Loria

Phenotype Meeting *December 1-2*

KIKA

Suggested Ontology for Pharmacogenomics (SO-Pharm), Modular Construction and Preliminary Testing

Adrien Coulet, Malika Smaïl-Tabbone, Amedeo Napoli and Marie-Dominique Devignes

Laboratoire Lorrain de Recherche en Informatique et ses Applications (CNRS, INRIA, Nancy University), Nancy, France

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

INRIA

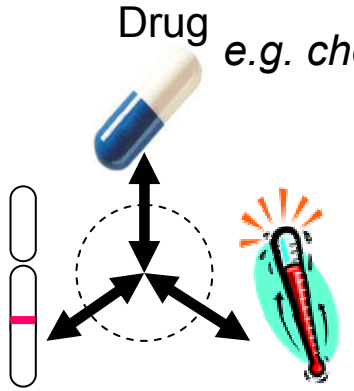
université HENRI POINCARÉ NANCY 1

Outline

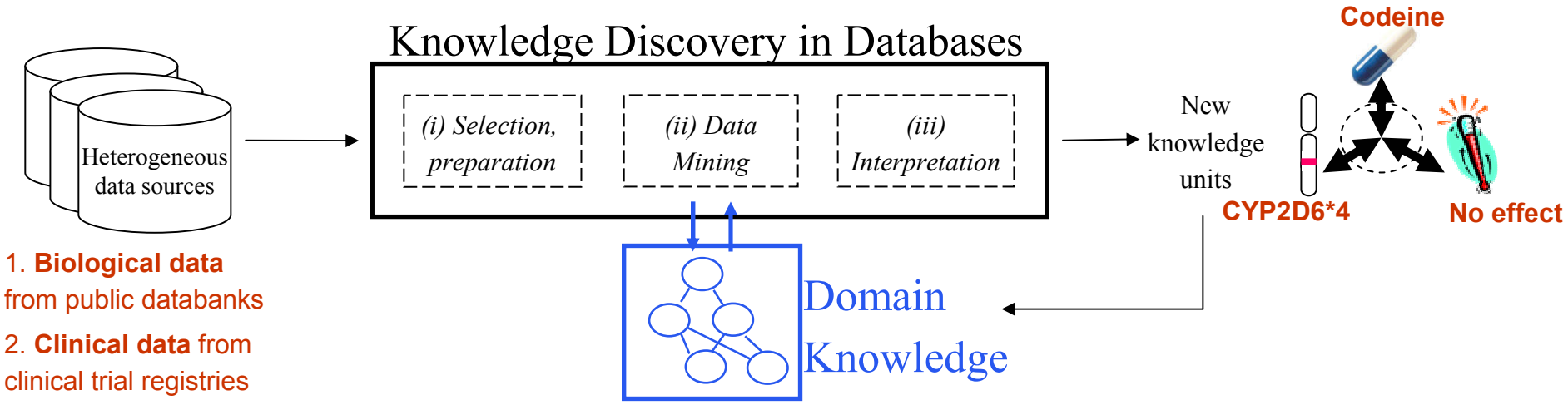
- Motivation
- Ontology Construction: Which methodology?
- SO–Pharm Construction Details
- SO–Pharm Validation
- Conclusion

Pharmacogenomics Definition

- How genomic variations lead to variations in drug response
- Example of Codeine and CYP2D6
  - Codeine (opioid) must be metabolized into morphine to be active
  - Gene CYP2D6 code for the enzyme that performs this metabolism
  - 7% of Caucasians have an inactive variant of CYP2D6 enzyme
  - Codeine stays inactive in these patients



Pharmacogenomic KDDK

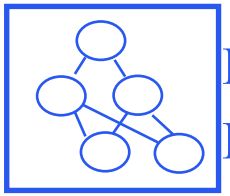


- Discovering new knowledge units thanks to
  - KDD process
  - -- guided by -- formal representation of the domain

Knowledge Discovery guided by Domain Knowledge

- Applied to pharmacogenomics

Ontology for Pharmacogenomics



- What we need ?
  - A formal representation of the pharmacogenomics domain knowledge
- What is existing ?
  - PharmGKB
    - XML schema
  - More specific bio-ontologies
    - Ontology describing genomics, phenomics and pharmacology
    - *e.g.* OBO portal, Protégé Ontologies Library, dispersed ontologies, etc.
- What we decide to do ?
  - Construct a formal ontology
    - based on existing information i.e. models, databases, etc. (PharmGKB, CDISC,etc.)
    - reusing existing ontologies as much as possible (PATO, SNP-Ontology, ChEBI, etc.)

