Leveraging Post-marketing Drug Safety Research through Semantic Technologies: The PharmacoVigilance Signal Detectors Ontology

Vassilis Koutkias and Marie-Christine Jaulent
INSERM, U1142, LIMICS - University Paris VI/XIII
CRC, 15 rue de l' École de Médecine, 75006 Paris, FRANCE
Email: vasileios.koutkias@inserm.fr
Web: http://safer-project.eu/

International SWAT4LS Workshop
Semantic web applications and tools for life sciences

Background: How is Drug Safety Being Assessed?

- **Pre-market** setting: clinical trials of drugs
  - Limitations: Time constraints, population size and potential bias

not sufficient to detect
all possible safety issues

- **Post-market** surveillance: continuous monitoring throughout the time a drug is actively prescribed

many heterogeneous data sources to investigate
(e.g. spontaneous reports, observational healthcare data, literature, social media, etc.)

Definitions (*in the scope of this work*)

- **Signal** (in Pharmacovigilance):
  - Any information (*reported or computationally extracted*) on an **unknown / incompletely documented <u>possible</u> causal relationship between a drug and an adverse effect**
- **Signal Detection Method**:
  - **Computational method** aiming to **identify pharmacovigilance signals**
  - Designed **to operate on a specific type of data** (i.e. contingency tables of drug-condition pairs)
  - Offers a **set of parameters / analysis options**, e.g. thresholds for decision making, ranking criteria, etc.
  - Results in a ranked **list of drug-condition pairs** along with features denoting their statistical significance
- **Signal Detector**:
  - A **software implementation** of a signal detection method executed based on a **specific parameterization**

Example: The GPS[1] Implementation (PhViD[2])
Input Data & Parameters

Usage

```
GPS(DATABASE,  RR0 = 1, MIN.n11 = 1, DECISION = 1, DECISION.THRES = 0.05,
RANKSTAT = 1, TRONC = FALSE, TRONC.THRES = 1,
PRIOR.INIT = c(alpha1 = 0.2, beta1 = 0.06, alpha2 = 1.4,
beta2 = 1.8, w = 0.1), PRIOR.PARAM = NULL)
```

Arguments

DATABASE	Object returned by the function as.PhViD.
RR0	Value of the tested risk. By default, RR0=1.
MIN.n11	Minimum number of notifications for a couple to be potentially considered as a signal. This option does not affect the calculation of the hyper parameters. By default, MIN.n11 = 1.

• • •

[1] ***Gamma Poisson Shrinker, used by FDA to screen reports gathered in FAERS.*** Details: W. DuMouchel, "Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system", The American Statistician, 53, 1999, 190–6.
[2] I. Ahmed and A. Poncet, "PhViD: an R package for PharmacoVigilance signal Detection", http://cran.r-project.org/web/packages/PhViD/

Example: GPS Implementation (PhViD)
Output Options

Value

ALLSIGNALS	Data.frame summarizing the results of all couples with at least MIN.n11 notifications ordered by RANKSTAT. It contains notably the labels, the cell counts, the expected counts, RANKSTAT, the ratios(count/expected count), the marginal counts and the estimations of FDR, FNR, Se et Sp. If RANKSTAT!=1, the last column is the posterior probability of the null hypothesis.
SIGNALS	Same Data.frame as ALLSIGNALS but restricted to the list of generated signals.
NB.SIGNALS	Number of generated signals.
INPUT.PARAM	Parameters entered in the function.
PARAM	A list that contains the prior hyper parameters (PRIOR.PARAM). Additionally if PRIOR.PARAM=NULL, it also contains the prior hyper parameters initialization (PRIOR.INIT) and the convergence code (see nlm()).

Our Ultimate Goal (SAFER project)

- **Develop a platform capable of exploiting and complementing evidence** obtained from **diverse, existing signal detection methods and data sources**
- Why:
  - **Uncertainty** in the results obtained from all types of computational signal detection methods
  - Accurate indications & timely decisions required

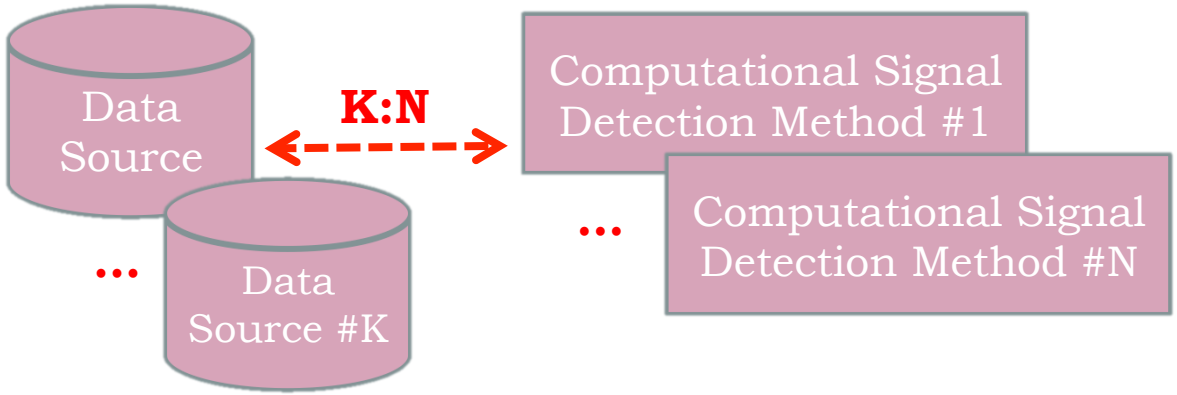
Position statement:

- Each method and data source may contribute at a different level in **complementing our knowledge on drug safety risks**

→ **An integrative perspective may add value!**

Some Challenges Introduced by the Integrated Perspective

- **Practical Perspective**:
  - Are **sufficient and appropriate methods and data** available to implement the integrated approach?
- **Technical Perspective**:
  - How can this integration be **systematic** and **large-scale**?
  - How to obtain a **combined, prioritized** list of the **findings**?
- **End-user Perspective**:
  - How can I use all these methods without having to be an **expert in statistics**?
  - How can I **setup an analysis experiment** based on my interest?
  - How can I make my **results understandable and easy to share** with other researchers?



V. Koutkias & M.-C. Jaulent, Leveraging Post-marketing Drug Safety Research through Semantic Technologies: The PharmacoVigilance Signal Detectors Ontology, SWAT4LS Workshop, Berlin, Dec. 10, 2014.

Practical Perspective: Availability of Data Sources for Signal Detection

FAERS Spontaneous Reports (Programmatic Access)

openFDA

Europe PubMed Central

Bibliographic Databases (Programmatic Access)

OMOP Simulated Data

Observational Medical Outcomes Partnership

Social Media Streams (Programmatic Access)

Trend toward the provision of programmatic access to data sources

V. Koutkias & M.-C. Jaulent, Leveraging Post-marketing Drug Safety Research through Semantic Technologies: The PharmacoVigilance Signal Detectors Ontology, SWAT4LS Workshop, Berlin, Dec. 10, 2014.

Practical Perspective: Availability of Open-Source Signal Detection Methods - *List of OMOP Implementations*[1]

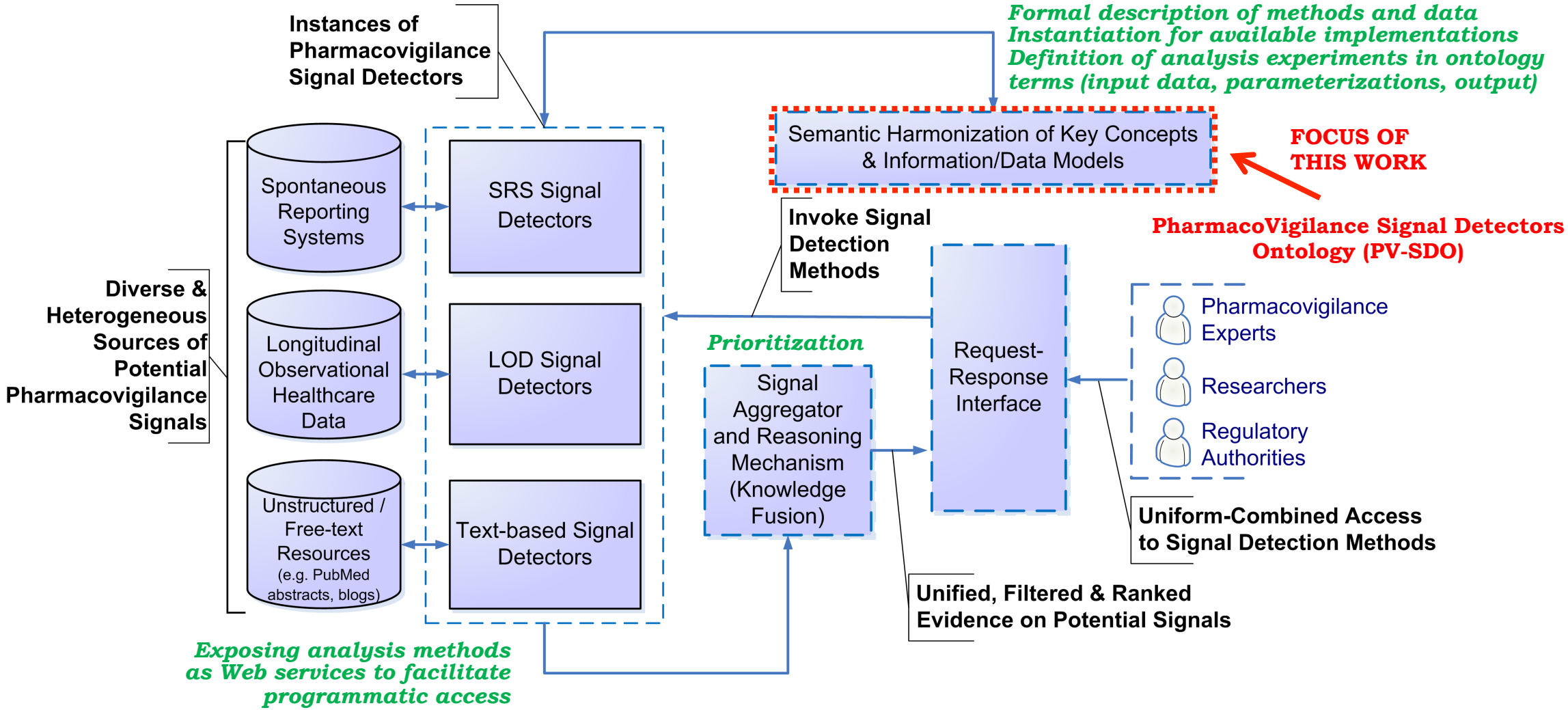
- Disproportionality Analysis (DP) - OMOP Research Team
- Univariate Self-Controlled Case Series (USCCS) - OMOP Research Team
- Observational Screening (OS) - ProSanos Corporation
- Multi-Set Case Control Estimation (MSCCE) - OMOP Research Team
- Bayesian Logistic Regression (BLR) - OMOP Research Team
- Case Control Surveillance (CCS) - Lilly
- IC Temporal Pattern Discovery (ICTPD) - Uppsala Monitoring Centre
- Case-Crossover (CCO) - University of Utah
- HSIU Population-Based Method - Indiana University
- Maximized Sequential Probability Ratio Test (MSPRT) - Harvard Pilgrim
- Conditional Sequential Sampling Procedure (CSSP) - Harvard Pilgrim
- High-Dimensional Propensity Score (HDPS) - OMOP Research Team
- Incident User Design (IUD-HOI) - M. Alan Brookhart

