

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

Vassilis Koutkias¹ and Marie-Christine Jaulent¹

¹ INSERM, U1142, LIMICS, F-75006, Paris, France;
Sorbonne Universités, UPMC Univ Paris 06, UMR_S 1142, LIMICS, F-75006, Paris, France;
Université Paris 13, Sorbonne Paris Cité, LIMICS, (UMR_S 1142), F-93430, Villetaneuse,
France
{vasileios.koutkias, marie-christine.jaulent}@inserm.fr

Abstract. Accurate and timely identification of post-marketing drug safety risks (the so-called “signals” in pharmacovigilance) is an important public health issue. While various computational methods have been proposed to analyze the diverse data sources employed for signal detection, still the challenge of effective drug monitoring and surveillance remains. On the other hand, there is an emerging belief that the synthesis of all possible information sources is necessary to achieve further advancements. Aiming to facilitate integrated signal detection by concurrently exploring various data sources via respective analysis methods in a systematic way, we propose the Pharmacovigilance Signal Detectors Ontology (PV-SDO). PV-SDO constitutes the backbone of a semantically-enriched platform for this integration and aims to: (a) semantically harmonize heterogeneous data sources and analysis methods in the field, (b) facilitate their joint exploitation through mappings between reference terminologies that the data sources rely on, and (c) provide an exploitable knowledge base of signal analysis methods, experiments and their outcomes including provenance information. PV-SDO has been populated with a significant number of individuals using data from open-source signal detection method implementations, and assessed via data-driven and logic-based techniques, while an evaluation with experts is currently being conducted with well-promising results.

Keywords: Pharmacovigilance, Heterogeneous Data Sources, Computational Signal Detection, Semantic Integration, Ontologies.

1 Introduction

Identification of post-marketing drug safety risks is an important public health issue requiring accurate and timely evidence [1]. Pharmacovigilance is the science that studies all the activities related with the detection, assessment, understanding and prevention of adverse effects or any other drug-related problem [2]. The detection of the so-called “signals”, which are typically defined as “information on a possible causal relationship between an adverse event and a drug, the relationship being

unknown or incompletely documented previously [1, 3]”, relies on the analysis of data originated from different sources [4]. These sources vary, spanning from spontaneous reporting systems (SRS) and the literature, to social media platforms. Nevertheless, while various computational methods have been proposed to analyze the diverse data sources employed for signal detection, optimal drug monitoring and surveillance remains an open issue [5]. On the other hand, there is an emerging belief that the synthesis of all possible information sources is necessary to achieve further advancements [6, 7].

Aiming to implement signal detection by concurrently exploring various data sources and the respective computational analysis methods in a systematic way, we present in this paper the Pharmacovigilance Signal Detectors Ontology (PV-SDO). First, we provide background information as regards post-marketing signal detection. Next, we present PV-SDO in terms of its scope, the employed material for its construction, its overall schema, main entities and attributes, referring also to development and evaluation aspects. We finally discuss the contribution that PV-SDO aspires to provide towards more automated, systematic tools for signal detection.

2 Post-marketing Signal Detection

2.1 Rationale

Systematic drug surveillance requires the ability to test all drugs for any outcome. However, the drugs for which surveillance is most crucial are those newly introduced in the market, having a limited safety profile from clinical studies. When a novel, credible signal is detected, it triggers an evaluation procedure performed by regulatory authorities that involves a detailed review of the indicated association. This may determine that the causal relationship is sufficiently supported to warrant some action (e.g. labelling amendment), that the relationship is non-causal, or that it is unclear (thus, further monitoring and/or analysis studies are required). As the volume of raw data for signal detection is enormous and various data sources are being considered for analysis, it is imperative to support signal detection via appropriate analysis methods.

Next, we provide background information on the typical data sources that are employed for signal detection, and on the respective methods that are available for their analysis.

