



Biomedical Ontologies

Ronald Cornet
Amsterdam UMC



[doi:10.5281/zenodo.14643760](https://doi.org/10.5281/zenodo.14643760)



Funded by
the European Union

Ontologies – some examples

- ORDO – Orphanet Rare Disease Ontology
- HPO – Human Phenotype Ontology
- BFO – Basic Formal Ontology
- GO – Gene Ontology
- SNOMED CT
- A.k.a. nomenclatures, vocabularies, ...

Statistics	
Ontologies	1,176
Classes	15,253,669
Properties	36,286
Mappings	101,110,817

<https://bioportal.bioontology.org/>

- Dependent on use case, often in combination

Ontologies – purposes

- Encode
- Describe
- Order

Ontologies – Encoding

- *Mr. John Doe; Oct 1st, 2001; RR 120/80; BMI 29; OI.*

Ontologies – Encoding

• *Mr. John Doe; Oct 1st, 2001; RR 120/80; BMI 29; OI.*

- Name: “John Doe”
- Gender: Male identity
- Date of birth: Oct 1st, 2001
- Blood pressure: 120/80 mmHg
- Body Mass Index: 29
- Diagnosis: Osteogenesis Imperfecta

Ontologies – Encoding

• *Mr. John Doe; Oct 1st, 2001; RR 120/80; BMI 29; OI.*

- **Name:** “John Doe”
- **Gender:** **Male identity**
- **Date of birth:** Oct 1st, 2001
- **Blood pressure:** 120/80 **mmHg**
- **Body Mass Index:** 29
- **Diagnosis:** **Osteogenesis Imperfecta**

Ontologies – Encoding

- *Mr. John Doe; Oct 1st, 2001; RR 120/80; BMI 29; OI.*

LOINC:54125-0 ← **Name:**

LOINC:76691-5 ← **Gender:**

LOINC:21112-8 ← **Date of birth:**

LOINC:35094-2 ← **Blood pressure:**

LOINC:39156-5 ← **Body Mass Index:**

LOINC:29308-4 ← **Diagnosis:**

SNOMED:446151000124109

“John Doe”

Male identity

Oct 1st, 2001

120/80 mmHg → UCUM:mm[Hg]

29

Osteogenesis Imperfecta

ORPHA:666

Ontologies – Encoding

• *Mr. John Doe; Oct 1st, 2001; RR 120/80; BMI 29; OI.*

• Unique **encoding (identification)** of

data elements

“blood pressure”

“date of birth”

...

categorical data values, units

Male identity

mmHg

Osteogenesis Imperfecta

...

To prevent **ambiguity**

Ontologies – Describing

- Provide **descriptions** for **humans** and **machines**
 - **Humans**: Multiple languages, synonyms, etc.

ORPHA:666

[Classification level: Disorder](#)

Synonym(s):

Brittle bone disease

Glass bone disease

Lobstein disease

OI

Osteopsathyrosis

Porak and Durante disease

Prevalence: 1-5 / 10 000

Inheritance: Autosomal dominant or Autosomal recessive

Age of onset: All ages



orphanet

Language

English

French

German

Dutch

Italian

Spanish

Portuguese

Polish

Czech

Ontologies – Describing

- Provide **descriptions** for **humans** and **machines**
 - **Machines**: Mappings.

ORPHA:666

[Classification level: Disorder](#)

Synonym(s):

Brittle bone disease

Glass bone disease

Lobstein disease

OI

Osteopsathyrosis

Porak and Durante disease

Prevalence: 1-5 / 10 000

Inheritance: Autosomal dominant or Autosomal recessive

Age of onset: All ages

ICD-10: Q78.0

OMIM: [166200](#) [166210](#) [166220](#)
[166230](#) [259420](#) [259440](#) [610682](#)

[610915](#) [610967](#) [610968](#) [613848](#)
[613849](#) [613982](#) [614856](#) [615066](#)
[615220](#) [616229](#) [616507](#)

UMLS: C0023931 C0029434
C0268360 C1859069

MeSH: D010013

GARD: [1017](#)

MedDRA: 10031243

Ontologies – Describing

- Provide **descriptions** for **humans** and **machines**
 - **Machines**: Properties.

