LEIBNIZ-INFORMATIONSZENTRUM
TECHNIK UND NATURWISSENSCHAFTEN
UNIVERSITÄTSBIBLIOTHEK

TIB

# Challenges for Transforming Big data into Knowledge: Precision Medicine- A Use Case

Maria-Esther Vidal
Scientific Data Management Group TIB

# Motivating Example

*A **53-year** old **male** patient with Non Squamous carcinoma of stage III. He is a **non-smoker** and **non-positive for EGFR, ROS1, and ALK**. He suffers from **high cholesterol** and takes **Simvastatin**. **No relatives with cancer**. His **father, mother, and older brother died** from **heart attacks.** He is prescribed with **Paclitaxel***

Medical images

Liquid Biopsies

Solid Biopsies
Clinical tests

# Motivating Example

*A **53-year** old **male** patient with Non Squamous carcinoma of stage III. He is a **non-smoker** and **non-positive for EGFR, ROS1, and ALK**. He suffers from **high cholesterol** and takes **Simvastatin**. **No relatives with cancer**. His **father, mother, and older brother died** from **heart attacks.** He is prescribed with **Paclitaxel***

Medical images

Liquid Biopsies

Solid Biopsies
Clinical tests

Factors that may impact on the effectiveness of an oncological treatment?

# Motivating Example

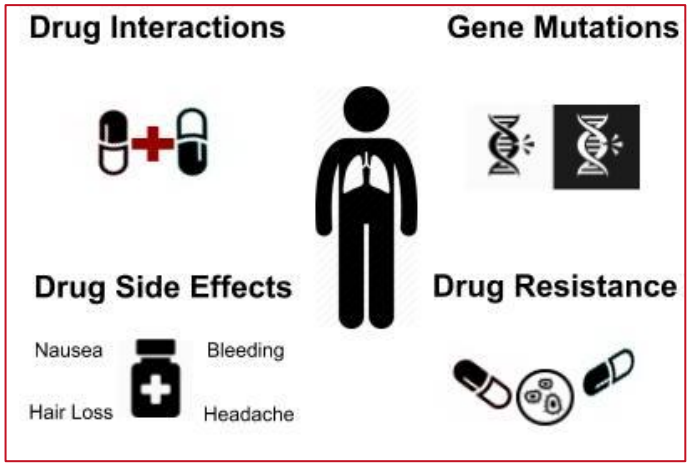
*A **53-year** old **male** patient with Non Squamous carcinoma of stage III. He is a **non-smoker** and **non-positive for EGFR, ROS1, and ALK**. He suffers from **high cholesterol** and takes **Simvastatin**. **No relatives with cancer**. His **father, mother, and older brother died** from **heart attacks.** He is prescribed with **Paclitaxel***

Medical images

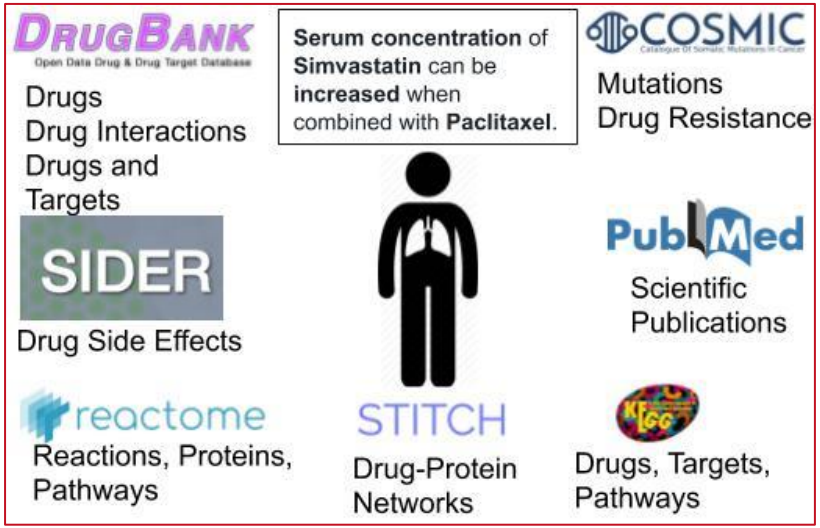
Liquid Biopsies

Solid Biopsies
Clinical tests

Factors that may impact on the effectiveness of an oncological treatment?



# Motivating Example




Heterogeneous Data Sources need to be searched in order to identify issues that may impact on the effectiveness of the treatments

# Motivating Example

Search Drug: Pacitaxel → Protein Selection: CYP3A4 → Transcript of the Protein → Search in Mutations → Extract Mutations

Collect **26 records of interactions** with proteins
**Extract Relevant** Information from the collected records

Select the **Proteins interacting** with the drug

Search for **the Entrez Id** of the Proteins
ENSP00000337915

Search for **the Entrez ID** of the transcript of the Protein
ENST00000336411

Search in the COSMIC dataset of Mutations

Select the mutations can impact the treatment effectiveness **Missense-substitution** results in altering the amino acid at that position **(and may end alternation of the protein which it is coding!)**
T363M, P474S, E412Q, L401F
**Nonsense-substitution** results in a termination codon and foreshortening the translated peptide **(an usually cause nonfunctional protein product!)**
Q265*

Genes Associated with Proteins

| Gene(s) | Condition(s) |
|---|---|
| COL10A1, NT5DC1 | Metaphyseal chondrodysplasia |
| FLNC | Cardiomyopathy, familial hypertrophic, 26, Myopathy, distal, 4, Dilated Cardiomyopathy, Dominant, Myofibrillar myopathy, filamin C-related |
| SLC13A5 | Seizures |
| CHRNA4 | Autosomal dominant nocturnal frontal lobe epilepsy |

# Big Data Dimensions

**Variety**
Source Heterogeneity

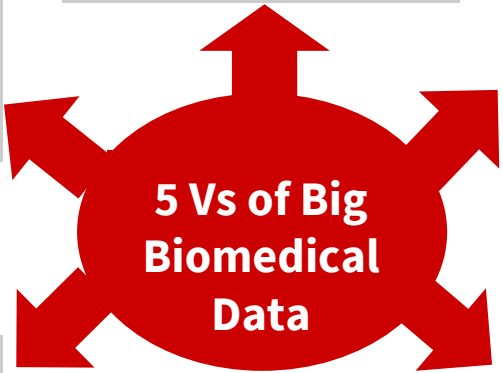
**Veracity**
Noise and Missing Data

**Volume**
Large scale data sets

What can be obtained
**Value**

Data on motion
**Velocity**

# Big Biomedical Data




### Volume

**Large-scale** datasets,
**TCGA** >32.000 **cases**,
**UK Biobank**>500.000 **cases**
**Public Images > 12 TB**

### Variety

**Various** formats of clinical and genomics data including:
**Diverse** health records, treatments, diagnostics, and outcomes. Tests and results in different forms: images, genomic and molecular test.

### Value

**Great potential** to improve and advance healthcare: **Personalised** medicine, **interpretation** of medical images; **unrevealing** of patterns in patient's conditions and outcomes, and health d**ecision making**

### Velocity

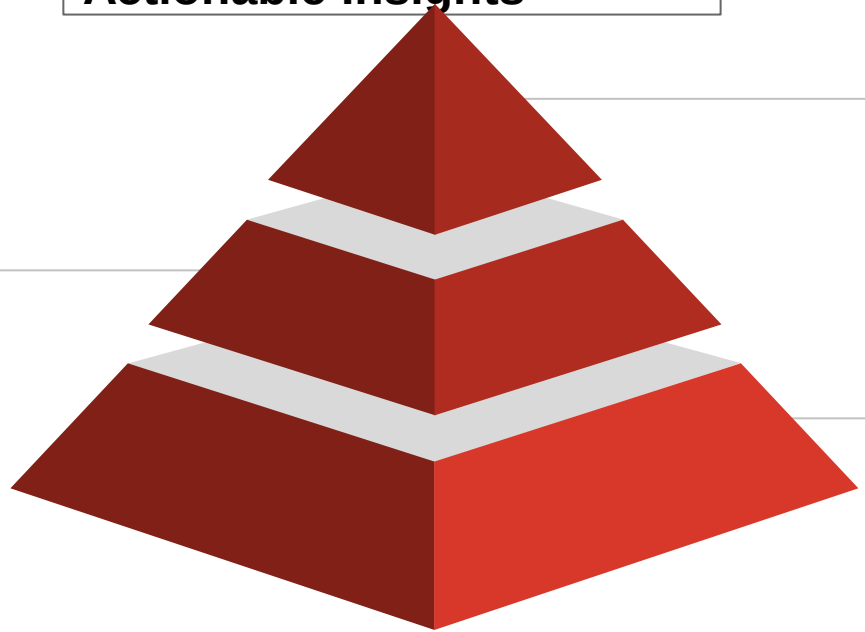
Data generated at various levels of **granularity**:
Patient's visits, interventions outcomes, genomic and molecular test results, clinical trials

### Veracity

**Incompleteness, deduplication** and **disambiguation** of data about patients, observations, intervention results, , and outcomes

# Big Data Challenges

## Transforming Big Data into Actionable Insights




### Representing Actionable Insights

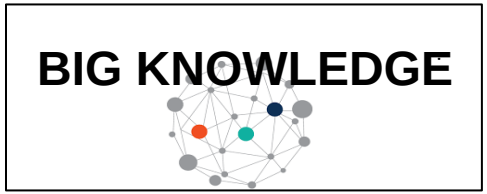
**Knowledge Discovery** methods are used to uncover patterns and hidden associations

### Ingesting and Describing Data

**Ontologies** are defined and utilized to describe data

### Semantic Data Integration

**Knowledge Graphs** allow for the representation of knowledge extracted from Big Data

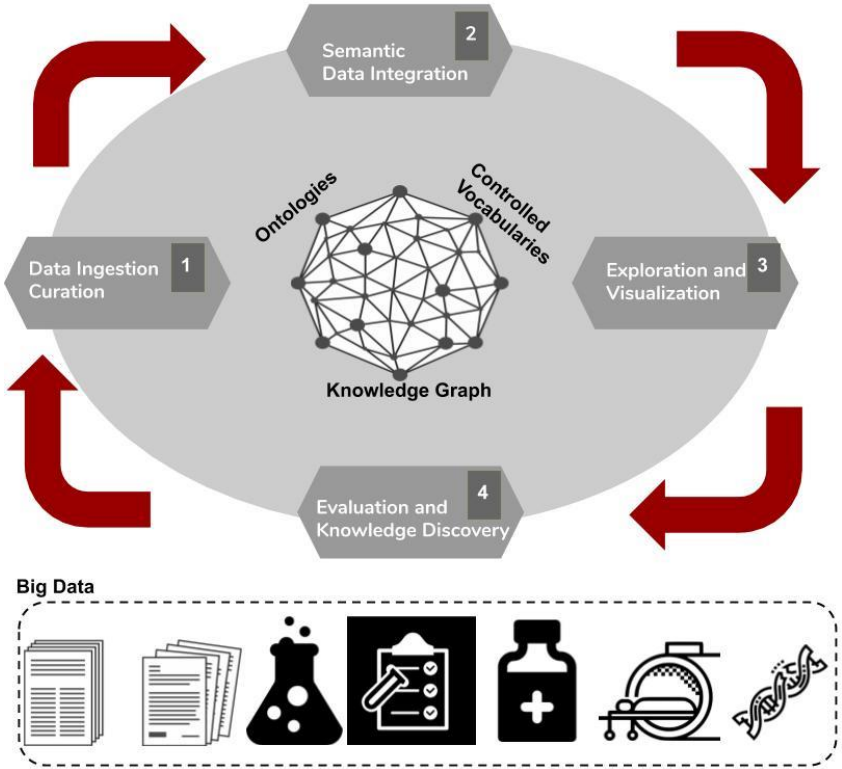


# Agenda

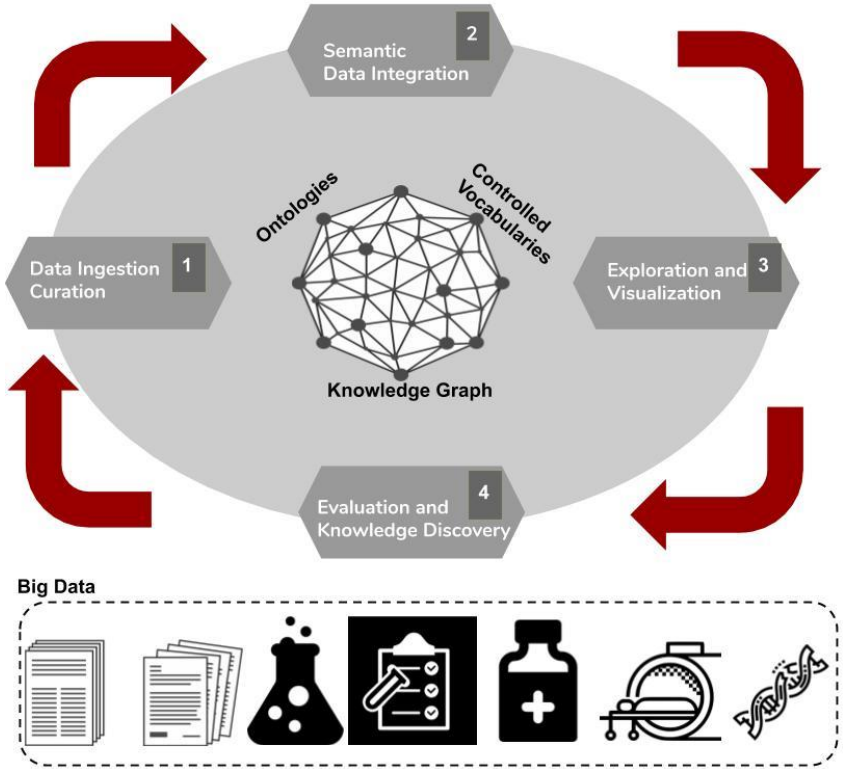
1. A Knowledge-Driven Approach
    a. Knowledge Graphs
2. Big Data Ingestion and Knowledge Extraction
3. Semantic Data Integration
4. Federated Query Processing
5. Knowledge Discovery
6. Applications
7. Grand Challenges and Opportunities

# A Knowledge-Driven Approach

# A Knowledge-driven Approach



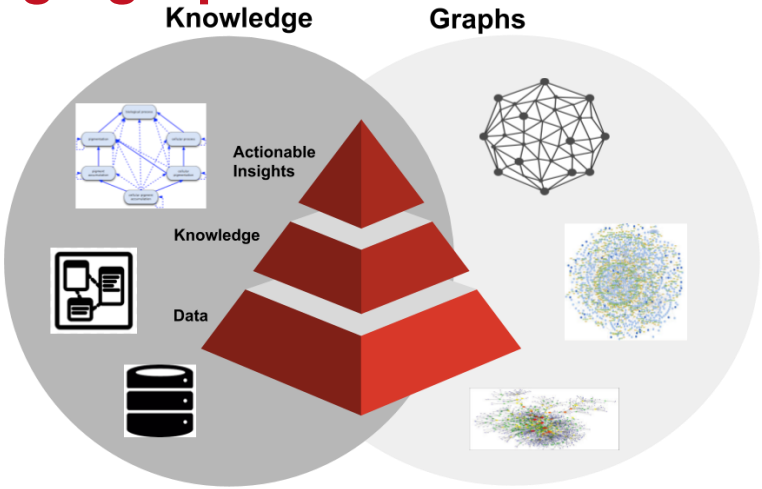
# A Knowledge-driven Approach



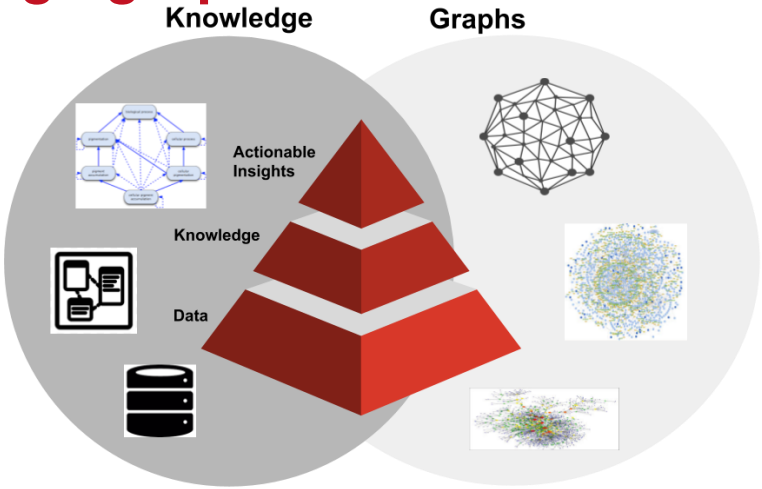

- **Big data sources** are ingested, curated, and integrated into a knowledge graph.
- Diverse **knowledge extraction methods** enable to transform unstructured data and describe the extracted facts using ontologies.
- **Federated query processing** and visualization tools enable the exploration of the knowledge graph
- **Knowledge discovery techniques** facilitate the uncovering of relevant **patterns** and **relations**.

