Healthcare-EHR-SynData

Alessia Artusi, Emma Salti

course of Knowledge Graph Engineering

Academic Year 2022/2023

Purpose and DoI

"A service which facilitates citizens to handle (their own) healthcare data, in a context in which the citizens moves around Europe "

- Improved productivity of physicians
- Improved patient care delivery
- Facilitated cross-border care
- Easier data processing

Resources

Knowledge resources:

- schema.org
- FHIR

Database used:

- Synthea
- EMRBOTS

Purpose formalization - personas, scenario, CQ

Pedro

cardiologist at La Fe university and polytechnic hospital in Valencia, Spain.

CQ: Pedro is a cardiologist at La Fe Hospital in Valencia. During its work time a foreign patient arrives to the hospital unconscious.To intervene the doctor needs to know the patient medical records in order to safely operate.

Sonia

Italian woman from Olmeneta, Italy.

CQ: Sonia is an Italian woman from Olmeneta, Italy. She is 67 years-old and decides to travel around Andalusia (Spain) for a couple of weeks. One day, during her trip she notices a cutaneous rush. She travels to the nearest emergency room but does not speak spanish.

Andreas

Biotechnologist at the university hospital in Berlin.

CQ: Andreas is a biotechnologist who works at the university hospital of Berlin. He usually perform analysis on the patients' lab samples.He need a records of the already performed tests in order to avoid running duplicate tests.

Inception phase - synthea

- Patient ID: unique identifier
- Name: name and surname of the patient
- Gender
- Birthdate
- Marital Status
- Address: complete address
- Language: spoken language
- Encounters: encounters conducted on the patient with the relative diagnosis description and code
- Careplans: description of the careplan for each encounter
- Devices: devices used by the patient (i.e.: wheelchair)
- Allergy
- Treatment: the treatment prescribed to the patient
- Dosage: dosage of the treatment
- Immunizations

Inception phase - EMRBOTS

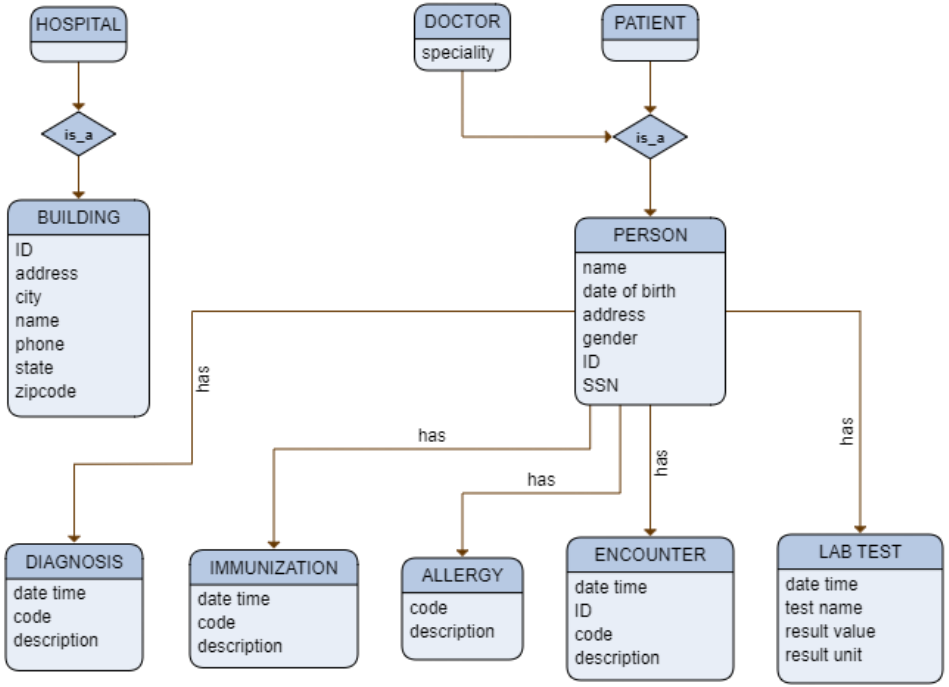
- Patient ID
- Gender
- Birthdate
- Population percentage below poverty
- Marital status
- Admission ID: unique identifier for the patient’s admission
- Admission start and end date and time
- Primary diagnosis code: ICD-10 code for the diagnosis
- Primary diagnosis description
- Lab test name
- Lab test result
- Lab test date

