OPDDR Knowledge Network (KN), Beta v2

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 molecular targets.

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 from ChEMBL. Biological networks can be extremely complex, and many further entity classes can be integrated in future (e.g. pathways), and will be facilitated by this initial KN.

Definitional Note

Defining the boundaries of the KN is somewhat arbitrary. By integrating ChEMBL targets, we allow for many further links and potentially additional external datasets. Likewise additional PubChem links, and links to NCBI and other external datasets. This imprecision is a benefit, a feature of linked open data.

Ontologies Used

The KN uses the following ontologies:

PubChem RDF[1]	Primary reference for OPDDR, 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#>
ChEMBL RDF[3]	ChEMBL, UniChem, Uniprot endpoint & downloads available. CCO = ChEMBL Core Ontology <http://rdf.ebi.ac.uk/terms/chembl#>
OBO[4]	Open Biological and Biomedical Ontologies BFO = Basic Formal Ontology <http://purl.obolibrary.org/obo/>
SIO[5]	Semanticscience Integrated Ontology

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

	<http://semanticscience.org/resource/>

Entities:

entity [abbr] namespace	example
substance <http://rdf.ncbi.nlm.nih.gov/pubchem/substance/>	SID124893119
compound <http://rdf.ncbi.nlm.nih.gov/pubchem/compound/>	CID1131
assay (bioassay) <http://rdf.ncbi.nlm.nih.gov/pubchem/bioassay/>	AID1117354
measuregroup (measureg) <http://rdf.ncbi.nlm.nih.gov/pubchem/measuregroup/>	AID1117354
endpoint <http://rdf.ncbi.nlm.nih.gov/pubchem/endpoint/>	SID124893119_AID1117354
protein <http://rdf.ncbi.nlm.nih.gov/pubchem/protein/>	GI124375976
target <http://rdf.ebi.ac.uk/resource/chembl/target/>	CHEMBL3038470
targetcomponent (target_cmpt) <http://rdf.ebi.ac.uk/resource/chembl/targetcomponent/>	CHEMBL_TC_1927
UniprotRef (uniprot) <http://rdf.ebi.ac.uk/terms/chembl#UniprotRef>	P53350
assay <http://rdf.ebi.ac.uk/resource/chembl/assay/>	CHEMBL987214
activity <http://rdf.ebi.ac.uk/resource/chembl/activity/>	CHEMBL_ACT_2470294
molecule <http://rdf.ebi.ac.uk/resource/chembl/molecule/>	CHEMBL44884

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. Note also that PubChem measuregroups are defined for each assay, for example, the measuregroup URI for AID12345 is http://rdf.ncbi.nlm.nih.gov/pubchem/measuregroup/AID12345. PubChem

endpoints represent activity outcomes. ChEMBL RDF represents bioactivities somewhat differently than PubChem, but we can rigorously link these data via chemical structure and CIDs.

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.
endpoint	2511*35	PubChem SID-AID pairs.
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.

Asserted triplets, patterns and examples

description	examples
assay to BAO class	bioassay:AID1117354 rdf:type bao:BAO_0000015
assay title	bioassay:AID1117354 dcterms:title "human JAK2 kinase inhibition-screen"@en
assay to measuregroup	bioassay:AID1117354 bao:BAO_0000209 measuregroup:AID1117354
substance to NCGC ID	substance:SID144206486 skos:exactMatch ncats_sample:NCGC00182710-02 .
substance to measure group	substance:SID124882766 obo:BFO_0000056 measureg:AID1117326
endpoint outcome (activity)	endpoint:SID170466632_AID743241 vocabulary:PubChemAssayOutcome vocabulary:inactive

endpoint class	endpoint:SID103164874_AID443491 rdf:type bao:BAO_0000190
substance to compound association	substance:SID124893119 sio:CHEMINF_000477 compound:CID1131
assay to OIDD ID	bioassay:AID1117350 skos:exactMatch oidd_assay:17
ChEMBL target to UniProt	chembl_target:CHEMBL5464 cco:targetXref uniprot:Q13546
ChEMBL target to assay	chembl_target:CHEMBL5464 cco:hasAssay assay:CHEMBL3110727
ChEMBL target to target component	chembl_target:CHEMBL1867 cco:hasTargetComponent chembl_targetcmpt:CHEMBL_TC_180
ChEMBL target component to Uniprot	chembl_targetcmpt:CHEMBL_TC_180 cco:targetCmptXref uniprot:P08913
ChEMBL assay to activity	assay:CHEMBL3110727 cco:hasActivity activity:CHEMBL_ACT_13890030
ChEMBL molecule to activity	chembl_molecule:CHEMBL313842 cco:hasActivity activity:CHEMBL_ACT_14447741
PubChem substance to ChEMBL molecule	substance:SID225144242 skos:exactMatch molecule:CHEMBL1474122

