OPDDR Knowledge Network (KN), Beta Version

AIM: The Knowledge Network (KN) demonstrates how a semantic version of the Open Phenotypic Drug Discovery Resource (OPDDR) data can be used to identify relationships between National Pharmaceutical Collection (NPC) compounds, phenotypic assays, ontological classes of assays, and associated public data on related gene targets and pathways.

Introduction

This initial version of the KN is intended to provide a clear and easily comprehensible first step of describing the OPDDR compounds and assays in accordance with standardized community ontologies and namespaces, and relating these to protein targets derived from ChEMBL, and pathways derived from Reactome. Biological networks can be extremely complex, and many further entity classes could be integrated in future.

Ontologies Used

The KN uses the following ontologies:

PubChem RDF[1]	Primary reference for this project. Mainly because assays and substances have been deposited into PubChem. <http://rdf.ncbi.nlm.nih.gov/pubchem/>
BAO[2]	Bioassay classification. Initially using a minimal set based on annotation template provided by AstraZeneca. Only bao_vocabulary_assay.owl is required currently. <http://www.bioassayontology.org/bao#>
OBO[3]	Open Biological and Biomedical Ontologies BFO = Basic Formal Ontology <http://purl.obolibrary.org/obo/>
ChEMBL RDF[4]	ChEMBL, Reactome, Uniprot RDF downloads available. CCO = ChEMBL Core Ontology <http://rdf.ebi.ac.uk/terms/chembl#>
SIO[5]	Semanticscience Integrated Ontology <http://semanticscience.org/resource/>

[1] https://pubchem.ncbi.nlm.nih.gov/rdf/
[2] http://bioassayontology.org/
[3] http://www.obofoundry.org/
[4] https://www.ebi.ac.uk/rdf/
[5] http://semanticscience.org/

Entities:

entity, namespace	abbr	example
substance <http://rdf.ncbi.nlm.nih.gov/pubchem/substance/>	substance	SID124893119
compound <http://rdf.ncbi.nlm.nih.gov/pubchem/compound/>	compound	CID1131
assay <http://rdf.ncbi.nlm.nih.gov/pubchem/bioassay/>	bioassay	AID1117354
measuregroup <http://rdf.ncbi.nlm.nih.gov/pubchem/measuregroup/>	measureg	AID1117354
protein <http://rdf.ncbi.nlm.nih.gov/pubchem/protein/>	protein	GI124375976
target <http://rdf.ebi.ac.uk/resource/chembl/target/>	target	CHEMBL3038470
targetcomponent <http://rdf.ebi.ac.uk/resource/chembl/targetcomponent/>	target_cmpt	CHEMBL_TC_1927
UniprotRef (protein) <http://rdf.ebi.ac.uk/terms/chembl#UniprotRef>	uniprot	P53350
ReactomeRef (pathway) <http://rdf.ebi.ac.uk/terms/chembl#ReactomeRef>	reactome	REACT_1016

Note that PubChem *compounds* are required in addition to *substances*. Compounds refer to canonically defined and identifiable chemical entities which can be linked across databases; Substances refer to specific samples of compounds as provided by a supplier. We thus include both, to be as comprehensive and specific as possible.

KN Statistics

type	count	notes
substance	2511	PubChem SIDs
compound	2511	PubChem CIDs
assay	35	PubChem AIDs. Summary AID is 36th.
measuregroup	35	PubChem AIDs. Default for assay.

