**connecting the dots**

# Solve Problems with Knowledge Graphs

by Paula Perez

## Goals for representing knowledge

· Make things (articles, genes, antibodies, etc.) easier to find · Answer questions · Generate hypotheses controlled vocabularies (MeSH) ontologies (Gene Ontology) knowledge graphs on the Web: the SPARQL query languageknowledge plus computation = inference, the ABC model

Attachments include information that complements slide content. The following uses playful “lorem ipsum” style text to illustrate presentation of additional information.

Abbervail Dream

Blue Diamonds

Bright amazing and wonderful

Dancing around the flames

Everybody knows bird is word

Caramel Sensation

Dairy Cream

Frosty the snowman is a boss

Girls just want to have fun

Got some popsicles in the cellar

Elusive Enchantment

Fat Chance Cinnamon

Insomnia gives me time to

Inspiration slaps me in the face

Last chance for one last dance

Good Luck Charm

Hershey's Kiss

Laugh all day for no reason

Life is a box of chocolates

Live like there is no tomorrow

Ice Cream Mix

Jack Daniels

Make it up as you go

Moms cookies make everything

My room is an organized mess

Kitty Hawk

Last Man Standing

Pluto is still a planet

Six words can mean the world

Sleeping with a giant bear

Made You Look

Nabisco Cracker

Sour candy makes me twitch

The sky is not the limit

There always gonna be another

One in a Million

Peach Blossom

There no place like grandmas

Why whisper what you shout

Your the apple to my pie

## Ontology

### The word comes from philosophy:

“the branch of metaphysics dealing with the nature of being”

### In practice they are:

A set of concepts, definitions and inter-relationships.

The dividing line between “controlled vocabulary”, “thesaurus”, “ontology” is hazy and not terribly important for practical purposes.

We have hundreds of ontologies in biology.

Attachments include information that complements slide content. The following uses playful “lorem ipsum” style text to illustrate presentation of additional information.

Abbervail Dream

Blue Diamonds

Bright amazing and wonderful

Dancing around the flames

Everybody knows bird is word

Caramel Sensation

Dairy Cream

Frosty the snowman is a boss

Girls just want to have fun

Got some popsicles in the cellar

Elusive Enchantment

Fat Chance Cinnamon

Insomnia gives me time to

Inspiration slaps me in the face

Last chance for one last dance

Good Luck Charm

Hershey's Kiss

Laugh all day for no reason

Life is a box of chocolates

Live like there is no tomorrow

Ice Cream Mix

Jack Daniels

Make it up as you go

Moms cookies make everything

My room is an organized mess

Kitty Hawk

Last Man Standing

Pluto is still a planet

Six words can mean the world

Sleeping with a giant bear

Made You Look

Nabisco Cracker

Sour candy makes me twitch

The sky is not the limit

There always gonna be another

One in a Million

Peach Blossom

There no place like grandmas

Why whisper what you shout

Your the apple to my pie

## Gene Ontology

### Started in 1999 as a collaboration among three Model Organism Databases

GO:0008150biological\_processGO:0022607cellular componentassemblyGO:0016043cellular componentorganizationGO:0009987cellular processGO:0071840cellular componentorganization orbiogenesisGO:0048856anatomical structuredevelopmentGO:0044763single-organism cellularprocessGO:0030198extracellular matrixorganizationGO:0043062extracellular structureorganizationGO:0009888tissue developmentGO:0071470cellular response toosmotic stressGO:0006972hyperosmotic responseGO:0071214cellular response toabiotic stimulusGO:0006970response to osmoticstressGO:0033554cellular response tostressGO:0009628response to abioticstimulusGO:0051716cellular response tostimulusGO:0006950response to stressGO:0050896response to stimulusMore specific conceptLess specific concept

## Gene Ontology (GO)

#### A way to capture biological knowledge for individual gene products in a computable form

#### A set of concepts and their relationships to each other arranged as a hierarchy

OntologySlide credit: Mélanie Courtot, Ph.D.