Various!

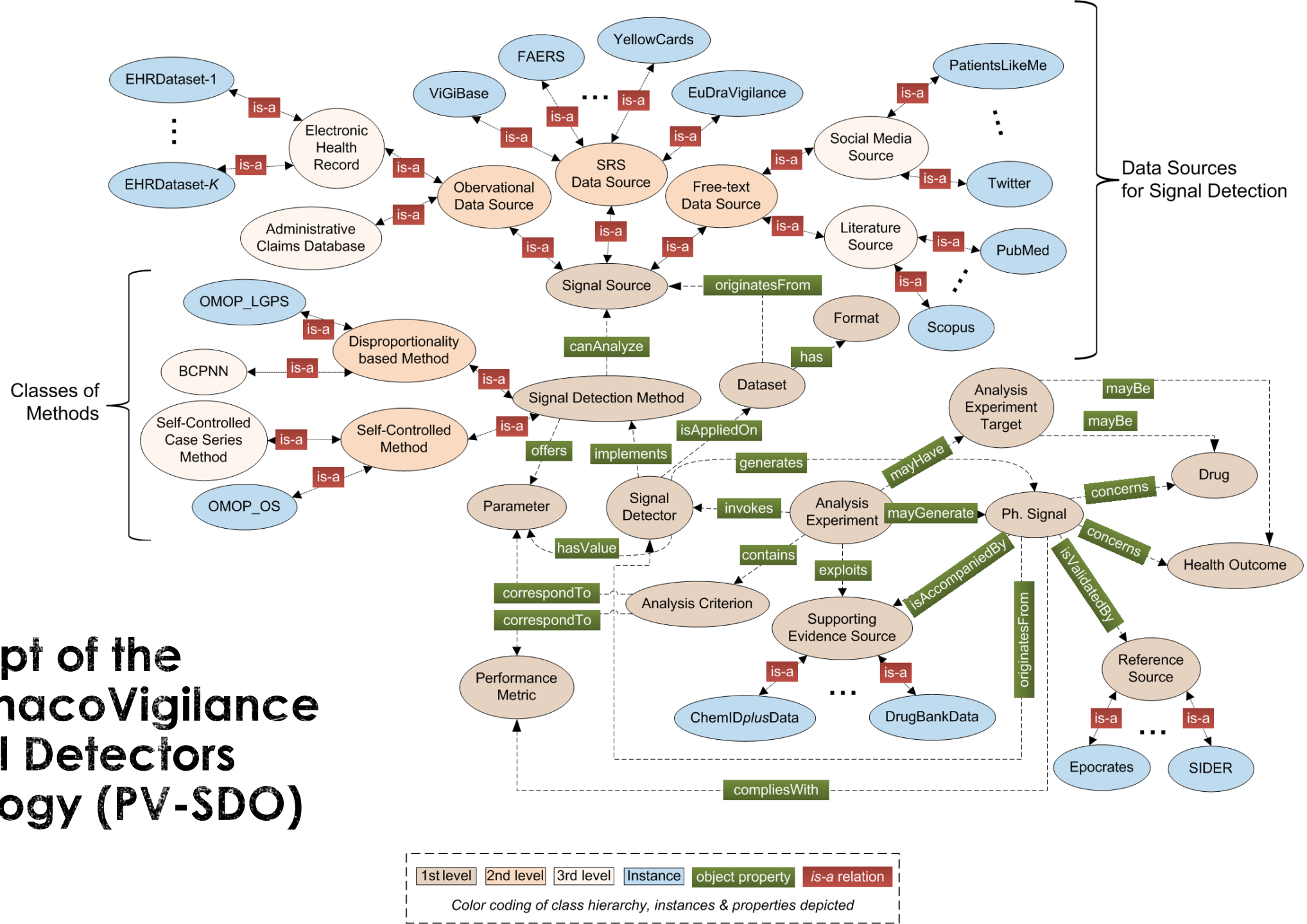
[1]**http://omop.org/MethodsLibrary**

V. Koutkias & M.-C. Jaulent, Leveraging Post-marketing Drug Safety Research through Semantic Technologies: The PharmacoVigilance Signal Detectors Ontology, SWAT4LS Workshop, Berlin, Dec. 10, 2014.