NOTE: provisional namespaces defined for this beta version:

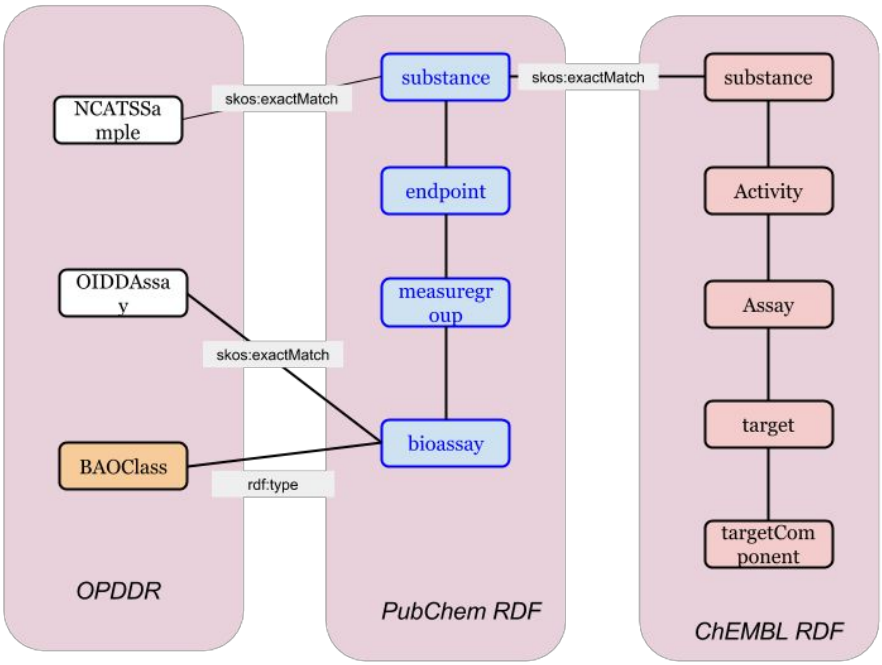
- ncats_sample: <http://rdf.ncats.nih.gov/ncgc/sample/>
- oidd_assay: <http://openinnovation.lilly.com/bioassay#>

Essential linking predicates:

(From PubChem and ChEMBL RDF)

predicate	description	links
bao:BAO_0000209	Has Measuregroup	assay-measuregroup
obo:BFO_0000056	Participates In	substance-measuregroup
obo:IAO_0000136	Is About	endpoint-substance
obo:OBI_0000299	Has Output	measuregroup-endpoint
sio:CHEMINF_000477	Has Stdized Cpd	substance-compound
cco:hasActivity		substance-activity assay-activity
cco:hasTarget		assay-target

cco:hasTargetComponent	target-targetcomponent



OPDDR RDF schema, simplified, showing source relationships.

Files:

The following files comprise this release. Files are grouped below by source, each file from one source only.

file	source	description
npcpd2_assay.ttl	OPDDR	Assay links to OIDD namespace. `bioassay:AID1117326 skos:exactMatch oidd_assay:4`
npcpd2_bao.ttl	OPDDR	Manually curated BAO classifications. `bioassay:AID1117352 rdf:type bao:BAO_0000219`
npcpd2_substance.ttl	OPDDR	Substance links to NCATS namespace. `substance:SID170465644 skos:exactMatch ncats_sample:NCGC00160518-03`
bao_vocabulary_assay.owl	BAO	BAO module with bioassay class hierarchy.
pubchem_vocabulary.owl	PubChem	PubChem module with bioactivity terms etc.
pubchem_assay.ttl	PubChem	PubChem RDF, includes titles, measuregroups. `bioassay:AID1117356 bao:BAO_0000209`