Genetic disease (disorder)

☰ Osteogenesis imperfecta (disorder) ☆ 📄

SCTID: 78314001

78314001 | Osteogenesis imperfecta (disorder) |

- en* Osteogenesis imperfecta
- en* Osteopsathyrosis
- en* Fragilitas ossium
- en* Osteogenesis imperfecta (disorder)
- en* Brittle bone syndrome
- en* OI - Osteogenesis imperfecta
- en* Brittle bone disease

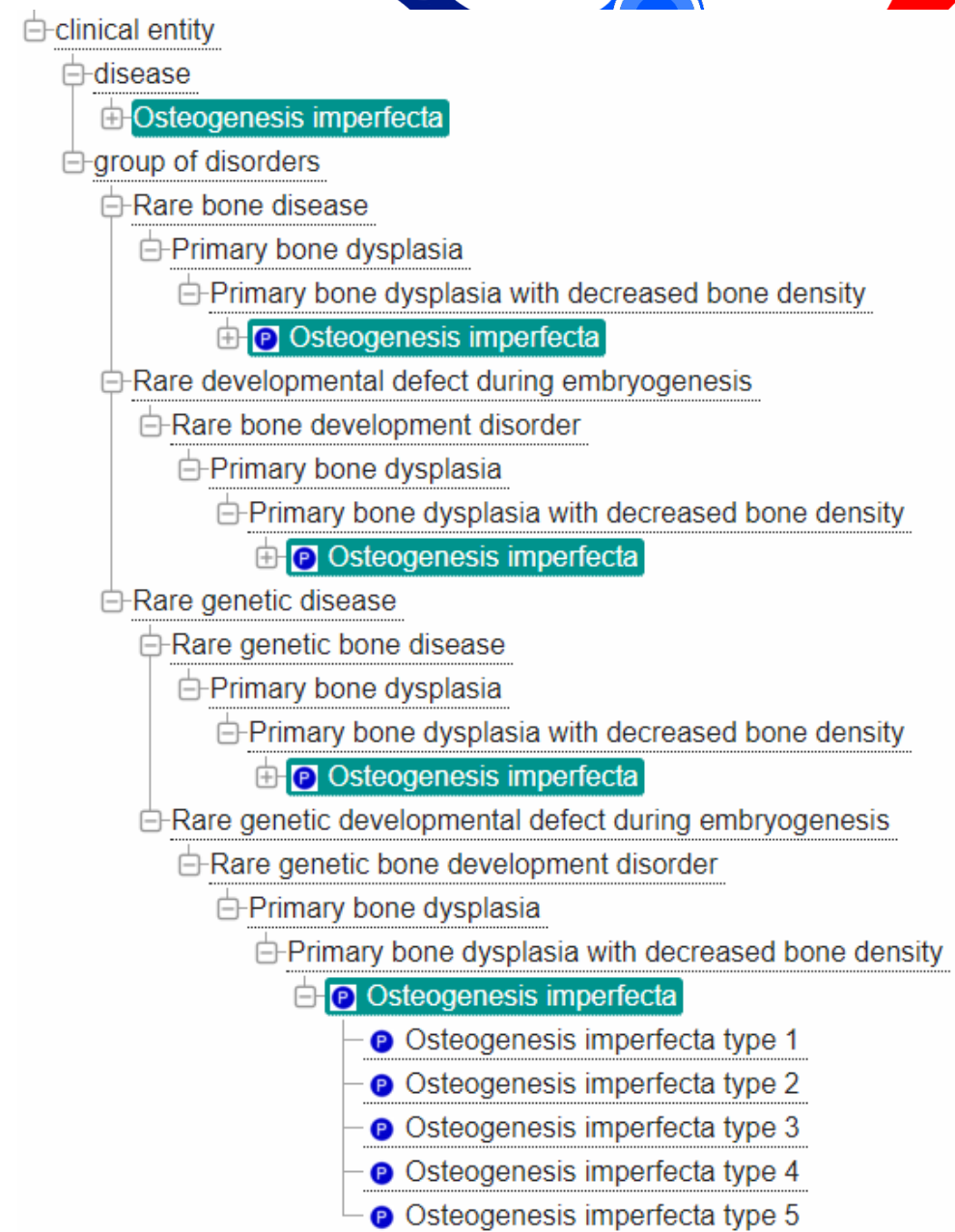
Axiom

- Finding site → Bone structure
- Occurrence → Congenital
- Associated morphology → Dysplasia
- Pathological process → Pathological developmental process

