Ontology
as a sense-making technology

Riichiro Mizoguchi

JAIST: Japan Advanced Institute of Science and Technology

What I have been doing

- Ontology (theory)
  1. Theories of roles/functions/objects/processes/events
  2. Building an upper ontology **YAMATO**
- Ontological engineering (practice)
  1. Modeling functional structures of artifacts and its deployment into industry **(SOFAST/OntoloGear)**
  2. Theory-aware authoring tool based on an ontology of learning/instructional theories **(SMARTIES)**
  3. Building a disease ontology and its **publication as a linked data** with links to some existing ontologies
  4. etc.

Agenda

- Ontology and Ontology Engineering
- Light-weight ontology
  - SW, LD and LOD
    - How to build an ontology
      - Top-down vs. Bottom-up
    - Metadata, RDF, OWL
  - Linked Jazz (an LOD)
  - Open vocabularies
    - Dublin core, FOAF, LOV, Schema.org
    - Knowledge Graph
- Heavy-weight ontology
  - A case in a medical ontology

Cultural difference?

- Some say “Ontologies depend on culture”
- But, I believe although vocabularies are culturally different, meaning is not
- A good example of “a false friend” of Japan and China

手　紙

- Japan and China share Kanji letters
- 手 means “hand” and 紙 means “paper” in both
- But, the meanings of 手紙 are very different
- In Japan, it means “a letter” but “??” in China
- Imagine a couple of a Japanese boy and a Chinese girl, and he asks her “Send me your 手紙”
- She got angry!! Why?

What is an ontology?

- **In philosophy**: “Theory of being”

- **An ontology is an explicit specification of entities and relations in the target world intended to share within a community and to use for building a model of the target world**

  It usually specifies
- **HOW you VIEW the WORLD**

An example

Ontology of vehicle

A simple ontology

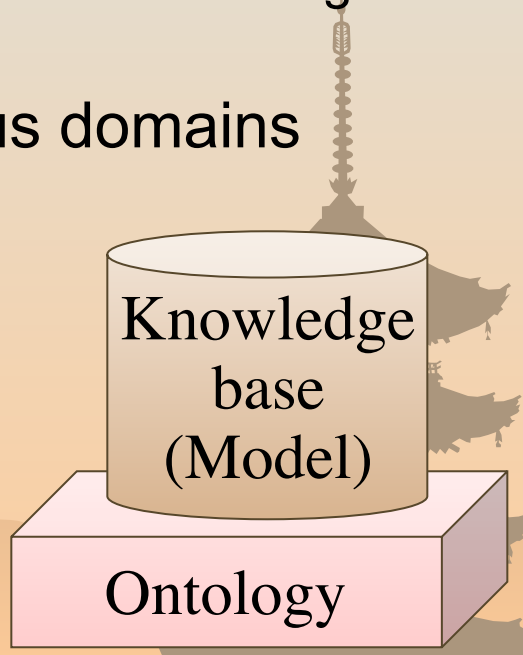
Vehicle
- ground vehicle
  - motor car
    - 4 or more wheel car
      - car
      - truck
    - motor bike
  - train
- ship
- air craft

An ontology of vehicle

- **Vehicle world**
  - Vehicle
    - ground vehicle
    - ship
    - air craft
  - function
    - to carry persons
    - to carry freights
  - attribute
    - power
    - size
  - component
    - engine
    - body
  - traffic system
  - ...

Ontology (Engineering) as Content Technology

- It enables you
  - to do consistent knowledge representation
    - by giving you an explicit viewpoint/guideline
    - Without consistency, computers cannot deal with knowledge properly
  - to share/reuse knowledge across various domains
  - to make pieces of knowledge interoperable
  - to build **ontology-aware** intelligent systems



Kinds of ontologies (not exhaustive)

- Upper ontology (Top-level ontology)
  - *DOLCE, BFO, UFO, YAMATO, etc.*
- Domain ontology
  - *Disease ontology, Gene ontology, etc.*
- Task ontology (coined by myself)
  - *Ontology of design, ontology of diagnosis, etc.*
- A system of vocabulary
  - *FOAF, Dublin core, etc.*

Dichotomy of ontology

- **Light-weight** Ontology
  - Something like FOAF, Dublin core, etc.
  - **Vocabulary** rather than concepts
  - Annotation-oriented ontology
  - Used as metadata for information search
  - **Practice-oriented**
- **Heavy-weight** Ontology
  - Something like an upper ontology
  - **Concepts** rather than vocabulary
  - for understanding the target world
  - for making what is left implicit explicit
  - for building ontology-aware system
  - **Philosophy-oriented**

Use/building of Ontology

Bottom-up

vs.

Top-down

Two approaches

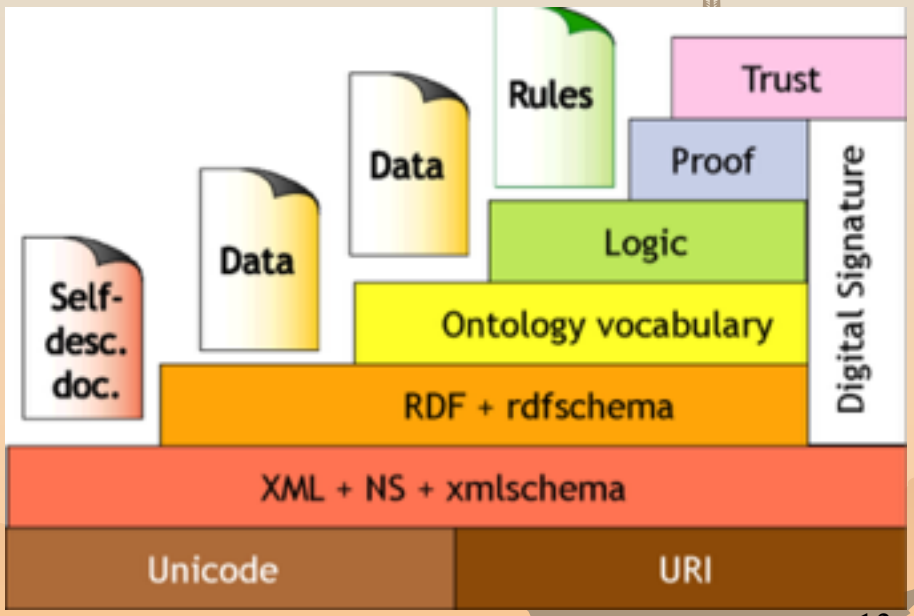
- **Bottom-up approach (Light-weight ontology)**
  - Policy
    - Distributed control (No centralized control)
  - Utilize existing various resources
    - How to find them?
  - Metadata
    - How to understand metadata?
  - Ontologies
    - How to make ontologies interoperate?
  - Ontology alignment
- **Top-down approach (Heavy-weight ontology)**
  - Policy
    - **Try to model the target world with a sophisticated ontology**
  - **FOIS conferences, Journal of Applied Ontology**

The Semantic Web

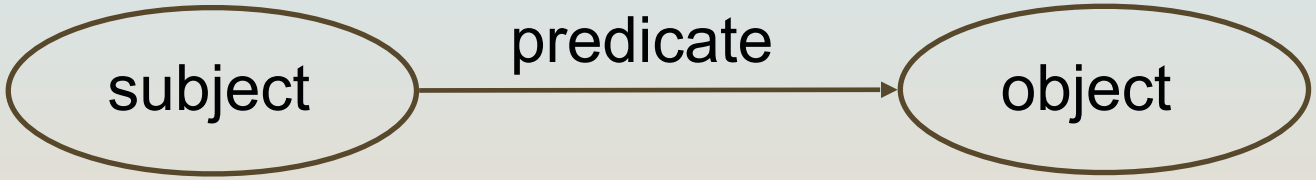
- Metadata needs ontology to make sense of them
- HTML-based Web
  - Links have **no semantics**, and they just allow you to jump to the specified page

- RDF-based Web including LD/LOD is different
- Links have **meanings** such as *is-member-of, is-capital-of, manufactured-by, etc.*

Copied from @W3C



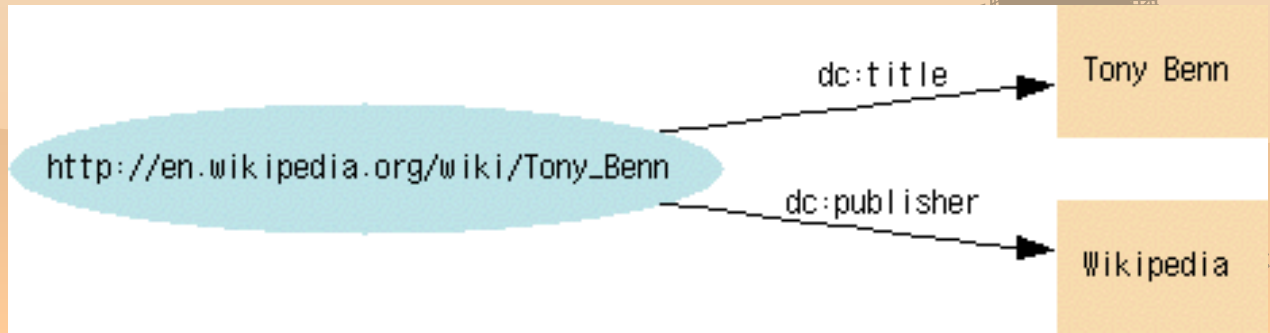
RDF (Resource Description Framework)




```
<rdf:RDF
 xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
 xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description rdf:about="http://en.wikipedia.org/wiki/Tony_Benn">
    <dc:title>Tony Benn</dc:title>
    <dc:publisher>Wikipedia</dc:publisher>
  </rdf:Description>
</rdf:RDF>
```

URI

URI

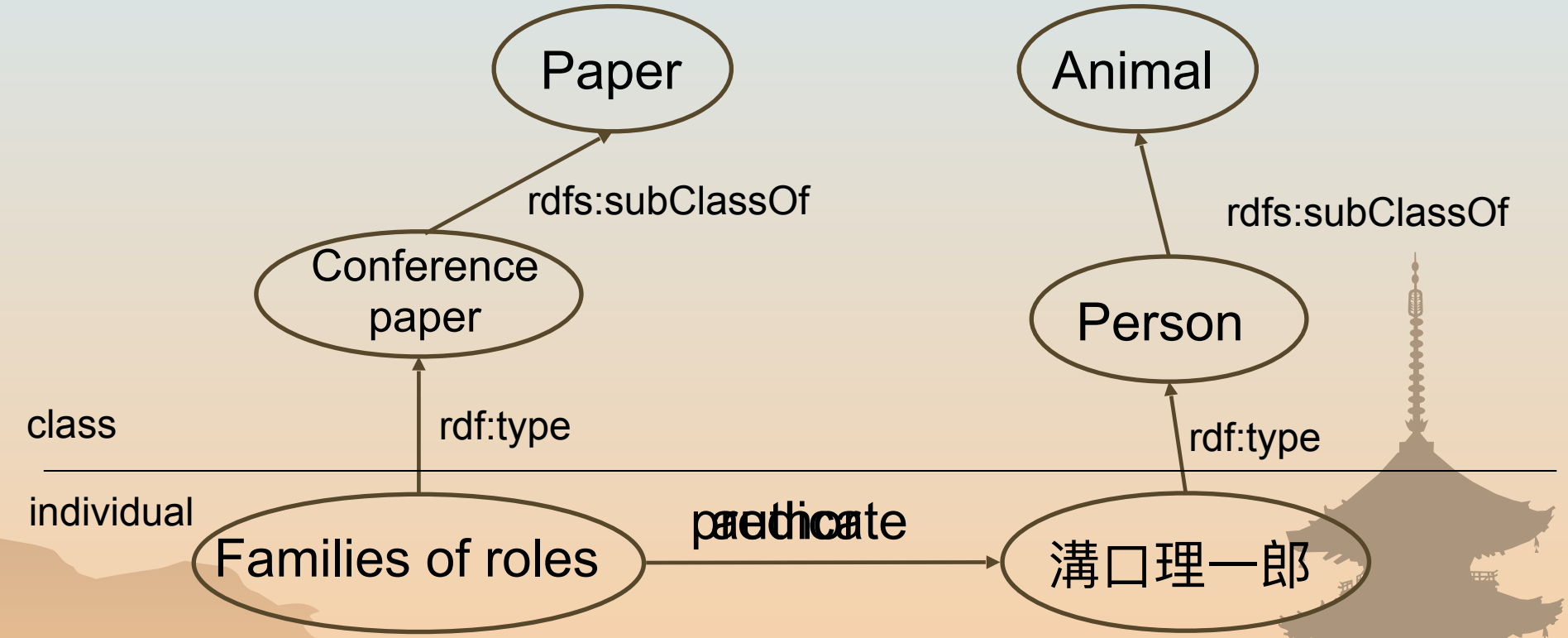


What is sense-making?

