# ONTOLOGY CONSTRUCTION:

# 2 EXAMPLES IN PHARMACOGENOMICS

Adrien Coulet

EQUIPE ORPAILLEUR

Loria
Laboratoire lorrain de recherche en informatique et ses applications

Inria
informatics mathematics

# PHARMACOGENOMICS (PGx)

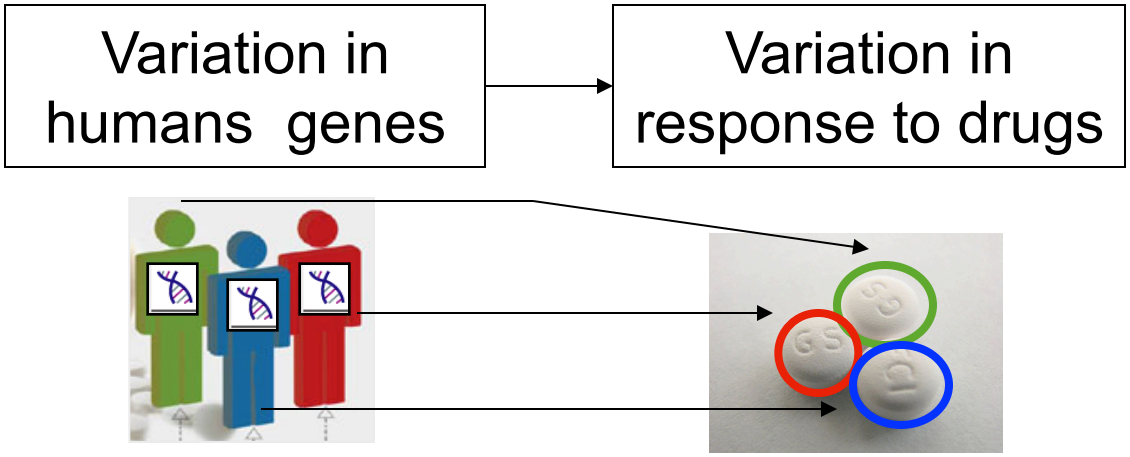



figure from Y. Garten

# OUTLINE

1. Principles of ontology construction

2. First example: SO-Pharm

   for data integration

3. Second example: PHARE

   for knowledge extraction from text

# METHODOLOGIES FOR ONTOLOGY CONSTRUCTION

- Inspirated from software engineering

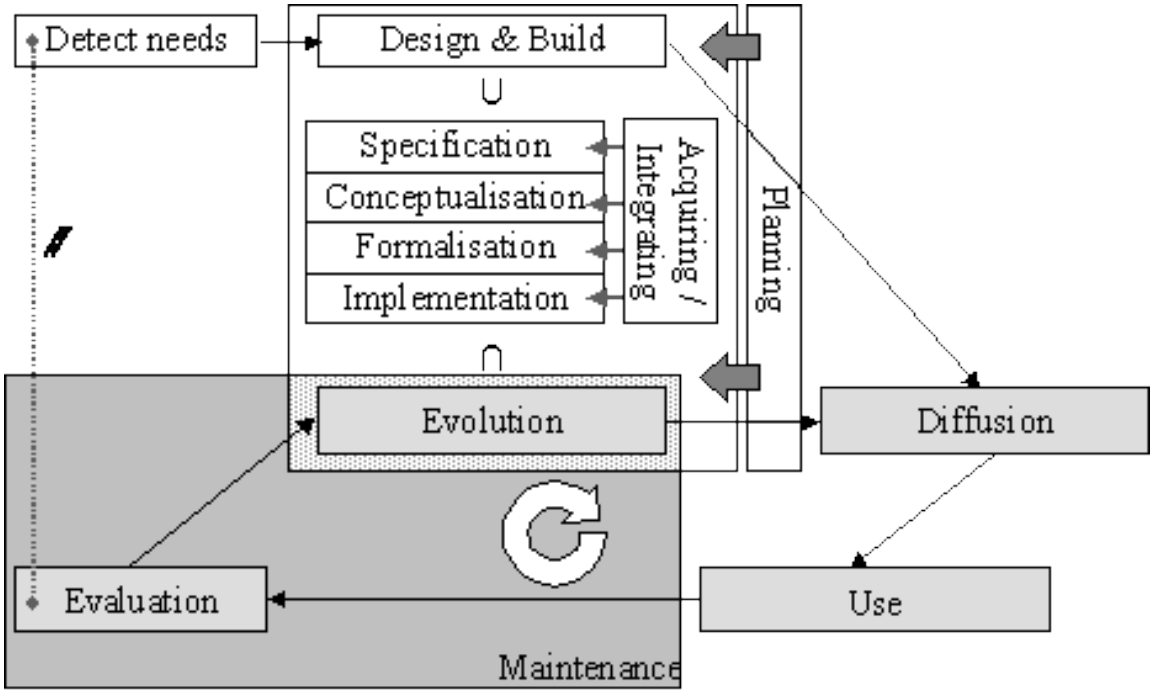


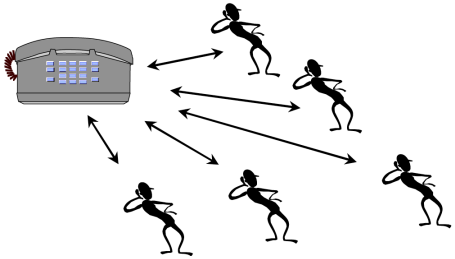

figure from F. Gandon

- For an overview of methodologies:

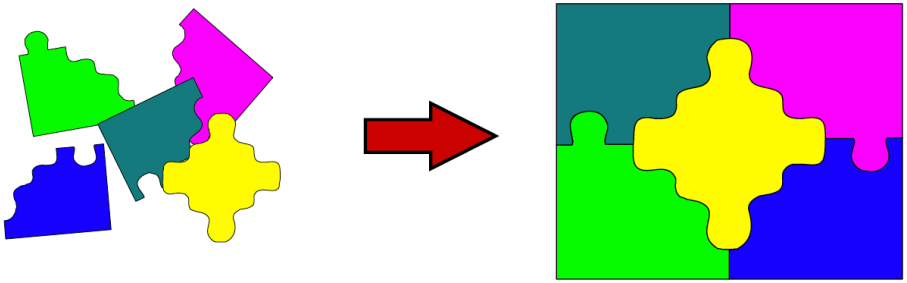
Asunción Gómez-Pérez *et al.:* Ontological Engineering, Springer (2004)

# TWO PRINCIPLES

- Ontology construction is a **COLLABORATIVE** process
  - Knowledge engineer and domain experts
  - Domain experts and domain experts
  - Knowledge engineer and knowledge engineer



  - enables the ontology to be reused by applications, projects
- Existing components must be **REUSED**



figures from A. Gómez-Pérez *et al.*

# **THREE** STEPS OF THE PROCESS

1. Specifications
   - ontology domain
   - granularity
   - **use +** (competency questions)
   - inventory of the domain data and knowledge resources
2. Conception
   - conceptualisation
   - formalisation
   - implementation
3. Evaluation

# THREE STEPS OF THE PROCESS

1. Specifications
   - ontology domain
   - granularity
   - **use +** (competency questions)
   - inventory of the domain data and knowledge resources
2. Conception
   - conceptualisation
   - formalisation
   - implementation
3. Evaluation

