

LJCD Ljubljana University Hospital Center

Faster and better results for people living with a [rare] disease by enabling computers to automatically Find, Access, Interoperate, and Reuse data from across the world

Marco Bošek  
Bioinformatics group, Human Genetics Department, LJCD  
Ljubljana, Slovenia, 10.10.2018, 10:00-10:30



## Why FAIR? Patient perspective



### SECTION: FAIR GAME TO FIND A TREATMENT

Why is FAIR important for data science?  
Identify real-life challenges and solutions for federated data analysis by finding a treatment for rare disease patients: a FAIR game.

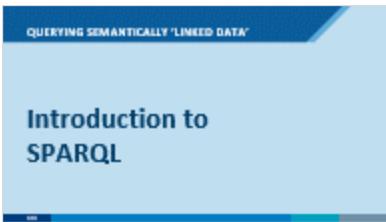


## The FAIR game explained An introduction to 'ontologised data'



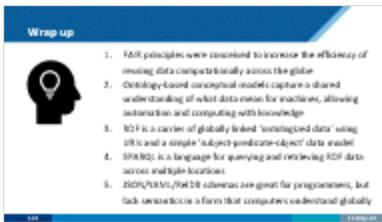
### QUERYING SEMANTICALLY 'LINKED DATA'

## Introduction to SPARQL

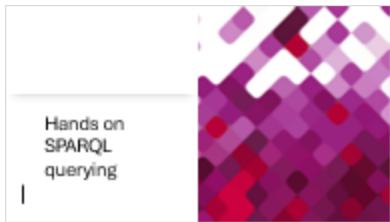


### Wrap up

1. FAIR principles were introduced to increase the efficiency of reusing data computationally across the globe
2. Ontology-based conceptual models capture a shared understanding of what data means for machines, allowing interaction and computing with knowledge
3. RDF is a carrier of globally linked 'ontologised data' using URIs and a simple 'subject-predicate-object' data model
4. SPARQL is a language for querying and retrieving RDF data across multiple locations
5. JSON/JSONML/JSONB structures are great for programmes, but lack semantics in a form that computers understand globally



### Hands on SPARQL querying



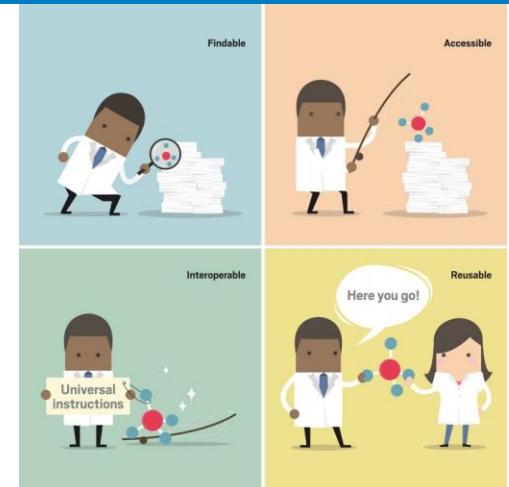
# Faster and better results for people living with a (rare) disease by enabling computers to automatically Find, Access, Interoperate, and Reuse data from across the world

Automated data science by creating and using F+A+I+R data for machines

Marco Roos

Biosemantics group, Human Genetics Department, LUMC

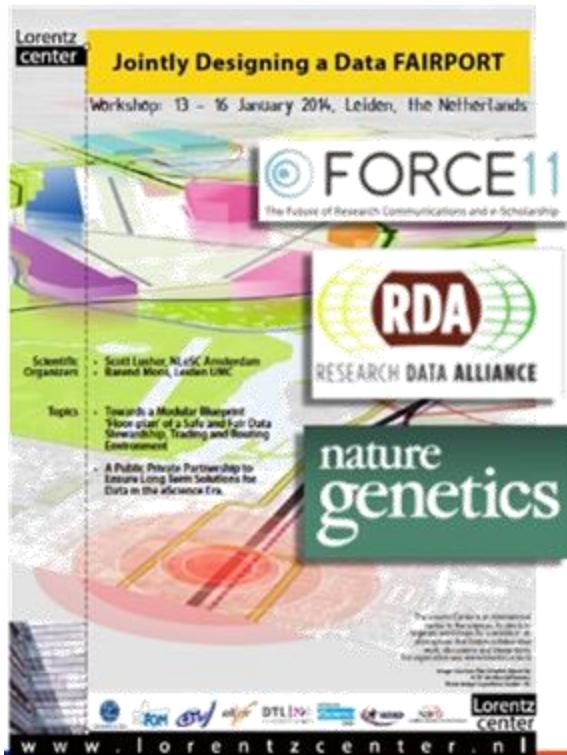
ROOM S4-42, EINTHOVENWEG 20, LEIDEN



<https://cen.acs.org/policy/publishing/Chemistry-data-should-FAIR-proponents/97/35>

# Lorentz workshop, Leiden 2014

## The motivation behind FAIR principles



## Global leaders in data science

Brainstorming solutions for

- Data use & reuse *unacceptably* inefficient
- Poorly reproducible
- Especially when multiple sources are needed

Starting point: efficiency of airports and harbours



## Outcome = requirements

Data should minimally be

**Findable, Accessible, Interoperable, Reusable**

*for humans **and machines***



## Motivation for *implementing FAIR principles*



Price Waterhouse Coopers for EC:

estimated *quantifiable* cost for the European economy of data **not** Findable, Accessible, Interoperable, Reusable

**> €10.2 billion per year in Europe**

(unquantifiable elements probably +€16 billion)

sources: [EC report cost of not having FAIR research data](#) (2018), DG Research and Innovation (EC) & PwC EU Services

€€€

# Learning goals



After this morning you, the next generation data scientist, can describe

1. why FAIR principles were conceived
2. how semantic modelling combined with web technology make the meaning of data unambiguous for humans and understandable '*for machines*'
3. what distinguishes FAIR for humans and programmers (web sites, schemas, relational models), and FAIR for machines (ontologies, Semantic Web/Linked Data)
4. that a language exists to query data across multiple locations across the globe in terms of a semantic model

Learning method:  
interactive, time to  
think & discuss,  
followed by input  
from teacher.

# Plan

15:15 Introduction

15:25 Why is FAIR important from a patient point of view?

Assess an interview with Lizanne, a woman living with Duchenne Muscular Dystrophy (you)

15:40 FAIR for Rare Diseases (Marco)

15:50 Why is FAIR important for data science and data science for FAIR?

Identify real-life challenges for federated data analysis by finding a treatment for rare disease patients in a game (you)

16:00 Share challenges and opportunities for health data science (you)

16:05 *Short break*

16:10 The FAIR game explained, introduction to ‘ontologising data’ (Marco)

16:30 SPARQL A language to query ontologised data across the world (Marco)

16:40 Wrap-up: what was FAIR all about? (All)

# Why FAIR? Patient perspective

# Instructions



[https://1drv.ms/v/s!Au77uNdV\\_FAgZR3Ws8tAvSP\\_CVqnQ](https://1drv.ms/v/s!Au77uNdV_FAgZR3Ws8tAvSP_CVqnQ)



Objective: to learn why FAIR principles exist from a patient's perspective.

1. Watch the video of Lizanne
2. Individual: during/after the video, write down keywords why FAIR principles may be important from Lizanne's point of view
3. Groups of two or three:
  1. Discuss your points and what Lizanne should expect from professionals generating or collecting health data
  2. Discuss what you would gain as a data scientist
4. Share via wooclap

# Why is FAIR important – patient perspective



# Instructions



[https://1drv.ms/v/s!Au77uNdV\\_FAgZR3Ws8tAvSP\\_CVqnQ](https://1drv.ms/v/s!Au77uNdV_FAgZR3Ws8tAvSP_CVqnQ)



<https://app.wooclap.com/TUFCJO>

Objective: to learn why FAIR principles exist from a patient's perspective.

1. Watch the video of Lizanne
2. Individual: during/after the video, write down keywords why FAIR principles may be important from Lizanne's point of view
3. Groups of two or three:
  1. Discuss your points and what Lizanne should expect from professionals generating or collecting health data
  2. Discuss what you would gain as a data scientist
4. Share via wooclap: <https://app.wooclap.com/TUFCJO>

**One in 17 people suffers from  
a rare disease**



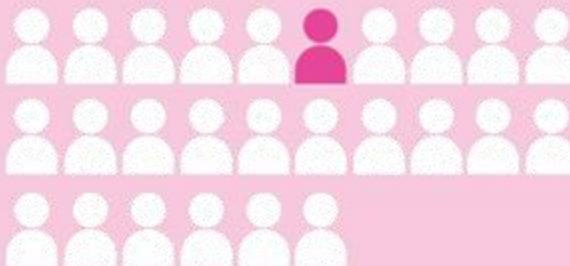
# FAIR and rare diseases

RARE  
DISEASES



7% OF THE  
POPULATION  
ARE AFFECTED BY  
RARE DISEASES

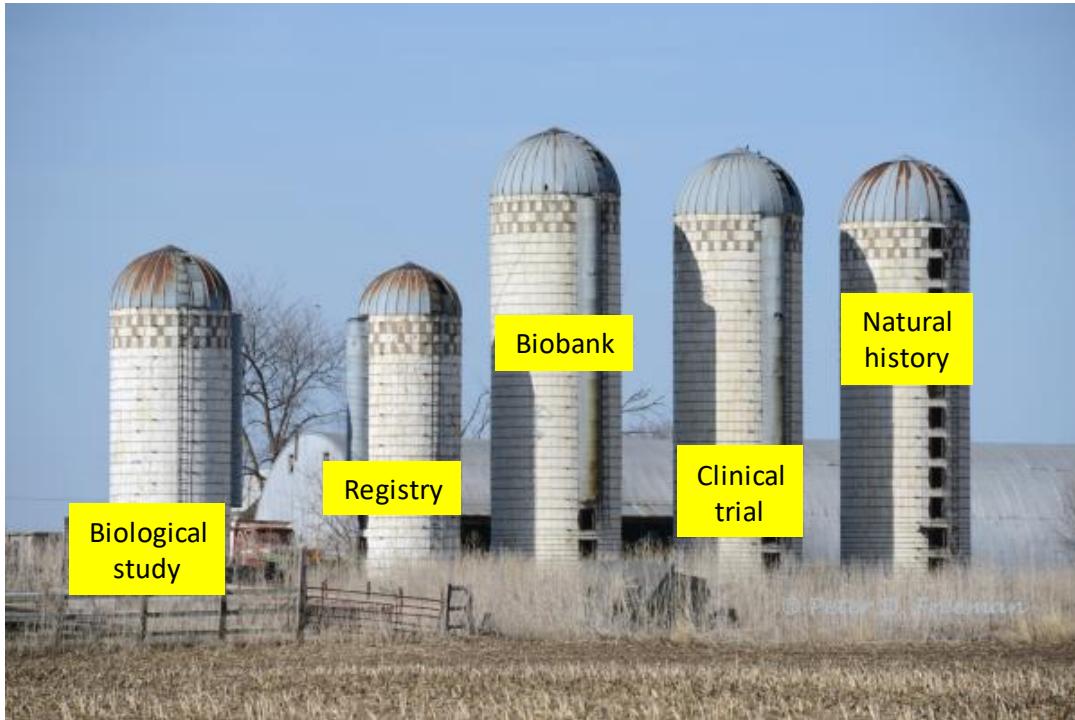
THE EU CLASSES A  
DISEASE AS 'RARE' WHEN  
LESS THAN  
1 IN 2000 SUFFER



OVER 7000  
DISEASES  
BIOSAMPLES,  
DISEASE &  
PATIENT INFO,  
OMICS,  
GENOTYPE-  
PHENOTYPE

DELAY TO TIME TO DIAGNOSIS ESTIMATED 4-5 YEARS.

# 7000 x N siloes



# Silos, silos, silos



# Silos, silos, silos



# Silos, silos, silos



# Absolute need to combine



# Achieving the global vision



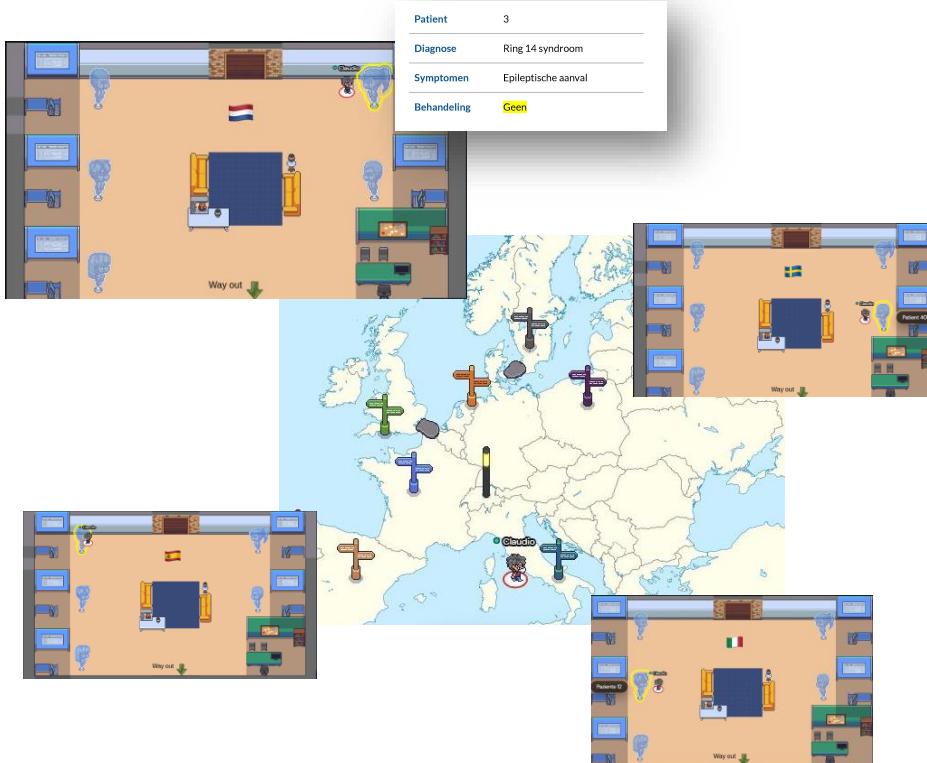
Usability of data for rare disease care and research **must be** brought to higher, global levels, i.e. more Findable, Accessible, Interoperable, and Reusable for automated computational analysis

## SECTION: FAIR GAME TO FIND A TREATMENT

### Why is FAIR important for data science?

Identify real-life challenges and solutions for federated data analysis by finding a treatment for rare disease patients: a FAIR game.

# Empathy with the algorithm Problem Solving – Gather.town Game



Objective: to identify bottlenecks in using distributed health data

## Steps

1. Go to a registry
2. Find a patient without a treatment
3. Navigate the other registries to find a potential treatment by matching symptoms

# Patient – FAIR Data Game

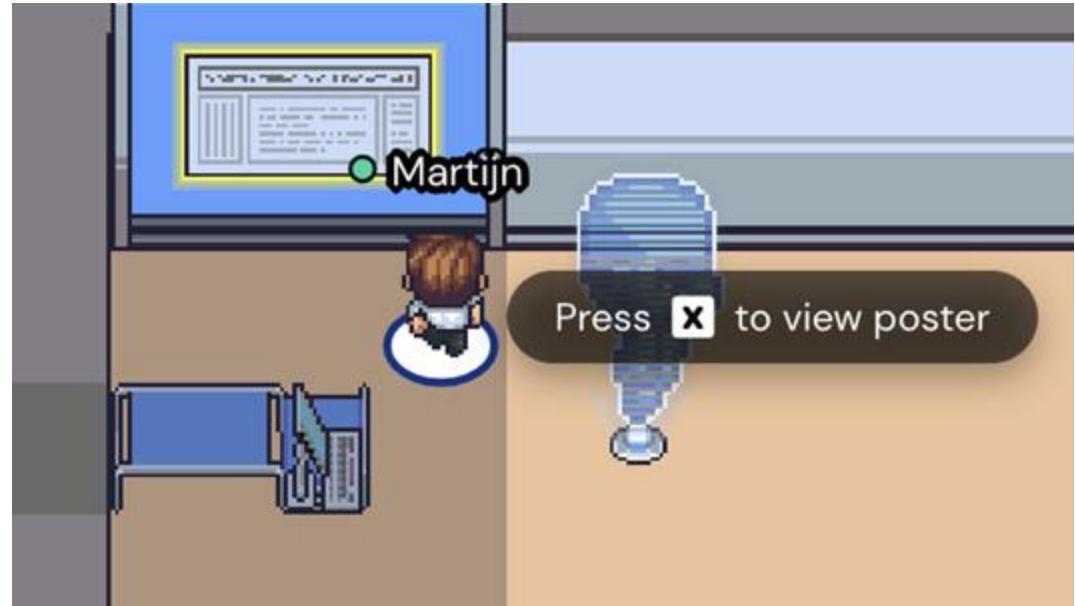
See patient's **number**



# Patient – FAIR Data Game

Open patient's  
**record** by:

- walking up to the **poster** behind the patient and
- **pressing X key.**



# Patient – FAIR Data Game

---

Patient 3

---

Diagnose Ring 14 syndroom

---

Symptomen Epileptische aanval

---

Behandeling Geen

# Instructions



[https://app.gather.town/invite  
?token=2v6u8drITGKbTB9sRAQq](https://app.gather.town/invite?token=2v6u8drITGKbTB9sRAQq)



[Go to wooclap.com  
Event code IYJPBR  
https://app.wooclap.com/IYJPBR?  
from=instruction-slide](https://app.wooclap.com/IYJPBR?from=instruction-slide)

Go to [gather.town](https://gather.town)  
Enter the FAIR game with small groups

While finding candidate treatments for symptoms of patients without a treatment write down

- Bottlenecks to finding a treatment
- Possible solutions

~10 minutes

# VISIT THE DATA ACCESS COMMITTEE FIRST

Exit

0 people are confused

How to participate?



