dayton children's medical data ontology

Anmol Saini * Jason Nolte * Spencer Seals

Outline

- Use case overview
- Problem
- Key Notions
- Schema Diagrams
- SPARQL Queries
- Future Work
- Retrospective

Use Case Overview

- Ontology is designed for a specific customer
  - Several researchers from Dayton Children's hospital and Wright State University Boonshoft School of Medicine
- Collect and analyze data for research purposes from large scale nationwide health databases

Use Case Overview

- Databases contain info from a variety of sources:
  - Patient-specific medical info
  - Billing records
  - Prescription drugs ordered
  - And more!

Problem Overview

- Each set is exported as a csv abstract
  - Most have some columns in common
  - NOT organized on the same keys
- Research group would like to analyze data that relies on information from multiple abstracts
- Often unclear how to merge these data sources
- Would be nice to have a common representation framework

Problem Overview

- Each set is exported as a csv abstract
  - Most have some columns in common
  - NOT organized on the same keys
- Research group would like to analyze data that relies on information from multiple abstracts
- Often unclear how to merge these data sources
- Would be nice to have a common representation framework

- **An extensible ontology**

Key Notions

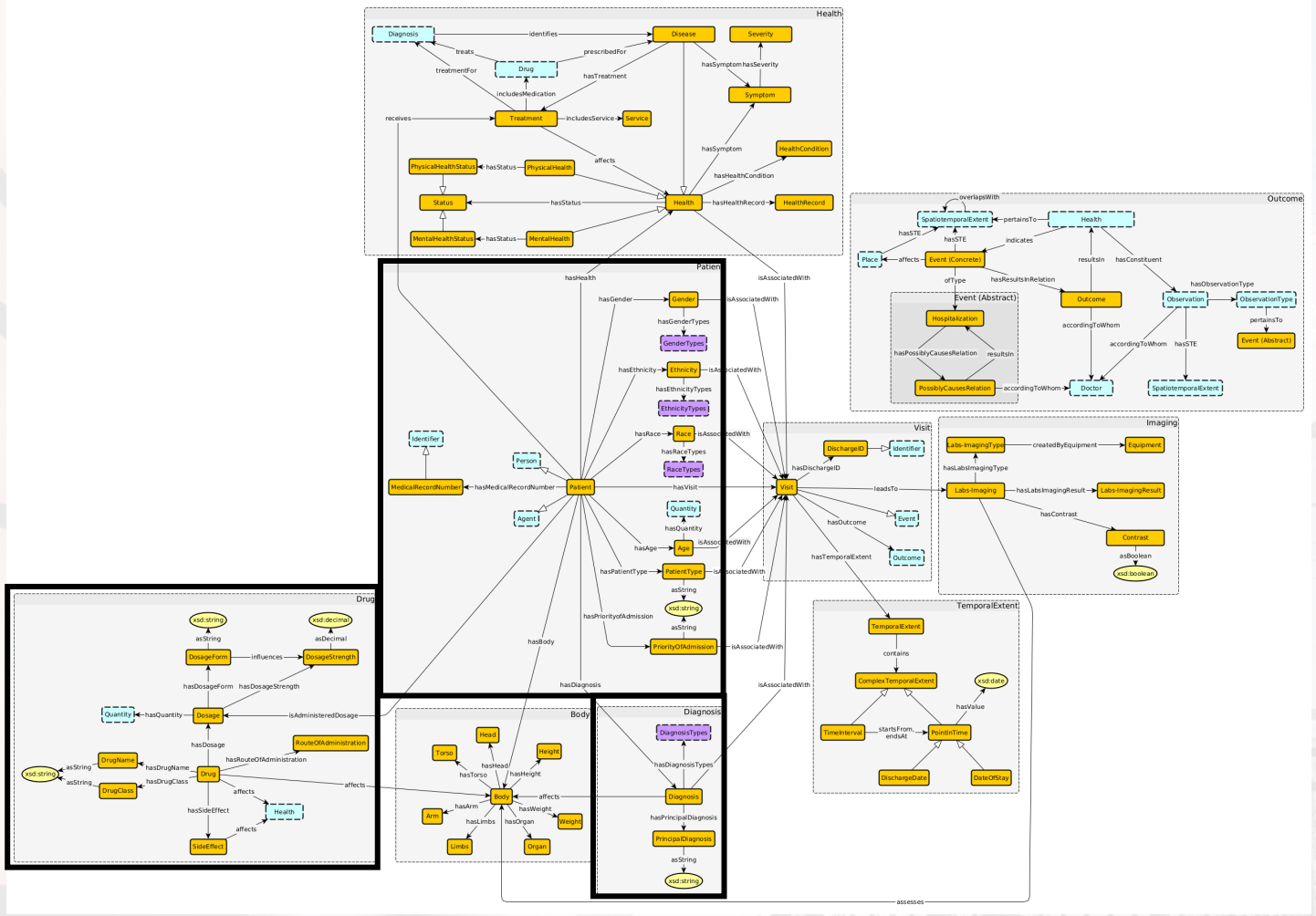
- Patient
- Drug
- Body
- Diagnosis
- Visit
- Temporal Extent
- Imaging
- Outcome
- Health