Maria-Esther Vidal, Kemele M. Endris, Samaneh Jozashoori, Ahmad Sakor, Ariam Rivas. Transforming Heterogeneous Data into Knowledge for Personalized Treatments- A Use Case. Journal Datenbank-Spektrum. 2019

# What is a knowledge graph?



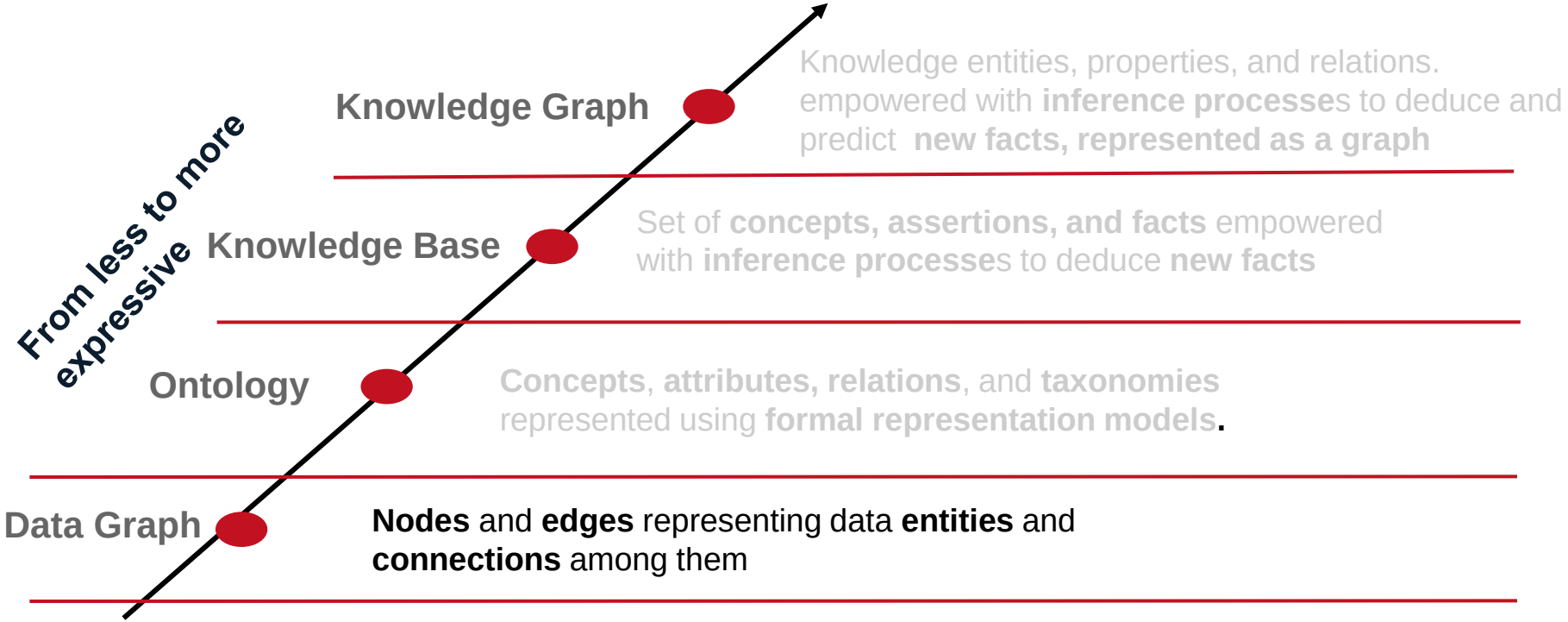
# What is a knowledge graph?



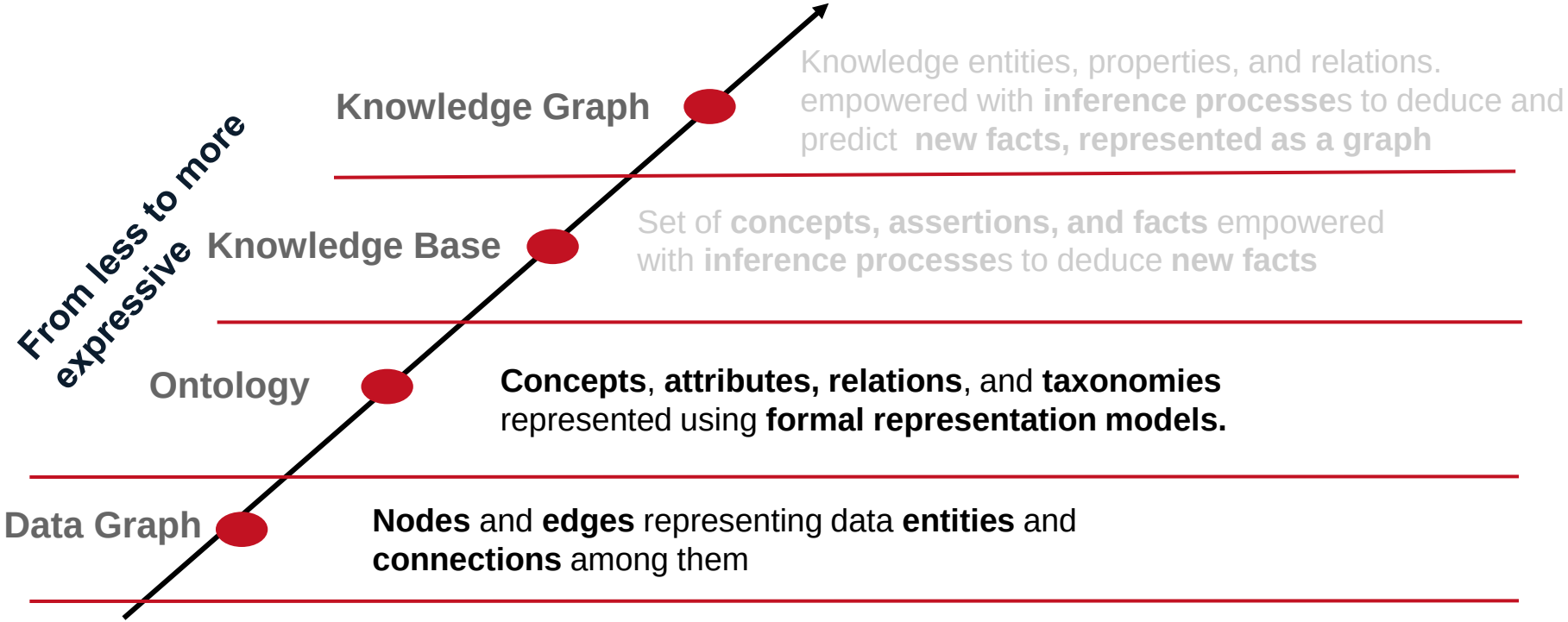

A **Knowledge Graph** is a graph **KG=(O,V,E)**:

- **O** is an ontology
- **V** is a set of entities representing data, information, or knowledge. Types of the entities in **V** are defined in **O**
- **E** is a set of edges between entities in **V. Edges** are labeled with predicates in **O**. The meaning of these predicates is also stated in **O.**

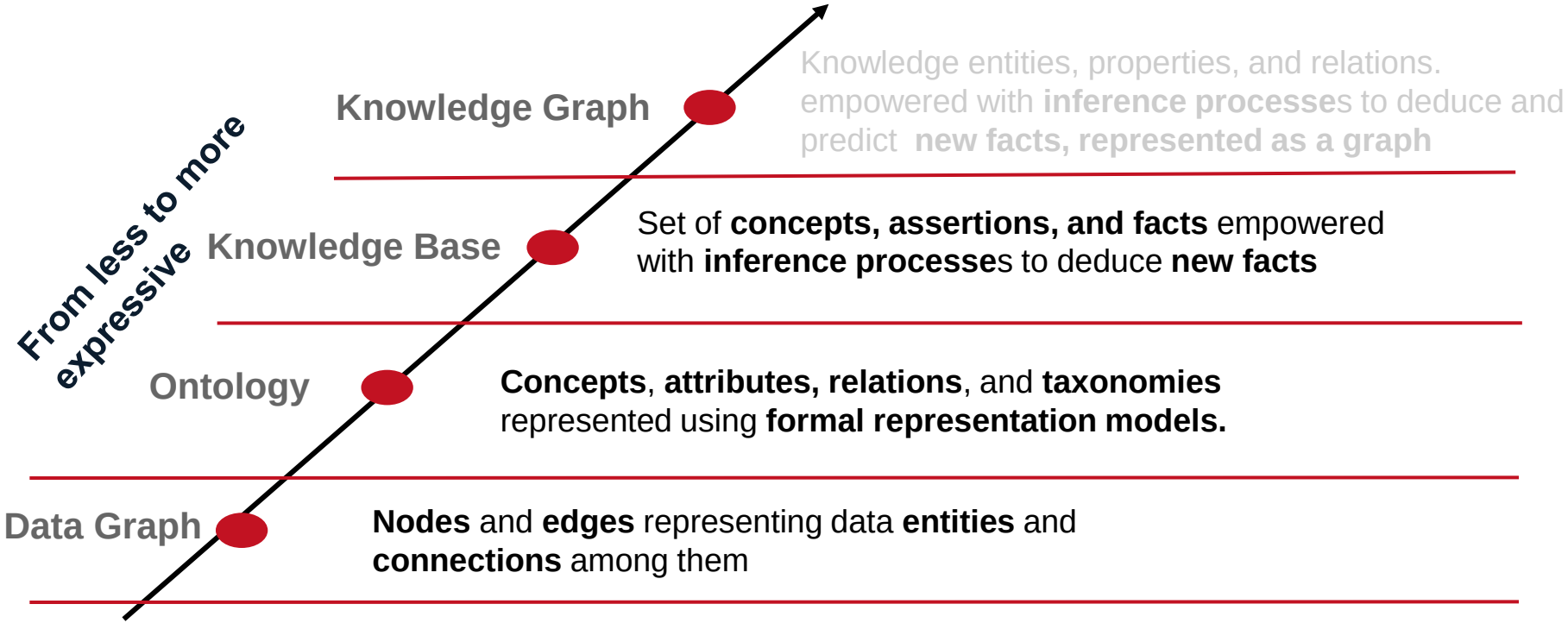
# Knowledge Graph Spectrum



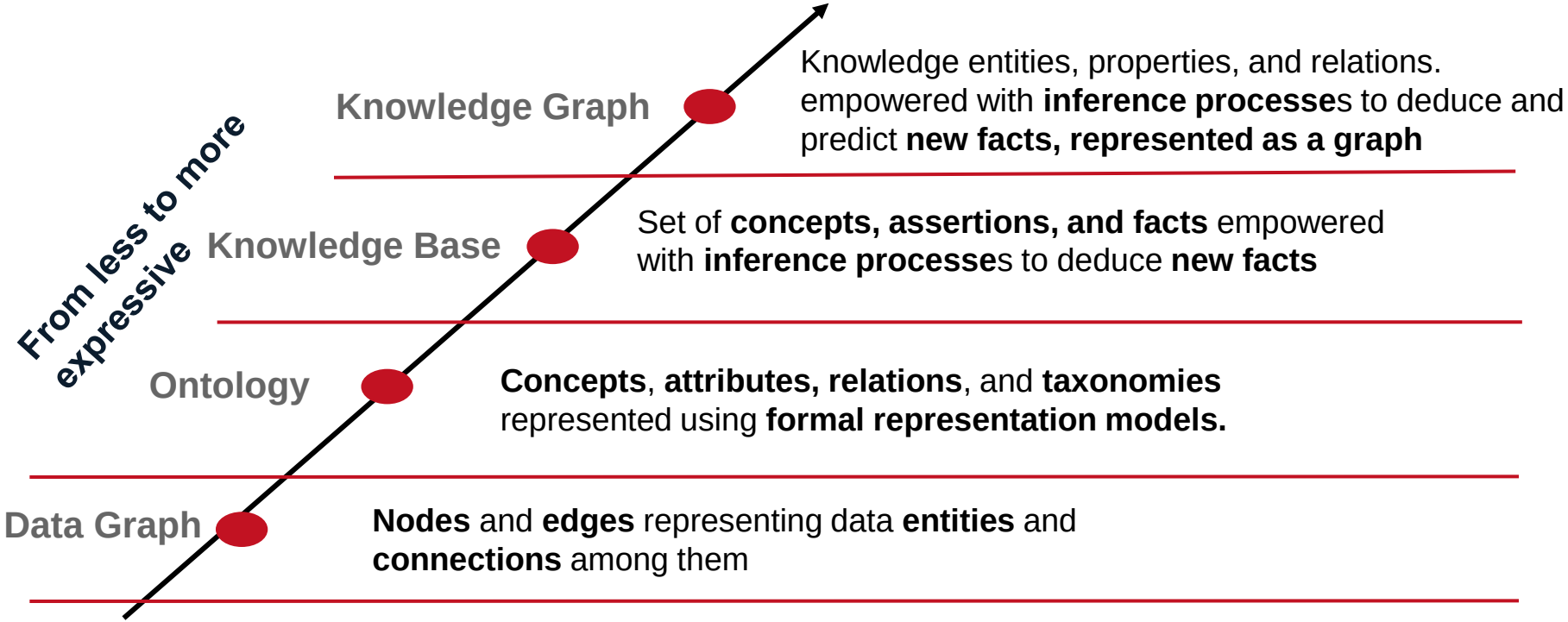
# Knowledge Graph Spectrum



# Knowledge Graph Spectrum

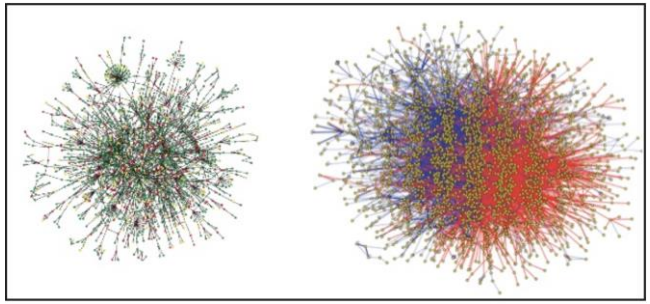


# Knowledge Graph Spectrum

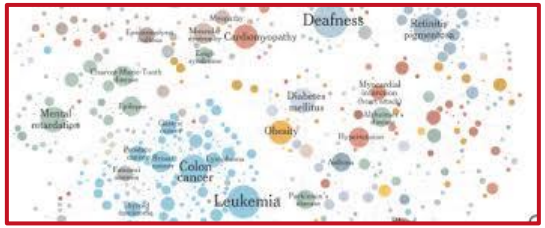


# Graph Databases in the Biomedical Domain

**Protein-Protein Interaction Networks** (https://www.ebi.ac.uk/training/online/course/network-analysis-protein-interaction-data-introduction/protein-protein-interaction-networks)