Technical Perspective: Systematic, Large-scale Integration and Prioritization of Outcomes (the SAFER framework)



Excerpt of the PharmacoVigilance Signal Detectors Ontology (PV-SDO)



1st level | 2nd level | 3rd level | Instance | object property | *is-a* relation

Color coding of class hierarchy, instances & properties depicted

End-user Perspective: Hiding Complexity

- **Semantic harmonization of domain concepts**:
  - classification and annotation of data sources for signal detection (either for discovery or filtering)
  - classification of signal detection methods
  - classification of analysis parameters, performance metrics and ranking criteria that can be set/used by signal detection methods
- Facilitate the **definition of drugs and health outcomes of interest** through semantic mappings

Semantic Harmonization
Classification of Signal Detection Method Implementations (excerpt)

- 'Signal Detection Method'
  - 'Bayesian Confidence Propagation Neural Network'
    - 'Bayesian Confidence Propagation Neural Network Extended To The Multiple Comparison Framework'
    - 'Original Bayesian Confidence Propagation Neural Network Original'
  - 'Case Control Signal Detection Method'
  - 'Cohort Signal Detection Method'
  - 'Gamma Poisson Shrinkage based Signal Detection Method'
    - 'Gamma Poisson Shrinkage Extended To The Multiple Comparison Framework'
    - 'Longitudinal Gamma Poisson Shrinker'
    - 'Multi-item Gamma Poisson Shrinker'
    - 'Original Gamma Poisson Shrinkage'
  - 'Proportional Reporting Ratio based Method'
    - 'Original Proportional Reporting Ratio Signal Detection Method'
    - 'Proportional Reporting Ratio Method Extended To The Multiple Comparison Framework'
  - 'Reporting Fisher's Exact Test Signal Detection Method'
  - 'Reporting Odds Ratio based Method'
    - 'Original Reporting Odds Ratio Method'
    - 'Reporting Odds Ratio Method Extended To The Multiple Comparison Framework'
  - 'Self-Controlled Signal Detection Method'
    - 'Self-Controlled Case Series Signal Detection Method'
    - 'Self-Controlled Cohort Design Signal Detection Method'

Indentation denotes class-subclass relation

Semantic Harmonization
Classification of Analysis Parameters (excerpt)

- Parameter
  - 'Apply Leopard Filtering For Protopathic Bias'
  - 'Case-Control Matching Strategy'
  - 'Comparator Population'
  - 'Control Related Parameter'
    - 'Control Period'
    - 'Controls PerCase'
    - 'Include Index Date In Control Period'
    - 'Use Control Period In Expected Calculation'
  - 'Covariate Related Parameter'
    - 'Additional Covariates Included In The Propensity Score Model'
    - 'Covariate Eligibility Window'
    - 'Covariate Selection Algorithm Additional Parameters'
    - 'Dimensions To Include As Potential Covariates'
  - 'Decision Related Parameter'
    - 'Decision Rule For The Signal Generation'
    - 'Threshold for Decision'
  - 'Exposures To Include'
  - 'Include All Drugs In A Multivariate Analysis'
  - 'Include Index Date In Time At Risk'
  - 'Minimum Number Of Notifications For A Couple To Be Potentially Co
  - 'Nesting Within Population With The Indication Of The Target Drug'
  - 'Outcomes To Include'
  - 'Prior Related Parameter'
    - 'Period Prior To Exposure In Expected Calculation'
      - 'Use One Day Prior To Exposure In Expected Calculation'
      - 'Use One Month Prior To Exposure In Expected Calculation'
    - 'Prior Distribution'
    - 'Prior Parameters Chosen'
    - 'Prior Parameters Initialization Vector'
    - 'Required Observation Time Prior To Exposure'
    - 'Required Observation Time Prior To Outcome'
    - 'Variance Of The Prior'

End-user Perspective: Facilitate Study Setup
Example: Mapping Terms across Diverse Data Sources

