

Ontology for biomedical models and data

Pierre Grenon, EMBL-EBI, UK
pgrenon@ebi.ac.uk

CrEDIBLE, Session 2
16 Oct 2012

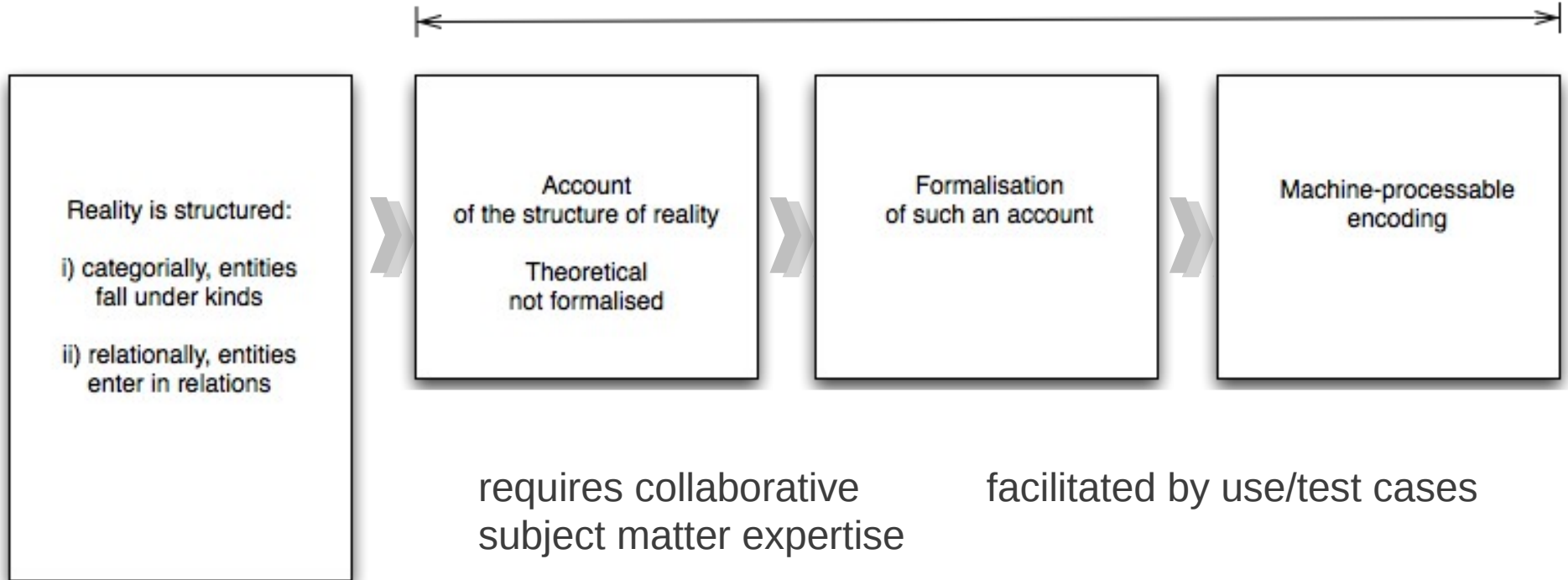
Ontology and ontologies

- *Ontology* is a field of activities.
- *An ontology* is a product in that field.

An ontology is a theory
of the *kinds of entities*
and *relations* there are
in a domain of reality .

Science / knowledge

Ontology



requires collaborative subject matter expertise

facilitated by use/test cases

Domains

Engineering, Physics, Biology, Medicine, Earth sciences, Social science...

Technological Applications in Computer and Information Sciences

Knowledge-based systems, Human-computer interaction, NLP, Database integration, Complex systems, Sensors, Web...

Ontology

- What Account in terms of categories and relations of a domain of reality
- Why Support Knowledge Representation, Reasoning and Management
- How Machine processable formal knowledge representation language