- Has interpretation → Abnormal
- Interprets → Bone formation, function

Ontologies – Ordering

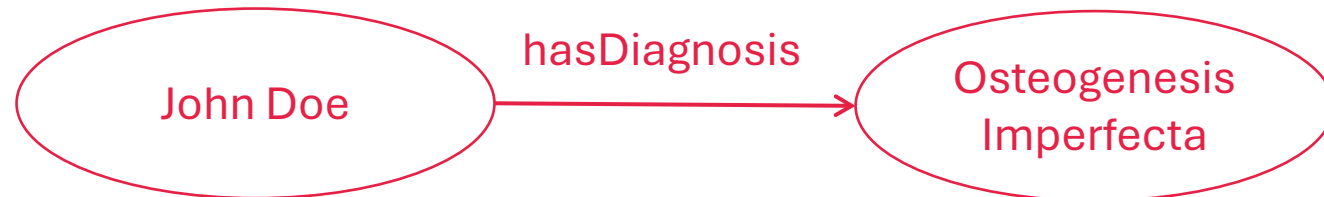
- To **record** data at varying detail
 - OI
 - OI type 5
- To **select** data at varying detail
 - All patients with a type of OI
 - All patients with a type of Rare bone disease



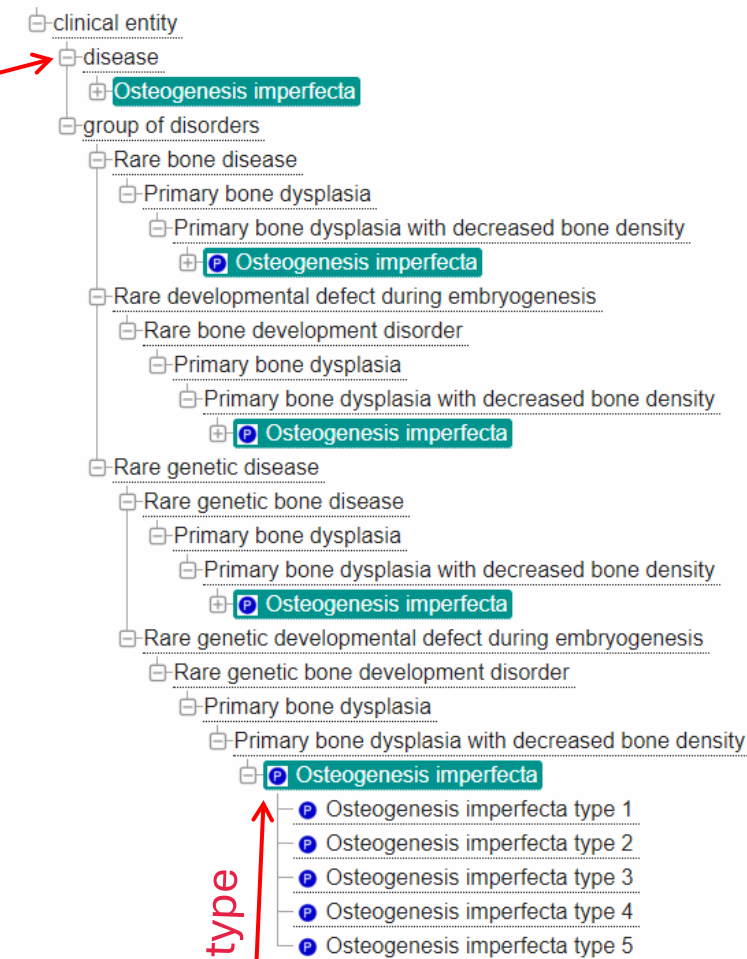
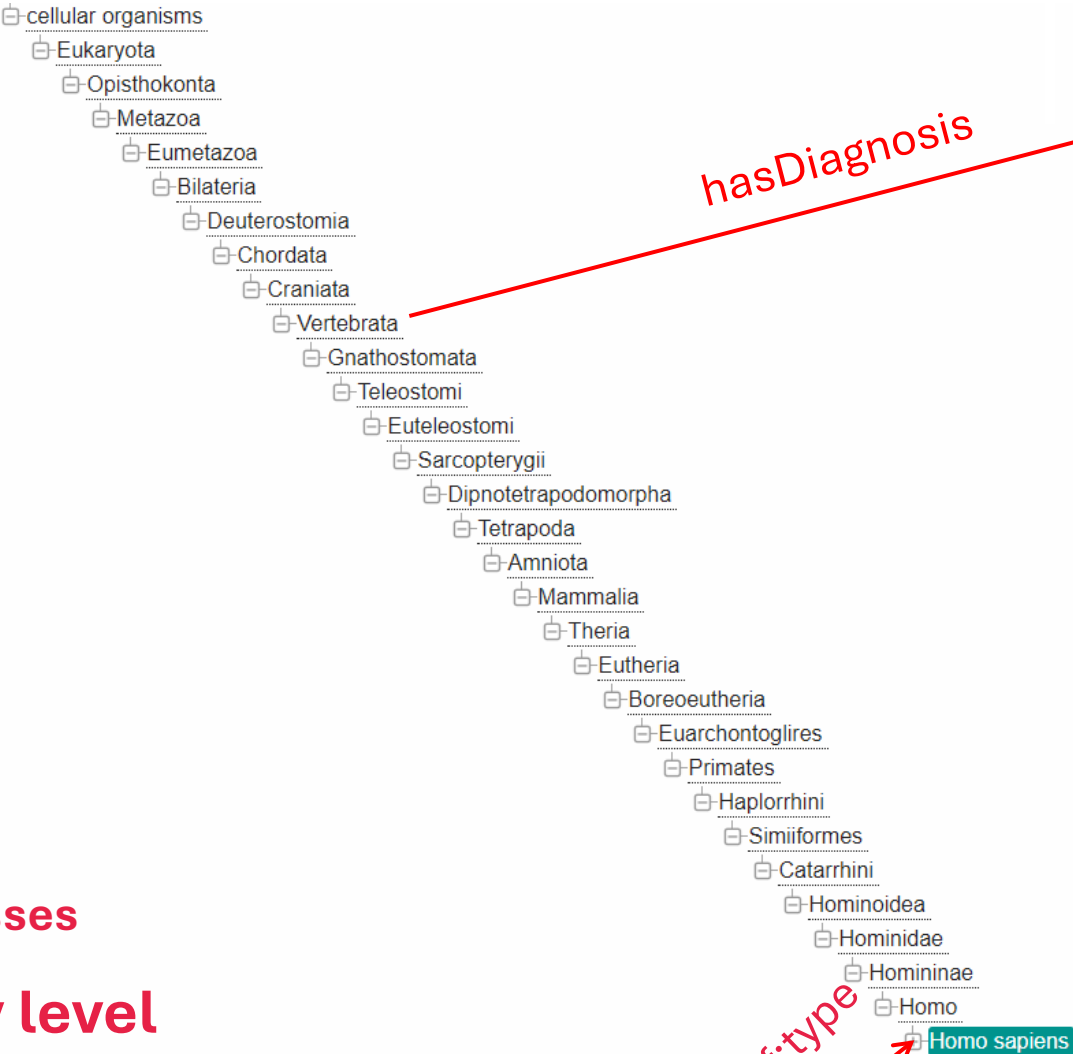
Data and ontologies