Copy participation link

1 Go to [wooclap.com](https://wooclap.com)

2 Enter the event code in the top banner

Event code  
**SFMPPQ**

Enable answers by SMS

Questions - / 7 + Messages 100 % 🔎

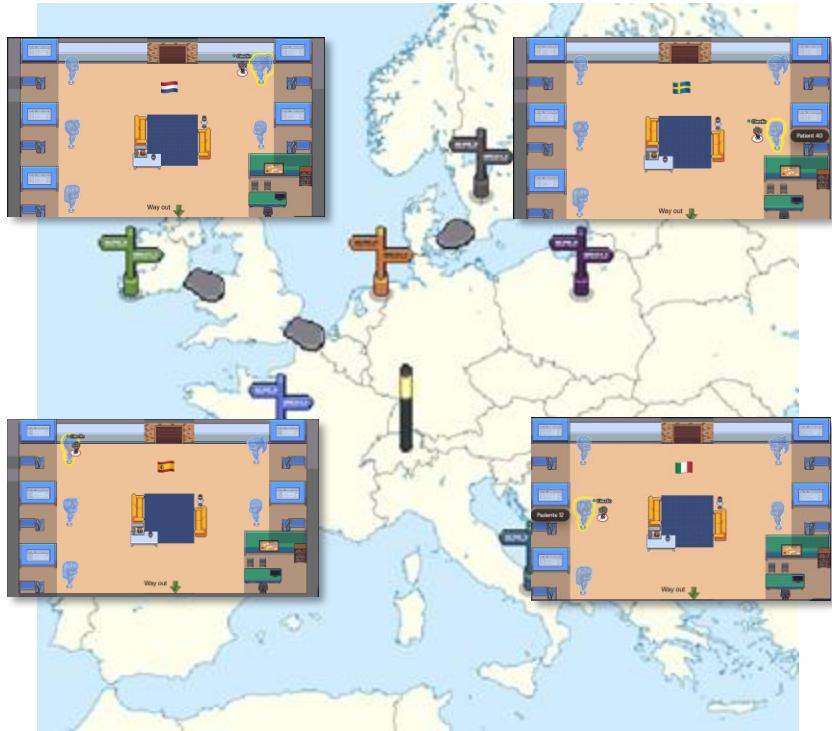
wooclap 0 people 0 messages

<https://app.wooclap.com/SFMPPQ?from=instruction-slide>

# The FAIR game explained

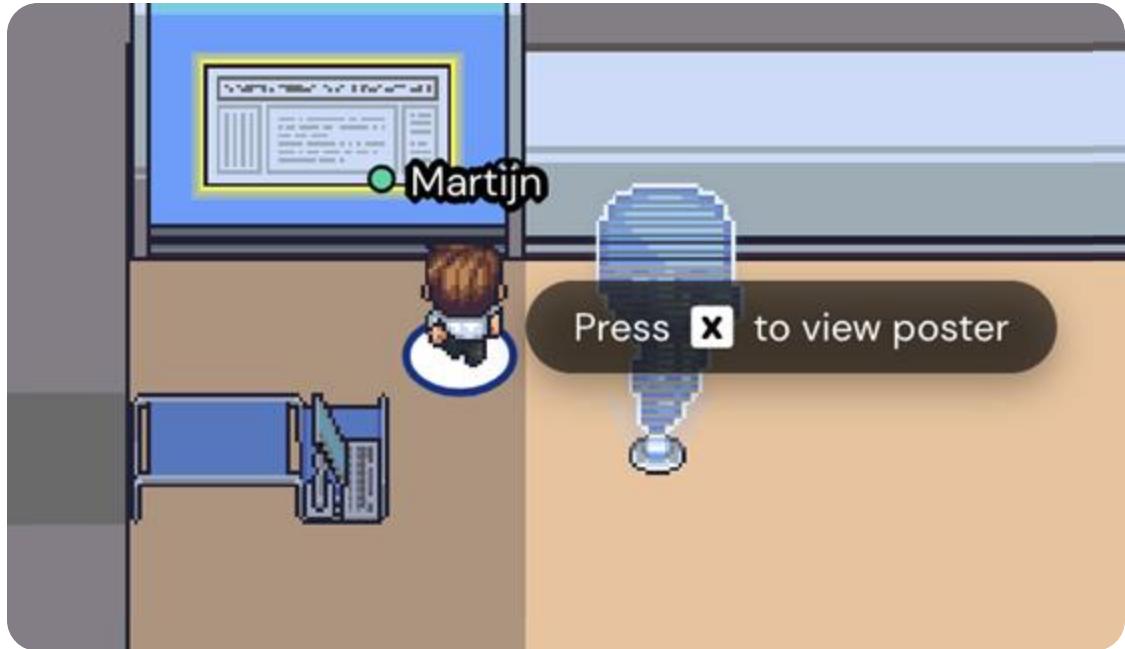
## An introduction to ‘ontologised data’

- 4 registries
  - All with their own data definitions (languages)
    - Patient No / Diagnosis / Symptoms / Treatment
- 5 patients per registry
- 1 patient without a possible treatment
- **Objective:**  
Figure out which drug treats your patient



Open patient's  
record by:

- walking up to the **poster** behind the patient and
- **pressing X key.**



Patient	1
Diagnose	Taaislijmziekte
Symptomen	Hoofd- en nekafwijking
Behandeling	(24S)-methylcalciol

# Four registries



Monika

Krankheit Ringbildung  
Chromosom 14, Salaam-  
Anfälle, (Keine  
Behandlung)



Rajaram

பெர்ரி நோய்க்குறி,  
வலிப்பு  
தாக்குதல்கள்,  
லாமோட்டரேஜின்

The ‘I’ of FAIR  
**Assuming data are  
findable and  
accessible:  
How can we find the  
candidate treatment  
efficiently?**



Annika

Ring-14-sjúkumynd,  
sankta Vitusar dansur,  
eingin viðgerð



Pietro

sindrome Perry, sbalzi  
d'umore estremi, ossalato

# Find the treatment experiment

 Monika	 Annika	 Rajaram	 Pietro	 FAIR
Krankheit Ringbildung Chromosom 14, Salaam-Anfälle, (Keine Behandlung)	Ring-14-sjúkumynd, sankta Vitusar dansur, eingin viðgerð	பெர்ரி நோய்க்குறி, வலிப்பு தாக்குதல்கள், லாமோட்டரஜின்	sindrome Perry, sbalzi d'umore estremi, ossalato	Local registry

# Translate to a language we can understand

Monika	Annika	Rajaram	Pietro	FAIR
Krankheit Ringbildung Chromosom 14, Salaam-Anfälle, (Keine Behandlung)	Ring-14-sjúkumynd, sankta Vitusar dansur, eingin viðgerð	ପେର୍ରି ନୋପକୁରି, ଲାକ୍ଷଣ୍ୟାବଳୀ, ଲାମୋଟାର୍ଜିଙ୍	sindrome Perry, sbalzi d'umore estremi, ossalato	Local registry
Ring-14 disease, Salaam seizures, (no treatment)	Ring-14 syndrome, Chorea, (no treatment)	Perry syndrome, Epileptic attacks, lamotrigine	Perry syndrome, extreme mood swings, oxalate	English

# Translate to a language we can understand

Monika	Annika	Rajaram	Pietro	FAIR
Krankheit Ringbildung Chromosom 14, Salaam-Anfälle, (Keine Behandlung)	Ring-14-sjúkumynd, sankta Vitusar dansur, eingin viðgerð	பெர்ரி நோய்க்குறி, வலிப்பு தாக்குதல்கள், லாமோட்டரஜின்	sindrome Perry, sbalzi d'umore estremi, ossalato	Local registry
Ring-14 disease, Salaam seizures, (no treatment)	Ring-14 syndrome, Chorea, (no treatment)	Perry syndrome, Epileptic attacks, lamotrigine	Perry syndrome, extreme mood swings, oxalate	English

Who can see the solution with  
100% precision?

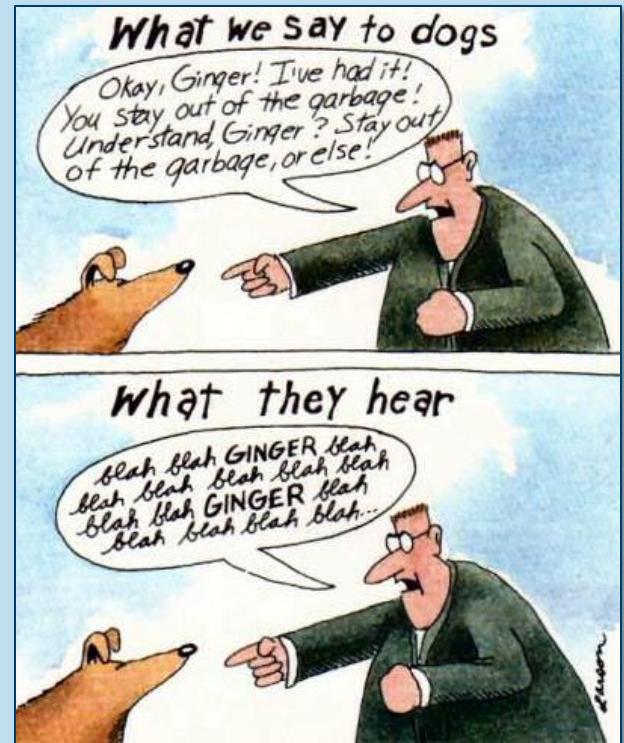
# Translate to a language we can understand

Monika	Annika	Rajaram	Pietro	FAIR
Krankheit Ringbildung Chromosom 14, Salaam-Anfälle, (Keine Behandlung)	Ring-14-sjúkumynd, sankta Vitusar dansur, eingin viðgerð	பெர்ரி நோய்க்குறி, வலிப்பு தாக்குதல்கள், லாமோட்டரஜின்	sindrome Perry, sbalzi d'umore estremi, ossalato	Local registry
Ring-14 disease, Salaam seizures, (no treatment)	Ring-14 syndrome, Chorea, (no treatment)	Perry syndrome, Epileptic attacks, lamotrigine	Perry syndrome, extreme mood swings, oxalate	English

Observation: human language is too ambiguous to be clear on what is what

Computers cannot solve our puzzle with this

# Conveying meaning to computers



## Steps



*"Now! That should clear up  
a few things around here!"*

1. Conceptual modelling: agree on what the data mean in human terms, terms for what the values denote, terms for the relations between them denote
2. Substitute ambiguous human terms with persistent, globally unique terms that computers and humans can understand
3. Solve the FAIR game by sending a query in terms of the computer understandable language

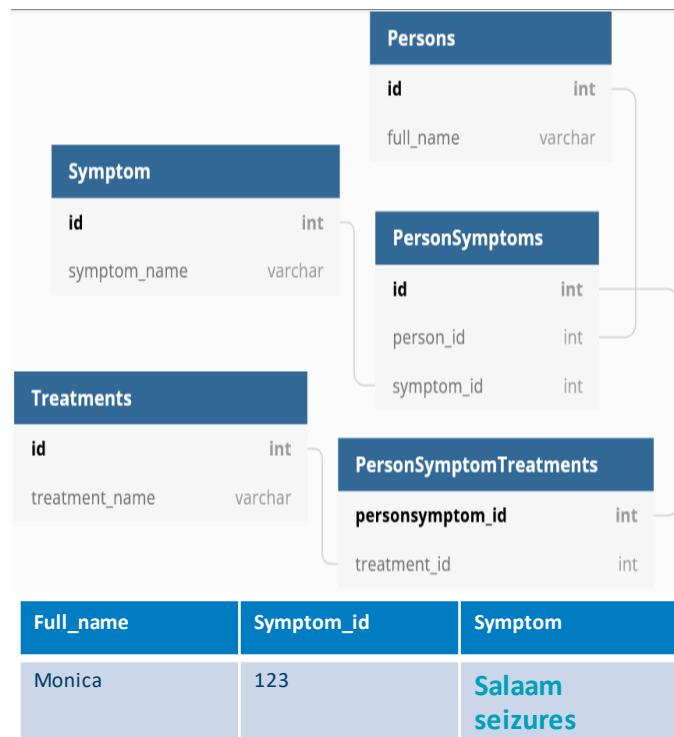
*...in a language that computers understand*

## Suggestions?

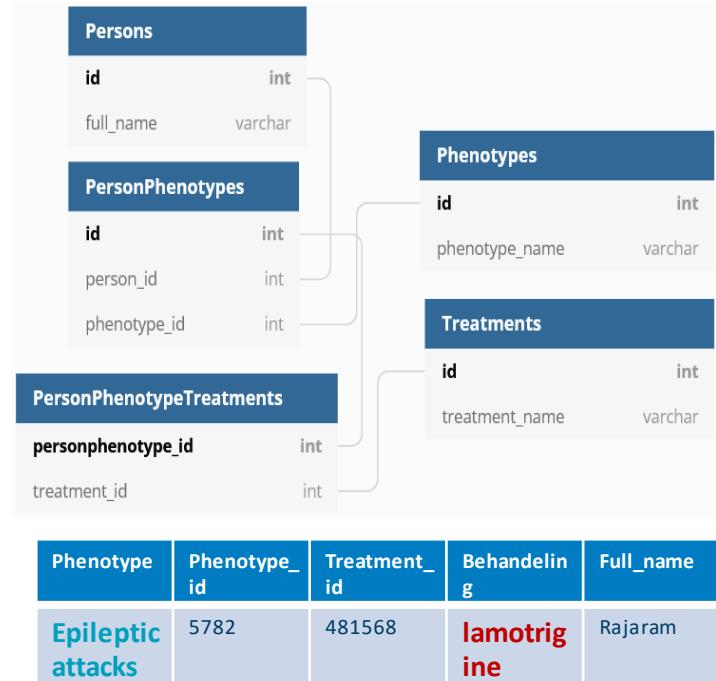
Students of essentials for data science say...

- ...

## Relational database 1 – by Annie in Australia

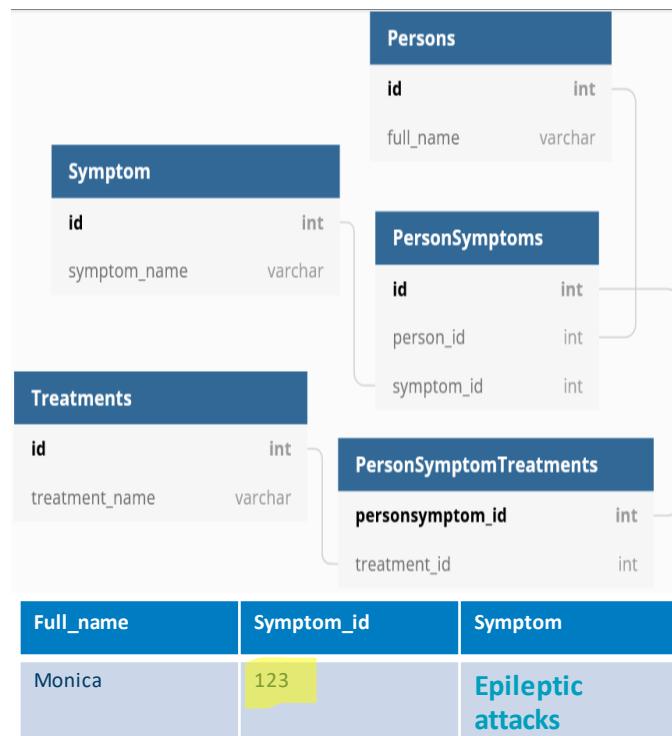


## Relational database 2 – by Bob in The Netherlands

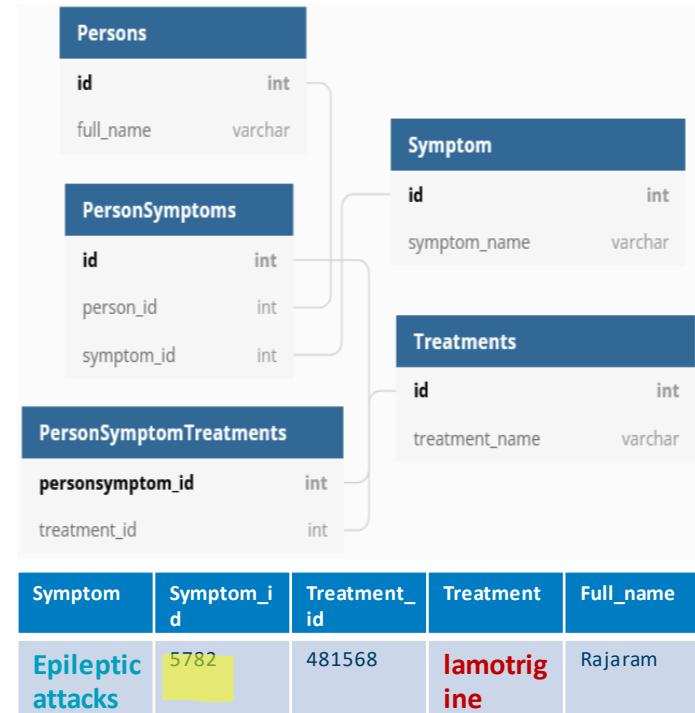


Can a computer  
know these are the  
same?

## Relational database 1 – by Annie in Australia



## Relational database 2 – by Bob in The Netherlands



Can a computer  
be sure these are  
the same?

## JSON schema?

### JSON schema 1 – by Annie in Australia

```
{  
  "$id": "https://example.com/person.schema.json",  
  "$schema": "https://json-schema.org/draft/2019-09/schema",  
  "title": "Person",  
  "type": "object",  
  "properties": {  
    "personName": {  
      "type": "string",  
      "description": "The person's name."  
    },  
    "symptom": {  
      "type": "string",  
      "description": "The person's symptom."  
    },  
    "treatment": {  
      "type": "string",  
      "description": "Treatment person is receiving"  
    }  
  }  
  
{  
  "personName": "Monica",  
  "symptom": "Salaam seizures",  
  "treatment": ""  
}
```

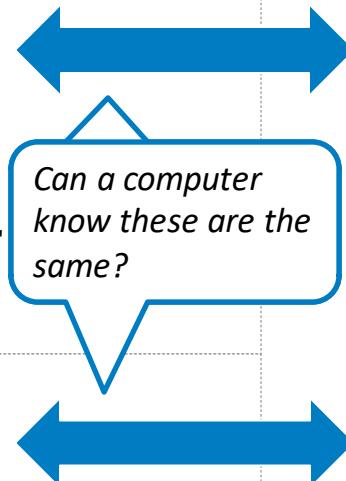
### JSON schema 2 – by Bob in The Netherlands

```
{  
  "$id": "https://example.com/persoon.schema.json",  
  "$schema": "https://json-schema.org/draft/2019-09/schema",  
  "title": "Person",  
  "type": "object",  
  "properties": {  
    "personName": {  
      "type": "string",  
      "description": "Naam van patient."  
    },  
    "phenotype": {  
      "type": "string",  
      "description": "Observeerbare afwijking."  
    },  
    "behandeling": {  
      "type": "string",  
      "description": "Gegeven behandeling"  
    }  
  }  
  
{  
  "personName": "Rajaram",  
  "phenotype": "Epileptic attacks",  
  "behandeling": "lamotrigine"  
}
```

# JSON schema?

## JSON schema 1 – by Annie in Australia

```
{  
  "$id": "https://example.com/person.schema.json",  
  "$schema": "https://json-schema.org/draft/2019-09/schema",  
  "title": "Person",  
  "type": "object",  
  "properties": {  
    "personName": {  
      "type": "string",  
      "description": "The person's name."  
    },  
    "symptom": {  
      "type": "string",  
      "description": "The person's symptom."  
    },  
    "treatment": {  
      "type": "string",  
      "description": "Treatment person is receiving"  
    }  
  }  
  
  {  
    "personName": "Monica",  
    "symptom": "Salaam seizures",  
    "treatment": ""  
  }
```



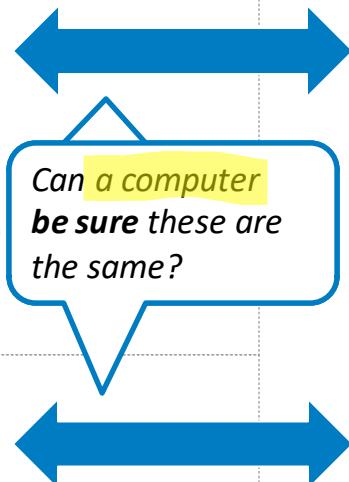
## JSON schema 2 – by Bob in The Netherlands

```
{  
  "$id": "https://example.com/persoon.schema.json",  
  "$schema": "https://json-schema.org/draft/2019-09/schema",  
  "title": "Person",  
  "type": "object",  
  "properties": {  
    "personName": {  
      "type": "string",  
      "description": "Naam van patient."  
    },  
    "phenotype": {  
      "type": "string",  
      "description": "Observeerbare afwijking."  
    },  
    "behandeling": {  
      "type": "string",  
      "description": "Gegeven behandeling"  
    }  
  }  
  
  {  
    "personName": "Rajaram",  
    "phenotype": "Epileptic attacks",  
    "behandeling": "lamotrigine"  
  }
```