Most Concrete – direct data and physical objects

Less Concrete – direct data available

Most abstract – interpreted from other data

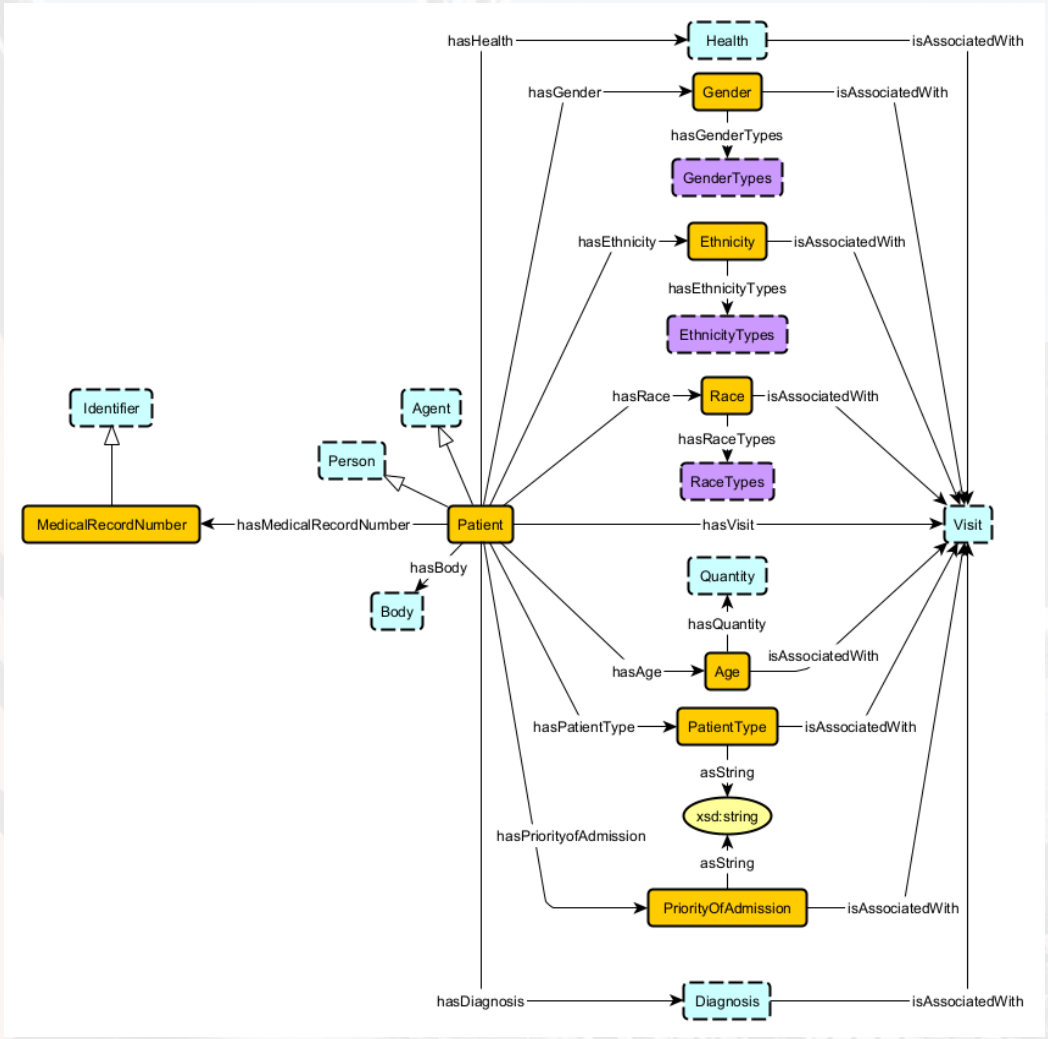
Overall Graph



Patient

```
gd = [
    "Patient hasVisit Visit",
    "Patient hasPatientType PatientType",
    "Patient hasPriorityofAdmission PriorityofAdmission",
    "Patient hasMedicalRecordNumber MedicalRecordNumber",
    "Patient hasDiagnosis Diagnosis",
    "Patient isAdministeredDosage Dosage",
    "Race hasRaceTypes RaceTypes",
    "Ethnicity hasEthnicityTypes EthnicityTypes",
    "Gender hasGenderTypes GenderTypes"
]

sr = [
    "Patient hasGender Gender",
    "Patient hasEthnicity Ethnicity",
    "Patient hasRace Race",
    "Patient hasAge Age",
    "Patient hasVisit Visit",
    "Patient hasPatientType PatientType",
    "Patient hasPriorityofAdmission PriorityofAdmission",
    "Patient hasMedicalRecordNumber MedicalRecordNumber",
    "Patient hasDiagnosis Diagnosis",
    "Patient isAdministeredDosage Dosage",
    "Age hasQuantity Quantity",
    "Race hasRaceTypes RaceTypes",
    "Ethnicity hasEthnicityTypes EthnicityTypes",
    "Gender hasGenderTypes GenderTypes"
]
```



Ontology metrics:

Metrics

Axiom	129
Logical axiom count	102
Declaration axioms count	27
Class count	16
Object property count	12
Data property count	0
Individual count	0
Annotation Property count	0