**Drug-Path: a database for drug-induced pathways (**http://www.cuilab.cn/drugpath)




**Diseasome: a network of gene-disease interactions.**
(http://www.kobic.kr/diseasome/)

# Gene Ontology

**Knowledge Model** of Biological Systems from **Molecular Level** to **pathways**, **cellular**, and **organism-level** systems

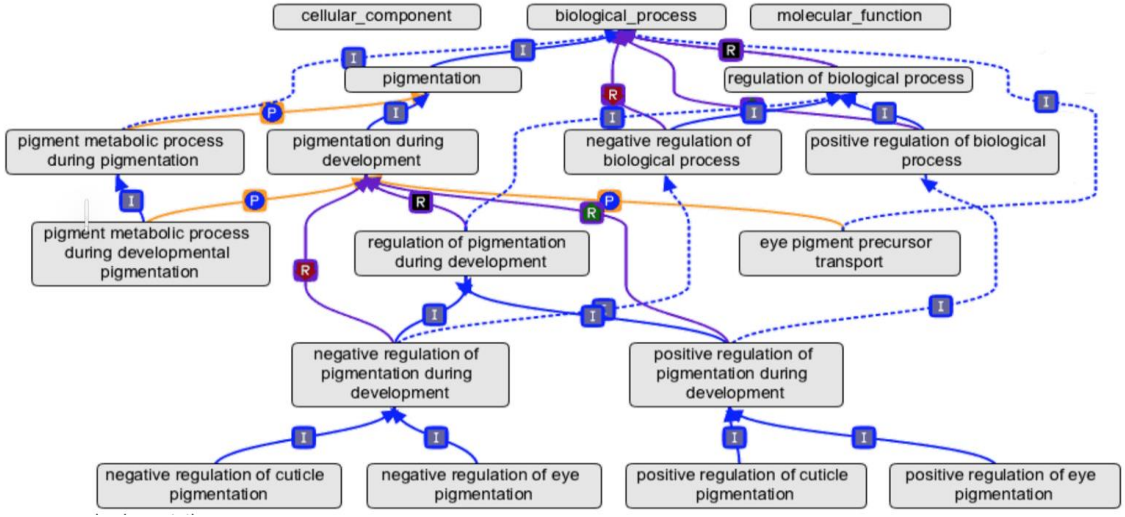
**Cellular Component: where gene products operate**

**Molecular Function: what gene products do**

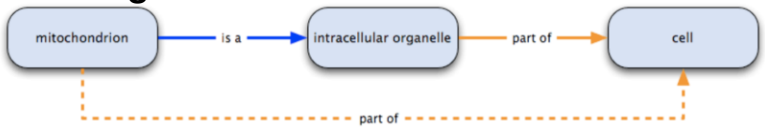
**Biological Process: pathways and processes that gene processes participate in**

http://www.geneontology.org/page/ontology-structure

## Classes and Relationships

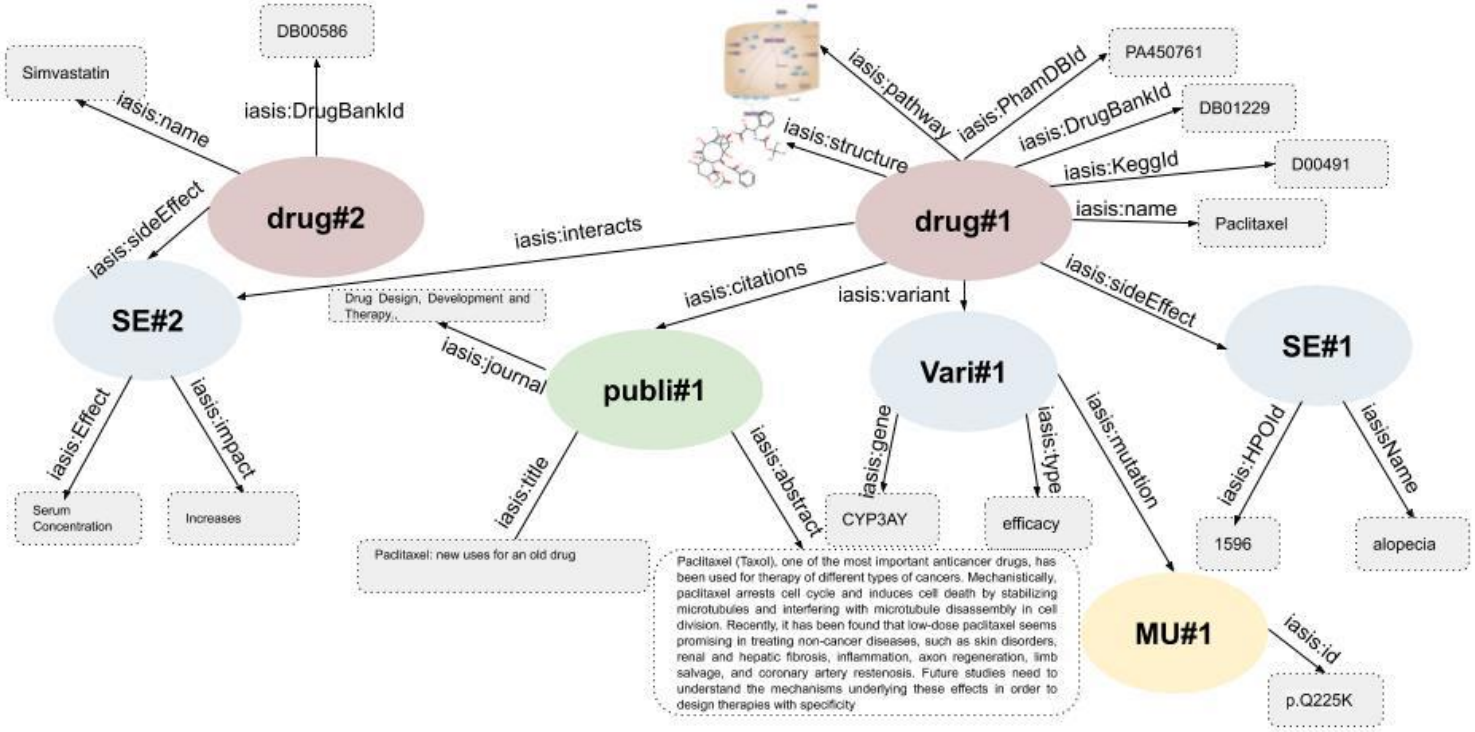



## Reasoning Rules



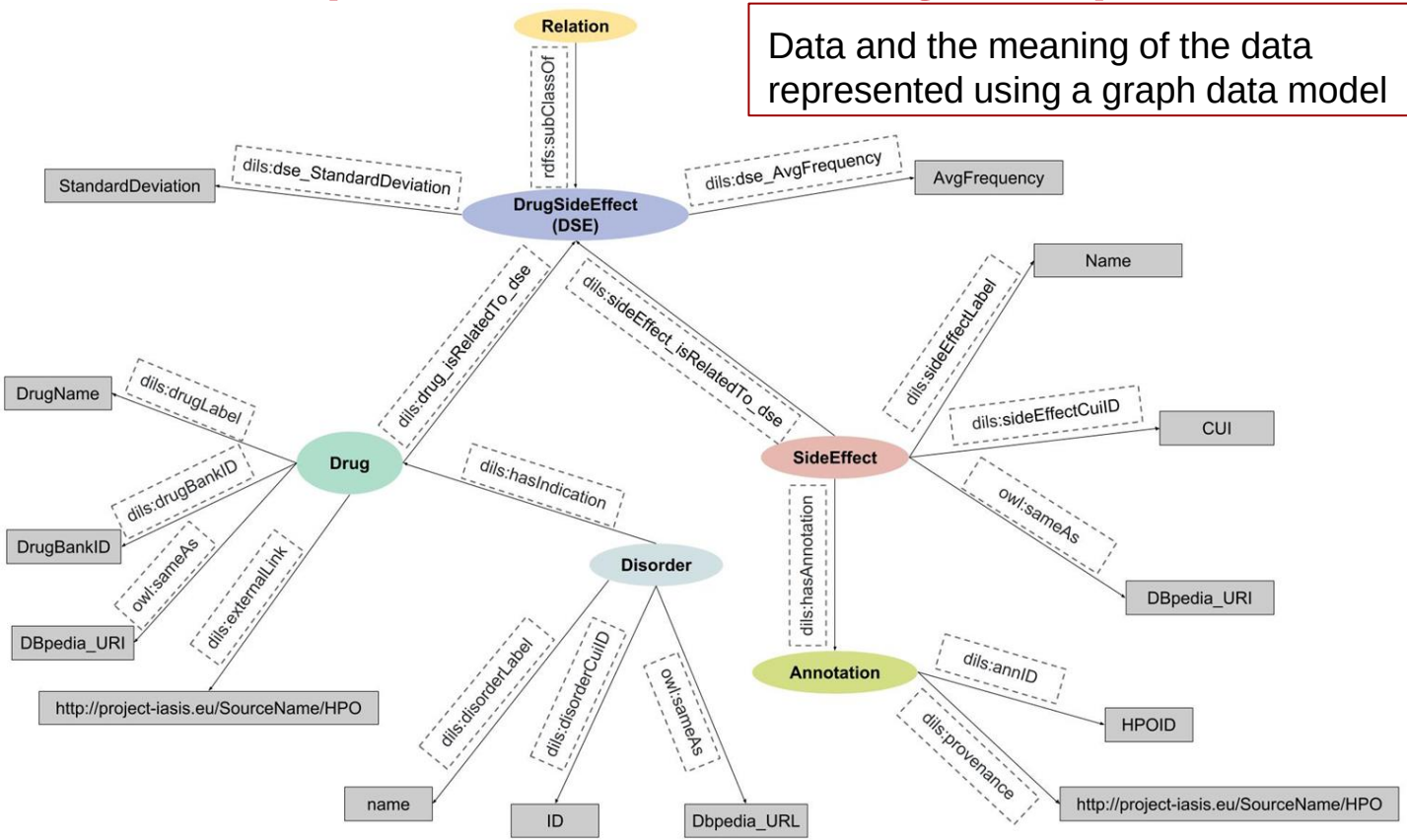
# What can we represent in a Knowledge Graph?

Data and the meaning of the data represented using a graph data model



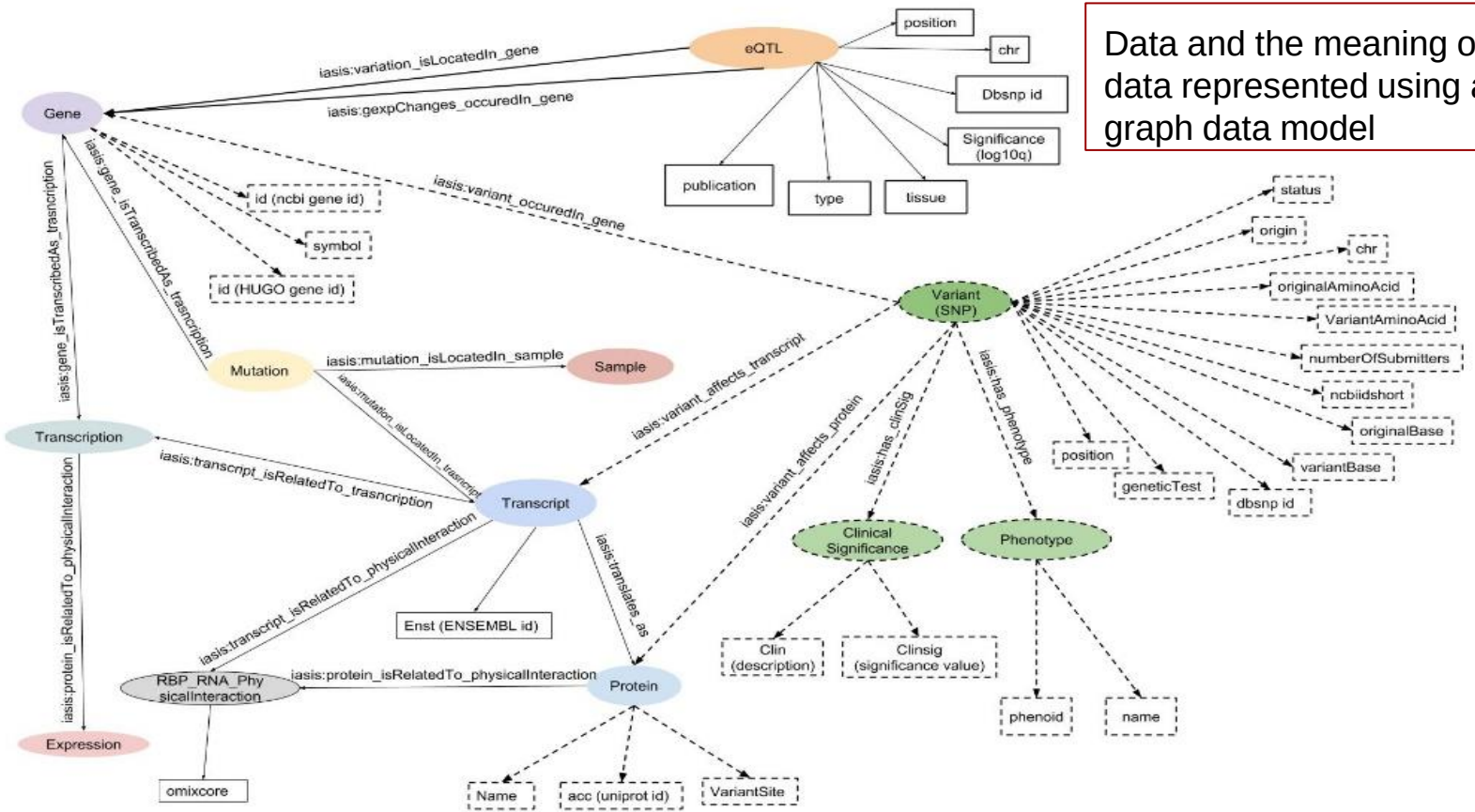
# What can we represent in a Knowledge Graph?

Data and the meaning of the data represented using a graph data model



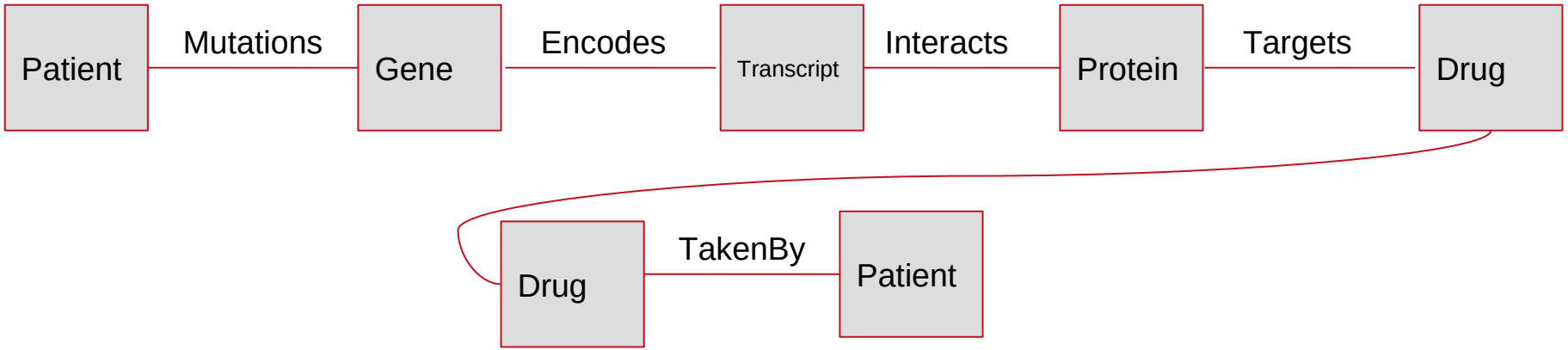
# What can we represent in a Knowledge?