## JSON schema?

### JSON schema 1 – by Annie in Australia

```
{  
  "$id": "https://example.com/person.schema.json",  
  "$schema": "https://json-schema.org/draft/2019-09/schema",  
  "title": "Person",  
  "type": "object",  
  "properties": {  
    "personName": {  
      "type": "string",  
      "description": "The person's name."  
    },  
    "symptom": {  
      "type": "string",  
      "description": "The person's symptom."  
    },  
    "treatment": {  
      "type": "string",  
      "description": "Treatment person is receiving"  
    }  
  }  
  
  {  
    "personName": "Monica",  
    "symptom": "Epileptic attacks",  
    "treatment": ""  
  }
```

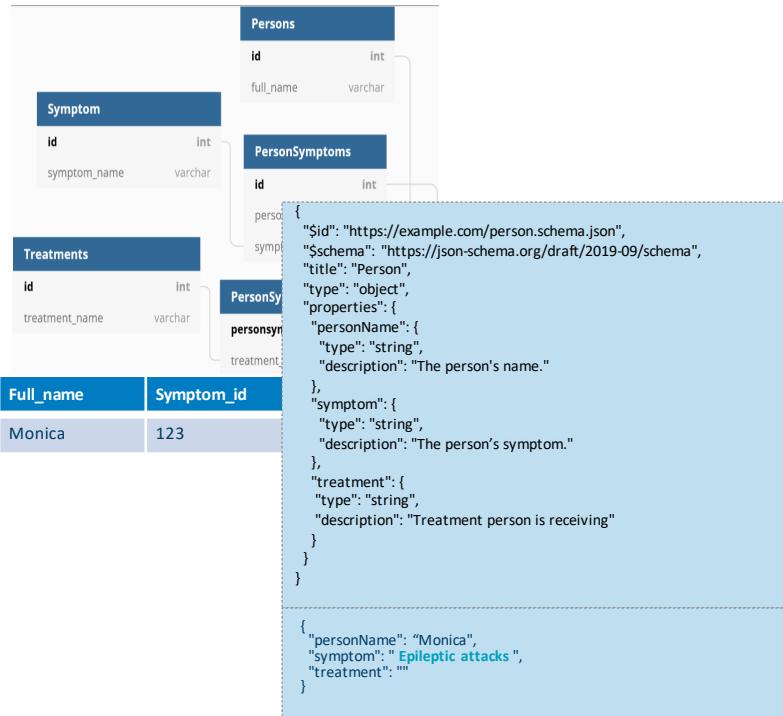


### JSON schema 2 – by Bob in The Netherlands

```
{  
  "$id": "https://example.com/persoon.schema.json",  
  "$schema": "https://json-schema.org/draft/2019-09/schema",  
  "title": "Person",  
  "type": "object",  
  "properties": {  
    "personName": {  
      "type": "string",  
      "description": "Naam van patient."  
    },  
    "symptom": {  
      "type": "string",  
      "description": "The person's symptoms."  
    },  
    "treatment": {  
      "type": "string",  
      "description": "Gegeven behandeling"  
    }  
  }  
  
  {  
    "personName": "Rajaram",  
    "symptom": "Epileptic attacks",  
    "treatment": "lamotrigine"  
  }
```

# What we say to computers programmers

## JSON and database schemas



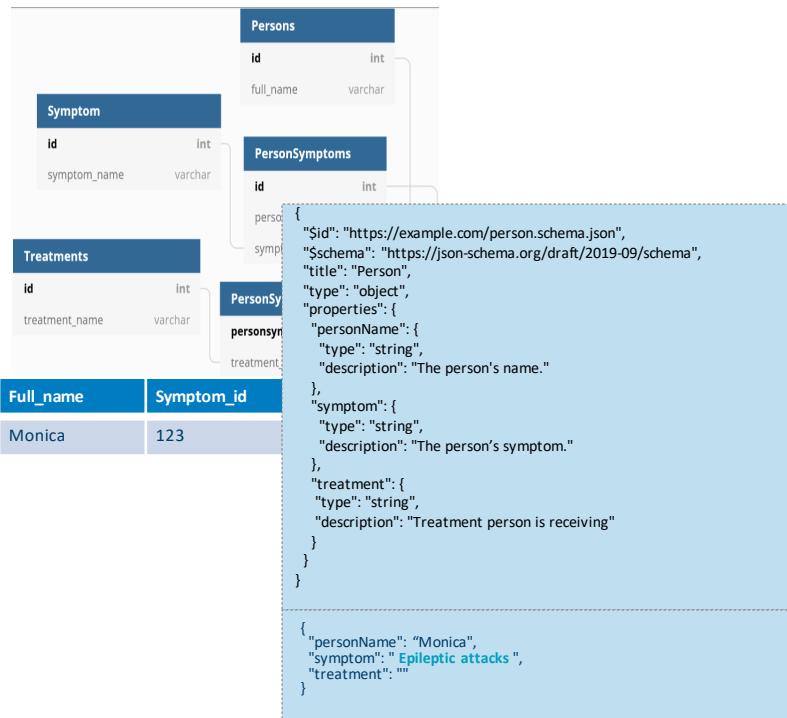
## Specifications

```
version: 2.0.0-draft.3
title: PersonSymptomTreatment
description: |
  Data of symptoms of persons and treatments they received.
type: object
required:
- id
- full_name
- symptom
- treatment
properties:
  id:
    type: string
    description: |
      Unique identifier of a person.
  full_name:
    type: string
    description: |
      Name of the person.
  symptom:
    type: string
    description: |
      symptom a person.
  treatment:
    type: string
    description: |
      treatment a person received.
```



# What the computer understands

## JSON and database schemas



## Specifications

version: 2.0.0-draft.3  
title: blah  
description: |  
blah blah blah blah.  
type: object  
required:

- blah
- bluh
- bleh
- bloh

properties:  
blah:  
  type: string  
  description: |  
    blah blah blah blah.  
bluh:  
  blah: blah  
  blah: |  
    blah blah blah blah.  
bleh:  
  type: string  
  description: |  
    blah blah blah.  
bloh:  
  type: string  
  description: |  
    blah blah blah blah.



## Are JSON and relational schemas computer understandable? – Not by default

- Data structures and APIs are useful for programming, programmers are humans, not machines
- They do not automatically make data unambiguously machine understandable
- Relations in Relational Database Schemas are semantically poor
  - Cardinality only
  - Primary – foreign key tables do not convey to computers what data values and relations between them denote
  - “Meaning” in models implicit, not enforced

**Making data and data  
schemas unambiguously  
actionable for computers  
using web technologies  
= The Semantic Web**



## Steps



"Now! *That* should clear up  
a few things around here!"

1. Conceptual modelling: agree on what the data mean in human terms, terms for what the values denote, terms for the relations between them (subject – predicate=relation – object)
2. Substitute ambiguous human terms with computer understandable, globally unique persistent identifiers from community-crafted semantic models (ontologies)
3. Use human terms as labels
4. Solve the FAIR game by sending a query in terms of the ontological model, use the labels for the user interface

*...in a language that computers understand*

## Translate to a language we can understand

Monika	Annika	Rajaram	Pietro	FAIR
Krankheit Ringbildung Chromosom 14, Salaam-Anfälle, (Keine Behandlung)	Ring-14-sjúkumynd, sankta Vitusar dansur, eingin viðgerð	பெர்ரி நோய்க்குறி, வலிப்பு தாக்குதல்கள், லாமோட்டரேஜின்	sindrome Perry, sbalzi d'umore estremi, ossalato	Local registry
Ring-14 disease, Salaam seizures, (no treatment)	Ring-14 syndrome, Chorea, (no treatment)	Perry syndrome, Epileptic attacks, lamotrigine	Perry syndrome, extreme mood swings, oxalate	English

Observation: human language is too ambiguous to be clear on what is what

Computers cannot solve our puzzle with this

## Conceptual modelling



# Core web component = the URI

Uniform Resource Identifier (URI)  
**100% Globally Unique!**

`http://rdf.biosemantics.org/owl/BioSemanticsConcepts#c3877...`



The diagram illustrates the structure of a Uniform Resource Identifier (URI). It consists of three main parts: 'Protocol' (blue), 'Address' (green), and 'data item' (red). The 'Protocol' part is 'http://'. The 'Address' part is 'rdf.biosemantics.org/owl/BioSemanticsConcepts#'. The 'data item' part is 'c3877...'. Brackets with arrows point from each label to its corresponding segment in the URI string.

Protocol  
for  
exchange  
by  
computers

“Address”

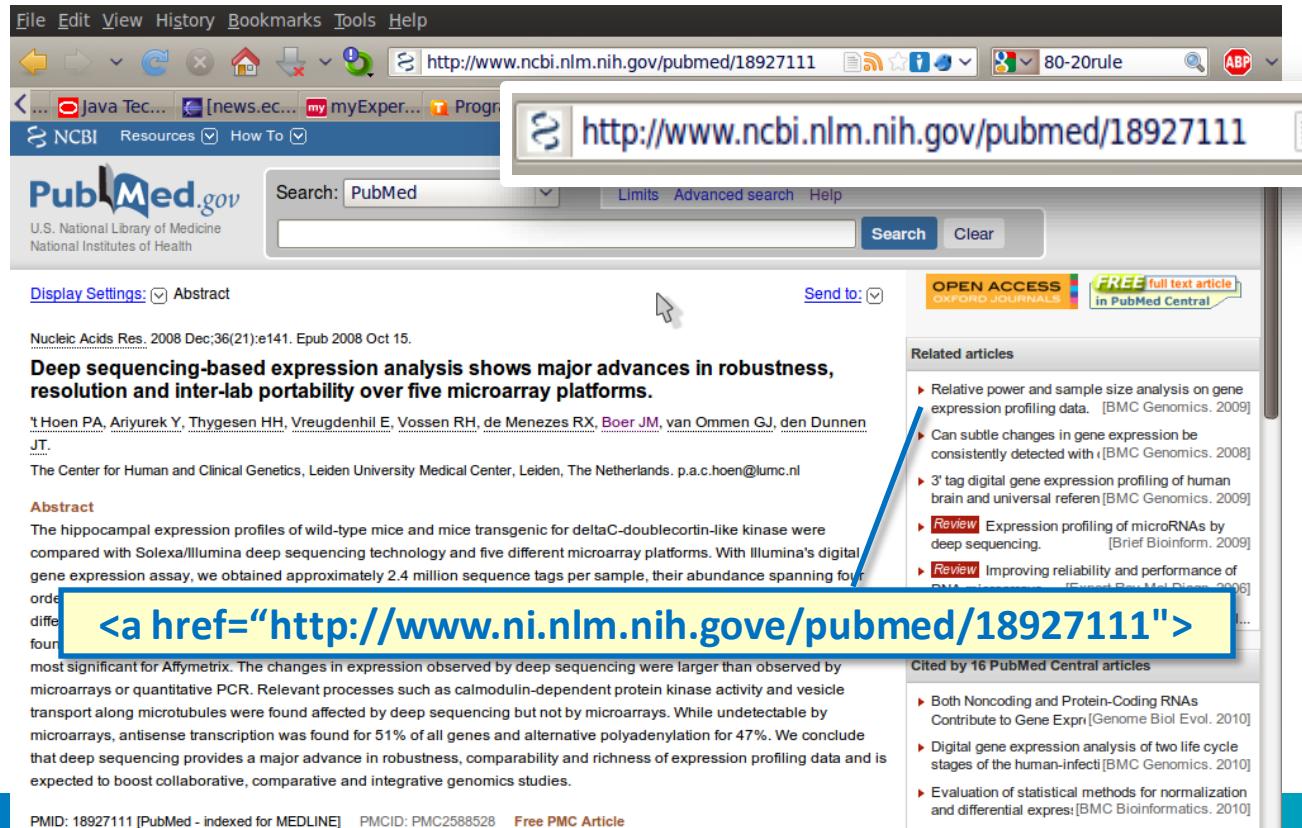
data item

# Uniform Resource Identifier

<http://rdf.biosemantics.org/owl/BioSemanticsConcepts#c3877...>

Unique **identifier** for a datum or concept  
Unique **reference** for a datum or concept  
**Computer readable**

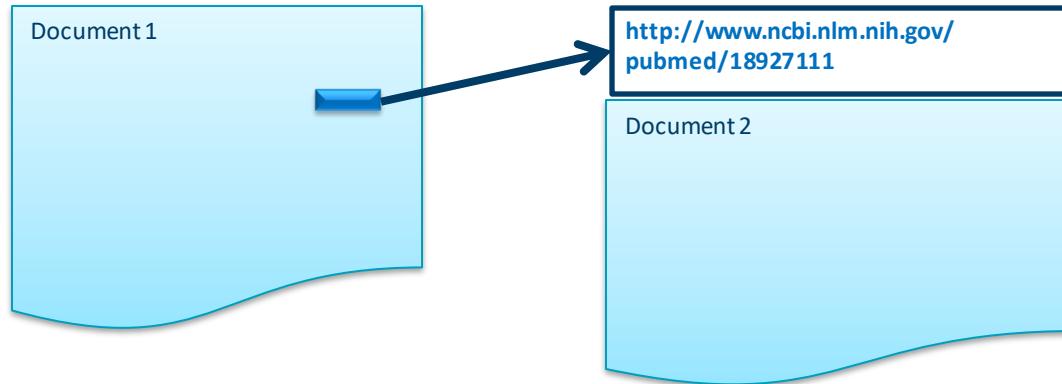
# Reuse of web technology: hyperlink



The screenshot shows a web browser window with the following details:

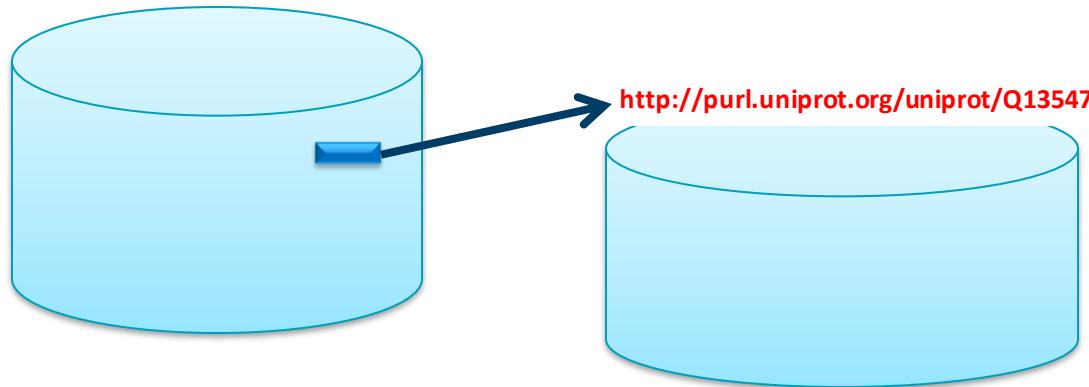
- Address Bar:** http://www.ncbi.nlm.nih.gov/pubmed/18927111
- Page Title:** PubMed
- Search Bar:** Search: PubMed
- Content Area:**
  - Article Summary:** Nucleic Acids Res. 2008 Dec;36(21):e141. Epub 2008 Oct 15.  
**Title:** Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms.  
**Authors:** Hoen PA, Ariyurek Y, Thygesen HH, Vreugdenhil E, Vossen RH, de Menezes RX, Boer JM, van Ommen GJ, den Dunnen JT.  
**Institution:** The Center for Human and Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands. p.a.c.hoen@lumc.nl  
**Abstract:** The hippocampal expression profiles of wild-type mice and mice transgenic for deltaC-doublecortin-like kinase were compared with Solexa/Illumina deep sequencing technology and five different microarray platforms. With Illumina's digital gene expression assay, we obtained approximately 2.4 million sequence tags per sample, their abundance spanning four orders of magnitude. The results were comparable to those from microarrays, but with significantly higher resolution and lower detection limits. The changes in expression observed by deep sequencing were larger than observed by microarrays or quantitative PCR. Relevant processes such as calmodulin-dependent protein kinase activity and vesicle transport along microtubules were found affected by deep sequencing but not by microarrays. While undetectable by microarrays, antisense transcription was found for 51% of all genes and alternative polyadenylation for 47%. We conclude that deep sequencing provides a major advance in robustness, comparability and richness of expression profiling data and is expected to boost collaborative, comparative and integrative genomics studies.
  - Hyperlink:** <a href="http://www.ncbi.nlm.nih.gov/pubmed/18927111">
  - Related Articles:** A sidebar listing several related articles with titles like "Relative power and sample size analysis on gene expression profiling data" and "Can subtle changes in gene expression be consistently detected with".
  - Cited By:** Cited by 16 PubMed Central articles.

# Hyperlinks (URIs) link documents



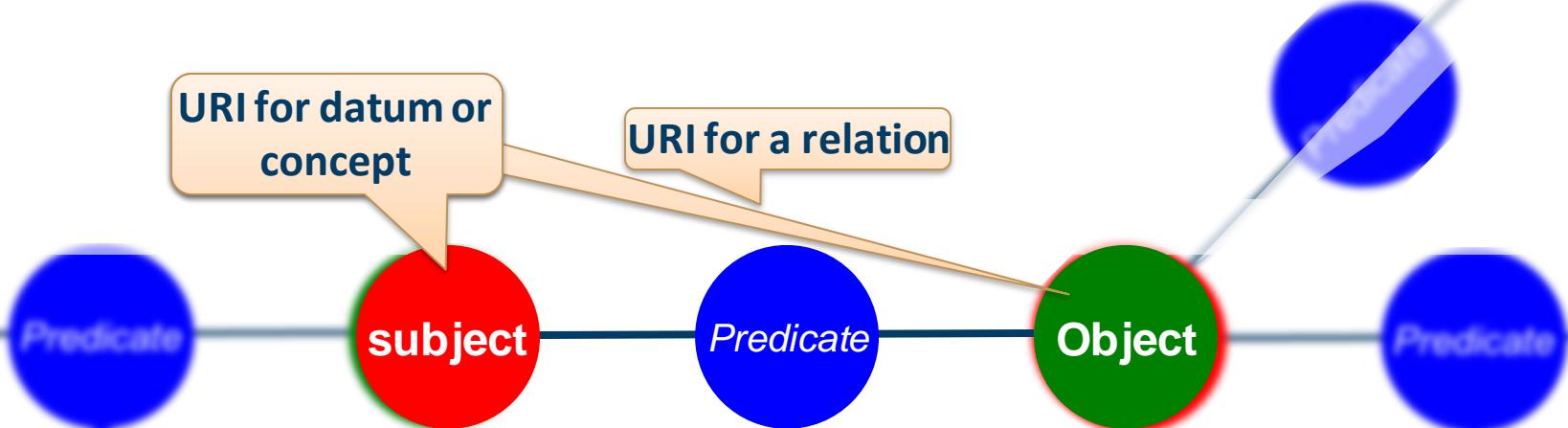
**Documents for human consumption**

# Hyperlinks (URIs) can link data



**Data referring to other data =  
Linked Data**

# Saying things with URIs in a language for computers



<URI for 'HDAC1'>	<URI for 'interacts with'>	<URI for 'ParvB'>
<URI for 'malaria'>	<URI for is 'transmitted by'>	<URI for 'mosquitos'>
<URI for 'mutation X'>	<URI for 'has frequency'>	<'0.25%'>

**Formal definition for humans & computers: Resource Description Framework (RDF)**

<https://www.w3.org/RDF/> (NB RDF is a data *model*, not a data *format*)

# Machine actionable knowledge

“HDAC1 *interacts with* Parvb”

<http://purl.uniprot.org/uniprot/Q13547>

[http://purl.obolibrary.org/obo/RO\\_0002434](http://purl.obolibrary.org/obo/RO_0002434)

<http://bio2rdf.org/geneid:29780>

# Linked data = computer readable knowledge

“HDAC1 *interacts with* Parvb”

<http://purl.uniprot.org/uniprot/Q13547>

[http://purl.obolibrary.org/obo/RO\\_0002434](http://purl.obolibrary.org/obo/RO_0002434)

<http://bio2rdf.org/geneid:29780>

Note: references point to  
different locations

# Linked data = computer readable knowledge

“HDAC1 *interacts with* Parvb”

<http://purl.uniprot.org/uniprot/Q13547>

[http://purl.obolibrary.org/obo/RO\\_0002434](http://purl.obolibrary.org/obo/RO_0002434)

<http://bio2rdf.org/geneid:29780>

Note: references point to  
different locations

Note: we actually said  
something meaningful!