2.2 Data Sources: Merits and Limitations

The dominant source of post-marketing drug safety signals is Spontaneous Reporting Systems (SRS) through which individual case safety reports are submitted by healthcare professionals and patients to pharmaceutical companies, regulatory authorities and other bodies. Well-known SRSs are VigiBase™ from the Uppsala Monitoring Centre of WHO that gathers reports from more than 100 countries [8], and FAERS (the Food and Drug Administration’s Adverse Event Reporting System)

in the US [9]. Despite their explicit focus on drug safety, underreporting, inconsistent reporting, bias and latency are important limitations of SRSs [10].

The biomedical literature has been often employed for signal detection, but also as a mean to support the evaluation of potential signals obtained by other sources [11]. Although significant, this source is limited by the study context (scope, papers selection protocol, etc.), reducing the generalization perspective of the findings.

Observational healthcare data obtained from Electronic Health Record systems and administrative claim databases, have been employed quite recently (as secondary use / repurposing) for signal detection [12]. Such data contain more information than spontaneous reports, e.g. drug exposure/non-exposure time and laboratory examination results, and they bring the potential to enable active and real-time surveillance. However, among the drawbacks for their adoption is the replication of findings due to parameters like the sample size and the considered population.

Patient self-reports about drug concerns/problems that are shared among networked communities using social media (blogs, messaging platforms, etc.) are an emerging source of signals [13], which can provide important insights in some cases, e.g. for drugs used for the treatment of rare diseases. Inherent limitations of patient self-reports for signal detection are the subjective information provided by the patients and the lack of quality control in the reporting.

2.2 Signal Detection Methods: Towards Combinatorial Analysis

For each one of the above data sources, various analysis methods have been proposed. For example, a typical strategy to explore SRS-based data is to pursue statistical investigations for measures of disproportionality, i.e. assess whether a specific event is reported more frequently with a specific drug than would be expected if randomly occurring [14]. Examples of computational methods proposed for the analysis of observational data are case-control series, cohort methods and self-controlled studies, to name a few [15]. For the analysis of biomedical literature and patient self-reports in social media, natural language processing (NLP) techniques have been employed, driven by standard terminologies and controlled vocabularies [16].

Given the inherent limitations of the data sources and the technical limitations of signal detection methods, comparative studies illustrated that there is no optimal method for signal detection and that all methods provide a high rate of false-positive indications [17, 18]. On the other hand, it is reasonable to argue that combining information across data sources can lead to more effective and accurate signal detection. This can be achieved by either increasing the evidence and the statistical power of findings, or by facilitating discoveries that could not be possible with the analysis of a single source [19].

However, the heterogeneity and fragmentation characterizing the available data sources and signal detection methods avert the realization of such a synthesis at large-scale. To reach this level, a semantic harmonization layer is required. Thus, we propose PV-SDO as the mean for semantic description of signal sources, detection methods and relevant analysis experiments.

3 The PV-SDO Ontology

3.1 Scope and Source Knowledge

The primary aim of PV-SDO is to support the development of an integrated platform for signal detection. The platform aspires to enable the exploration of data originated from different sources with diverse computational signal detection methods in a combinatorial fashion [20]. In this regard, the sub-objectives of PV-SDO are manifold, as it shall:

- (a) classify and annotate through quality attributes (e.g. the coverage, the system organ classes included, the population period, etc.) the data sources that can be used in the detection of potential signals, either for discovery or filtering;
- (b) classify currently available signal detection methods, along with the analysis parameters that they offer;
- (c) embody performance metrics and ranking criteria/measures that can be employed for combinatorial signal detection;
- (d) support the conduction of analysis experiments involving multiple sources and detection methods, through mappings among the input data and the output of signal detection methods, e.g. drugs in RxNorm¹ vs. ATC² (Anatomical Therapeutic Chemical classification system), health outcomes of interest in ICD³ (International Classification of Diseases) vs. MedDRA⁴ (Medical Dictionary for Regulatory Activities);
- (e) support the annotation of signal detection analysis experiments with provenance information (i.e. from which dataset, method(s) and parameterization(s) were the outcomes generated) and, overall,
- (f) provide an exploitable knowledge base of available data sources, signal detection methods, analysis experiments and outcomes.