- Ontology plays a role of sense-making technique for semantic processing by giving “meaning” of terms used in SW/LD/LOD

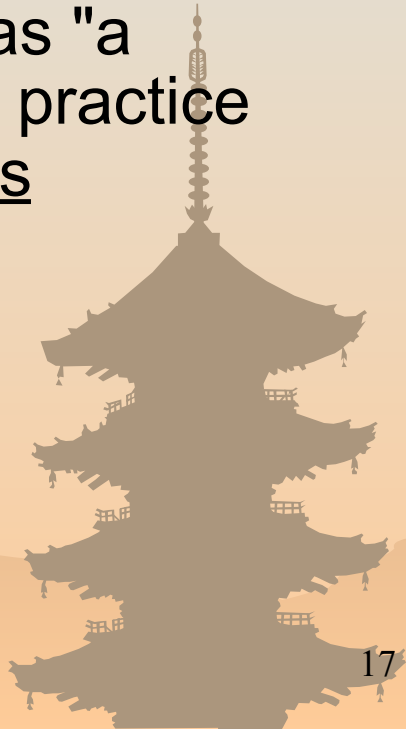
1. Identification of the ***type/class*** of the **individual** under consideration
2. Identification of ***is-a*** hierarchy of those types/ classes
3. An upper ontology can explain what exist out there

An example of sense-making



LD/LOD

- Linked Data is about using the Web to **connect related data** that wasn't previously linked, or using the Web to lower the barriers to linking data currently linked using other methods. More specifically, Wikipedia defines Linked Data as "a term used to describe a recommended best practice for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web **using URIs and RDF**
- **Excerpt from** http://linkeddata.org/



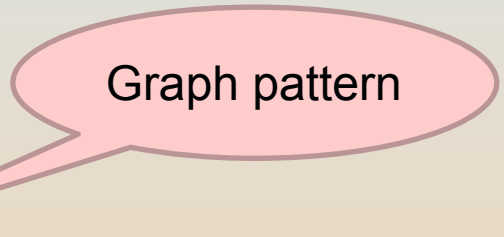
Four conditions

1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI, provide **useful information**, using the standards (RDF, SPARQL)
4. Include links to other URIs so that they can discover more things

*What data are there must be understood by computer, which is enabled by **SPARQL endpoint** supported by an ontology*

SPARQL

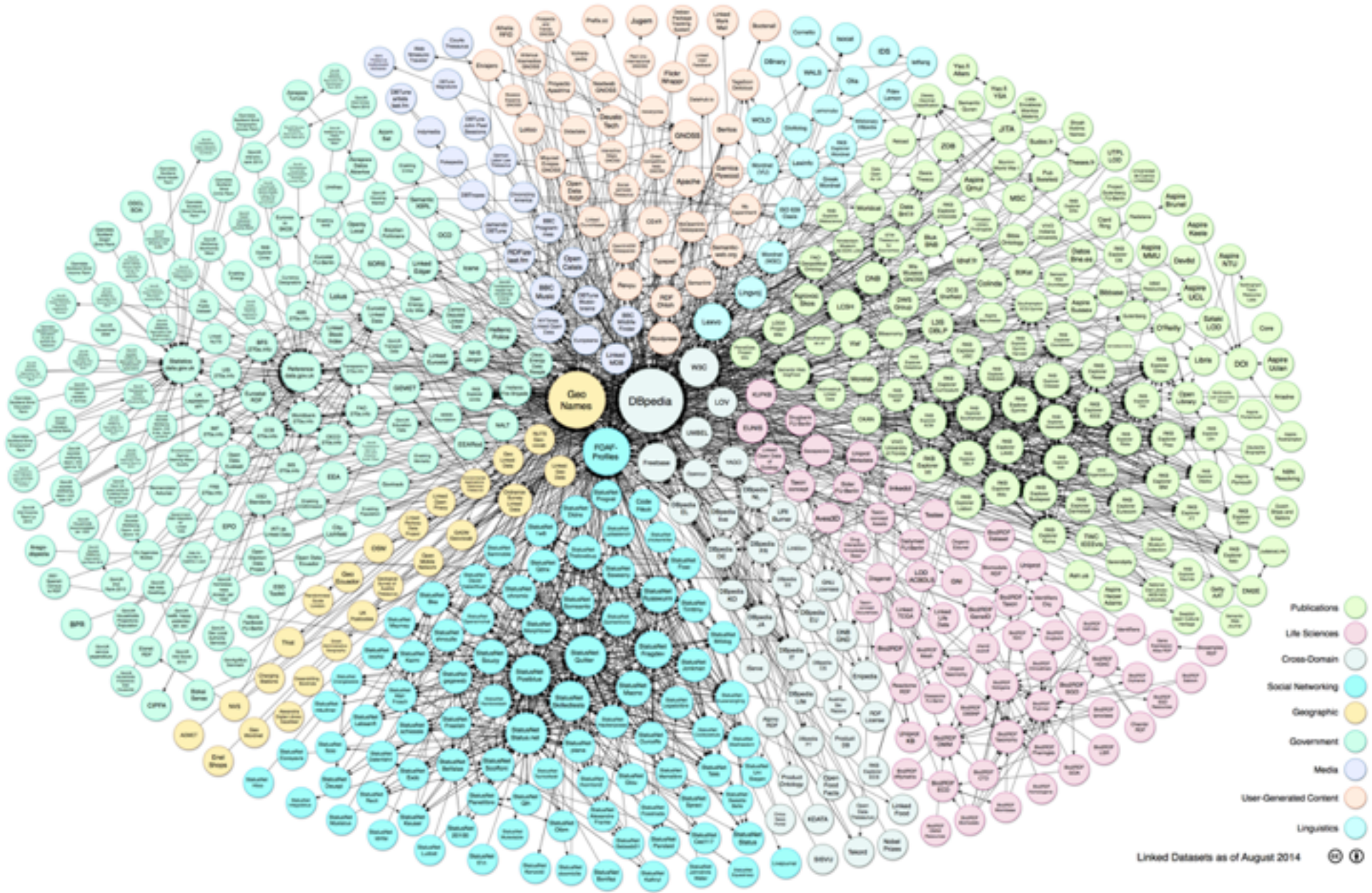
- Query language for RDF




```
SELECT ?s/p/v
WHERE {
?s ?p ?v .
}
```

```
SELECT ?title
WHERE {
<http://example.org/book/book1> <http://purl.org/dc/elements/1.1/title> ?title .
}
```

The LOD cloud diagram



Dados Abertos Conectados

by Seiji and Ig Ibert

- *http://ceweb.br/publicacao/livro-dados-abertos/*
- http://ceweb.br/publicacao/livro-dados-abertos/

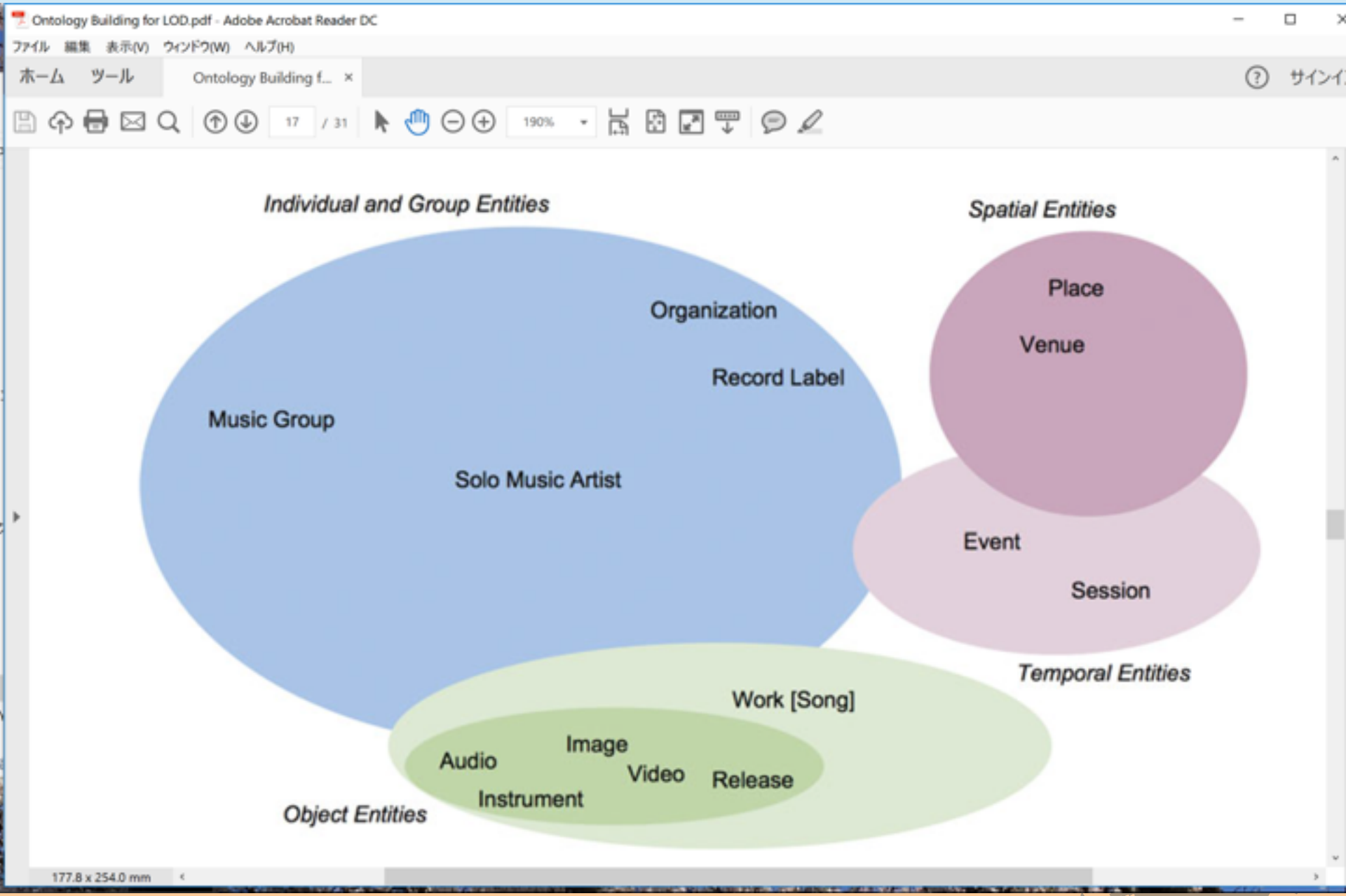
- A tool for developing LDs is also provided

- It stresses the importance of ontology to develop LD/LOD

Light-Weight ontology used in LD/LOD

- Linked **Jazz LOD**
- https://linkedjazz.org/
- https://linkedjazz.org/
- Ontology Building for Linked Open Data:
  A Pragmatic Perspective
- http://dx.doi.org/10.1080/19386389.2015.1099979