Kinds of ontology

Sophistication

List, tree, axiomatised

Specification

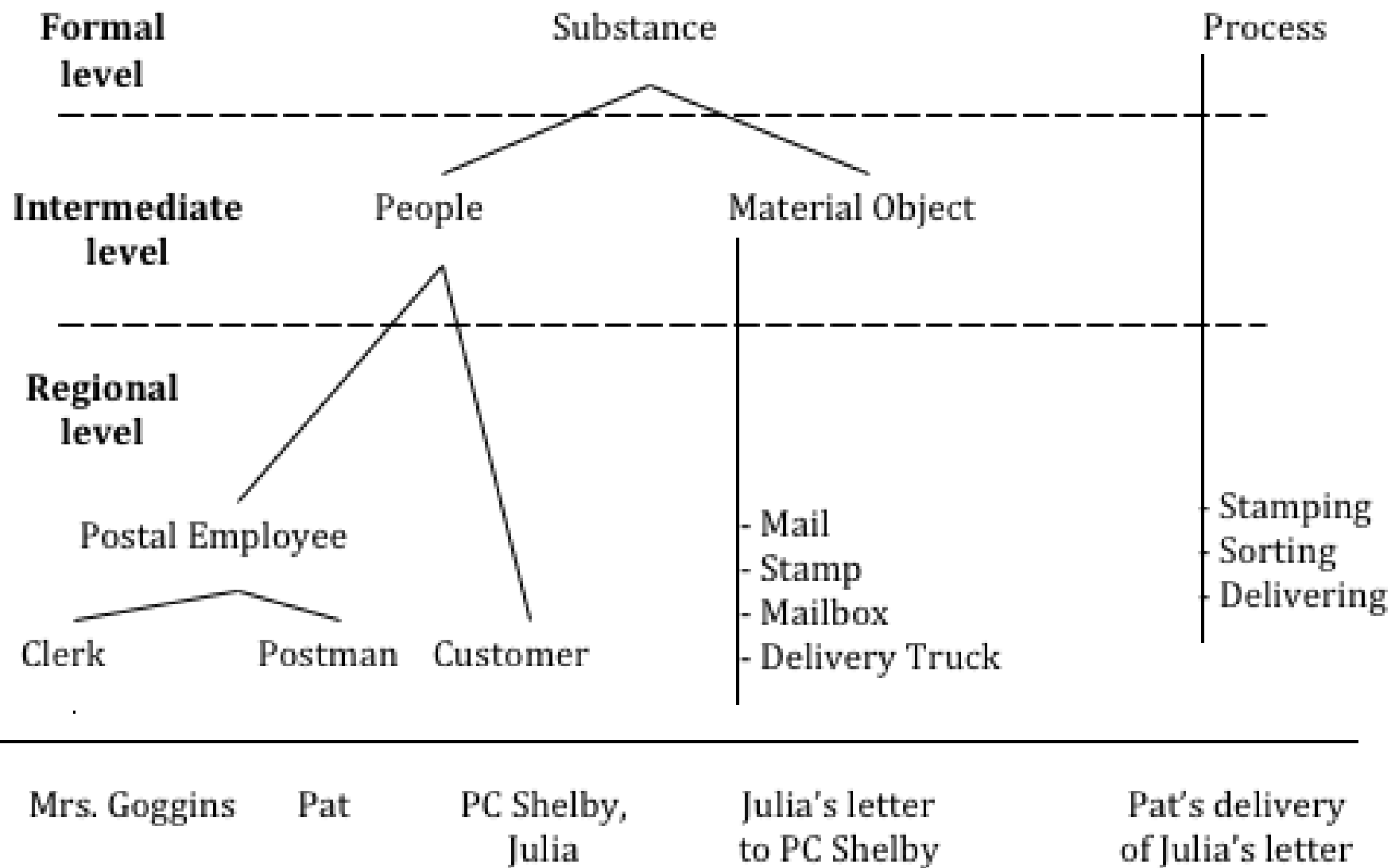
Informal, formalised, encoded in a variety of languages

Scope

Breadth and depth

Generality

Domain non specific, specific or intermediate



The specification, sophistication,
scope and generality of an ontology
is partly defined by

Purpose and Application

also community of practice, legacy...

Ontology *for* biomedical models and data

- resource =def model or data
- Ontology for resources in the biomedical domain

What for?

An ontology for managing biomedical resources?

An ontology for biomedical resource use?

An ontology of activities using resources?

An ontology of multiple kinds of resources?

An ontology of resources?

All of it, with subsumption

An ontology for managing biomedical resources

An ontology for biomedical resource use

An ontology of activities using resources

An ontology of multiple kinds of resources

An ontology of resources

Ontology *of*

Models

- What is a model?
- What sort of entity is it? Is it mathematical? Is it computational? Is it data? Something else?
- What sort of part does it have?
- What sort of property does it have?
- What sort of process does it participate in?
- What kinds of model are there?
 - What makes a given model *this* model?
 - What makes a given model of *model of a given type*?
- How to classify them?