Data and the meaning of the data represented using a graph data model



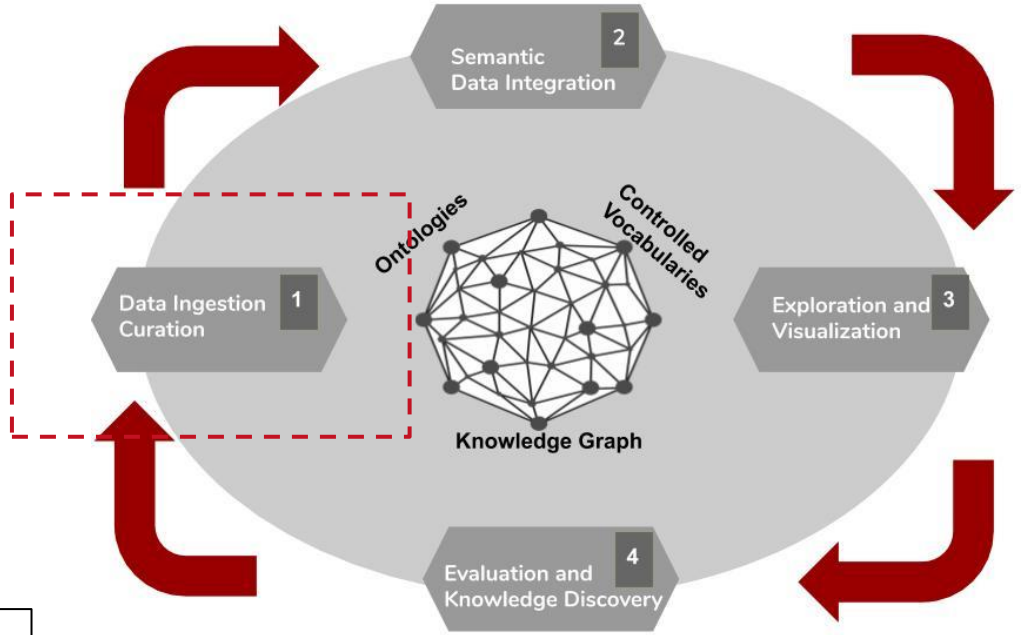
# Generation of Research Hypotheses

Factors that may impact on the effectiveness of an oncological treatment?



# Big Data Ingestion and Knowledge Extraction

# Knowledge-driven Approach




Extraction of knowledge encoded in data presented in different formats

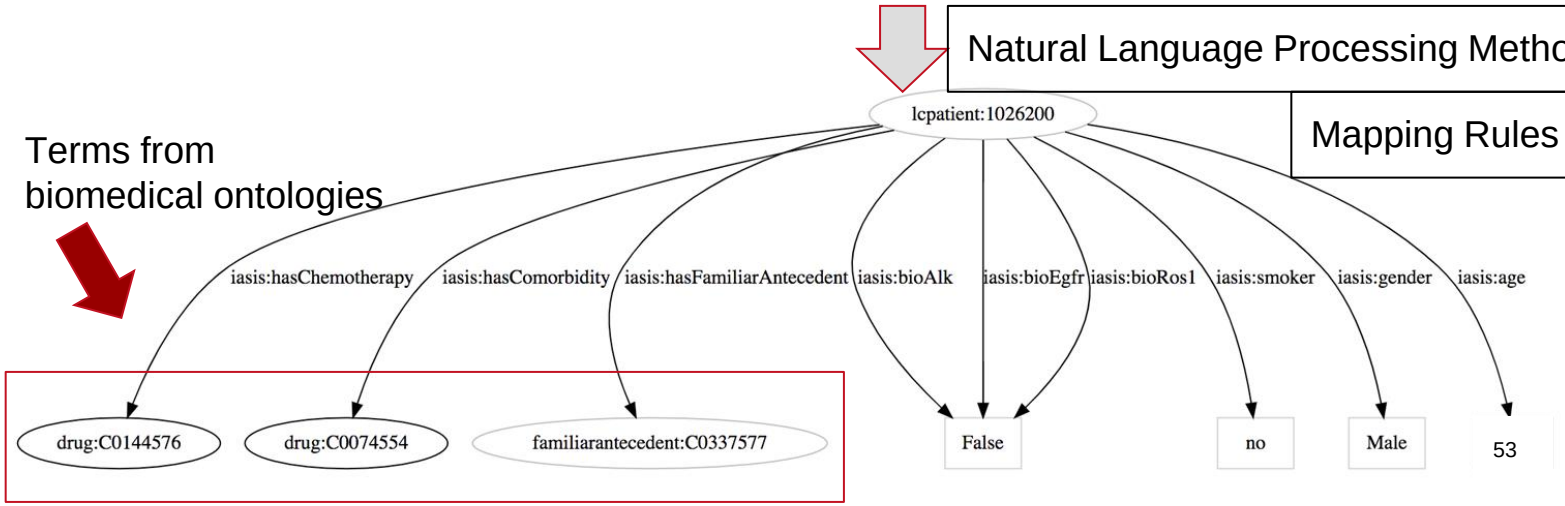
# Knowledge Extraction

*A **53-year** old **male** patient with Non Squamous carcinoma of stage III. He is a **non-smoker** and **non-positive for EGFR, ROS1, and ALK**. He suffers from **high cholesterol** and takes **Simvastatin**. **No relatives with cancer**. His **father, mother, and older brother died** from **heart attacks.** He is prescribed with **Paclitaxel***

Natural Language Processing Methods

Mapping Rules

Terms from biomedical ontologies



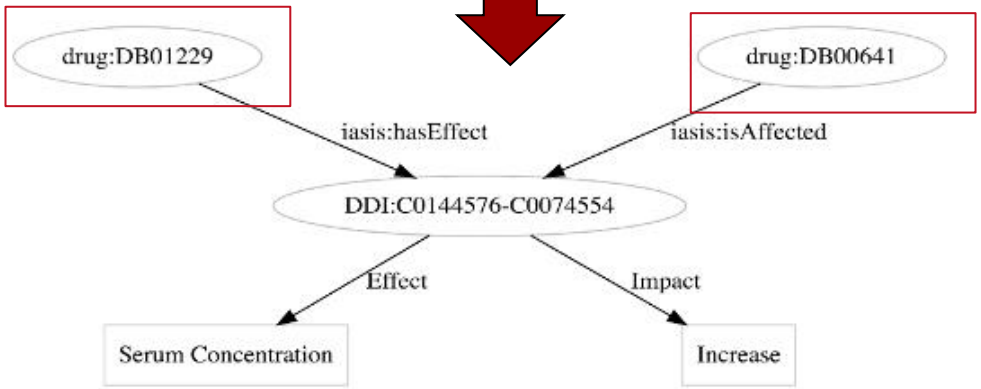
# Knowledge Extraction

| DrugBankId1 | DrugBankId2 | Interaction Description |
|---|---|---|
| DB01229 | DB00641 | **Serum concentration** of **Simvastatin** can be **increased** when combined with **Paclitaxel** |

Terms from controlled vocabularies

Natural Language Processing

Mapping Rules



# Entity Recognition and Linking

**Pacitaxel** is a well-established anti-mitotic chemotherapy medication that works by interfering with cell division. Docetaxel is approved by the FDA for treatment of locally advanced or **metastatic breast cancer**, **head and neck cancer**, **gastric cancer**, **hormone-refractory prostate cancer** and **non small-cell lung cancer**. Docetaxel can be used as a single agent or in combination with other chemotherapeutic drugs as indicated depending on specific cancer type and stage. (en)

C0024623

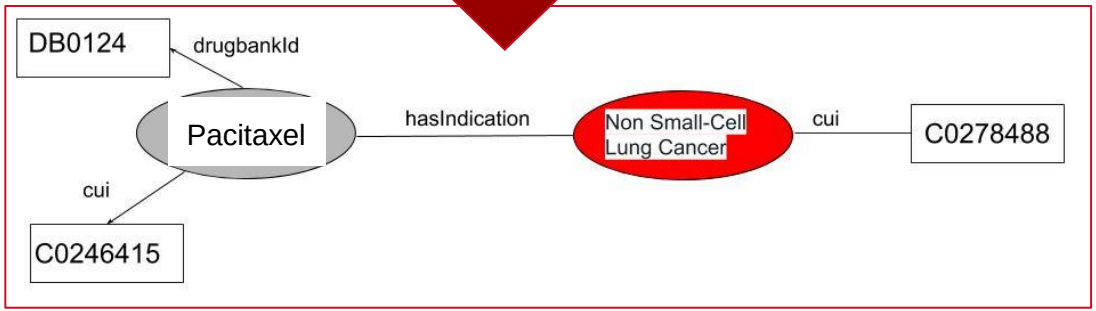
C0007131

C0278488

C0278488

C0278996

Indication of the disorders approved to be treated with Pacitaxel



# Entity and Prediction Linking Over KGs

1. **Knowledge graph extension**
   a. An **extended KG** is created based on the alignments of the labels in KGs, e.g., DBpedia and UMLS.
   b. Alignments correspond to ***owl:SameAs*** and equivalent associations in UMLS.
   c. Synonyms of relations are considered (e.g., UMLS)
2. **Entity and Relation extraction**
   a. Rules are used to extract **entities & relations** (e.g., drugs and interactions)

**3. Knowledge graph extension**

a. **Entity and Relation Linking**
b. Find the **smallest subset** of entities and relations
c. **Candidate entities** & **relations** are **ranked** based on the number of **matchings** in the KG.
d. **Candidate entities** with the **highest score** are selected and given as an output.

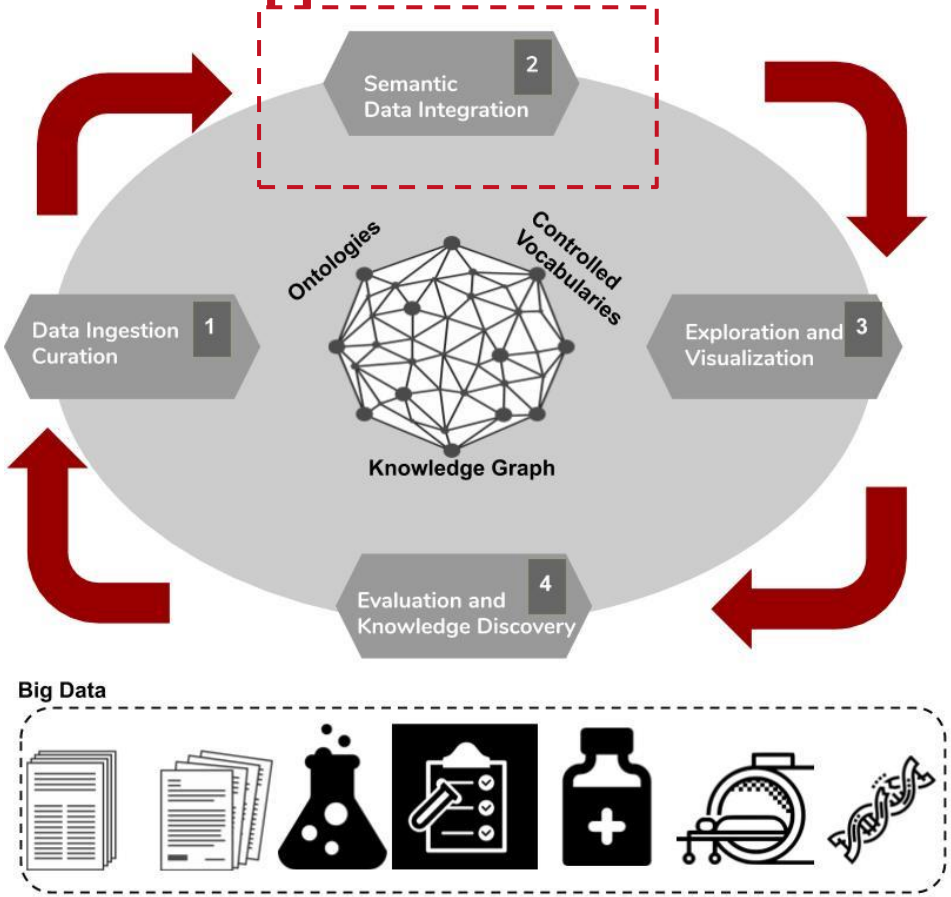
https://labs.tib.eu/falcon

Falcon

A. Sakor, I. O. Mulang', K. Singh, S. Shekarpour, M.-E. Vidal, J. Lehmann, and S. Auer. Old is gold: Linguistic driven approach for entity and relation linking of short text. In Proceedings of the NAACL HLT, 2019.

# Semantic Data Integration

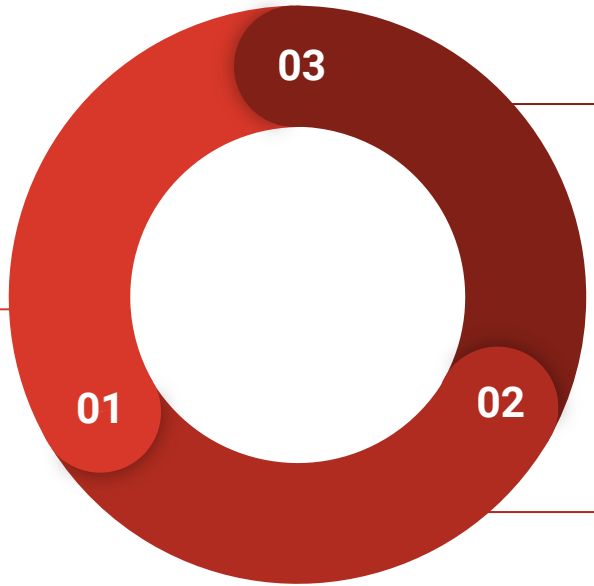
# Knowledge-driven Approach



# Steps for Semantic Data Integration

**Knowledge Extraction**

NLP Techniques extract relevant concepts and annotate them using existing vocabularies.