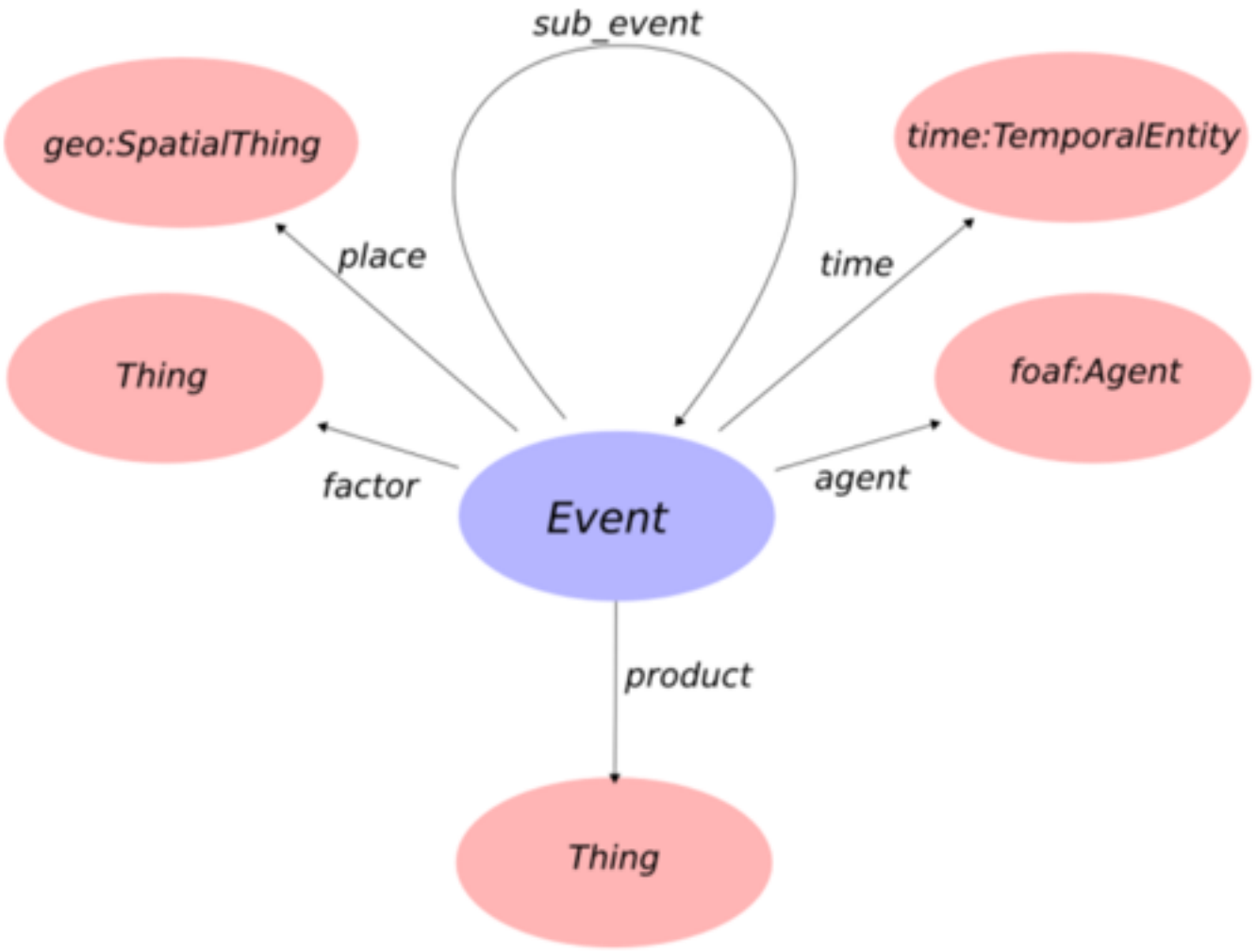
Entity types in Linked Jazz



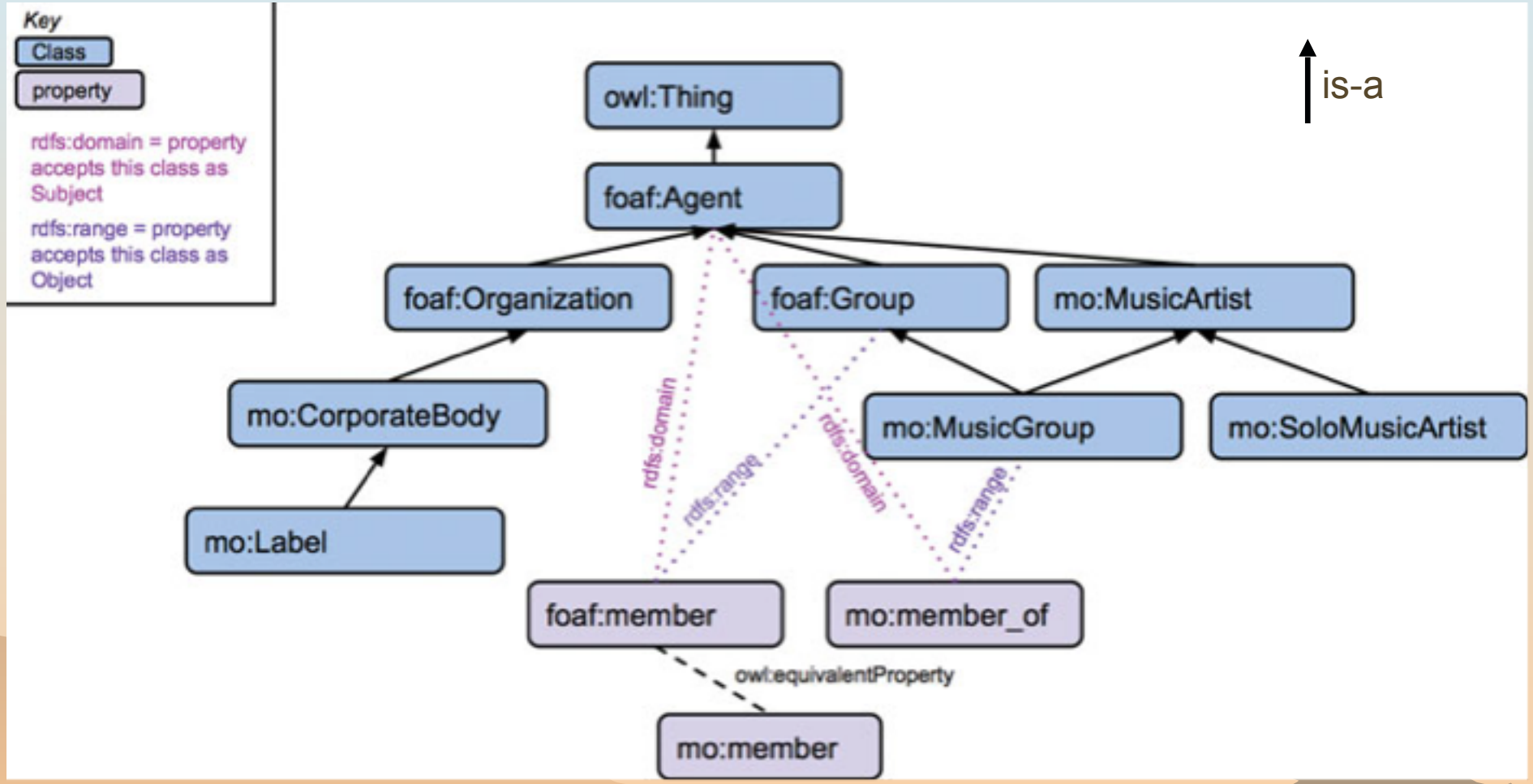
Event ontology

http://motools.sourceforge.net/event/event.html

active agents, factors and products, as depicted below.



Portion of the ontology in Linked Jazz LOD

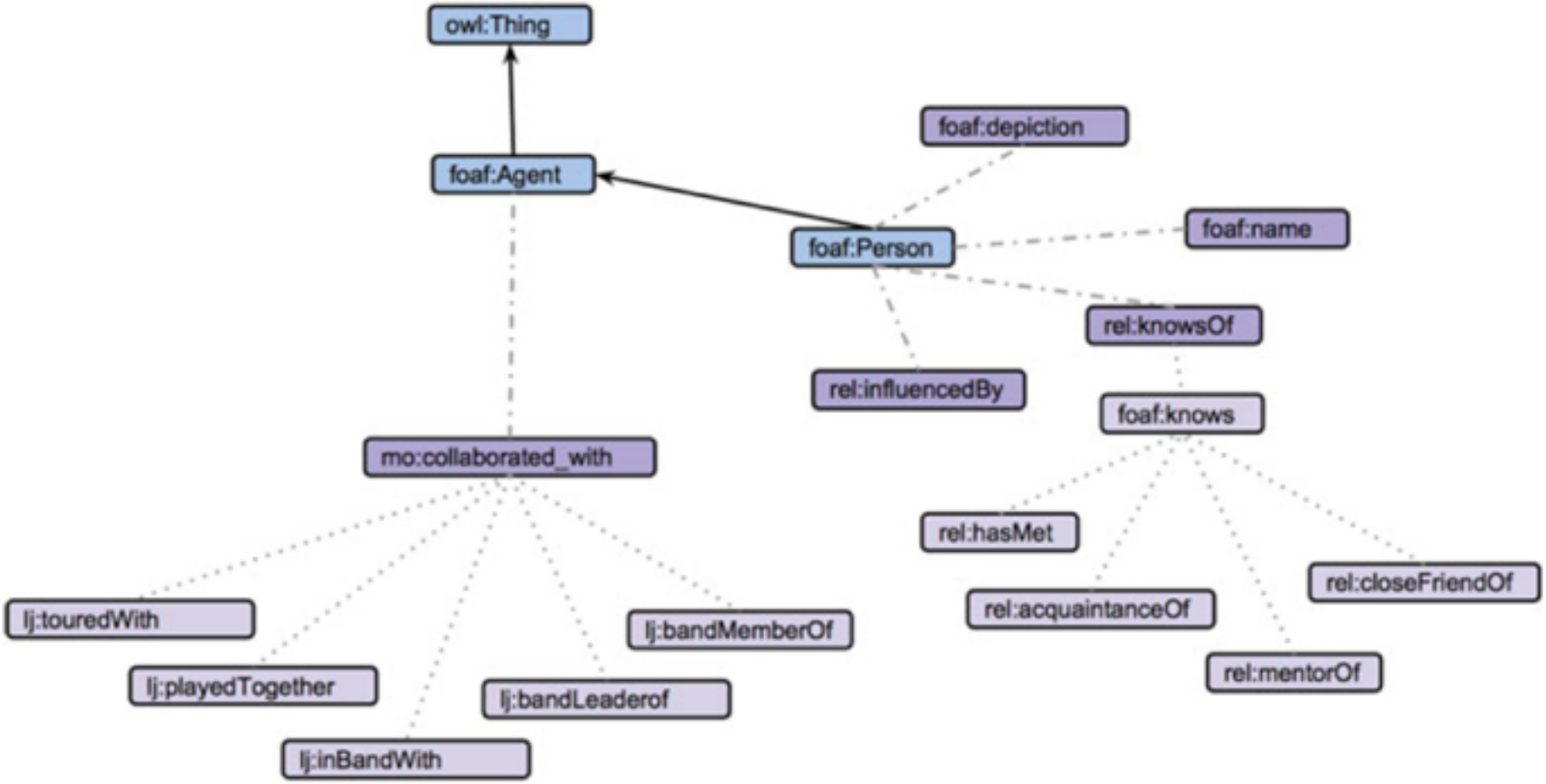


Excerpt from Journal of Library Metadata, 15:3-4, 265-294, DOI: 10.1080/19386389.2015.1099979

Properties

- *rel:hasMet,*
- *rel:acquaintanceOf,*
- *rel:closeFriendOf,*
- *rel:influencedBy,*
- *rel:mentorOf.*
- *lj:touredWith,*
- *lj:playedTogether,*
- *lj:inBandWith,*
- *lj:bandLeaderOf,*
- *lj:bandMemberOf,*

Specialization of properties



Excerpt from Journal of Library Metadata, 15:3-4, 265-294, DOI: 10.1080/19386389.2015.1099979

FOAF (Friend Of A Friend)

- **Classes**:
  | Agent | Document | Group | Image | LabelProperty | OnlineAccount | OnlineChatAccount | OnlineEcommerceAccount | OnlineGamingAccount | Organization | **Person** | PersonalProfileDocument | Project |

- **Properties**:
  | account | accountName | accountServiceHomepage | age | aimChatID | based_near | birthday | currentProject | depiction | depicts | dnaChecksum | familyName | family_name | firstName | focus | fundedBy | geekcode | gender | givenName | givenname | holdsAccount | homepage | icqChatID | img | interest | isPrimaryTopicOf | jabberID | **knows** | lastName | logo | made | maker | mbox | mbox_sha1sum | member | membershipClass | msnChatID | myersBriggs | name | nick | openid | page | pastProject | phone | plan | primaryTopic | publications | schoolHomepage | sha1 | skypeID | status | surname | theme | thumbnail | tipjar | title | topic | topic_interest | weblog | workInfoHomepage | workplaceHomepage | yahooChatID |

