

SWAT4LS Hackathon Ideas and Programme

Instructions for contributing to this document:

- Add **Datasets** that we can use in an activity with sample SPARQL query
 - Add proposals for **Activities**
 - We will create a **poll** leading up to the day of the Hackathon to sign up for activities of interest (possibly more than one)
-

Program - Day 1

- 9:00 Arrive
 - 9:30 Welcome and Introduction
 - Short presentations on activities per cluster
 - hack hack hack
 - 12:30 Lunch PIZZA
 - hack hack hack
 - document document document
 - Short presentations about progress, lessons learned, etc.
 - Food, drink, merriment, ..
 - Building Closes 8PM
-

Welcome to the SWAT4LS Hackathon 2013 discussion group. We will be discussing ideas here leading up to the hackathon on Wed. Dec. 11. Some of the data that we will have to work with: Clinical data from MAASTRO (SPARQL endpoint), artificial clinical data from VU Amsterdam (SPARQL endpoint), EBI RDF and SPARQL endpoints, WikiPathways RDF and SPARQL endpoints, Disgenet RDF and SPARQL endpoints

EBI has already provided an RDFApp challenge. Details:

<http://www.ebi.ac.uk/rdf/rdfapps>

A few ideas for additional themes and challenges:

- mixed-level policy guided access control for SPARQL

- federated lookup of labels from medical terminologies
- data catalog based on W3C Note on data description (draft)
- validating data descriptions with modified OpenPhacts tool
- validating data descriptions with Shapes (Eric's RDF validator)
- query builders
- query template pipeline

Feel free to suggest ideas.

See also last year's wiki for inspiration:

<http://www.w3.org/wiki/HCLS/SWAT4LS2012/Hackathon>

Program - Day 1

<http://www.ebi.ac.uk/rdf/rdfapps>

Datasets

Provided by VU Amsterdam:

[Advanced Patient Data Generator for 10K breast cancer females:](#)

[4665 clinical trials of breast cancer](#)

[Adverse Event data](#)

[MAASTRO clinic data](#)

[WikiPathways](#)

[Open PHACTS Linked Data API](#)

[EBI RDF](#)

[EBI Biosamples \(http://www.ebi.ac.uk/rdf/services/biosamples/\)](http://www.ebi.ac.uk/rdf/services/biosamples/)

[DisGeNET RDF](#)

Activities

[Create a Data Map for the Hackathon Data Sources](#)

[Data Federation with data description](#)

[Create Dataset Description RDF for a few of the Hackathon data sources](#)

[Gathering Dataset Descriptions](#)

[Validation](#)

[Dataset description visualization](#)

[Access Control](#)

[EBI-WikiPathways-Disgenet](#)

[Cancer Therapy Drugs, Targets, Interactions and Pathways](#)

[Adverse Drug Event application](#)

[COEUS Data Integration](#)

Datasets

Provided by VU Amsterdam:

Advanced Patient Data Generator for 10K breast cancer females:

<http://wasp.cs.vu.nl/apdg/download/female-breast-cancer-2013a-10000.zip>

The above 10,000 patient data of breast cancer are available at the following SPARQL endpoint:

<http://ops.few.vu.nl:8890/sparql>

The graph is: <http://female-breast-cancer>

Here is an example of the SPARQL queries:

```
SELECT *  
FROM http://female-breast-cancer  
WHERE  
{?s ?p ?o}  
LIMIT 100
```

4665 clinical trials of breast cancer

The 4665 clinical trials of breast cancer are available at the following SPARQL endpoints:

<http://ops.few.vu.nl:8890/sparql>

The graph is: <http://nct>

Here is an example of SPARQL query.

```
PREFIX rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns#  
PREFIX foaf: http://xmlns.com/foaf/0.1/  
PREFIX sct: http://wasp.cs.vu.nl/sct/sct#  
PREFIX rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns#  
PREFIX foaf: http://xmlns.com/foaf/0.1/  
PREFIX sct: http://wasp.cs.vu.nl/sct/sct#  
PREFIX sctid: http://wasp.cs.vu.nl/sct/id#  
SELECT DISTINCT ?nctid ?minage ?maxage ?summary  
FROM http://nct
```

```

where {
    ?ct rdf:type sct:ClinicalTrial.
    ?ct sct:NCTID ?nctid.
    ?ct sct:OverallStatus "Recruiting".
    ?ct sct:Phase "Phase 3".
    ?ct sct:Intervention ?if.
    ?if sct:InterventionType "Drug".
    ?ct sct:EligibilityGender "Female".
    ?ct sct:EligibilityMinAge ?minage.
    ?ct sct:EligibilityMaxAge ?maxage.
    ?ct sct:BriefSummary ?summary.
    FILTER(?minage <= "70 Years" &&
    ?maxage >= "75 Years").
}
ORDER BY ?nctid
LIMIT 100