Informal Modeling

10 ETypes

2 Parent classes
building, person

3 Subclasses
hospital, patient, doctor

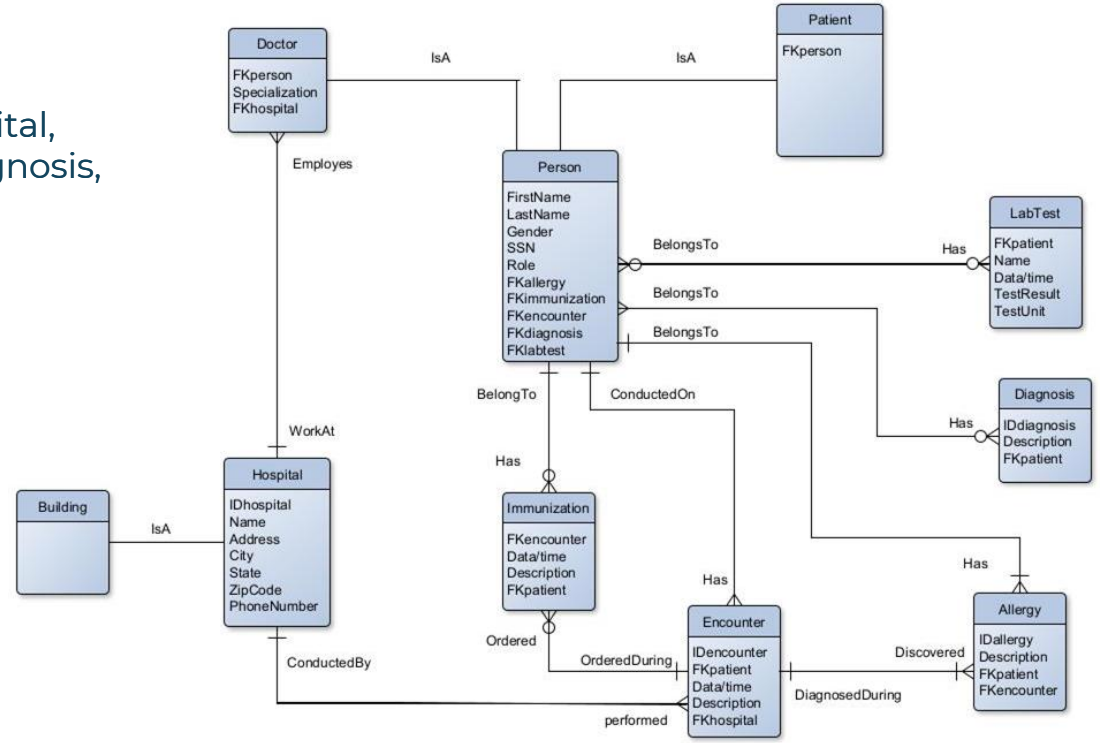


Informal Modeling

10 ETypes
person, patient, doctor, building, hospital, allergy, immunization, encounter, diagnosis, lab test

19 Object Properties

27 Data Properties

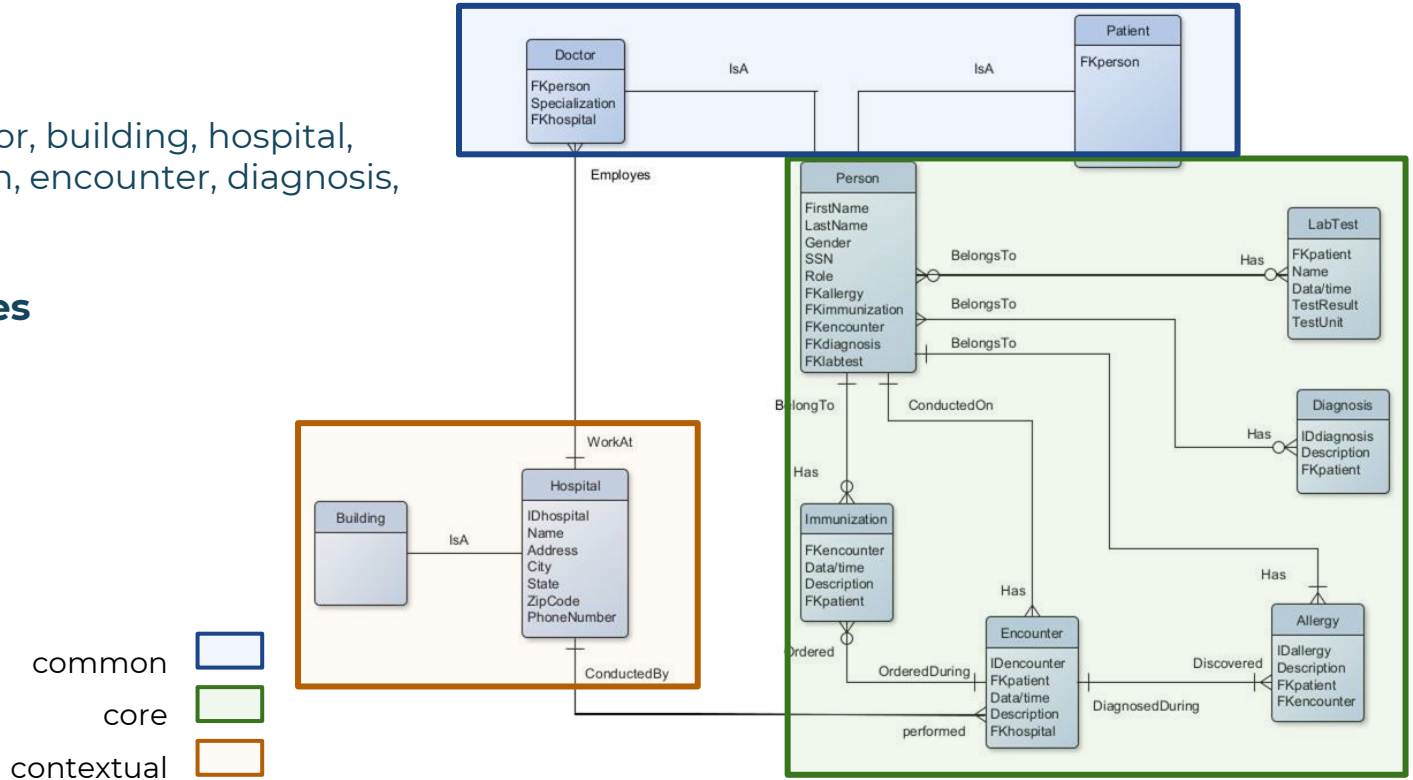


Informal Modeling