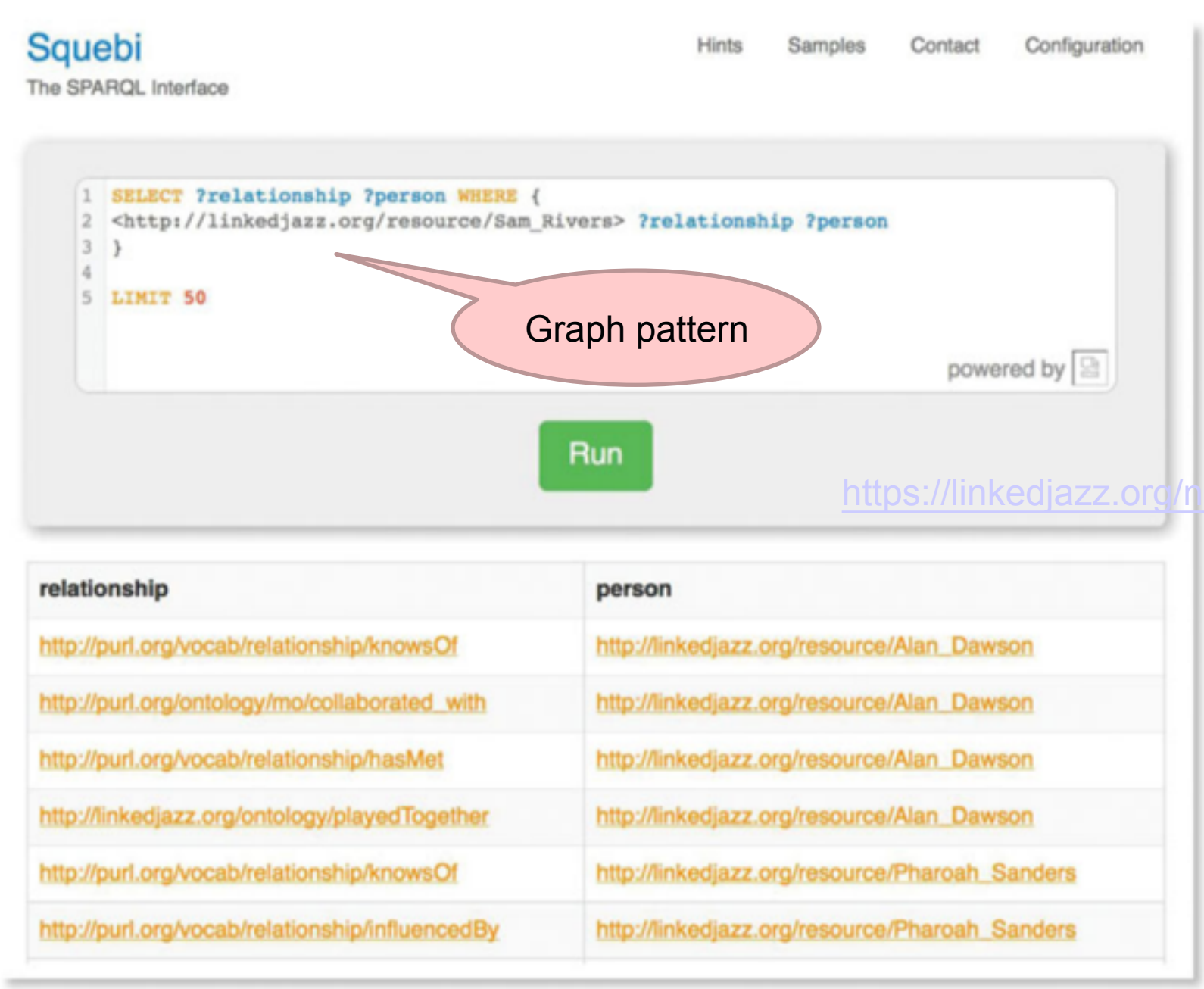


FIGURE 10 SPARQL query returning Sam Rivers's relationships.

LOV: Linked Open Vocabulary

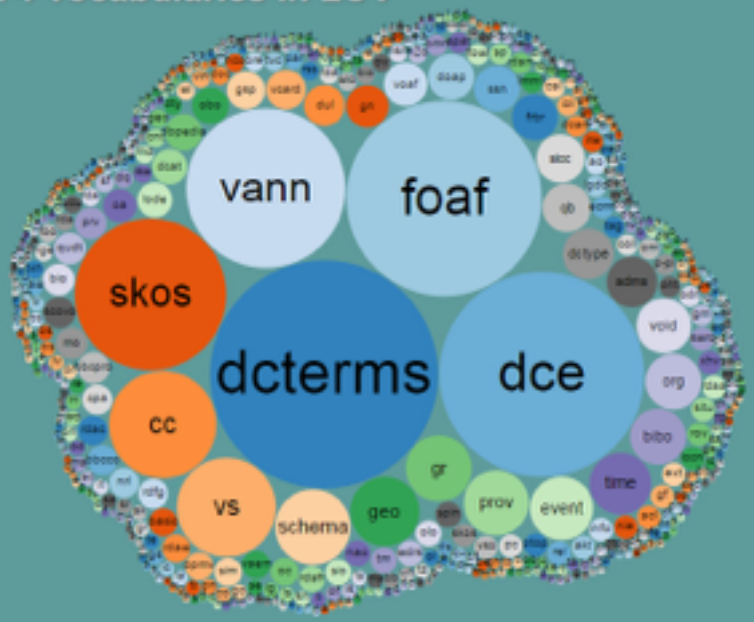
- Maintained by the Open Knowledge Foundation (OKF)
- http://lov.okfn.org/dataset/lov/
- http://lov.okfn.org/dataset/lov/

VOCABS TERMS AGENTS SPARQL/DUMP

Linked Open Vocabularies (LOV)

+ Suggest | Documentation | Follow | 🔍 | 🐱

574 Vocabularies in LOV



Category Tags

Methods | Metadata | Catalogs | Geography | API | Society | Support | Quality | RDF | People | Services | Vocabularies | General & Upper | Time | Environment | Industry | Multimedia | Geometry | Events | Biology | FRBR | W3C Rec | SPAR | Government | PLM | Academy | eBusiness | Tag | Travel | Security

Latest insertion

trao - Transport Administration Ontology
2016-09-09

cfrl - Corporate Financial Reports and Loans Ontology
2016-09-09

duv - Dataset Usage Vocabulary
2016-09-09

hosp - Hospital Vocabulary
2016-09-09

dbm - DBM Ontology
2016-09-09

Latest Updates

dbpedia-owl - The DBpedia Ontology
2016-09-12

duv - Dataset Usage Vocabulary
2016-09-09

cfrl - Corporate Financial Reports and Loans Ontology
2016-09-09

veo - Vehicle Emissions Ontology
2016-09-09

trao - Transport Administration Ontology

Dublin core

http://dublincore.org/

http://www.kanzaki.com/docs/sw/dublin-core.html#dcterms

DCMI Type Vocabulary	Collection , Dataset , Event , Image , InteractiveResource , MovingImage , PhysicalObject , Service , Software , Sound , StillImage ,Text
Properties in the/ elements/1.1/ namespace **(DC elements:15)**	**contributor , coverage , creator , date , description , format , identifier , language , publisher , relation , rights , source , subject, title, type**
Properties in the / terms/namespace **(DC terms:55)**	abstract , accessRights , accrualMethod, accrualPeriodicity, accrualPolicy , alternative, audience, available , bibliographicCitation ,conformsTo , contributor , coverage , created , creator , date , dateAccepted , dateCopyrighted , dateSubmitted , description ,educationLevel , extent , format , hasFormat , hasPart , hasVersion , identifier , instructionalMethod , isFormatOf , isPartOf , isReferencedBy , isReplacedBy , isRequiredBy , issued , isVersionOf , language , license , mediator , medium , modified , provenance, publisher , references , relation , replaces , requires , rights , rightsHolder , source , spatial , subject , tableOfContents , temporal , title , type , valid
Classes	Agent, AgentClass, BibliographicResource, FileFormat, Frequency, Jurisdiction, LicenseDocument, LinguisticSystem, Location,LocationPeriodOrJurisdiction, MediaType, MediaTypeOrExtent, MethodOfAccrual, MethodOfInstruction, PeriodOfTime,PhysicalMedium, PhysicalResource, Policy, ProvenanceStatement, RightsStatement, SizeOrDuration , Standard

Google Knowledge Graph

derived from many sources, including
the CIA World *Factbook*, *Wikidata,* and *Wikipedia*

1. Find the right thing

   [Taj Mahal]: *a monument in India, a musician, casino in Atlantic City, etc.*

   It contains **500 M** objects, as well as more than **3.5 B** facts

2. Get the best summary
3. Go deeper and broader

Schema.org

- ***Schema.org*** is a collaborative, community activity with a mission to create, maintain, and promote ***schemas*** for structured data on the Internet, on web pages, in email messages, and beyond
- Over **10 M** sites use ***Schema.org*** to markup their web pages
- Essentially, a set of vocabulary for metadata of web sites

schema.org types

- Creative works:
  CreativeWork, Book, Movie, MusicRecording, Recipe, TVSeries ...
- Embedded non-text objects:
  AudioObject, ImageObject, VideoObject
- *Event*
- *Organization*
- ***Person***
- *Place, LocalBusiness, Restaurant ...*
- *Product, Offer, AggregateOffer*
- *Review, AggregateRating*

Properties of Person
(not exhaustive)

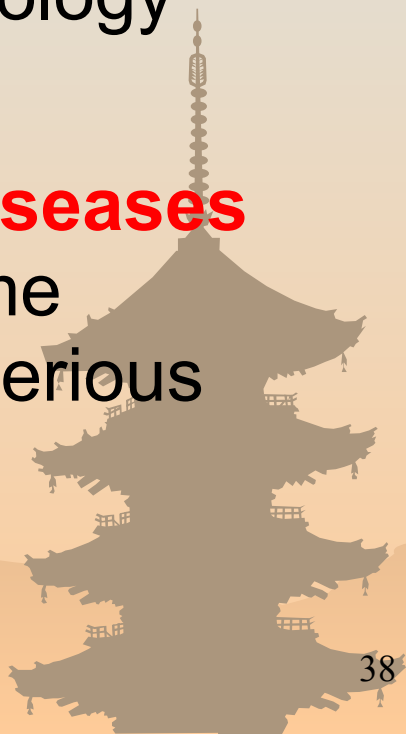
Property	Expected Type	Description
Properties from Person		
additionalName	Text	An additional name for a Person, can be used for a middle name.
address	PostalAddress or Text	Physical address of the item.
affiliation	Organization	An organization that this person is affiliated with. For example, a school/ university, a club, or a team.
alumniOf	EducationalOrganization or Organization	An organization that the person is an alumni of. Inverse property: alumni.
award	Text	An award won by or for this item. Supersedes awards.
birthDate	Date	Date of birth.
birthPlace	Place	The place where the person was born.
brand	Brand or Organization	The brand(s) associated with a product or service, or the brand(s) maintained by an organization or business person.
children	Person	A child of the person.
colleague	Person or URL	A colleague of the person. Supersedes colleagues.