```

Adverse Event data

<http://aers.data2semantics.org/>

<http://adepedia.org>

MAASTRO clinic data

De-identified patient data from VATE project (contributed by Johan van Soest and Andre Dekker):

```

PREFIX vate:<http://vocab.maastro.nl/store/repositories/vate/ontology#>
SELECT (STR(?patientlabel) AS ?patientlabel) ?gendercode ?stagecode
WHERE {
    SERVICE <http://vocab.maastro.nl/store/repositories/vate> {
        ?patient vate:hasDisease ?disease.
        ?patient rdf:label ?patientlabel.
        ?patient vate:hasBiologicalSex ?gendercode.
        ?disease vate:hasStaging ?stageFinding.
        ?stageFinding vate:hasAjccv6StageFinding ?stagecode.
    }
}

```

```
}
```

WikiPathways

Pathway content from www.wikipathways.org as RDF through sparql endpoint (<http://sparql.wikipathways.org>). A list of example queries: http://www.wikipathways.org/index.php/Help:WikiPathways_Sparql_queries

For the genes differentially expressed in asthma, get the gene products associated to a WikiPathways pathway. (Built upon example query 5 in: <http://www.ebi.ac.uk/rdf/services/atlas/sparql>). You can substitute the [EFO number for other disease codes](#).

```
PREFIX identifiers:<http://identifiers.org/ensembl/>
PREFIX atlas: <http://rdf.ebi.ac.uk/resource/atlas/>
PREFIX atlasterms: <http://rdf.ebi.ac.uk/terms/atlas/>
PREFIX efo: <http://www.ebi.ac.uk/efo/>
```

```
SELECT DISTINCT ?wpURL ?pwTitle ?expressionValue ?pvalue where {
```

```
SERVICE <http://www.ebi.ac.uk/rdf/services/atlas/sparql> {
  ?factor rdf:type efo:EFO_0000270 .
  ?value atlasterms:hasFactorValue ?factor .
  ?value atlasterms:isMeasurementOf ?probe .
  ?value atlasterms:pValue ?pvalue .
  ?value rdfs:label ?expressionValue .
  ?probe atlasterms:dbXref ?dbXref .
}
  ?pwElement dcterms:isPartOf ?pathway .
  ?pathway dc:title ?pwTitle .
  ?pathway dc:identifier ?wpURL .
  ?pwElement wp:bdbEnsembl ?dbXref .
}
ORDER BY ASC(?pvalue)
```

Open PHACTS Linked Data API

Build with the Open PHACTS Discovery Platform. The platform provides a convenient API to query across multiple pharmacology datasets. (<https://dev.openphacts.org/docs/1.3>)

Helpful libraries for processing the data

- Javascript <https://github.com/openphacts/ops.js>
- Ruby https://github.com/openphacts/ops_gems
- Java <https://github.com/openphacts/JavaLDAClient>

The datasets that are integrated are detailed at

<http://support.openphacts.org/support/solutions/articles/131704-what-data-is-in-the-system-exactly->

Also provides a service for mapping between equivalent URIs in the various datasets, this is available through the Map URL method in the LDAPI.

EBI RDF

Data available from here:

<http://www.ebi.ac.uk/rdf/services>

Includes:

Gene expression experiments (Atlas), Pathways (Reactome), Chemical compounds and drug targets (ChEMBL), Biological samples (BioSamples), Computational models (BioModels), plus proteins (UniProt).

EBI Biosamples (<http://www.ebi.ac.uk/rdf/services/biosamples/>)

Just a few ideas of interesting things that could be done with this dataset and other EBI datasets or other similar datasets mentioned hereby.