Ontology Construction: which methodology

- Semi automatic construction VS Manual construction
  - The domain is complex
  - We want a consistent and well sounded ontology
  - We collaborate with domain experts
- Manual
  - …but not from scratch
- Need of a precise methodology
  - Based on existing methodologies
    - *e.g.* METHONTOLOGY, DILIGENT

    (i) Specification ***(ii)*** Conceptualization ***(iii)*** Formalization ***(iv)*** Implementation
  - Adapted to our case
    - Favoring reuse and iterative process

SO-Pharm Specification

- Ontology definition
  - What domain?

    Clinical trials in pharmacogenomics
  - Which purpose?

    Knowledge resource useful for
    - integrating data
    - guiding data mining
    - interpreting results
- Competency questions
  - Definition: Questions to which the information in the ontology should answer?
  - Example:

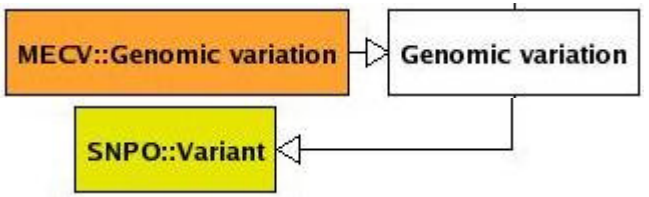
    What kind of genomic variations are present in patients with

    no response to codeine?

SO-Pharm Conceptualization (1/3)

- From *linked concepts* … to *linked ontologies*
- Linking existing ontologies
  - *Embedding*

    e.g. The genomic variation concept from the Mutation Event Controlled Vocabulary is embedded by the SO-Pharm concept of genomic variation

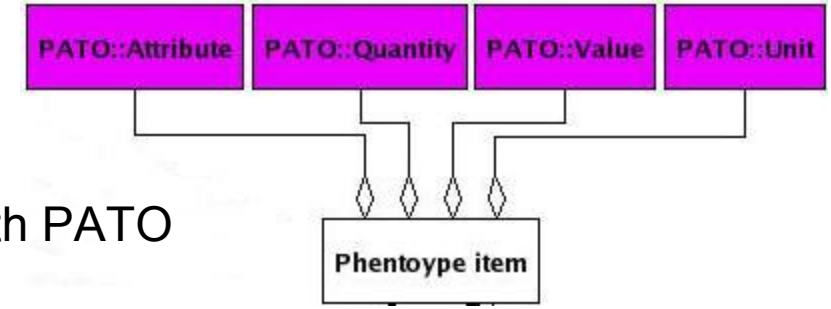


  - *Extending*

    e.g The SO-Pharm concept of genomic variation is extended by the variant concept defined in SNP-Ontology

