

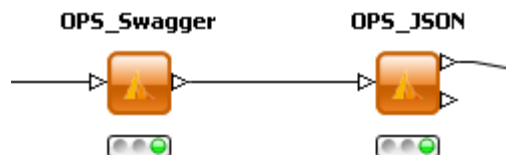


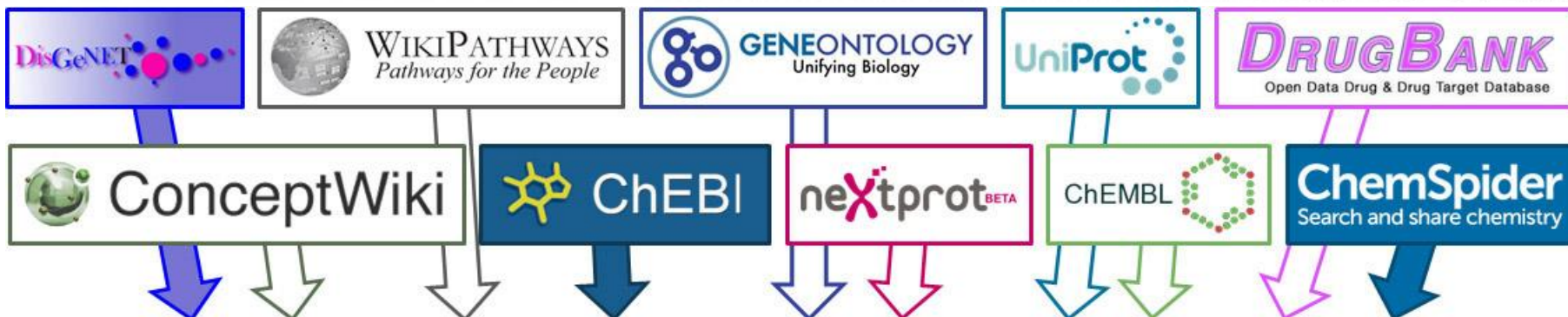
Creating workflows for drug-discovery with Open PHACTS and KNIME