		measuregroup:AID1117356 bioassay:AID1117351 dcterms:title "Increased HeLa cells with 4N DNA content-IC50"@en
pubchem_substance.ttl	PubChem	PubChem RDF, includes CIDs, measuregroups. substance:SID124882766 obo:BFO_0000056 measuregroup:AID1117326 . endpoint:SID124882766_AID1117342 obo:IAO_0000136 substance:SID124882766 .
pubchem_endpoint.ttl	PubChem	PubChem RDF, includes endpoints, activity results. endpoint:SID170464708_AID1117354 obo:IAO_0000136 substance:SID170464708 ; vocabulary:PubChemAssayOutcome vocabulary:inactive . measuregroup:AID1117354 obo:OBI_0000299 endpoint:SID170464708_AID1117354 .
chembl_cco.ttl	ChEMBL	ChEMBL Core Ontology
chembl_target.ttl	ChEMBL	ChEMBL protein targets, target components, Uniprots. chembl_target:CHEMBL2366239 a cco:SingleProtein ; dcterms:title "KLE" . chembl_target:CHEMBL218 cco:hasTargetComponent chembl_targetcmpt:CHEMBL_TC_172 . chembl_targetcmpt:CHEMBL_TC_172 cco:targetCmptXref uniprot:P21554 .
chembl_activity.ttl	ChEMBL	PubChem substance links to ChEMBL molecules, activities, assays, targets. substance:SID170466134 skos:exactMatch chembl_molecule:CHEMBL1230222 . chembl_molecule:CHEMBL44884 cco:hasActivity chembl_activity:CHEMBL_ACT_7667167 . chembl_target:CHEMBL218 cco:hasAssay chembl_assay:CHEMBL1909122 .

Obtaining PubChem RDF

The RESTful API was used, specifically the following resources

resource	example
substance	http://rdf.ncbi.nlm.nih.gov/pubchem/substance/SID70464708.rdf
bioassay	http://rdf.ncbi.nlm.nih.gov/pubchem/bioassay/AID1117354.rdf
endpoint	http://rdf.ncbi.nlm.nih.gov/pubchem/endpoint/SID70464708_AID1117354.rdf

Obtaining ChEMBL RDF

The Sparql endpoint was used, specifically https://www.ebi.ac.uk/rdf/services/chembl/sparql, to obtain activities against protein targets for all OPDDR substances.

Sparql

```
SELECT "chembl_molecule:${chembl_id}" ?activity ?assay ?target
?targetcmpt ?uniprot
WHERE {
  ?activity a cco:Activity ;
     cco:hasMolecule chembl_molecule:${chembl_id} ;
     cco:hasAssay ?assay .
  ?assay cco:hasTarget ?target .
  ?target cco:hasTargetComponent ?targetcmpt .
  ?targetcmpt cco:targetCmptXref ?uniprot .
  ?uniprot a cco:UniprotRef .
}
```

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 endpoint: <http://rdf.ncbi.nlm.nih.gov/pubchem/endpoint/>
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 vocabulary: <http://rdf.ncbi.nlm.nih.gov/pubchem/vocabulary#>
PREFIX cco: <http://rdf.ebi.ac.uk/terms/chembl#>
PREFIX chembl_molecule: <http://rdf.ebi.ac.uk/resource/chembl/molecule/>
PREFIX chembl_activity: <http://rdf.ebi.ac.uk/resource/chembl/activity/>
PREFIX chembl_assay: <http://rdf.ebi.ac.uk/resource/chembl/assay/>
PREFIX chembl_target: <http://rdf.ebi.ac.uk/resource/chembl/target/>
PREFIX chembl_targetcmpt:
<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 assays (PubChem AIDs).	✔

SELECT
 ?name ?assay ?oidd ?assaytype ?alabel
WHERE
{
  ?assay skos:exactMatch ?oidd .
  ?assay dcterms:title ?name .
  ?assay rdf:type ?assaytype .
  GRAPH ?b {
  ?assaytype rdfs:label ?alabel .
  ?assaytype rdfs:subClassOf bao:BAO_0000015 .
  FILTER(REGEX(?alabel,"binding","i"))
  }
}
	Find binding assays; include OIDD IDs.	✔