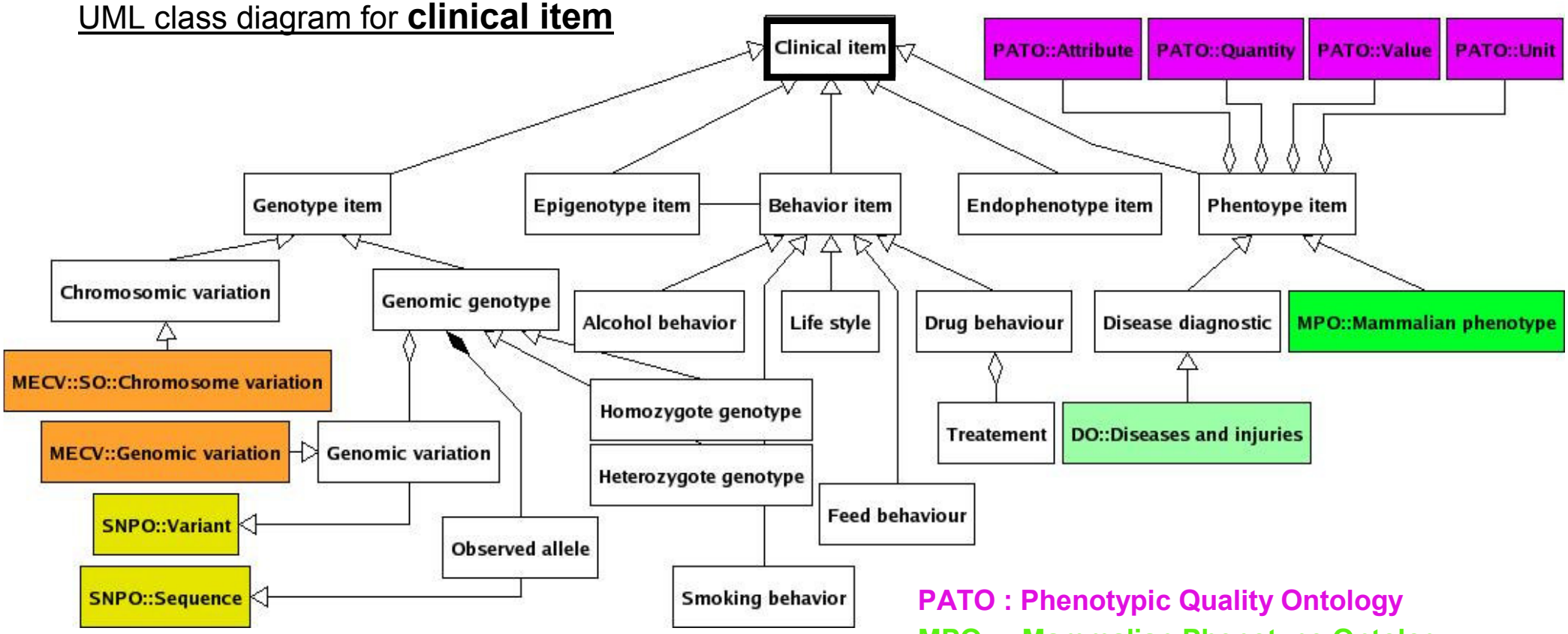
  - *Composition*

    For describing phenotype measurements with PATO



SO-Pharm Conceptualization (2/3)

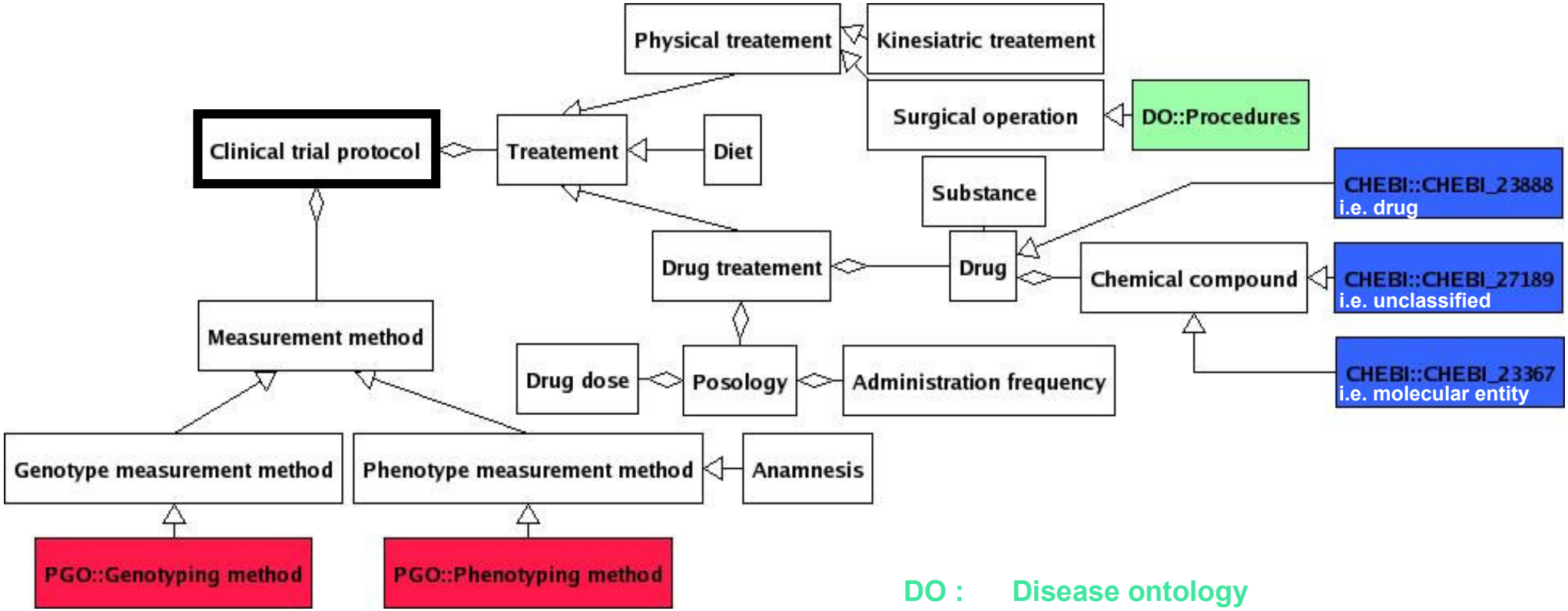
UML class diagram for **clinical item**



PATO : Phenotypic Quality Ontology
MPO : Mammalian Phenotype Ontology
DO : Disease Ontology
MECV : Mutation Event Controlled Vocabulary
SNPO : SNP-Ontology