Daniela Digles

support@openphacts.org

KNIME Spring Summit 2016 – Berlin






Open PHACTS

Physicochemical data

Molecular weight & formula

H-Bond acceptors / donors

Polar surface area, AlogP

Melting point

...and more

Identifiers

Synonyms

SMILES

InChI / InChIkey

ChemSpider ID

...and more

Pharmacological data

Activity type, value, concentration

Assay description

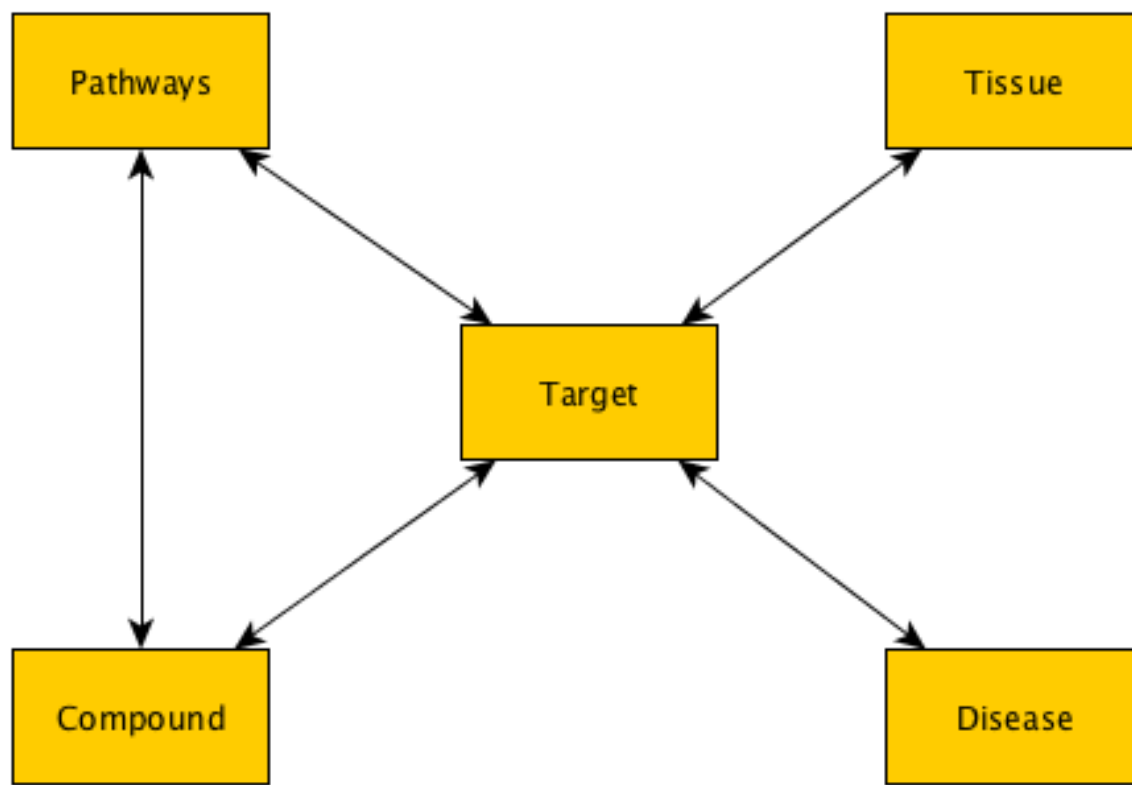
Target organism

Target name

...and more



Connections of different concept types





URL centred queries

- ❖ Examples:
 - <http://www.chemspider.com/2157>
 - <http://purl.uniprot.org/uniprot/Q9Y5Y9>
- ❖ See support portal for lists of queries available for each type of URL:
<http://support.openphacts.org/support/solutions/articles/4000037993-which-api-calls-can-i-use-starting-with->



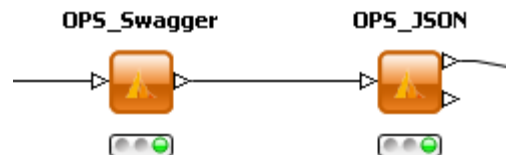
Useful links

- ❖ API: <https://dev.openphacts.org/>
- ❖ Support portal: <http://support.openphacts.org/>
- ❖ Open PHACTS Explorer: <http://explorer.openphacts.org/>
- ❖ Example Workflows: <http://www.myexperiment.org/groups/1125.html>



OPS-Knime nodes

- ❖ Originally created by Ronald Siebes, VU Amsterdam. Current developer: Evan Tzanis, QMUL London.
- ❖ No predefined set of nodes for each API call.
- ❖ OPS_Swagger:
 - creates the API call
 - Swagger file is used to automatically provide available API calls and parameters
- ❖ OPS_JSON (deprecated):
 - executes the API call
 - transforms the output into a flattened spreadsheet format
- ❖ available from <https://github.com/openphacts/OPS-Knime>





Installing the Open PHACTS KNIME nodes

- ❖ <https://github.com/openphacts/OPS-Knime>
- ❖ Download the latest version of the KNIME nodes:
 - Click on the zip file (currently the latest version is org.openphacts.utils.json_1.1.0.zip)
 - Click on Raw to start the Download (save it anywhere on your computer).
 - Unzip it into a folder called org.openphacts.utils.json_1.1.0 in the plugins folder of your KNIME installation.
- OR
- Rename to org.openphacts.utils.json_1.1.0.jar and place the file in the plugins folder of your KNIME installation
- ❖ Start KNIME



Swagger

- ❖ Structured format for the generation of API documentation.
(<https://helloverb.com/developers/swagger>)
- ❖ https://raw.githubusercontent.com/openphacts/OPS_LinkedDataApi/1.5.0/api-config-files/swagger.json

```
{
  "basePath": "https://beta.openphacts.org/1.3",
  "apiVersion": "v1.3",
  "apis": [
    {
      "path": "/compound" ,
      "operations": [
        {
          "httpMethod": "GET",
          "summary": "Compound Information" ,
          "description": "Information about a single compound.
          ...

          "group": "Compound" ,
          "parameters": [
            {
              "name": "uri" ,
              "description": "A compound URI. e.g.: http://www.conceptwiki.org/concept/38932552-111f-4a4e-a46a-4ed1
              "dataType": "string",
              "required": true,
              "paramType": "query"
            },
            {
```




<https://dev.openphacts.org/docs/1.5>

OpenPHACTS API v1.5

Compound Information

/compound **GET**

Description

Returns information about a single compound including (but not limited to): molecular weight, biotransformation, protein binding and toxicity.

...

PARAMETER	VALUE	DESCRIPTION
uri	<input type="text" value="(required)"/>	A compound URI. e.g.: http://www.conceptwiki.org/concept/38932552-111f-4a4e-a46a-4ed1d7bdf9d5
app_id	<input type="text"/>	Your access application id
app_key	<input type="text"/>	Your access application key
_format	<input type="text" value=""/>	The desired result format.