data level

(RDF) Instances



Data ε



(RDFS) Classes

ontology level

data level

(RDF) Instances



hasDiagnosis



Summary - Ontologies

Ontologies serve three main purposes:

- Encoding
 - A unique identifier (per ontology) to represent entity types (concepts; data elements / properties)
- Describing
 - **Human**-readable labels (names, synonyms) in (m)any languages
 - **Machine**-readable properties and mappings to other
- Ordering
 - Storage and retrieval at “arbitrary” level of detail

Demo – powers and pitfalls

1. [wikipedia.org](https://www.wikipedia.org)
 - Geared towards **humans**
2. [wikidata.org](https://www.wikidata.org)
 - **Human**-readable version of **machine** representation
 - With human interfacing, e.g., search functionality, human-readable labels, all based on Resource Description Framework (RDF)
3. query.wikidata.org
 - **Machine**-readable representation and search functionality, with “tricks” to support human-readability

Demo – powers and pitfalls

1. [wikipedia.org](https://www.wikipedia.org)
 - Geared towards **humans**
2. [wikidata.org](https://www.wikidata.org) **Pages:** [RareDisease](#) [OsteogenesisImperfecta](#)
 - **Human**-readable version of **machine** representation
 - With human interfacing, e.g., search functionality, human-readable labels, all based on Resource Description Framework (RDF)
3. query.wikidata.org **Queries:** [RDs](#) [RD-names](#) [RD-death_cause](#)
 - **Machine**-readable representation and search functionality, with “tricks” to support human-readability