https://www.ebi.ac.uk/QuickGO/

Abbervail Dream

Blue Diamonds

Bright amazing and wonderful

Dancing around the flames

Everybody knows bird is word

Caramel Sensation

Dairy Cream

Frosty the snowman is a boss

Girls just want to have fun

Got some popsicles in the cellar

Elusive Enchantment

Fat Chance Cinnamon

Insomnia gives me time to

Inspiration slaps me in the face

Last chance for one last dance

Good Luck Charm

Hershey's Kiss

Laugh all day for no reason

Life is a box of chocolates

Live like there is no tomorrow

Ice Cream Mix

Jack Daniels

Make it up as you go

Moms cookies make everything

My room is an organized mess

Kitty Hawk

Last Man Standing

Pluto is still a planet

Six words can mean the world

Sleeping with a giant bear

Made You Look

Nabisco Cracker

Sour candy makes me twitch

The sky is not the limit

There always gonna be another

One in a Million

Peach Blossom

There no place like grandmas

Why whisper what you shout

Your the apple to my pie

· protein kinase activity· insulin receptor activity· mitochondrion· mitochondrial matrix· mitochondrial inner membraneinsulin receptorinsulin

## GO Branches

### subtitle

#### Molecular Function

An elemental activity or task or job

#### Biological Process

A commonly recognized series of events

#### Cellular Component

Where a gene product is located

OntologySlide credit: Mélanie Courtot, Ph.D.

https://www.ebi.ac.uk/QuickGO/

Abbervail Dream

Blue Diamonds

Bright amazing and wonderful

Dancing around the flames

Everybody knows bird is word

Caramel Sensation

Dairy Cream

Frosty the snowman is a boss

Girls just want to have fun

Got some popsicles in the cellar

Elusive Enchantment

Fat Chance Cinnamon

Insomnia gives me time to

Inspiration slaps me in the face

Last chance for one last dance

Good Luck Charm

Hershey's Kiss

Laugh all day for no reason

Life is a box of chocolates

Live like there is no tomorrow

Ice Cream Mix

Jack Daniels

Make it up as you go

Moms cookies make everything

My room is an organized mess

Kitty Hawk

Last Man Standing

Pluto is still a planet

Six words can mean the world

Sleeping with a giant bear

Made You Look

Nabisco Cracker

Sour candy makes me twitch

The sky is not the limit

There always gonna be another

One in a Million

Peach Blossom

There no place like grandmas

Why whisper what you shout

Your the apple to my pie

WISDOMKNOWLEDGEINFORMATIONDATAKnowing what to do with Drug A ...Drug A caused Gene X to be expressedGene X is expressedRNASeq reads

## Knowledge Graphs

#### Also called “knowledge bases” to distinguish them from databases.

#### An integrated collection of assertions or claims represented in something that can be visualized as a graph and is technically very much like a database.

## Example knowledge graphs

#### Wikidata: The structured equivalent of Wikipedia

#### UniProt Knowledge Base: Manually curated Protein knowledge base

#### Microsoft Knowledge Graph (“Satori”)

#### Google Knowledge Graph

Attachments include information that complements slide content. The following uses playful “lorem ipsum” style text to illustrate presentation of additional information.

Abbervail Dream

Blue Diamonds

Bright amazing and wonderful

Dancing around the flames

Everybody knows bird is word

Caramel Sensation

Dairy Cream

Frosty the snowman is a boss

Girls just want to have fun

Got some popsicles in the cellar

Elusive Enchantment

Fat Chance Cinnamon

Insomnia gives me time to

Inspiration slaps me in the face

Last chance for one last dance

Good Luck Charm

Hershey's Kiss

Laugh all day for no reason

Life is a box of chocolates

Live like there is no tomorrow

Ice Cream Mix

Jack Daniels

Make it up as you go

Moms cookies make everything

My room is an organized mess

Kitty Hawk

Last Man Standing

Pluto is still a planet

Six words can mean the world

Sleeping with a giant bear

Made You Look

Nabisco Cracker

Sour candy makes me twitch

The sky is not the limit

There always gonna be another

One in a Million

Peach Blossom

There no place like grandmas

Why whisper what you shout

Your the apple to my pie

## Example: Google Knowledge Graph

#### search for: Vemurafenib

#### 2,310,000 results

#### 1 infobox

#### 1 node in Google Knowledge Graph

## Why Knowledge Graphs?

### Answer explicit questions; uncover implicit relations

?