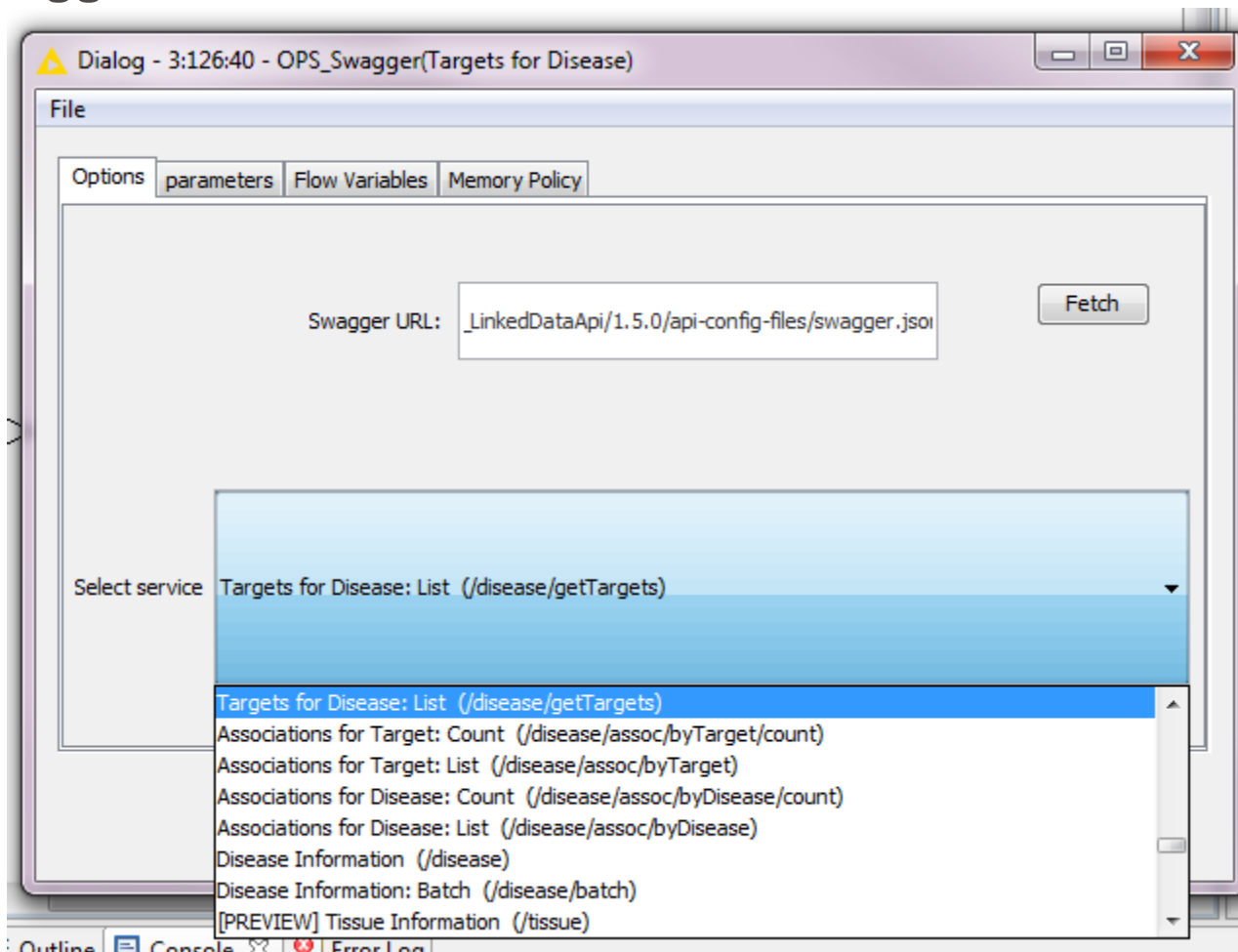


OPS_Swagger details

- ❖ Knime node where the user provides a url to a Swagger file (default: Open PHACTS API, v1.4)
- ❖ File is parsed and provides a list of the available API calls.
- ❖ Parameters tab is updated to the available parameters.
- ❖ Parameters can be set in the parameters tab or in the input table.
- ❖ Output of the node is an executable API call.



OPS_Swagger details



Dialog - 3:126:40 - OPS_Swagger(Targets for Disease)

File

Options parameters Flow Variables Memory Policy

Swagger URL: Fetch

Select service

- Targets for Disease: List (/disease/getTargets)
- Targets for Disease: List (/disease/getTargets)
- Associations for Target: Count (/disease/assoc/byTarget/count)
- Associations for Target: List (/disease/assoc/byTarget)
- Associations for Disease: Count (/disease/assoc/byDisease/count)
- Associations for Disease: List (/disease/assoc/byDisease)
- Disease Information (/disease)
- Disease Information: Batch (/disease/batch)
- [PREVIEW] Tissue Information (/tissue)

Outline Console Error Log



OPS_Swagger details

Dialog - 3:126:40 - OPS_Swagger(Targets for Disease)

File

Options parameters Flow Variables Memory Policy

Default URL parameters

uri	<input type="text"/>
app_id	15a18100
app_key	i272f1cd961d215f318a0315dd3d
_page	<input type="text"/>
_pageSize	all
_orderBy	<input type="text"/>
_format	<input type="text"/>
_callback	<input type="text"/>
_metadata	<input type="text"/>

OK Apply Cancel ?