10 ETypes
person, patient, doctor, building, hospital, allergy, immunization, encounter, diagnosis, lab test

19 Object Properties

27 Data Properties

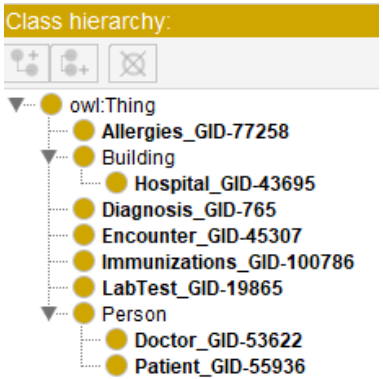


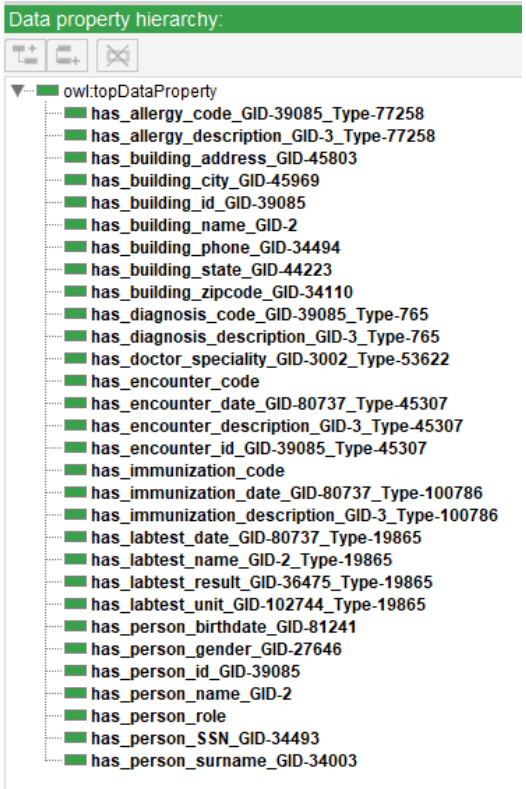
Formal Modeling

- **Schema building**

- Addressing **Language Misalignment**
  - data types misalignment
  - data value format misalignment