The primary knowledge sources that were employed for defining PV-SDO were:

- (a) the scientific literature with major emphasis on papers describing signal detection methods and the results obtained from their use in various analysis experiments, and
- (b) documentation of signal detection method implementations, such as those available in the PhViD R package [21], and the OMOP⁵ (Observational Medical Outcomes Partnership) methods library [22].

It is important to note that classification schemas related with the field of signal detection in pharmacovigilance are quite missing. In the scope of the literature review that we conducted, we were only able to identify the so-called “taxonomy for monitoring methods within a medical product safety surveillance system” that has been proposed in the Mini-Sentinel project [23]. This taxonomy, however, considers only methods appropriate for the analysis of observational data.

¹ <http://www.nlm.nih.gov/research/umls/rxnorm/>

² http://www.whooc.no/atc_ddd_index/

³ <http://www.who.int/classifications/icd/en/>

⁴ <http://www.meddra.org/>

⁵ <http://omop.org/>

3.2 Overall Schema, Key Entities and Properties

In order to fulfill its aims, PV-SDO includes concepts that are related with: (a) the domain of signal detection (e.g. Drug, Health Outcome, Pharmacovigilance Signal, etc.), and (b) the construction of the integrated platform for signal detection (e.g. Signal Source, Signal Detector, Analysis Experiment, Analysis Experiment Target, etc.). Figure 1 illustrates an overview of PV-SDO. The schema depicts the 1st-level entities of the hierarchy, some example 2nd- and 3rd-level entities, the main linkage among entities via object properties, as well as example individuals that have been defined. In more detail, Table 1 provides (in alphabetical order) definitions of the key PV-SDO concepts in natural language, including examples. Likewise, Table 2 provides a list of key object properties.

Considering a platform in which Datasets originated from diverse types of Signal Sources and respective Signal Detection Methods will be registered as accessible resources for signal detection, PV-SDO may support a matchmaking mechanism (through ontology querying) between the user's requirements for a particular analysis and the resources that are appropriate from those available. These requirements are either implicit or explicit. For example, selecting the Parameter "Time at Risk" for an Analysis Experiment, would implicitly result in invoking only Signal Detectors that support this Parameter, or selecting disproportionality-based methods would launch all the available Signal Detectors taking advantage of the class-subclass hierarchy. Of course, the user can explicitly define resources to be used for their analysis experiments.

PV-SDO can express provenance information as regards a potential signal like in the following example: Pharmacovigilance Signal⁶: "*Drug:Drg_134 – Health Outcome:Acute Myocardial Infarction*" is result of Analysis Experiment:AB456⁷, generated by Signal Detector:BCPNN⁸_45 and Signal Detector:MGPS⁹_103, validated by Reference Source:Epocrates¹⁰ and accompanied by Supporting Evidence Source:ChemIDPlus¹¹.

3.3 Design and Development Aspects

Taking into account the multi-facet scope of PV-SDO, we carefully analyzed the requirements that it has to fulfill with respect to the application logic that the integrated platform shall support [20]. For the design of PV-SDO we followed an incremental strategy and the post-coordination principle [24]. We also applied common practices in ontology modeling that span from formatting and naming conventions, to version control. In addition, we investigated the potential for ontology reuse that was nevertheless restricted to the adoption of some property definitions and

⁶ The data cited in this example are fictive.

⁷ Codes like these are expected to be attributed by the proposed signal detection platform.

⁸ Denotes the Bayesian Confidence Neural Network, a well-known signal detection method.

⁹ Denotes the Multi-Item Gamma Poisson Shrinkage, a well-known signal detection method.