OPS_Swagger details

Chunked input - 3:126:65 - Chunk Loop Start

File

Table "default" - Rows: 1 | Spec - Column: 1 | Properties | Flow Variables

Row ID	s uri
Row0	http://linkedlifedata.com/resource/umls/id/C0030567

OPS_Swagger

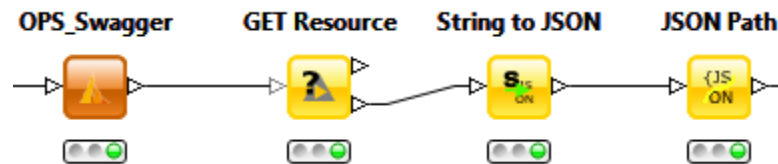


executable API call



Parsing the results

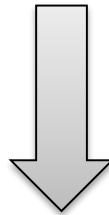
- ❖ Either use OPS_Json (deprecated) or the REST and JSON nodes available as add-in from KNIME.
- ❖ GET Resource: retrieves the actual data from the server. Configure the node to use the column url as input. Response representation cell type: Autodetection.
- ❖ String to JSON: transforms the result to a JSON column type.
- ❖ JSON Path: allows the individual selection of the data which is transformed into a tabular structure.





JSON path details

```
{
  "format": "linked-data-api",
  "version": "1.5",
  "result": {
    "_about": "https://beta.openphacts.org/1.5/compound/pharmacology/count?app_id=15a18100&app_key=52",
    "definition": "https://beta.openphacts.org/api-config",
    "extendedMetadataVersion": "https://beta.openphacts.org/1.5/compound/pharmacology/count?app_id=15a18100&app_key=52",
    "linkPredicate": "http://www.w3.org/2004/02/skos/core#exactMatch",
    "activeLens": "Default",
    "primaryTopic": {
      "_about": "http://rdf.ebi.ac.uk/resource/chembl/molecule/CHEMBL1945801",
      "compoundPharmacologyTotalResults": 4,
      "isPrimaryTopicOf": "https://beta.openphacts.org/1.5/compound/pharmacology/count?app_id=15a18100&app_key=52"
    }
  }
}
```



Row ID	S count	S url
Row 1	4	https://beta.openphacts.org/1.5/compound/pharmacology/count?app_id=15a18100&app_key=52



JSON path configuration

Dialog - 0:106:121 - JSON Path

File

Settings | Flow Variables | Memory Policy

Input
JSON Representation

Remove source column

Outputs

Output column	JSONPath	List	Paths
S count	<code>[\$[result][primaryTopic][compoundPharmacologyTotalResults]]</code>	<input type="checkbox"/>	<input type="checkbox"/>
S url	<code>[\$[result][_about]]</code>	<input type="checkbox"/>	<input type="checkbox"/>

Add single query Add collection query Add JSONPath Edit JSONPath Remove JSONPath

JSON-Cell Preview

```
1 {
2   "format" : "linked-data-api",
3   "version" : "1.5",
4   "result" : {
5     "_about" : "https://beta.openphacts.org/1.5/compound/pharmacology/count?app_id=15a18100&app_key=528a8272f1cd961d",
6     "definition" : "https://beta.openphacts.org/api-config",
7     "extendedMetadataVersion" : "https://beta.openphacts.org/1.5/compound/pharmacology/count?app_id=15a18100&app_key=528a8272f1cd961d",
8     "linkPredicate" : "http://www.w3.org/2004/02/skos/core#exactMatch",
9     "name" : "Pharmacology Count",
10    "url" : "https://beta.openphacts.org/1.5/compound/pharmacology/count?app_id=15a18100&app_key=528a8272f1cd961d"
11  }
```

OK Apply Cancel ?



Example JSON path queries

- ❖ To easily generate a query, click on the wanted property in the JSON-Cell Preview and click on Add single query. If the data is actually a list, and you want to retrieve all entities, click on Add collection query instead.
- ❖ Simple query:
 - `$['result']['primaryTopic']['compoundPharmacologyTotalResults']`
 - OR
 - `$.result.primaryTopic.compoundPharmacologyTotalResults`
 - OR
 - `$..compoundPharmacologyTotalResults`
- ❖ Be aware that the path might change depending on the used query. The last version is therefore the preferred one.



Example JSON path queries – advanced queries

- ❖ Retrieving one property, while filtering for another one
 - Example for compound classification API call: retrieve the labels of the classification, but only when the classification is of the type “has role”.