Summary of what we have discussed thus far

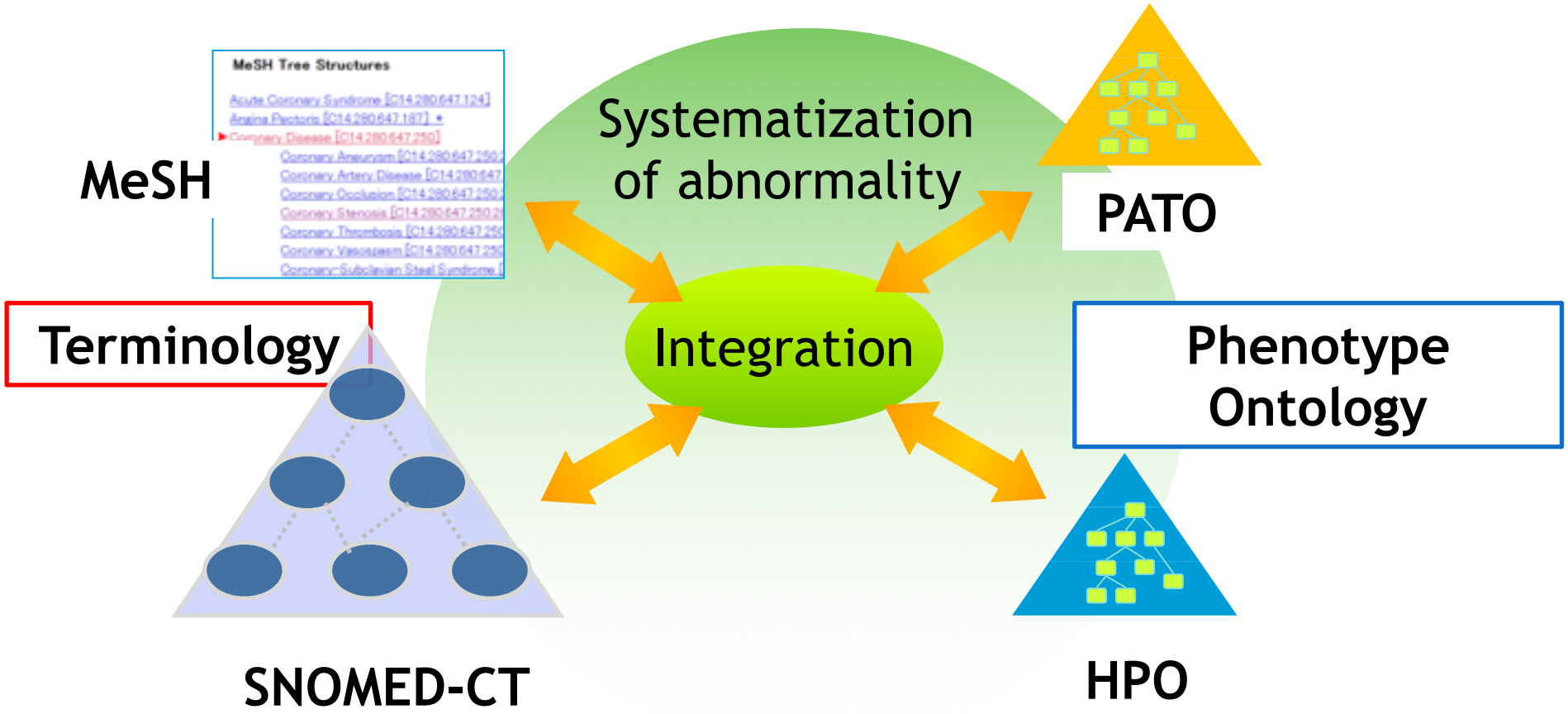
- Ontology is necessary for LD/LOD
- It helps you to **make sense of data** in it
- A lot of vocabularies are available
  - You reuse what you need among them
- You don't need heavy-weight ontology
- **Really?**
- **Of course, NOT!**

A case of a medical ontology

- On the WWW, several medical ontologies exist
- They are based on their own upper ontology
  - Not **interoperable**
- **For example, abnormal states** and **diseases** are considered as belonging to the same category, say, disorder, which causes serious problems



Linking existing resources via ontology of abnormal states



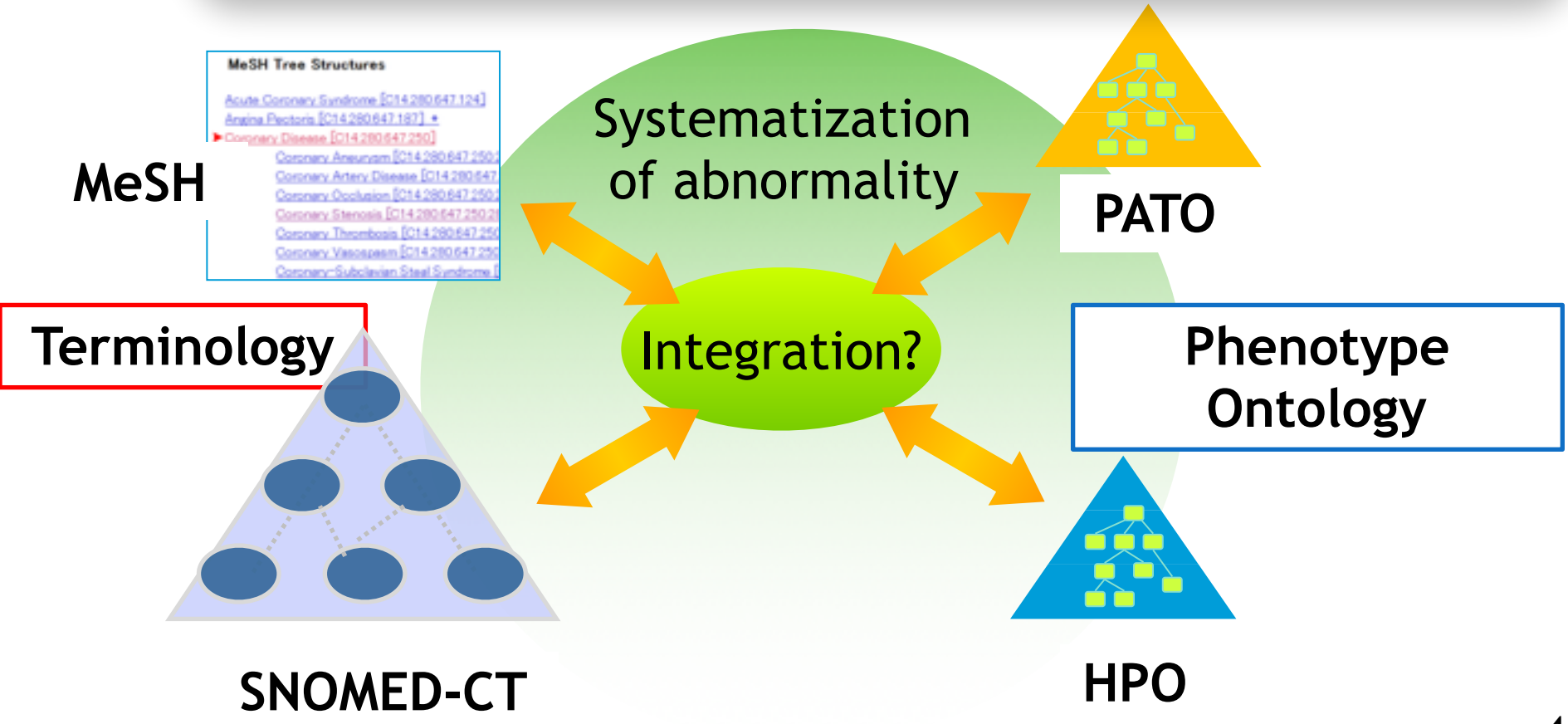
Existing resources

- **SNOMED-CT** (Systematized Nomenclature Of Medicine−Clinical Terms)
  - Multiple inheritance
  - Abnormal states are mixed up with diseases
- **MeSH** (Medical Subject Headings)
  - Multiple inheritance
  - Abnormal state is subsumed by disease class
- **PATO** (Phenotypic Quality Ontology)
  - Multiple inheritance
  - Property-oriented <E,P>, not compatible with <E,A,V>
- HPO (Human Phenotype Ontology)
  - Organized according to the body structure
  - Abnormal states a
  - No attribute value

正しいis-aもありますので，misuse of is-a relation などにとどめておくほうがよいかもと思いました.

Issues of integration

1. Granularity of concepts to be linked
2. Abnormal states are mixed up with diseases
3. Incompatibility between property-orientated <E,P> and Attribute-orientated <E,A,V> representation



Sense making of…

- State vs. Property
- Property vs. Attribute
- Disease
  - abnormal state vs. disposition vs. causal chain

Property and State

"A state is a time-indexed property"

Tall and hungry

- John is tall
  - John possesses tall property which inheres in John
- John is hungry
  - John is in a hungry state in which John participates (at time t)

- <John, tall> = <John, height, large>
  <E, P> <E, A, V>
  - tall = <height, large>
  - **P**roperty = <**A**ttribute, **V**alue>
- <John, hungry> = <John, (degree of) hunger, large>
  - Hungry = <(degree of) hunger, large>

Ontology of Property/Quality

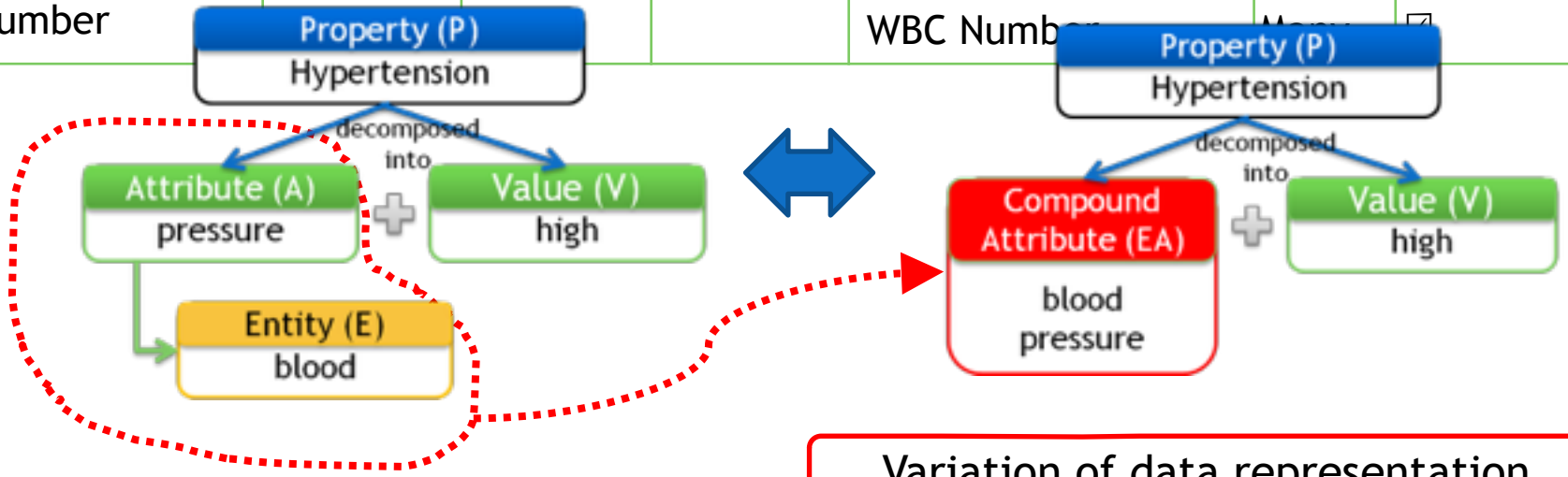
- **BFO** = <E, P>-based
- **DOLCE** = <E, A, V>-based