*<u>Example:</u> a Protein ontology*

*1.*

- *Human protein kinase family*
- *domain and subdomain, no physicochemical properties, no AA sequences*
- *to classify kinase proteins*
- *UniProt, PDB, PO, etc.*

*2.*

- *kinase, transmembrane domain, tyrosine kynase domain, etc.*
- $\mathtt{TyrosineKinase} \sqsubseteq \mathtt{Kinase}$

  $\mathtt{TyrosineKinase} \equiv \exists \mathtt{contains.TransmembraneDomain} \sqcap \exists \mathtt{contains.TyrosineKinaseDomain}$

...

```xml
<?xml version="1.0"?>
<rdf:RDF
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:protege="http://protege.stanford.edu/plugins/owl/protege#"
    xmlns:xsp="http://www.owl-ontologies.com/2005/08/07/xsp.owl#"
    xmlns:owl="http://www.w3.org/2002/07/owl#"
    xmlns="http://www.owl-ontologies.com/Ontology1318875168.owl#"
    xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
    xmlns:swrl="http://www.w3.org/2003/11/swrl#"
    xmlns:swrlb="http://www.w3.org/2003/11/swrlb#"
    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xml:base="http://www.owl-ontologies.com/Ontology1318875168.owl">
  <owl:Ontology rdf:about=""/>
  <owl:Class rdf:ID="Kinase"/>
  <owl:Class rdf:ID="TransmembraneDomain"/>
  <owl:Class rdf:ID="TyrosineKinase">
    <rdfs:subClassOf rdf:resource="#Kinase"/>
    <owl:equivalentClass>
      <owl:Class>
        <owl:intersectionOf rdf:parseType="Collection">
          <owl:Restriction>
            <owl:onProperty>
              <owl:ObjectProperty rdf:ID="contains"/>
            </owl:onProperty>
            <owl:someValuesFrom rdf:resource="#TransmembraneDomain"/>
          </owl:Restriction>
          <owl:Restriction>
            <owl:someValuesFrom>
              <owl:Class rdf:ID="TyrosineKinaseDomain"/>
            </owl:someValuesFrom>
            <owl:onProperty rdf:resource="#contains"/>
          </owl:Restriction>
        </owl:intersectionOf>
      </owl:Class>
    </owl:equivalentClass>
  </owl:Class>
</rdf:RDF>

<!-- Created with Protege (with OWL Plugin 3.4.6, Build 613)  http://protege.stanford.edu -->
```

File Edit Project OWL Reasoning Code Tools Window Collaboration Help

protégé

Metadata(Ontology1318875168.owl) | OWLClasses | Properties | Individuals | Forms

SUBCLASS EXPLORER

For Project:

Asserted Hier

- owl:Thing
  - Kinase
    - TyrosineKinase
  - TransmembraneDomain
  - TyrosineKinaseDomain

CLASS EDITOR for TyrosineKinase (instance of owl:Class)

For Class: TyrosineKinase

Inferred View

Annotations

| Property | Value | Lang |
|---|---|---|
| rdfs:comment | | |

Asserted Conditions

NECESSARY & SUFFICIENT

contains **some** TransmembraneDomain

contains **some** TyrosineKinaseDomain

NECESSARY

Kinase

# MANUAL *VS.* AUTOMATIC CONSTRUCTION (1/2)

1. Specifications
   - ontology domain
   - granularity
   - **use +** (competency questions)
   - inventory of the domain data and knowledge resources
2. Conception
   - conceptualisation
   - formalisation
   - implementation
3. Evaluation

# MANUAL *VS.* AUTOMATIC CONSTRUCTION (1/2)

1. Specifications
   - ontology domain
   - granularity
   - **use +** (competency questions)
   - inventory of the domain data and knowledge resources
2. Conception
   - conceptualisation
   - formalisation
   - implementation
3. Evaluation

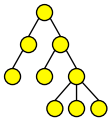
these steps can be automated

# MANUAL *VS.* AUTOMATIC CONSTRUCTION (2/2)

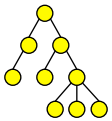
- The automation depends on either
  - representing what's in expert minds: *top-down approach*
  - representing what's in data (Data Bases, literatures, etc.): *bottom-up approach*