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
  ?substance ?bioassay ?oidd ?assayname
WHERE
{
  ?substance obo:BFO_0000056 ?measureg .
  ?bioassay bao:BAO_0000209 ?measureg .
  ?bioassay dcterms:title ?assayname .
  ?bioassay skos:exactMatch ?oidd .
  FILTER(REGEX(?assayname, "HeLa cell","i")) .
  ?measureg obo:OBI_0000299 ?endpoint .
  ?endpoint obo:IAO_0000136 ?substance .
  ?endpoint vocabulary:PubChemAssayOutcome vocabulary:active .
}
	Find active substances in “HeLa cell” assays.	✔

SELECT
  ?substance ?assay ?assayname
WHERE
{
  ?substance obo:BFO_0000056 ?measureg .
  ?bioassay bao:BAO_0000209 ?measureg .
  ?measureg obo:OBI_0000299 ?endpoint .
  ?endpoint obo:IAO_0000136 ?substance .
  ?endpoint vocabulary:PubChemAssayOutcome vocabulary:active .
  ?assay rdfs:label ?assayname .
  FILTER(REGEX(?assayname, "Hela Cell","i")) .
}
	Find substances active in “Hela Cell” assays.	✔

SELECT DISTINCT
  ?substance ?mol ?target ?targetname ?uniprot
WHERE
{
  ?substance skos:exactMatch ?mol .
  ?mol cco:hasActivity ?activity .
  ?assay cco:hasActivity ?activity .
  ?target cco:hasAssay ?assay .
  ?target cco:hasTargetComponent ?tc .
  ?tc cco:targetCmptXref ?uniprot .
  ?uniprot a cco:UniprotRef .
  ?target dcterms:title ?targetname .
  FILTER (regex(?targetname, "Caspase", "i")) .
}
	Find substances which are active against a caspase protein target.	✔

SELECT DISTINCT
  ?target ?targetname ?assay ?assayname
WHERE
{
  #phenotypic
  ?substance obo:BFO_0000056 ?measureg .
  ?assay bao:BAO_0000209 ?measureg .
  ?measureg obo:OBI_0000299 ?endpoint .
  ?endpoint obo:IAO_0000136 ?substance .
 	
  #biochemical
  ?substance skos:exactMatch ?mol .
  ?mol cco:hasActivity ?activity .
  ?chembl_assay cco:hasActivity ?activity .
  ?target cco:hasAssay ?chembl_assay .
 	
  ?target dcterms:title ?targetname .
  FILTER (regex(?targetname, "Caspase-1", "i")) .
  ?assay dcterms:title ?assayname .
}
	Find phenotypic assays associated with ChEMBL "Caspase-1" protein target via shared active substances.	✔