¹⁰ <http://www.epocrates.com/>

¹¹ <http://chem.sis.nlm.nih.gov/chemidplus/>

modeling approaches, e.g. the approach of representing provenance information with the PROV ontology (PROV-O) [25].

A first version of the ontology has been outlined based on which brainstorming sessions took place with ontology engineers from our organization. As a result, where necessary, ambiguities in concept definitions have been addressed in order to clarify the conceptualization of the model. In this regard, PV-SDO evolved and instantiated with sample data. In the current stage, PV-SDO comprises of 1,312 axioms, 101 classes, 34 object properties, 32 data properties and 168 individuals.

PV-SDO has been encoded in OWL2 (Web Ontology Language 2) using Protégé [26], in the entire lifecycle of ontology modeling, population and assessment.

Table 1. List of key entities defined in PV-SDO (entities marked in bold).

Entity Name	Description
Analysis Criterion	A constraint/requirement that has been defined either implicitly (through computational inference) or explicitly (by the user of the platform) for an Analysis Experiment . It can be a Dataset , a Parameter , and an Analysis Experiment Target . <i>Example:</i> “ <i>Decision Rule for Signal Generation in SRS: > 3 reports</i> ”.
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 by users of the platform.
Analysis Experiment Target	A specific Drug (or Drug Class), Health Outcome , “ Drug – Health Outcome ” pair or “ Drug - Drug - Health Outcome ” triplet that is being investigated in the scope of an Analysis Experiment for Pharmacovigilance Signal detection. <i>Example:</i> “ <i>Acute Myocardial Infarction - DrugX</i> ”.
Dataset	A collection of data originated in the general case from a combination of Signal Sources . <i>Example:</i> “ <i>FAERS_ASCII_2013q3.zip and PubMed_Extract_2012-14.txt</i> ”.
Drug	A pharmaceutical product (or class of products). It may be referred using standard terminologies/classifications, like ATC and RxNorm. <i>Example:</i> “ <i>Heparin group / ATC: B01AB01</i> ”.
Health Outcome	An observation in the health condition of a subject. It may be expressed using standard terminologies such as UMLS (Unified Medical Language System), MedDRA (preferred terms), and LOINC (Logical Observation Identifiers Names and Codes) value ranges. <i>Example:</i> “ <i>Gastrointestinal bleeding / MedDRA: 10017936, 10005116</i> ”.
Parameter	A configurable option that is available for the instantiation and execution of a Signal Detection Method . <i>Examples:</i> “ <i>Time at Risk</i> ”, “ <i>Decision rule for signal generation</i> ”.
Performance Metric	Measure for assessing the effectiveness of Signal Detectors . <i>Examples:</i> “ <i>Area Under the Curve</i> ”, “ <i>F-score</i> ”.
Pharmacovigilance Signal	A “ Drug - Health Outcome ” pair or a “ Drug - Drug - Health Outcome ” triplet indicating a <u>possible</u> causal relation among drug(s) and outcome, generated within an Analysis Experiment by one or more Signal Detectors .
Ranking Criterion	A feature that is used to sort (and implicitly filter) potential Pharmacovigilance Signals generated in the scope of an Analysis Experiment . It is specialized into two types, Domain-specific Ranking Criterion and Computation Ranking Criterion .