- From geo-located samples (samples annotated with latitude/longitude) to Google maps, e.g. by using Exhibit (<http://www.simile-widgets.org/exhibit>).
 - Real hacky code on <https://github.com/ianwdunlop/biosamples-geo-rdf> using google maps and sparql to draw biosamples on a map - only static at the moment fixed around some point in Liverpool.
- Take similar datasets (e.g., MAASTRO, Breast Cancer Data, your data), unify the schemas (e.g., using CONSTRUCT and this approach: <http://tinyurl.com/ncjc3c3>), define federated queries.
- Use the Shape or OpenPHACTS validator to define sensible rules for BioSD and similar data-sets, e.g., must contain an organism, should have a treatment.
- Design/build an App (or Web widget) that asks for eligibility criterion, i.e., pairs of attribute value/type, and translate it into a SPARQL query (or a more complex search based on SPARQL) to find samples.
 - Use common ontologies for auto-completion over property types
 - Use string-based auto-completion for values
 - Consider numerical values, interval, units
 - Do approximate matching, i.e., matching 8/10 of specified pairs is good.
 - after an initial simple version, expand searches via (downward) transitive closure of subClassOf

DisGeNET RDF

Gene-disease association data from DisGeNET (<http://ibi.imim.es/DisGeNET>) available in RDF via the SPARQL endpoint: <http://rdf.imim.es/sparql/>

The graph is <<http://disgenet>>

Some example queries and the RDF Schema @ <http://rdf.imim.es/DisGeNET.html>

Activities

Create a Data Map for the Hackathon Data Sources

Like this:

http://www.w3.org/wiki/HCLS/SWAT4LS2012/Hackathon#Data_Map

http://www.w3.org/wiki/HCLS/SWAT4LS2012/Hackathon#.23_Void

http://www.w3.org/wiki/HCLS/SWAT4LS2012/Hackathon#Triplestore_map

Data Federation with data description

Create Dataset Description RDF for a few of the Hackathon data sources

One idea for a task at the hackathon is to describe contributed datasets (from MAASTRO clinic and VU) using the vocabularies selected for dataset descriptions. There is also a link in that document to the spreadsheet that is gradually being transferred into the note (see below). Demonstration queries for data discovery could form the follow up task (e.g. I only want open data with a CC license of type X that is updated at least 2X per year.)

Try using the prototype void editor from uniman - <http://voideditor.cs.man.ac.uk/voidEditor/>

W3C Note draft:

https://docs.google.com/document/d/1zGQJ9bO_dSc8taINTNHdnjYEzUyYkbjglrcuUPuoITw/edit?pli=1#heading=h.wyc73yp7c8jz

10,000 patient data of female breast cancer: VOID description:

<http://wasp.cs.vu.nl/apdg/femalebreastcancer2013a-void.ttl>

Spreadsheet with vocabulary decisions for describing datasets in RDF:

https://docs.google.com/spreadsheets/ccc?key=0Aoy0zfdRviKsdFJWTDfPbINXc3BtelhrdEpNYTdVbXc&usp=drive_web#gid=1

Gathering Dataset Descriptions

Google form for gathering dataset descriptions is available from

<https://docs.google.com/forms/d/1-C8vZ67NQm6BDaMhd4gJtasZTngfqvop0jEhb--bA/viewform>

The form focuses on the MUST and SHOULD properties. The idea is to do a csv2rdf conversion of the data gathered.

Johan van Soest has created some dataset descriptions by hand and added them to VATE:

PREFIX dct:<<http://purl.org/dc/terms/>>

SELECT ?pred ?obj

WHERE {

SERVICE <<http://vocab.maastro.nl/store/repositories/vate>> {

 ?dataset rdf:type dct:Dataset.

 ?dataset ?pred ?obj.

 }

}

Alasdair is in the process of developing a GUI for creating dataset descriptions for the Open PHACTS specification.

Validation

The Manchester Open PHACTS team (Alasdair) has written a validator for Open PHACTS data that makes use of MUST, SHOULD, MAY qualifications in a specification to validate.

Possible activity: Adapt this to the Dataset Description specification from HCLS Linked Life Data task force. (Essentially a joint OWL writing session; would stop Alasdair being a bottleneck!)

From Eric to semweb: "I tooled away on <http://www.w3.org/2013/ShEx/SimpleShExDemo> to compile validating SPARQL queries. If you 'v'alidate, you'll see a link for a <SPARQL validating query> which captures very nearly all of the ShEx functionality. More work to do, but it might be useful to some of y'all in its current form." See also, <http://www.w3.org/2013/ShEx/Primer>

Initial creation of a HCLS Validator based on the Shape Expressions is being developed at <https://github.com/AlasdairGray/HCLSValidator>

Dataset description visualization

Develop an intuitive way of presenting descriptions which allows a user to understand the datasets that are available.