## Implicit relations for hypothesis generation

### ABC model

Co-occurs in an article withRaynaud’s SyndromeACo-occurs in an article withBC?Dietary fish oil· platelet inhibition· vasodilation· lower blood viscosity

Attachments include information that complements slide content. The following uses playful “lorem ipsum” style text to illustrate presentation of additional information.

Abbervail Dream

Blue Diamonds

Bright amazing and wonderful

Dancing around the flames

Everybody knows bird is word

Caramel Sensation

Dairy Cream

Frosty the snowman is a boss

Girls just want to have fun

Got some popsicles in the cellar

Elusive Enchantment

Fat Chance Cinnamon

Insomnia gives me time to

Inspiration slaps me in the face

Last chance for one last dance

Good Luck Charm

Hershey's Kiss

Laugh all day for no reason

Life is a box of chocolates

Live like there is no tomorrow

Ice Cream Mix

Jack Daniels

Make it up as you go

Moms cookies make everything

My room is an organized mess

Kitty Hawk

Last Man Standing

Pluto is still a planet

Six words can mean the world

Sleeping with a giant bear

Made You Look

Nabisco Cracker

Sour candy makes me twitch

The sky is not the limit

There always gonna be another

One in a Million

Peach Blossom

There no place like grandmas

Why whisper what you shout

Your the apple to my pie

## Open Discovery and Closed Discovery

#### Open, you don’t know what C or B is (e.g. disease -> ?drug)

#### Closed, you know what C is and are looking for B (e.g. disease – why? – drug)

ABC?

DRUGNameRegistered IndicationNew Indicationrepurposing potentialDISEASEKnown Drug in DrugBankNumber of Medline AbstractsNumber in Clinical Trial RegistryRepurposing Potential

## Example question: drug repurposing

#### For a given drug, what diseases might it be used to treat?

Attachments include information that complements slide content. The following uses playful “lorem ipsum” style text to illustrate presentation of additional information.

Abbervail Dream

Blue Diamonds

Bright amazing and wonderful

Dancing around the flames

Everybody knows bird is word

Caramel Sensation

Dairy Cream

Frosty the snowman is a boss

Girls just want to have fun

Got some popsicles in the cellar

Elusive Enchantment

Fat Chance Cinnamon

Insomnia gives me time to

Inspiration slaps me in the face

Last chance for one last dance

Good Luck Charm

Hershey's Kiss

Laugh all day for no reason

Life is a box of chocolates

Live like there is no tomorrow

Ice Cream Mix

Jack Daniels

Make it up as you go

Moms cookies make everything

My room is an organized mess

Kitty Hawk

Last Man Standing

Pluto is still a planet

Six words can mean the world

Sleeping with a giant bear

Made You Look

Nabisco Cracker

Sour candy makes me twitch

The sky is not the limit

There always gonna be another

One in a Million

Peach Blossom

There no place like grandmas

Why whisper what you shout

Your the apple to my pie

## Implicit relations for hypothesis generation

### ABC model for drug repurposing

physical interactiondrugAgenetic associationBC?diseasegenes

Asking forMetforminMetformin?diseasetreatstreatsConstraintsResult:Type 2 diabetes

## Query a knowledge graph with SPARQL

#### SPARQL protocol and RDF query language

#### RDF: Resource Description Framework (common standard for storing knowledge graphs)

#### A SPARQL query = a partially completed graph

?’s show what you are looking for

remainder constrains the search

## Metformin’s unique id: Q19484 Treats property id: P2175

Metformintreats?disease=

## Metformin’s unique id: Q19484 Treats property id: P2175

Metformintreats?disease

## Example question: drug repurposing

### “For a given drug, what diseases might it be used to treat?”

mode of actionAgonistAntagonistOtherUnknownGENESNPsymbolassociationP-ValueOdds RatioGWAS or PheWASDRUGNameRegistered IndicationNew Indicationrepurposing potentialDISEASEKnown Drug in DrugBankNumber of Medline AbstractsNumber in Clinical Trial RegistryRepurposing Potential

interacts withencoded bygenetic associationtreats???drugprotein?diseasegene

## Example question: repurposing Metformin

interacts withencoded bygenetic associationtreats??MetforminSolute carrier family 22member 3prostate cancerSLC22A3

Attachments include information that complements slide content. The following uses playful “lorem ipsum” style text to illustrate presentation of additional information.