Entity Name	Description
	<u>Examples</u> : Computational Ranking Criterion: “ <i>Precision at K</i> ”; Domain-specific Ranking Criterion: “ <i>Adverse Drug Reaction seriousness</i> ”.
Reference Source	An information repository that is used for validating the novelty of a potential Pharmacovigilance Signal .
Signal Detection Method	An implementation of a computational method aiming to identify potential Pharmacovigilance Signals via its application to data originated from an appropriate type of Signal Source . A Signal Detection Method offers a set of Parameters for its fine-tuning that can be set. <u>Example</u> : “ <i>The BCPNN (Bayesian Confidence Propagation Neural Network) implementation contained in [21]</i> ”.
Signal Detector	The concrete runtime instantiation of a Signal Detection Method , according to specific Parameter values and input Dataset(s) . <u>Example</u> : “ <i>The MGPS (Multi-Item Gamma Poisson Shrinkage) implementation contained in [21] with its default parameter values applied on the French, national SRS database</i> ”.
Signal Source	A data repository that can be explored for Pharmacovigilance Signal detection. It has specific characteristics with respect to the type of data that it offers (structured/unstructured), the quality of its data, the type of Signal Detection Methods that can be used for its analysis, etc. <u>Examples</u> : “ <i>Spontaneous Reporting System</i> ”, “ <i>Electronic Health Record system</i> ”, “ <i>Bibliographic database</i> ”, “ <i>Social media platform</i> ”.
Supporting Evidence Source	Information source that can either facilitate the understanding, or complement the indications of Pharmacovigilance Signals generated by an Analysis Experiment . <u>Examples</u> : “ <i>Drugbank</i> ¹² ”, “ <i>ChemIDPlus</i> ”.

4 PV-SDO Evaluation

Evaluation is an integral part of ontology design and development. While various approaches and methods have been proposed in the literature, there is no single best or preferred evaluation methodology/approach in general. The selection relies primarily on the purpose/focus of the evaluation per se and the application scope of the ontology. Brank et al. proposed four types of ontology evaluation approaches, namely, “gold standard based”, “application-based”, “data-driven” and “assessment by humans” [27].

For the evaluation of PV-SDO the “gold standard based” approach cannot be employed, since no standard exists that covers its scope and purpose. The “application-based” approach could be used at a later stage, once the envisaged integrated platform for signal detection is fully operational. Thus, in the current stage we employed: (a) the “data-driven” approach and (b) “assessment by humans”. Details about these methods are provided in the following. In addition, we employed the reasoning mechanisms embodied in Protégé to assess PV-SDO for its logical consistency and the correctness of its taxonomy, besides for obtaining inferred types.

¹² <http://www.drugbank.ca/>

Table 2. List of indicative properties defined in PV-SDO (properties marked in bold).

Object Properties (label, domain and range)
A Signal Detection Method can analyze multiple Signal Sources
A Dataset originates from Signal Source(s)
A Pharmacovigilance Signal concerns a “Drug - Health Outcome” pair or a “Drug - Drug - Health Outcome” triplet
A Signal Detection Method is implemented for a Computing Environment
A Pharmacovigilance Signal is generated by Signal Detector(s)
A Signal Detector has parameter value Parameter
A Signal Detector instantiates a Signal Detection Method
A Signal Detector is used in Analysis Experiment(s)
An Analysis Experiment may have Analysis Experiment Target
An Analysis Experiment uses for ranking Ranking Criterion
A Pharmacovigilance Signal is validated by Reference Source(s)
Datatype Properties (label, domain and range)
(Clinical Narrative or Literature Source or Observational Data Source or SRS Data Source) encodes conditions in {"ICD9", ...}
Observational Data Source has population size integer
Drug has ATC code / has OMOP code / has RxNORM code / has SNOMED-CT¹³ code / has UMLS code string
(Dataset or Signal Detection Method or Signal Source) has license {"Apache License Version 2.0", "GPL-2", ...}
(Dataset) has format {"JSON (JavaScript Object Notation)", "OMOP CDM (Common Data Model) v.4.0", "OMOP CDM v3.0", ...}
Analysis Experiment has result XMLLiteral
Analysis Experiment was launched/terminated on dateTimeStamp

4.1 Data-driven Assessment

In the “data-driven” assessment we investigated whether PV-SDO is sufficient to describe signal detection methods, which were not part of the source knowledge that we employed in our design. In particular, we explored the potential of populating PV-SDO with methods that are not included in the PhViD R package [21], or in the methods library and the analysis results obtained in the 2011-2012 Experiments of the OMOP project [22].