- **YAMATO** = Integration of both and realizes interoperability between them

Representation of abnormal states

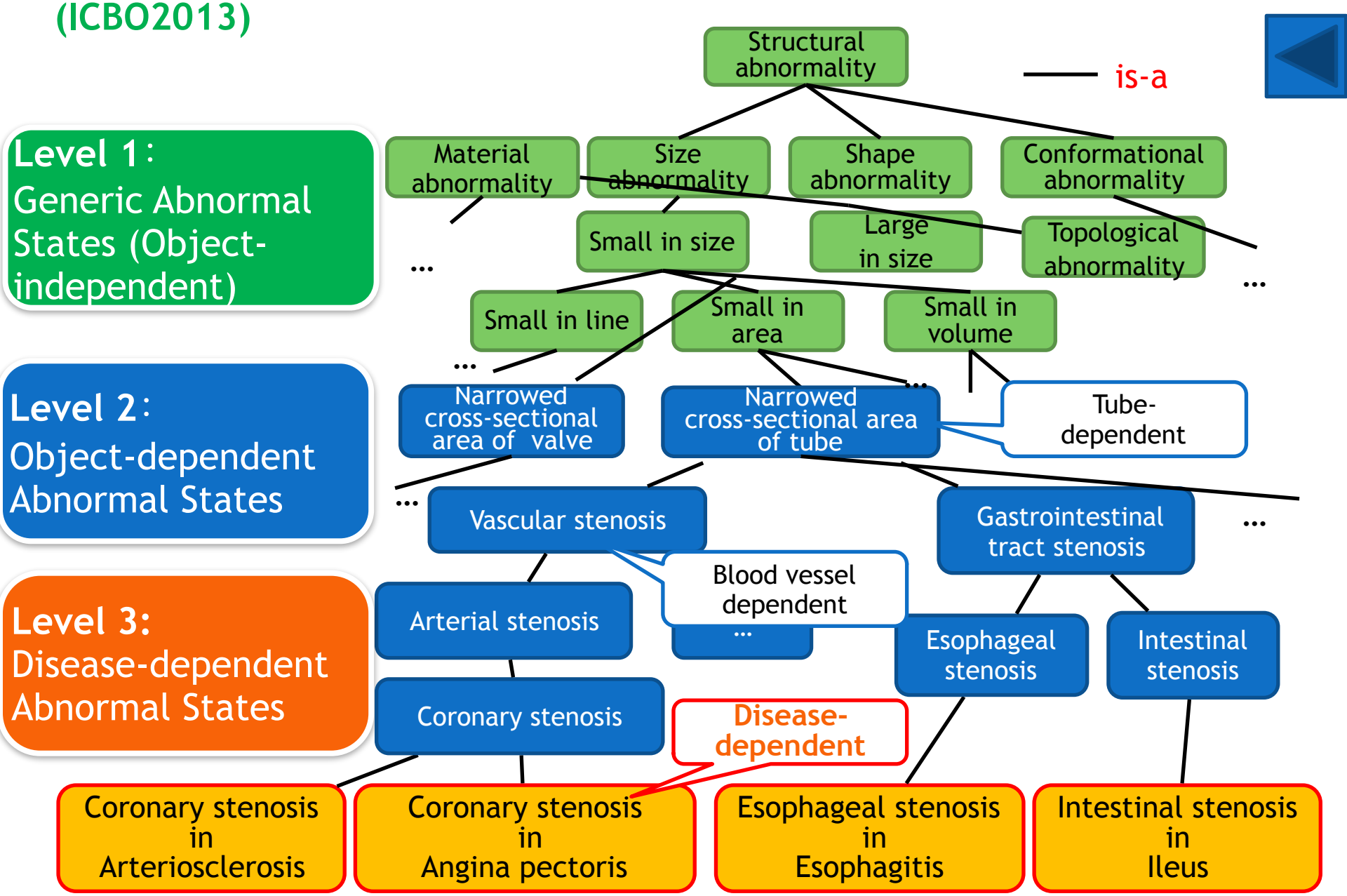
Abnormal states		E	Se	A	V	Solved
Nausea	EAV	Human		Nausea (non-decomposable)		☑
Chest pain	ESeAV	Human	Chest	Pain (non-decomposable)		☑
Arterial stenosis	EAV	Artery		Cross-section Area	Small	
Hypertension	EAV	Blood		Pressure	High	
	EA,V			Blood pressure	High	☐
Hyperglycemia	ESeAV	Blood	Glucose	Concentration	High	☑
	ESeA,V			Blood Glucose Level	High	☑
Increased WBC number	ESeAV	Blood	WBC	Number	Many	☑
				WBC Number		

Compound attribute



Variation of data representation

Is-a hierarchy of abnormal state ontology-1
(ICBO2013)



Abnormality

- We don’t go into philosophical discussion about what is abnormal

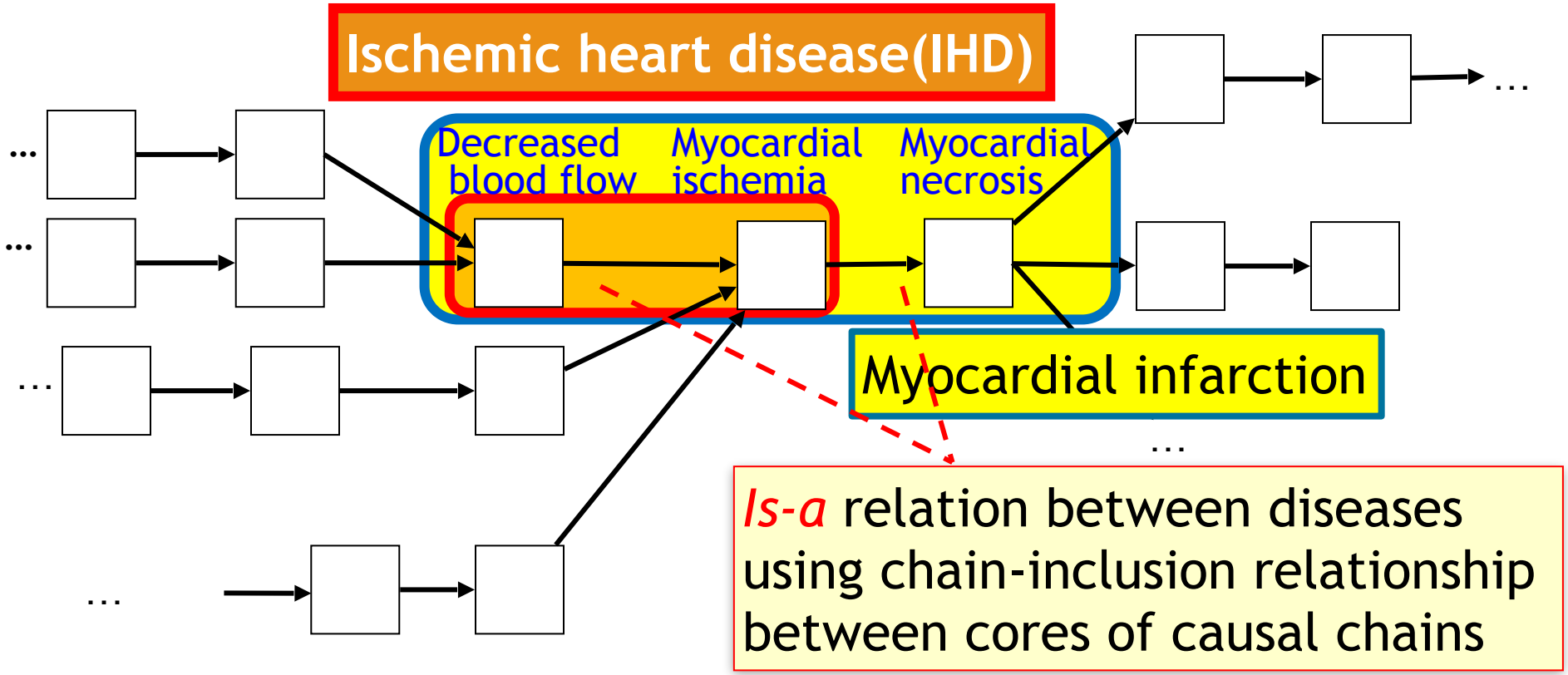
- We consider any states ***referred to in definitions of diseases*** as abnormal states

What is a disease?

- Abnormal state?
- Disposition?
  - *A disease is a disposition (i) to undergo pathological processes that (ii) exist in an organism because of one or more disorders in that organism (excerpt from OGMS paper)*
  - *A disposition has manifestation which corresponds to disease course*
- Causal chain of abnormal states?

An Example of Causal Chains of Ischemic heart disease in RFM (ICBO2012)

•A disease is *constituted of* one or more causal chains of disorders (abnormal states)



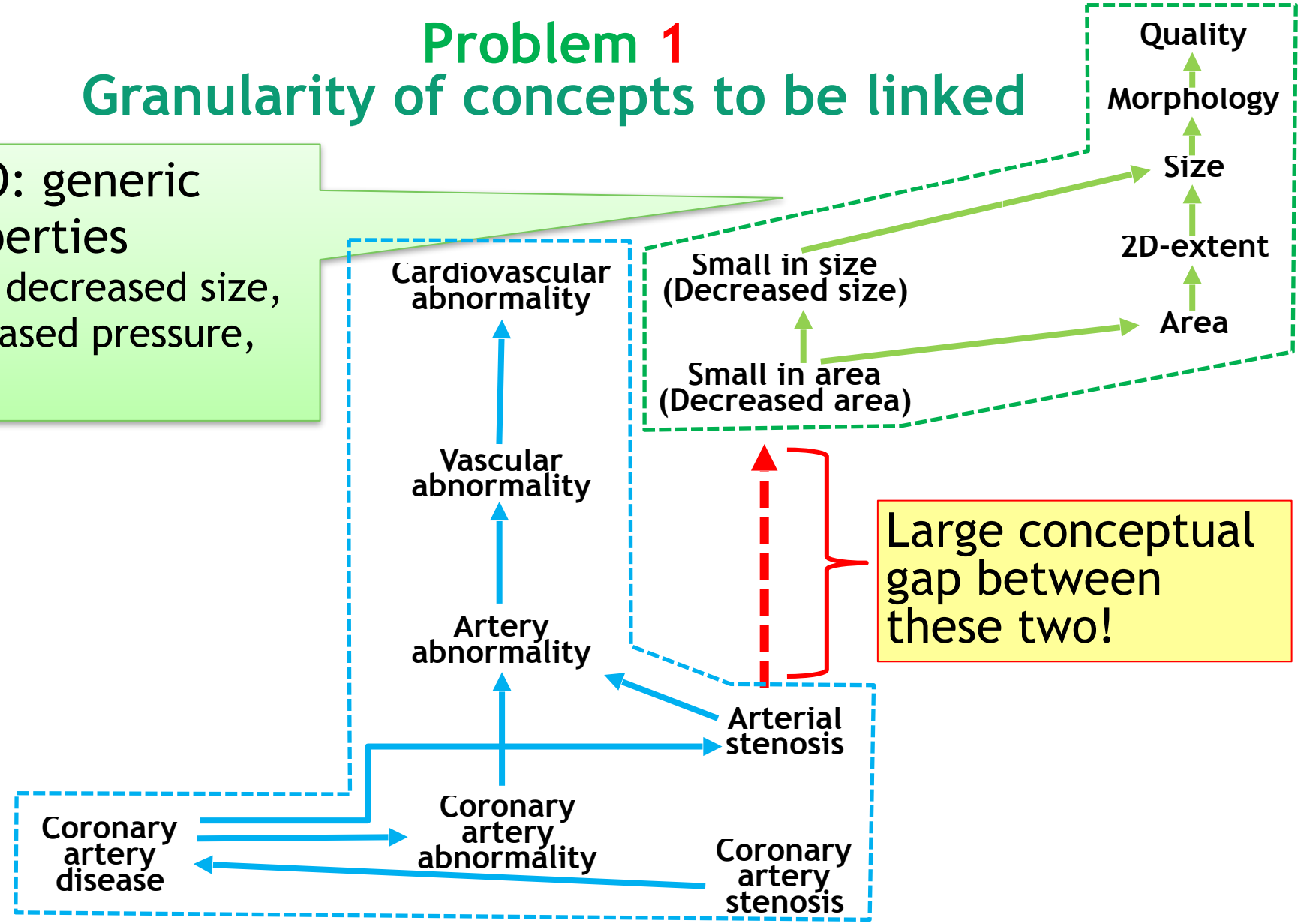
A disease is not an abnormal state

- A disease is constituted of (causal chains of) abnormal states
- A disease is a **whole** whose **constituents** are abnormal states
- Even if a disease is defined in terms of **one abnormal state**, it doesn't mean the disease and the abnormal state is equivalent

Problems and solutions in linking data

Problem 1
Granularity of concepts to be linked

PATO: generic properties
e.g., decreased size, increased pressure, etc.