Abbervail Dream

Blue Diamonds

Bright amazing and wonderful

Dancing around the flames

Everybody knows bird is word

Caramel Sensation

Dairy Cream

Frosty the snowman is a boss

Girls just want to have fun

Got some popsicles in the cellar

Elusive Enchantment

Fat Chance Cinnamon

Insomnia gives me time to

Inspiration slaps me in the face

Last chance for one last dance

Good Luck Charm

Hershey's Kiss

Laugh all day for no reason

Life is a box of chocolates

Live like there is no tomorrow

Ice Cream Mix

Jack Daniels

Make it up as you go

Moms cookies make everything

My room is an organized mess

Kitty Hawk

Last Man Standing

Pluto is still a planet

Six words can mean the world

Sleeping with a giant bear

Made You Look

Nabisco Cracker

Sour candy makes me twitch

The sky is not the limit

There always gonna be another

One in a Million

Peach Blossom

There no place like grandmas

Why whisper what you shout

Your the apple to my pie

## PubChem

#### transition text

## Identify a unique item

Atorvastatin (CID60823)

identified by InChIKey XUKUURHRXDUEBC-KAYWLYCHSA-N

as defined by molar mass unit (obo:UO\_0000055)

expressed as double floating point

also identified in ChEBI (SID103554720)

compound:CID60823

descr:CID60823\_Molecular\_Weight

obo:UO\_0000055

“558.639803”^^xsd:double

sio:CHEMINF\_000334

substance:SID103554720

syno:MD5\_9a05646d461669f86de312d88ab5748a

ChEBI:39548

“atorvastatin”@en

sio:CHEMINF\_000339

sio:has-attribute

sio:is-attribute-of

sio:has-attribute

sio:is-attribute-of

sio:has-unit

rdf:type

rdf:type

rdf:type

sio:has-value

sio:has-value

The following uses playful “lorem ipsum” style text to illustrate presentation of additional information.

Abbervail Dream

Blue Diamonds

Bright amazing and wonderful

Dancing around the flames

Everybody knows bird is word

Caramel Sensation

Dairy Cream

Frosty the snowman is a boss

Girls just want to have fun

Got some popsicles in the cellar

Elusive Enchantment

Fat Chance Cinnamon

Insomnia gives me time to

Inspiration slaps me in the face

Last chance for one last dance

Good Luck Charm

Hershey's Kiss

Laugh all day for no reason

Life is a box of chocolates

Live like there is no tomorrow

Ice Cream Mix

Jack Daniels

Make it up as you go

Moms cookies make everything

My room is an organized mess

Kitty Hawk

Last Man Standing

Pluto is still a planet

Six words can mean the world

Sleeping with a giant bear

Made You Look

Nabisco Cracker

Sour candy makes me twitch

The sky is not the limit

There always gonna be another

One in a Million

Peach Blossom

There no place like grandmas

Why whisper what you shout

Your the apple to my pie

## Link an item to similar items

Express similarity:

by molecular formula & bonds

by isotopic composition

by connectivity

by 2 dimension shape

by 3 dimension shape

compound:CID10507504

compound:CID60823

compound:CID23665101

compound:CID10507504

compound:CID10507504

compound:CID10030610

compound:CID11330946

sio:is\_stereoisomer\_of

sio:has\_component

More about linking similar items.