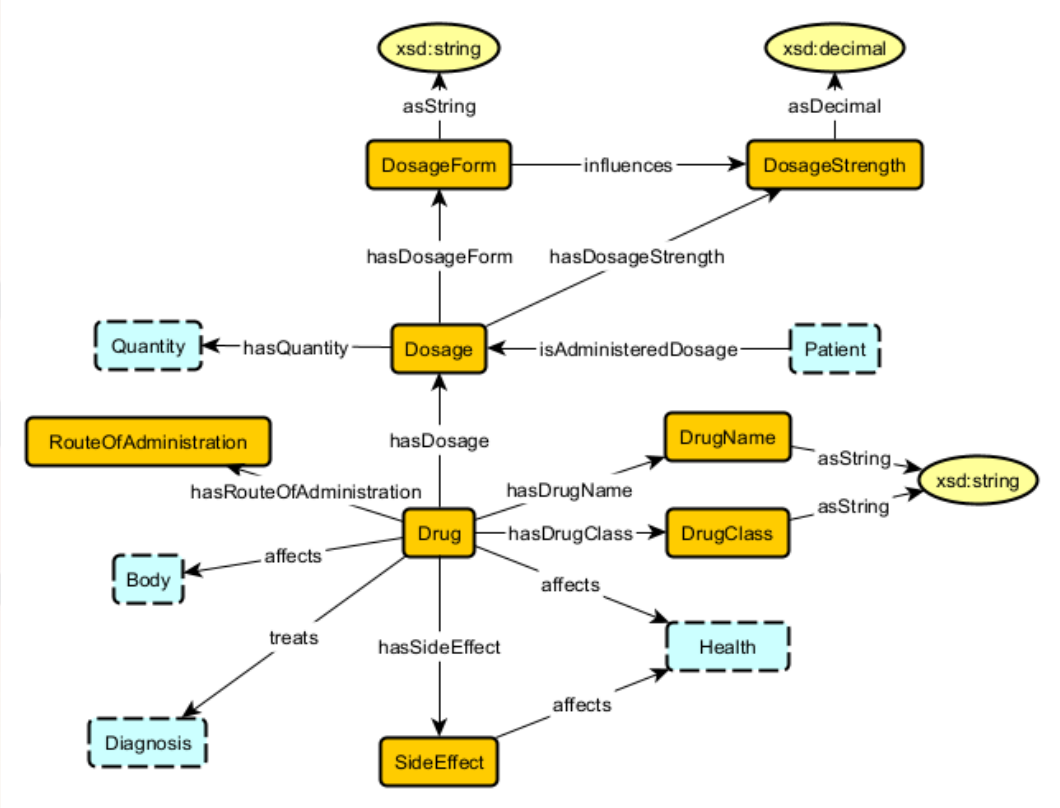
Class axioms

SubClassOf	87
EquivalentClasses	0
DisjointClasses	15
GCI count	12
Hidden GCI Count	0

Drug

```
gd = [
    "Drug hasDosage Dosage",
    "Drug hasRouteOfAdministration
RouteOfAdministration",
    "patient isAdministered Dosage",
    "Dosage hasDosageStrength DosageStrength",
    "Drug hasSideEffect SideEffect",
    "Dosage hasDosageForm DosageForm",
    "Drug hasName DrugName",
    "Drug hasDrugClass DrugClass"
]

sr = [
    "Dosage hasDosageStrength DosageStrength",
    "Dosage hasDosageForm DosageForm",
    "Drug hasDosage Dosage",
    "Drug hasRouteOfAdministration
RouteOfAdministration",
    "Drug affects Body_or_Health",
    "Drug hasSideEffect SideEffect",
    "Drug hasDrugName DrugName",
    "Drug hasDrugClass DrugClass",
    "SideEffect affects Health",
    "Dosage hasQuantity Quantity"
]
```



Ontology metrics:

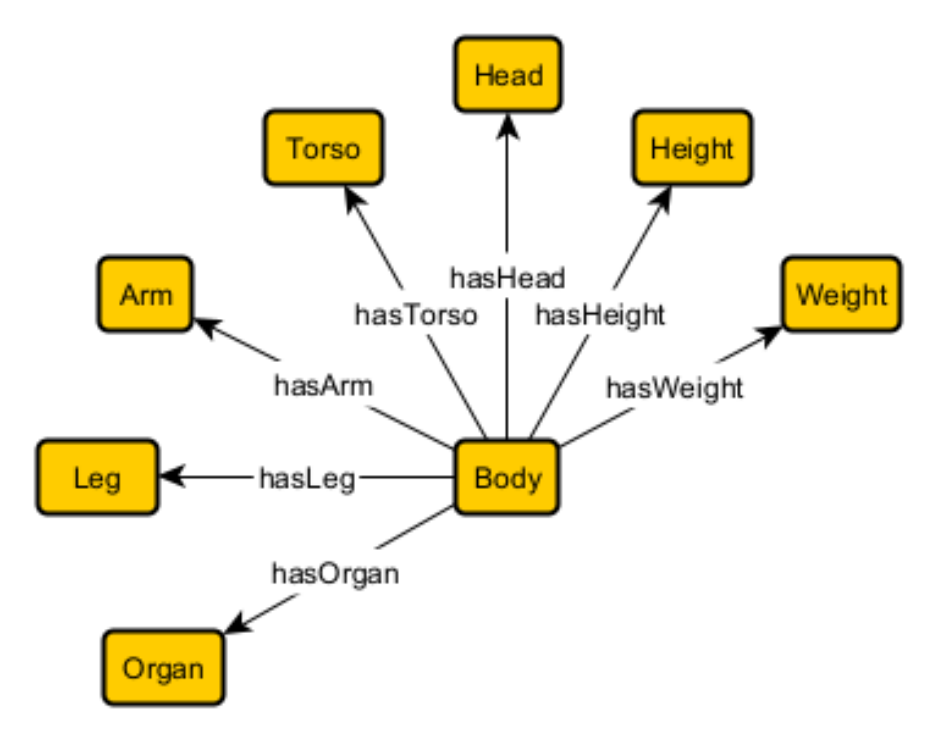
Metrics

Axiom	87
Logical axiom count	62
Declaration axioms count	25
Class count	14
Object property count	12
Data property count	0
Individual count	0
Annotation Property count	0

Class axioms

SubClassOf	51
EquivalentClasses	0
DisjointClasses	11
GCI count	15
Hidden GCI Count	0

Body



Ontology metrics:

Metrics

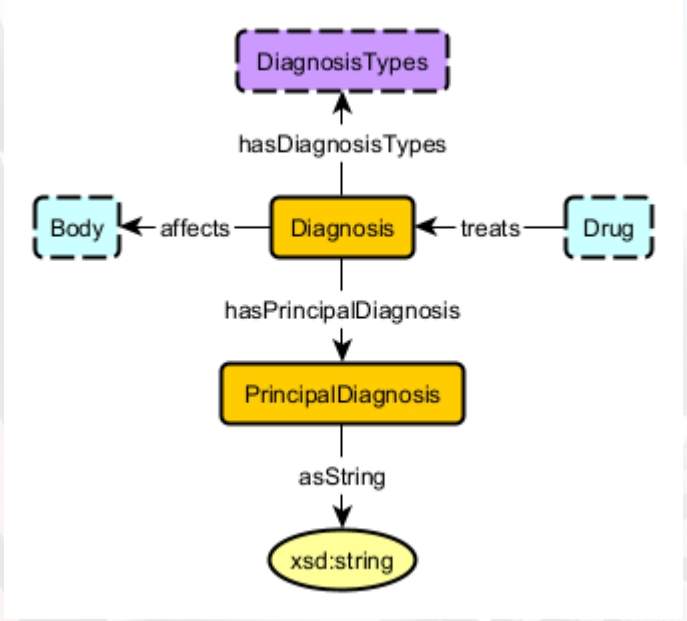
Axiom	34
Logical axiom count	19
Declaration axioms count	15
Class count	9
Object property count	7
Data property count	0
Individual count	0
Annotation Property count	0

Class axioms

SubClassOf	19
EquivalentClasses	0
DisjointClasses	0
GCI count	2
Hidden GCI Count	0

Diagnosis

```
gd = [
    "Diagnosis hasPrincipalDiagnosis PrincipalDiagnosis",
    "Diagnosis hasDiagnosisTypes DiagnosisTypes",
    "Diagnosis identifies Disease",
    "Patient hasDiagnosis Diagnosis",
    "Treatment treatmentFor Diagnosis"
]
sr = [
    "Diagnosis hasPrincipalDiagnosis PrincipalDiagnosis",
    "Diagnosis hasDiagnosisTypes DiagnosisTypes",
    "Diagnosis identifies Disease",
    "Diagnosis affects Body",
    "Diagnosis isAssociatedWith Visit",
    "Patient hasDiagnosis Diagnosis",
]
```



Ontology metrics:

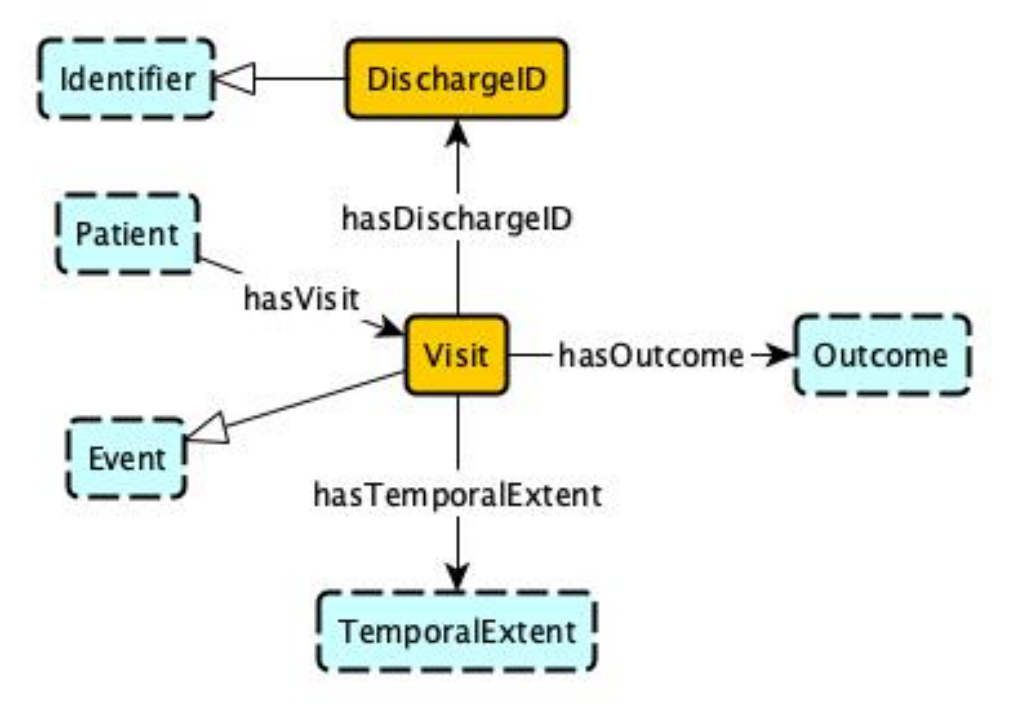
Metrics

Axiom	52
Logical axiom count	38
Declaration axioms count	14
Class count	8
Object property count	7
Data property count	0
Individual count	0
Annotation Property count	0

Class axioms

SubClassOf	32
EquivalentClasses	0
DisjointClasses	6
GCI count	8
Hidden GCI Count	0

Visit



Ontology metrics:

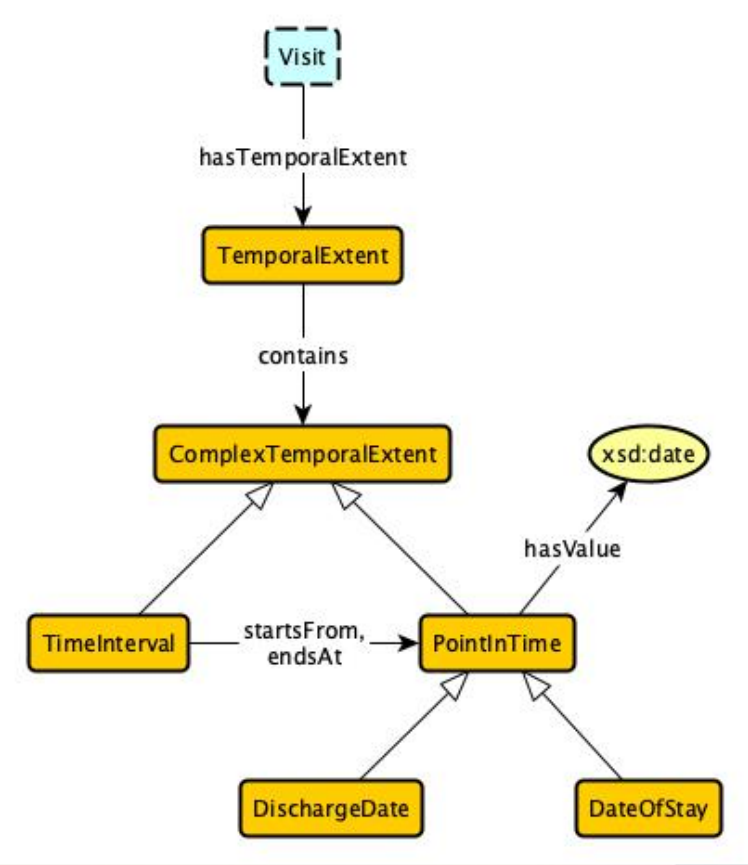
Metrics

Axiom	115
Logical axiom count	93
Declaration axioms count	22
Class count	17
Object property count	6
Data property count	0
Individual count	0
Annotation Property count	0

Class axioms

SubClassOf	80
EquivalentClasses	0
DisjointClasses	13
GCI count	3
Hidden GCI Count	0

Temporal Event



Ontology metrics:

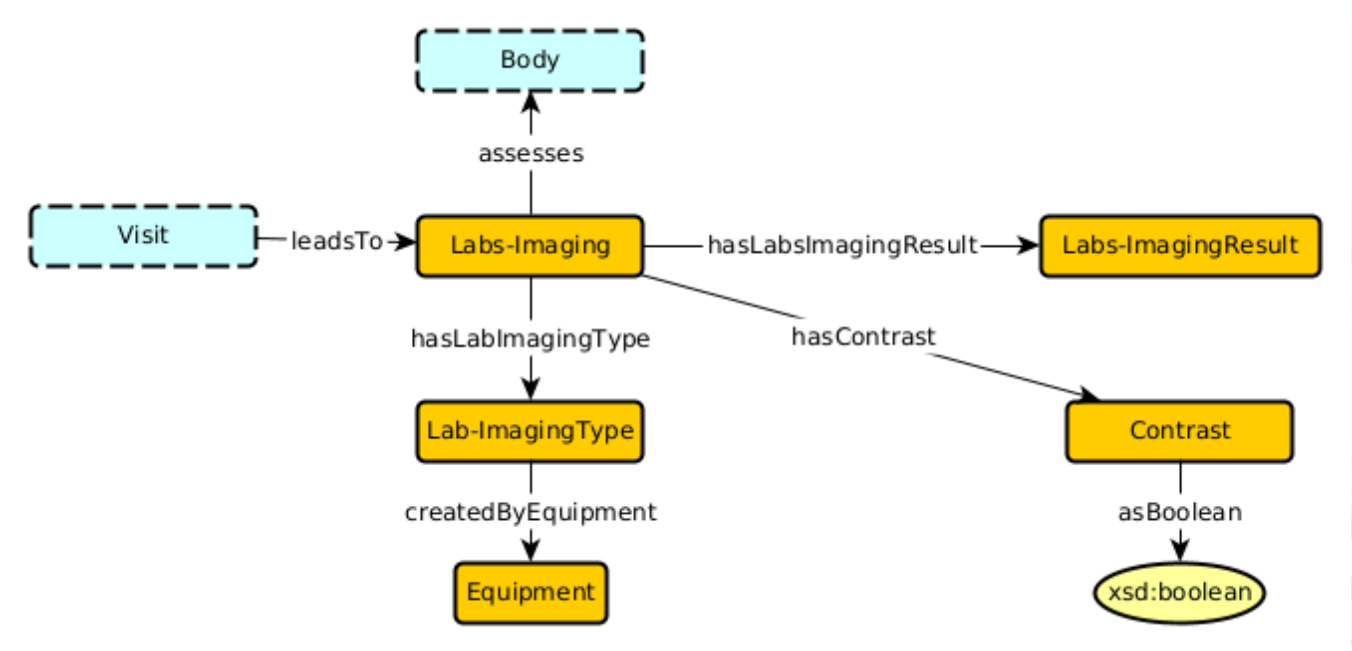
Metrics

Axiom	43
Logical axiom count	33
Declaration axioms count	10
Class count	7
Object property count	3
Data property count	1
Individual count	0
Annotation Property count	0

Class axioms

SubClassOf	31
EquivalentClasses	0
DisjointClasses	2
GCI count	5
Hidden GCI Count	0

Labs and Imaging



Ontology metrics:

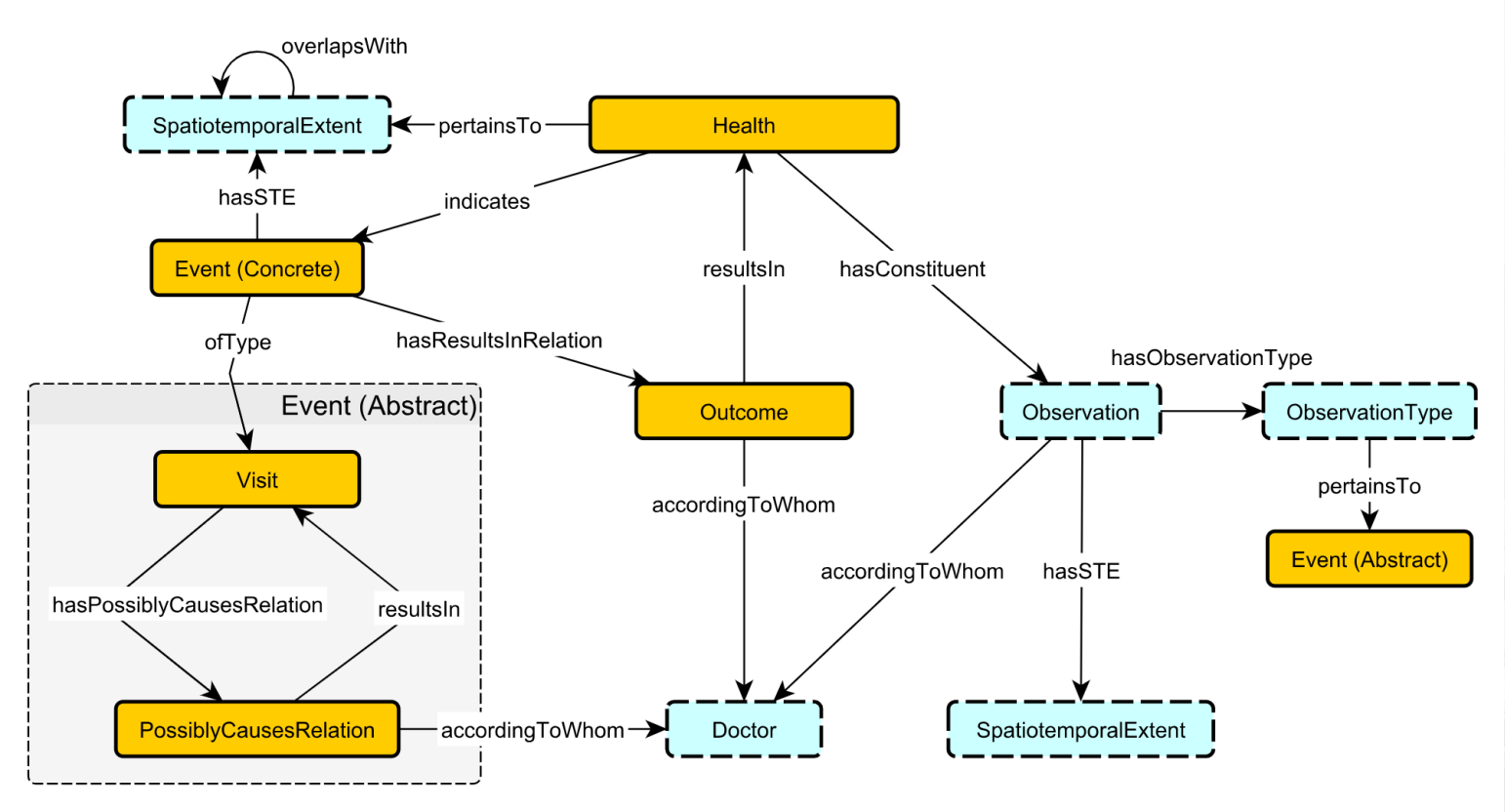
Metrics

Axiom	79
Logical axiom count	66
Declaration axioms count	13
Class count	8
Object property count	6
Data property count	0
Individual count	0
Annotation Property count	0

Class axioms

SubClassOf	60
EquivalentClasses	0
DisjointClasses	6
GCI count	9
Hidden GCI Count	0

Outcome



Ontology metrics:

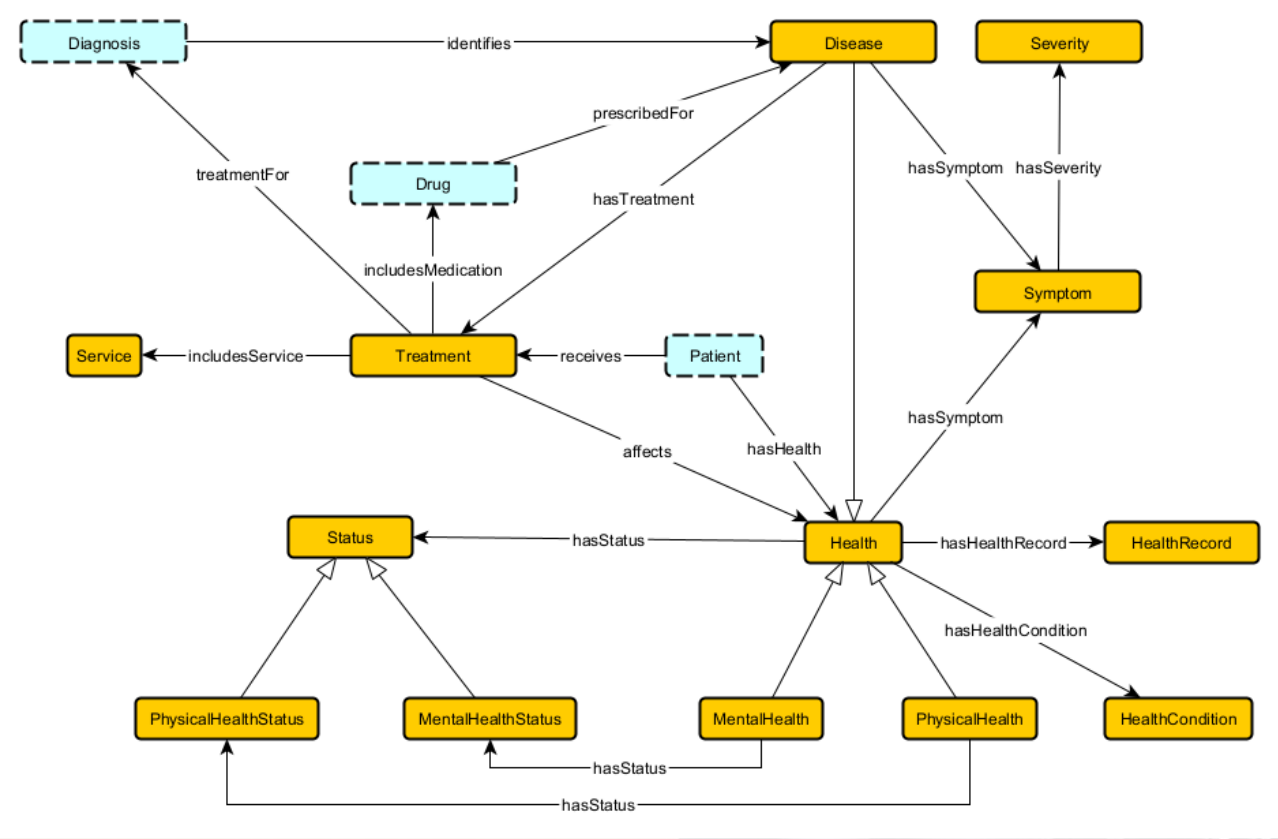
Metrics

Axiom	24
Logical axiom count	12
Declaration axioms count	12
Class count	7
Object property count	5
Data property count	0
Individual count	0
Annotation Property count	0

Class axioms

SubClassOf	7
EquivalentClasses	0
DisjointClasses	5
GCI count	0
Hidden GCI Count	0

Health



Ontology metrics:

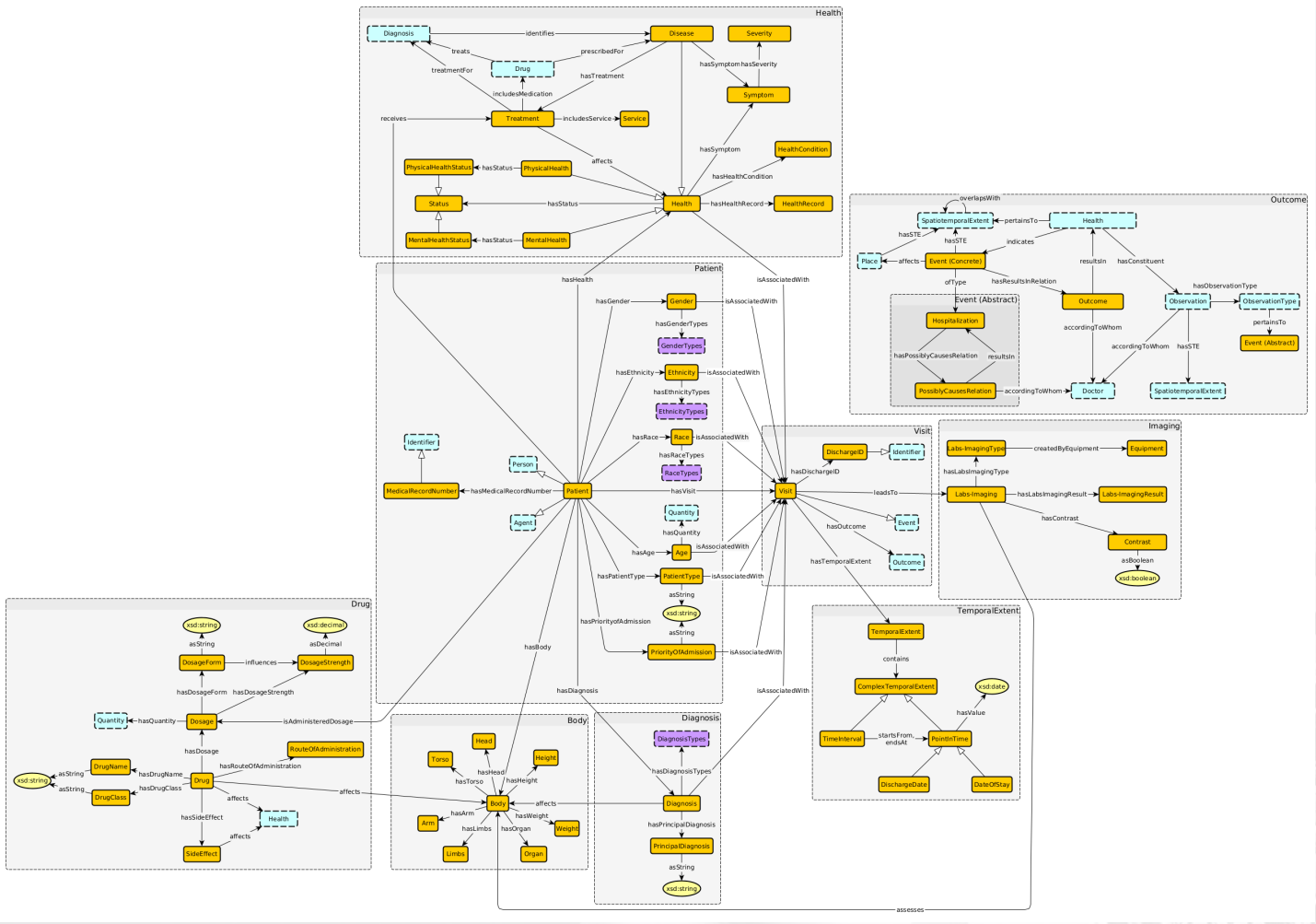
Metrics

Axiom	134
Logical axiom count	108
Declaration axioms count	26
Class count	16
Object property count	11
Data property count	0
Individual count	0
Annotation Property count	0