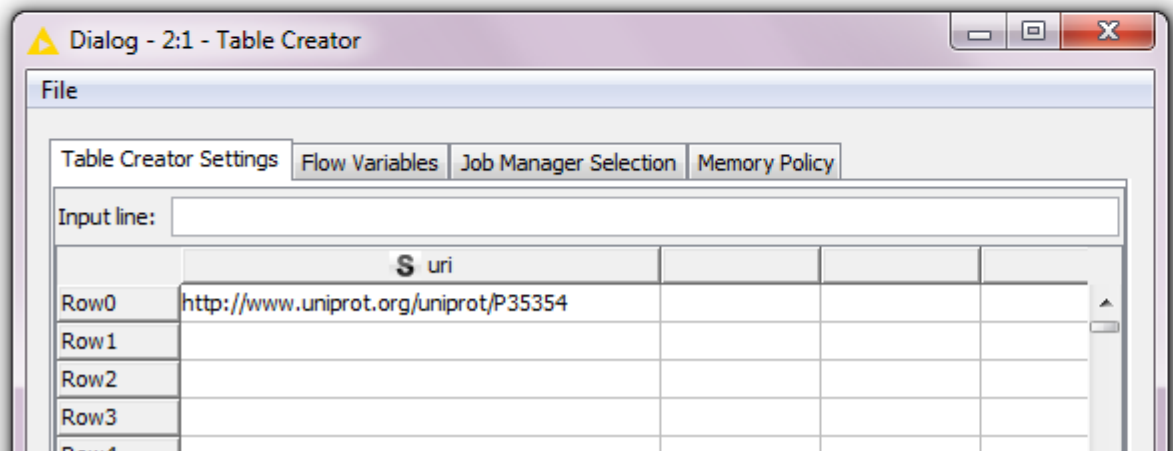
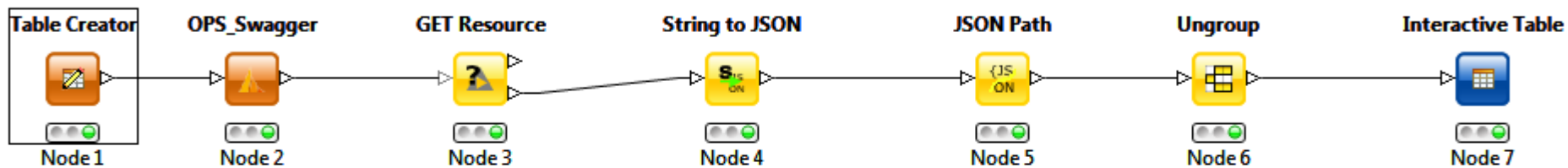
```
$..hasChebiClassification[?(@.classificationType.prefLabel=='has role')].prefLabel
```

- Example for any API call returning data from Concept Wiki: retrieves the URI from Conceptwiki.

```
$..[?(@.inDataset=='http://www.conceptwiki.org')]._about
```

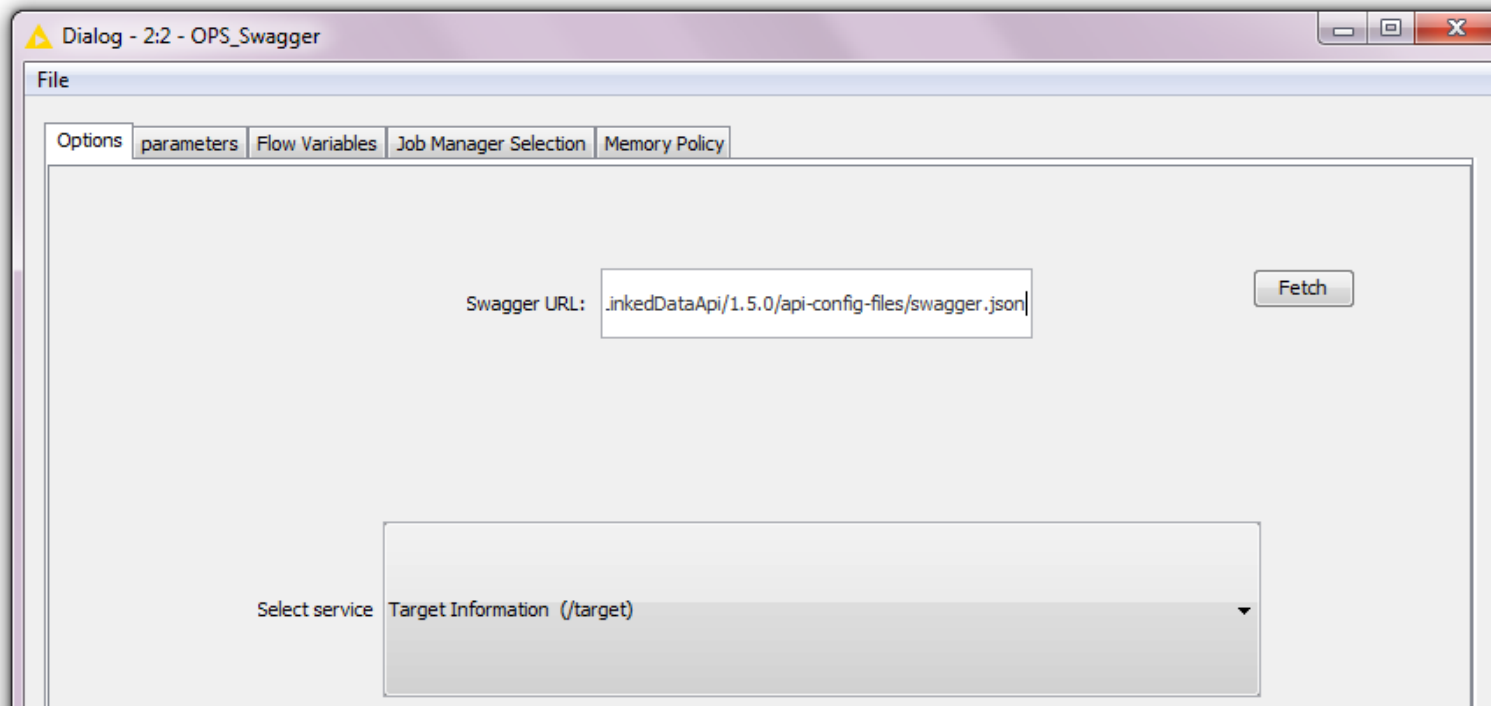
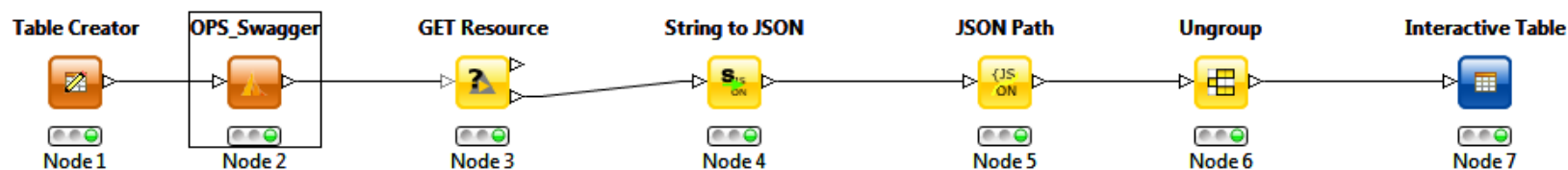


