



PGxO: A very lite ontology to reconcile pharmacogenomic knowledge units

Pierre Monnin¹, Clément Jonquet^{2,3}, Joël Legrand¹, Amedeo Napoli¹ and Adrien Coulet^{1,3}

October 18, 2017

¹LORIA (CNRS, Inria Nancy-Grand Est, Université de Lorraine), Nancy, France

²LIRMM (CNRS, Université de Montpellier), Montpellier, France

³BMIR (Stanford University), Stanford, USA



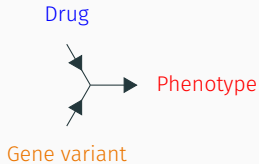
Stanford
MEDICINE

Context & Motivation

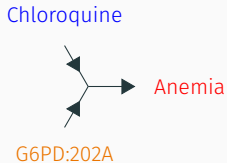
Context & Motivation

Pharmacogenomics (PGx): studies the influence of gene variants in drug response

Knowledge unit of pharmacogenomics:



Example: influence of a variant of the gene G6PD in patient reactions to a chloroquine treatment



Knowledge sources for PGx relationships:

- State-of-the-art knowledge:
 - Reference databases (*e.g.*, PharmGKB)
 - Scientific biomedical literature
- May also be discovered by mining Electronic Health Records (EHRs)

⇒ Knowledge of various quality, granularity and vocabulary

Knowledge sources for PGx relationships:

- State-of-the-art knowledge:
 - Reference databases (*e.g.*, PharmGKB)
 - Scientific biomedical literature
- May also be discovered by mining Electronic Health Records (EHRs)

⇒ Knowledge of various quality, granularity and vocabulary

Available PGx knowledge is increasing:

- In 2008, from 17,000,000 abstracts in Medline, extraction of 40,000 PGx relationships
- Now, 27,000,000 abstracts available

Motivation: building a framework for comparing PGx knowledge of various origins (DBs, literature, EHRs)

One objective of the PractiKPharma project (funded by the ANR)¹

Advantages:

- Asset for personalized medicine
- Confirming / moderating / contextualizing PGx knowledge of different sources

¹<http://practikpharma.loria.fr/>

Specifications:

- Representing basic elements of PGx knowledge units (drugs, genomic variations, phenotypes)

Specifications:

- Representing basic elements of PGx knowledge units (drugs, genomic variations, phenotypes)
- Associating provenance to PGx relationships:
 - Source(s) of the relationships
 - Confidence of the relationships
 - Confidence given in the source (*e.g.*, level of evidence in PharmGKB)
 - Confidence given by the mining algorithm
 - Version of sources and algorithms
 - Date of extraction
 - ...

Specifications:

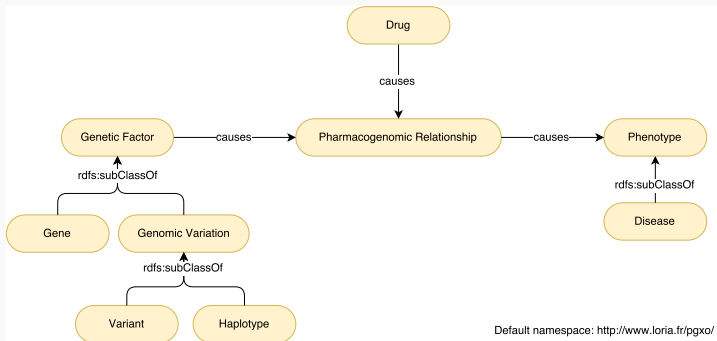
- Representing basic elements of PGx knowledge units (drugs, genomic variations, phenotypes)
- Associating provenance to PGx relationships:
 - Source(s) of the relationships
 - Confidence of the relationships
 - Confidence given in the source (*e.g.*, level of evidence in PharmGKB)
 - Confidence given by the mining algorithm
 - Version of sources and algorithms
 - Date of extraction
 - ...
- Defining rules to confront PGx relationships:
 - Two relationships from two sources can represent the same knowledge unit
 - One can be a more precise version of the other

Representing pharmacogenomic knowledge units with PGxO

Representing pharmacogenomic knowledge units with PGxO

PGxO:

- Implemented in OWL
- Lightweight ontology: 9 concepts, 4 relations
- Ability to represent ternary relationships between:
 - One (or more) genetic factor
 - One (or more) drug treatment
 - One (or more) phenotype



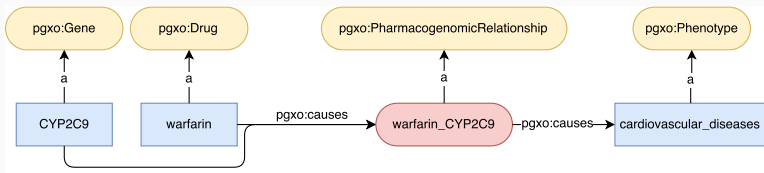
Representing pharmacogenomic knowledge units with PGxO