# Linked data = computer readable knowledge

“HDAC1 *interacts with* Parvb”

<http://purl.uniprot.org/uniprot/Q13547>

[http://purl.obolibrary.org/obo/RO\\_0002434](http://purl.obolibrary.org/obo/RO_0002434)

<http://bio2rdf.org/geneid:29780>

Note: references point to  
different locations

Note: we actually said  
something meaningful!

Is that all we said?

```
- <rdf:RDF>
  - <owl:Ontology rdf:about="">
    <owl:imports rdf:resource="http://purl.uniprot.org/core/" />
  </owl:Ontology>
  - <rdf:Description rdf:about="http://purl.uniprot.org/uniprot/Q13547">
    <rdf:type rdf:resource="http://purl.uniprot.org/core/Protein"/>
    <reviewed rdf:datatype="http://www.w3.org/2001/XMLSchema#boolean">true</reviewed>
    <created rdf:datatype="http://www.w3.org/2001/XMLSchema#date">1997-11-01</created>
    <modified rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2012-05-16</modified>
    <version rdf:datatype="http://www.w3.org/2001/XMLSchema#int">151</version>
    <mnemonic>HDAC1_HUMAN</mnemonic>
    <oldMnemonic>HDA1_HUMAN</oldMnemonic>
    <oldMnemonic>HIDA_HUMAN</oldMnemonic>
    <replaces rdf:resource="http://purl.uniprot.org/uniprot/Q92534"/>
    <citation rdf:resource="http://purl.uniprot.org/citations/8602529" rdf:ID="_513133353437001"/>
    <citation rdf:resource="http://purl.uniprot.org/citations/8646880" rdf:ID="_513133353437003"/>
    <citation rdf:resource="http://purl.uniprot.org/citations/15489334" rdf:ID="_513133353437005"/>
    <citation rdf:resource="http://purl.uniprot.org/citations/10487760" rdf:ID="_513133353437007"/>
    <citation rdf:resource="http://purl.uniprot.org/citations/10898795" rdf:ID="_513133353437008"/>
    <citation rdf:resource="http://purl.uniprot.org/citations/10846170" rdf:ID="_513133353437009"/>
    <citation rdf:resource="http://purl.uniprot.org/citations/11006275" rdf:ID="_51313335343700A"/>
    <citation rdf:resource="http://purl.uniprot.org/citations/10669754" rdf:ID="_51313335343700B"/>
    <citation rdf:resource="http://purl.uniprot.org/citations/10655483" rdf:ID="_51313335343700C"/>
    <citation rdf:resource="http://purl.uniprot.org/citations/11602581" rdf:ID="_51313335343700D"/>
    <citation rdf:resource="http://purl.uniprot.org/citations/11331609" rdf:ID="_51313335343700E"/>
```

“HDAC1”

## Things we can say

<http://purl.uniprot.org/uniprot/Q13547>

We said all that ‘by proxy’  
by just using this reference

*URIs are references: UniProt data can be  
used ‘by reference’*

Using Semantic Web-Linked Data does not  
require downloading to your own site

# Things we can say: relation

<URI for the subject of the relation>  
<URI for a type of relation>  
<URI for the object of the relation>

<http://purl.uniprot.org/uniprot/Q13547>  
[http://purl.obolibrary.org/obo/RO\\_0002434](http://purl.obolibrary.org/obo/RO_0002434)  
<http://bio2rdf.org/geneid:29780>

# Things we can say: type information for machines

<http://purl.uniprot.org/uniprot/Q13547>

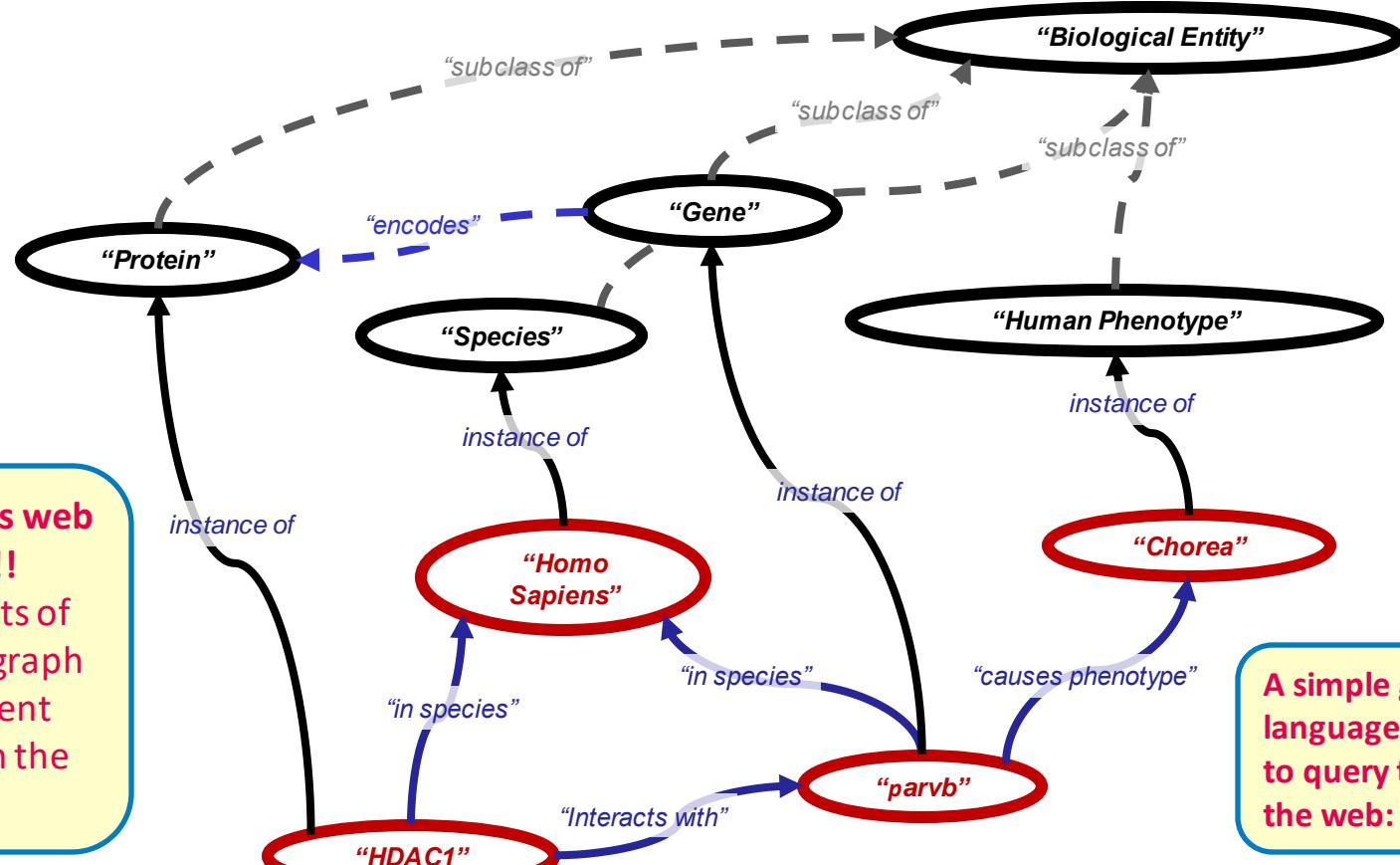
“HDAC1”

<URI for “is instance of type”>

<URI for class ‘Protein’>

# Knowledge graph

(Mock example, each entity has a URI)



# Back to the four registries

Solving the FAIR game with web technologies

1. Reach common understanding
2. Capture by URIs and ontologies  
→ virtual knowledge graph over 4 registries
3. Query the knowledge graph with SPARQL



Krankheit Ringbildung  
Chromosom 14, Salaam-  
Anfälle, (Keine  
Behandlung)



Ring-14-sjúkumynd,  
sankta Vitusar dansur,  
eingin viðgerð



பெர்ரி நோய்க்குறி,  
வளிப்பு  
தாக்குதல்கள்,  
லாமோட்டரஜின்

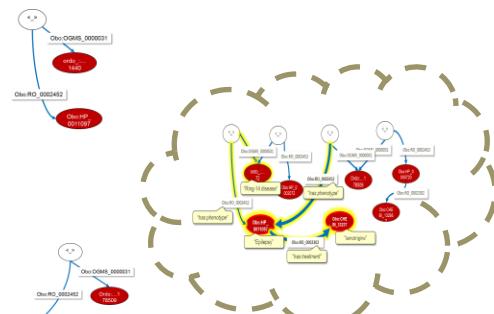


sindrome Perry, sbalzi  
d'umore estremi, ossalato

# FAIR data landscape

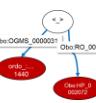
Monika

Krankheit Ringbildung Chromosom 14, Salaam-Anfälle, (Keine Behandlung)



Annika

Ring-14-sjúkumynd, sankta Vitusar dansur, eingin viðgerð



Rajaram

பெர்ரி நோய்க்குறி, வலிப்பு தாக்குதல்கள், ஸாமோட்டரஜின்



Pietro

sindrome Perry, sbalzi d'umore estremi, ossalato



# Reach common understanding: data types

Monika	Annika	Rajaram	Pietro	FAIR
Krankheit Ringbildung Chromosom 14, Salaam-Anfälle, (Keine Behandlung)	Ring-14-sjúkumynd, sankta Vitusar dansur, eingin viðgerð	ପେର୍ରି ନୋପକୁରୀ, ଲାକ୍ଷଣ୍ୟାବଳୀ, ଲାମୋଟାରେଜିଣ୍	sindrome Perry, sbalzi d'umore estremi, ossalato	Local registry
Ring-14 disease, Salaam seizures, (no treatment)	Ring-14 syndrome, Chorea, (no treatment)	Perry syndrome, Epileptic attacks, lamotrigine	Perry syndrome, extreme mood swings, oxalate	English

Semantics:  
Person, Disease, Phenotype, Treatment

# Reach common understanding: relations

Monika	Annika	Rajaram	Pietro	FAIR
ORPHA1440, HP:0011097	ORPHA72, HP:00027072	ORPHA178509, HP:0011097, CHEBI_6367	ORPHA178509, HP:0000720 CHEBI_132952	Coded
Monika <i>has disease</i> Ring-14 disease, and <i>has phenotype</i> Salaam seizures	Annika <i>has disease</i> Ring-14 disease, and <i>has phenotype</i> Chorea	Rajaram <i>has disease</i> Perry syndrome, and <i>has phenotype</i> Epileptic seizures. Epileptic seizures <i>are treated by</i> lamotrigine	Pietro <i>has disease</i> Ring-14 disease, and <i>has phenotype</i> Extreme mood swings. Extreme mood swings <i>are treated by</i> the drug Oxalate	full meaning

Semantics:

Person *has disease* Disease,  
Person *has phenotype* Phenotype,  
Phenotype *is treated by* Treatment

# Capture common understanding by URIs

Monika	Annika	Rajaram	Pietro	FAIR
Monika <b>has disease</b> Ring-14 disease, and <b>has phenotype</b> Salaam seizures	Annika <b>has disease</b> Ring-14 disease, and <b>has phenotype</b> Chorea	Rajaram <b>has disease</b> Perry syndrome, and <b>has phenotype</b> Epileptic seizures. Epileptic seizures <b>are treated by</b> lamotrigine	Pietro <b>has disease</b> Ring-14 disease, and <b>has phenotype</b> Extreme mood swings. Extreme mood swings <b>are treated by</b> the drug Oxalate	Full meaning
<_> <i>obo:OGMS_0000031</i> ordo:Orphanet_1440 <i>obo:RO_0002452</i> obo:HP_0011097.	<_> <i>obo:OGMS_0000031</i> ordo:Orphanet_1440, <i>obo:RO_0002452</i> obo:HP_0002072.	<_> <i>obo:OGMS_0000031</i> ordo:Orphanet_178509, <i>obo:RO_0002452</i> obo:HP_0011097 <i>obo:RO_0002302</i> obo:CHEBI_33237	<_> <i>obo:OGMS_0000031</i> ordo:Orphanet_178509, <i>obo:RO_0002452</i> obo:HP_0000720 <i>obo:RO_0002302</i> obo:CHEBI_132952	Interoperable & Machine readable

Machine readable semantics (RDF) for:  
Person **has disease** Disease,  
Person **has phenotype** Phenotype,  
Phenotype **is treated by** Treatment

## Ontologies

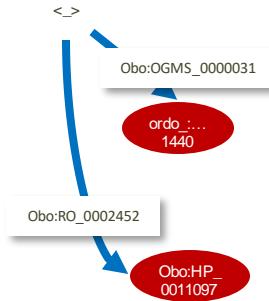
*obo:*<http://purl.obolibrary.org/obo/>

(general medical science, human phenotypes, relations, chemical compounds)

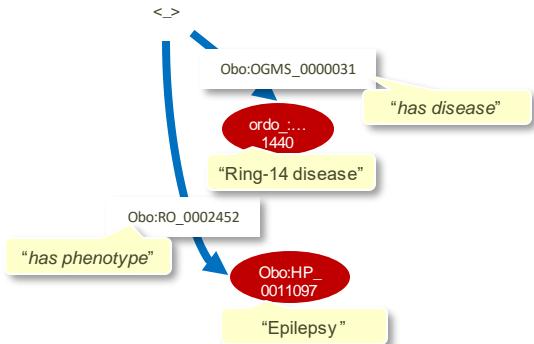
*ordo:*<http://www.orpha.net/ORDO/>

(rare diseases)

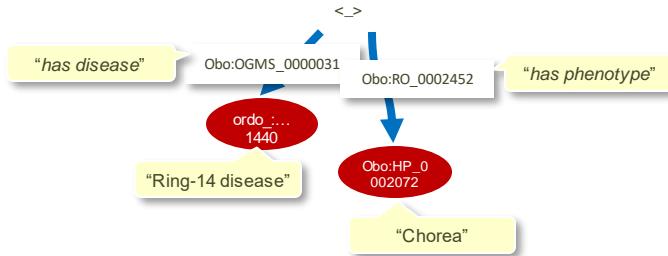
Monika	Annika	Rajaram	Pietro
<p>&lt;_&gt;</p> <p><i>obo:OGMS_0000031</i>  <i>ordo:Orphanet_1440</i>  <i>obo:RO_0002452</i>  <i>obo:HP_0011097.</i></p>	<p>&lt;_&gt;</p> <p><i>obo:OGMS_0000031</i>  <i>ordo:Orphanet_1440,</i>  <i>obo:RO_0002452</i>  <i>obo:HP_0002072.</i></p>	<p>&lt;_&gt;</p> <p><i>obo:OGMS_0000031</i>  <i>ordo:Orphanet_178509,</i>  <i>obo:RO_0002452</i>  <i>obo:HP_0011097</i>  <i>obo:RO_0002302</i>  <i>obo:CHEBI_33237</i></p>	<p>&lt;_&gt;</p> <p><i>obo:OGMS_0000031</i>  <i>ordo:Orphanet_178509,</i>  <i>obo:RO_0002452</i>  <i>obo:HP_0000720</i>  <i>obo:RO_0002302</i>  <i>obo:CHEBI_132952</i></p>



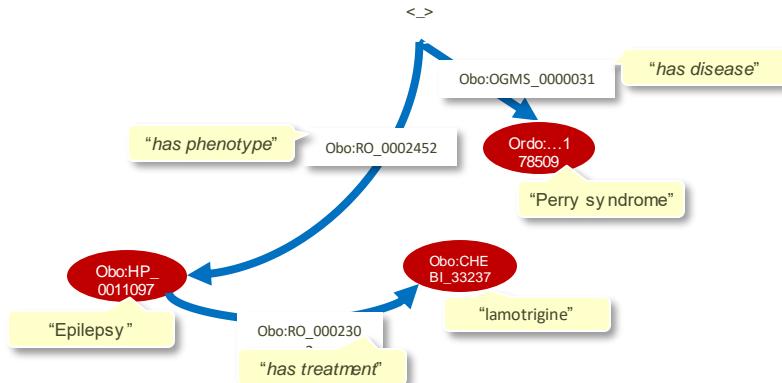
Monika	Annika	Rajaram	Pietro
$\langle \rangle$ obo:OGMS_0000031 ordo:Orphanet_1440 obo:RO_0002452 obo:HP_0011097.	$\langle \rangle$ obo:OGMS_0000031 ordo:Orphanet_1440, obo:RO_0002452 obo:HP_0002072.	$\langle \rangle$ obo:OGMS_0000031 ordo:Orphanet_178509, obo:RO_0002452 obo:HP_0011097 obo:RO_0002302 obo:CHEBI_33237	$\langle \rangle$ obo:OGMS_0000031 ordo:Orphanet_178509, obo:RO_0002452 obo:HP_0000720 obo:RO_0002302 obo:CHEBI_132952



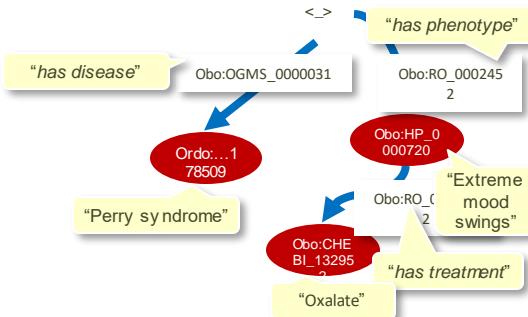
Monika	Annika	Rajaram	Pietro
$\langle \rangle$ obo:OGMS_0000031 ordo:Orphanet_1440 obo:RO_0002452 obo:HP_0011097.	$\langle \rangle$ obo:OGMS_0000031 ordo:Orphanet_1440, obo:RO_0002452 obo:HP_0002072.	$\langle \rangle$ obo:OGMS_0000031 ordo:Orphanet_178509, obo:RO_0002452 obo:HP_0011097 obo:RO_0002302 obo:CHEBI_33237	$\langle \rangle$ obo:OGMS_0000031 ordo:Orphanet_178509, obo:RO_0002452 obo:HP_0000720 obo:RO_0002302 obo:CHEBI_132952



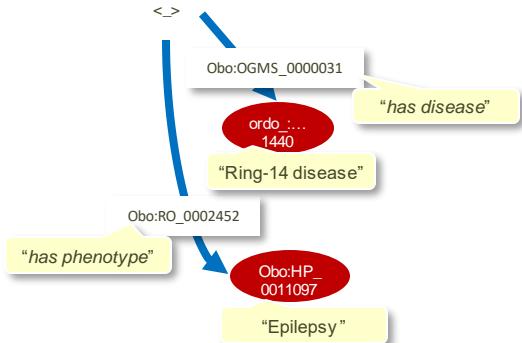
Monika	Annika	Rajaram	Pietro
$\langle \rangle$ obo:OGMS_0000031 ordo:Orphanet_1440 obo:RO_0002452 obo:HP_0011097.	$\langle \rangle$ obo:OGMS_0000031 ordo:Orphanet_1440, obo:RO_0002452 obo:HP_0002072.	$\langle \rangle$ obo:OGMS_0000031 ordo:Orphanet_178509, obo:RO_0002452 obo:HP_0011097 obo:RO_0002302 obo:CHEBI_33237	$\langle \rangle$ obo:OGMS_0000031 ordo:Orphanet_178509, obo:RO_0002452 obo:HP_0000720 obo:RO_0002302 obo:CHEBI_132952