# TOP-DOWN VS. BOTTOM-UP

Expert



ontology

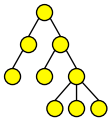


ontology

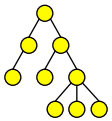
Data and knowledge resources

# TOP-DOWN (SO-PHARM) VS. BOTTOM-UP(PHARE)

PGx expert



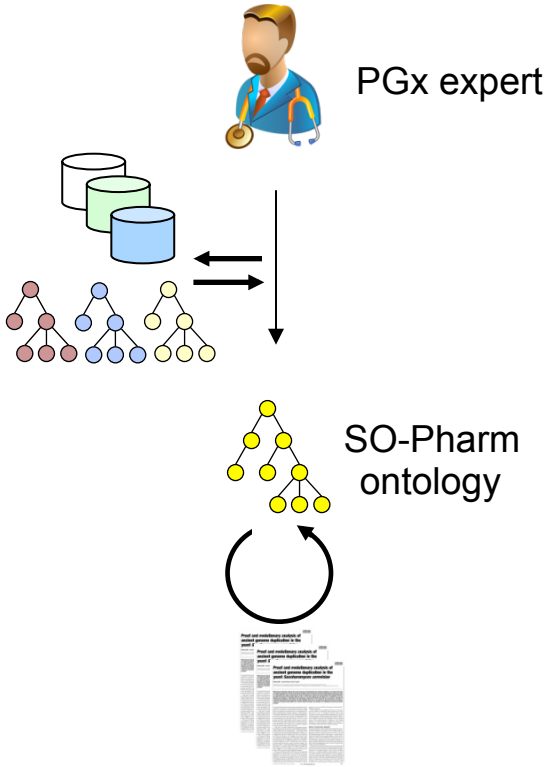
SO-Pharm ontology

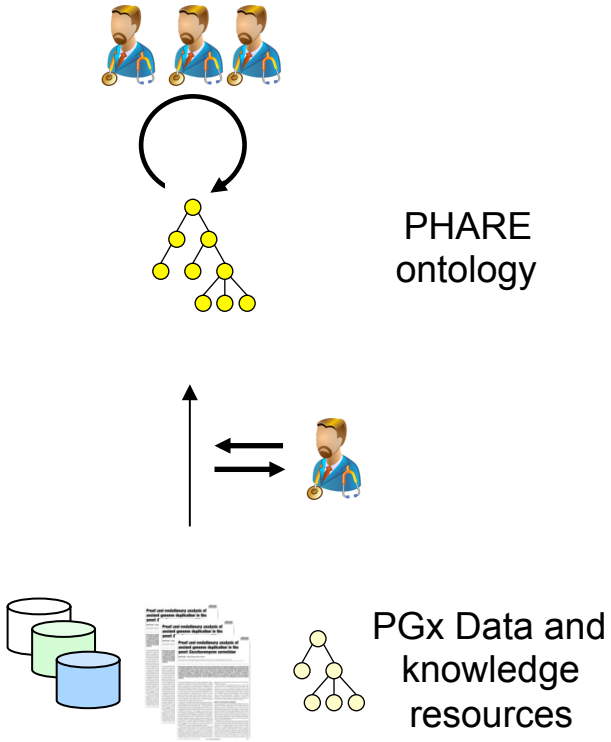


PHARE ontology

PGx Data and knowledge resources

# TOP-DOWN (SO-PHARM) VS. BOTTOM-UP(PHARE)




- Consensus on semi-automatic approaches

# OUTLINE

1. Principles of ontology construction

2. First example: SO-Pharm

   for data integration

3. Second example: PHARE

   for knowledge extraction from text

# WHY A PGx ONTOLOGY?

- To integrate
  - phenotype data
  - genotype data
  - drug data

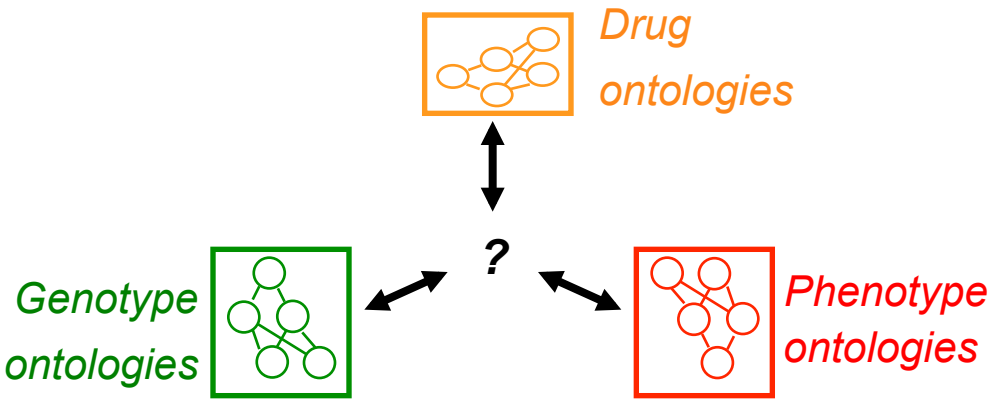
  both from clinical trial and biomedical resource

- And to serve as a guide to discover genotype-phenotype-drug relationships

# INVENTORY OF RESOURCES (1/2)

| *Resource name* | *Resource type* | *Domain* | *URL* |
|---|---|---|---|
| dbSNP | XML schema, data model | genotype | http://www.ncbi.nlm.nih.gov/projects/SNP/ |
| HapMap | XML schema | genotype | http://www.hapmap.org/ |
| HGVBase | DTD, data model | genotype | http://hgvbase.cgb.ki.se/ |
| OMIM | Data resource | genotype, phenotype | http://www.ncbi.nlm.nih.gov/omim/ |
| OMG SNP | Data model | genotype | http://www.omg.org/technology/documents/formal/snp.htm |
| MECV | Controlled vocabulary | genotype | http://www.ebi.ac.uk/mutations/ |
| PharmGKB | XML schema, data model | genotype, drug, phenotype | http://www.pharmgkb.org/ |
| Pharmacogenetics Ontology | Controlled vocabulary | genotype, phenotype | http://www.pharmgkb.org/home/projects/project-po.jsp |
| Sequence Ontology | Controlled vocabulary* | genotype | http://song.sourceforge.net/ |
| Gene Ontology | Controlled vocabulary* | genotype | http://www.geneontology.org/ |
| PubChem | Data resource | drug | http://pubchem.ncbi.nlm.nih.gov/ |
| RX-Norm | Controlled vocabulary | drug | http://www.nlm.nih.gov/research/umls/rxnorm/index.html |
| CDISC | XML schema | phenotype | http://www.cdisc.org/ |
| ICD-10 | Controlled vocabulary | phenotype | http://www.who.int/classifications/icd/ |
| Disease Ontology | Controlled vocabulary* | phenotype | http://diseaseontology.sourceforge.net |
| Mammalian Phenotype | Controlled vocabulary* | phenotype | http://www.informatics.jax.org/searches/MP_form.shtml |
| PATO | Controlled vocabulary* | phenotype | http://obo.sourceforge.net/ |
| ChEBI | Controlled vocabulary* | drug | http://www.ebi.ac.uk/chebi/ |
| Pathway Ontology | Controlled vocabulary* | genotype, phenotype | http://rgd.mcw.edu/tools/ontology |
| SNOMED-Clinical | Controlled vocabulary | phenotype | http://www.snomed.org/snomedct/glossary.html |