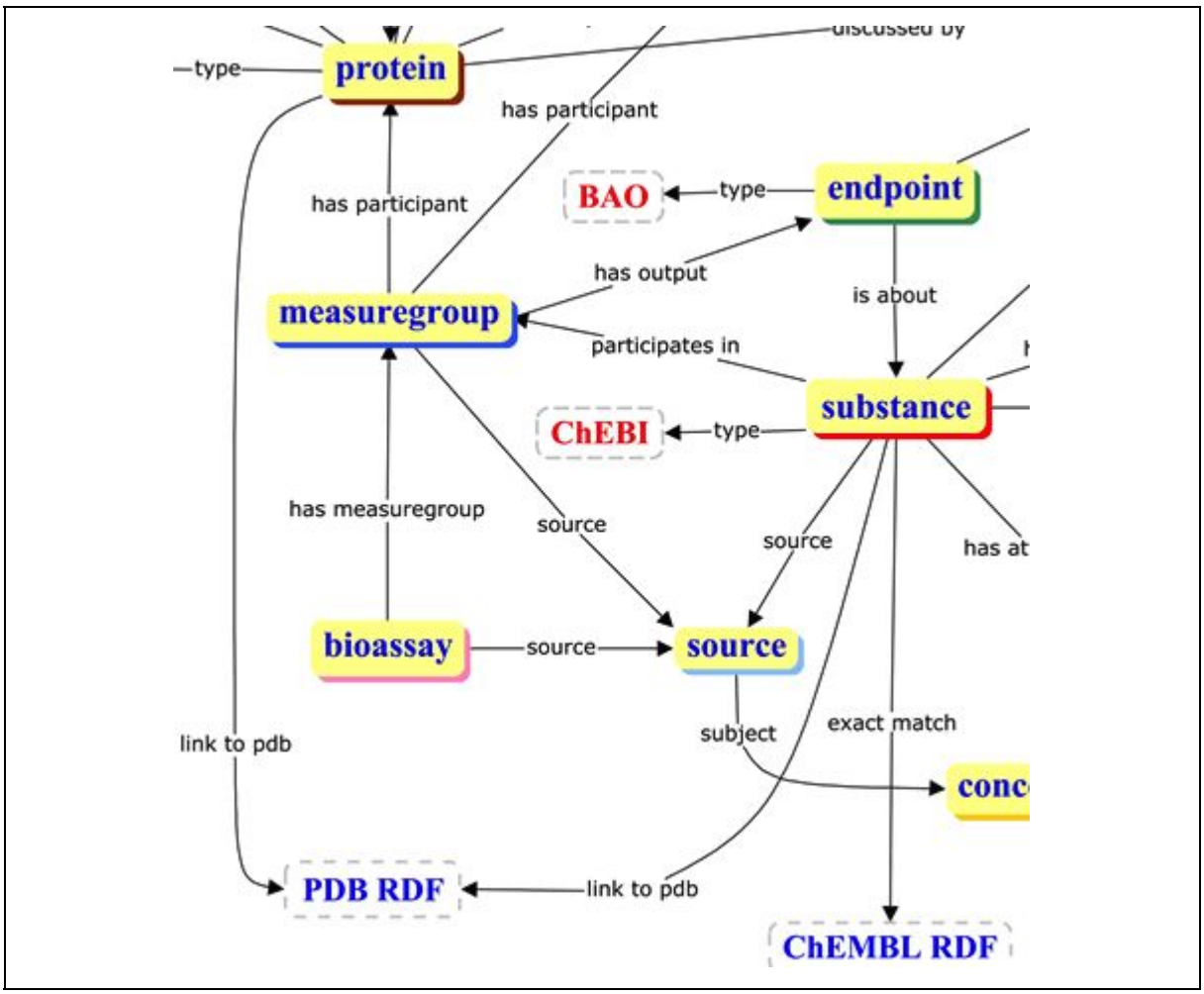
SELECT DISTINCT
  ?assay ?assayname ?target ?targetname
WHERE
{
  ?substance obo:BFO_0000056 ?measureg .
  ?assay bao:BAO_0000209 ?measureg .
  ?measureg obo:OBI_0000299 ?endpoint .
  ?endpoint obo:IAO_0000136 ?substance .
	Find ChEMBL protein kinase targets associated with OIDD "Hela Cell" phenotypic assays via shared active substances.	✔