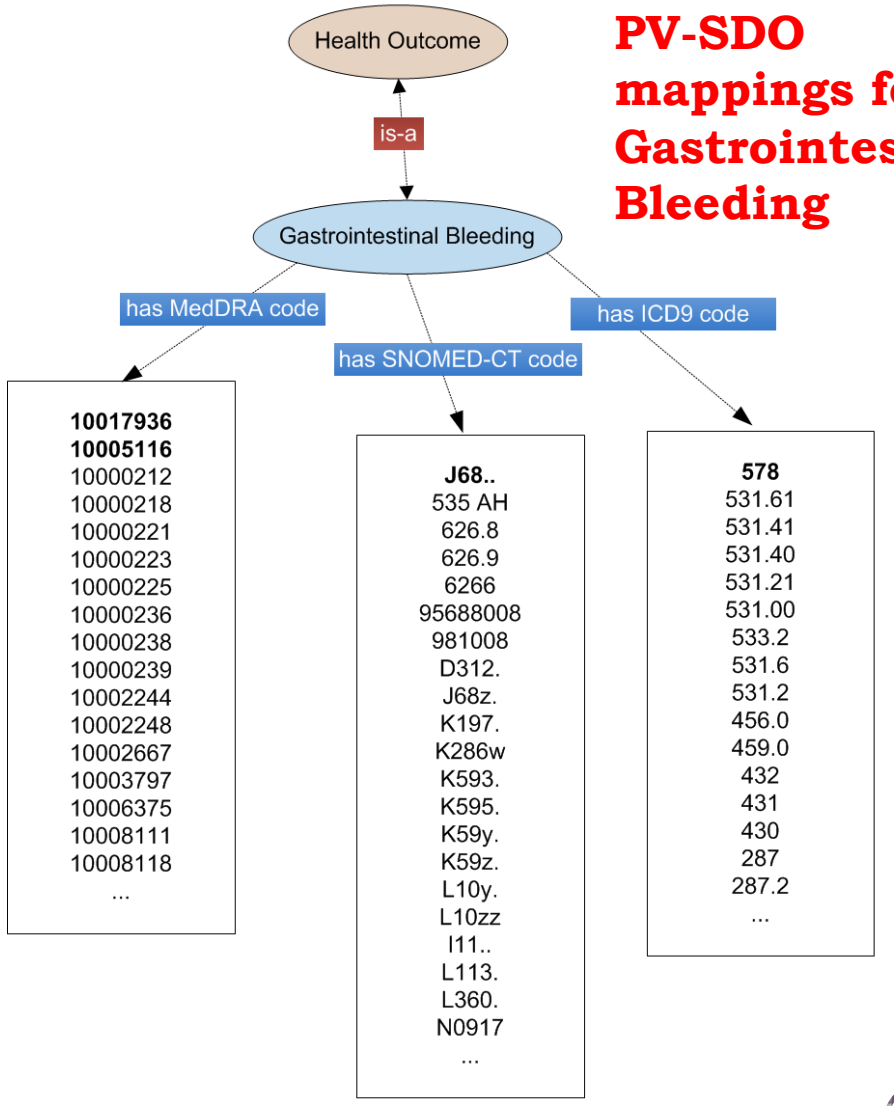
Illustration of the problem

Table 1 AERS and EHR—event and outcome definitions

Event	AERS MedDRA PT event definitions	EHR Outcome definitions
Pancreatitis	Pancreatitis acute	Lab tests: amylase >300 U/l or lipase >120 U/l
Rhabdomyolysis	Rhabdomyolysis, blood creatine phosphokinase MM increased	Lab test: 5× normal levels of creatine kinase
QT	Long QT syndrome, ECG QT prolonged, Torsade de pointes, ECG QT interval abnormal, ventricular tachycardia	UMLS codes: C0023976, C0151878, C0743431, C0855333, C1560305

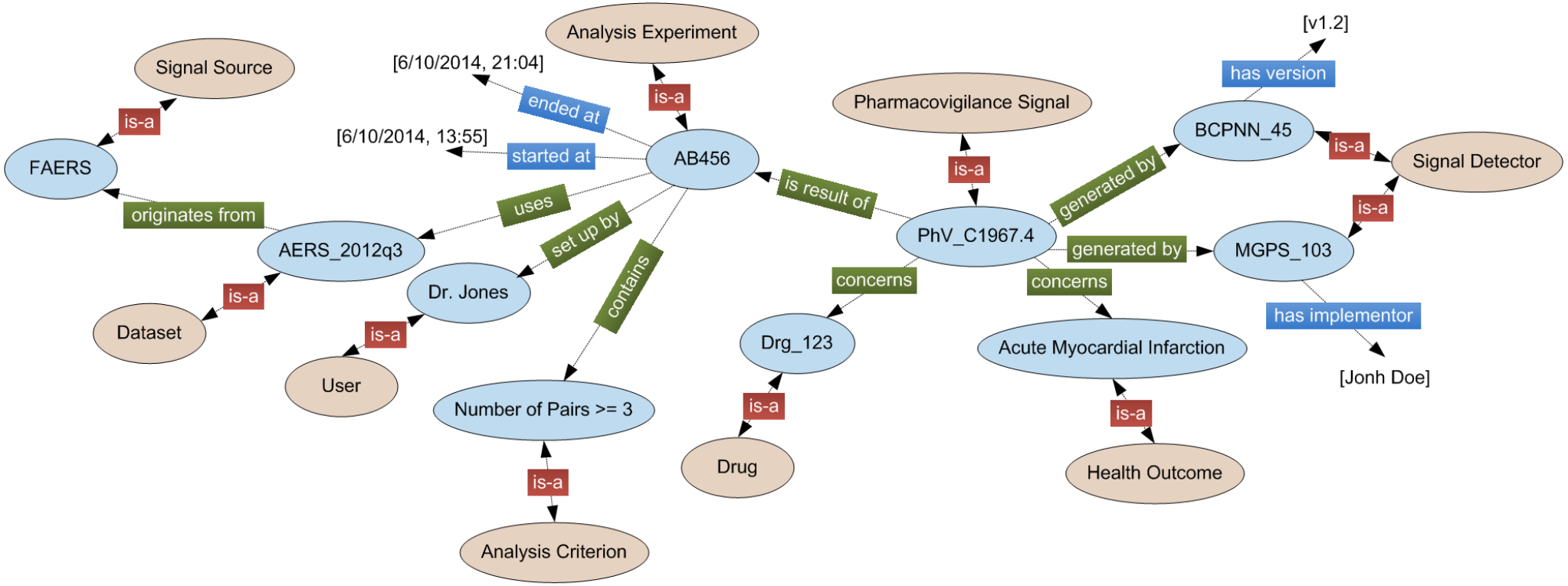
AERS, adverse event reporting system; EHR, electronic health record; PT, preferred term.