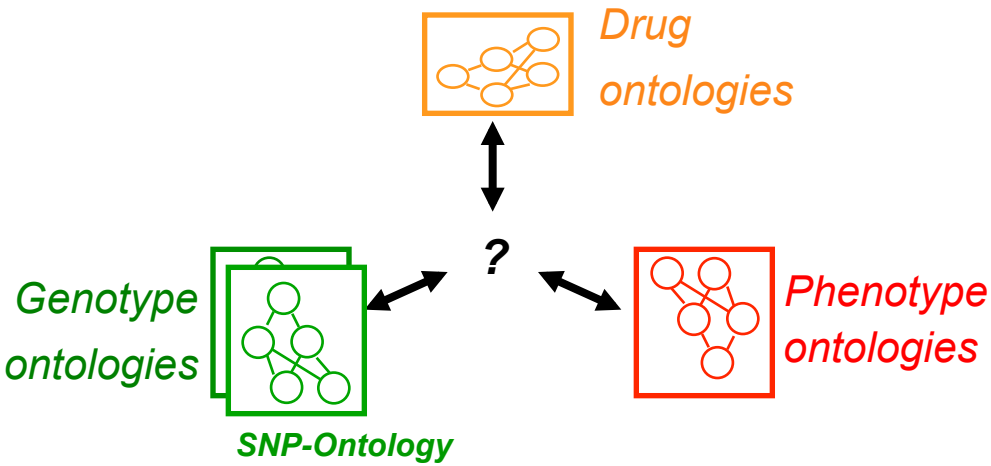
# NEED FOR A PGx ONTOLOGY




No ontology for genomic variations

No ontology to articulate genotype, phenotype and drug concepts

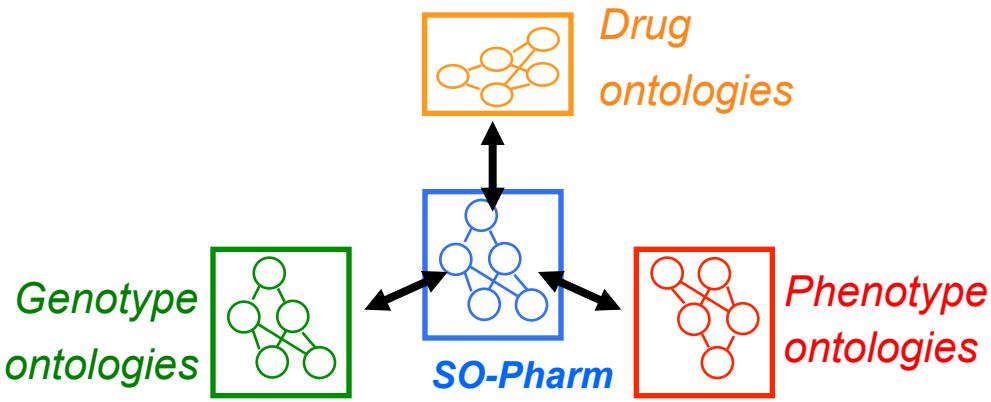
# NEED FOR A PGx ONTOLOGY




No ontology for genomic variations

No ontology to articulate genotype, phenotype and drug concepts

# NEED FOR A PGx ONTOLOGY




No ontology for genomic variations

No ontology to articulate genotype, phenotype and drug concepts

# INVENTORY OF RESOURCES (2/2)

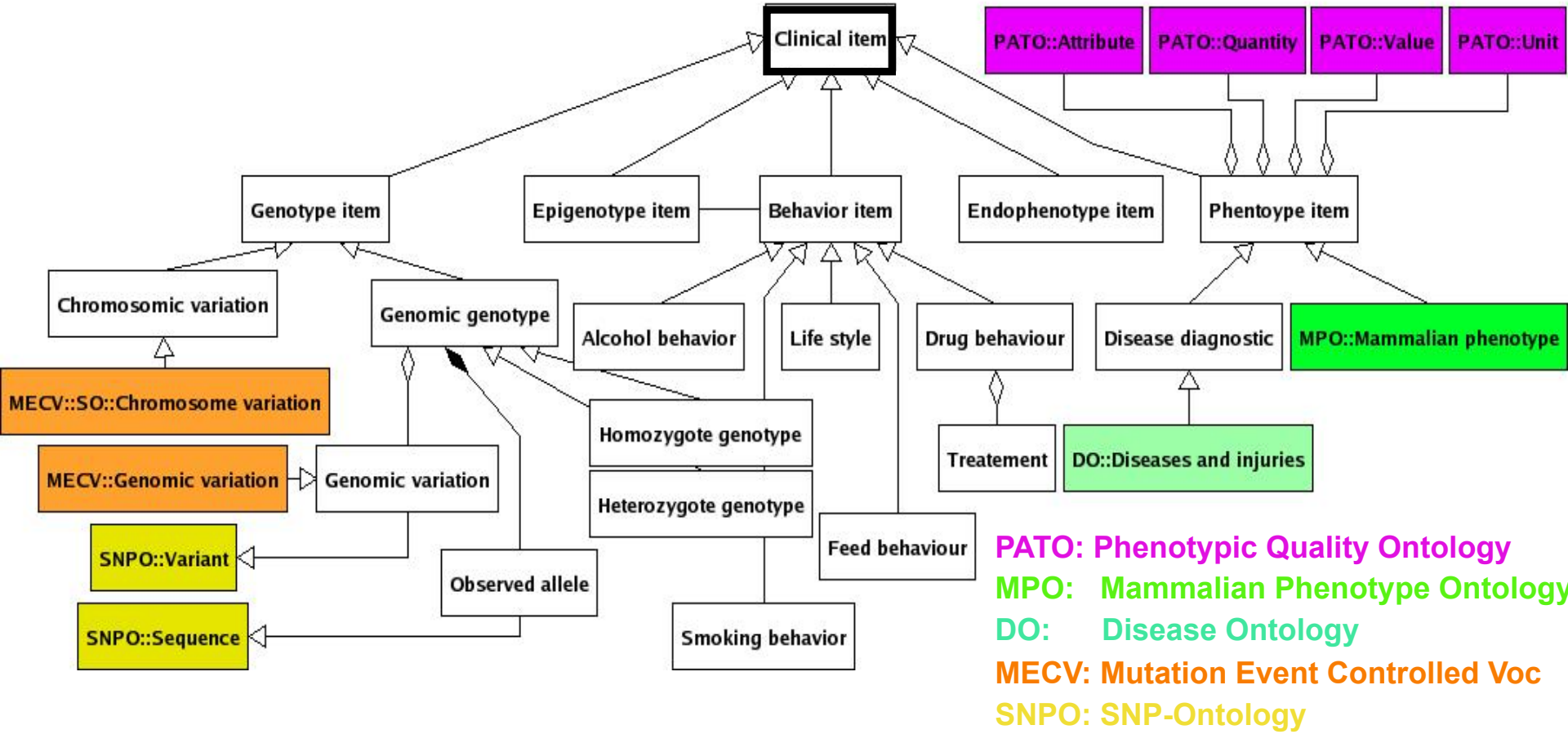
- List of ontologies to reuse (partial)

| *Ontology name* | *Description* | *Prefix* | *Namespace* |
|---|---|---|---|
| MECV | genomic variation classification | MECV | http://www.loria.fr/~coulet/ontology/mecv.owl |
| SNP-Ontology | genomic variations | SNPO | ~/ontology/snpontology.owl |
| Pharmacogenetics Ontology | describes genotyping and phenotyping methods | PO | ~/ontology/pharmacogeneticsontology.owl |
| Disease Ontology | a classification of disease | DO | ~/ontology/diseaseontology.owl |
| Mammalian Phenotype | phenotype features | MPO | ~/ontology/mammalianphenotypeontology.owl |
| PATO | attributes and values for phenotype description | PATO | ~/ontology/pato.owl |
| ChEBI | molecular compounds | CHEBI | ~/ontology/chebi.owl |