Ontology of Biomedical Resources

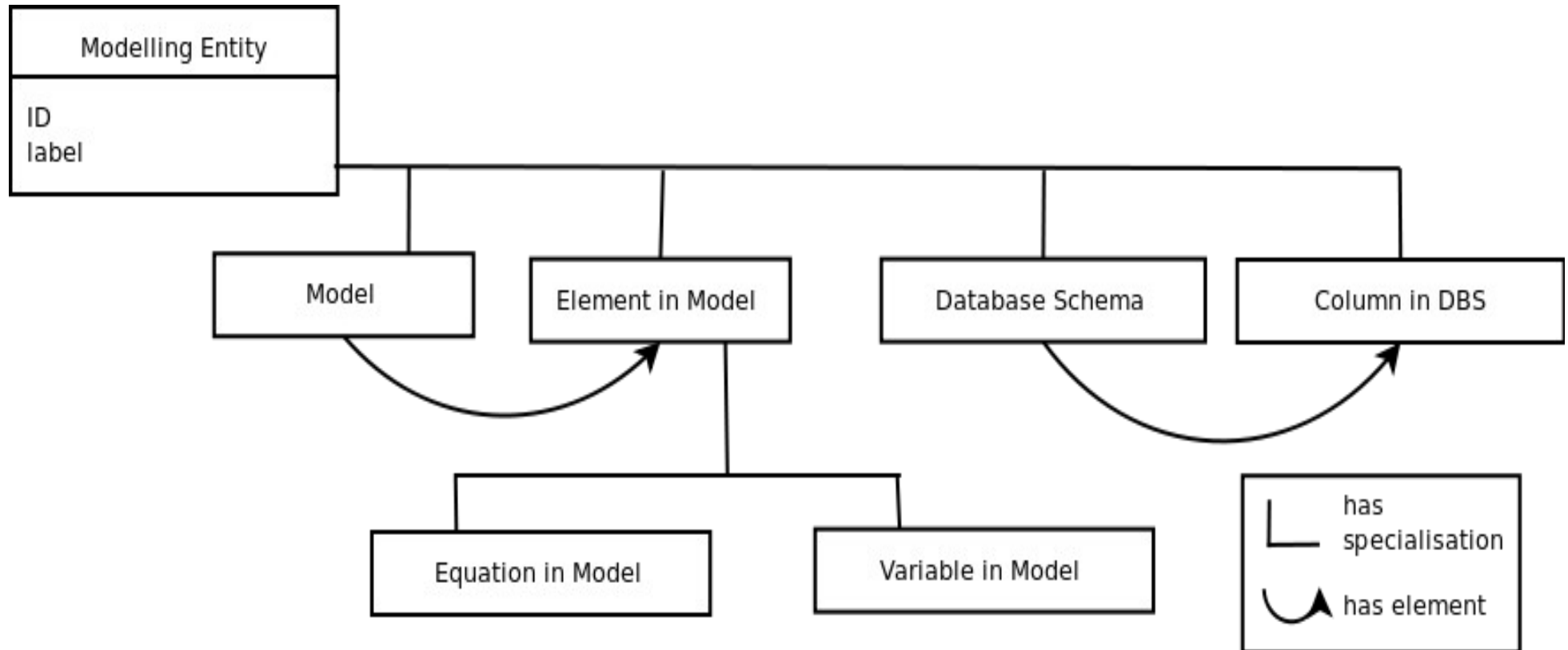
- Are there biomedical resources? What makes them biomedical?
- What makes them domain specific or relevant? Circumscribed by domain (application)
- What is specific to biomedical resources?
- Are formats, structures and so on specific to application domain?
 - Phenomena modelled are biomedical (Interpretation 1)
 - Elements represent biomedical things and properties (Interpretation 2)
 - Data is biomedical (Interpretation 3)

Formal (high) level

- Resources are entities which
- i) can be decomposed in parts (elements),
- ii) can be aggregated (e.g. in repositories),
- iii) can participate in a variety of processual entities (basic operations and workflows).

	BFO	Model
Endurant	Substance Fiat part Aggregate Trope	Resource Resource element Repository Attribute
Perdurant	Process Aggregate Process	Workflow Workflow element

Ontology of (modelling and data) resources



A theoretical schema of classes and relationships.

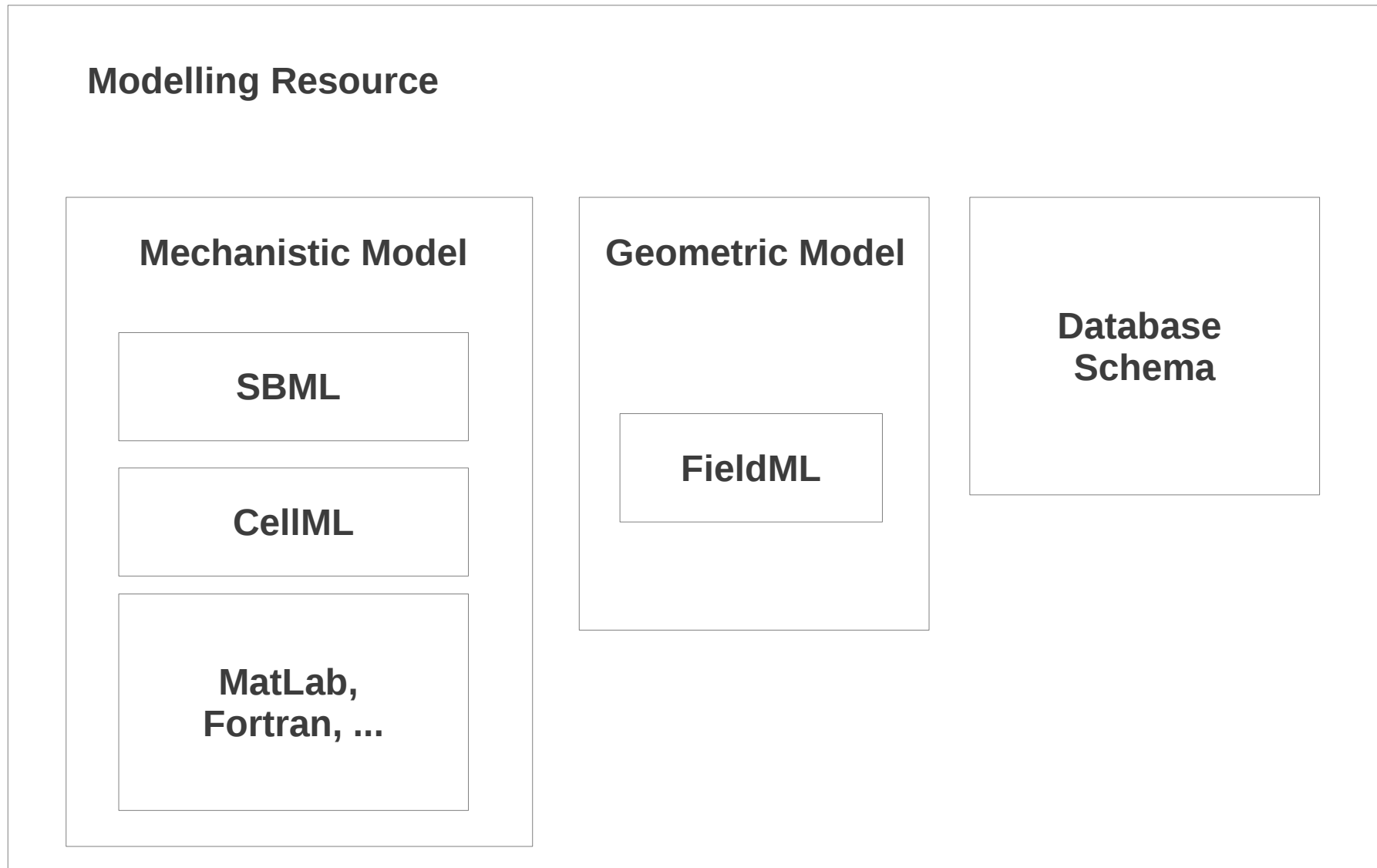
Objects, a specific model, a specific column in a specific database fall under a type in the ontological schema.