SPARQL translation of data

Team: M. Scott Marshall, Eric Prud'hommeaux

Create a dynamic translation of a SPARQL query from one vocabulary (VATE) to the other (EuroCAT) using [SWObjects](#). For a nice SWObjects tutorial, try this from Helena Deus in SWAT4LS 2011 (probably handy to download):

https://docs.google.com/document/d/170GFtfwcuXt5vvhzvZLyCJChMnDeWJbrLd4_7LuUQGQ/edit

An example piece of an SWObjects map (SPARQL CONSTRUCT) that 'translates' a code for gender in VATE to a label via a federated lookup:

```
PREFIX vate: <http://vocab.maastro.nl/store/repositories/vate/ontology#>
```

```
LABEL <gender> CONSTRUCT {
  ?patient vate:hasBiologicalSex ?genderlabel.
}
WHERE {
  SERVICE <http://192.168.100.51/store/repositories/vate> {
    ?patient vate:hasBiologicalSex ?gendercode.
  }
  SERVICE <http://linked.opendata.cz/sparql> {
    GRAPH <http://linked.opendata.cz/resource/dataset/nci-thesaurus> {
      ?gendercode rdfs:label ?genderlabel.
    }
  }
}
```

Part of the SWObjects map file:

```
PREFIX vate: <http://vocab.maastro.nl/store/repositories/vate/ontology#>
PREFIX eurocat: <http://vocab.maastro.nl/store/repositories/eurocat/ontology#>
```

```
LABEL <gender> CONSTRUCT {
  ?patient eurocat:Gender_PREFERRED_Name ?genderlabel.
}
WHERE {
  SERVICE <http://192.168.100.51/store/repositories/vate> {
    ?patient vate:hasBiologicalSex ?gendercode.
  }
  SERVICE <http://linked.opendata.cz/sparql> {
    GRAPH <http://linked.opendata.cz/resource/dataset/nci-thesaurus> {
      ?gendercode rdfs:label ?genderlabel.
    }
  }
}
```

```

}

LABEL <stage> CONSTRUCT {
?patient eurocat:StageClinical ?stagelabel.
}
WHERE {
    SERVICE <http://localhost/store/repositories/vate> {
        ?patient vate:hasDisease ?disease.
        ?patient rdf:label ?patientlabel.
        ?disease vate:hasStaging ?stageFinding.
        ?stageFinding vate:hasAjccv6StageFinding ?stagecode.
    }
    SERVICE <http://linked.opendata.cz/sparql> {
        GRAPH <http://linked.opendata.cz/resource/dataset/nci-thesaurus> {
            ?stagecode rdfs:label ?stagelabel.
        }
    }
}

```

The query that works on it:

```

PREFIX eurocat:<http://vocab.maastro.nl/store/repositories/eurocat/ontology#>
SELECT ?patient (STR(?sex) AS ?gender) (STR(?stageFinding) as ?stage)
WHERE {
    SERVICE <http://vocab.maastro.nl:8086/sparql> {
        ?patient eurocat:Gender_PREFERRED_Name ?sex.
        ?patient eurocat:StageClinical ?stageFinding.
    }
}

```

In SNORQL:

http://data.semanticweb.org/snorql/?query=PREFIX+eurocat%3A%3Chttp%3A%2F%2Fvocab.maastro.nl%2Fstore%2Frepositories%2Feurocat%2Fontology%23%3E%0D%0ASELECT+%3Fpatient+%28STR%28%3Fsex%29+AS+%3Fgender%29+%28STR%28%3FstageFinding%29+as+%3Fstage%29%0D%0AWHERE+%7B%0D%0A%09SERVICE+%3Chttp%3A%2F%2Fvocab.maastro.nl%3A8086%2Fsparql%3E+%7B%0D%0A+++++++%3Fpatient+eurocat%3AGender_PREFERRED_Name+%3Fsex.%0D%0A+++++++%3Fpatient+eurocat%3AStageClinical+%3FstageFinding.%0D%0A%09%7D%0D%0A%7D

Access Control

Team: Eric Prud'hommeaux, Kasper van den Berg, M. Scott Marshall, Andre Dekker
[http://www.w3.org/2009/Talks/0504-swobjects-ep/acls#\(5\)](http://www.w3.org/2009/Talks/0504-swobjects-ep/acls#(5))

EBI-WikiPathways-Disgenet

Team: Alasdair Gray, Nuria Queralt Rosinach, Simon Jupp, James Malone, Guoqian Jiang (?), Chris Evelo, Andra Waagmeester.