For example, we elaborated on instantiating PV-SDO for a recently published signal detection method, namely, *vigiRank* [28]. In the *vigiRank* algorithm, a signal detection measure originated from disproportionality analysis is being used as input parameter; a case that had not been taken into account in PV-SDO beforehand. Thus, appropriate revisions have been applied in PV-SDO. Overall, while PV-SDO is still evolving, the data-driven assessment demonstrated that there is a solid basis for the description of various types of signal detection methods.

¹³ Systematized Nomenclature of Medicine--Clinical Terms: <http://ihtsdo.org/snomed-ct/>

4.2 Human Assessment

This involves the conduction of an international, anonymous survey by inviting experts to provide their feedback on PV-SDO through an online questionnaire. The invitees have been organized in two groups:

(a) Experts in the field of knowledge engineering and ontologies, expected to provide feedback on the survey setup and its comprehension, besides the pure ontology modeling aspects.

(b) Experts in the field of signal detection with know-how in the design and development of computational signal detection methods. These experts have been identified from their participation in relevant scientific publications.

Ten knowledge engineers and 15 signal detection experts with different levels of awareness concerning the scope of our work have been invited. For the moment, the survey with the knowledge engineers has been concluded (response rate: 80%), while the survey with signal detection experts is still in progress. The questionnaire was structured in sections containing:

- Part I with questions concerning the main concepts definition in PV-SDO;
- Part II with questions on object property definitions;
- Part III with questions on data property definitions;
- Part IV referring to the major parts of the class hierarchy that was defined in PV-SDO, and
- Part V referring to an overall assessment of PV-SDO with respect to its usefulness and fulfillment of its aims.

The most important remarks of the survey with the knowledge engineers group are summarized below:

- Part I: The overall agreement with the definitions of concepts was 72%. In the rest 28% of the cases, the experts hesitated to express an opinion for some definitions (selected the “I don’t know” answer), as they did not know the exact details underlying signal detection. Regarding the definition of Pharmacovigilance Signal one expert commented that a signal could be rather a mathematical variable (or set of variables) the value of which (under the circumstances considered) suggest(s) a possible causal relation between drug(s) and health outcome.

- Part II: The overall agreement with the provided definitions of object properties was 78%. Again, in the rest of the cases, the experts hesitated to express an opinion. For example, concerning the relation “A Signal Detection Method *can analyze* multiple Signal Sources”, there was a comment that this depends on what the term “multiple Signal Sources” exactly means. Also, some experts faced difficulty in discriminating the concepts Dataset and Signal Source.

- Part III: The overall agreement with the provided definitions of data properties was 96%, and no significant objections were expressed.

- Part IV: The overall agreement concerning the defined hierarchy was 67%. In the remaining 33% the experts expressed their hesitation due to their lack of knowledge in the domain of signal detection. Among the important remarks, one expert proposed to distinguish between private SRS signal sources (e.g. the SRSs maintained by pharmaceutical companies) and public ones, which is indeed applicable.

- Part V: Experts ranked with 4.5/5 the statement “PV-SDO provides the basis to semantically describe signal detection methods and analysis experiments”, while they

were quite neutral in their answers regarding the comprehension of PV-SDO. They were clearly positive (4/5) in the statement “The concepts/properties of PV-SDO cover the required scope”.

4.3 Revisions

Based on the evaluation outcomes up to now, the main corrective actions that have been applied in PV-SDO involved:

- 1) Addressing ambiguous domain and range definitions, as indicated by the knowledge engineers assessment group.
- 2) Revising misleading names of class labels and property names.
- 3) Assessing all the annotations that were included in PV-SDO for clarity and comprehension.

For the assessment phase with the participation of signal detection experts, we adopted a shorter, more targeted version of the questionnaire. We particularly focused on receiving feedback as regards the organization of the Parameter class and the Signal Detection Methods classification, which appeared to be the most specialized aspects for the knowledge engineers to assess. This group corresponds to the ultimate user target for our evaluation, since they represent providers of the signal detection methods that our platform aims to integrate, as well as potential end-users of the platform.