Schemas – examples

- Set of common data elements for Rare Diseases Registration
 - Specific for rare disease registries

EUROPEAN PLATFORM ON RARE DISEASE REGISTRATION (EU RD Platform)

SET OF COMMON DATA ELEMENTS FOR RARE DISEASES REGISTRATION

GROUP	ELEMENT N°	ELEMENT NAME	ELEMENT DESCRIPTION	CODING	COMMENT
1. Pseudonym	1.1.	Pseudonym	Patient's pseudonym	• String	https://eu-rd-platform.irc.ec.europa.eu/spider
	2.1.	Date of birth	Patient's date of birth	• Date (dd/mm/yyyy)	
2. Personal information	2.2.	Sex	Patient's sex at birth	• Female • Male • Undetermined • Foetus (Unknown)	
	3.1.	Patient's status	Patient alive or dead	• Alive • Dead • Lost in follow-up • Opted-out	If dead then answer question 3.2
3. Patient Status	3.2.	Date of death	Patient's date of death	• Date (dd/mm/yyyy)	
	4.1.	First contact with specialised centre	Date of first contact with specialised centre	• Date (dd/mm/yyyy)	
4. Care pathway					
5. Disease history	5.1.	Age at onset	Age at which symptoms/signs first appeared	• Antenatal • At birth • Date (dd/mm/yyyy) • Undetermined	
	5.2.	Age at diagnosis	Age at which diagnosis was made	• Antenatal • At birth • Date (dd/mm/yyyy) • Undetermined	
6. Diagnosis	6.1.	Diagnosis of the rare disease	Diagnosis retained by the specialised centre	Orpha code (strongly recommended – see link) / Alpha code/ ICD-9 code/ ICD-9-CM code / ICD-10 code	http://www.orphadata.org/cgi-bin/inc/product1.inc.php
	6.2.	Genetic diagnosis	Genetic diagnosis retained by the specialised centre	International classification of mutations (HGVS) (strongly recommended – see link) / HGNC / OMIM code	http://www.hgvs.org
	6.3.	Undiagnosed case	How the undiagnosed case is defined	• Phenotype (HPO) • Genotype (HGVS)	
7. Research	7.1.	Agreement to be contacted for research purposes	Patient's permission exists for being contacted for research purposes	• YES • NO	
	7.2.	Consent to the reuse of data	Patient's consent exists for his/her data to be reused for other research purposes	• YES • NO	
	7.3.	Biological sample	Patient's biological sample available for research	• YES • NO	If YES answer question 7.4
	7.4.	Link to a biobank	Biological sample stored in a biobank	• YES (if appropriate use link) • NO	https://directory.bbmri-eric.eu
8. Disability	8.1.	Classification of functioning/disability	Patient's disability profile according to International Classification of Functioning and Disability (ICF)	• Disability profile / Score	http://www.who.int/classifications/icf/whodasii/en/

Schemas – examples

- Set of common data elements for Rare Diseases Registration
- Condition – HL7 FHIR
 - HL7 FHIR: primarily used for data exchange

Name	Flags	Card.	Type
Condition	TU		DomainResource
identifier	Σ	0..*	Identifier
clinicalStatus	?! Σ C	1..1	CodeableConcept
verificationStatus	?! Σ	0..1	CodeableConcept
category	C	0..*	CodeableConcept
severity		0..1	CodeableConcept
code	Σ	0..1	CodeableConcept
bodySite	Σ	0..*	CodeableConcept
subject	Σ	1..1	Reference(Patient Group)
encounter	Σ	0..1	Reference(Encounter)
onset[x]	Σ	0..1	
abatement[x]	C	0..1	
recordedDate	Σ	0..1	dateTime
participant	Σ	0..*	BackboneElement
stage	C TU	0..*	BackboneElement
evidence	Σ TU	0..*	CodeableReference(Any)
note		0..*	Annotation

Schemas – examples

- Set of common data elements for Rare Diseases Registration
- Condition – HL7 FHIR
- [OMOP-CDM](#) – Condition_occurrence
 - OMOP-CDM: Focus on real-world