**Semantic Data Integration**

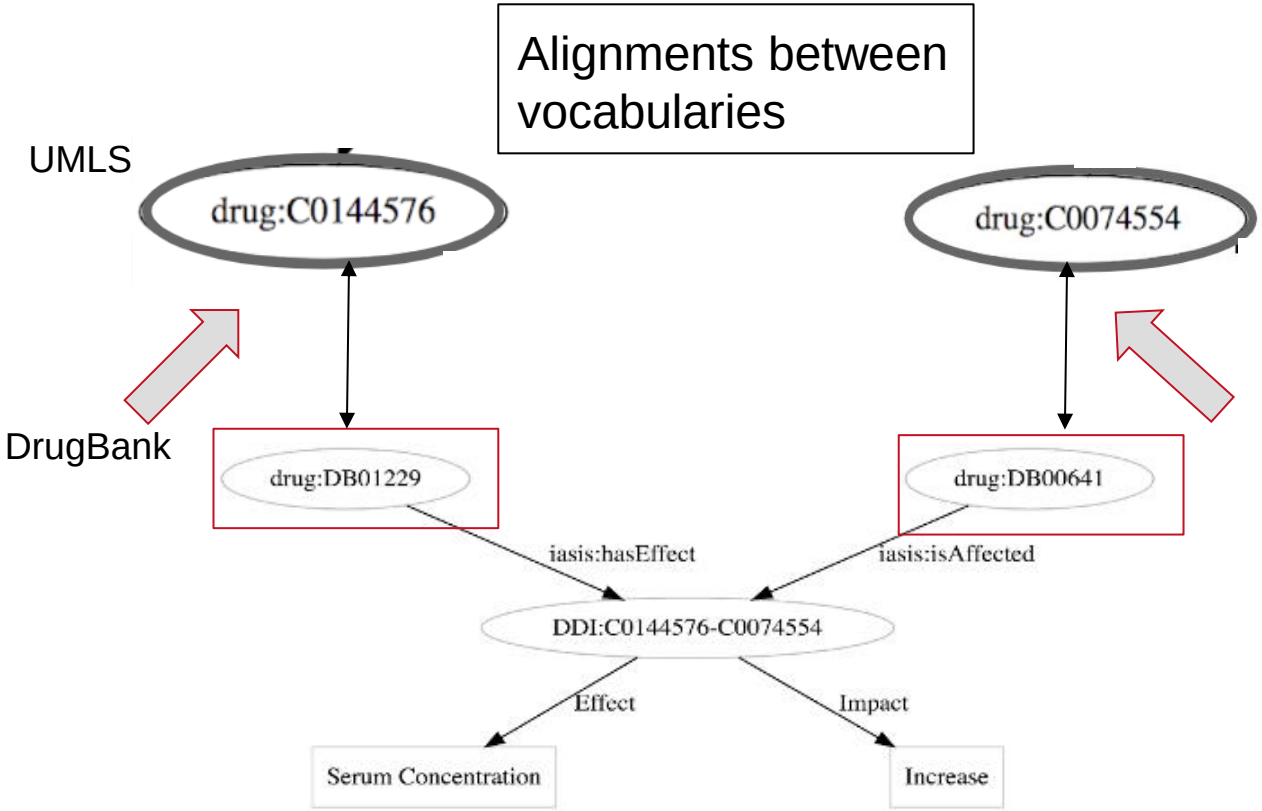
Alignments among annotations of the concepts are used for entity matching and data integration

**Alignments among vocabularies**

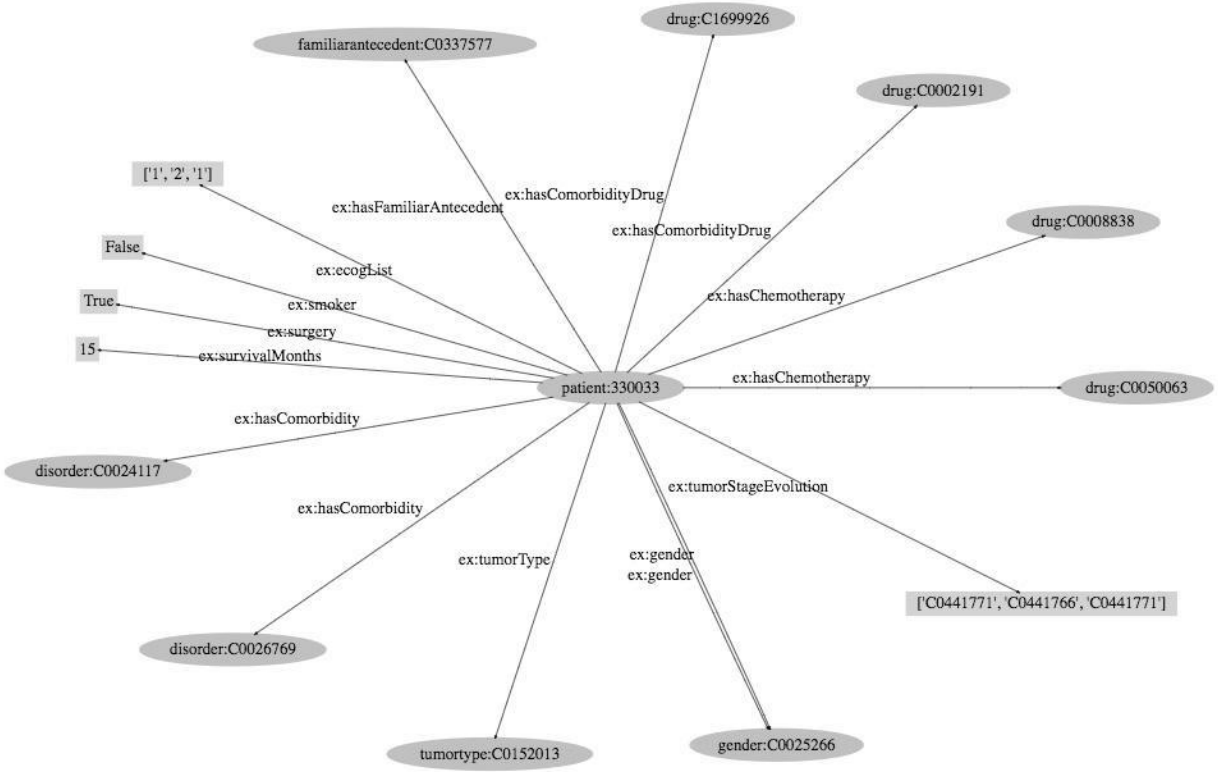
Concepts from different vocabularies are aligned

Maria-Esther Vidal, Kemele M. Endris, Samaneh Jozashoori, Ahmad Sakor, Ariam Rivas. Transforming Heterogeneous Data into Knowledge for Personalized Treatments- A Use Case. Journal Datenbank-Spektrum. 2019

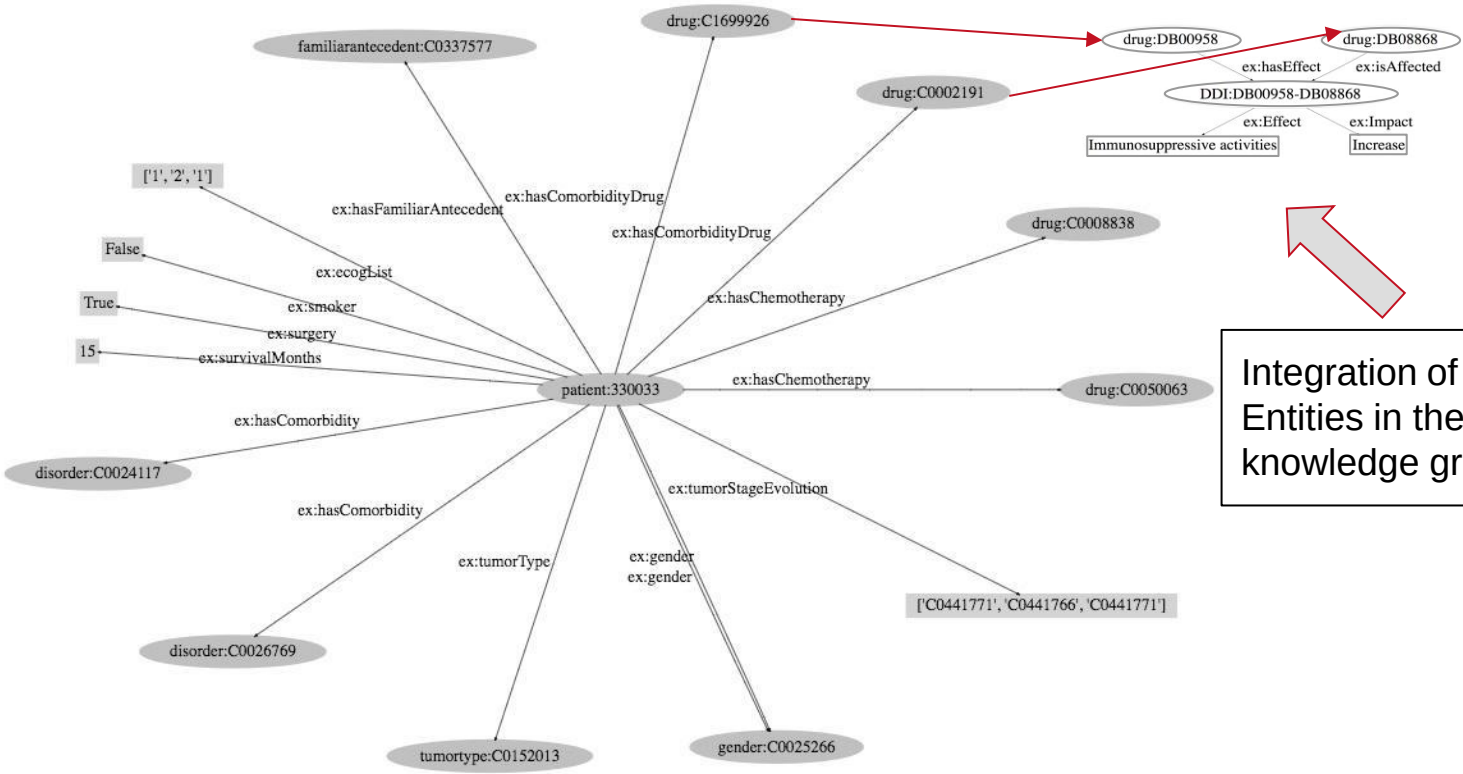
# Semantic Data Integration



# Semantic Data Integration



# Semantic Data Integration

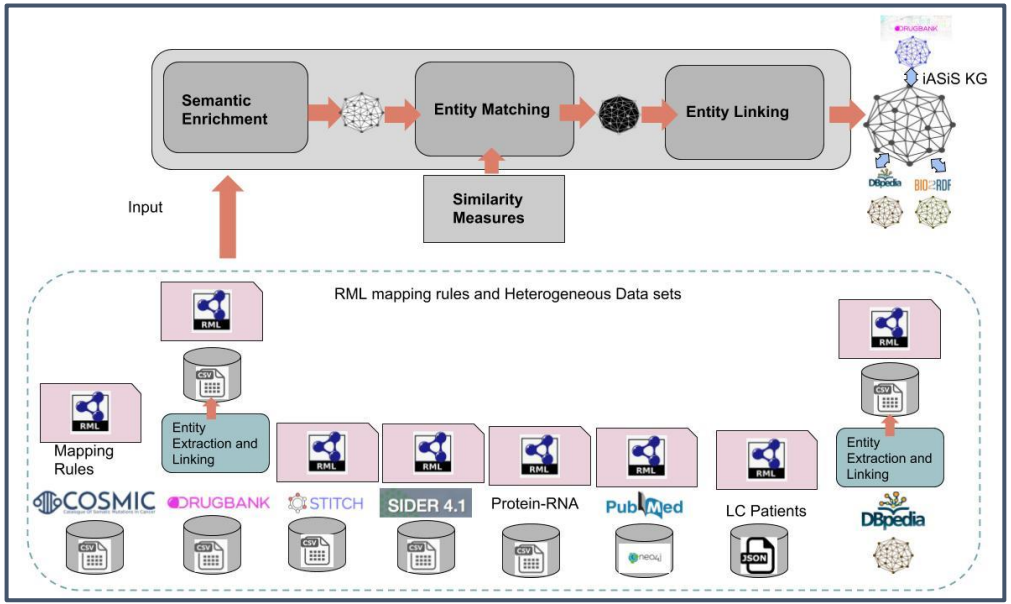



Integration of Entities in the knowledge graph

# Knowledge Graph Creation Pipeline

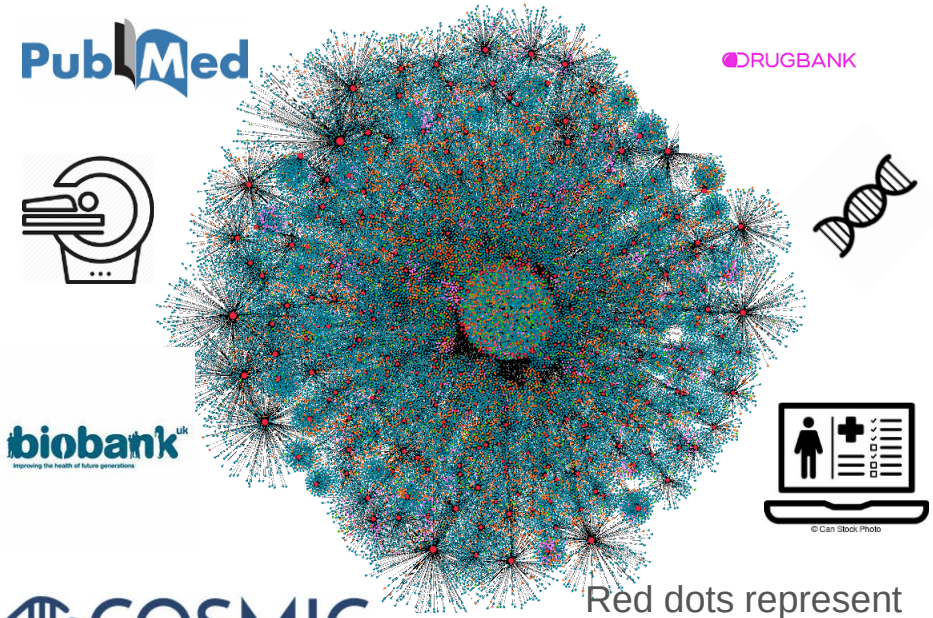
The **Knowledge Graph** creation pipeline:

- Datasets are in **various formats**
- **Entity Extraction** and **Linking** allow for the transformation of short text into **CUI terms**
- RDF Mapping Language (RML) **mapping rules** represent **transformations** from data sets into the iASiS KG
- **Entity Matching** resorts to **similarity measures** to match **equivalent entities** in the iASiS KG
- **Entity Linking** enables the linking of entities in the iASiS KG with **entities in other knowledge graphs**, e.g., DBpedia, Bio2RDF



# The iASiS Knowledge Graph

Total number of Triples: **1.207.930.891**
Data integrated from 35 data sources




Red dots represent patients

Each Patient **is associated** with **3,5** Drugs (average)
- 823 **patients are linked** to at least **1** Drug
- Maximum **15** Drugs