- List of concepts not represented

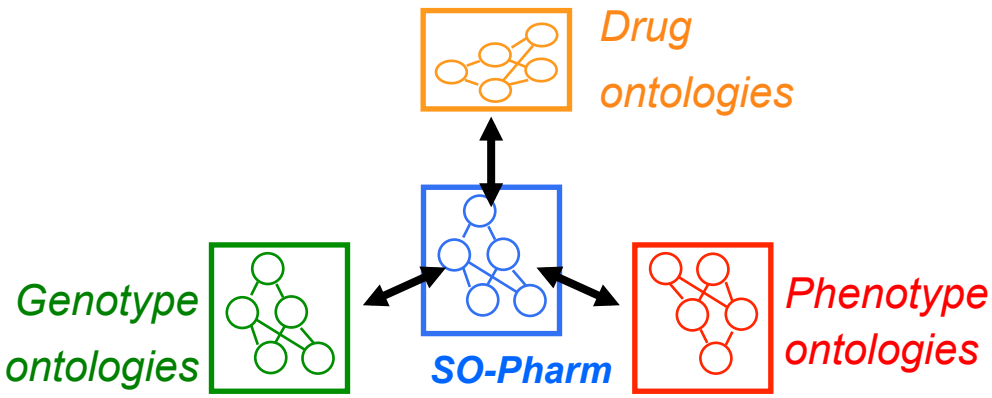
# CONCEPTUALISATION

UML class diagram around the concept of Clinical Item




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

# SO-PHARM




## Formalisation and implementation

- OWL DL with Protégé
- 70 concepts that articulate 15 ontologies

## Diffusion

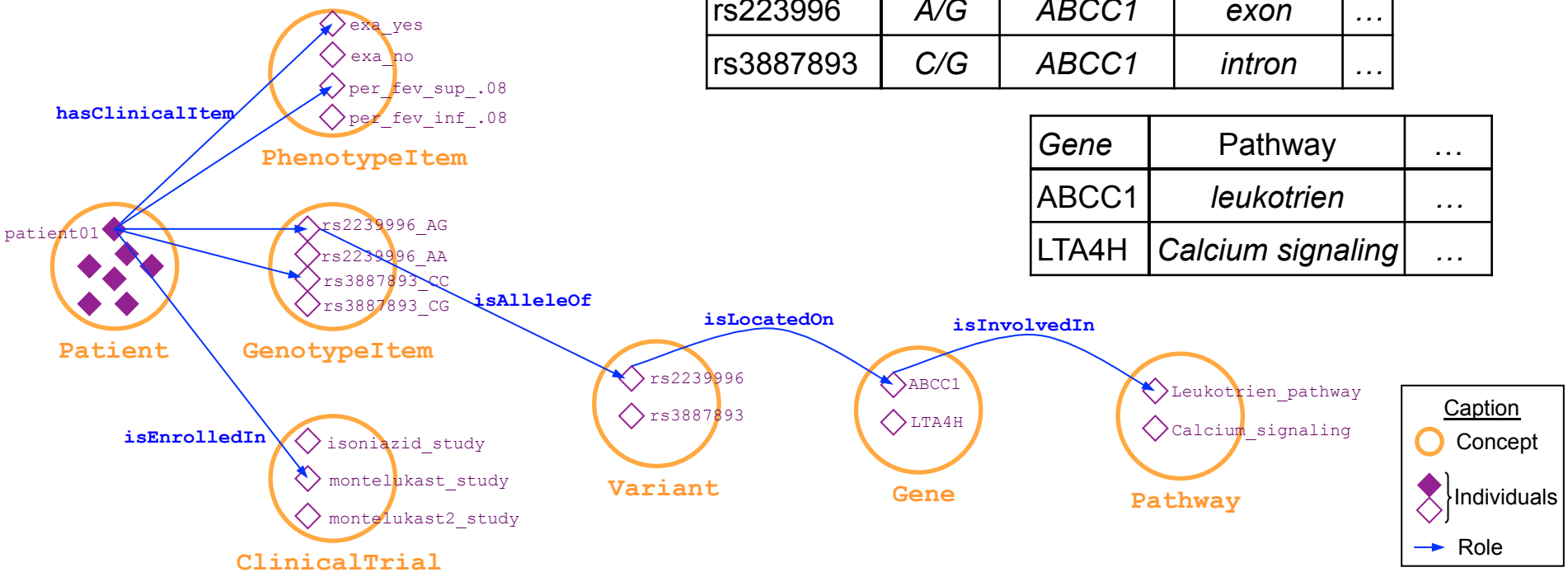
- OBO Foundry
- BioPortal

# USE OF SO-PHARM (1/2)

| *Montelukast Study* | Phenotype | | Genotype | | | |
|---|---|---|---|---|---|---|
| | Exacerbation (Crise d'asthme) | %Change in FEV (VEMS) | rs2239996 | rs3887893 | rs4148356 | … |
| patient01 | yes | 0.01 | A/G | C/C | G/G | … |
| patient02 | no | 0.1 | A/G | C/G | G/G | … |

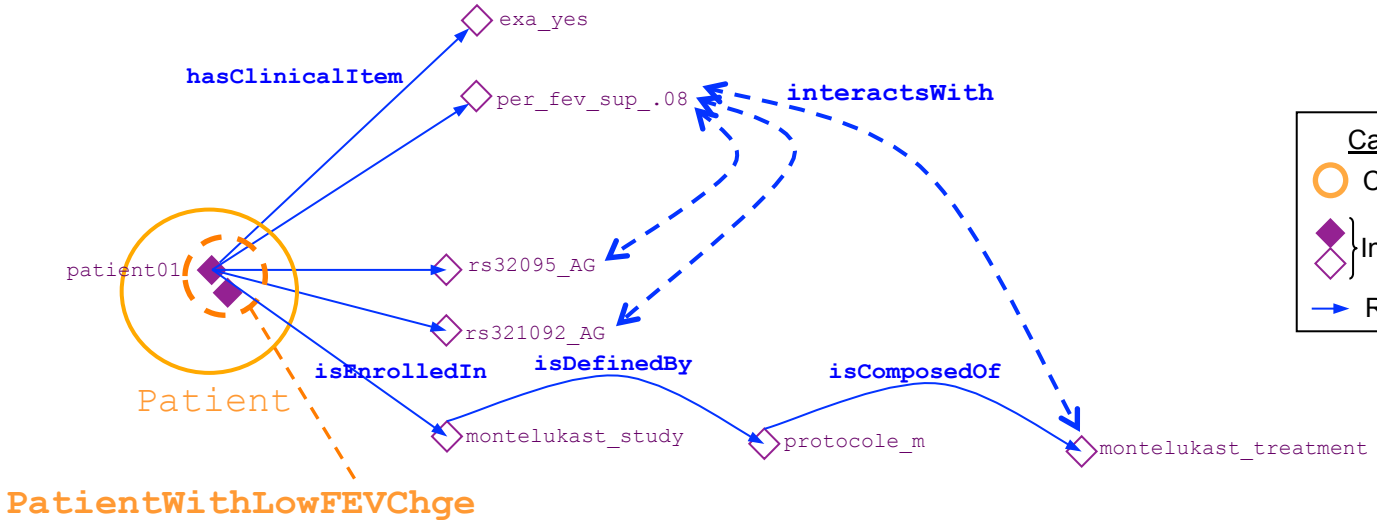
| *dbSNP* | Variant | Gene | Region | … |
|---|---|---|---|---|
| rs223996 | *A/G* | *ABCC1* | *exon* | … |
| rs3887893 | *C/G* | *ABCC1* | *intron* | … |

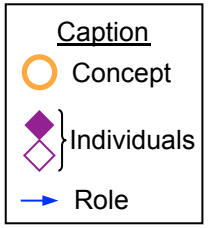
| *Gene* | Pathway | … |
|---|---|---|
| ABCC1 | *leukotrien* | … |
| LTA4H | *Calcium signaling* | … |



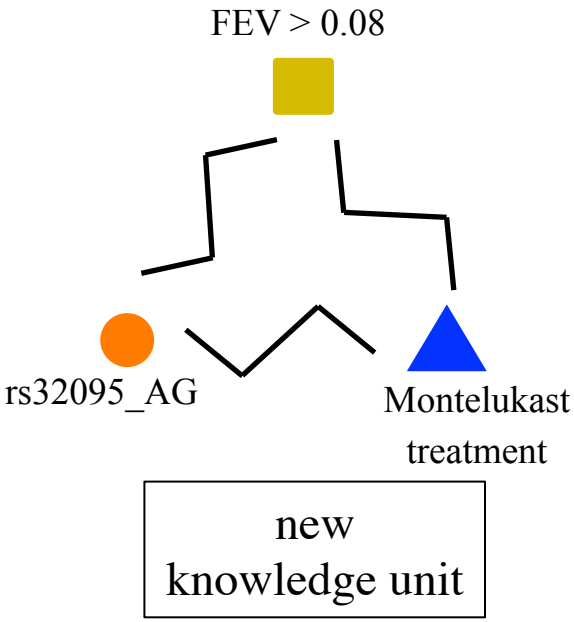
# USE OF SO-PHARM (2/2)

[COULET ET AL., *ADV EXP MED BIOL*, 2011]




- New & pertinent knowledge
  - Description of patient sub-groups
  - Description of relationship between SNPs
  - Description of genotype – disease – drug relationships
- Extend results from initial clinical study



