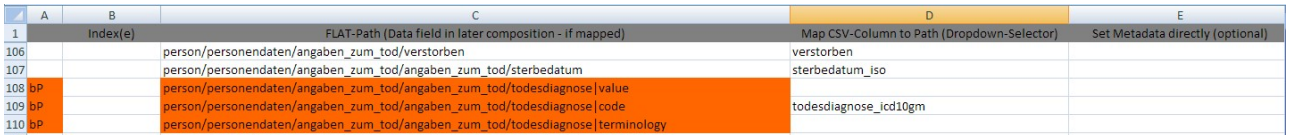
    - Note, setting the Values directly for a mandatory field should not result in problems, but setting values directly for non-mandatory field may result in compositions where for example a terminology name is defined but no termninology value or code is present.
- The entries that you need for Terminology-Names, Values from Terminologies, etc. can be looked up in the CKM, WebTemplate, Online, OpenEHR-Specification or the Example-Composition.

- If you want to duplicate a path you have to have an eye on the indexes
  - Make sure index 0 always exists in a data set before also mapping the path for index 1, otherwise the server cant store the resulting compositions.



	A	B	C	D	E
1		Index(e)	FLAT-Path (Data field in later composition - if mapped)	Map CSV-Column to Path (Dropdown-Selector)	Set Metadata directly (optional)
31		0	person/personendaten/person/name/vorname:<<index>>	person_name_vorname	
32		1	person/personendaten/person/name/vorname:<<index>>	person_name_vorname2	

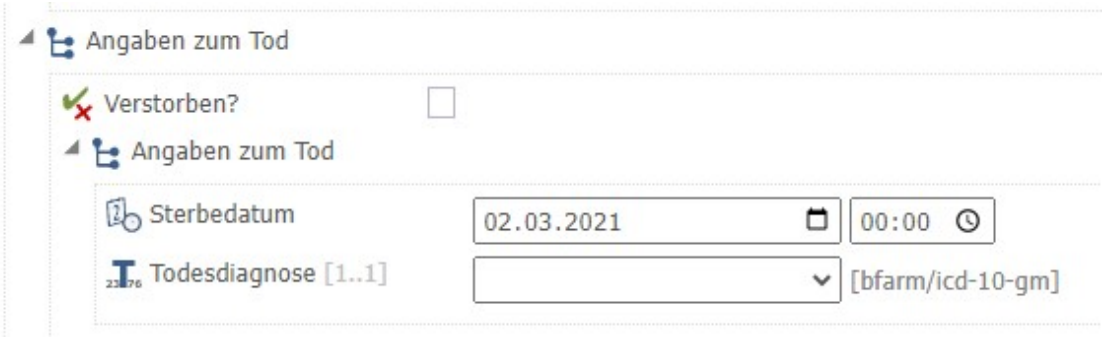
- Todesdiagnose is a great example for enriching of the CSV:



	A	B	C	D	E
1		Index(e)	FLAT-Path (Data field in later composition - if mapped)	Map CSV-Column to Path (Dropdown-Selector)	Set Metadata directly (optional)
106			person/personendaten/angaben_zum_tod/verstorben	verstorben	
107			person/personendaten/angaben_zum_tod/angaben_zum_tod/sterbedatum	sterbedatum_iso	
108	bP		person/personendaten/angaben_zum_tod/angaben_zum_tod/todesdiagnose|value		
109	bP		person/personendaten/angaben_zum_tod/angaben_zum_tod/todesdiagnose|code	todesdiagnose_icd10gm	
110	bP		person/personendaten/angaben_zum_tod/angaben_zum_tod/todesdiagnose|terminology		

  - The Mapping requires not only the code from the terminology but also its name and the corresponding value

- In the CKM the Template has "bfarm/icd-10-gm" defined as the name for our terminology



  - We add this as the name of the terminology to our CSV.
- For our subject/entry with "I22.1" as a code from this terminolgy we now lookup the Display-Value for this code (e.g. using Google).