Model, Element in Model, Database Schema are subtypes of the generic *Modelling Entity*.
Every *Model Element* is an element in (part of) a *Model*

Abstract vs concrete

- Models are objects
- With elements, e.g.
 - Variables
 - Equations
 - Compartments
- Encodings have structure too
- How well do they map to an abstract level and to alternative encoding formats?

The modelling resource ontology can be specialised and modularised



Boxes within a box indicate more specialised ontological treatments of the sort of resources indicated by labels.

Specialisations of elements

- SBML
 - compartment
 - species
 - parameter
 - reaction
 - ...
- CellML
 - component
 - variable
 - math
 - connection
 - ...

Counting compartments in SBML models (using RDF)

```
SELECT ?o (count(?s) as ?comp)
WHERE {
  ?s <http://www.ebi.ac.uk/ricordo/toolbox/sbmlo#compartmentElementOf> ?o
  . ?o rdf:type <http://www.ebi.ac.uk/ricordo/model#model>
} group by ?o order by ?o
```

<http://www.ebi.ac.uk/ricordo/toolbox/sbmlo#BIOMD0000000242>	1	<http://www.w3.org/2001/XMLSchema#integer>
<http://www.ebi.ac.uk/ricordo/toolbox/sbmlo#BIOMD0000000243>	"1"	^^<http://www.w3.org/2001/XMLSchema#integer>
<http://www.ebi.ac.uk/ricordo/toolbox/sbmlo#BIOMD0000000244>	"2"	^^<http://www.w3.org/2001/XMLSchema#integer>
<http://www.ebi.ac.uk/ricordo/toolbox/sbmlo#BIOMD0000000245>	"2"	^^<http://www.w3.org/2001/XMLSchema#integer>
<http://www.ebi.ac.uk/ricordo/toolbox/sbmlo#BIOMD0000000246>	"4"	^^<http://www.w3.org/2001/XMLSchema#integer>
<http://www.ebi.ac.uk/ricordo/toolbox/sbmlo#BIOMD0000000247>	"2"	^^<http://www.w3.org/2001/XMLSchema#integer>
<http://www.ebi.ac.uk/ricordo/toolbox/sbmlo#BIOMD0000000248>	"2"	^^<http://www.w3.org/2001/XMLSchema#integer>
<http://www.ebi.ac.uk/ricordo/toolbox/sbmlo#BIOMD0000000249>	"1"	^^<http://www.w3.org/2001/XMLSchema#integer>
<http://www.ebi.ac.uk/ricordo/toolbox/sbmlo#BIOMD0000000250>	"3"	^^<http://www.w3.org/2001/XMLSchema#integer>
<http://www.ebi.ac.uk/ricordo/toolbox/sbmlo#BIOMD0000000251>	"1"	^^<http://www.w3.org/2001/XMLSchema#integer>
<http://www.ebi.ac.uk/ricordo/toolbox/sbmlo#BIOMD0000000252>	"1"	^^<http://www.w3.org/2001/XMLSchema#integer>
<http://www.ebi.ac.uk/ricordo/toolbox/sbmlo#BIOMD0000000253>	"1"	^^<http://www.w3.org/2001/XMLSchema#integer>
<http://www.ebi.ac.uk/ricordo/toolbox/sbmlo#BIOMD0000000254>	"1"	^^<http://www.w3.org/2001/XMLSchema#integer>
<http://www.ebi.ac.uk/ricordo/toolbox/sbmlo#BIOMD0000000255>	"6"	^^<http://www.w3.org/2001/XMLSchema#integer>
<http://www.ebi.ac.uk/ricordo/toolbox/sbmlo#BIOMD0000000256>	"2"	^^<http://www.w3.org/2001/XMLSchema#integer>
<http://www.ebi.ac.uk/ricordo/toolbox/sbmlo#BIOMD0000000257>	"1"	^^<http://www.w3.org/2001/XMLSchema#integer>
<http://www.ebi.ac.uk/ricordo/toolbox/sbmlo#BIOMD0000000258>	"1"	^^<http://www.w3.org/2001/XMLSchema#integer>
<http://www.ebi.ac.uk/ricordo/toolbox/sbmlo#BIOMD0000000259>	"2"	^^<http://www.w3.org/2001/XMLSchema#integer>