# OUTLINE

1. Principles of ontology construction

2. First example: SO-Pharm

   for data integration

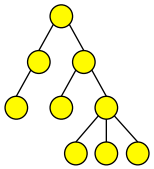
3. Second example: PHARE

   for knowledge extraction from text

# NEED TO NORMALISE PGx RELATIONSHIPS

"SNPs in ABCB1 increase methotrexate sensitivity" →

increase (ABCB1_SNPs, methotrexate_sensitivity)



"P-gp variants augment the intolerance to methopterin" →

augment (P-gp_variants, methopterin_intolerance)

→ `increases` (*ABCB1_variant*, *methotrexate_sensitivity*)

# WE OBTAINED RELATIONS AND CONCEPTS AUTOMATICALLY

increase ( ABCB1_gene, methotrexate_effect )

modifier head entity modifier head entity

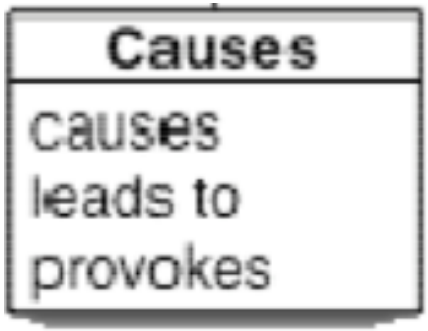
type subject object

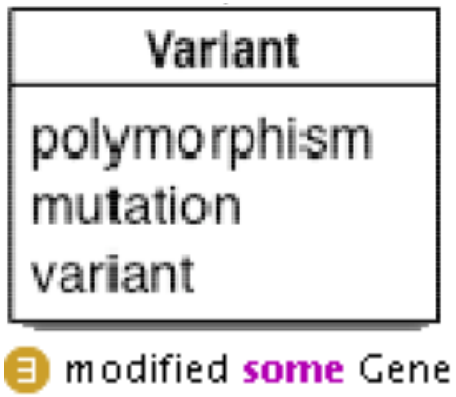
…out of Medline

| Relationship types | Entities modified by | | |
|---|---|---|---|
| | Genes | Drugs | Phenotypes |
| 2538 associate | 1237 *gene* | 377 *metabolism* | 304 *cell* |
| 1017 increase | 1000 *inhibitor* | 358 *activity* | 114 *line* |
| 985 inhibit | 935 *polymorphism* | 298 *inhibitor* | 101 *patient* |
| 825 induce | 775 *expression* | 267 *effect* | 71 *risk* |
| 763 metabolize | 773 *activity* | 263 *administration* | 35 *tissue* |
| 666 involve | 689 *mutation* | 246 *channel* | 34 *specimen* |
| 643 reduce | 685 *genotype* | 242 *treatment* | 33 *case* |
| 547 catalyze | 393 *inhibition* | 193 *antagonist* | 27 *treatment* |
| 515 cause | 329 *level* | 178 *concentration* | 26 *rate* |
| 509 affect | 245 *gene_mutation* | 172 *dose* | 26 *effect* |

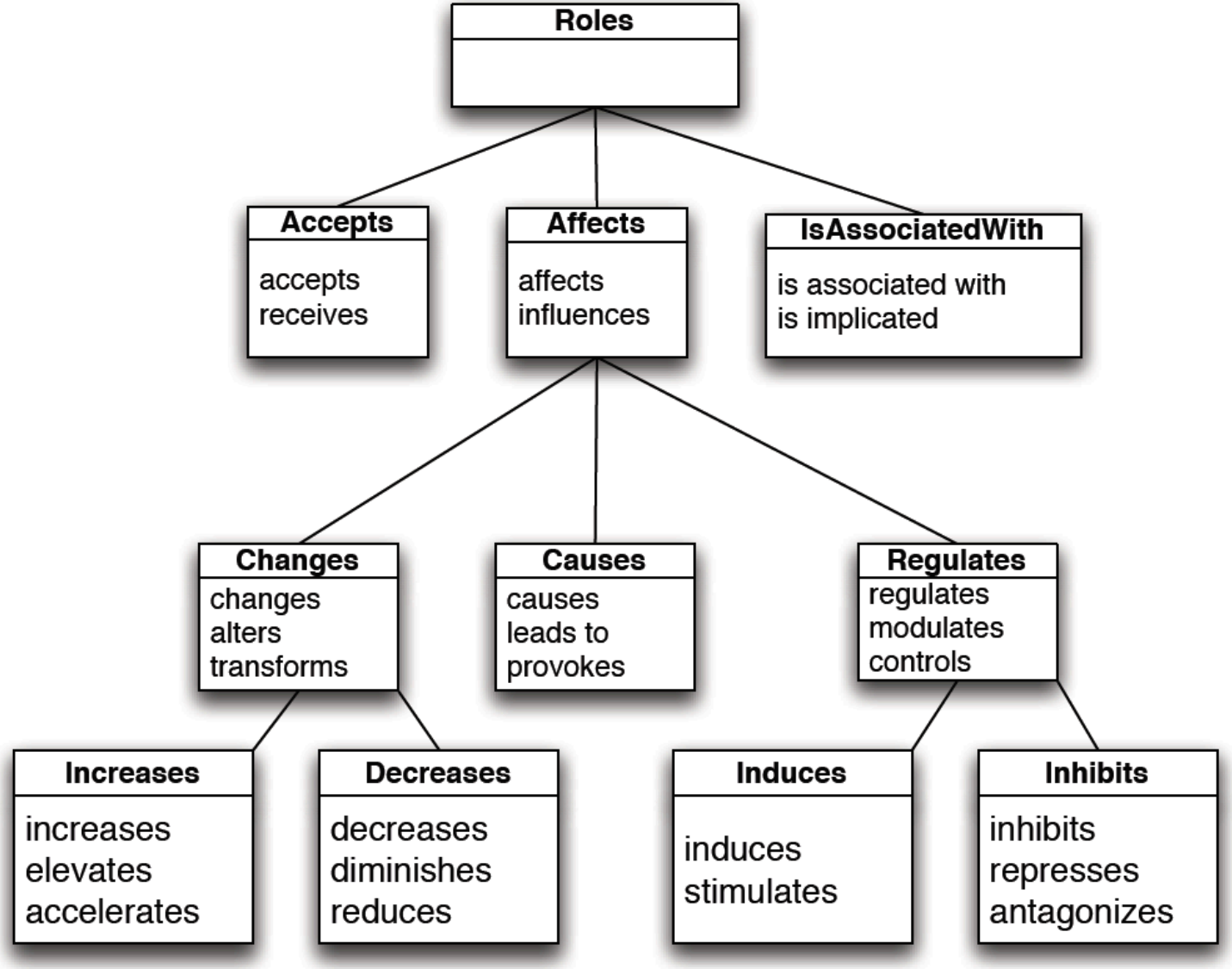
# WE GROUPED SYNONYMS MANUALLY

| Relationship types | Entities modified by | | |
|---|---|---|---|
| | Genes | Drugs | Phenotypes |
| 2538 associate | 1237 *gene* | 377 *metabolism* | 304 *cell* |
| 1017 increase | 1000 *inhibitor* | 358 *activity* | 114 *line* |
| 985 inhibit | 935 *polymorphism* | 298 *inhibitor* | 101 *patient* |
| 825 induce | 775 *expression* | 267 *effect* | 71 *risk* |
| 763 metabolize | 773 *activity* | 263 *administration* | 35 *tissue* |
| 666 involve | 689 *mutation* | 246 *channel* | 34 *specimen* |
| 643 reduce | 685 *genotype* | 242 *treatment* | 33 *case* |
| 547 catalyze | 393 *inhibition* | 193 *antagonist* | 27 *treatment* |
| 515 cause | 329 *level* | 178 *concentration* | 26 *rate* |
| 509 affect | 245 *gene_mutation* | 172 *dose* | 26 *effect* |