SO-Pharm Conceptualization (3/3)

UML class diagram for **clinical trial protocol**



DO : Disease ontology

ChEBI : Chemical Entities of Biological Interest

PGO : Pharmacogenetics Ontology

SO-Pharm Formalization and Implementation

- Manually formalized and implemented from UML to OWL DL

- Articulation of existing ontologies
  - OWL translation
    - Directly downloaded in OWL, manually / automatically translated in OWL

    N.B.: obo2owl and [Obo-format] mailing list discussions

  - Existing ontologies are then imported and linked

$$MECV : genomic\_variation \sqsubseteq genomic\_variation \sqsubseteq SNPO : variant$$

SO-Pharm Validation

- A preliminary validation
  - By expressing pharmacogenomics properties with SO-Pharm concepts
    - *i.e.* instantiate the ontology with patients, their clinical items and genotype items that correspond to a “pharmacogenomic profile”
  - Using 10 clearly defined pharmacogenomic properties from litterature
    - TPMT, CYP2D6, etc.
- Validation results
  - Lead to refine SO-Pharm in an iterative process
    - new concepts: *e.g.* anamnesis, poor/rich metabolizer, etc.
    - New relationships

Conclusion (1/3)

- Manual construction
  - (+) Sensible reuse: avoid co-existing concepts and unclear concepts which lead problems for reasoning upon merged ontologies
  - (-) Time consuming, manual update
- Results
  - 64 concepts articulating more than 30 000 external concepts (from 8 ontologies)
  - 18 defined concepts
  - Genotype, phenotype and pharmacology concepts are articulated
  - Take advantage of various type of ontologies: controlled vocabularies (that define classes) and formal ontologies (that allow classification)
    - *e.g.:* a *genomic variation* can be specified as a *MECV: insertion*, *MECV: deletion*, etc.
    - *e.g.:* a *genomic variation* benefits from the formal definition of the *SNPO: variant*
- Available at http://www.loria.fr/~coulet/ontology/sopharm_description.html
- Submittion to the new NCBO BioPortal

Conclusion (2/3)

- Illustration: SO–Pharm for guiding data mining tasks
  - Using SO-Pharm for filtering data mining results
  - A set of association rules
    - 1. ~~[Morpine]$_{blood}$ < 1,5 nmol/l => Paris hospital~~ *no interest*
    - 2. ~~codeine treatment => opioid treatment~~ *evident*
    - 3. rs1065852, codeine treatment => no analgesic effect *interesting*
    - 4. ~~CYP2D6*10, codeine treatment => no analgesic effect~~ *equivalent*
    - 5. ~~CYP2D6 variant, opioid treatment => no analgesic effect~~ *generalization*

Conclusion (3/3)

- On going work
  - Development of wrappers to populate a knowledge base
  - KDDK on real data

- About PATO
  - Efficiency of PATO for expressing phenotype data
  - Use PATO for integrating phenotype data sources?
  - Finding generalization or composition relationships in phenotype data?
  - Dynamic phenotype

Thanks

- INSERM UMRS 538
  - Dr Pascale Benlian
- KIKA Medical
  - Benoît Lopez
- Phenosystems
  - David Atlan
- ORPAILLEUR group in LORIA
  - Amedeo Napoli
  - Mathieu d'Aquin
  - Marie-Dominique Devignes
  - Nicolas Jay
  - Jean Lieber
  - Sandy Maumus
  - Malika Smaïl-Tabbone

SO-Pharm Specification (details)