May 31, 2015

targets	4977	ChEMBL IDs. All single-component.
protein	4977	A.k.a. target component. With UniprotRefs.
protein activity	584,157	From ChEMBL, but includes PubChem data.
PD2 activity	5320	All “ACTIVE” outcomes from results.
assay classifications	155	Manually curated PD2 to BAO associations. Exported from worksheet.
pathway	*?*	*Reactome IDs.*
pathway-protein associations	*?*	*From ChEMBL, via Sparql or download.*

Asserted triplets, patterns and examples

description	examples
assay class membership	bioassay:AID1117354 rdf:type bao:BAO_0000015
assay title	bioassay:AID1117354 dcterms:title "human JAK2 kinase inhibition-screen"@en
NCGC substance to PubChem substance	ncats_sample:NCGC00182710-02 skos:exactMatch substance:SID144206486 .
substance to measure group association	substance:SID124882766 obo:BFO_0000056 measureg:AID1117326
substance to compound association	substance:SID124893119 sio:CHEMINF_000477 compound:CID1131
substance to PD2 assay asssociation	ncgc_substance:NCGC00091195-08 opddr:activeInAssay oidd_assay:17
OIDD assay ID to AID association	oidd_assay:17 skos:exactMatch bioassay:AID1117350
ChEMBL target to UniProt	chembl_target:CHEMBL5464 cco:targetXref uniprot:Q13546

NOTE: provisional namespaces defined for this beta version, subject to change:

- oidd_assay: <http://openinnovation.lilly.com/bioassay#>
- ncgc_substance: <http://ncats.nih.gov/ncgc/substance#>
- opddr: <http://ncats.nih.gov/opddr/>

Files:

The following files comprise this release.

file	description
bao_vocabulary_assay.owl	BAO module with bioassay class hierarchy.
chembl_activity.ttl	ChEMBL activity associations on protein targets.
chembl_target.ttl	ChEMBL protein targets.
chembl2uniprot.ttl	ChEMBL to UniProt target mappings.
npcpd2_bao_annotations.ttl	Manually curated BAO classifications.
npcpd2.ttl	PD2 activity associations.
oidd2pubchem_assay.ttl	OIDD to PubChem assay ID mapping.
ncats2pubchem_substance.ttl	NCATS to PubChem substance ID mapping.
pubchem_pd2_assay.ttl	PubChem RDF, includes titles, measure group associations.
pubchem_pd2_substance.ttl	PubChem RDF, includes CIDs, measure groups, AIDs.

Example Sparql Queries:

Prologue

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX obo: <http://purl.obolibrary.org/obo/>
PREFIX sio: <http://semanticscience.org/resource/>
PREFIX oidd_assay: <http://openinnovation.lilly.com/bioassay#>
PREFIX ncats_sample: <http://rdf.ncats.nih.gov/ncgc/sample/>
PREFIX opddr: <http://rdf.ncats.nih.gov/opddr/>
PREFIX bao: <http://www.bioassayontology.org/bao#>
PREFIX bioassay: <http://rdf.ncbi.nlm.nih.gov/pubchem/bioassay/>
PREFIX substance: <http://rdf.ncbi.nlm.nih.gov/pubchem/substance/>
PREFIX compound: <http://rdf.ncbi.nlm.nih.gov/pubchem/compound/>
PREFIX measureg: <http://rdf.ncbi.nlm.nih.gov/pubchem/measuregroup/>
PREFIX source: <http://rdf.ncbi.nlm.nih.gov/pubchem/source/>
PREFIX descr: <http://rdf.ncbi.nlm.nih.gov/pubchem/descriptor/>
PREFIX syno: <http://rdf.ncbi.nlm.nih.gov/pubchem/synonym/>
PREFIX reference: <http://rdf.ncbi.nlm.nih.gov/pubchem/reference/>
PREFIX cco: <http://rdf.ebi.ac.uk/terms/chembl#>
```