- This results in the following entries for "Todesdiagnose"

S	T	U
todesdiagnose_icd10gm	todesdiagnose_terminology	todesdiagnose_value
I22.1	bfarm/icd-10-gm	Rezidivierender Myokardinfarkt der

  - Note that there is only one subject with entry about death. Therefore it is the right way to supply these metadata diretly in the CSV and not as Metadata set directly in the mapping, since subject 1+2+4+5 do not have a diagnosis and therefore should not have any metadata mapped about it.

> **Caution:**
> After enriching metadata newly added CSV-columns are not available in the mapping.
> You may enter the names directly.
>
> ! If you generate new mapping the existing file is overwritten - save your old mapping if needed !

Generating ressources and uploading them to the server

- Run "Step2_BuildAndUpload.bat"

  - At first the resources are generated and stored in the Output-Folder

```
BuildComp is running.
    Erstelle 5 Composition-Ressourcen.
    5 / 5 Ressourcen erstellt und im Ordner "Output" gespeichert.
    buildComp finished.
```

- After that the EHRs are created at the server

```
Create 5 EHRs:
    Hindernis beim EHR erstellen mit Status: 409
      EHR existierte bereits mit ehrID: 2ff4c284-a8b1-49e6-af0b-cf44a9a0b3ac

    Hindernis beim EHR erstellen mit Status: 409
      EHR existierte bereits mit ehrID: a80477cd-54d8-4552-a27d-2f3dd3d24ca9

    Hindernis beim EHR erstellen mit Status: 409
      EHR existierte bereits mit ehrID: 489ee219-012a-404b-b5d3-470a8a9ed6bb

    Hindernis beim EHR erstellen mit Status: 409
      EHR existierte bereits mit ehrID: b7aada04-2b38-4ec7-9c81-961b0565e4ee

    Hindernis beim EHR erstellen mit Status: 409
      EHR existierte bereits mit ehrID: b514ab79-78f9-4988-a942-37f733a16590

EHRs for 5 / 5 Subjects have been created successfully.
```

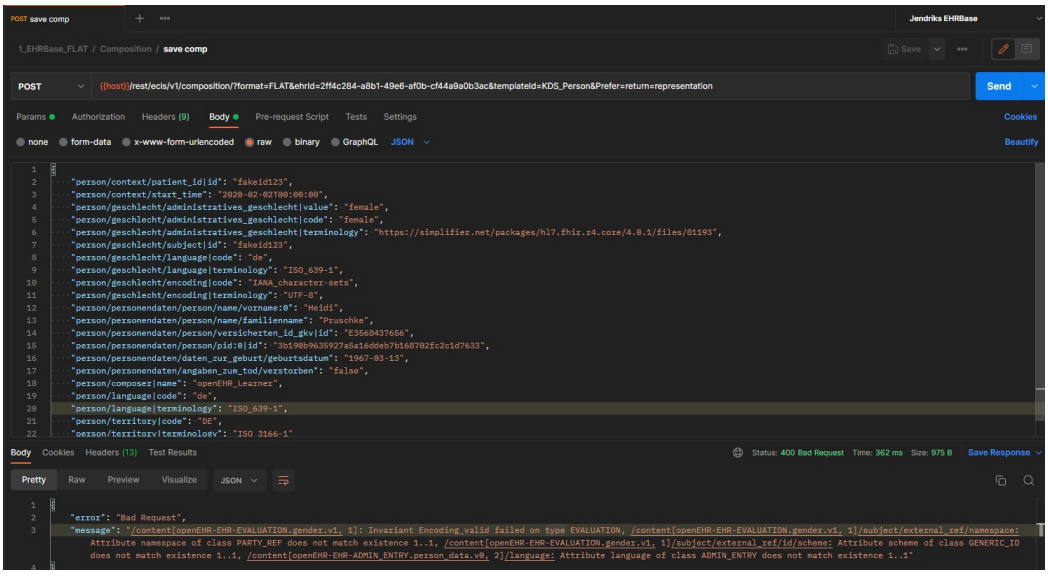
- At last the ressources are uploaded to the server

  - This step either suceeds and the ressources are available in the openEHR-Repo.

```
    Status beim Upload der Composition: 201
        {'meta': {'href': {'url': 'http://141.5.100.115/ehrbase/rest/ecis/v1/composition/?format=FLAT&ehrId=b514ab79-78f
9-4988-a942-37f733a16590&templateId=KDS_Person/rest/ecis/v1/composition/088a55d2-1e83-4111-9f7e-229f423051f2::local.ehrb
ase.org::1'}}, 'action': 'CREATE', 'compositionUid': '088a55d2-1e83-4111-9f7e-229f423051f2::local.ehrbase.org::1'}
        CompositionUid: 088a55d2-1e83-4111-9f7e-229f423051f2::local.ehrbase.org::1

5 / 5 Compositions have been created successfully.

Upload finished. Great Success.
```

  - Or you are presented an error message by the server that has to be dealt with.