Class axioms

SubClassOf	94
EquivalentClasses	0
DisjointClasses	14
GCI count	20
Hidden GCI Count	0

Overall Graph



Ontology metrics:

Metrics

Axiom	605
Logical axiom count	490
Declaration axioms count	115
Class count	60
Object property count	55
Data property count	1
Individual count	0
Annotation Property count	0

Class axioms

SubClassOf	433
EquivalentClasses	0
DisjointClasses	57
GCI count	70
Hidden GCI Count	0

Competency Questions

- Has the incidence of Kawasaki disease decreased during the pandemic?
- How have trends in pediatric drowning changed over the past 5-10 years? Are certain populations affected disproportionately?
- Does a patient's sociodemographic factors affect outcomes of pediatric meningitis cases?
- How have trends in pediatric open globe injuries changed over the past 5-10 years? Are certain populations affected disproportionately?

Competency Questions

- **How many Asian people have been diagnosed with Kawasaki disease?**
- **In what race has Kawasaki disease been most prevalent? How many cases have there been?**
- **How many patients that have been diagnosed with Kawasaki disease have also been diagnosed with a metabolic condition?**

SPARQL Query

- Count of cases of Kawasaki for a specific race

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX dc-ont: <https://www.childrensdayton.org/lod/ontology/>

SELECT (COUNT(?patient) AS ?count)
WHERE {
  ?patient a dc-ont:Patient ;
           dc-ont:hasDiagnosis ?diagnosis ;
           dc-ont:hasRace ?raceEntity .
  ?diagnosis dc-ont:hasPrincipalDiagnosis ?principalDiagnosis .
  ?principalDiagnosis dc-ont:asString "Mucocutaneous Lymph Node Syndrome [kawasaki]"^^xsd:string .
  ?raceEntity dc-ont:raceAsian "Y" .
}
```

	count
1	"53"^^<http://www.w3.org/2001/XMLSchema#integer>

SPARQL Query

- Identification of race with most cases of Kawasaki and count of cases

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX dc-ont: <https://www.childrensdayton.org/lod/ontology/>

SELECT ?raceName (COUNT(?patient) AS ?count)
WHERE {
  ?patient a dc-ont:Patient ;
           dc-ont:hasDiagnosis ?diagnosis ;
           dc-ont:hasRace ?raceEntity .
  ?diagnosis dc-ont:hasPrincipalDiagnosis ?principalDiagnosis .
  ?principalDiagnosis dc-ont:asString "Mucocutaneous Lymph Node Syndrome [kawasaki]"^^xsd:string .

  # Match races where the value is "Y"
  ?raceEntity ?race "Y" .

  # Map race properties to human-readable labels
  VALUES (?race ?raceName) {
    (dc-ont:raceWhite "White")
    (dc-ont:raceBlack "Black")
    (dc-ont:raceAsian "Asian")
    (dc-ont:racePacificIslander "Pacific Islander")
    (dc-ont:raceAmericanIndian "American Indian")
    (dc-ont:raceOther "Other")
  }
}
GROUP BY ?raceName
ORDER BY DESC(?count)
LIMIT 1
```

	raceName	count
1	White	"79"^^<http://www.w3.org/2001/XMLSchema#integer>

Future Work

- Refine ontology to address errors and concerns noted in axiomatization and materialization phases
- Expansion of ontology to represent more concepts or better represent existing concepts
  - Separate Labs and Imaging
  - PatientType and PriorityOfAdmission as controlled vocabularies
- Expand or finetune use cases by incorporating more data
  - E.g., specific patient demographics, mental health information, etc
- Bridge gap between abstract concepts and instance-level data
  - Health pattern
  - Outcome pattern

Retrospective

- Anmol Saini
  - Pros
    - Greater insight and practice into process of creating a KG. Beginning to understand axioms
    - Flexibility in meeting schedule and tasking
    - Assigned topic with significant application
    - Application-oriented course
  - Cons
    - Suboptimal workload distribution over course of semester. Little time for materialization and axioms in Protégé
    - Limited feedback and guidance, especially later in semester
    - Somewhat disorganized with no grades assigned throughout the class

Retrospective

- Jason Nolte
  - Part of a great team. Anmol and Spencer were awesome
  - Helped create an ontology
    - Group discussions about the project, key notions, and axioms taught me a lot
    - Learned a bit about Protégé and axioms while entering the axioms in Protégé
    - Not sure I could explain what I did outside this class
  - Having an assigned topic helped
  - More interaction with the professor in the second half of the course would have been nice
    - Could teach some concepts/software closer to when used in the project
    - I understand the irony of saying this while virtual

Retrospective

- Spencer Seals
  - Our team was great
  - Flexibility in meeting schedule and tasking was great
  - More prepared to create and defend an ontology (semi) independently
    - Group discussions taught me a lot
    - Learned a lot more about the practical aspects of creating KGs
  - Distribution of workload over the semester wasn’t very uniform
  - Teaching some of the concepts closer to when they were used would be nice
  - Would like more exposure to reasoning algorithms and how they interact with the axioms

Thank you!

Competency Questions

- How have hospitalizations for motor vehicle vs bicycle injuries changed over the past 5-10 years? Does a patient's sociodemographic factors affect outcomes of hospitalization?
- Are there differences in surgical trauma outcomes between those who are trans vs cis?
- How have patterns of blood transfusions changed over the past 5-10 years? Has the use of whole blood increase over the past 5-10 years and if so in which kinds of cases are these?
- How often do pediatric trauma hospitalizations return to the emergency department within 28 days and are there factors that might help us predict these occurrences?