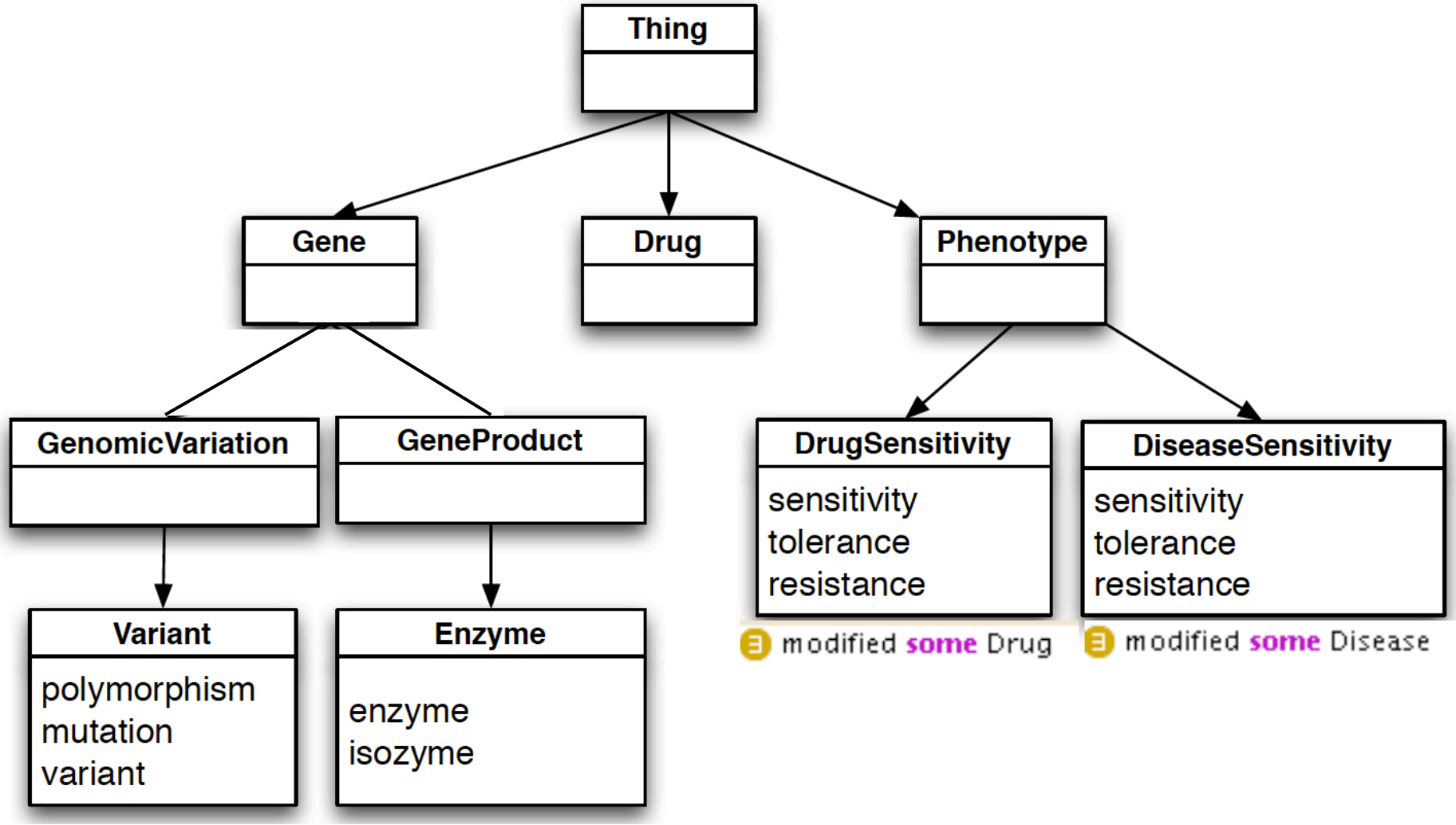


# SNAPSHOT OF THE ROLE HIERARCHY



# SNAPSHOT OF THE CONCEPT HIERARCHY

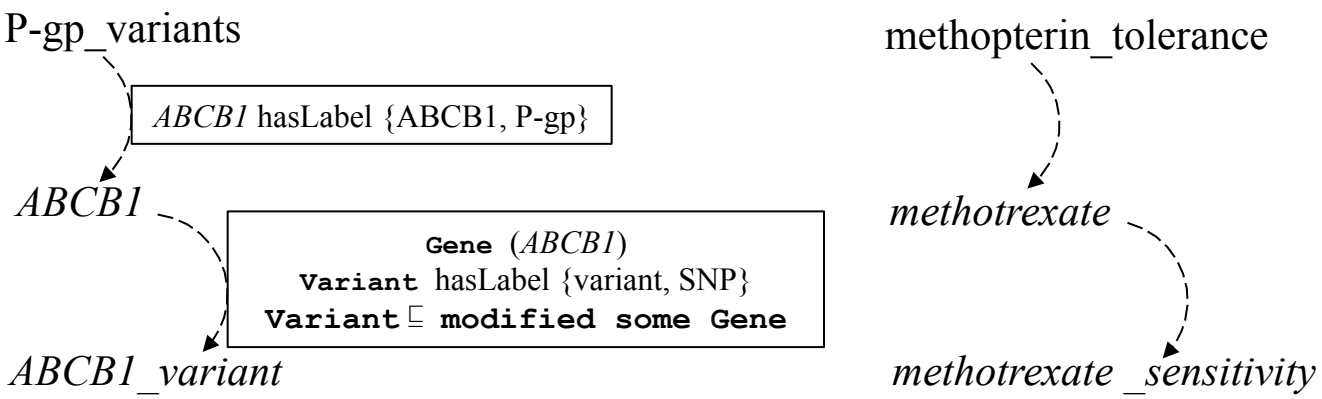


# NORMALISATION

**Start**: affect ( P-gp_variants, methopterin_tolerance )

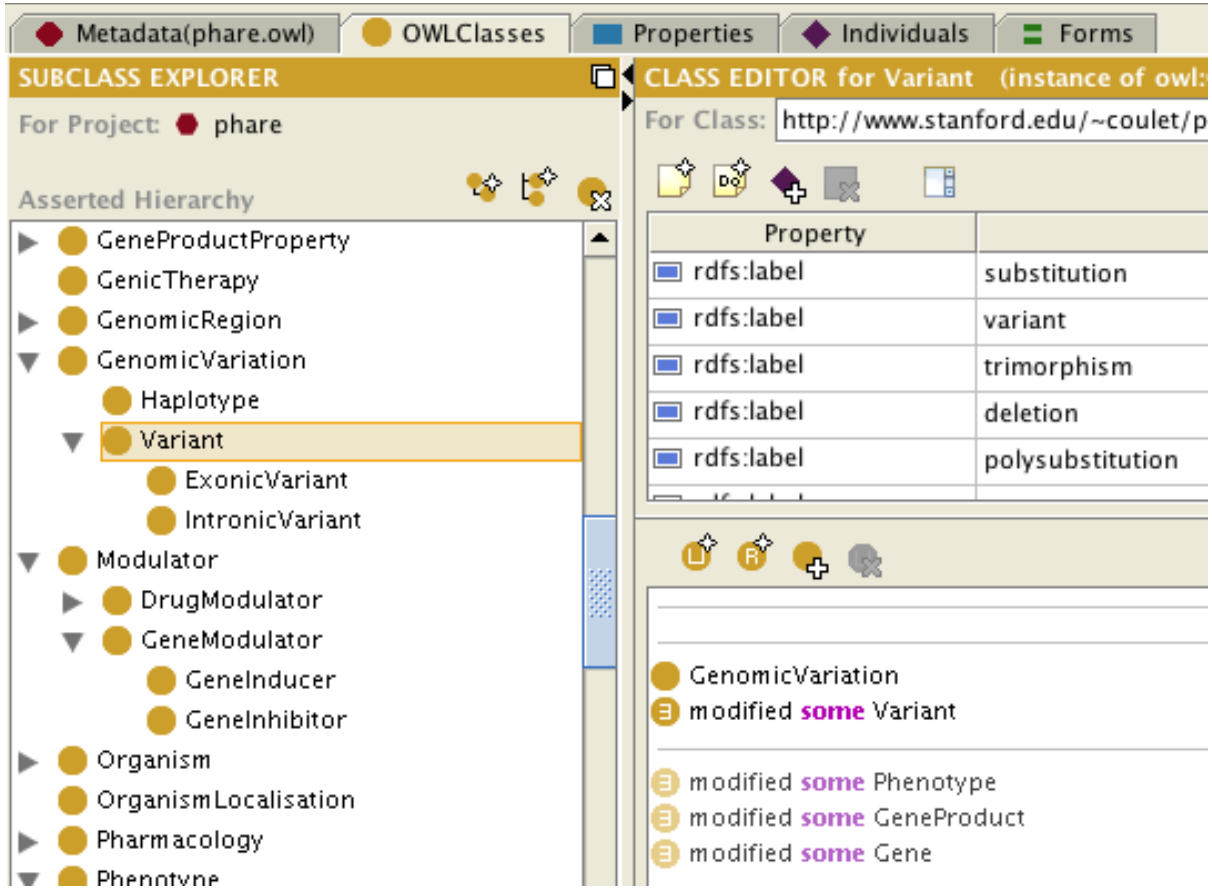
- Normalisation of the relation type


- Normalisation of the subject and of the object




**End**: increases ( *ABCB1_variant*, *methotrexate_sensitivity* )

# PHARE ONTOLOGY: GENE VARIANT



# PHARE ONTOLOGY: DRUG SENSITIVITY

Metadata(phare.owl) | OWLClasses | Properties | Individuals | Forms

SUBCLASS EXPLORER

For Project: phare

Asserted Hierarchy

- Phenotype
  - ChemicalModification ▸ DrugTransfor
  - Death
  - Disease
  - DiseaseAbsence
  - DiseaseCause
  - DiseaseDuration
  - DiseaseEffect
  - DiseaseExacerbation
  - DiseaseRelapse
  - DiseaseRelief
  - DiseaseSensitivity
  - DiseaseSeverity
  - DiseaseSurvival
  - DrugEffect ▸ DrugPharmacodynamics
  - DrugSensitivity
  - Expression
  - GeneOrGeneProductFunction
  - GeneProductActivityChange

CLASS EDITOR for DrugSensitivity (instance of owl:Class)

For Class: http://www.stanford.edu/~coulet/phare.owl#DrugSensitivity

| Property | |
|---|---|
| rdfs:label | susceptibilty |
| rdfs:label | sensitivity |
| rdfs:label | sensitization |
| rdfs:label | desensitization |