Integration of already present elements and new elements generated ad hoc using KOS

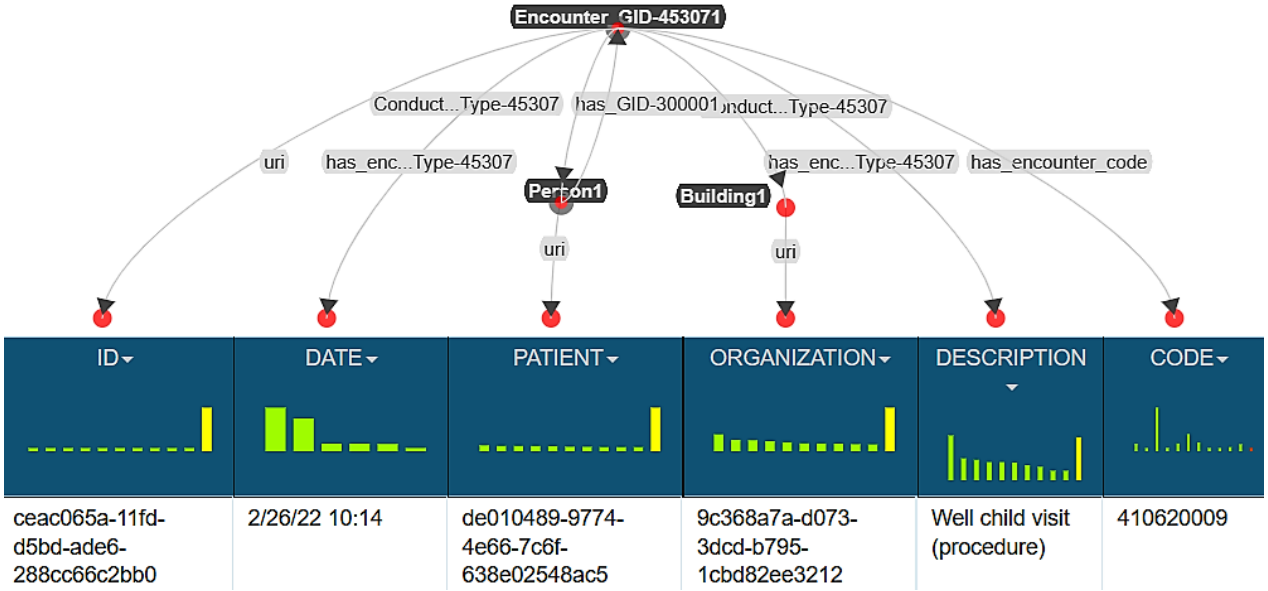




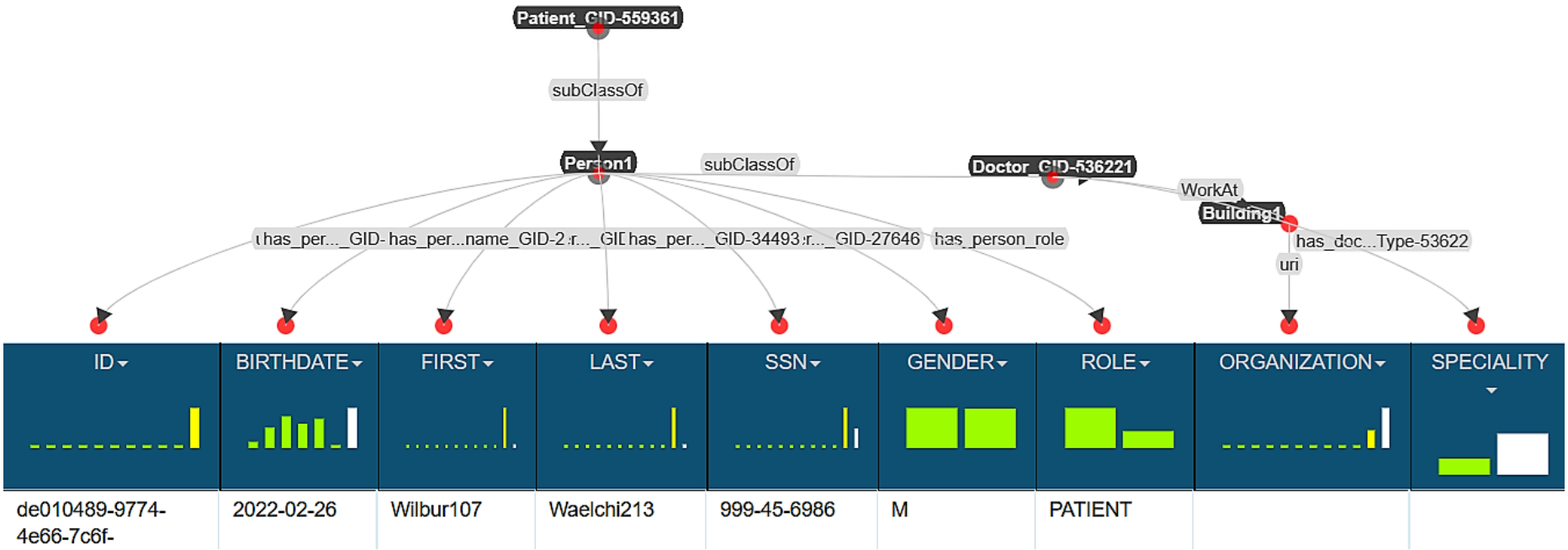
Data Integration

Mapping of **7** different **datasets** to the ETG
people, buildings, immunization, lab test, allergy, encounter, diagnosis

Karmalinker



Data Integration



SPARQL Query

Pedro 1.3 CQ

A list of the most recent blood test for a specific patient is returned with values associated

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

SELECT ?PersonName ?PersonSurname ?LabTestName ?LabTestValue ?LabtestUnit ?LabTestDate WHERE {
    ?Person a <http://knowdive.disi.unitn.it/etype#Person> ;
            <http://knowdive.disi.unitn.it/etype#has_person_name_GID-2> ?PersonName ;
            <http://knowdive.disi.unitn.it/etype#has_person_surname_GID-34003> ?PersonSurname.
    ?LabTest a <http://www.semanticweb.org/simone/ontologies/2021/10/untitled-ontology-34#LabTest_GID-19865> ;
             <http://www.semanticweb.org/simone/ontologies/2021/10/untitled-ontology-34#has_labtest_name_GID-2_Type-19865> ?LabTestName
;
             <http://www.semanticweb.org/simone/ontologies/2021/10/untitled-ontology-34#has_labtest_result_GID-36475_Type-19865> ?
LabTestValue ;