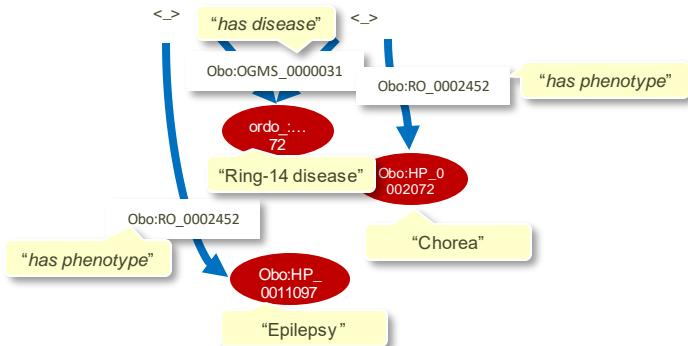
Monika	Annika	Rajaram	Pietro
<code>&lt;_&gt;</code> <code>obo:OGMS_0000031</code> <code>ordo:Orphanet_1440</code> <code>obo:RO_0002452</code> <code>obo:HP_0011097.</code>	<code>&lt;_&gt;</code> <code>obo:OGMS_0000031</code> <code>ordo:Orphanet_1440,</code> <code>obo:RO_0002452</code> <code>obo:HP_0002072.</code>	<code>&lt;_&gt;</code> <code>obo:OGMS_0000031</code> <code>ordo:Orphanet_178509,</code> <code>obo:RO_0002452</code> <code>obo:HP_0011097</code> <code>obo:RO_0002302</code> <code>obo:CHEBI_33237</code>	<code>&lt;_&gt;</code> <code>obo:OGMS_0000031</code> <code>ordo:Orphanet_178509,</code> <code>obo:RO_0002452</code> <code>obo:HP_0000720</code> <code>obo:RO_0002302</code> <code>obo:CHEBI_132952</code>



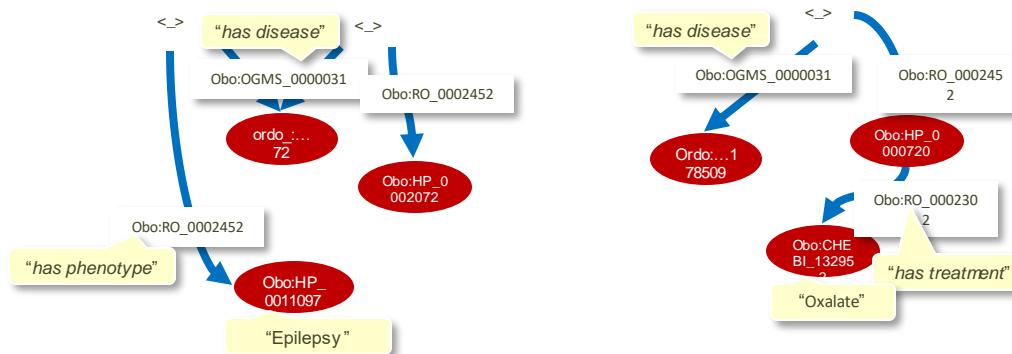
Monika	Annika	Rajaram	Pietro
$\langle \rangle$ obo:OGMS_0000031 ordo:Orphanet_1440 obo:RO_0002452 obo:HP_0011097.	$\langle \rangle$ obo:OGMS_0000031 ordo:Orphanet_1440, obo:RO_0002452 obo:HP_0002072.	$\langle \rangle$ obo:OGMS_0000031 ordo:Orphanet_178509, obo:RO_0002452 obo:HP_0011097 obo:RO_0002302 obo:CHEBI_33237	$\langle \rangle$ obo:OGMS_0000031 ordo:Orphanet_178509, obo:RO_0002452 obo:HP_0000720 obo:RO_0002302 obo:CHEBI_132952



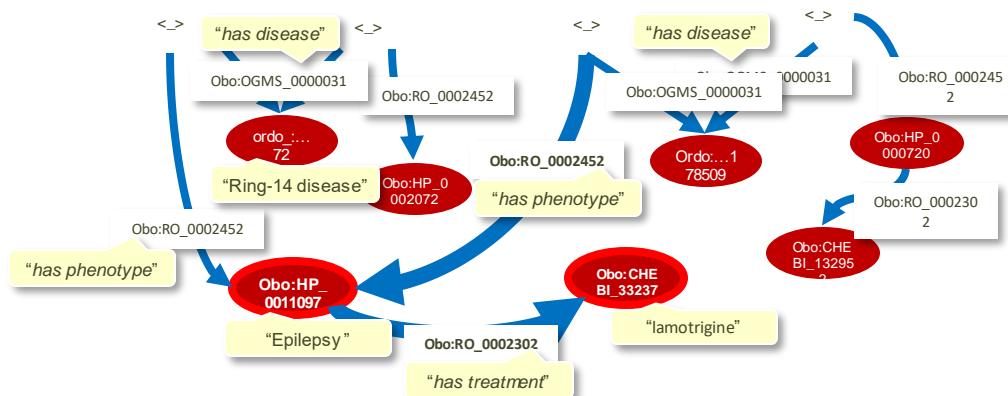
Monika	Annika	Rajaram	Pietro
$\langle \rangle$ obo:OGMS_0000031 ordo:Orphanet_1440 obo:RO_0002452 obo:HP_0011097.	$\langle \rangle$ obo:OGMS_0000031 ordo:Orphanet_1440, obo:RO_0002452 obo:HP_0002072.	$\langle \rangle$ obo:OGMS_0000031 ordo:Orphanet_178509, obo:RO_0002452 obo:HP_0011097 obo:RO_0002302 obo:CHEBI_33237	$\langle \rangle$ obo:OGMS_0000031 ordo:Orphanet_178509, obo:RO_0002452 obo:HP_0000720 obo:RO_0002302 obo:CHEBI_132952



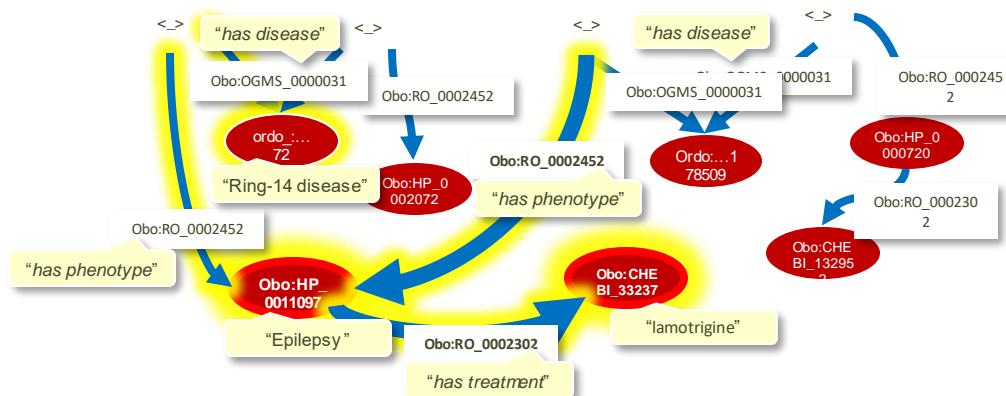
Monika	Annika	Rajaram	Pietro
<p>&lt;_&gt;</p> <p><i>obo:OGMS_0000031</i>  <i>ordo:Orphanet_1440</i>  <i>obo:RO_0002452</i>  <i>obo:HP_0011097.</i></p>	<p>&lt;_&gt;</p> <p><i>obo:OGMS_0000031</i>  <i>ordo:Orphanet_1440,</i>  <i>obo:RO_0002452</i>  <i>obo:HP_0002072.</i></p>	<p>&lt;_&gt;</p> <p><i>obo:OGMS_0000031</i>  <i>ordo:Orphanet_178509,</i>  <i>obo:RO_0002452</i>  <i>obo:HP_0011097</i>  <i>obo:RO_0002302</i>  <i>obo:CHEBI_33237</i></p>	<p>&lt;_&gt;</p> <p><i>obo:OGMS_0000031</i>  <i>ordo:Orphanet_178509,</i>  <i>obo:RO_0002452</i>  <i>obo:HP_0000720</i>  <i>obo:RO_0002302</i>  <i>obo:CHEBI_132952</i></p>



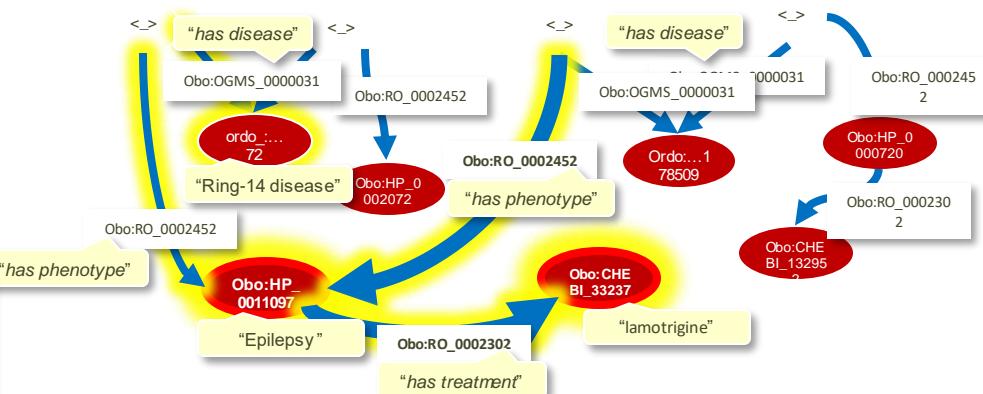
Monika	Annika	Rajaram	Pietro
<p>&lt;_&gt;</p> <p><i>obo:OGMS_0000031</i>  <i>ordo:Orphanet_1440</i>  <i>obo:RO_0002452</i>  <i>obo:HP_0011097</i>.</p>	<p>&lt;_&gt;</p> <p><i>obo:OGMS_0000031</i>  <i>ordo:Orphanet_1440</i>,  <i>obo:RO_0002452</i>  <i>obo:HP_0002072.</i></p>	<p>&lt;_&gt;</p> <p><i>obo:OGMS_0000031</i>  <i>ordo:Orphanet_178509,</i>  <i>obo:RO_0002452</i>  <i>obo:HP_0011097</i>  <i>obo:RO_0002302</i>  <i>obo:CHEBI_33237</i></p>	<p>&lt;_&gt;</p> <p><i>obo:OGMS_0000031</i>  <i>ordo:Orphanet_178509,</i>  <i>obo:RO_0002452</i>  <i>obo:HP_0000720</i>  <i>obo:RO_0002302</i>  <i>obo:CHEBI_132952</i></p>



Monika	Annika	Rajaram	Pietro
$\langle \rangle$ obo:OGMS_0000031 ordo:Orphanet_1440 obo:RO_0002452 obo:HP_0011097.	$\langle \rangle$ obo:OGMS_0000031 ordo:Orphanet_1440, obo:RO_0002452 obo:HP_0002072.	$\langle \rangle$ obo:OGMS_0000031 ordo:Orphanet_178509, obo:RO_0002452 obo:HP_0011097 obo:RO_0002302 obo:CHEBI_33237	$\langle \rangle$ obo:OGMS_0000031 ordo:Orphanet_178509, obo:RO_0002452 obo:HP_0000720 obo:RO_0002302 obo:CHEBI_132952



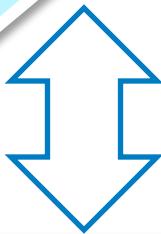
Monika	Annika	Rajaram	Pietro
$\langle \rangle$ obo:OGMS_0000031 ordo:Orphanet_1440 obo:RO_0002452 obo:HP_0011097.	$\langle \rangle$ obo:OGMS_0000031 ordo:Orphanet_1440, obo:RO_0002452 obo:HP_0002072.	$\langle \rangle$ obo:OGMS_0000031 ordo:Orphanet_178509, obo:RO_0002452 obo:HP_0011097 obo:RO_0002302 obo:CHEBI_33237	$\langle \rangle$ obo:OGMS_0000031 ordo:Orphanet_178509, obo:RO_0002452 obo:HP_0000720 obo:RO_0002302 obo:CHEBI_132952



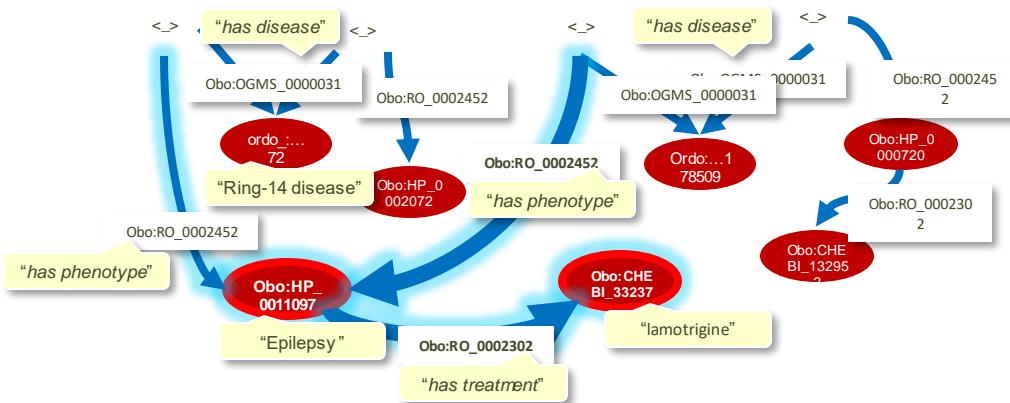
lamotrigine **is treating** the phenotype Epilepsy that **is a** phenotype for person Monika who **has disease** Ring-14 disease

How do we extract that answer from the graph?

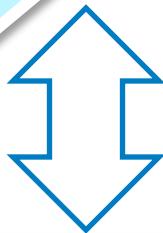
Which treatment **is**  
**treating** the phenotype  
that **is a** phenotype for  
Person Monica who **has**  
**disease** Ring-14 disease



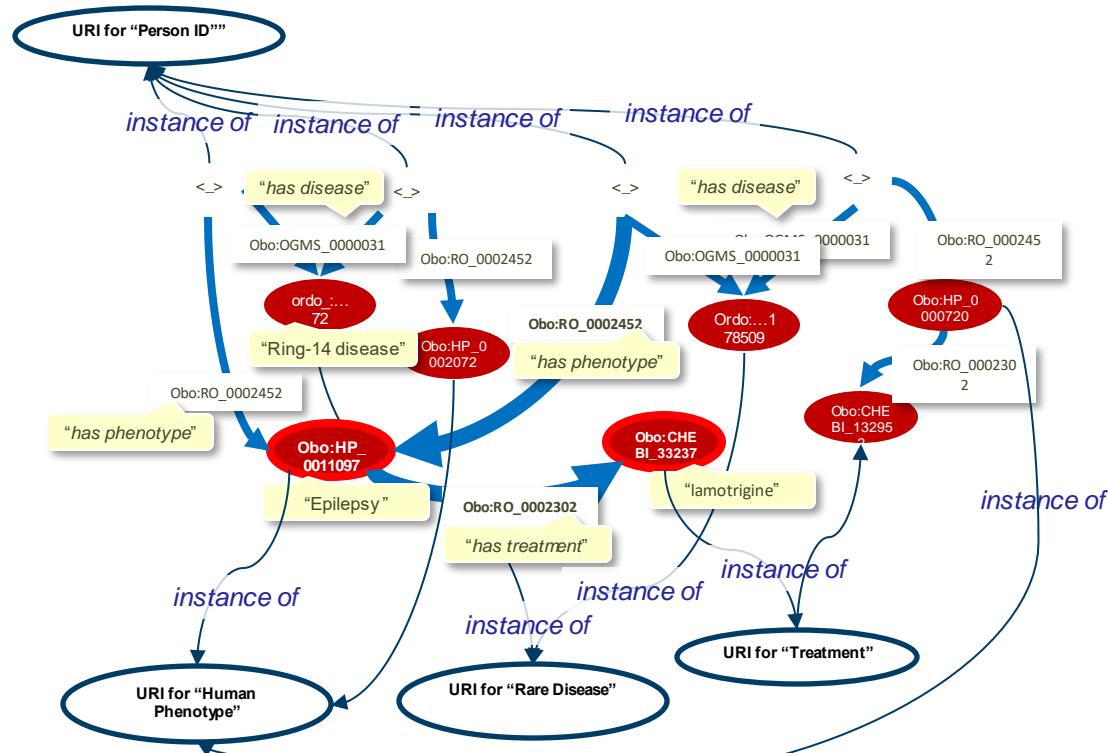
lamotrigine **is treating**  
the phenotype Epilepsy  
that **is a** phenotype for  
person Monica who **has**  
**disease** Ring-14 disease



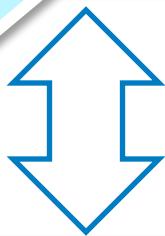
Which treatment **is**  
**treating** the phenotype  
that **is a** phenotype for  
Person Monica who **has**  
**disease** Ring-14 disease



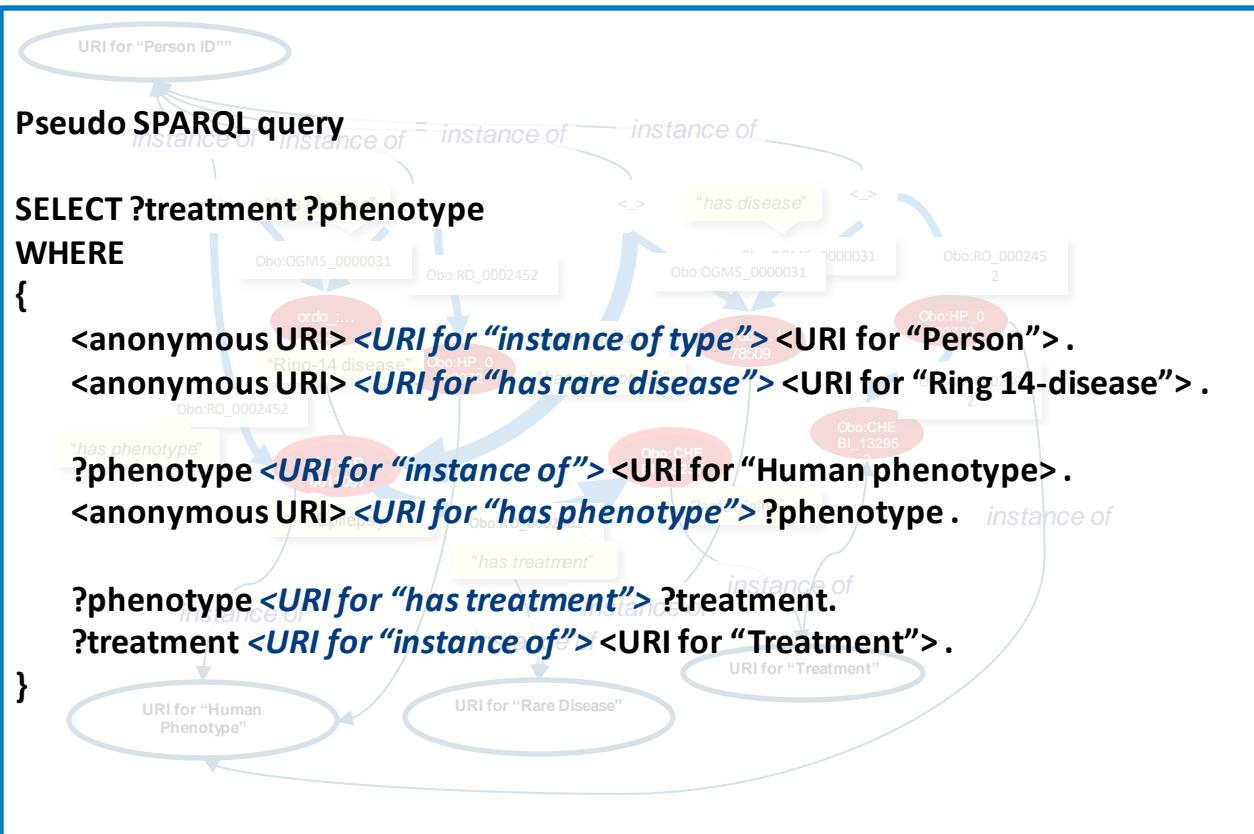
lamotrigine **is treating**  
the phenotype Epilepsy  
that **is a** phenotype for  
person Monica who **has**  
**disease** Ring-14 disease