Similarity to other substances

item-1

(3R,5R)-7-[2-(4-fluorophenyl)-3-phenyl-4-(phenylcarbamoyl)-5-propan-2-yl(314C)pyrrol-1-yl]-3,5-dihydroxyheptanoic acid

XUKUURHRXDUEBC-IEILOPHXSA-N

item-2

(3R,5R)-7-[2-cyclopropyl-5-(4-fluorophenyl)-4-phenyl-3-(phenylcarbamoyl)pyrrol-1-yl]-3,5-dihydroxyheptanoic acid

WDKDRVHIGPVFCW-KAYWLYCHSA-N

item-3

(E,3R,5S)-7-[5-[(3-chlorophenyl)carbamoyl]-3,4-bis(4-fluorophenyl)-1-propan-2-ylpyrrol-2-yl]-3,5-dihydroxyhept-6-enoic acid

JNJZDKOQTIGNNB-FDFRSNTISA-N

Has component

Atorvastatin sodium

VVRPOCPLIUDBSA-CNZCJKERSA-M

## Define measurement to link similar items

Define items:

similar to Atorvastatin

by chemical structure

by fingerprint

using TanimotoScore

expressed as a double precision decimal

compound:CID60823

sio:CHEMINF\_000333

“0.98“^^xsd:double

nbr:CID60823\_CID10030610\_2DSimilarity

compound:CID10030610

vocab:2D\_structural\_similarity

vocab:2D\_Fingerprint\_TanimotoScore

nbr:CID60823\_CID10030610\_2DTanimotoScore

rdf: type

rdf: type

sio:is-output-of

sio:has-measurement-value

sio:has-value

The following uses playful “lorem ipsum” style text to illustrate presentation of additional information.

Abbervail Dream

Blue Diamonds

Bright amazing and wonderful

Dancing around the flames

Everybody knows bird is word

Caramel Sensation

Dairy Cream

Frosty the snowman is a boss

Girls just want to have fun

Got some popsicles in the cellar

Elusive Enchantment

Fat Chance Cinnamon

Insomnia gives me time to

Inspiration slaps me in the face

Last chance for one last dance

Good Luck Charm

Hershey's Kiss

Laugh all day for no reason

Life is a box of chocolates

Live like there is no tomorrow

Ice Cream Mix

Jack Daniels

Make it up as you go

Moms cookies make everything

My room is an organized mess

Kitty Hawk

Last Man Standing

Pluto is still a planet

Six words can mean the world

Sleeping with a giant bear

Made You Look

Nabisco Cracker

Sour candy makes me twitch

The sky is not the limit

There always gonna be another

One in a Million

Peach Blossom

There no place like grandmas

Why whisper what you shout

Your the apple to my pie

## Identify a specific dataset from an investigation

A dataset about …

Atorvastatin

recorded for the ChEMBL data source

provided by the EBI organization

about a concept assigned the label:

“Biological Properties”

substance:SID103554720

dcterms:Dataset

skos:Concept

source:ChEMBL

http://www.ebi.ac.uk/chembldb/

source:European\_Bioinformatics\_Institute\_EBI\_ChEMBL

skos:ConceptScheme

vocab:Biological\_Properties

vocab:Substance\_Categorization\_Classification

“Biological Properties“

sio:has-value

foaf:homepage

dcterms:source

pav:providedBy

rdf:type

rdf:type

The following uses playful “lorem ipsum” style text to illustrate presentation of additional information.

Abbervail Dream

Blue Diamonds

Bright amazing and wonderful

Dancing around the flames

Everybody knows bird is word

Caramel Sensation

Dairy Cream

Frosty the snowman is a boss

Girls just want to have fun

Got some popsicles in the cellar

Elusive Enchantment

Fat Chance Cinnamon

Insomnia gives me time to

Inspiration slaps me in the face

Last chance for one last dance

Good Luck Charm

Hershey's Kiss

Laugh all day for no reason

Life is a box of chocolates

Live like there is no tomorrow

Ice Cream Mix

Jack Daniels

Make it up as you go

Moms cookies make everything

My room is an organized mess

Kitty Hawk

Last Man Standing

Pluto is still a planet

Six words can mean the world

Sleeping with a giant bear

Made You Look

Nabisco Cracker

Sour candy makes me twitch

The sky is not the limit

There always gonna be another

One in a Million

Peach Blossom

There no place like grandmas

Why whisper what you shout

Your the apple to my pie

## Identify a specific study & its evidence

Study summarized as: “A new class of anti-inflammatory ester prodrugs”