Large conceptual gap between these two!

is-a relation
PATO HPO

HPO: human-specific abnormality

No existing ontologies can resolve this problem

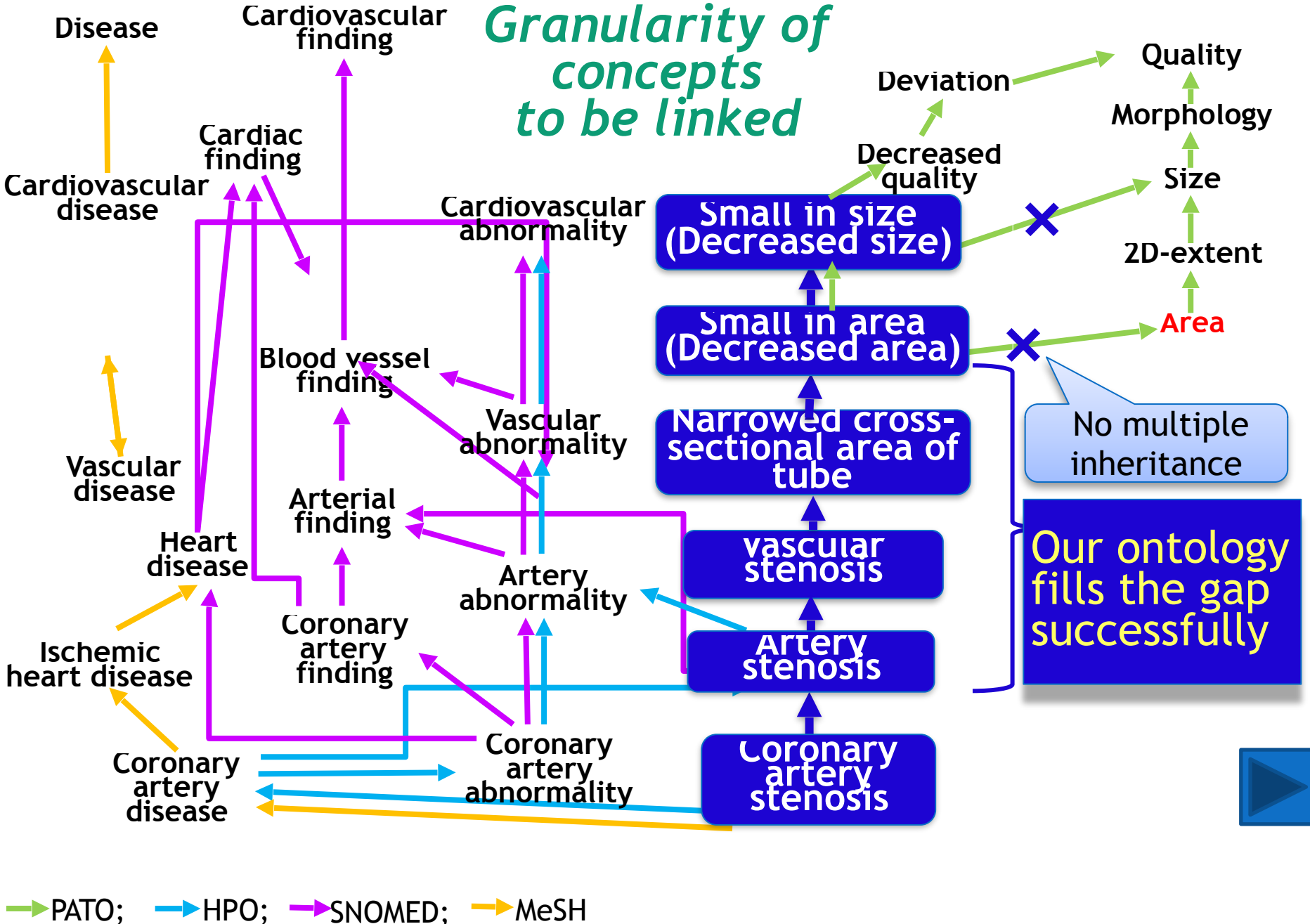
Various categories are mixed up:

e.g. Findings, Diseases, abnormal states…

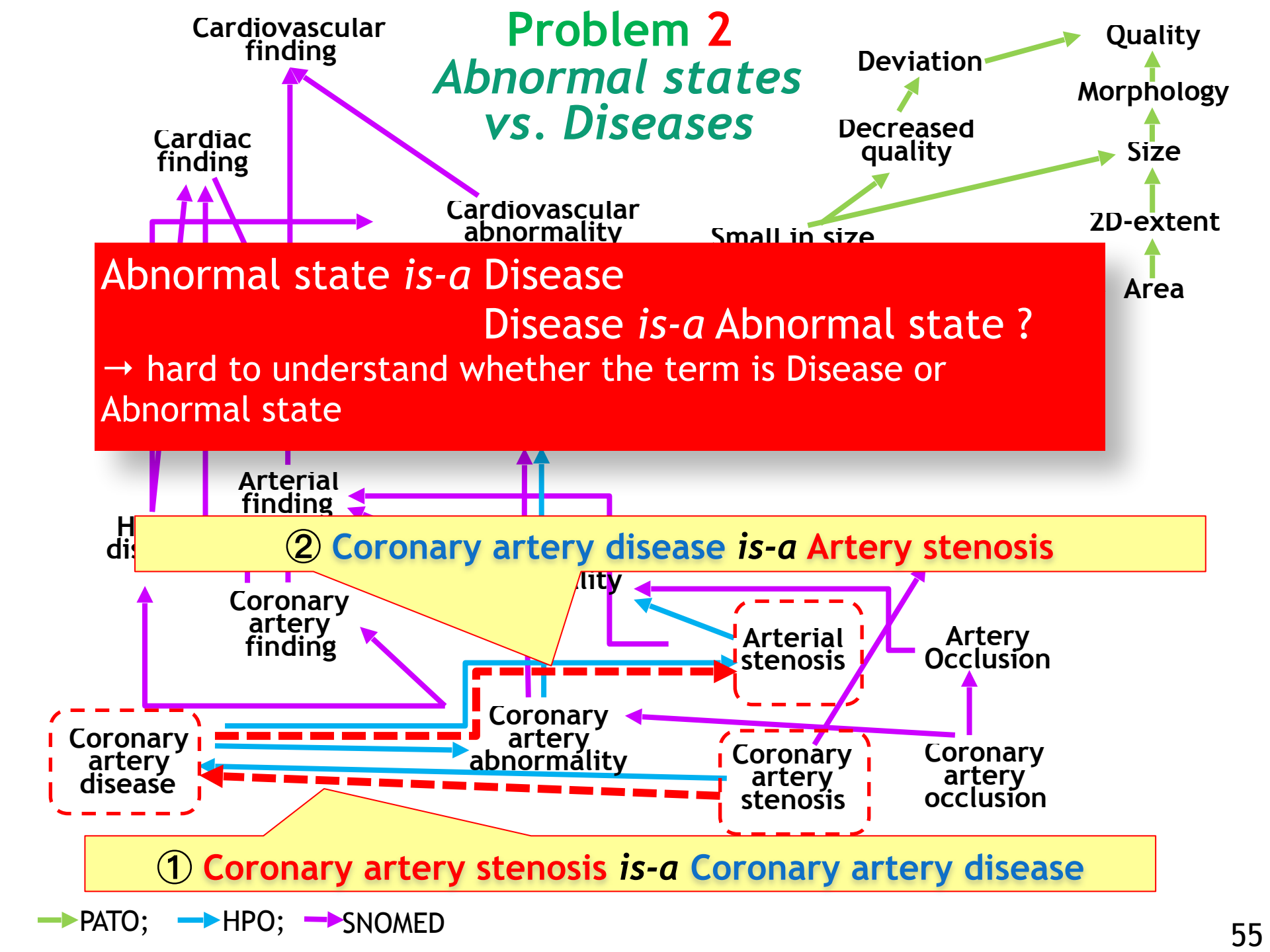
Multiple inheritance

→ hard to understand the intrinsic nature of concepts

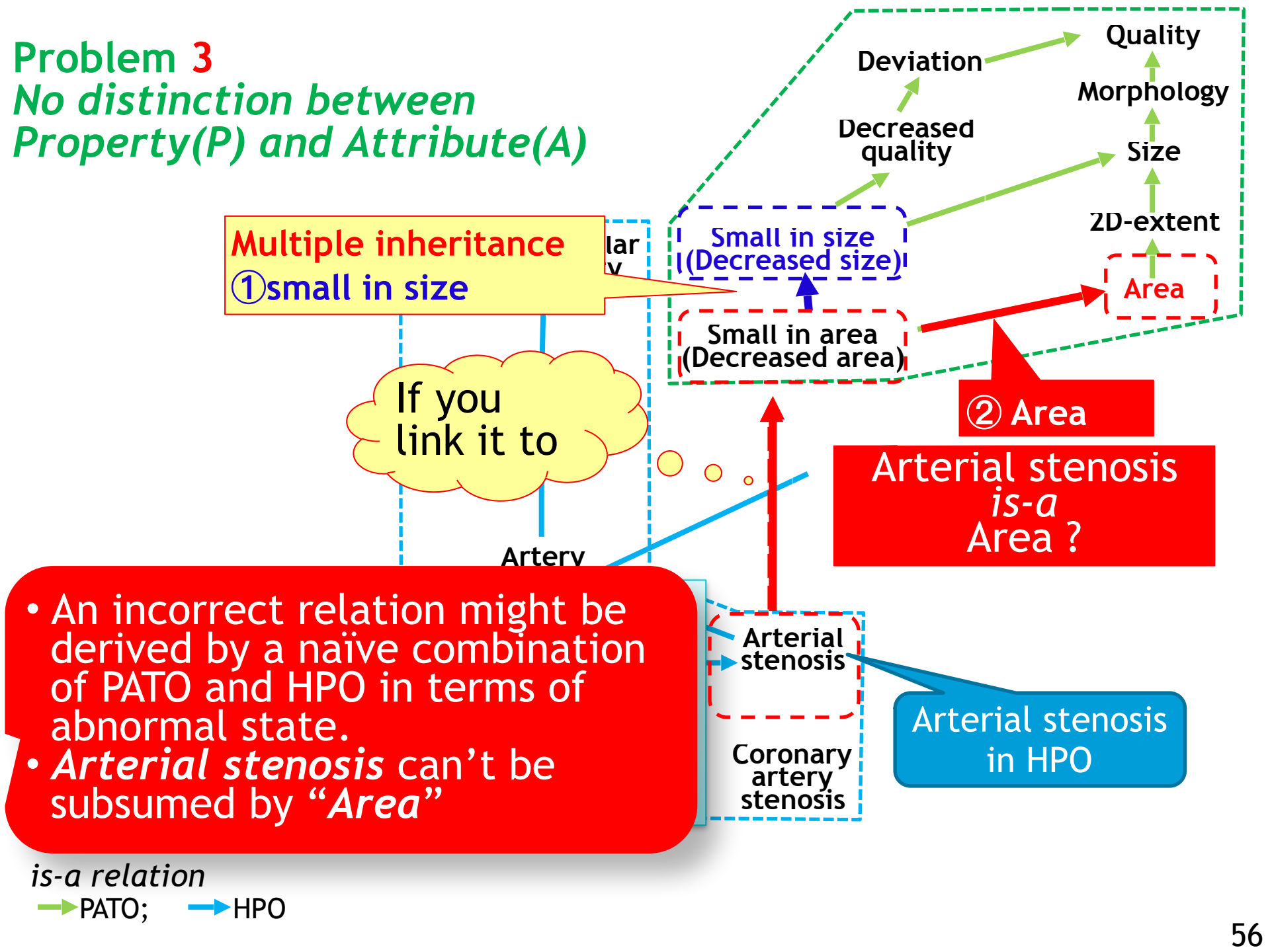
Problem 1
Granularity of concepts to be linked



Problem 2
Abnormal states vs. Diseases



Problem 3
No distinction between Property(P) and Attribute(A)