[11-11-2013 15:32:54] Andra Waagmeester: I would love to work on some EBI - WikiPathways explorations in terms of service calls

Some examples:

http://www.wikipathways.org/index.php/Help:WikiPathways_Sparql_queries#WikiPathways_with_EBI_Atlas_RDF

It is about linking diseases to pathways. Would be nice to compare with other disease resources (Disgenet)

Núria Queralt Rosinach [DisGeNET/OPS]: I'd like to work on crossing DisGeNET data with Wikipathways and EBI datasets in order to retrieve, for instance, information about tissue specific gene expression for some gene diseases, ...

Simon Jupp: This sounds good, I'd like to see how much ontology coverage both Wikipathways and Disgenet have for disease terms and see if we can align with EFO.

james: I'd be interested in expanding our enrichment package using this.

Cancer Therapy Drugs, Targets, Interactions and Pathways

Team: Guoqian Jiang, Adrien Coulet, Oliver Cure

Guoqian Jiang (Mayo Clinic): I would like to work on a subset of cancer therapy drugs extracted from NCI Thesaurus and explore the EBI RDF platform and other resources (eg, OpenPHACT) for their molecular targets, interactions and pathways.

<http://www.cancer.gov/cancertopics/factsheet/Therapy/targeted>

Oliver - tamoxifen

<http://www.cancer.gov/cancertopics/druginfo/tamoxifencitrate>

Several drugs that interfere with estrogen binding to the ER have been approved by the FDA for the treatment of ER-positive breast cancer. Drugs called selective estrogen receptor modulators ([SERMs](#)), including [tamoxifen](#) and [toremifene \(Fareston®\)](#), bind to the ER and prevent estrogen binding. Another drug, [fulvestrant \(Faslodex®\)](#), binds to the ER and promotes its destruction, thereby reducing ER levels inside cells.

Retrieving info on targets of Tamoxifen

On drugBank:

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

PREFIX dbv: <http://bio2rdf.org/drugbank_vocabulary:>

PREFIX db: <http://bio2rdf.org/drugbank:>

SELECT ?geneName ?category ?label

WHERE {

db:DB00675 dbv:category ?category.

db:DB00675 dbv:target ?target.

db:DB00675 rdfs:label ?label.

?target dbv:gene-name ?geneName

}

ORDER BY ASC(?geneName)

on chEMBL:

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 cco: <http://rdf.ebi.ac.uk/terms/chembl#>

PREFIX chembl_molecule: <http://rdf.ebi.ac.uk/resource/chembl/molecule/>

SELECT ?activity ?assay ?target ?targetcmpt ?uniprot

WHERE {

?activity a cco:Activity ;

cco:hasMolecule chembl_molecule:CHEMBL83 ;

cco:hasAssay ?assay .

?assay cco:hasTarget ?target .

?target cco:hasTargetComponent ?targetcmpt .

?targetcmpt cco:targetCmptXref ?uniprot .

?uniprot a cco:UniprotRef

}

I asked for openPHACTS but have not received the API key that would allow me to call the API

Adrien

<http://www.cancer.gov/cancertopics/druginfo/bevacizumab>

- [Bevacizumab \(Avastin®\)](#) is a monoclonal antibody that is approved for the treatment of [glioblastoma](#). The therapy is also approved to treat some patients with non-small cell lung cancer, metastatic colorectal cancer, and metastatic kidney cancer. Bevacizumab binds to VEGF and prevents it from interacting with receptors on [endothelial cells](#), blocking a step that is necessary for the initiation of new blood vessel growth.

```
#7 GET UNIPROT ID OF TARGETS OF Bevacizumab (DB00112)
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX dbv:<http://bio2rdf.org/drugbank_vocabulary:>
PREFIX db:<http://bio2rdf.org/drugbank:>
SELECT ?targetg
CONCAT("http://purl.uniprot.org/uniprot/",REPLACE (STR(?xref) ,
"http://bio2rdf.org/uniprot:", "")) AS ?x WHERE {

db:DB00112 dbv:category ?category.
db:DB00112 dbv:target ?target.
db:DB00112 rdfs:label ?label.
FILTER (?category != "Approved" && ?category != "Biotech" &&
?category != "Small Molecule")
?target dbv:gene-name ?targetg.
?target dbv:xref ?xref.
FILTER REGEX(str(?xref) , "^http://bio2rdf.org/uniprot:")
}

##results
VEGFA      http://purl.uniprot.org/uniprot/P15692
C1QA http://purl.uniprot.org/uniprot/P02745
C1QB http://purl.uniprot.org/uniprot/P02746
C1QC http://purl.uniprot.org/uniprot/P02747
FCGR2A     http://purl.uniprot.org/uniprot/P12318
FCGR2C     http://purl.uniprot.org/uniprot/P31995
FCGR1A     http://purl.uniprot.org/uniprot/P12314
FCGR3B     http://purl.uniprot.org/uniprot/O75015
C1R  http://purl.uniprot.org/uniprot/P00736
FCGR3A     http://purl.uniprot.org/uniprot/P08637
```