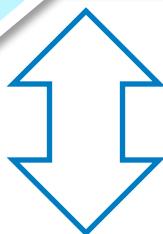
Which treatment **is treating** the phenotype  
that **is a** phenotype for a  
Person who **has disease**  
Ring-14 disease



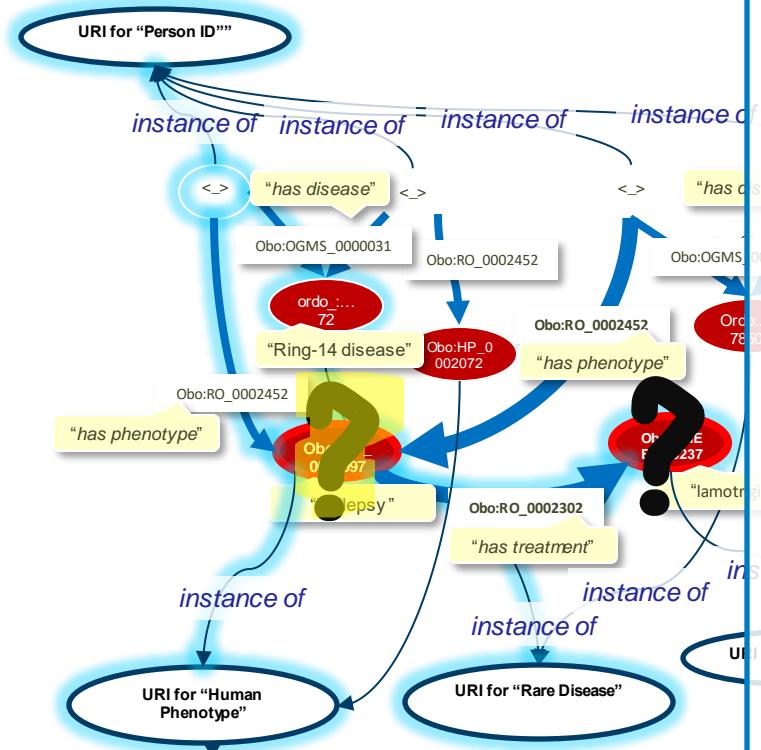
lamotrigine **is treating**  
the phenotype Epilepsy  
that **is a** phenotype for a  
person who **has disease**  
Ring-14 disease



Which treatment **is treating** the phenotype that **is a** phenotype for a Person who **has disease**  
Ring-14 disease



lamotrigine **is treating** the phenotype Epilepsy that **is a** phenotype for a person who **has disease**  
Ring-14 disease



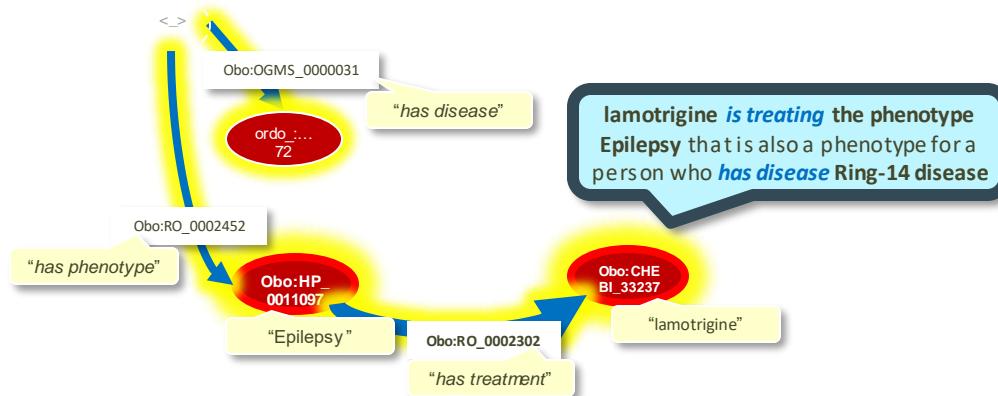
### Pseudo SPARQL query

```

SELECT ?treatment ?phenotype
WHERE
{
  <anonymous URI> <URI for "instance of type"> <URI for "Person"> .
  <anonymous URI> <URI for "has rare disease"> <URI for "Ring 14-disease"> .
  ?phenotype <URI for "instance of"> <URI for "Human phenotype"> .
  <anonymous URI> <URI for "has phenotype"> ?phenotype .
  ?phenotype <URI for "has treatment"> ?treatment .
  ?treatment <URI for "instance of"> <URI for "Treatment"> .
}
  
```

# Result

Disease   global machine readable code (URI)	Treatment   URI	Phenotype   URI
Ring-14 disease ordo:Orphanet_1440	Lamotrigine Obo:CHEBI_33237	Epilepsy Obo:HP_0011097



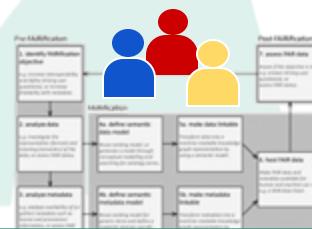


# **EUROPEAN JOINT PROGRAMME RARE DISEASES**



# Tooling by the EJP RD

Your data  
in original  
form



FAIRification

R&D of tools for  
FAIR data  
stewards

FAIR VP compliant resource  
with FAIR data

Web User  
Interface

Application  
Programming  
Interface(s)

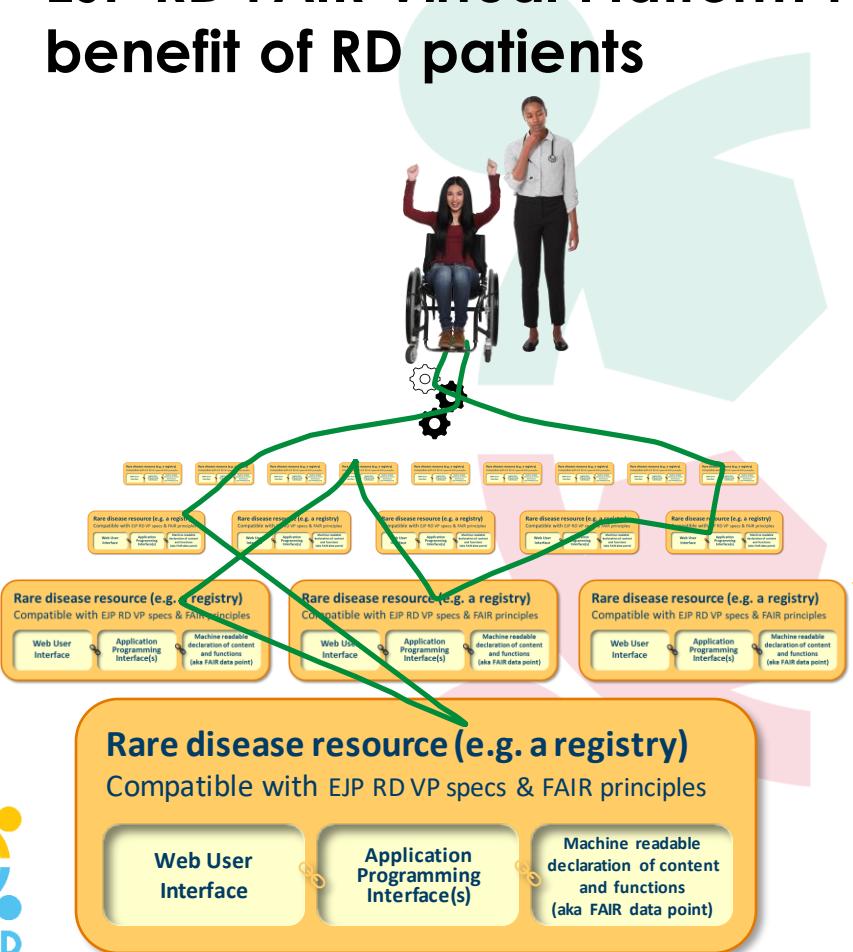
Machine readable  
declaration of content  
and functions  
(aka FAIR data point)

Your data  
in original  
form

Ontological  
Information about  
resource &  
access

Ontological  
model describing  
data elements

# EJP RD FAIR Virtual Platform to enhance your research for benefit of RD patients



- Rare disease patients would like you to contribute to the VP network
- Resources create the Virtual Platform network
- By what resources **declare** about their content and functions *in a language that computers understand* (= FAIR for machines)
- Analysis algorithms make use of FAIR, researchers make use of these algorithms

# for engineers... (a special type of human)

A resource discovery API (programmatic access)

Rare disease resource (e.g. a registry)

Compatible with EJP RD VP specs & FAIR principles

Web User Interface

Application Programming Interface(s)

Machine understandable declaration of content

The screenshot shows a GitHub repository page for 'query\_builder\_api'. The repository has 1 branch and 3 tags. The README.md file is updated by davidReinert. The repository focuses on the query builder API, which is described as a web service for rare disease data sharing. The API specification is provided in a specification.yaml file, which is linked from the README.md and versions/1.0/specification.yaml files. The specification includes details like openapi version, info, version, title, description, contact, license, and externalDocs. The code repository has 99 commits and 1 contributor.

Code

Issues 2

Pull requests

Actions

Projects 1

Wiki

Security

Insights

Notifications

Star 2

Fork 1

master 1 branch 3 tags

Code

About

This project focusses on the query builder API.

davidReinert Update README.md

52a9381 on 6 Jul 99 commits

Code Issues 2 Pull requests Actions Projects 1 Wiki Security Insights

versions Update specification.yaml

LICENSE Create LICENSE

README.md Update README.md

README.md

## Query Builder API

This project focusses on the query builder API. It will build up the catalogue APIs.

288 lines (288 sloc) 8.74 KB

```
1  openapi: "3.0.0"
2  info:
3    version: "1.0"
4    title: "Federated Resource Query API Specification"
5    description: '>
6      "Federated Resource Query API is a web service for rare disease data sharing
7      that can be queried for information about specific rare disease data."
8    contact:
9      name: "EJP-RD Pillar 2 Query Builder Work Focus"
10     license:
11       name: "Apache 2.0"
12       url: "http://www.apache.org/licenses/LICENSE-2.0.html"
13     externalDocs:
14       description: "EJP-RD Website"
```

Raw Blame

Go to file

Code

query\_builder\_api / versions / v1.0 / specification.yaml

davidReinert Update specification.yaml

Latest commit fcb9388 on 2 Jul History

At 1 contributor

Raw Blame

Code

Issues 2

Pull requests

Actions

Projects 1

Wiki

Security

Insights

Raw Blame

Code

Issues 2

Pull requests

Actions

Projects 1

Wiki

Security

Insights



# FAIR for machines

example: FAIR data point Duchenne patient platform  
(DCAT metadata provisioning service)

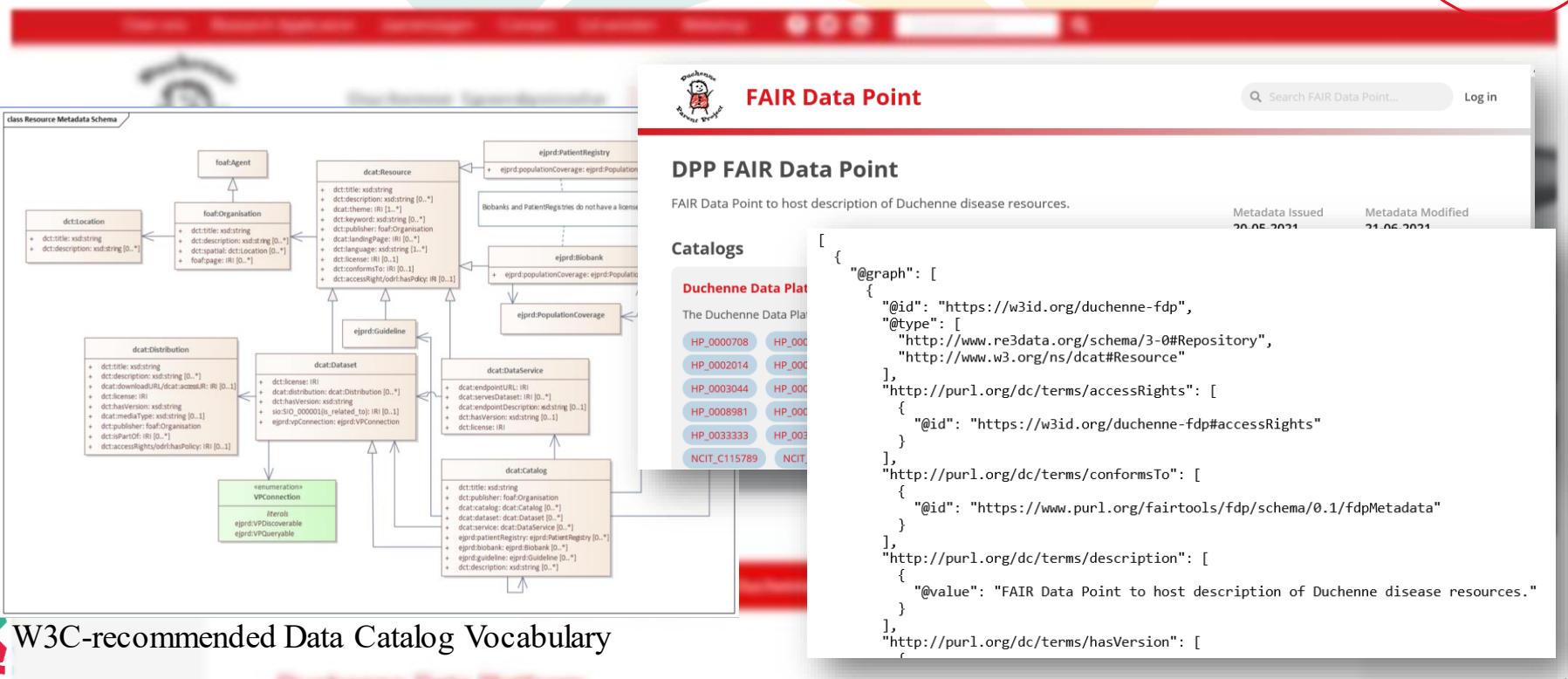
Rare disease resource (e.g. a registry)

Compatible with EJP RD VP specs & FAIR principles

Web User Interface

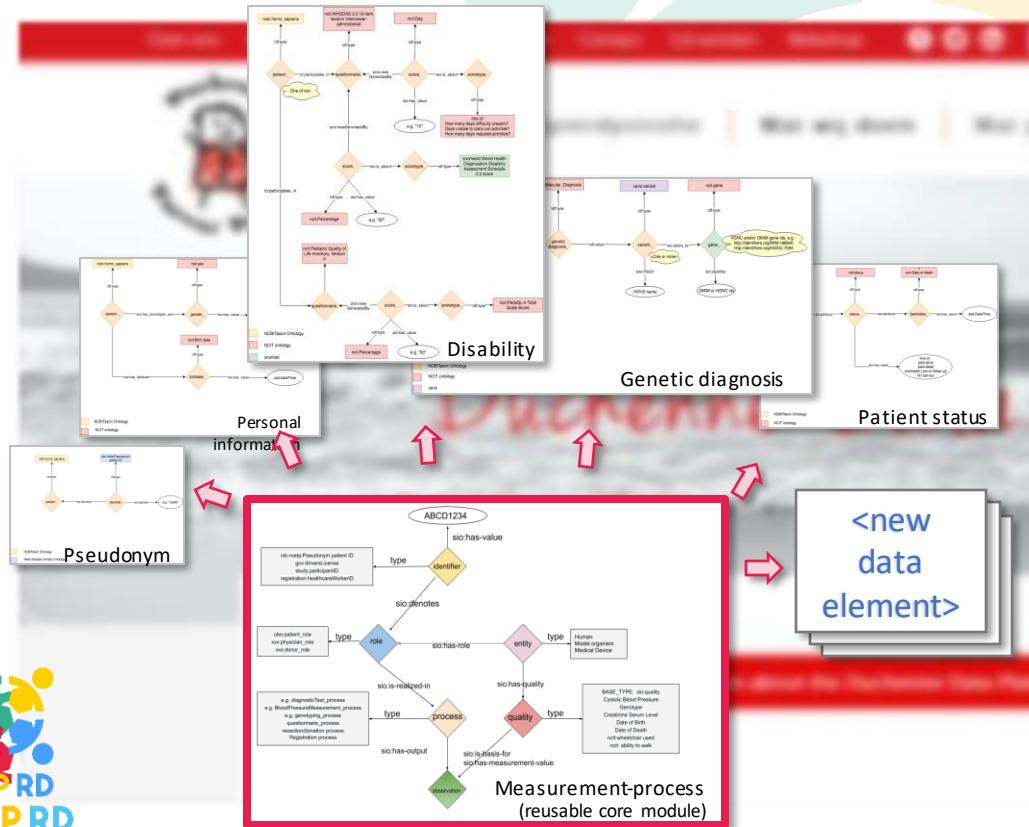
Application Programming Interface(s)

Machine understandable declaration of content



# FAIR for machines

Modelling pattern, aka “DCAT” for data elements



Rare disease resource (e.g. a registry)

Compatible with EJP RD VP specs & FAIR principles

Web User Interface

Application Programming Interface(s)

Machine understandable declaration of content

Semantic ‘measurement-process’ pattern for observational data

- A reusable, scalable, and queryable data model ‘for machines’ using standard ontologies, applied at source
- Worked out for 16 common data elements for patient registries
- Mappings and bridging solutions to other semantic frameworks & formats (e.g. FHIR, OMOP, C-DISC, OBO Foundry, GA4GH)
- Complements resource-level metadata model

# FAIR for machines

example: web interface Duchenne Data Platform

Rare disease resource (e.g. a registry)

Compatible with EJP RD VP specs & FAIR principles

Web User  
Interface

Application  
Programming  
Interface(s)

Machine  
understandable  
declaration of  
content

The screenshot shows the homepage of the Duchenne Data Platform. At the top, there is a red navigation bar with links: Over ons, Research Application, Jaarverslagen, Contact, Lid worden, Webshop, and a search bar. Below the navigation bar is a logo for 'Duchenne Parent Project' featuring a cartoon character. To the right of the logo are menu items: Duchenne Spierdystrofie, Wat wij doen, Wat jij kunt doen, and Nieuws. The main background image is a black and white photograph of a beach with people in the water. Overlaid on the image is the text 'Duchenne Data Platform' in large, red, hand-drawn style letters. At the bottom of the page, there is a red button with the text 'Information about the Duchenne Data Platform in English'.



Funded by the  
European  
Union  
GA n°825575

# FAIR Duchenne (Patient) Platform



## Duchenne Parent Project

FAIR Metadata



Search FAIR Data Point...