```
PREFIX chembl_target: <http://rdf.ebi.ac.uk/resource/chembl/target/>
PREFIX chembl_targetcomponent:
<http://rdf.ebi.ac.uk/resource/chembl/targetcomponent/>
PREFIX biopax3: <http://www.biopax.org/release/biopax-level3.owl#>
PREFIX reactome: <http://identifiers.org/reactome/>
PREFIX uniprot: <http://purl.uniprot.org/uniprot/>
```

Sparql	Description

SELECT
  ?assay ?assaytype ?alabel
WHERE
{
  ?assay rdf:type ?assaytype .
  GRAPH ?g {
  ?assaytype rdfs:subClassOf bao:BAO_0000015 .
  ?assaytype rdfs:label ?alabel .
  }
  FILTER(REGEX(?alabel,"cell","i"))
}
	Find cell based PD2 assays.

SELECT
  ?substance ?oidd ?bioassay ?assayname
WHERE
{
  ?substance opddr:activeInAssay ?oidd .
  ?oidd skos:exactMatch ?bioassay .
  ?oidd rdfs:label ?assayname .
  FILTER(REGEX(?assayname, "Basal_Viability","i"))
}
	Find active substances in “Basal Viability” assays.

SELECT
  ?assay ?assaytype
WHERE
{
  ?assay rdf:type ?assaytype .
  GRAPH ?g {
  ?assaytype rdfs:label "cell growth assay" .
  }
}
	Find all PD2 assays of BAO type “cell growth assay”.

SELECT
  ?assay ?assayname ?activesubstance
WHERE
{
  ?activesubstance opddr:activeInAssay ?assay .
  ?assay rdfs:label ?assayname .
  FILTER(REGEX(?assayname, "Hela Cell","i")) .
}
	Find substances active in OIDD assays with "Hela Cell" in name.

SELECT ?name ?oidd ?assay ?assaytype ?alabel	Find OIDD assays, “binding” type.

WHERE
{
  ?oidd rdfs:label ?name .
  ?oidd skos:exactMatch ?assay .
  ?assay rdf:type ?assaytype .
  GRAPH ?b {
  ?assaytype rdfs:label ?alabel .
  ?assaytype rdfs:subClassOf bao:BAO_0000015 .
  FILTER(REGEX(?alabel,"binding","i"))
  }
}	

SELECT
  ?substance_ncgc ?substance_pc ?compound ?target ?targetname ?uniprot
WHERE
{
  ?substance_ncgc skos:exactMatch ?substance_pc .
  ?substance_pc sio:CHEMINF_000477 ?compound .
  ?compound opddr:activeOnTarget ?target .
  ?target dcterms:title ?targetname .
  ?target cco:targetXref ?uniprot .
  FILTER (regex(?targetname, "Caspase", "i"))
}
	Find OIDD PD2 substances which are active against a caspase protein target.

SELECT DISTINCT
  ?target ?targetname ?oidd_assay ?oidd_assayname
WHERE
{
  ?substance_ncgc skos:exactMatch ?substance_pc .
  ?substance_pc sio:CHEMINF_000477 ?compound .
  ?compound opddr:activeOnTarget ?target .
  ?target dcterms:title ?targetname .
  FILTER (regex(?targetname, "Caspase-1", "i")) .
  ?activesubstance opddr:activeInAssay ?assay .
  ?oidd_assay rdfs:label ?oidd_assayname .
}
	Find phenotypic assays associated with ChEMBL "Caspase-1" protein target via shared active substances.

SELECT DISTINCT
  ?oidd_assay ?oidd_assayname ?target ?targetname
WHERE
{
  ?substance_ncgc opddr:activeInAssay ?assay .
  ?oidd_assay rdfs:label ?oidd_assayname .
  FILTER(REGEX(?oidd_assayname, "Hela Cell","i")) .
  ?substance_ncgc skos:exactMatch ?substance_pc .
  ?substance_pc sio:CHEMINF_000477 ?compound .
  ?compound opddr:activeOnTarget ?target .
  ?target dcterms:title ?targetname .
  FILTER(REGEX(?targetname, "kinase","i")) .
}
	Find ChEMBL protein kinase targets associated with OIDD "Hela Cell" phenotypic assays via shared active substances.

Project Status, Next Steps

This initial KN Beta Version provides sufficient associations for semantic exploration across PD2 phenotypic and public biochemical assays for the NPC substances. However, testing, debugging and creation of illustrative Sparql queries is an essential next step. In addition, one or more queryable KN triple store instances will be implemented, for validation and usability. Addition of Reactome pathways also planned.