published by DiscoveryGate

about SID103164874 (CHEBI:15365) and related compound – aspirin

based on a bioassay with the title “Inhibition of ovine COX1 by enzyme immunoassay”

bao:measure-group

reference:PMID19500994

http://www.ncbi.nlm.nih.gov/pubmed/19500994

“ProstaglandinG/H synthase 1”@en

protein:GI548481

rdf:type

rdf:type

dcterms:title

pr:PR\_000006933

“0.35”^^xsd:double

fabio:JournalArticle

“Inhibition of ovine COX1by enzyme immunoassay”@en

bao:ic50

measuregroup:AID447528

ops:Micromolar

substance:SID103164874

iao:is-about

endpoint:SID10316487\_AID447528

qudt:unit

The following uses playful “lorem ipsum” style text to illustrate presentation of additional information.

Abbervail Dream

Blue Diamonds

Bright amazing and wonderful

Dancing around the flames

Everybody knows bird is word

Caramel Sensation

Dairy Cream

Frosty the snowman is a boss

Girls just want to have fun

Got some popsicles in the cellar

Elusive Enchantment

Fat Chance Cinnamon

Insomnia gives me time to

Inspiration slaps me in the face

Last chance for one last dance

Good Luck Charm

Hershey's Kiss

Laugh all day for no reason

Life is a box of chocolates

Live like there is no tomorrow

Ice Cream Mix

Jack Daniels

Make it up as you go

Moms cookies make everything

My room is an organized mess

Kitty Hawk

Last Man Standing

Pluto is still a planet

Six words can mean the world

Sleeping with a giant bear

Made You Look

Nabisco Cracker

Sour candy makes me twitch

The sky is not the limit

There always gonna be another

One in a Million

Peach Blossom

There no place like grandmas

Why whisper what you shout

Your the apple to my pie

## Provide additional information about a bioassay

Bioassay title: “Primary Cell-based assay for inhibitors of the Retinoic Acid Receptor-related orphan receptor A (RORA)”

summarized as: “probe development efforts to identify novel modulators of ROR” as measured by dose-response cell-based assay for RORA inhibitors

bao:bioassay

bioassay:AID2139

bao:primary-assay

bao:confirmatory-assay

bao:has-assay-stage

bao:has-summary-assay

bao:has-assay-stage

bao:has-assay-stage

bao:summary-assay

measuregroup:AID610

bioassay:AID610

bioassay:AID561

bao:measure-group

rdf:type

rdf:type

bao:has-measure-group

bao:is-measure-group-of

The following uses playful “lorem ipsum” style text to illustrate presentation of additional information.

Abbervail Dream

Blue Diamonds

Bright amazing and wonderful

Dancing around the flames

Everybody knows bird is word

Caramel Sensation

Dairy Cream

Frosty the snowman is a boss

Girls just want to have fun

Got some popsicles in the cellar

Elusive Enchantment

Fat Chance Cinnamon

Insomnia gives me time to

Inspiration slaps me in the face

Last chance for one last dance

Good Luck Charm

Hershey's Kiss

Laugh all day for no reason

Life is a box of chocolates

Live like there is no tomorrow

Ice Cream Mix

Jack Daniels

Make it up as you go

Moms cookies make everything

My room is an organized mess

Kitty Hawk

Last Man Standing

Pluto is still a planet

Six words can mean the world

Sleeping with a giant bear

Made You Look

Nabisco Cracker

Sour candy makes me twitch

The sky is not the limit

There always gonna be another

One in a Million

Peach Blossom

There no place like grandmas

Why whisper what you shout

Your the apple to my pie

## Links among PubChem subdomains

At the center: measurments and evidence (bioassays)

Biologics are proteins and genes from specific sources

Items are chemical compounds and substances

Evidence is documented in a study (reference)

similar protein

protein

biosystem

domain

neighbor

gene

measuregroup

endpoint

reference

bioassay

source

substance

synonym

descriptor

inchikey

compound

neighbor

component compound

parent compound

same connectivity

isotopomer

stereoisomer

similar compound

## Summary: PubChem Data Model

More about PubChem RDF