Log in

Advanced

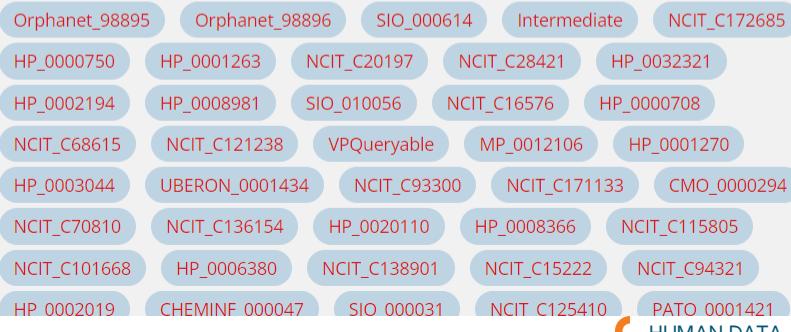
## Duchenne Parent Project: FAIR Data

Clinical and Patient Reported data for Duchenne and Becker muscular dystrophy patients. Conceptual metadata only. For data access, speak to the publisher of the dataset.

### Catalogs

#### Duchenne Data Platform Catalog

The Duchenne Data Platform (DDP) Catalog



Conforms to

- [FAIR Data Point Profile](#)

Metadata modified

13-04-2024

Metadata issued

26-01-2024

Metadata identifier  
[identifier](#)

Logo

[dpp-logo.png](#)

Gdpr personal information?

true

HUMAN DATA



# For machines

## Metadata

### Keywords

- [Becker Muscular Dystrophy](#)
- [Duchenne Muscular Dystrophy](#)

### Version

1.0

### Endpoint url

<https://w3id.org/duchenne-fdp>

### Ontological description

- [Orphanet 98895](#)
- [Orphanet 98896](#)

### RDF metadata formats

[ttl](#) [rdf+xml](#) [json-ld](#)

# FAIR Data Point for machines!

```
[ {  
  "@graph" : [ {  
    "@id" : "_:b336853dce19490ca18f1253cf9b84493781",  
    "@type" : [ "http://www.w3.org/2006/vcard/ns#Agent" ],  
    "http://www.w3.org/2006/vcard/ns#url" : [ {  
      "@id" : "https://www.worldduchenne.org/"  
    } ]  
  }, {  
    "@id" : "https://w3id.org/duchenne-fdp",  
    "@type" : [ "http://www.w3.org/ns/dcat#Resource", "http://www.w3.org/ns/dcat#DataService", "https://w3id.org/fdp/fdp-o#MetadataService", "https://w3id.org/fdp/fdp-o#FAIRDataPoint" ],  
    "http://purl.org/dc/terms/accessRights" : [ {  
      "@id" : "https://w3id.org/duchenne-fdp#accessRights"  
    } ],  
    "http://purl.org/dc/terms/conformsTo" : [ {  
      "@id" : "https://w3id.org/duchenne-fdp/profile/77aaad6a-0136-4c6e-88b9-07ffcc0ee4c"  
    } ],  
    "http://purl.org/dc/terms/description" : [ {  
      "@value" : "Clinical and Patient Reported data for Duchenne and Becker muscular dystrophy patients. Conceptual metadata only. For data access, speak to the publisher of the dataset."  
    } ],  
    "http://purl.org/dc/terms-hasVersion" : [ {  
      "@type" : "http://www.w3.org/2001/XMLSchema#float",  
      "@value" : "1.0"  
    } ],  
    "http://purl.org/dc/terms/keyword" : [ {  
      "@value" : "Becker Muscular Dystrophy"  
    }, {  
      "@value" : "Duchenne Muscular Dystrophy"  
    } ],  
    "http://purl.org/dc/terms/language" : [ {  
      "@id" : "http://id.loc.gov/vocabulary/iso639-1/en"  
    } ],  
    "http://purl.org/dc/terms/license" : [ {  
      "@id" : "http://rdflicense.appspot.com/rdflicense/cc-by-nc-nd3.0"  
    } ],  
    "http://purl.org/dc/terms/publisher" : [ {  
      "@id" : "https://w3id.org/duchenne-fdp#publisher"  
    } ],  
    "http://purl.org/dc/terms/title" : [ {  
      "@value" : "Duchenne Parent Project: FAIR Data"  
    } ],  
    "http://semanticscience.org/resource/SIO_000628" : [ {  
      "@id" : "https://w3id.org/duchenne-fdp/metrics/445c0a70d1e214e545b261559e2842f4"  
    }, {  
      "@id" : "https://w3id.org/duchenne-fdp/metrics/5d27e854a9e78eb3f663331cd47cdc13"  
    } ]  
  } ]
```

Metadata conform to the  
Data Catalog Vocabulary

World wide web recommendation  
Machine actionable

# Machine action on FAIR data sources

'Time to diagnosis' example courtesy of Mark Wilkinson (UPM, FAIR Data Systems) and Nawel Lalout (RUMC, Duchenne Parent Project)

## Call the registry data service interface

We need call the URL of the "time-to-diagnosis" service for each registry. We will then do a bit of processing if you want to join this demo yourself, add the abbreviation of your registry, and the URL to your data done for you automatically!

```
registries = [
    "OPP" => "https://www.fairdata.services/proxy/shallot/api-local/kpi-tts",
    "ENVO" => "https://rki-docker.uk3.uni-freiburg.de/gic-euronord/api-local/kpi-tts",
    # 'YOUR_REGISTRY_HERE' => "http://your.registry/service/kpi-tts"
]

disease_hash = Hash.new

registries.each do |registry, url|
  begin
    csv = RestClient.get(url)
    rescue
    next
  end

  puts "# Data Structure is:
# {CSV::Header} diagnosis,year,offset
# http://www.orpha.net/ORDO/Orphanet_08896_2006_334
```

## Use FAIR to get more data

Gather some data about the diseases, for example, the disease name.

```
puts "Getting FAIR data about Orphanet diseases"

# do an intersection over all registries
diseasenames = Hash.new
registries.each do |registry, url|
  next unless disease_hash[registry]
  years = disease_hash[registry].keys.each do |year|
    years <-> disease_hash[registry].keys.each do |year|
      disease_hash[registry][year].each do |disease, offset|
        disease_hash[registry][year].each do |disease, offset|
          disease_hash[registry][year][disease] = disease
        end
      end
    end
  end
end

orphanet_call = "https://www.orpha.net/espina/default/graph-api?query=PREFIX Orphanet: > SELECT ?label WHERE ?label = ?"
getts = HashClient.get(orphanet_call)
label = CSV::Parser.parse(getts['body']).first['label']
diseasenames[disease] = label
end
```

Setting FAIR data about Orphanet diseases  
URL: https://www.orpha.net/ORDO/Orphanet\_08896 Name: "Duchenne muscular dystrophy"  
URL: https://www.orpha.net/ORDO/Orphanet\_08895 Name: "Becker muscular dystrophy"

## Analytics

Here is a simple plot of the time-to-diagnosis for a specific disease over time

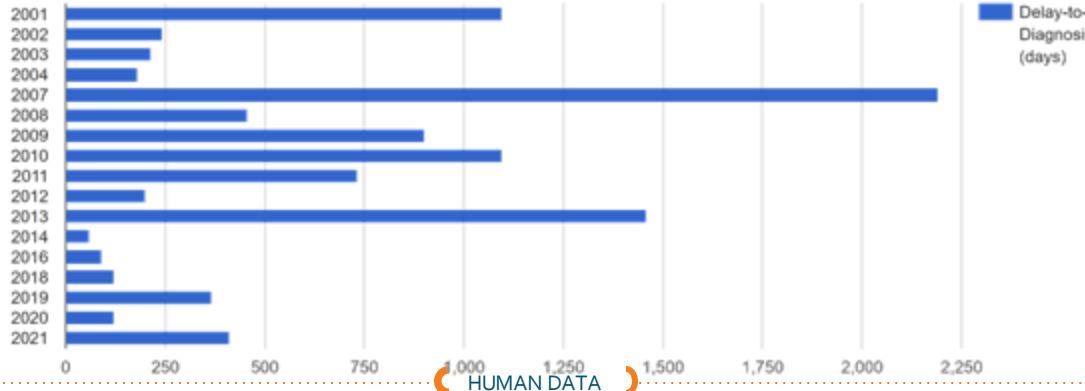
```
registry = "OPP"
diseasecode = "Orphanet_08896"

data_rows = []
label = "Duchenne muscular dystrophy"
disease_hash[registry].keys.each do |year|
  disease_hash[registry][year].each do |disease, offset|
    match = disease.match(/\d{4}/)
    next unless match[1] == diseasecode
    label = diseasenames[disease]
    data_rows.append [year, offset]
  end
end

Index = Doru::Index.new ["Year", "Delay-to-Diagnosis (days)"]
frame = Doru::DataFrame.rows(data_rows)
frame.vectors = Index
table = Doru::View::Table.new(frame)
more data_rows.inspect
options = { title: "Time to diagnosis for #{label} in the #{registry} registry",
            type: "bar",
            height: 500
          }

chart = Doru::View::Plot.new(table.table, options)
chart.show_in_iruby
```

Time to diagnosis for "Duchenne muscular dystrophy" in the DPP registry



# FAIR Data Point Index (for machines)



## FAIR Data Point

Metadata for machines



Search FAIR Data Point...

Log in

Advanced

## FAIR Data Points

Filter:

All 24 Active 18 Inactive 0 Unreachable 1 Invalid 5 Unknown 0

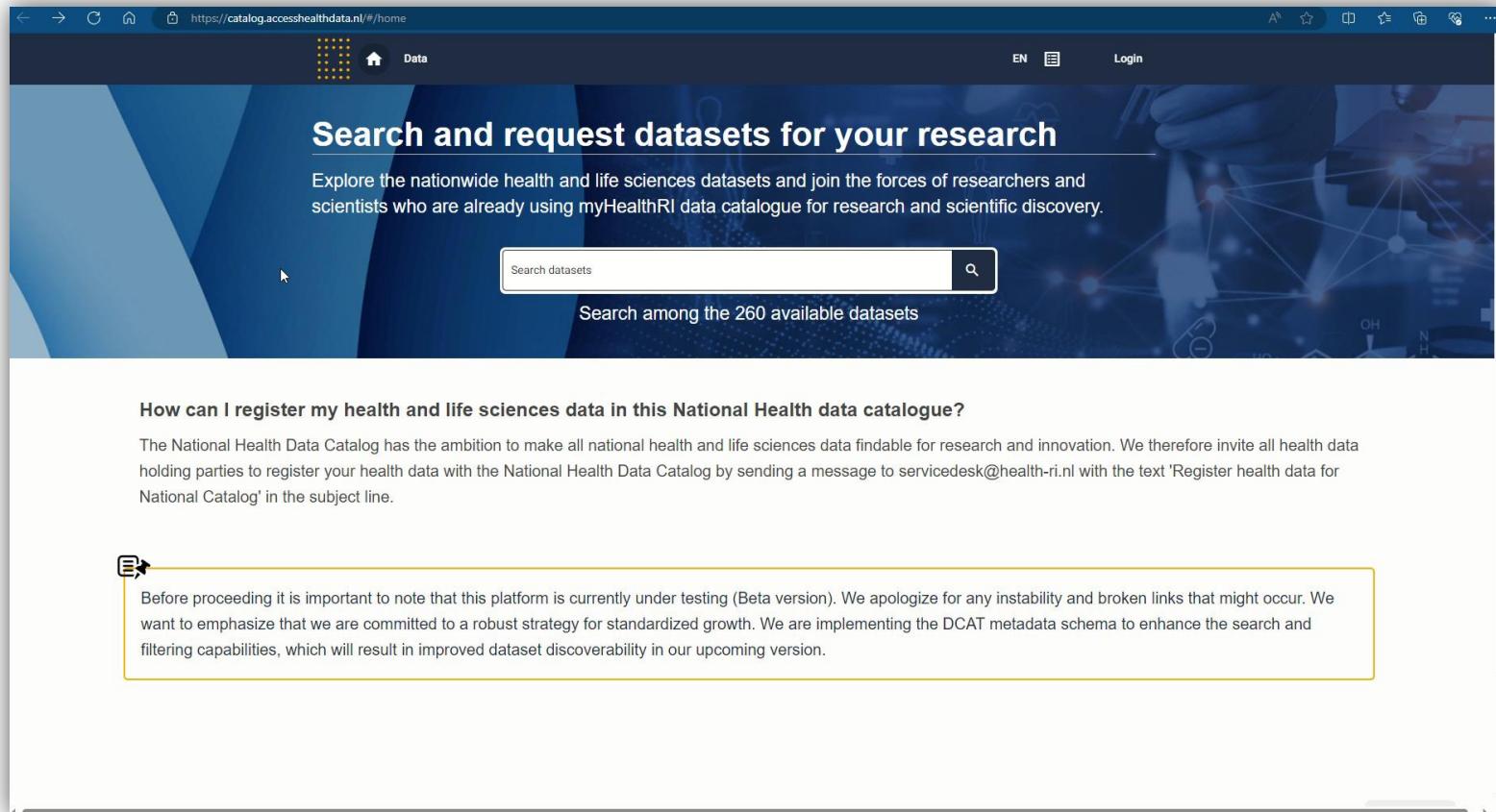
Endpoint ▲▼	Registration ▲▼	Modification ▲▼	Status
<a href="https://directory.bbmri-eric.eu/api/fdp">https://directory.bbmri-eric.eu/api/fdp</a>	04-10-2023, 16:12:15	17-04-2024, 14:00:00	ACTIVE
<a href="https://w3id.org/ctsr-fdp/">https://w3id.org/ctsr-fdp/</a>	15-12-2023, 06:00:00	17-04-2024, 06:00:00	ACTIVE
<a href="https://w3id.org/sympathetic/fdp">https://w3id.org/sympathetic/fdp</a>	11-07-2023, 13:23:57	16-04-2024, 13:23:57	ACTIVE
<a href="https://w3id.org/ctsr-fdp">https://w3id.org/ctsr-fdp</a>	27-02-2023, 09:23:29	16-04-2024, 13:16:50	ACTIVE
<a href="https://patient-registries.fdps.ejprd.semlab-leiden.nl/">https://patient-registries.fdps.ejprd.semlab-leiden.nl/</a>	17-05-2023, 10:41:43	16-04-2024, 13:16:47	ACTIVE
<a href="https://wp13.fdps.ejprd.semlab-leiden.nl/">https://wp13.fdps.ejprd.semlab-leiden.nl/</a>	17-05-2023, 10:41:29	16-04-2024, 13:16:44	ACTIVE
<a href="https://fdp.wikipathways.org/">https://fdp.wikipathways.org/</a>	04-07-2023, 16:36:52	16-04-2024, 13:16:43	ACTIVE
<a href="https://ejp-rd-fdp.ega-archive.org">https://ejp-rd-fdp.ega-archive.org</a>	02-03-2024, 11:53:09	16-04-2024, 12:53:05	ACTIVE



HUMAN DATA



# From 'for humans' to 'for machines' – Health RI portal showing ontologised metadata



The screenshot shows the homepage of the myHealthRI data catalogue. At the top, there is a navigation bar with icons for back, forward, refresh, and a search bar containing the URL <https://catalog.accesshealthdata.nl/#/home>. The main header features a yellow square icon with dots and the word "Data". To the right are language settings (EN), a grid icon, and a "Login" button. Below the header, a large banner with a blue and white abstract background displays the text "Search and request datasets for your research" and "Explore the nationwide health and life sciences datasets and join the forces of researchers and scientists who are already using myHealthRI data catalogue for research and scientific discovery." A search bar with the placeholder "Search datasets" and a magnifying glass icon is centered. Below the search bar, a button says "Search among the 260 available datasets". The main content area contains a section titled "How can I register my health and life sciences data in this National Health data catalogue?". It includes a paragraph about the catalog's ambition to make all national health and life sciences data findable and instructions for registering data by sending an email to [servicedesk@health-ri.nl](mailto:servicedesk@health-ri.nl). A callout box in the bottom left corner contains a note about the platform being in beta and the implementation of DCAT metadata.

How can I register my health and life sciences data in this National Health data catalogue?

The National Health Data Catalog has the ambition to make all national health and life sciences data findable for research and innovation. We therefore invite all health data holding parties to register your health data with the National Health Data Catalog by sending a message to [servicedesk@health-ri.nl](mailto:servicedesk@health-ri.nl) with the text 'Register health data for National Catalog' in the subject line.

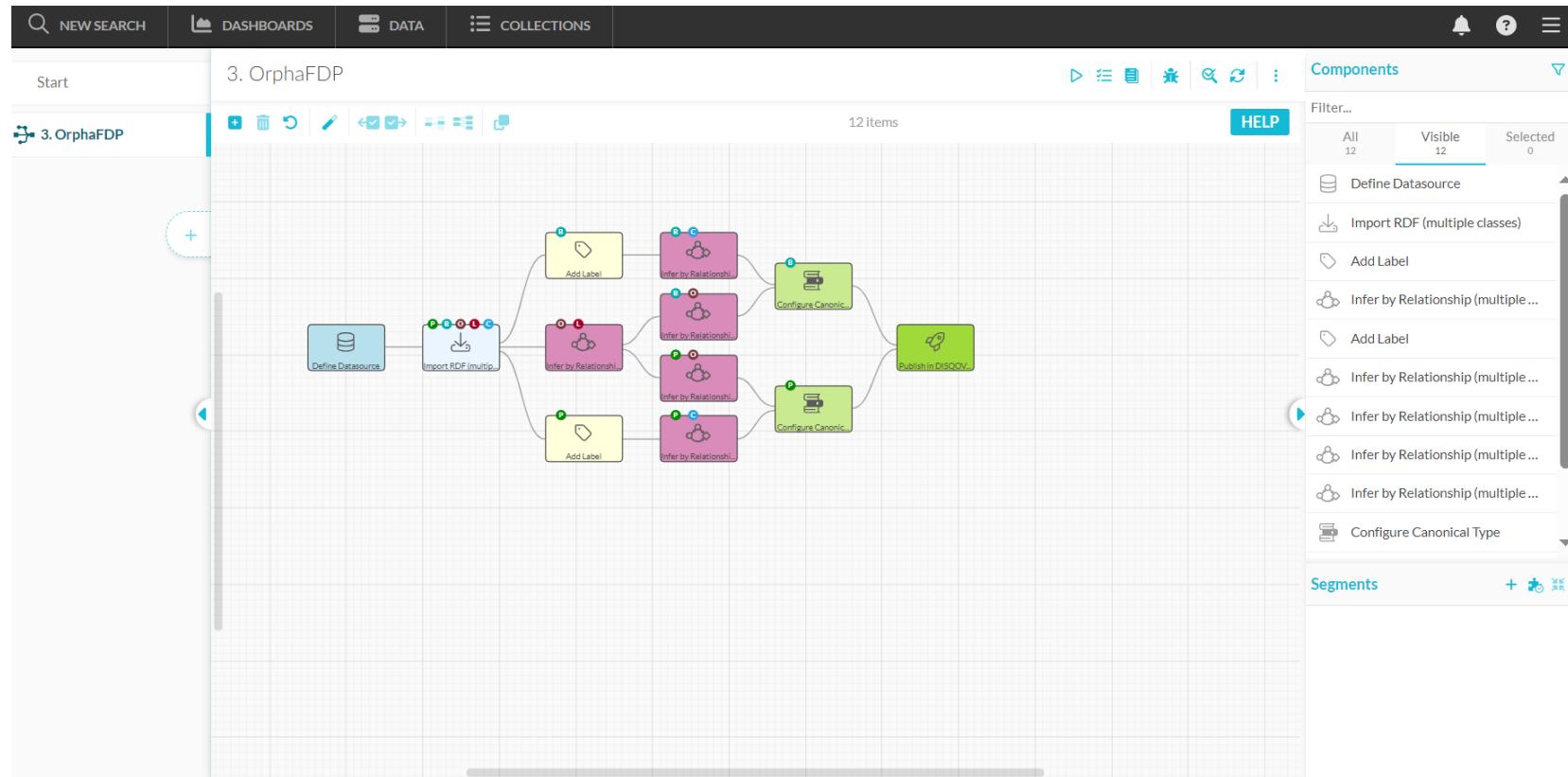
 Before proceeding it is important to note that this platform is currently under testing (Beta version). We apologize for any instability and broken links that might occur. We want to emphasize that we are committed to a robust strategy for standardized growth. We are implementing the DCAT metadata schema to enhance the search and filtering capabilities, which will result in improved dataset discoverability in our upcoming version.