Querying at the abstract level (first step towards integration)

- SPARQL Query on a simple RDF store

```
PREFIX rcms: <http://www.ebi.ac.uk/ricordo/model#>
SELECT ?s ?o
WHERE
{ ?s rcms:elementOf ?o }
```

Find any element
and what it is an element of

s	o
<http://www.ebi.ac.uk/biomodels/models-main/publ/BIOMD0000000064#_0000026>	<http://www.ebi.ac.uk/biomodels/models-main/publ/BIOMD0000000064#_0000002>
<http://www.ebi.ac.uk/biomodels/models-main/publ/BIOMD0000000064#_0000023>	<http://www.ebi.ac.uk/biomodels/models-main/publ/BIOMD0000000064#_0000002>
<http://www.ebi.ac.uk/biomodels/models-main/publ/BIOMD0000000003#_961192>	<http://www.ebi.ac.uk/biomodels/models-main/publ/BIOMD0000000003#_180340>
<http://www.ebi.ac.uk/biomodels/models-main/publ/BIOMD0000000064#_0000008>	<http://www.ebi.ac.uk/biomodels/models-main/publ/BIOMD0000000064#_0000002>
<http://www.ebi.ac.uk/ricordo/ersatz/barcelona/CVM0000000001#_000012>	<http://www.ebi.ac.uk/ricordo/ersatz/barcelona/CVM0000000001#_000001>
<http://www.ebi.ac.uk/biomodels/models-main/publ/BIOMD0000000064#_0000016>	<http://www.ebi.ac.uk/biomodels/models-main/publ/BIOMD0000000064#_0000002>
<http://www.ebi.ac.uk/ricordo/annotation/kb#DBS1C4>	<http://www.ebi.ac.uk/ricordo/annotation/kb#DBS1>

ontology *for*

Institutional
ontology

- Regulatory agencies
- Development phases
- Quality assurance
- Validation
- Documentation
- Submission