    - The error messages may range from unvalid encodings or missing attributes in a composition to Internal Server Errors.
    - It is often a good process to select some composition that has been created and try to upload it via e.g. Postman until all mistakes in the data-CSV and Mapping are fixed.



      - The errors shown above were due to
        - the subjectID being incorrectly mapped to subject|id instead of subject|name
        - one of the language|terminolgy-Paths being mappend to CSV-Column "territory_terminology"

        - "namensart" being mapped for the geburtsname-elemment (person/personendaten/person/geburtsname) instead the name-element (person/personendaten/person/name)
    - Please note that after dealing with the errors that were returned with the next request to the server the validation process goes further and more errors may be returned by the openEHR-Repo.

Querying uploaded Compositions

- To access the data in the openEHR-Repository we need to send an AQL-Query via the REST-API
    - Postman is a nice tool to send REST-Request
    - To build an AQL-Request:
        - Select the Item in the Template that you want to query (e.g. Vorname + Nachname)
            - Get the paths and IDs of these items from the CKM
            - Then get the aqlPath for this field from the WebTemplate.

              Vorname:

              > Vorname-Path in CKM = at [openEHR-EHR-COMPOSITION.person.v0]/content[openEHR-EHR-ADMIN_ENTRY.person_data.v0]/data[at0001]/items[openEHR-EHR-CLUSTER.person.v1 and name/value='Person']/items[openEHR-EHR-CLUSTER.structured_name.v0 and name/value='Name']/items[at0002 and name/value='Vorname']

              > AQLPath in WebTemplate = /content[openEHR-EHR-ADMIN_ENTRY.person_data.v0]/data[at0001]/items[openEHR-EHR-CLUSTER.person.v1 and name/value='Person']/items[openEHR-EHR-CLUSTER.structured_name.v0 and name/value='Name']/items[at0002 and name/value='Vorname']/value

              Nachname:

              > Nachname = at [openEHR-EHR-COMPOSITION.person.v0]/content[openEHR-EHR-ADMIN_ENTRY.person_data.v0]/data[at0001]/items[openEHR-EHR-CLUSTER.person.v1 and name/value='Person']/items[openEHR-EHR-CLUSTER.structured_name.v0 and name/value='Name']/items[at0005 and name/value='Familienname']

              > AQLPath in WebTemplate = /content[openEHR-EHR-ADMIN_ENTRY.person_data.v0]/data[at0001]/items[openEHR-EHR-CLUSTER.person.v1 and name/value='Person']/items[openEHR-EHR-CLUSTER.structured_name.v0 and name/value='Name']/items[at0005 and name/value='Familienname']/value

- Remove the parts with the pattern "items[openEHR-EHR-CLUSTER.person.v1 and name/value='Person']" and substitute them with items[openEHR-EHR-CLUSTER.person.v1,'Person'] since this is how it works with the EHRBase
- Add a /value at the end of the AQL-Paths for vorname and nachname because the base-path gives the following element:

```
{
  "_type": "DV_TEXT",
  "value": "Hertmann"
}
```