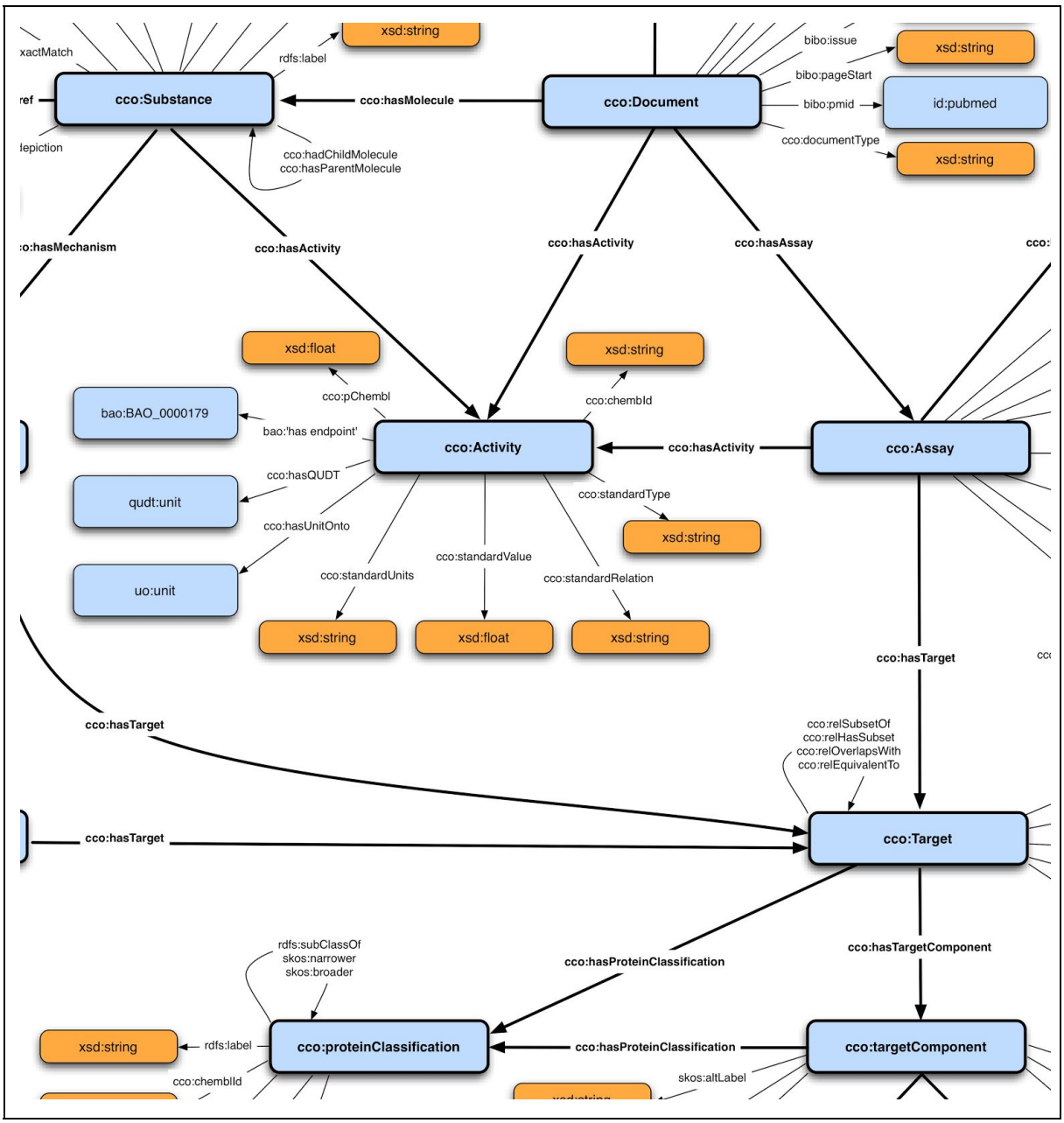
```
  ?substance skos:exactMatch ?mol .
  FILTER(REGEX(?assayname, "Hela Cell","i")) .
  ?assay dcterms:title ?assayname .
  ?mol cco:hasActivity ?activity .
  ?chembl_assay cco:hasActivity ?activity .
  ?target cco:hasAssay ?chembl_assay .

  ?target dcterms:title ?targetname .
  ?target dcterms:title ?targetname .
  FILTER(REGEX(?targetname, "kinase","i")) .
}
```

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.


protein
type
has participant
has participant
BAO
type
endpoint
has output
is about
measuregroup
participates in
substance
ChEBI
type
has measuregroup
source
source
has at
bioassay
source
source
subject
exact match
link to pdb
conc
PDB RDF
link to pdb
ChEMBL RDF


xactMatch
rdfs:label
xsd:string
cco:Substance
cco:hasMolecule
cco:Document
bibo:issue
xsd:string
bibo:pageStart
bibo:pmid
id:pubmed
cco:documentType
xsd:string
lepiction
cco:hadChildMolecule
cco:hasParentMolecule
:o:hasMechanism
cco:hasActivity
cco:hasActivity
cco:hasAssay
cco:
xsd:float
xsd:string
cco:pChembl
cco:chembId
bao:BAO_0000179
bao:'has endpoint'
cco:Activity
cco:hasActivity
cco:Assay
cco:hasQUDT
qudt:unit
cco:standardType
xsd:string
cco:hasUnitOnto
uo:unit
cco:standardUnits
cco:standardValue
cco:standardRelation
xsd:string
xsd:float
xsd:string
cco:hasTarget
cco:hasTarget
cco:relSubsetOf
cco:relHasSubset
cco:relOverlapsWith
cco:relEquivalentTo
cco:hasTarget
cco:Target
rdfs:subClassOf
skos:narrower
skos:broader
cco:hasProteinClassification
cco:hasTargetComponent
xsd:string
rdfs:label
cco:proteinClassification
cco:hasProteinClassification
cco:targetComponent
cco:chembIId
skos:altLabel