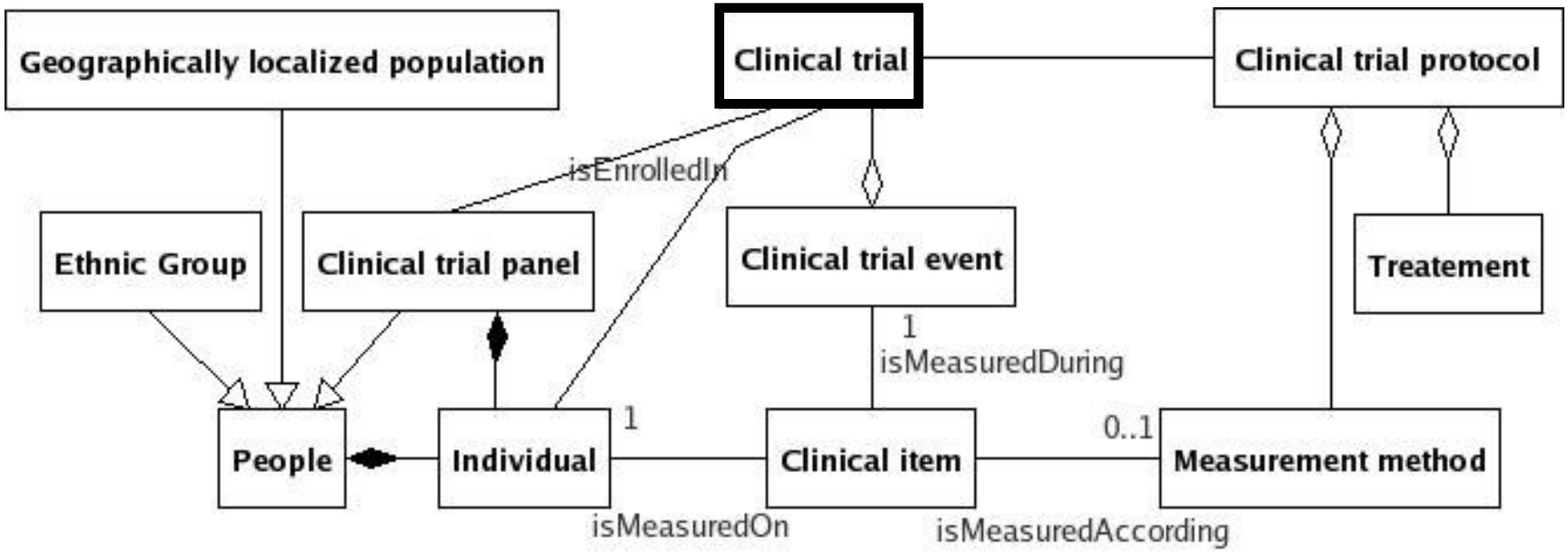
- Construct a formal ontology
  - based on existing information (i.e. models, databases, etc.)
    - List of useful data resources and knowledge resources
    - Various resource type: Databases, XML schemas, ontologies, etc.
    - *e.g.* PharmGKB, CDISC, ChEBI, SNP-Ontology, dbSNP, etc.

    N. B.: EBI Ontology Lookup Service

  - reusing existing ontologies as much as possible
    - List of reusable knowledge resources
    - *e.g.* PATO, SNP-Ontology, Disease Ontology, Mammalian Phenotype Ontology, etc.
    - May be extended with others (GO, Pathway Ontology, NCI, eVOC, etc.)

SO-Pharm Conceptualization

UML class diagram for **clinical trial**



SO-Pharm Formalization and Implementation(1/2)

- Manually formalized and implemented from UML to OWL
  - No automatic mechanism
  - Some guidelines exist: Ontology Design Patterns, reification
  - Complex classes need particular attention

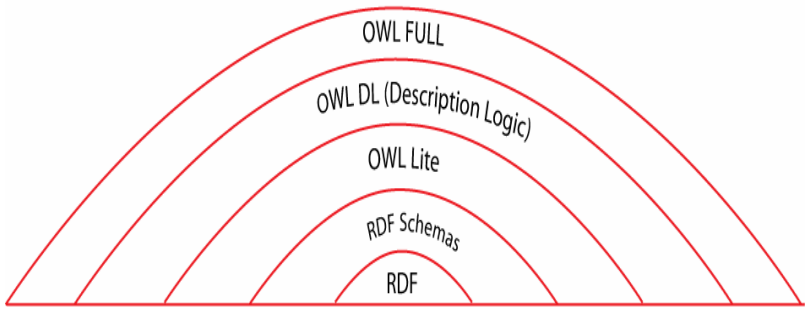


Image W3C/Ivan Herman

- Implemented in OWL DL
  - Expressivity > OWL Lite
    - *e.g.:* cardinality = 2 “a genomic variation is described by exactly 2 alleles”
  - Decidability of reasoning procedure ≠ OWL Full

- Implemented with Protégé and checked by Racer

SO-Pharm Validation (example)

- Example
  - *Patients* that are
    - homozygote for allele T for variant Chr22.26332887:C>T (rs1065852 in dbSNP)
    - treated with Codeine (orally, 100 mg, etc.)
    - blood concentration in morphine < 1.5nmol/l (according to a define protocol)
  - are *Poor metabolizers*