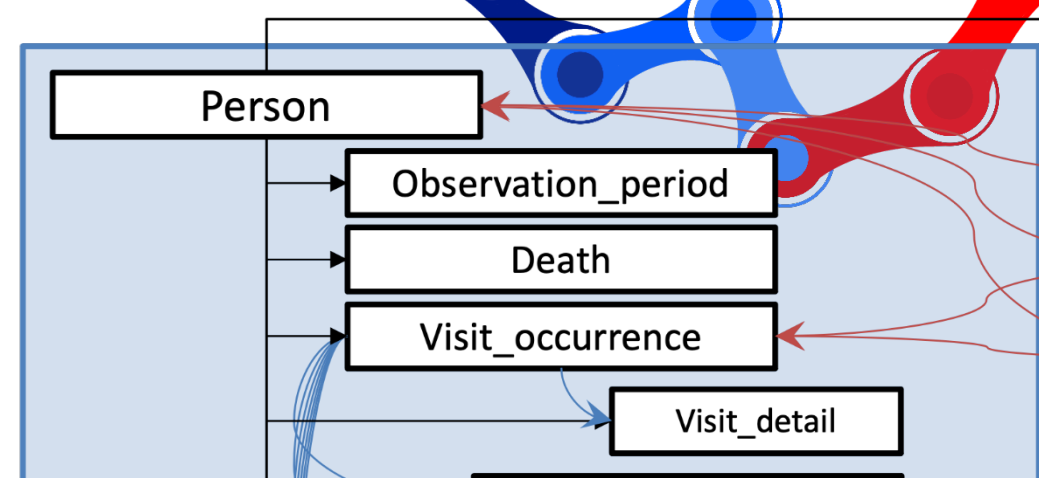
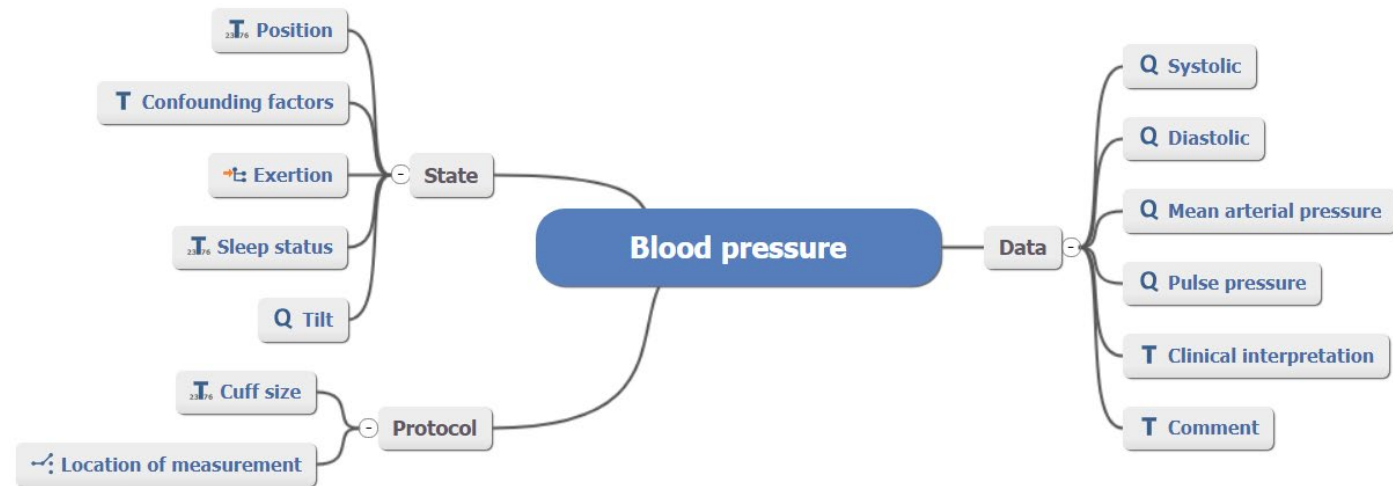


Table 4.7: The CONDITION_OCCURRENCE table.

Column name	Value	Explanation
CONDITION_OCCURRENCE_ID	964	This is typically an autogenerated value creating a unique identifier for each record.
PERSON_ID	1	This is a foreign key to Laura's record in the PERSON table and links PERSON to CONDITION_OCCURRENCE.
CONDITION_CONCEPT_ID	194696	A foreign key referring to the SNOMED code 266599000: 194696.
CONDITION_START_DATE	2010-01-06	The date when the instance of the Condition is recorded.
CONDITION_START_DATETIME	2010-01-06 00:00:00	The date and time when the instance of the Condition is recorded. Midnight is used since the time is unknown.
CONDITION_END_DATE	NULL	This is the date when the instance of the Condition is considered to have ended, but this is rarely recorded.
CONDITION_END_DATETIME	NULL	If known, this is the date and time when the instance of the Condition is considered to have ended.