FCGR2B <http://purl.uniprot.org/uniprot/P31994>

##CANNOT USE BIND in Bio2RDF TO GO TO REACTOME

#7 GET REACTOME PATHWAYS FOR EACH TARGET

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 biopax3:

<<http://www.biopax.org/release/biopax-level3.owl#>>

SELECT DISTINCT ?pathway ?pathwayname

WHERE

{?pathway rdf:type biopax3:Pathway .

?pathway biopax3:displayName ?pathwayname .

?pathway biopax3:pathwayComponent ?reaction .

?reaction rdf:type biopax3:BiochemicalReaction .

{

{?reaction ?rel ?protein .}

UNION

{

?reaction ?rel ?complex .

?complex rdf:type biopax3:Complex .

?complex ?comp ?protein .

}}

?protein rdf:type biopax3:Protein .

?protein biopax3:entityReference

<<http://purl.uniprot.org/uniprot/P15692>>

}

results ONLY one pathway

VEGFA <http://purl.uniprot.org/uniprot/P15692> #no results

C1QA <http://purl.uniprot.org/uniprot/P02745> #no results

C1QB <http://purl.uniprot.org/uniprot/P02746> #no results

```

C1QC http://purl.uniprot.org/uniprot/P02747 #no results
FCGR2A http://purl.uniprot.org/uniprot/P12318 ##
<http://identifiers.org/reactome/REACT_160274.2> FCGR
activation
FCGR2C http://purl.uniprot.org/uniprot/P31995 #no results
FCGR1A http://purl.uniprot.org/uniprot/P12314 ##
<http://identifiers.org/reactome/REACT_160274.2> FCGR
activation
FCGR3B http://purl.uniprot.org/uniprot/O75015 #no results
C1R http://purl.uniprot.org/uniprot/P00736 #no results
FCGR3A http://purl.uniprot.org/uniprot/P08637 ##
<http://identifiers.org/reactome/REACT_160274.2> FCGR
activation
FCGR2B http://purl.uniprot.org/uniprot/P31994 #no results

```

#LET'S TRY TO DO THE SAME WITH ChEMBL!

ChEMBL 1

```

PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
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 cco: <http://rdf.ebi.ac.uk/terms/chembl#>

```

```

SELECT distinct ?drug ?name ?synonyms
WHERE {
  ?drug cco:highestDevelopmentPhase ?phase ;
    rdfs:label ?name;
    skos:altLabel ?synonyms .
  FILTER(?phase = 4 )
  FILTER (regex(?name, "Bevacizumab", "i") || regex( ?synonyms,
    "Bevacizumab", "i" ))
}