             <http://www.semanticweb.org/simone/ontologies/2021/10/untitled-ontology-34#has_labtest_unit_GID-102744_Type-19865> ?
LabtestUnit ;
             <http://www.semanticweb.org/simone/ontologies/2021/10/untitled-ontology-34#has_labtest_date_GID-80737_Type-19865> ?
LabTestDate .

    FILTER REGEX (?LabTestName, "^CBC*") .
    FILTER(?PersonName = "Bess947") .
}
```

SPARQL Query

Pedro 1.3 CQ

A list of the most recent blood test for a specific patient is returned with values associated

PersonName	PersonSurname	LabTestName	LabTestValue	LabtestUnit	LabTestDate
Bess947	Huels583	CBC: HEMATOCRIT	440	%	9/30/97 9:38
Bess947	Huels583	CBC: HEMATOCRIT	324	%	9/30/97 22:57
Bess947	Huels583	CBC: WHITE BLOOD CELL COUNT	91	k/cumm	9/30/97 16:56
Bess947	Huels583	CBC: MEAN CORPUSCULAR VOLUME	778	fl	9/30/97 16:52
Bess947	Huels583	CBC: ABSOLUTE NEUTROPHILS	728	%	9/30/97 16:41

Evaluation

- no polysemous attributes
- no repeated data properties
- classes are distincted

After the adjustments the graph is able to solve most of the CQs.

ENTITIES COVERAGE	10/10	1
PROPERTIES COVERAGE	(45 + 20 - 7)/(45 + 20)	0.89

Open Issues

- Datasets not fully representative DoI
- ETG had to be modified to make GraphDB work
- date data type issue

THANK YOU FOR YOUR ATTENTION !