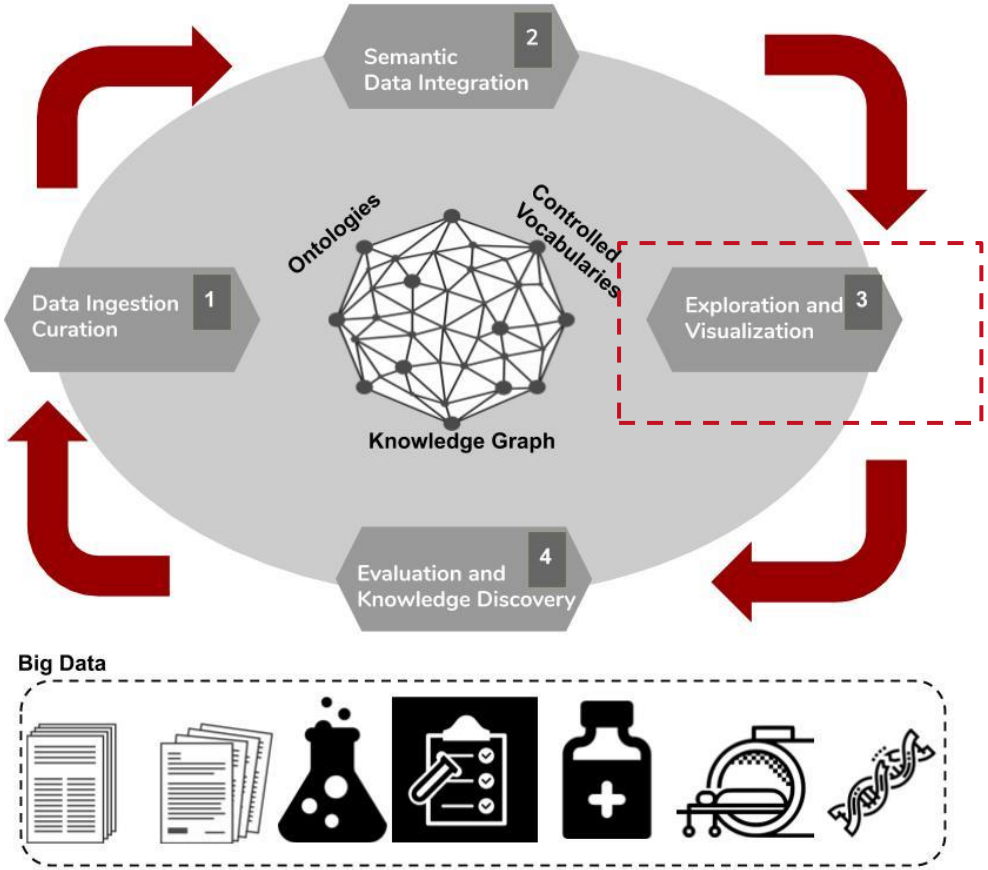
Each Patient **is associated** with **16** Drug Side Effects (average)
- 677 **patients are linked** to at least **3 Drug Side Effect**
- Maximum **68** Drug Side Effects

Each Patient **is associated** with **199** Publications (average)
- 809 **patients are linked** to at least **1** Publication
- Maximum **1.759** Publications

Each Patient **has 7** Toxicities (average)
- 738 **patients are linked** to at least **1** Toxicity
- Maximum **17** Toxicities

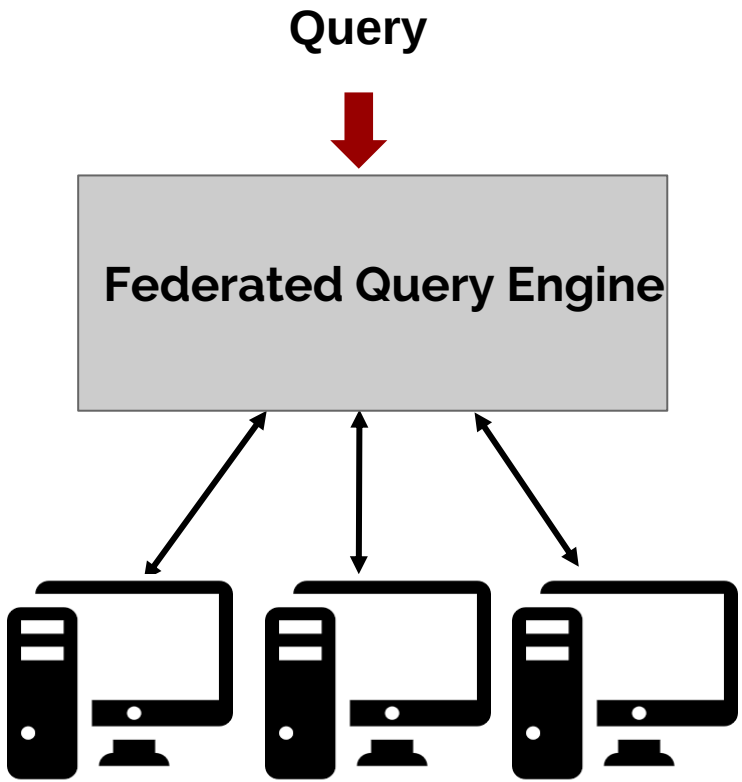
# Federated Query Engine

# Knowledge-driven Approach

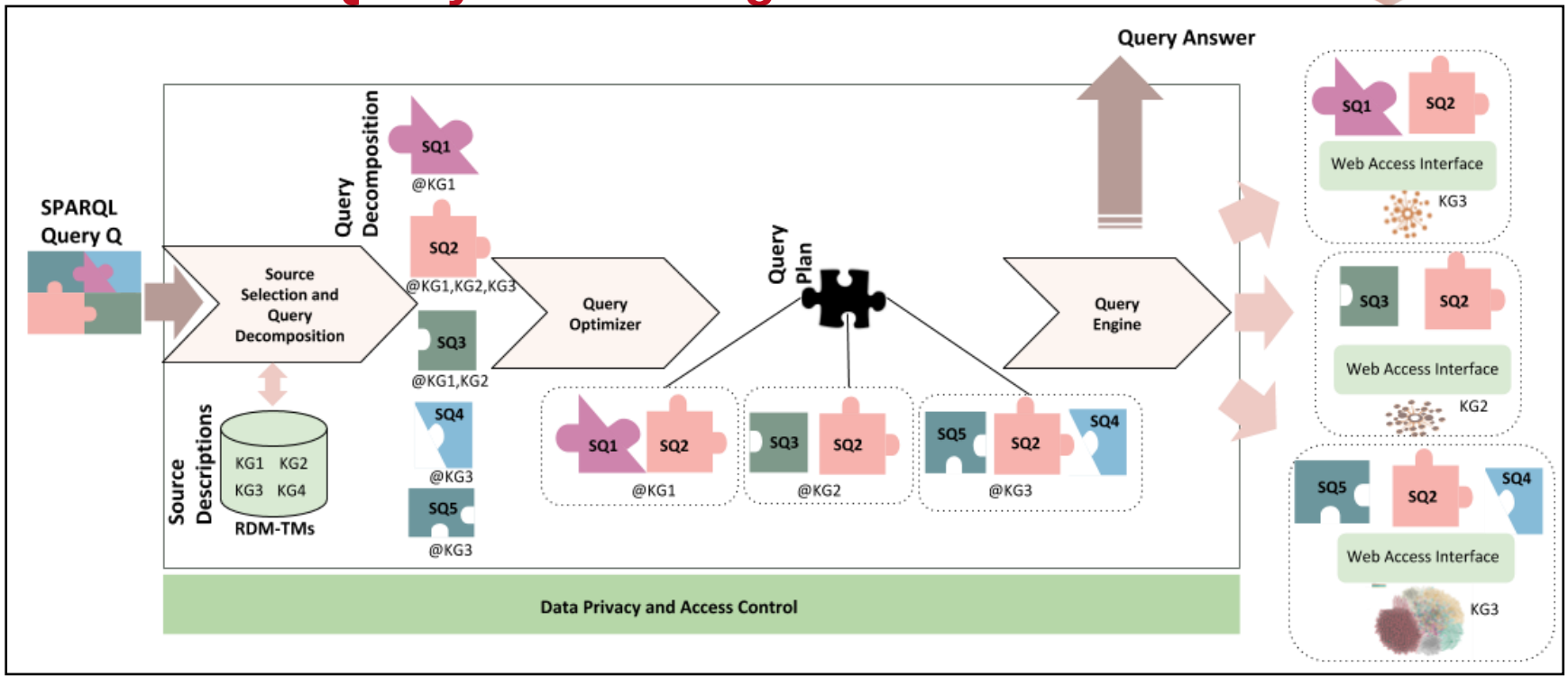


# Federated Query Processing

- SPARQL queries are **posed against** the federated engine
- The federated engine exports a **uniform interface** for query processing
- Input queries are **decomposed** into queries over SPARQL endpoints in the federation
- Data collected from the **evaluation of the subqueries** is merged to build the answer of the input query



# Federated Query Processing



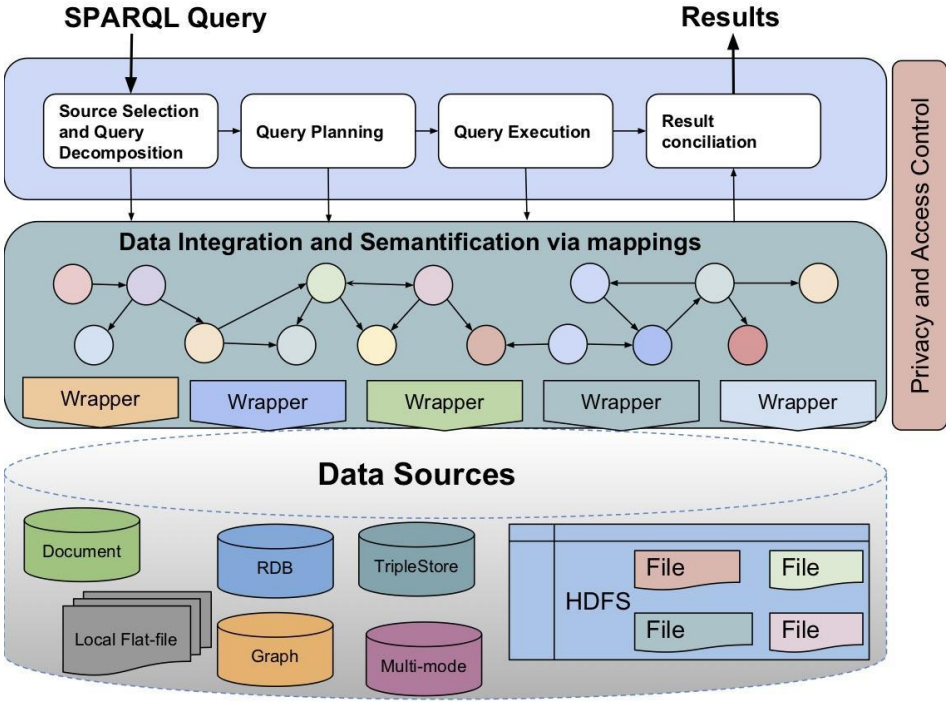

K. Endris, M. Galkin, I. Lytra, M. Nadjib Mami, M-E Vidal, S. Auer: Querying Interlinked Data by Bridging RDF Molecule Templates. T. Large-Scale Data- and Knowledge-Centered Systems 39: 1-42, 2018

K. Endris, Z. Almhithawi, I. Lytra, M-E Vidal, S. Auer: BOUNCER: Privacy-Aware Query Processing over Federations of RDF Datasets. DEXA 2018

# Federated Query Processing

Ontario is a **federated query engine**:

- Integrates a **federation** of RDF knowledge graphs and data lakes
- Provides a **unified interface** to query a federation using SPARQL
- Implements **source selection**, **query decomposition**, and **query planning**
- Implements **join operators** to merge results from **subqueries**




K. Endris, P. Rohde, M-E Vidal, S. Auer: Ontario: Federated Query Processing against a Semantic Data Lake. DEXA 2019.

# Federated Query Processing

*"**Mutations** of the type **Confirmed somatic variant** located in transcripts which are translated as proteins that are transporters of the drug **Pacitaxel**"*

**Q:**

```
PREFIX rdf:      <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX iasis:    <http://project-iasis.eu/vocab/>
PREFIX drugbank: <http://bio2rdf.org/drugbank_vocabulary:>
PREFIX owl:       <http://www.w3.org/2002/07/owl#>

SELECT DISTINCT ?mutation
WHERE {
          ?mutation    rdf:type                                iasis:Mutation .
          ?mutation    iasis:mutation_somatic_status           'Confirmed somatic variant'.
          ?mutation    iasis:mutation_isLocatedIn_transcript   ?transcript .
          ?transcript  iasis:translates_as                     ?protein .
          ?drug        iasis:drug_interactsWith_protein        ?protein .
          ?protein     iasis:label                             ?proteinName .
          ?drug        iasis:label                             'pacitaxel' .
          ?drug        owl:sameAs                              ?drug1 .
          ?drug1       drugbank:transporter                    ?transporter .
          ?transporter drugbank:gene-name                      ?proteinName .
        }
```

# Federated Query Processing

**SQ1:**

```
PREFIX rdf:      <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX iasis:    <http://project-iasis.eu/vocab/>
PREFIX owl:      <http://www.w3.org/2002/07/owl#>
SELECT DISTINCT ?mutation ?proteinName  ?drug1
WHERE {
     ?mutation    rdf:type                              iasis:Mutation .
     ?mutation    iasis:mutation_somatic_status         'Confirmed somatic variant'.
     ?mutation    iasis:mutation_isLocatedIn_transcript ?transcript .
     ?transcript  iasis:translates_as                   ?protein .
     ?drug        iasis:drug_interactsWith_protein      ?protein .
     ?protein     iasis:label                           ?proteinName .
     ?drug        iasis:label                           'pacitaxel' .
     ?drug        owl:sameAs                            ?drug1 .
  }
```

**SQ2:**

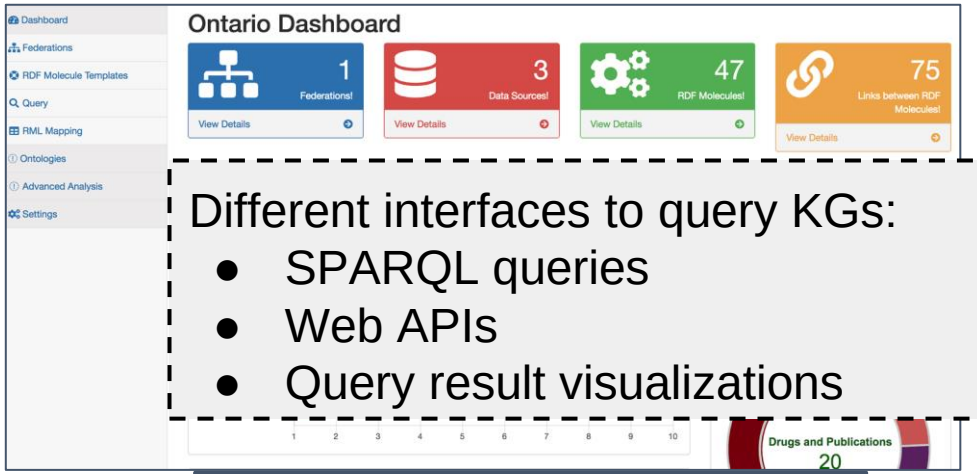
```
PREFIX drugbank: <http://bio2rdf.org/drugbank_vocabulary:>
SELECT DISTINCT ?proteinName  ?drug1
WHERE {
      ?drug1         drugbank:transporter     ?transporter .
      ?transporter   drugbank:gene-name       ?proteinName .
 }
```