Table source: R. Harpaz et al., Combing signals from spontaneous reports and electronic health records for detection of adverse drug reactions, J Am Med Inform Assoc. 2013;20(3):413-9.



PV-SDO mappings for Gastrointestinal Bleeding

End-user Perspective: PV-SDO Annotation of Experiments – Results Sharing

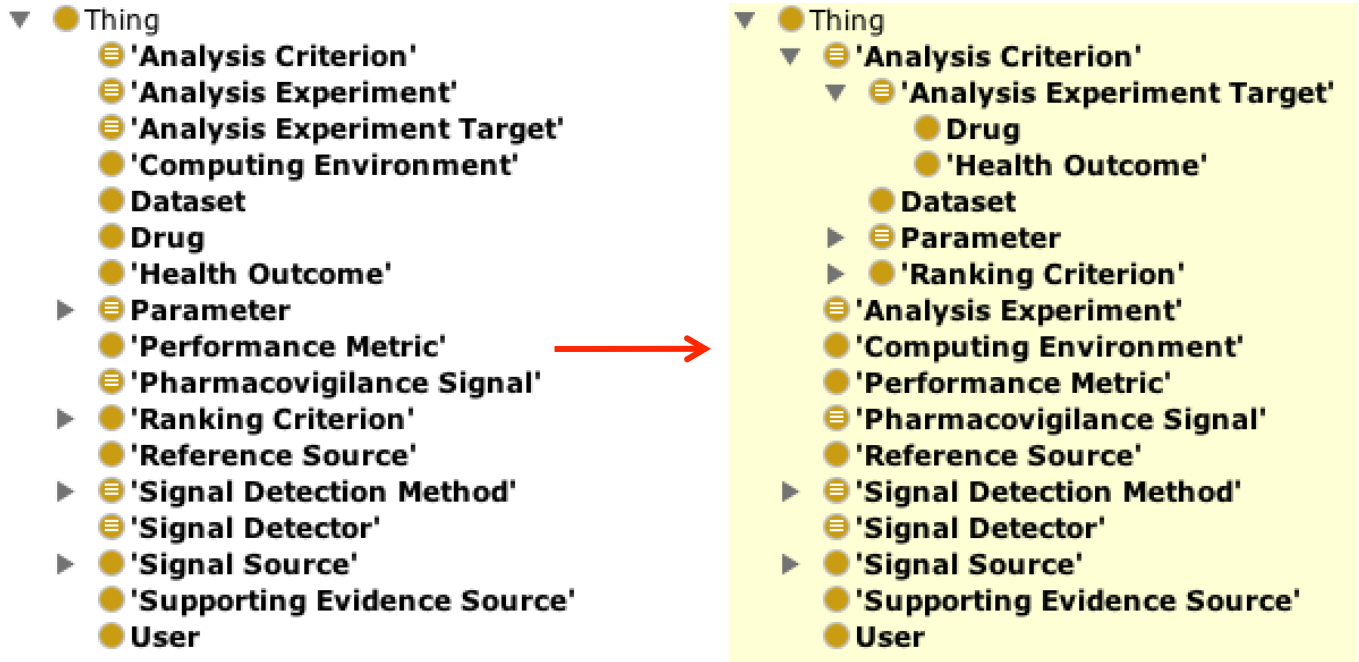


***Example: Provenance of analysis outcomes* (fictive data)**

Results: PV-SDO Evaluation
Data-driven & Automatically Using a Reasoner

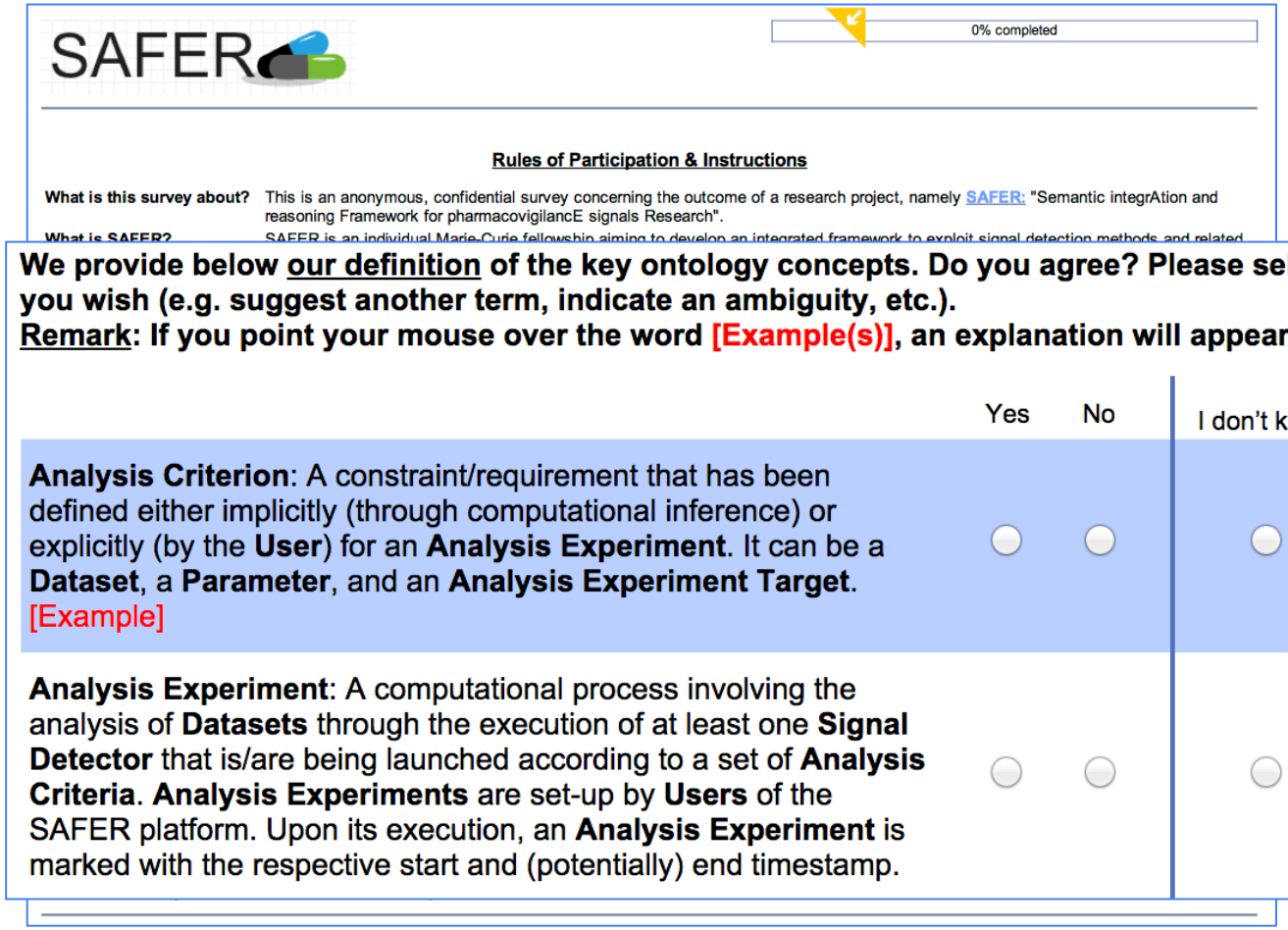
Is the ontology sufficient to describe methods which were not part of the source knowledge employed in the design?