| rdfs:label | tolerability |

Phenotype

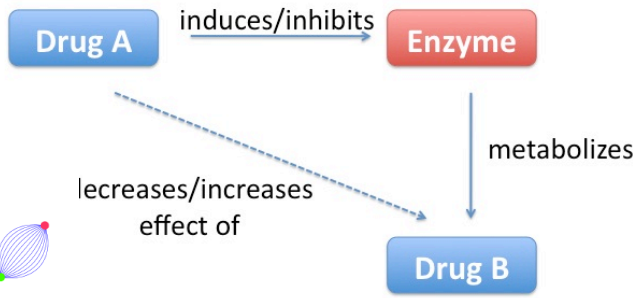
∃ modified **some** Drug

# PHARE TRIPLET

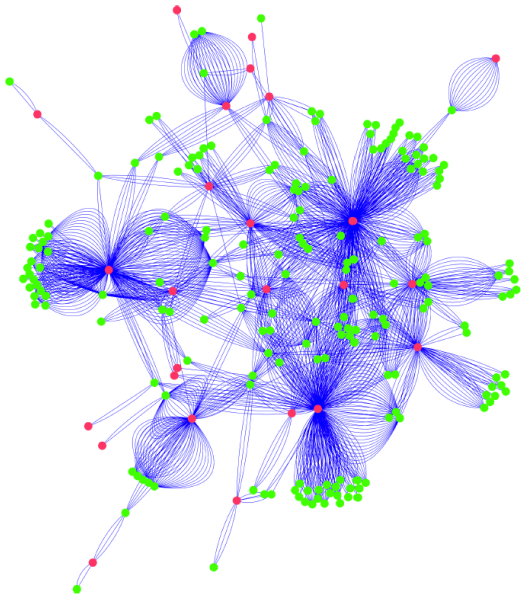
| Property | Value |
| --- | --- |
| owl:object | phare:warfarin_effect |
| owl:predicate | phare:affect |
| owl:subject | phare:vkorc1_variant |
| rdfs:comment | [15930419 , Variants in the gene encoding vitamin K epoxide reductase complex 1 (VKORC1) may affect the response to warfarin] |

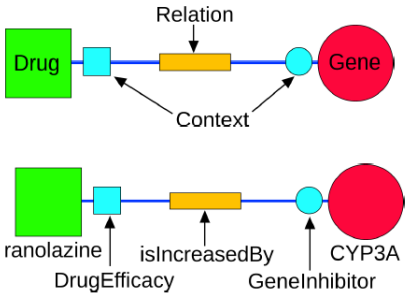
# USE OF PHARE

- Knowledge extraction from text
  - to guide PharmGKB curators
  - to propose PharmGKB users additional content
- Discover new knowledge
  - using reasoning




  - using machine learning






Bethany Percha *et al., PSB 2012*
*with drug-drug interactions*

# CONCLUSIONS

- Not perfect examples but illustrative
  - lots of construction choices can be discussed

    *"top-down for data integration?"*

    *"why not using WordNet for relationship normalization?"*
- New tools make construction easier
  - the *Bioportal* for diffusion

    for searching knowledge resources
  - the *NCBO Resource Index* for searching data reasources
  - the *NCBO Ontology Recommender*
- A never ending process:
  - Attempt to build a PGx ontology for other use cases
  - with the HCLS Interest Group @ w3c

# THANKS