**24 mutations** of the protein ABCB1

**11 mutations** of the protein ABCG2

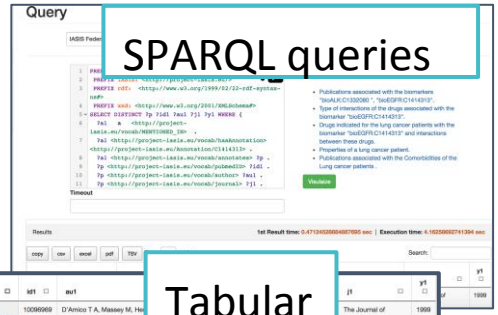
**FLNC** associated with Cardiomyopathy

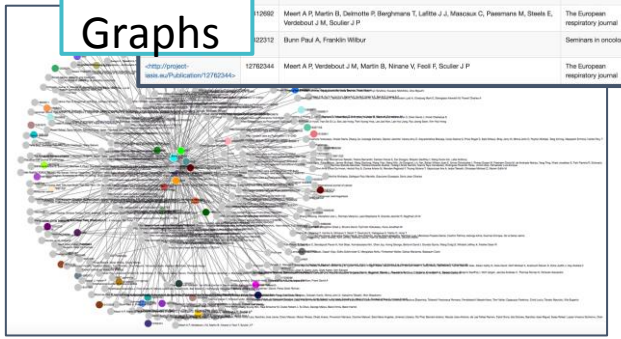
# Exploring Knowledge Graphs

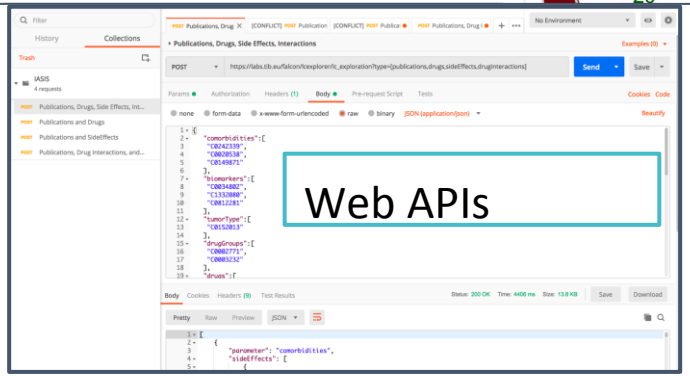



Different interfaces to query KGs:

- SPARQL queries
- Web APIs
- Query result visualizations



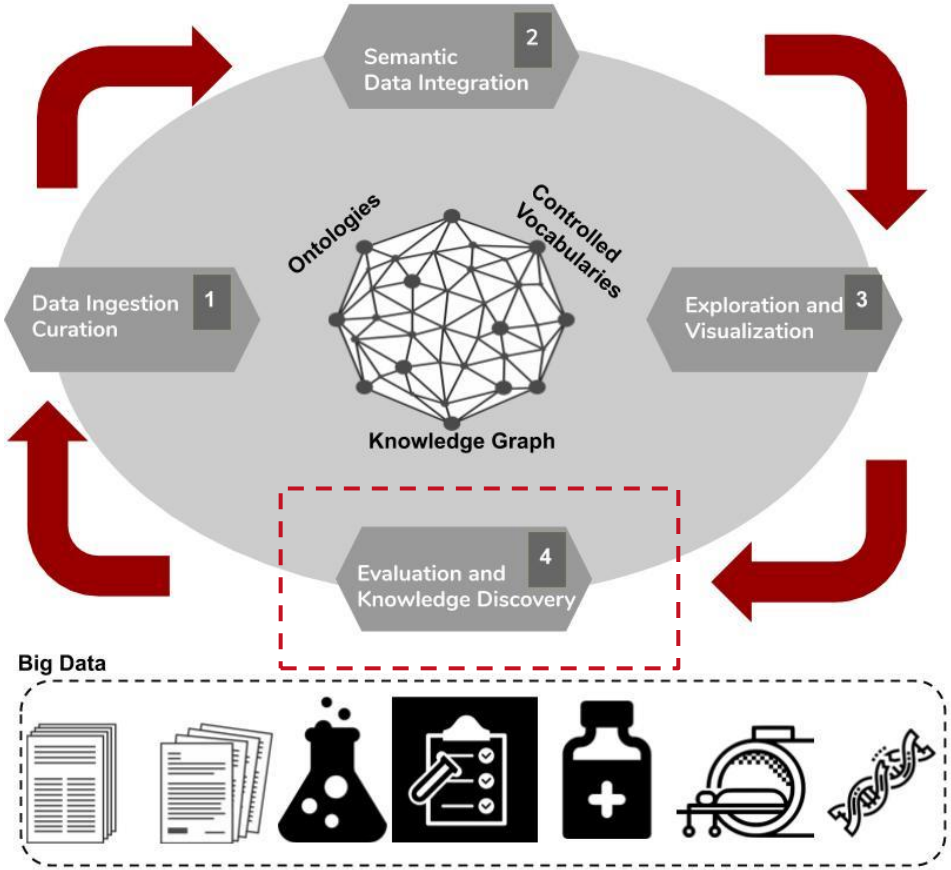


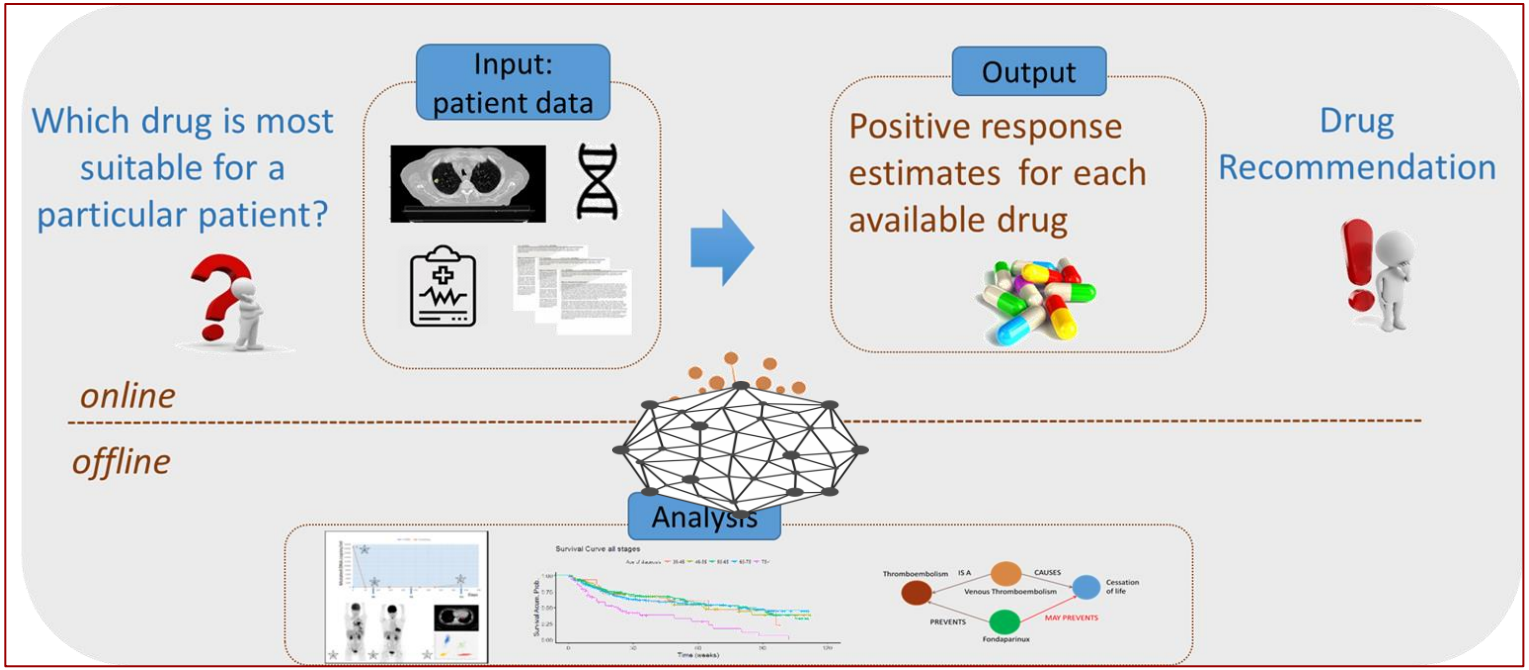

https://labs.tib.eu/ontario/dashboard/

# Knowledge Discovery

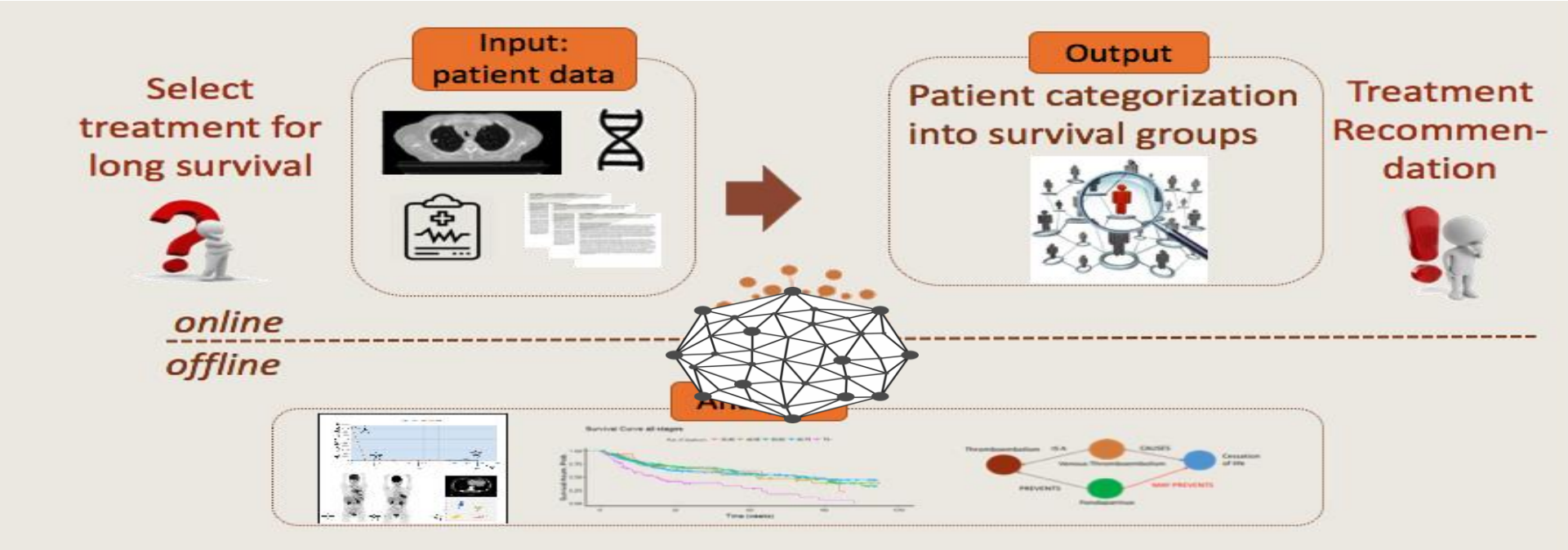
# Knowledge-driven Approach



# Lung Cancer Treatment Schema



# Lung Cancer Treatment Schema



# Prediction Methods

**Classification Task:**

Determine the category of lung cancer patients based on class labels.

- **Class label:** ordinal encoding of survival months of patients by applying quantile to have equal number of patients in each bin.
- **Evaluation Metrics:** Mean accuracy, recall, precision, f1-score

**Goal**: Classify lung cancer patients according to survival time (months)

**Machine Learning Methods (supervised learning):**

- **Random Forest Classifier**: various decision trees based on different sub sampling of dataset using ‘gini’ criterion and averaging them for prediction
- **K-Neighbors Classifier**: determine the label of query point based on k nearest nearest neighbours where k is integer value.
- **Linear Support Vector Classifier**: determine class of a query point by utilizing linear support vector machine
- **SGD Classifier**: linear classifier support vector machine with stochastic gradient descent

# Prediction Methods-Survival Time

Random Forest Classifier

80% for training and 20% for testing

K -Neighbor Classifier

| Class | Precision | Recall | f1-Score |
|---|---|---|---|
| [0.0, 6.0) | **0.45** | **0.59** | **0.51** |
| [6.0, 14.0) | 0.35 | 0.35 | 0.35 |
| [14.0, 30.0) | 0.30 | 0.21 | 0.25 |
| [30.0, 263.0) | 0.45 | 0.44 | 0.44 |

| Class | Precision | Recall | f1-Score |
|---|---|---|---|
| [0.0, 6.0) | **0.36** | **0.74** | **0.48** |
| [6.0, 14.0) | 0.30 | 0.28 | 0.29 |
| [14.0, 30.0) | 0.25 | 0.16 | 0.20 |
| [30.0, 263.0) | 0.48 | 0.21 | 0.29 |

The **highest accuracy** is achieved in class [0.0, 06)

# Prediction Methods-Survival Time

Linear Support Vector Classifier

80% for training and 20% for testing

SGD Classifier

| Class | Precision | Recall | f1-Score |
|---|---|---|---|
| [0.0, 6.0) | **0.45** | **0.59** | **0.51** |
| [6.0, 14.0) | 0.35 | 0.35 | 0.35 |
| [14.0, 30.0) | 0.30 | 0.21 | 0.25 |
| [30.0, 263.0) | 0.45 | 0.44 | 0.44 |

| Class | Precision | Recall | f1-Score |
|---|---|---|---|
| [0.0, 6.0) | **0.36** | **0.74** | **0.48** |
| [6.0, 14.0) | 0.30 | 0.28 | 0.29 |
| [14.0, 30.0) | 0.25 | 0.16 | 0.20 |
| [30.0, 263.0) | 0.48 | 0.21 | 0.29 |

The **highest accuracy** is achieved in class [0.0, 06)

# Applications

iASiS
Big Data for Precision Medicine

# H2020 Research and Innovation Action

TIB

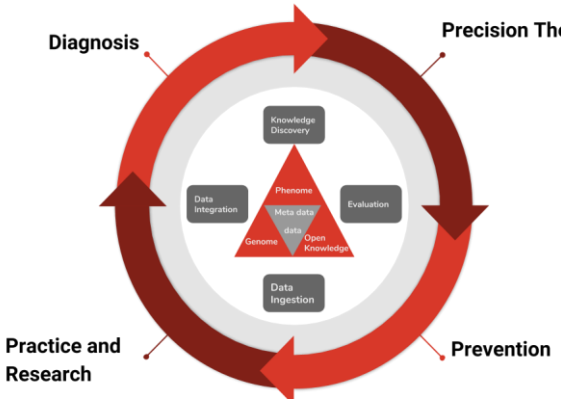
**http://project-iasis.eu/**

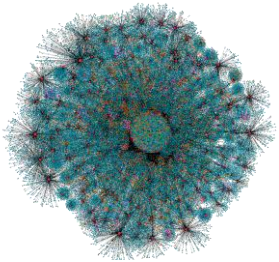
Coordinator: NCRS Demokritos, Greece (George Paliouras)

Project Period: 04/2017-04/2020

Project Vision:

- Turn **clinical and pharcogenomics** Big Data into **knowledge**
- Exploit **knowledge** for the discovery of **unknown** patterns
- Support **precision medicine** and health **policy making**



Knowledge Graph

# H2020 Big Data PPP: Large Scale Pilot actions in sectors best benefitting from data-driven innovation

**Coordinator:** Philips Electronics Nederland B.V.
(Supriyo Chatterjea)

**Project Period**: 03/2018- 03/2021

**35 Partners** including hospitals, companies, universities, and research centers

**Project Vision**:

- Application of **Big Data technologies** in the Healthcare sector to deliver **low cost** and **high quality care**:
  - Increase **healthcare productivity** (20%-60%)
  - Increase **market share** of **Big Data** providers on **Oncology**, **Cardiology**, **Radiology**, and **Hospital Logistics** (25%)

https://www.bigmedilytics.eu/

# H2020 Big Data PPP: Large Scale Pilot actions in sectors best benefitting from data-driven innovation

Coordinator: PHILIPS

Project Period: 03/2018-03/2021

**Theme 1**: Population Health & Chronic Diseases

**Pilot 1**: Comorbidities
**Pilot 2**: Kidney Disease
**Pilot 3**: Diabetes
**Pilot 4**: Asthma
**Pilot 5**: Heart Failure

**Theme 2**: Oncology

**Pilot 6**: Prostate Cancer
**Pilot 7**: Lung Cancer
**Pilot 8**: Breast Cancer

**Theme 3**: Industrialization and Healthcare Services

**Pilot 9**: Stroke Workflows
**Pilot 10**: Sepsis workflows
**Pilot 11**: Asset management workflows
**Pilot 12**: Radiology workflows