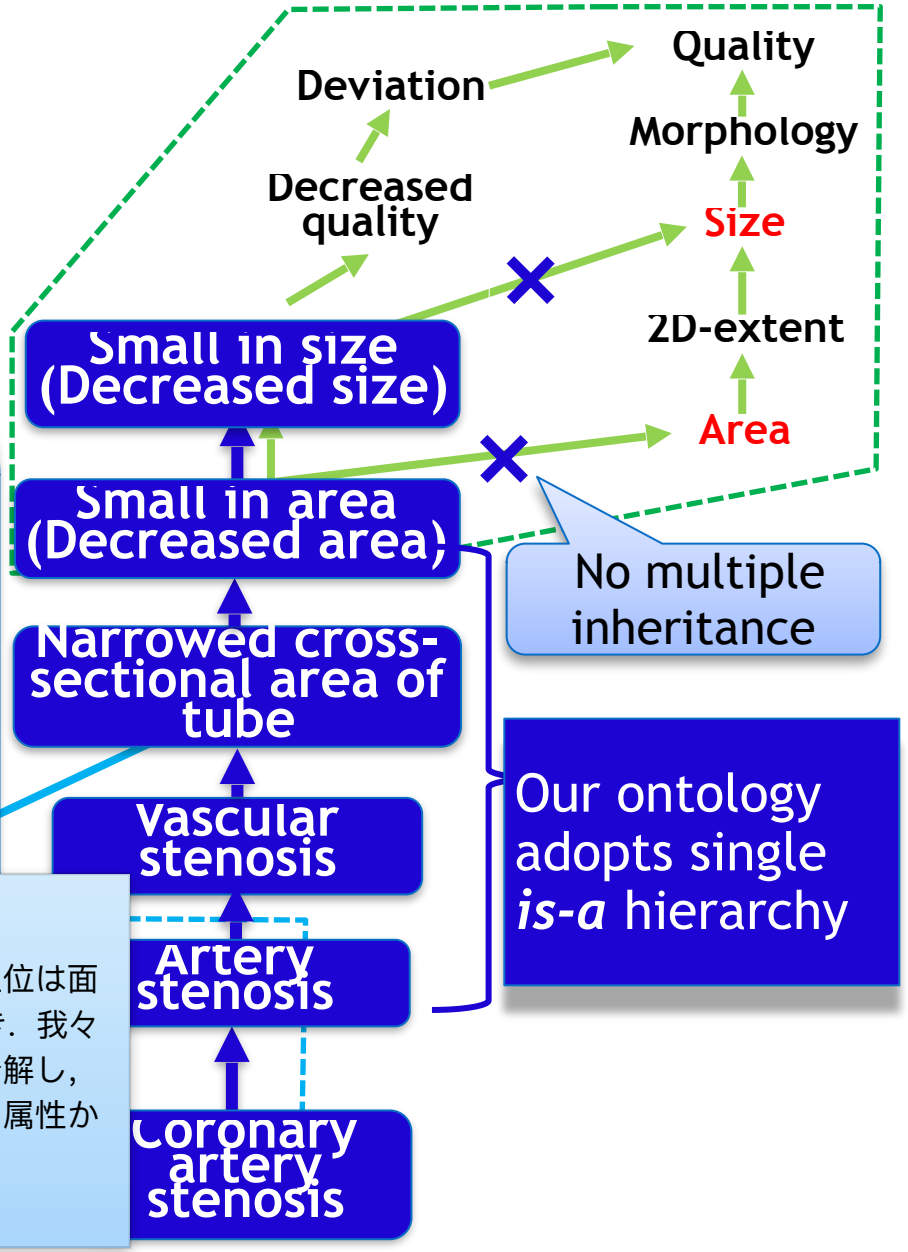


Problem 3
No distinction between Property(P) and Attribute(A)

Abnormality Ontology based on **single *is-a* hierarchy**

- The proper *upper class* of "Small in area" should be "Small in size (P) <size(A), small (V)>" rather than area (A).
- Clear distinction between diseases and abnormal states
- Our model diffe
  (P) and Attribut
  P =def <A,V>
- An engineering
  the issue of stat

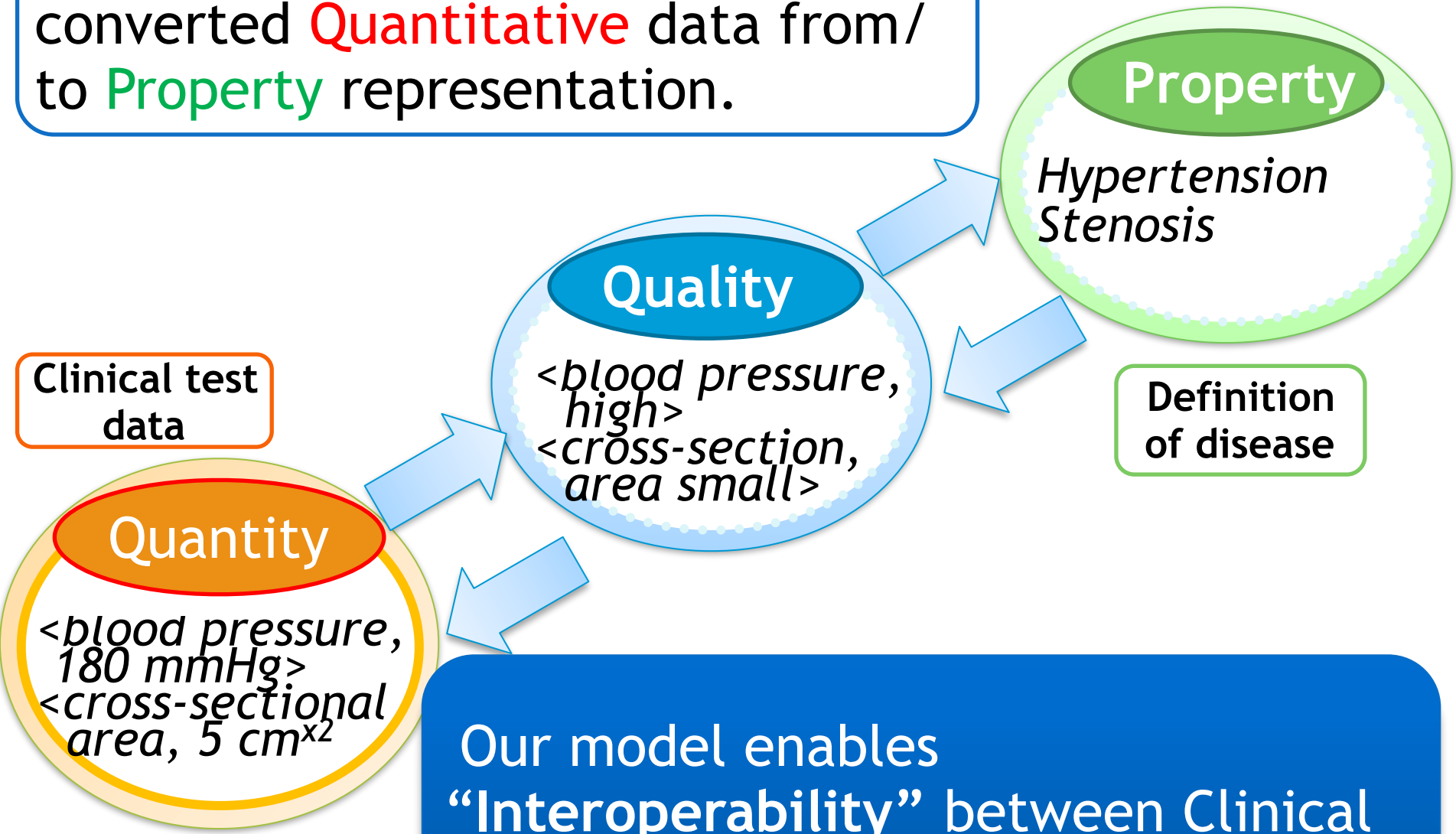
異常状態で考えた場合，狭窄の上位は面積（属性）ではなく，特性とすべき．我々の基本方針では，特性を属性に分解し，特性と属性を区別しているから，属性か



is-a relation

PATO; HPO; Ontology of Abnormal States

Qualitative representation can be converted Quantitative data from/ to Property representation.



Our model enables "**Interoperability**" between Clinical test data and abnormal States.

Linking our ontology as a hub of existing resources will bring benefits:

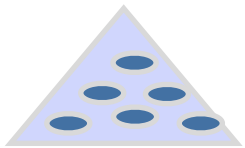
SNOMEDの下に隠れています．利点
Linking our ontology as a hub of existing resources will bring benefits:
・Complement knowledge levels

Concepts of Other resources

Link for mapping

Abnormality Ontology

is-a



SNOMED-CT

MeSH

Small in Area

HPO
Human phenotype ontology

Human Arterial Stenosis

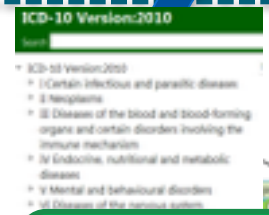
Mouse Arterial Stenosis

Mouse phenotype Ontology

Human Coronary artery stenosis

Human Coronary artery stenosis in Ischemic Heart Disease

Disease

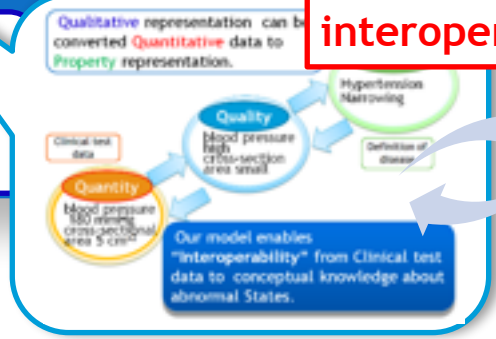


ICD10

Disease ontology

Abnormality Ontology
is-a hierarchy

Property⇔ Attribute interoperability



Clinical DB

Concluding remarks

- **Ontology is important as a sense-making technology**
- Light-weight ontologies work well for LD/LOD of common sense or daily life domains for **making sense of data**
- A lot of vocabularies are already available
- You would need heavy-weight ontologies for LD/LOD in special domains for **making sense of concepts**
- Such data are not interoperable and require a careful analyses of the ontological assumptions
- Heavy-weight (philosophy-oriented) ontologies contribute to such cases

Thank you!

語彙

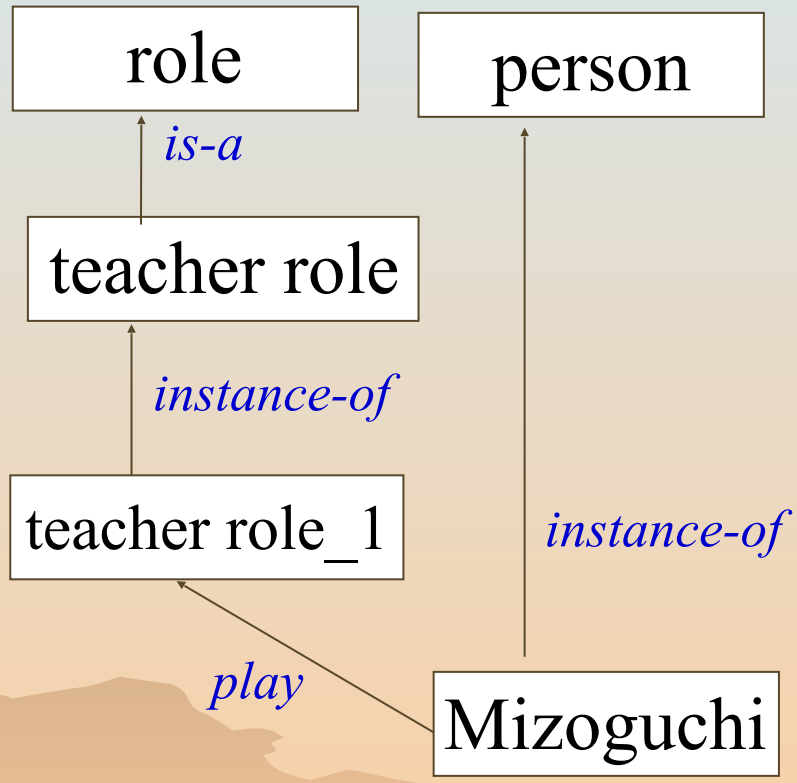
- contraction, stricture (狭窄)
- *stricture* of the esophagus (食道狭窄)
- angiostenosis, vascular constriction (血管狭窄)

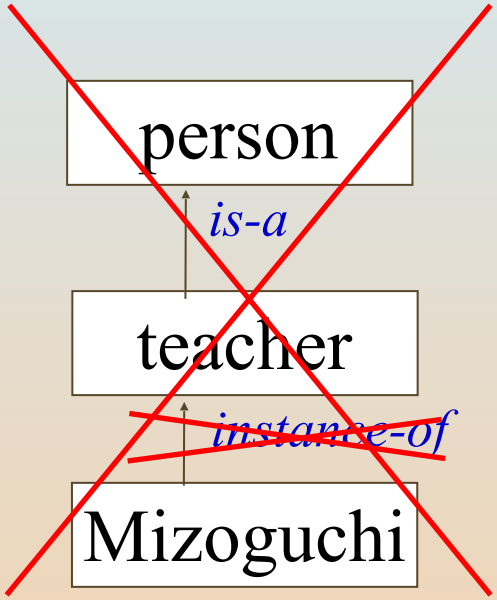
Entity stacking

an amount of clay **vs.** a vase

- They are different entities
- Two entities can't share a spatiotemporal region.
- How can they exist?

Why roles?





Mizoguchi ***instance-of*** person