```

CHEMBL1201583

ChEMBL 2

```

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 cco: <http://rdf.ebi.ac.uk/terms/chembl#>
```

```
PREFIX chembl_molecule:
<http://rdf.ebi.ac.uk/resource/chembl/molecule/>
```

```
SELECT ?activity ?assay ?target ?targetcmpt ?uniprot
WHERE {
    ?activity a cco:Activity ;
    cco:hasMolecule chembl_molecule:CHEMBL1201583 ;
    cco:hasAssay ?assay .
    ?assay cco:hasTarget ?target .
    ?target cco:hasTargetComponent ?targetcmpt .
    ?targetcmpt cco:targetCmptXref ?uniprot .
    ?uniprot a cco:UniprotRef
}
```

```
##NO RESULTS
```

```
# CHEMBL 3
```

```
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 cco: <http://rdf.ebi.ac.uk/terms/chembl#>
PREFIX
biopax3:<http://www.biopax.org/release/biopax-level3.owl#>
```



```

PREFIX chembl_molecule:
<http://rdf.ebi.ac.uk/resource/chembl/molecule/>

SELECT ?activity ?assay ?target ?targetcmpt ?uniprot
?pathwayname
WHERE {
    ?activity a cco:Activity ;
    cco:hasMolecule chembl_molecule:CHEMBL1201583 ;
    cco:hasAssay ?assay .
    ?assay cco:hasTarget ?target .
    ?target cco:hasTargetComponent ?targetcmpt .
    ?targetcmpt cco:targetCmptXref ?uniprot .
    ?uniprot a cco:UniprotRef

    SERVICE <http://www.ebi.ac.uk/rdf/services/reactome/sparql> {
?uniprot rdf:type biopax3:Protein .
?uniprot biopax3:memberPhysicalEntity
[biopax3:entityReference ?dbXrefProt] .
?pathway biopax3:displayName ?pathwayname .
?pathway biopax3:pathwayComponent ?reaction .
?reaction ?rel ?uniprot
    }
}

##NO RESULTS
##
##FINAL RESULTS
## DrugBank + Reactome => 1 pathway
## ChEMBL +Reactome => 0 pathway
## DrugBank + KEGG Pathway => 0 pathway

```

Guoqian - Paclitaxel (Taxol)

<http://www.cancer.gov/cancertopics/druginfo/paclitaxel>

Paclitaxel is approved to be used alone or with other [drugs](#) to treat:

- [AIDS-related Kaposi sarcoma](#).
- [Breast cancer](#).
- [Non-small cell lung cancer](#).
- [Ovarian cancer](#).

Paclitaxel is also being studied in the treatment of other types of [cancer](#).

1. Checked three therapy drugs above in OpenPHACTS explorer

<https://explorer.openphacts.org>

Out of three drugs, only found information for Tamoxifen with two drug targets:

- a) MCF7
- b) Estrogen receptor beta

Impression: 1) the explorer asks compound names instead of drug generic name/brand name. This is an obvious gap when use the openPhacts for clinical drug targets exploration. 2) OpenPhacts does not provide a SPARQL endpoint, if using the API, there will be an additional learning curve.

2. Talked with James Malone from EBI and he helped to define a SPARQL query that can retrieve all synonyms for a generic drug name in ChEMBL.

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
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 cco: <http://rdf.ebi.ac.uk/terms/chembl#>
```

```
SELECT distinct ?drug ?name ?synonyms
WHERE {
  ?drug cco:highestDevelopmentPhase ?phase ;
  rdfs:label ?name;
  skos:altLabel ?synonyms .
  FILTER(?phase = 4 ) .
}
```

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
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 cco: <http://rdf.ebi.ac.uk/terms/chembl#>
```

```

SELECT distinct ?drug ?name ?synonyms ?phase
WHERE {
  ?drug cco:highestDevelopmentPhase ?phase ;
  rdfs:label ?name;
  skos:altLabel ?synonyms .
  FILTER regex(?synonyms, "^tamoxifen", "i") .
}

```

Tamoxifen

http://rdf.ebi.ac.uk/resource/chembl/molecule/CHEMBL1213783	TAMOXIFEN METHYL IODIDE	0
http://rdf.ebi.ac.uk/resource/chembl/molecule/CHEMBL1213785	TAMOXIFEN ISOPROPYL BROMIDE	0
http://rdf.ebi.ac.uk/resource/chembl/molecule/CHEMBL1213786	TAMOXIFEN BUTYL BROMIDE	0
http://rdf.ebi.ac.uk/resource/chembl/molecule/CHEMBL1213784	TAMOXIFEN ETHYL BROMIDE	0
http://rdf.ebi.ac.uk/resource/chembl/molecule/CHEMBL1213848	TAMOXIFEN PROPYL BROMIDE	0
http://rdf.ebi.ac.uk/resource/chembl/molecule/CHEMBL83	TAMOXIFEN	4
http://rdf.ebi.ac.uk/resource/chembl/molecule/CHEMBL786	TAMOXIFEN CITRATE	4

```

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 cco: <http://rdf.ebi.ac.uk/terms/chembl#>

```

PREFIX chembl_molecule: <http://rdf.ebi.ac.uk/resource/chembl/molecule/>

```
SELECT DISTINCT ?target ?targetLabel
WHERE {
  ?activity a cco:Activity ;
  cco:hasMolecule chembl_molecule:CHEMBL84 ;
  cco:hasAssay ?assay .
  ?assay cco:hasTarget ?target .
  ?target rdfs:label ?targetLabel .
}
```