Example 1: Target information workflow



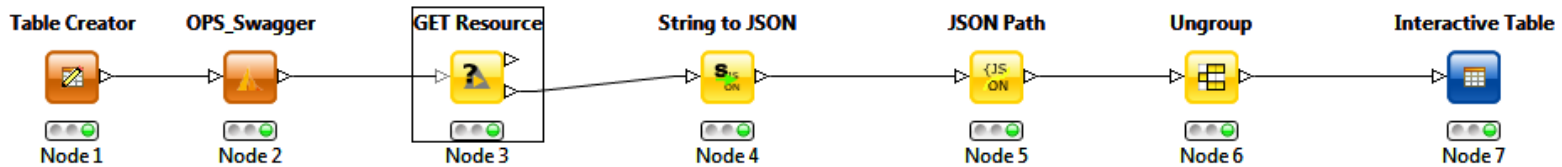


Example 1: Target information workflow





Example 1: Target information workflow



Dialog - 2:3 - GET Resource

File

Base Settings | Header Settings | Flow Variables | Job Manager Selection | Memory Policy

URL

Manual Input:

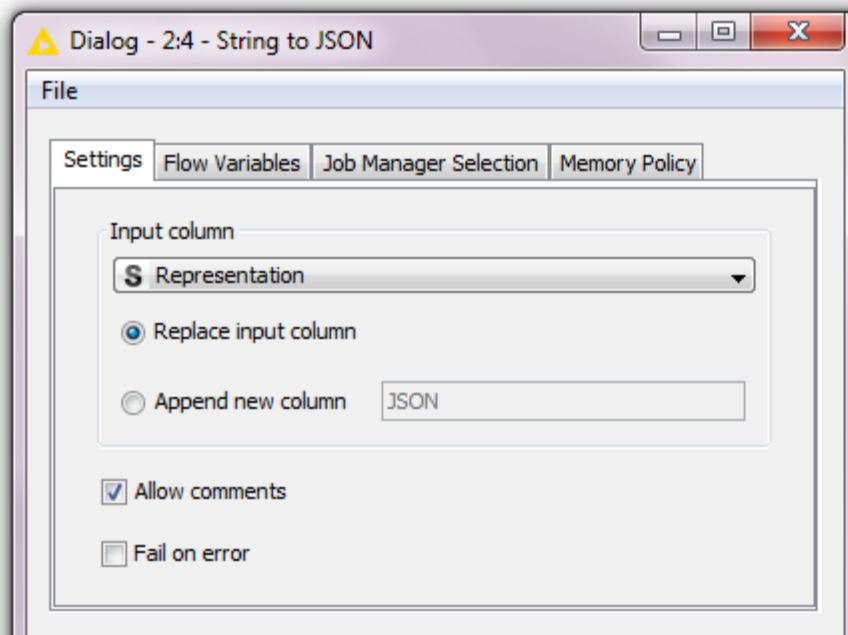
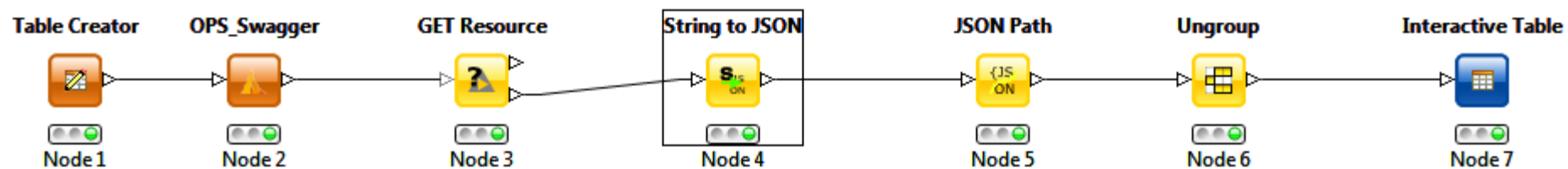
Column Selection:

Response Body Settings

Response representation cell type:

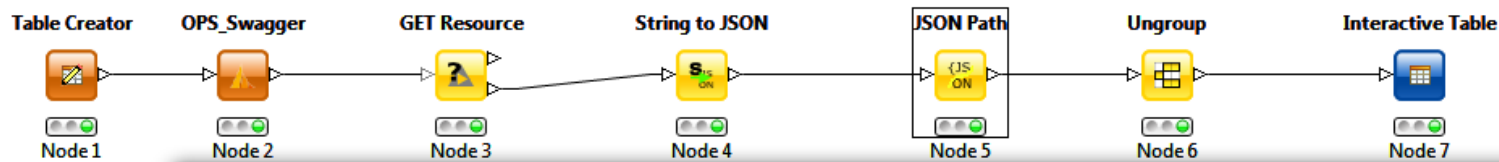


Example 1: Target information workflow





Example 1: Target information workflow



Dialog - 2:5 - JSON Path

File

Settings | Flow Variables | Job Manager Selection | Memory Policy

Input
 {JSON} Representation

Remove source column

Outputs

Output column	JSONPath	List	Paths
S prefLabel_en	\$.[prefLabel_en]	<input type="checkbox"/>	<input type="checkbox"/>
S Function_Annotation	\$.[Function_Annotation]	<input type="checkbox"/>	<input type="checkbox"/>
S genericName_en	\$(result)[primaryTopic][exactMatch][*][targetForDrug][*][genericName_en]	<input checked="" type="checkbox"/>	<input type="checkbox"/>

Add single query | Add collection query | Add JSONPath | Edit JSONPath | Remove JSONPath

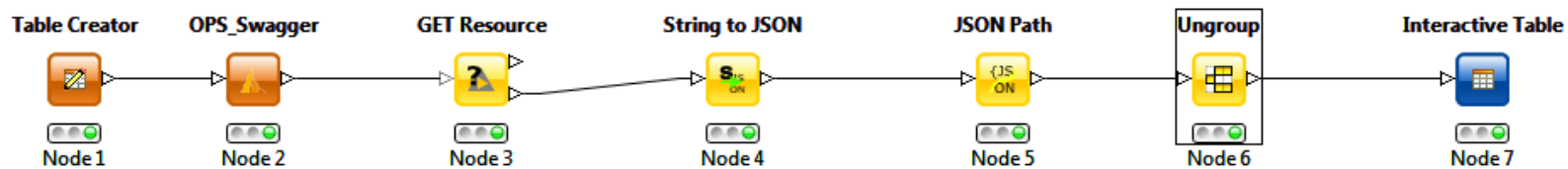
JSON-Cell Preview

```

    $(result)[primaryTopic][exactMatch][*][targetForDrug][*][genericName_en]
  16   "inDataset" : "http://www.openphacts.org/bio2rdf/drugbank",
  17   "targetForDrug" : [ {
  18     "_about" : "http://bio2rdf.org/drugbank:DB00605",
  19     "drug_type_en" : "approved [drugbank_resource:Approved]",
  20     "drug_type" : "approved [drugbank_resource:Approved]",
  21     "genericName_en" : "Sulindac",
  22     "specificName" : "Sulindac"
  