Test with new published methods



Results: PV-SDO Evaluation
By Knowledge Engineers & Signal Detection Experts

Online survey

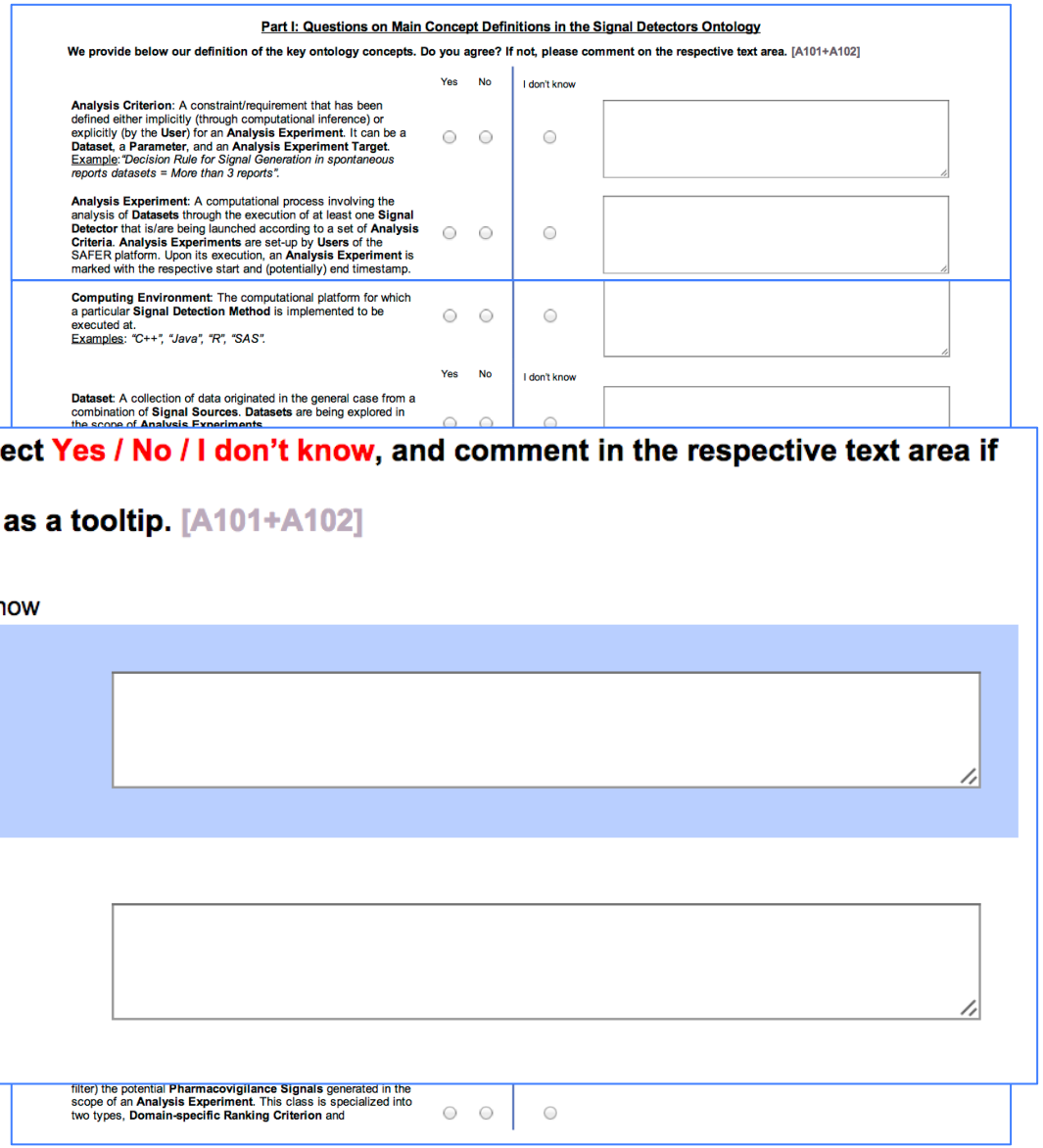

SAFER

0% completed

Rules of Participation & Instructions

What is this survey about? This is an anonymous, confidential survey concerning the outcome of a research project, namely SAFER: "Semantic integrAtion and reasoning Framework for pharmacovigilancE signals Research".

What is SAFER? SAFER is an individual Marie-Curie fellowship aiming to develop an integrated framework to exploit signal detection methods and related


Part I: Questions on Main Concept Definitions in the Signal Detectors Ontology

We provide below our definition of the key ontology concepts. Do you agree? If not, please comment on the respective text area. [A101+A102]

Yes No I don't know

Analysis Criterion: A constraint/requirement that has been defined either implicitly (through computational inference) or explicitly (by the **User**) for an **Analysis Experiment**. It can be a **Dataset**, a **Parameter**, and an **Analysis Experiment Target**. Example: *"Decision Rule for Signal Generation in spontaneous reports datasets = More than 3 reports"*.

Analysis Experiment: A computational process involving the analysis of **Datasets** through the execution of at least one **Signal Detector** that is/are being launched according to a set of **Analysis Criteria**. **Analysis Experiments** are set-up by **Users** of the SAFER platform. Upon its execution, an **Analysis Experiment** is marked with the respective start and (potentially) end timestamp.

Computing Environment: The computational platform for which a particular **Signal Detection Method** is implemented to be executed at. Examples: *"C++", "Java", "R", "SAS"*.

Yes No I don't know

Dataset: A collection of data originated in the general case from a combination of **Signal Sources**. **Datasets** are being explored in the scope of **Analysis Experiments**

filter) the potential **Pharmacovigilance Signals** generated in the scope of an **Analysis Experiment**. This class is specialized into two types, **Domain-specific Ranking Criterion** and

We provide below our definition of the key ontology concepts. Do you agree? Please select Yes / No / I don't know, and comment in the respective text area if you wish (e.g. suggest another term, indicate an ambiguity, etc.).
Remark: If you point your mouse over the word [Example(s)], an explanation will appear as a tooltip. [A101+A102]

	Yes	No	I don't know	