- Resulting AQL-Query that need to be sent to `{{host}}/rest/openehr/v1/query/aql`

```
  SELECT
    DISTINCT c/uid/value as compId,
    e/ehr_status/subject/external_ref/id/value as subjectId,
    c/content[openEHR-EHR-
ADMIN_ENTRY.person_data.v0]/data[at0001]/items[openEHR-EHR-
CLUSTER.person.v1,'Person']/items[openEHR-EHR-
CLUSTER.structured_name.v0,'Name']/items[at0002]/value/value as vorname,
    c/content[openEHR-EHR-
ADMIN_ENTRY.person_data.v0]/data[at0001]/items[openEHR-EHR-
CLUSTER.person.v1,'Person']/items[openEHR-EHR-
CLUSTER.structured_name.v0,'Name']/items[at0005]/value as nachname
    FROM EHR e
    CONTAINS COMPOSITION c
        CONTAINS ADMIN_ENTRY a
            CONTAINS CLUSTER[openEHR-EHR-CLUSTER.person.v1]
```

- Result:

```
{
    "q": "\n    SELECT \n        DISTINCT c/uid/value as compId,\n
e/ehr_status/subject/external_ref/id/value as subjectId,\n
c/content[openEHR-EHR-
ADMIN_ENTRY.person_data.v0]/data[at0001]/items[openEHR-EHR-
CLUSTER.person.v1,'Person']/items[openEHR-EHR-
CLUSTER.structured_name.v0,'Name']/items[at0002]/value/value as vorname,\n
c/content[openEHR-EHR-
ADMIN_ENTRY.person_data.v0]/data[at0001]/items[openEHR-EHR-
CLUSTER.person.v1,'Person']/items[openEHR-EHR-
CLUSTER.structured_name.v0,'Name']/items[at0005]/value/value as nachname\n
FROM EHR e\n    CONTAINS COMPOSITION c \n        CONTAINS ADMIN_ENTRY a\n
CONTAINS CLUSTER[openEHR-EHR-CLUSTER.person.v1]\n    ",
    "columns": [
        {
            "path": "/uid/value",
            "name": "compId"
```

```
        },
        {
            "path": "/ehr_status/subject/external_ref/id/value",
            "name": "subjectId"
        },
        {
            "path": "/content[openEHR-EHR-
ADMIN_ENTRY.person_data.v0]/data[at0001]/items[openEHR-EHR-
CLUSTER.person.v1,'Person']/items[openEHR-EHR-
CLUSTER.structured_name.v0,'Name']/items[at0002]/value/value",
            "name": "vorname"
        },
        {
            "path": "/content[openEHR-EHR-
ADMIN_ENTRY.person_data.v0]/data[at0001]/items[openEHR-EHR-
CLUSTER.person.v1,'Person']/items[openEHR-EHR-
CLUSTER.structured_name.v0,'Name']/items[at0005]/value/value",
            "name": "nachname"
        }
    ],
    "rows": [
        [
            "2ff4c284-a8b1-49e6-af0b-cf44a9a0b3ac",
            "89a21f4c-ed1b-4e0d-877a-4d03c5f772f6::local.ehrbase.org::1",
            "fakeid123",
            "Sigmar",
            "Täsche"
        ],
        [
            "489ee219-012a-404b-b5d3-470a8a9ed6bb",
            "aef91da6-ecb6-40a6-aee0-53a653358539::local.ehrbase.org::1",
            "fakeid125",
            "Sigmar",
            "Täsche"
        ],
        [
            "a80477cd-54d8-4552-a27d-2f3dd3d24ca9",
            "3d252a4e-7a6f-4895-8152-3efa4b5e399c::local.ehrbase.org::1",
            "fakeid124",
            "Rudi",
            "Klapp"
        ],
        [
            "b514ab79-78f9-4988-a942-37f733a16590",
            "088a55d2-1e83-4111-9f7e-229f423051f2::local.ehrbase.org::1",
            "fakeid127",
            "Ferdi",
            "Hertmann"
        ],
        [
            "b7aada04-2b38-4ec7-9c81-961b0565e4ee",
            "73180ccb-3065-49fe-b9f8-ec39a4e1e7cc::local.ehrbase.org::1",
            "fakeid126",
            "Kristin",
```

```
            "Löffler"
        ]
    ]
}
```

- The result holds all the information that we queried:

> EHR-ID: "b7aada04-2b38-4ec7-9c81-961b0565e4ee",
> Composition-ID: "73180ccb-3065-49fe-b9f8-ec39a4e1e7cc::local.ehrbase.org::1",
> Subject-ID: "fakeid126",
> Vorname: "Kristin",
> Familienname: "Löffler"