Mathematical
/formalisation
Ontology

- ODE
- PDE
- Stochastic

Software
Ontology

- Simulation
- Format
- Repository
- Datatype
- Provenance

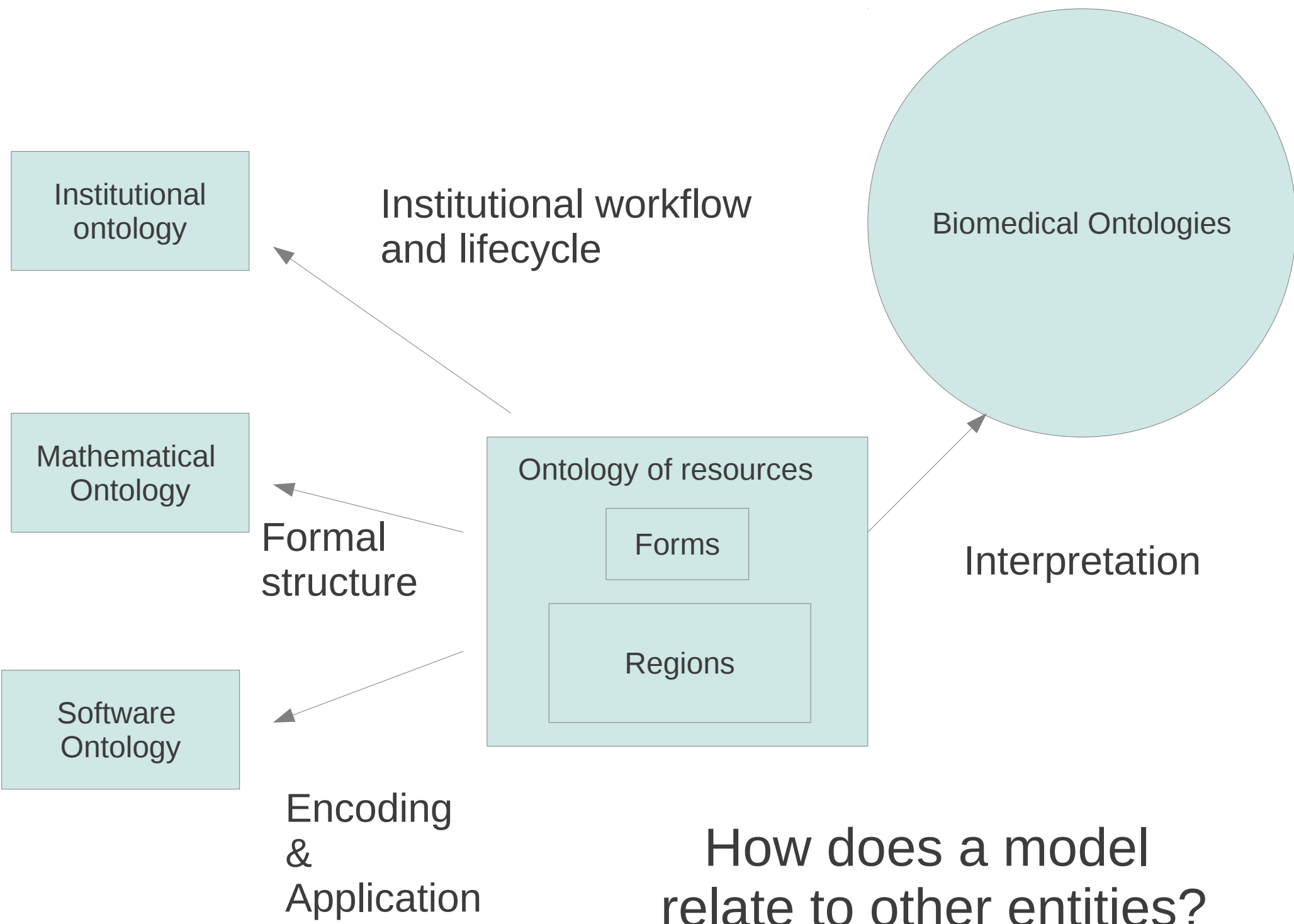
Ontology of resources

Forms

Regions

Biomedical Ontologies

- Protein
- Gene
- Clinical trial
- Lab tests
- Medical intervention
- Patient data (ramified)
- Drug
- Disease
- Microbiology



Result:

description and classification of resources and how they relate together based on

- Thematic domain
- Interpretation of model elements
- Biologically relevant ranges of spatial and temporal scales
- Kind of mathematics
- Data integration
- Quality measures
- ...

Reuse

Integration (merging, linking)

Extensions

Few new developments

Class hierarchy (inferred)

Class hierarchy: MULTIPLE-DOSE-LINEAR-ELIMINATION-BOLUS-1CPT-PK-MODEL

- MODEL-PARAMETER
 - PK-MODEL-PARAMETER
 - 1-CPT-PK-MODEL-PARAMETER
 - ABSORPTION-DURATION-FOR-ZERO-ORDER-ABSORPTION-1-CPT-PK-
 - ABSORPTION-RATE-CONSTANT-FOR-ORAL-ADMINISTRATION-1-CPT-
 - APPARENT-VOLUME-OF-DISTRIBUTION-FOR-EXTRA-VASCULAR-ADMINIS
 - ELIMINATION-RATE-CONSTANT-1-CPT-PK-MODEL-PARAMETER
 - LAG-TIME-FOR-ZERO-OR-FIRST-ORDER-ADMINISTRATION-1-CPT-PK-I
 - MAXIMUM-ELIMINATION-RATE-FOR-MICHAELIS-MENTEN-ELIMINATION
 - MICHAELIS-MENTEN-CONSTANT-1-CPT-PK-MODEL-PARAMETER
 - ORAL-CLEARANCE-OF-ELIMINATION-FOR-EXTRA-VASCULAR-ADMINIS
 - 2-CPT-PK-MODEL-PARAMETER
 - 3-CPT-PK-MODEL-PARAMETER
 - ABSORPTION-DURATION-FOR-ZERO-ORDER-ABSORPTION-PK-MODEL-P
 - ABSORPTION-RATE-CONSTANT-PK-MODEL-PARAMETER
 - APPARENT-VOLUME-OF-DISTRIBUTION-OF-COMPARTMENT-FOR-EXTRA-
 - LAG-TIME-FOR-ZERO-OR-FIRST-ORDER-ADMINISTRATION-PK-MODEL-PA
 - MODEL-VARIABLE-TYPE-PK
 - PK-MODEL
 - 1CPT-PK-MODEL
 - BOLUS-1CPT-PK-MODEL
 - LINEAR-ELIMINATION-BOLUS-1CPT-PK-MODEL
 - MULTIPLE-DOSE-LINEAR-ELIMINATION-BOLUS-1CPT-PK-MODEL
 - SINGLE-DOSE-LINEAR-ELIMINATION-BOLUS-1CPT-PK-MODEL
 - STEADY-STATE-LINEAR-ELIMINATION-BOLUS-1CPT-PK-MODEL
 - MICHAELIS-MENTEN-ELIMINATION-BOLUS-1CPT-PK-MODEL
 - FIRST-ORDER-ABSORPTION-1CPT-PK-MODEL
 - INFUSION-1CPT-PK-MODEL
 - ZERO-ORDER-ABSORPTION-1CPT-PK-MODEL
 - 2CPT-PK-MODEL
 - 3CPT-PK-MODEL