```



Example 1: Target information workflow



Dialog - 2:6 - Ungroup

File

Options | Flow Variables | Job Manager Selection | Memory Policy

Collection columns

Manual Selection Wildcard/Regex Selection

Exclude

Column(s): Search

Select all search hits

Enforce exclusion

Select

add >>

add all >>

<< remove

<< remove all

Include

Column(s): Search

Select all search hits

(...) genericName_en

Enforce inclusion

Additional Settings

Remove selected collection column Skip missing values Enable hilighting



Example 1: Target information workflow

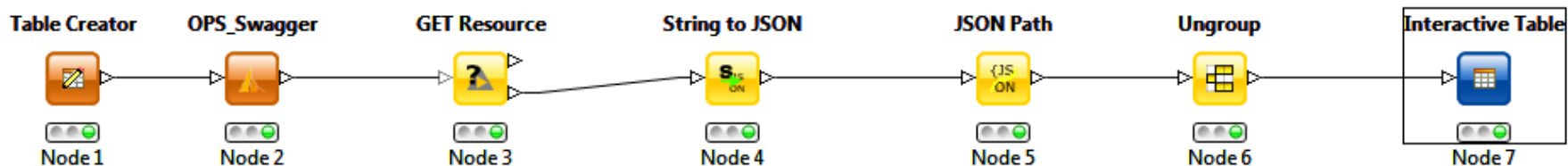


Table View - 2:7 - Interactive Table

File Hilite Navigation View Output

Row ID	S prefLabel_en	S Function_Annotation	S genericName_en
Row 1_1	Prostaglandin G/H synthase 2 (Homo sapiens)	Converts arachidonate to pro...	Sulindac
Row 1_2	Prostaglandin G/H synthase 2 (Homo sapiens)	Converts arachidonate to pro...	Niflumic Acid
Row 1_3	Prostaglandin G/H synthase 2 (Homo sapiens)	Converts arachidonate to pro...	Acetylsalicylic acid
Row 1_4	Prostaglandin G/H synthase 2 (Homo sapiens)	Converts arachidonate to pro...	Antrafenine
Row 1_5	Prostaglandin G/H synthase 2 (Homo sapiens)	Converts arachidonate to pro...	Indomethacin
Row 1_6	Prostaglandin G/H synthase 2 (Homo sapiens)	Converts arachidonate to pro...	Ginseng
Row 1_7	Prostaglandin G/H synthase 2 (Homo sapiens)	Converts arachidonate to pro...	Tiaprofenic acid
Row 1_8	Prostaglandin G/H synthase 2 (Homo sapiens)	Converts arachidonate to pro...	1-Phenylsulfonamide-3-Trifluoromethyl-5...
Row 1_9	Prostaglandin G/H synthase 2 (Homo sapiens)	Converts arachidonate to pro...	Flufenamic Acid
Row 1_10	Prostaglandin G/H synthase 2 (Homo sapiens)	Converts arachidonate to pro...	Licofelone
Row 1_11	Prostaglandin G/H synthase 2 (Homo sapiens)	Converts arachidonate to pro...	Aminosalicylic Acid
Row 1_12	Prostaglandin G/H synthase 2 (Homo sapiens)	Converts arachidonate to pro...	Mesalazine
Row 1_13	Prostaglandin G/H synthase 2 (Homo sapiens)	Converts arachidonate to pro...	Acetaminophen
Row 1_14	Prostaglandin G/H synthase 2 (Homo sapiens)	Converts arachidonate to pro...	Ketorolac
Row 1_15	Prostaglandin G/H synthase 2 (Homo sapiens)	Converts arachidonate to pro...	Resveratrol



General remarks

- ❖ Paginated API calls (with List in the name) return only 10 items as a default. To get all results:
 - Use `_pageSize = all`
 - Use the exact number of items (retrieved from the corresponding count API call)
 - Loop through the pages
- ❖ If no data is found, a 404 error is returned (exception: structure API calls return 500). Please take care that your workflow does not fail in such an event.
- ❖ Depending on the source of your input URI, the structure of the result JSON might be slightly different. Try using general queries, rather than exact paths.



Example 2: Compound classification

- ❖ Connection to Example 1: Retrieve compound URIs for DrugBank compounds in JSON Path.
 - Select drugbank:DB URI, click on add collection query, rename output column to uri.
- ❖ Add Column filter node after the Ungroup node, keep uri column only.
- ❖ Use Chunk loop start to get data for all compounds.
- ❖ Add OPS_Swagger node after the Column filter node, select Compound Classification from the dropdown.
- ❖ Add GET Resource, String to JSON and JSON Path nodes
- ❖ JSON Path: add preferred label. To get all compound classifications with “has_role” definition:
 - Click add JSONPath
 - `$.hasChebiClassification[?(@.classificationType.prefLabel=='has role')].prefLabel`
 - Activate “List”
- ❖ Loop End



Do you want to stay in contact?

- ❖ Newsletter at <http://www.openphactsfoundation.org/>
- ❖ Forum at <http://support.openphacts.org/support/home>
- ❖ E-mail to support@openphacts.org or info@openphactsfoundation.org