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 cco: <http://rdf.ebi.ac.uk/terms/chembl#>

PREFIX chembl_molecule: <http://rdf.ebi.ac.uk/resource/chembl/molecule/>

```
SELECT (COUNT(DISTINCT ?target) AS ?targetCount)
WHERE
{
  {
    ?activity a cco:Activity ;
    cco:hasMolecule chembl_molecule:CHEMBL83 ;
    cco:hasAssay ?assay .
    ?assay cco:hasTarget ?target .
    ?target rdfs:label ?targetLabel .
  }
```

```
UNION
{
  ?activity a cco:Activity ;
  cco:hasMolecule chembl_molecule:CHEMBL786 ;
  cco:hasAssay ?assay .
  ?assay cco:hasTarget ?target .
}
```

?target rdfs:label ?targetLabel .

}
}

Results: 25

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 cco: <http://rdf.ebi.ac.uk/terms/chembl#>

PREFIX chembl_molecule: <http://rdf.ebi.ac.uk/resource/chembl/molecule/>

SELECT DISTINCT ?target ?targetType ?targetLabel ?dbXref

WHERE

{
{

 ?activity a cco:Activity ;

 cco:hasMolecule chembl_molecule:CHEMBL83 ;

 cco:hasAssay ?assay .

 ?assay cco:hasTarget ?target .

 ?target rdfs:label ?targetLabel .

 ?target a cco:SingleProtein .

 ?target cco:hasTargetComponent ?targetcmpt .

 ?targetcmpt cco:targetCmptXref ?dbXref .

 ?dbXref a cco:UniprotRef

}

}

Adverse Drug Event application

[Núria Queralt Rosinach from DisGeNET/OPS]: I'd like to work on OPS data in order to adapt a web service workflow developed in our group to work with linked data.

HIV Mutation Browser

Venkata Satagopam (LCSB, University of Luxembourg): I would like to create the triple store and SPARQL endpoint for HIV Mutation Browser (<http://hivmut.org>)

COEUS Data Integration

Continue the work started on COEUS' hands-on tutorial and create a knowledge integration mashup combining data from the available datasets (or from other custom data sources).

Tasks:

- Create abstract data integration model
- Import data from sources
- Create sample client applications using APIs

Data structure recovery from RDF triple store (Jesse van Dam)

I created a programs, which is based on java and a large set of sparql queries to recover the structure of the structure of a RDF database.

This tool can be used to generate an overview what is present in the database, without to have to depend on the database owner creating a good/proper usable OWL file. This can also be used to validate the quality of your rdf resource.

The folder below contains the cytoscape network representation. The properties of the classes and links are included in the cytoscape properties tables. Trick to get the dataproperties of a class you can select a class and presse alt-E, please not that also incoming links are shown. To help youn can sort on the source column.

Recovery for reactome and uniprot can be found here.

<https://drive.google.com/?authuser=0#folders/0BxJB8U-rzEh-UmVMQWFWMkZGcVE>

The tool needs the ?subj a ?type and ?type rdfs:subClassOf ?supertype definitions. Some of the subject in the uniprot do not have type definition and so are missing in the overview. Please note that the quads encoding is also not visible in the overview. There is secondary network that give an overview of the errors detected.

The program uses the following stratetagy.

1. Collect all predicate present in the database
 - a. For each of them collect how often they occur
2. Collect all the class types used in the database (?subj a ?type)
 - a. Get for each class
 - i. All the predicate going out to other class types
 1. count how many
 - ii. All the predicates going into the class from other class types
 1. count how many
3. Get all subClassOf definitions from the database
 - a. Then a filter step is performed removing all the classes that do not have any properties as these blur the overview of the structure
 - b. Then there is simplication step, that search for common parent of the class properties.

All this statistics are stored into a new triple store database holding all this information.

Then from here it can be exported to cytoscape overview and shex shape expression. Support for OWL still have to be created.

Work done during hackathon

Created a shapes definition using the Shex (<http://www.w3.org/2013/ShEx/Definition#>) definition from ericP for the structure recovered from uniprot database. See file here <https://drive.google.com/?authuser=0#folders/0BxJB8U-rzEh-UmVMQWFWMkZGcVE>. No subclass/subshape definition yet, now busy how these could be encoded.