Annotations Usage

Annotations: MULTIPLE-DOSE-LINEAR-ELIMINATION-BOLUS-1CPT-PK-MODEL

Annotations +

Description: MULTIPLE-DOSE-LINEAR-ELIMINATION-BOLUS-1CPT-PK-MODEL

Equivalent classes +

- LINEAR-ELIMINATION-BOLUS-1CPT-PK-MODEL
and (MODEL-HAS-ADMINISTRATION-PROFILE-ATTRIBUTE value MULTIPLE-DOSE-ADMINISTRATION-PROFILE-MODEL)

Superclasses +

- LINEAR-ELIMINATION-BOLUS-1CPT-PK-MODEL
- MULTIPLE-DOSE-PK-MODEL

Inherited anonymous classes

- 1CPT-PK-MODEL
and (MODEL-HAS-ROUTE-OF-ADMINISTRATION-ATTRIBUTE value INTRAVENOUS-BOLUS-ROUTE-OF-ADMINISTRATION-MODEL)
- PK-MODEL
and (MODEL-HAS-COMPARTMENTAL-ATTRIBUTE value 1-COMPARTMENT-MODEL)
- BOLUS-1CPT-PK-MODEL
and (MODEL-HAS-ELIMINATION-PROCESS-ATTRIBUTE value



PharmaGUI – Annotate

Enter an ID for the model

Assert

Result

Enter a module name, e.g. TEST

Assert

Result

Enter a name for the model:

Assert

Result

Ascribe an interpretation type:

(DISEASE-PROGRESSION-MODEL TIMETOEVENT PKPDEFFECTS PKMODEL)

Assert

Result

[NONMEM] Ascribe a PREDPP subroutine:

Assert

Result

Ascribe a drug (using DrugBank ID):

<http://www.drugbank.ca/search/advanced>

Assert

Result

Ascribe a number of compartments

Assert

Result

Assert

Assert

Result

PharmaGUI -- Basic queries

<p>Models count: nil list: integer or 'all'</p> <input type="text"/>	<input type="button" value="#"/> <input type="button" value="list"/>	78
<p>Types of models count: nil list: integer or 'all'</p> <input type="text" value="all"/>	<input type="button" value="#"/> <input type="button" value="list"/>	("DISEASE-PROGRESSION-MODEL" "TIMETOEVENT" "PKPDEFFECTS" "PKMODEL")
<p>Models by ?type (use query 2)</p> <input type="text" value="pkmodel"/>	<input type="button" value="#"/> <input type="button" value="list"/>	"WARFARIN-PK-ALLOMCL-WTV" "WARFARIN-PK-WT-ON-VPWR" "WARFARIN-PK-BASE" "ONECP-ATOVAQUONE")
<p>Models by ?number of compartment (1, 2, 3)</p> <input type="text" value="1"/>	<input type="button" value="#"/> <input type="button" value="list"/>	8
<p>Models by ?drug (drugbank.ca id) E.g. Warfarin: DB00682</p> <input type="text" value="DB00682"/>	<input type="button" value="#"/> <input type="button" value="list"/>	27
<p>Intersect (non nil)</p> <input type="text"/>	<input type="button" value="#"/> <input type="button" value="list"/>	"WARFARIN-PK-WTV" "WARFARIN-PK-WTCL" "WARFARIN-PK-ALLOMCL" "WARFARIN-PK-ALLOMCL-WTV" "WARFARIN-PK-WT-ON-VPWR" "WARFARIN-PK-BASE")
<p>Code format</p> <input type="text" value="all"/>	<input type="button" value="#"/> <input type="button" value="list"/>	("WINNONLIN-SOFTWARE" "NONMEM-SOFTWARE" "MONOLIX-SOFTWARE" "MATLAB-SOFTWARE")
<p>Models by code format</p> <input type="text" value="NONMEM-software"/>	<input type="button" value="#"/> <input type="button" value="list"/>	10