Schemas – examples

- Set of common data elements for Rare Diseases Registration - EU_RD_Platform
- Condition – HL7 FHIR
- OMOP-CDM – Condition_occurrence
- Blood pressure – OpenEHR
OpenEHR: capture of health data conform ISO-standard



Schemas – examples

- Set of common data elements for Rare Diseases Registration - EU_RD_Platform
- Condition – HL7 FHIR
- OMOP-CDM – Condition_occurrence
- Blood pressure – OpenEHR
- Phenopackets
 - Focus on Bioinformation
- And many others!

Disease

The word *phenotype* is used with many different meanings, including “the observable traits of an organism”. In medicine, the word can be used with at least two different meanings. It is used to refer to some **observed** deviation from normal morphology, physiology, or behavior. In contrast, the *disease* is a diagnosis, i.e., an inference or hypothesis about the cause underlying the observed phenotypic abnormalities. Occasionally, physicians use the word *phenotype* to refer to a disease, but we do not use this meaning here.

Data model

Field	Type	Multiplicity	Description
term	OntologyClass	1..1	An ontology class that represents the disease. REQUIRED.
excluded	boolean	0..1	Flag to indicate whether the disease was observed or not.
onset	TimeElement	0..1	an element representing the age of onset of the disease
resolution	TimeElement	0..1	an element representing the age of resolution (abatement) of the disease
disease_stage	OntologyClass (List)	0..*	List of terms representing the disease stage e.g. AJCC stage group.
clinical_tnm_finding	OntologyClass (List)	0..*	List of terms representing the tumor TNM score
primary_site	OntologyClass	0..1	the primary site of disease
laterality	OntologyClass	0..1	laterality (left or right) of the primary site of sites if applicable

Schemas – purpose

- To describe **properties** of **classes** (a.k.a. “types”)
- To **characterize** those properties
 - Cardinality (1 or more)
 - Optionality (mandatory or not)
 - Allowed values

Schemas – Characterizing

- *Mr. John Doe; Oct 1st, 2001; RR 120/80; BMI 29; OI.*

LOINC:54125-0 ← **Name:**
LOINC:76691-5 ← **Gender:**
LOINC:21112-8 ← **Date of birth:**
LOINC:35094-2 ← **Blood pressure:**
LOINC:39156-5 ← **Body Mass Index:**
LOINC:29308-4 ← **Diagnosis:**

“John Doe” → Exactly 1 String
Male identity → Exactly 1 SNOMED-coded Gender identity finding
Oct 1st, 2001 → Exactly 1 Date
120/80 **mmHg** → 1 or more syst/diast values + unit
29 → At most 1 integer
Osteogenesis Imperfecta → 1 or more Orpha-code

Schemas and ontologies

- Ideally, schemas identify their elements and the allowed (categorical) values using (third-party) ontologies
- In practice, mappings are defined among schemas, such as HL7 FHIR, OMOP-CDM, OpenEHR, Phenopackets

Conclusion

- Ontologies aim to:
 - Encode / identify concepts
 - Describe concepts with codes and human/machine descriptions
 - Order concepts in a hierarchy, to query at arbitrary level of detail
- Schemas aim to:
 - Describe relevant properties of classes
 - Characterize those properties (e.g., cardinality, optionality, allowed values)

HemaFAIR surveys

- Please complete the **Data Protection and Consent survey** to receive personalised links for two follow-up surveys
 - Knowledge on FAIR principles
 - Training gaps and requirements



Prof. Ronald Cornet, PhD

Full Professor ; Principal Investigator ; Principal Educator ;
Education director Medical informatics; Program Leader APH Digital Health
Department of Medical Informatics - Reusable Health Data
Amsterdam Public Health Research Institute
Amsterdam UMC

Amsterdam Public Health Research Institute

Location AMC | Meibergdreef 15, 1105 AZ Amsterdam

 [+31 20-566 52 05](tel:+31205665205) |  r.cornet@amsterdamumc.nl

kik.amsterdamumc.org/home/rcornet | www.amsterdamumc.nl