Example:



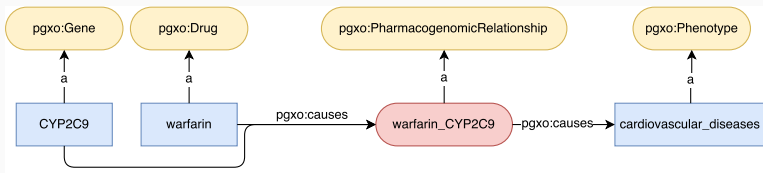
Representing pharmacogenomic knowledge units with PGxO

Example:



Representing pharmacogenomic knowledge units with PGxO

Example:



How to associate knowledge units with their provenance?

Associating provenance with PROV-O

PROV-O, an ontology for provenance

PROV-O:

- W3C recommendation (April 30, 2013)
- Represents provenance information with different granularity levels

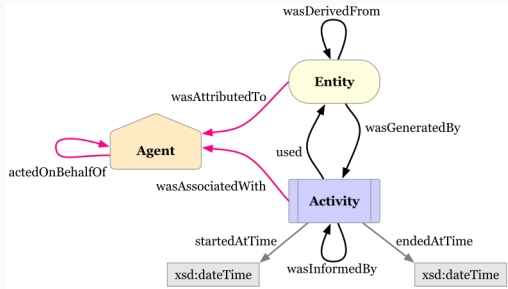
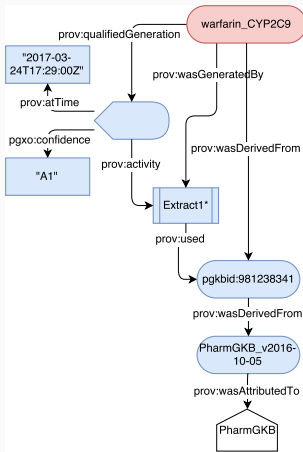


Figure 1: PROV-O main entities and relationships (from <https://www.w3.org/TR/prov-o/>)

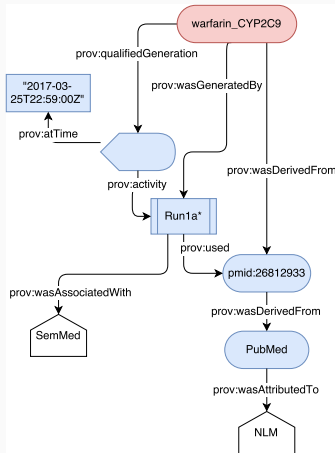
PROV-O, an ontology for provenance

Example: Represent provenance of a knowledge unit defined in PharmGKB



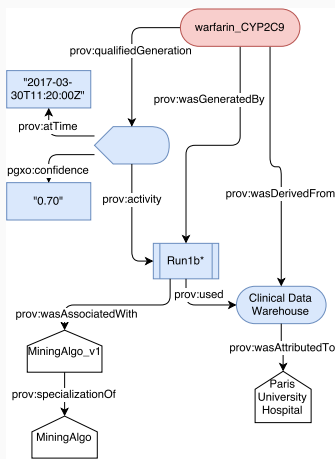
PROV-O, an ontology for provenance

Example: Represent provenance of a knowledge unit extracted from the literature



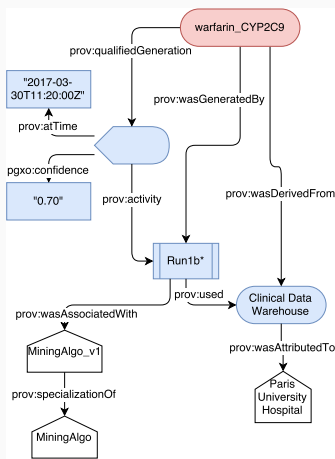
PROV-O, an ontology for provenance

Example: Represent provenance of a knowledge unit extracted from EHRs



PROV-O, an ontology for provenance

Example: Represent provenance of a knowledge unit extracted from EHRs



How to reconcile identical knowledge units from different sources?

Reconciling pharmacogenomic knowledge units

Reconciling pharmacogenomic knowledge units

Considering two PGx relationships r_1 and r_2 .