CDISC (partly) Ontologised

```
(ASSERT (BASIC-CDISC-UNIT CDISC-C70470))
(ASSERT (MULTIPLE-DECOMPOSITION DECIMAL-MULTIPLIER-PICO CDISC-C41139 CDISC-C69148))
(ASSERT (BASIC-CDISC-UNIT CDISC-C68915))
(ASSERT (MULTIPLE-DECOMPOSITION DECIMAL-MULTIPLIER-ATTO CDISC-C42539 CDISC-C68855))
(ASSERT (MULTIPLE-DECOMPOSITION DECIMAL-MULTIPLIER-FEMTO CDISC-C42539 CDISC-C68854))
(ASSERT (MULTIPLE-DECOMPOSITION DECIMAL-MULTIPLIER-CENTI CDISC-C42539 CDISC-C68687))
(ASSERT (MULTIPLE-DECOMPOSITION DECIMAL-MULTIPLIER-DECI CDISC-C42539 CDISC-C68685))
(ASSERT (MULTIPLE-DECOMPOSITION DECIMAL-MULTIPLIER-NANO CDISC-C48466 CDISC-C67352))
(ASSERT (BASIC-CDISC-UNIT CDISC-C67330))
(ASSERT (MULTIPLE-DECOMPOSITION DECIMAL-MULTIPLIER-NANO CDISC-C41139 CDISC-C67328))
(ASSERT (MULTIPLE-DECOMPOSITION DECIMAL-MULTIPLIER-MILLI CDISC-C42551 CDISC-C67324))
(ASSERT (BASIC-CDISC-UNIT CDISC-C67194))
(ASSERT (BASIC-CDISC-UNIT CDISC-C67193))
(ASSERT (MULTIPLE-DECOMPOSITION DECIMAL-MULTIPLIER-PICO CDISC-C42539 CDISC-C65045))
(ASSERT (MULTIPLE-DECOMPOSITION DECIMAL-MULTIPLIER-CENTI CDISC-C48155 CDISC-C64554))
(ASSERT (MULTIPLE-DECOMPOSITION DECIMAL-MULTIPLIER-ATTO CDISC-C48155 CDISC-C64553))
(ASSERT (MULTIPLE-DECOMPOSITION DECIMAL-MULTIPLIER-FEMTO CDISC-C48155 CDISC-C64552))
(ASSERT (MULTIPLE-DECOMPOSITION DECIMAL-MULTIPLIER-PICO CDISC-C48155 CDISC-C64551))
(ASSERT (BASIC-CDISC-UNIT CDISC-C54711))
(ASSERT (MULTIPLE-DECOMPOSITION DECIMAL-MULTIPLIER-CENTI CDISC-C41139 CDISC-C49668))
(ASSERT (BASIC-CDISC-UNIT CDISC-C48553))
(ASSERT (BASIC-CDISC-UNIT CDISC-C48531))
(ASSERT (BASIC-CDISC-UNIT CDISC-C48519))
(ASSERT (MULTIPLE-DECOMPOSITION DECIMAL-MULTIPLIER-NANO CDISC-C42539 CDISC-C48517))
(ASSERT (MULTIPLE-DECOMPOSITION DECIMAL-MULTIPLIER-MILLI CDISC-C42539 CDISC-C48513))
```

Ground facts about data/models

```
40 ;;;
41 ;;; ATOQUAVONE
42 |
43 (assert (dataset atoquavone-csv))
44 ; AGE Year
45 ; Weight Kg
46 ; DODE mg cdisc-C28253
47 ; time hour
48 ; sex
49
50 ; time
51 (assert (dataset-has-declared-quantity-type-in-unit atoquavone-csv time-span-pharma cdisc-C25529))
52 (assert (dataset-has-declared-unit atoquavone-csv CDISC-C25529))
53
54 ; weight
55 (assert (dataset-has-declared-quantity-type-in-unit atoquavone-csv CDISC-C25208 CDISC-C28252))
56 (assert (dataset-has-declared-unit atoquavone-csv CDISC-C28252))
57
58 ; age
59 (assert (dataset-has-declared-quantity-type-in-unit atoquavone-csv age-of-patient CDISC-C29848))
60 (assert (dataset-has-declared-unit atoquavone-csv CDISC-C29848))
61
62 ; sex
63 (assert (dataset-has-declared-unit atoquavone-csv CDISC-C66731))
64
65 ;amt
66 (assert (dataset-has-declared-quantity-type-in-unit atoquavone-csv cdisc-C25488 cdisc-C28253))
67 (assert (dataset-has-declared-unit atoquavone-csv cdisc-C28253))
68
69 ;
70 (assert (model-has-associated-dataset ONECP-ATOVAQUONE atoquavone-csv))
71
```

2 models with some compatible units (irrespective of quantity)

- (and (not (= ?x ?y))
 (model-has-declared-unit-gen ?x ?u)
 (model-has-declared-unit-gen ?y ?v)
 (unit-compatible-with-unit ?u ?v)
 (cdisc-label ?u ?lu) (cdisc-label ?v ?lv))
- ONECP-ATOVAQUONE
 CDISC-C28253 "Milligram"

- WARFARIN-PK-WT-ON-V
 CDISC-C28252 "Kilogram"

2 models:

- i) with units compatible for the same quantity
- ii) do not reference the same dataset

- (and
(not (= ?x ?y))
(model-has-declared-quantity-type-in-unit-gen ?x ?q ?u)
(model-has-declared-quantity-type-in-unit-gen ?y ?q ?v)
(model-has-associated-dataset ?x ?dx)
(model-has-associated-dataset ?y ?dy) (not (= ?dx ?dy))
(or (unit-compatible-with-unit ?u ?v) (= ?u ?v))
(cdisc-label ?u ?lu) (cdisc-label ?v ?lv))

- WARFARIN-PK-WT-ON-V model
CDISC-C28252 "Kilogram" unit
WARFARIN-CSV data

- ONECP-ATOVAQUONE model
CDISC-C28252 "Kilogram" unit
ATOQUAVONE-CSV data

- CDISC-C25208 ("Weight") quantity

Operationalisation

1. Model construction

- Using ontology directly within model specification
- Ontology linked to code library
- Tools capable of 'filling in' specification details
- Combination

2. Workflow description and parameterisation

- Representation using ontology
- Execution

Summary

- Ontology as theory of a domain
- Ontology for a domain may include and articulate ontologies in multiple domains at different levels
- Extent of specification depends on purpose
- Specifications (encoding and so on) for different, articulated ontologies do not have to be homogeneous in complexity nor used in the same way by the same tools