**The Health Continuum**

https://www.bigmedilytics.eu/

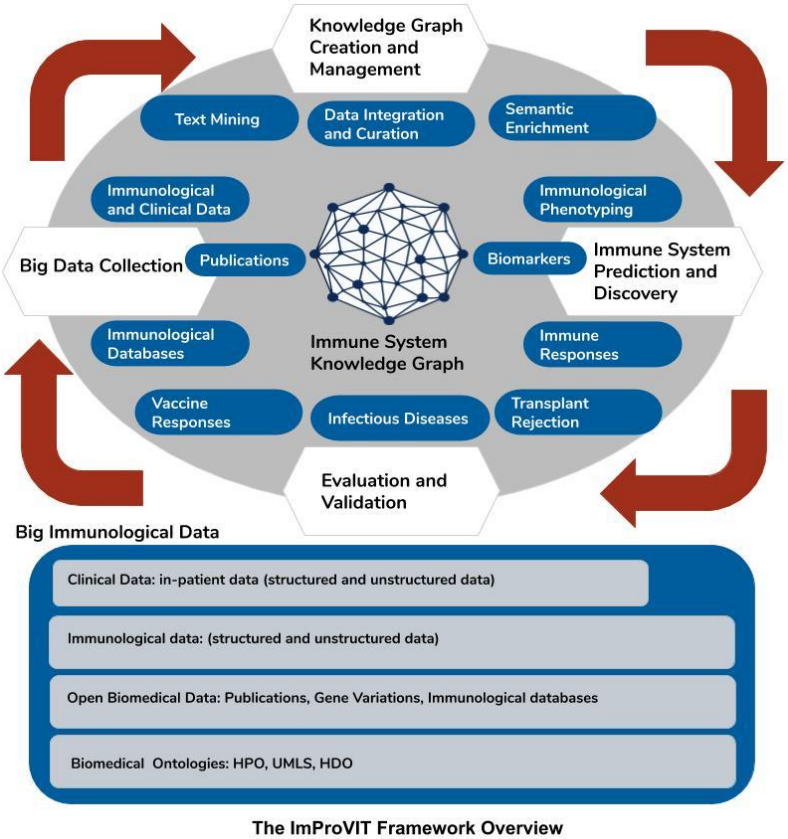
# ImProVIT: Transforming big data into knowledge: for deep immunoprofiling in vaccination, infectious diseases, and transplantation

**Coordinator**: TWINCORE
**Project Period**: 01/2020-12/2022
**Partners**: TIB and Medical School Hannover
**Call**: *Big Data in den Lebenswissenschaften – Chancen datenintensiver Forschung und personalisierter Medizin*

Björn Thümler Niedersächsischer Minister für Wissenschaft und Kultur

Project Goal

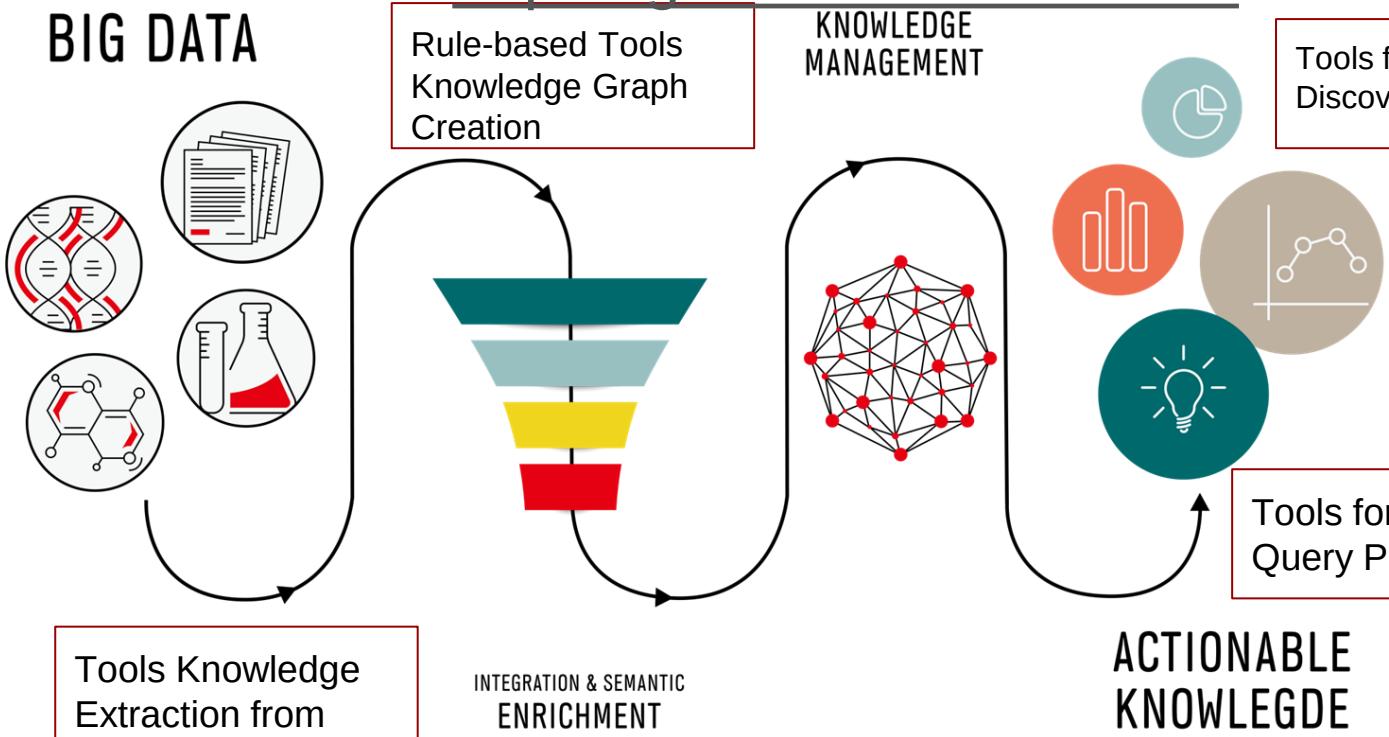
- Holistic understanding of the **human immune system** under conditions of health and disease
- Evaluation of **Vaccine Response**, **Infectious Diseases**, and T**ransplant Rejection**



The ImProVIT Framework Overview

# Results so far….

https://github.com/SDM-TIB




https://iasisdev.atc.gr/new-pattern

# Grand Challenges and Opportunities

# Gaps and Needs in Big Data and Knowledge

| Needs | Research Challenges |
|---|---|
| **Consuming** Big Data | **Ingesting** data from Big Data sources<br>**Describing** quality-aware data acquisition techniques from Big Data sources |
| **Extracting** Knowledge | **Mining** knowledge from unstructured data sources<br>**Describing** extracted knowledge |
| **Representing and Managing** High-Quality and Actionable Knowledge | Designing **storage techniques** to scale up to large and heterogeneous data<br>Devising **quality-aware knowledge extraction** from Big Data<br>Describing Knowledge in **symbolic and subsymbolic** models<br>Integrating **quality management processes** into the Big Data value chain |

# Gaps and Needs in Big Data and Knowledge

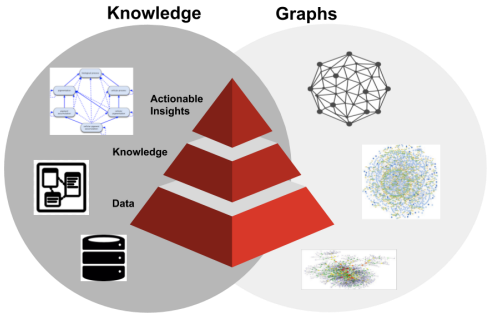
| Needs | Research Challenges |
|---|---|
| **Processing** Big Knowledge | Defining **languages and models** to manage Knowledge Graphs<br>Devising **querying techniques to scale up** to large Knowledge Graphs<br>Executing **queries over decentralized or federated** Knowledge Graphs |
| **Understanding** Knowledge Graphs | Unravelling **hidden patterns** or associations in Knowledge Graphs<br>**Creating value** from Knowledge Graphs<br>Providing **scalable visualization** of Knowledge Graphs |

# Gaps and Needs in Big Data and Knowledge

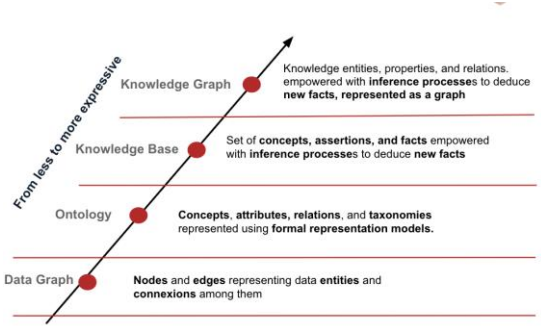
| Needs | Research Challenges |
| --- | --- |
| **Ensuring** Privacy Regulations | Modeling **privacy-aware access** control<br>Implementing **privacy-preserving operations**<br>Enabling **traceability through audits** at different granularities |

# Conclusions

**Knowledge graphs** represent actionable knowledge, data facts, and information and connections among them. Graphs are utilized as formal structures of representation.



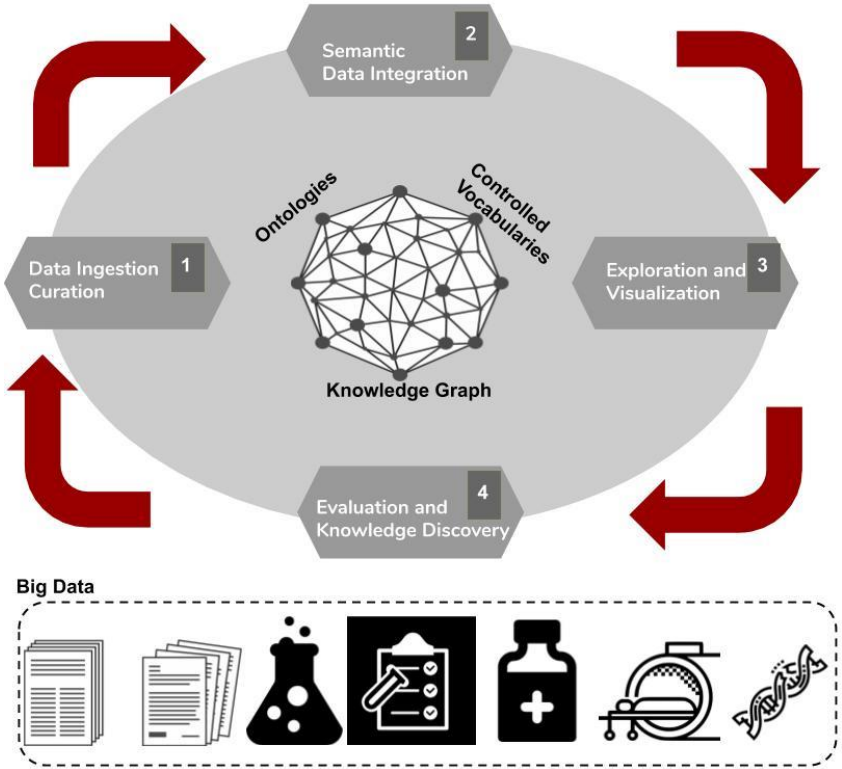

The **knowledge graph spectrum** ranges from data graphs, ontologies, knowledge bases, and knowledge graphs, allowing for **expressive representations** of knowledge



# Conclusions

**Knowledge-driven** pipeline allow for

- **Integration and semantic** description of Big Data
- **Extraction of knowledge** from unstructured data
- **Exploration of knowledge** and **unknown** connections
- Discovery of **patterns and prediction** of unknown relations



# Conclusions

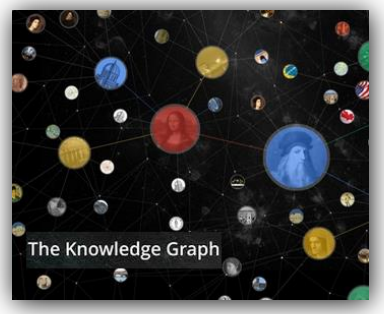
**Knowledge graphs** in real-world applications:

- Google Knowledge Graph
- AirBnB
- ORKG at TIB

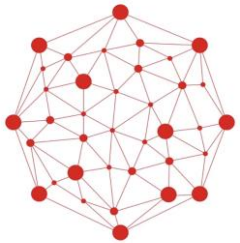
**Knowledge graphs** allow for

- **Describing entities** and a large **number of connections** among them
- **Traversing** and **exploring** entities and their connections
- Discovery of **patterns and prediction** of unknown connections

https://developers.google.com/knowledge-graph/




https://medium.com/airbnb-engineering/contextualizing-airbnb-by-building-knowledge-graph-b7077e268d5a



https://projects.tib.eu/orkg/

Leibniz
Universität
Hannover
LEIBNIZ INFORMATION CENTRE
FOR SCIENCE AND TECHNOLOGY
UNIVERSITY LIBRARY
TIB

# Our Team at the Scientific Data Management Group

**Scientific Data Management Group**

Prof, Dr. Maria-Esther Vidal

## PostDoc

Dr. Ingo Keck

## Senior Researcher

Akhilesh Vyas

## Research Assistants

Samaneh Jozashoor

Ariam Rivas

Maria Isabel Castellanos

Kemele Endris

Farah Karim

Ahmad Sakor

## Visiting Researchers

Lucie-Aimée Kaffee

David Chaves

## Master Students

Enrique Iglesias

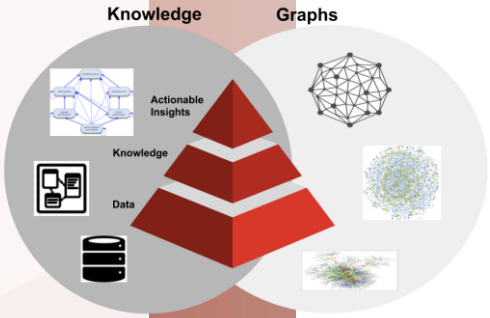
Monica Figuera

Philipp Rohde

LEIBNIZ INFORMATION CENTRE
FOR SCIENCE AND TECHNOLOGY
UNIVERSITY LIBRARY

TIB

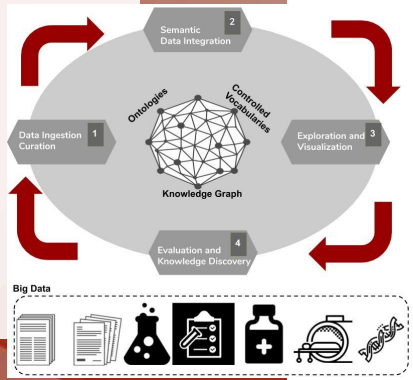
# MANY THANKS




# Questions?

**Contact:**

Maria-Esther Vidal
Maria.Vidal@tib.eu




Creative Commons Attribution 3.0 Germany
https://creativecommons.org/licenses/by/3.0/de/deed.en

# Entity and Prediction Linking Over KGs

**Benchmark**: DrugBank -- description of interactions between drugs; 348,689

**Gold Standard:** Annotations done by 13 participants

**KGs**: UMLS and DBpedia

| | |
|---|---|
| Number of annotations | 2406 |
| Number of patterns | 44 |
| Participants | 13 |
| Unique number of interactions | 1203 |
| Annotated as correct | 2392 |
| Annotated as wrong | 14 |
| Correct patterns | 41 |
| Wrong patterns | 3 |
| Number of disagreement | 4 |
| Accuracy | **99.41** |