We define the following classes:

- Drugs:
 - $D_1 \equiv Drug \sqcap \exists causes.\{r_1\}$
 - $D_2 \equiv Drug \sqcap \exists causes.\{r_2\}$
- Genetic Factors:
 - $G_1 \equiv GeneticFactor \sqcap \exists causes.\{r_1\}$
 - $G_2 \equiv GeneticFactor \sqcap \exists causes.\{r_2\}$
- Phenotypes:
 - $P_1 \equiv Phenotype \sqcap \exists isCausedBy.\{r_1\}$
 - $P_2 \equiv Phenotype \sqcap \exists isCausedBy.\{r_2\}$

Reconciling pharmacogenomic knowledge units

$D_i \equiv \text{Drug} \sqcap \exists \text{causes}.\{r_i\}$



$G_i \equiv \text{GeneticFactor} \sqcap \exists \text{causes}.\{r_i\}$

Inference rules:

1. $D_1 \equiv D_2 \sqcap G_1 \equiv G_2 \sqcap P_1 \equiv P_2 \Rightarrow \{r_1\} \equiv \{r_2\}$
 r_1 and r_2 are referring to the same PGx relationship

Reconciling pharmacogenomic knowledge units

$D_i \equiv \text{Drug} \sqcap \exists \text{causes}.\{r_i\}$



$G_i \equiv \text{GeneticFactor} \sqcap \exists \text{causes}.\{r_i\}$

Inference rules:

1. $D_1 \equiv D_2 \sqcap G_1 \equiv G_2 \sqcap P_1 \equiv P_2 \Rightarrow \{r_1\} \equiv \{r_2\}$
 r_1 and r_2 are referring to the same PGx relationship
2. $D_1 \sqsubseteq D_2 \sqcap G_1 \sqsubseteq G_2 \sqcap P_1 \sqsubseteq P_2 \Rightarrow \{r_1\} \sqsubseteq \{r_2\}$
 r_1 is more specific than r_2

Reconciling pharmacogenomic knowledge units

$$D_i \equiv \text{Drug} \sqcap \exists \text{causes}.\{r_i\}$$



$$G_i \equiv \text{GeneticFactor} \sqcap \exists \text{causes}.\{r_i\}$$

Inference rules:

- $D_1 \equiv D_2 \sqcap G_1 \equiv G_2 \sqcap P_1 \equiv P_2 \Rightarrow \{r_1\} \equiv \{r_2\}$
 r_1 and r_2 are referring to the same PGx relationship
- $D_1 \sqsubseteq D_2 \sqcap G_1 \sqsubseteq G_2 \sqcap P_1 \sqsubseteq P_2 \Rightarrow \{r_1\} \sqsubseteq \{r_2\}$
 r_1 is more specific than r_2
- $(D_1 \sqsubseteq D_2 \sqcap G_1 \sqsubseteq G_2 \sqcap P_2 \equiv \perp) \sqcup$
 $(D_1 \sqsubseteq D_2 \sqcap G_2 \equiv \perp \sqcap P_1 \sqsubseteq P_2) \sqcup$
 $(D_2 \equiv \perp \sqcap G_1 \sqsubseteq G_2 \sqcap P_1 \sqsubseteq P_2)$
 $\Rightarrow \{r_1\} \sqsubseteq \{r_2\}$ r_1 is more specific than r_2

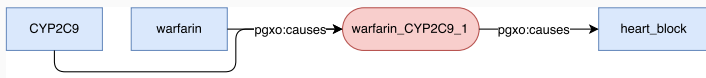
Reconciling pharmacogenomic knowledge units

Example:

From PharmGKB:



In the literature:



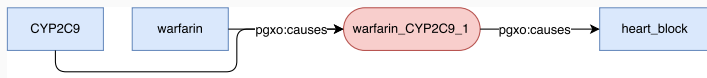
Reconciling pharmacogenomic knowledge units

Example:

From PharmGKB:



In the literature:



$warfarin \sqsubseteq warfarin \sqcap$

$CYP2C9 \sqsubseteq CYP2C9 \sqcap$

$heart_block \sqsubseteq cardiovascular_diseases$

(MeSH)

$\Rightarrow \{CYP2C9_warfarin_1\} \sqsubseteq \{CYP2C9_warfarin\}$ (rule 2)

Conclusion & Perspectives

Conclusion & Perspectives

- Main features:
 - Representing PGx relationships
 - Associating metadata (provenance, confidence...)
 - Reconciling various sources thanks to inference rules
- Ontology available on GitHub² and the NCBO Bioportal³
- Also available on GitHub:
 - Mappings to MeSH, NCit, SNOMED CT, Genomics CDS, SO-Pharm, PHARE and PO
 - Instantiated examples (figures and *.owl files)

²<https://github.com/practikpharma/PGx0>

³<https://bioportal.bioontology.org/ontologies/PGX0>

Work in progress:

- Populating a LOD data set with PGx relationships
 - (i) From PharmGKB
 - (ii) Extracted from abstracts of PubMed
 - (iii) Mined from Electronic Health Records (of Paris and Stanford Hospitals)
- Integrating inference rules in the ontology with SWRL

Thank you for your attention
Questions?

pierre.monnin@loria.fr

<http://practikpharma.loria.fr/>