Analysis Criterion: A constraint/requirement that has been defined either implicitly (through computational inference) or explicitly (by the **User**) for an **Analysis Experiment**. It can be a **Dataset**, a **Parameter**, and an **Analysis Experiment Target**. [Example]	○	○	○	
Analysis Experiment: A computational process involving the analysis of **Datasets** through the execution of at least one **Signal Detector** that is/are being launched according to a set of **Analysis Criteria**. **Analysis Experiments** are set-up by **Users** of the SAFER platform. Upon its execution, an **Analysis Experiment** is marked with the respective start and (potentially) end timestamp.	○	○	○	

Current version of PV-SDO: 1,312 axioms, 101 classes, 34 object properties, 32 data properties and 168 individuals

Discussion: Foreseen Value of Combining Data Sources & Detection Methods

- Support the **detection of true associations – reduce false positive findings** through replicated signaling;
- Support **timely decisions**: the "many sources – many methods" setting may provide faster indications;
- Address the **event-based** and **data-source based differential performance** of methods;
- Address the fact that no individual approach to detect signals is adequate and **the concurrent use of multiple methods is essential,**

by hiding the complexity for end-users (i.e. drug safety experts and regulatory authorities)

Next Steps

- **PV-SDO**:
  - Complete **revisions** based on signal detection experts' feedback
  - Investigate **extension** with domain knowledge, e.g. for method selection
  - Provide a **comprehensive resource** for end-users
- **PV-SDO availability:** http://bioportal.bioontology.org/ontologies/PV-SDO, currently private access – publicly available soon!
- Focus on favorable **test cases** (adverse effects and/or drugs) and **scenarios** (e.g. targeted vs. strong signal identification)
- A **proof-of-concept** end-to-end integration and **evaluation**

Discussion

Acknowledgement

This research was supported by a Marie Curie Intra European Fellowship within the 7th European Community Framework Programme FP7/2007-2013 under REA grant agreement n° 330422 – the SAFER project

Thank you!

Leveraging Post-marketing Drug Safety Research through Semantic Technologies: The PharmacoVigilance Signal Detectors Ontology

Vassilis Koutkias and Marie-Christine Jaulent
INSERM, U1142, LIMICS - University Paris VI/XIII, Paris, FRANCE
Email: vasileios.koutkias@inserm.fr, Web: http://www.safer-project.eu/

SWAT4LS International SWAT4LS Workshop
Semantic web applications and tools for life sciences

MARIE CURIE ACTIONS