# From 'for humans' to 'for machines' – the EJP RD VP Portal & its FAIR Data Points (FDP)

The screenshot shows the homepage of the EJP RD VP Portal. At the top, there's a navigation bar with links like 'My Proposal(s)', 'EJP RD', 'Open tabs', 'Science', 'uAccess Home', 'Getting Started', 'LIST OF PUBLISHERS', 'Vacature businessanalist...', 'Pillar2-CENTRAL Note...', 'WeevilScout', and 'Other Bookmarks'. Below the navigation is the European Joint Programme Rare Diseases VP-PORTAL logo. The main content area has a dark blue header with 'RESOURCE DISCOVERY', 'VP NETWORK RESOURCES', and a user profile for 'MARCO ROOS'. The main heading is 'The EJP-RD Virtual Platform'. Below it, text describes the VP as a growing network of FAIR resources and includes a paragraph about its features, followed by a 'Learn More' link. To the right is a colorful illustration of clouds connected by a winding path. Below this is a search bar with placeholder text 'Search for a disease name (e.g. ADPKD), gene (e.g. PKD1), or Orphacode (e.g. 730)' and a search button. A link 'Advanced Search' is also present. The bottom section is titled 'Tools and Resources' and features three icons: a red 'EXPLORE' button with a right-pointing arrow, a yellow 'GUIDANCE' button with a circular arrow, and a green 'CONNECT' button with a hand icon. Below each button is a brief description: 'Explore Rare Disease resources with the EJP RD Mind Map', 'Find out how to make your data more FAIR', and 'Contact us for information or feedback'. At the very bottom, there's a decorative footer with the text 'HUMAN DATA' and the Elixir logo.

# From 'for humans' to 'for machines' – smart use of ontological metadata

Example: FDPs automatically ingested in OntoForce's semantic data navigation tool 'Disqover'



# From 'for humans' to 'for machines' – smart use of ontological metadata

Example: FDPs automatically ingested in OntoForce's semantic data navigation tool 'Disqover'

The screenshot shows the Disqover interface with a dark header bar containing 'NEW SEARCH', 'DASHBOARDS', 'DATA', 'COLLECTIONS', a search icon, a bell icon, a question mark icon, and a three-dot menu icon. To the right of the header are 'Default' and a gear icon. The main area features a central 'DISQOVER' logo with a magnifying glass icon. Below it is a search bar with the placeholder 'I am looking for...'. A grid of 3x6 circles, each representing a dataset with an icon and a name. The data points are:

Category	Dataset	Count
Biobank	Biobank	174
Patient Registry	Patient Registry	850
Clinical Protocol	Clinical Protocol	898k
Facility	Facility	889k
Journal	Journal	43.8k
Location	Location	113k
Ontology	Ontology	969k
Active Substance	Active Substance	11.1k
Adverse Event	Adverse Event	16.7M
Antibody	Antibody	1.98M
Assay	Assay	20.3M
Biospecimen	Biospecimen	4.17k
Cell Line	Cell Line	363k
Chemical	Chemical	3.05M
Clinical Study	Clinical Study	796k
Disease	Disease	228k
Drug Treatment	Drug Treatment	34.1M
Enzyme	Enzyme	12.4k
Gene	Gene	49.2M
Homology	Homology	44.2k
Medicine	Medicine	742k
Model Organism	Model Organism	87.3k
Organism	Organism	2.55M
Organization	Organization	449k
Patent	Patent	6.69M
Plasmid	Plasmid	29.6k
Pathway	Pathway	23.7k
Person	Person	45.6M
Project	Project	3.22M
Protein	Protein	613k
Publication	Publication	37.1M
Transcript	Transcript	11.3M
Medical Device	Medical Device	5.05M
Variant	Variant	22.7M

At the bottom left, it says 'DISQOVER powered by ONTOFORCE • © 2022 LUMC • Cookie Policy • License agreement'. At the bottom right is the Elixir logo.

# From 'for humans' to 'for machines' – smart use of ontological metadata

Example: FDPs automatically ingested in OntoForce's semantic data navigation tool 'Disqover'

The screenshot shows the OntoForce Disqover interface for a dataset named 'Biobank'. The top navigation bar includes 'NEW SEARCH', 'DASHBOARDS', 'DATA', 'COLLECTIONS', and a notification bell. The main search bar is set to 'Biobank'. On the left, a sidebar lists various categories such as 'All', 'Clinical Study', 'LUMC Facility', and 'Leiomyosarcoma Cell Line'. The central area features several visualizations: a 'Count' gauge showing 174 results, a 'Text Search' input field, a 'Key words' section listing URLs, and two pie charts under 'from location' and 'Coverage'. A sidebar on the right lists 174 results, each with a checkbox and a 'Label' description. A large blue '+' button is located at the bottom right of the visualization area.

Count

Text Search

I am looking for...

Key words

174 [http://edamontology.org/topic\\_3337](http://edamontology.org/topic_3337)

174 <http://www.wikidata.org/entity/Q864217>

7 [http://www.orpha.net/ORDO/Orphanet\\_98053](http://www.orpha.net/ORDO/Orphanet_98053)

6 [http://www.orpha.net/ORDO/Orphanet\\_273](http://www.orpha.net/ORDO/Orphanet_273)

6 [http://www.orpha.net/ORDO/Orphanet\\_399](http://www.orpha.net/ORDO/Orphanet_399)

6 [http://www.orpha.net/ORDO/Orphanet\\_803](http://www.orpha.net/ORDO/Orphanet_803)

5 [http://www.orpha.net/ORDO/Orphanet\\_68381](http://www.orpha.net/ORDO/Orphanet_68381)

5 [http://www.orpha.net/ORDO/Orphanet\\_797](http://www.orpha.net/ORDO/Orphanet_797)

5 [http://www.orpha.net/ORDO/Orphanet\\_98896](http://www.orpha.net/ORDO/Orphanet_98896)

Filter 414 values

1 / 46

from location

Multi-valued

Coverage

174 results

Search in 174 results

Export

Unsaved template

Label Enroll-HD Biobank

Label Adrenocortical Carcinoma Biobank - part of the ENS

Label HUSEC reference strain collection

Label PXE Biobank (Plasma, serum, DNA, skin fibroblasts)

Label Huntington disease biobank

Label Movement Disorders Biobank (EuroBioBank partner)

Label Arrhythmogenic right ventricular cardiomyopathy/dysplasia: biobank of DNA and tissue samples

Label MPN Registry: German MPN Register and Biomaterial Bank for BCR-ABL1-negative myeloid neoplasias

HUMAN DATA



# From 'for humans' to 'for machines' – mining the ontologised metadata of FDPs

The screenshot shows a web browser window for the FLAIR-GG Virtual Platform. The URL is https://vp.bgv.cbgp.upm.es/flair-gg-vp-server/resources. The page displays a list of discovered resources and provides search and ontology URI functions.

**Discovered VP Resources**

- Resource: [EURO-NMD Registry Catalog](#) (Catalog)
- Resource: [CTSR catalog](#) (Catalog)
- Resource: [SMarCARE Catalog](#) (Catalog)
- Resource: [FAIRVASC](#) (Catalog)
- Resource: [a tool that will create Box Whisker plots](#) (Dataservice)
- Resource: [CTSR dataset](#) (Dataset)
- Resource: [SMarCARE Dataset](#) (Dataset)

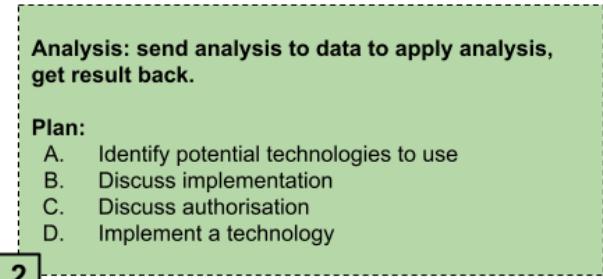
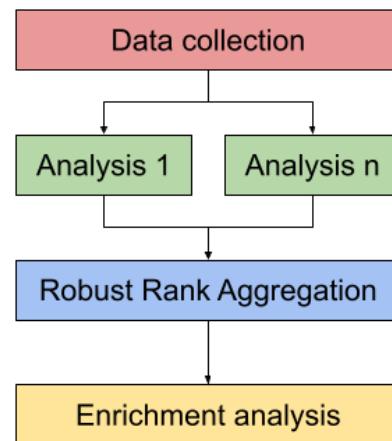
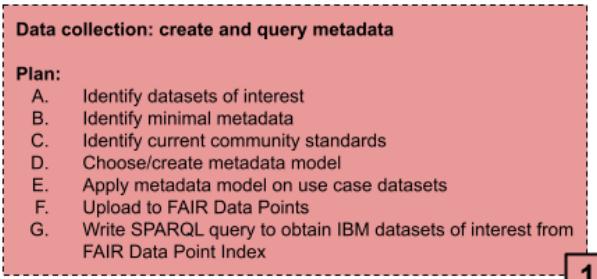
**Keyword Search:**

**Ontology URI:**

© 2023 Mark D Wilkinson

# Federated analysis plan

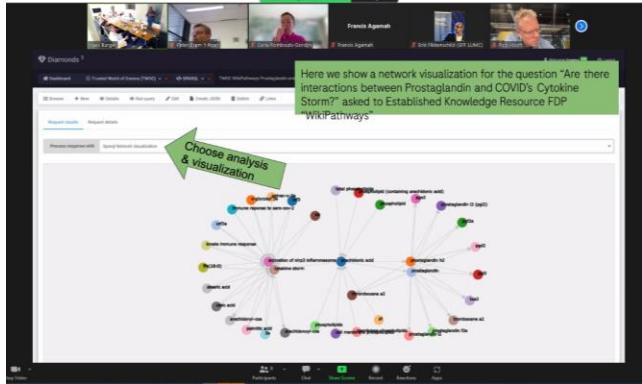
## Plan for federated analytics use case: differential gene expression meta-analysis for Inclusion Body Myositis



Metadata schema has been defined  
↓  
Allows identification of relevant datasets

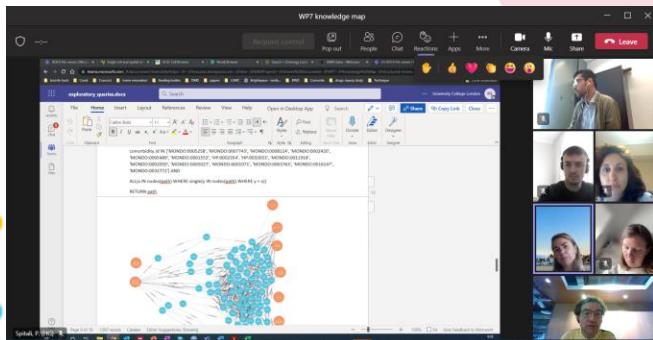
Federated analysis is being implemented  
↓  
Will allow privacy preserving analysis

# Interdisciplinary, real-time collaboration enabled by FAIR data



Exploration of synthetic Covid-19 real world observations enabled by 'on the fly' SPARQL queries on FAIR data.

Courtesy of Eugene van Someren (TNO). and the 'Trusted World of Corona' project. Synthetic data and ontological data models by Núria Queralt-Rosinach (reuses EJPRD data element model)



Exploration of Duchenne data and knowledge enabled by 'on the fly' Neo4J queries on FAIR data.

Courtesy of Pietro Spitali, Núria Queralt-Rosinach and the BIND project. Reuse of EJP RD ontological data model.



Funded by the European Union  
GA n°825575

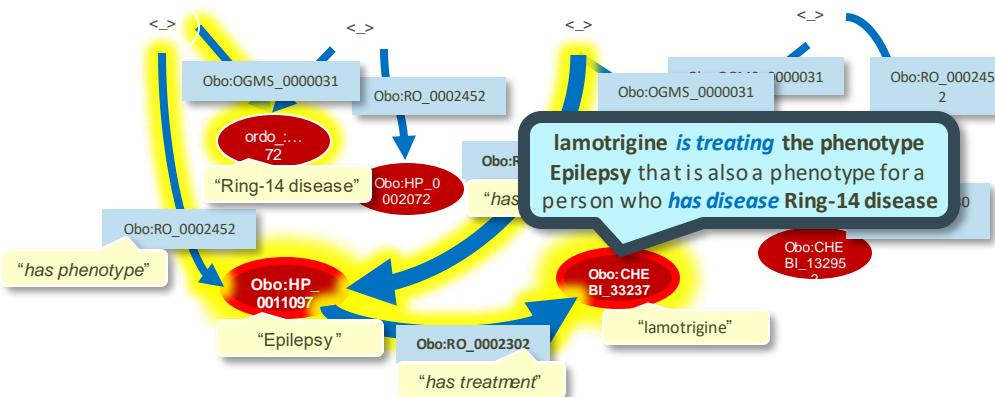
# Introduction to SPARQL

SPARQL = SPARQL Protocol and RDF Query Language

- A language for querying knowledge graphs consisting of subject-predicate-object triples made from URIs
- Syntax similar to SQL, in no other way similar
- SPARQL queries are graph matching queries
  - Queries are graphs with some subject/predicate/objects as variables. They are matched with the existing graphs; where the graph fits the variables are filled in.
- Useful for extracting specific information from distributed RDF data
- Useful for retrieving data sets across multiple data sets as if they were one database

Which treatment *is treating* the phenotype that is also a phenotype for Monica who *has disease* Ring-14 disease

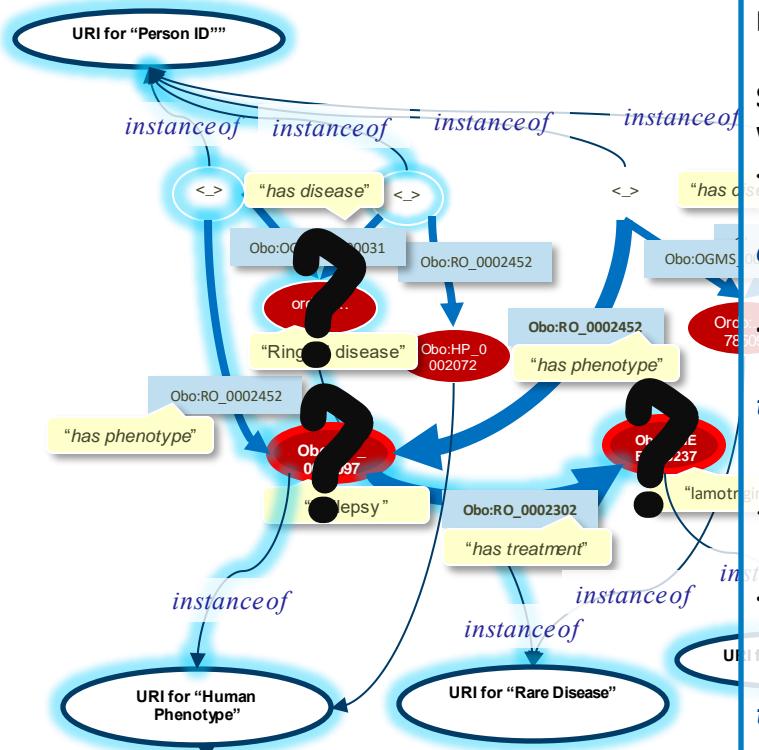
Monika	Annika	Rajaram	Pietro
$\langle \rangle$ obo:OGMS_0000031 ordo:Orphanet_1440 obo:RO_0002452 obo:HP_0011097.	$\langle \rangle$ obo:OGMS_0000031 ordo:Orphanet_1440, obo:RO_0002452 obo:HP_0002072.	$\langle \rangle$ obo:OGMS_0000031 ordo:Orphanet_178509, obo:RO_0002452 obo:HP_0011097 obo:RO_0002302 obo:CHEBI_33237	$\langle \rangle$ obo:OGMS_0000031 ordo:Orphanet_178509, obo:RO_0002452 obo:HP_0000720 obo:RO_0002302 obo:CHEBI_132952



Which treatment **is treating** the phenotype  
that **is a** phenotype for a  
Person who **has disease**  
**Disease with the same**  
**phenotype**



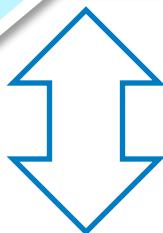
lamotrigine **is treating**  
the phenotype Epilepsy  
that **is a** phenotype for  
person Monica who **has**  
**disease** Disease **with the**  
**same phenotype**



### Pseudo SPARQL query

```
SELECT ?disease ?treatment ?phenotype
WHERE
{
    ?person <URI for "has rare disease"> ?disease .
    ?person <URI for "instance of type"> <URI for "Person"> .
    ?disease <URI for "instance of type"> <URI for "Rare Disease"> .
    ?person <URI for "has phenotype"> ?phenotype .
    ?phenotype <URI for "instance of"> <URI for "Human phenotype"> .
    ?phenotype <URI for "has treatment"> ?treatment .
    ?treatment <URI for "instance of"> <URI for "Treatment"> .
}
```

Which treatment **is treating** the phenotype that **is a** phenotype for a Person who **has disease**  
Ring-14 disease



lamotrigine **is treating** the phenotype Epilepsy that **is a** phenotype for a person who **has disease**  
Ring-14 disease

### Pseudo SPARQL query

SELECT ?disease ?treatment ?phenotype

WHERE

{

?person\_uri <URI for "has rare disease"> <URI for "Ring 14-disease"> .  
?person\_uri <URI for "instance of type"> <URI for "Person"> .

?person\_uri <URI for "has phenotype"> ?phenotype\_uri .  
?phenotype\_uri <URI for "instance of type"> <URI for "Human phenotype"> .  
?phenotype\_uri <URI for "human readable label"> ?phenotype .

?phenotype\_uri <URI for "has treatment"> ?treatment\_uri .  
?treatment\_uri <URI for "instance of type"> <URI for "Treatment"> .  
?treatment\_uri <URI for "human readable label"> ?treatment .

}

URI for "Human Phenotype"

URI for "Treatment"

URI for "Rare Disease"

URI for "has treatment"

URI for "instance of type"

URI for "has phenotype"

URI for "instance of type"

URI for "has rare disease"

URI for "instance of type"

URI for "has disease"

URI for "instance of type"

URI for "Person ID"

## SPARQL conclusions

SPARQL enables to query across data sets across the world in terms of globally understood semantic models (ontologies)

SPARQL can also be used to retrieve data sets from across multiple sources for use in other data analysis tools (e.g. R)

SPARQL is not an implementation of FAIR principles!

Essential for FAIR is the guarantee that an algorithm can know what data mean on the inside of a data source if allowed in.

# Wrap up



1. FAIR principles were conceived to increase the efficiency of reusing data computationally across the globe
2. Ontology-based conceptual models capture a shared understanding of what data mean for machines, allowing automation and computing with knowledge
3. RDF is a carrier of globally linked ‘ontologized data’ using URIs and a simple ‘subject-predicate-object’ data model
4. SPARQL is a language for querying and retrieving RDF data across multiple locations
5. JSON/YAML/RelDB schemas are great for programmers, but lack semantics in a form that computers understand globally