5 Discussion and Conclusions

Accurate and timely signal detection poses important research challenges. As new data sources are being considered as useful in providing evidence for pharmacovigilance, such as observational databases and even social media platforms, the field is lately very active [4]. Despite the well-promising results, it has become evident that each data source exhibits specific potential and limitations and, similarly, each computational detection method has strengths and weaknesses [5, 17, 18]. Thus, combined signal detection strategies are emerging [6, 7].

In order to leverage such strategies at large-scale, the harmonization among the underlying information models corresponding to the respective data sources and signal detection methods shall be performed. Information Technologies (IT) play an important role towards this aim by developing and applying appropriate data and knowledge engineering methods. Aligned with combined signal detection strategies, we propose PV-SDO as the backbone for the development of such a semantically-enriched, integrated IT platform. PV-SDO is the first attempt for constructing a formal model to describe data sources and signal detection methods in terms of their types, underlying computational model, input, output, analysis parameters, requirements of use, strengths and weaknesses.

To the best of our knowledge, the only systematic effort to categorize signal detection methods is the taxonomy of observational analysis methods elaborated in the scope of Mini-Sentinel [23]. Interestingly, the Mini-Sentinel taxonomy has been employed to design a tool (available in a spreadsheet form) for guiding researchers in

the appropriate selection of signal detection methods. Notably, it remains questionable at which level this selection process can be generic, since most of the findings supporting this approach originate from empirical studies. Nevertheless, we aim to explore this perspective and embody such a feature for the users of our platform, in order to support them in the selection of the appropriate signal detection method(s), given their analysis scope, e.g. drug of interest, health outcome of interest, available data for analysis etc. This feature will be implemented via semantic rules that will be introduced in PV-SDO.

PV-SDO will be enriched during our platform implementation phase, taking into account the forthcoming evaluation outcomes. It has been uploaded in BioPortal¹⁴ [29], currently in private access. Our plan is to make it publicly available, right after the conclusion of its evaluation and the revisions, in order to pursue synergies with other teams developing ontologies in the domain of pharmacovigilance, but also to enable its maintenance and evolution as a reference resource for describing signal detection.

The availability of drug-related datasets through semantic Web standard formats (e.g. via Bio2RDF [30] and the EBI RDF platform [31]) and open drug safety data (like FAERS data published through the openFDA initiative [32]), illustrates that there is a strong potential for application of semantic technologies in the field of drug safety. Given the fact that signal detection is a very active research field, PV-SDO constitutes a novel contribution in the domain of computational signal detection, aspiring to reinforce existing efforts in a systematic way.

Acknowledgments. 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. We would like to thank the knowledge engineers and signal detection experts who participated in the evaluation of PV-SDO.

References

1. Institute of Medicine: Preventing Medication Errors. The National Academic Press, Washington DC (2007)
2. World Health Organization: A practical handbook on the pharmacovigilance of antimalarial medicines. Geneva, Switzerland (2008)
3. Linquist, M.: The Need for Definitions in Pharmacovigilance. *Drug Saf.* 30(10) (2007) 825–830
4. Harpaz, R., DuMouchel, W., Shah, N.H., Madigan, D., Ryan, P., Friedman, C.: Novel data mining methodologies for adverse drug event discovery and analysis. *Clin. Pharmacol. Ther.* 91(6) (2012) 1010–1021
5. Hauben, M., Norén, G.N.: A decade of data mining and still counting. *Drug Saf.* 33 (2010), 527–534
6. Harpaz, R., et al.: Combing signals from spontaneous reports and electronic health records for detection of adverse drug reactions. *JAMIA* 20 (2013) 413–419

¹⁴ <http://bioportal.bioontology.org/ontologies/PV-SDO>

7. Liu, M., et al.: Large-scale prediction of adverse drug reactions using chemical, biological, and phenotypic properties of drugs. *JAMIA* 19 (2012) e28–e35
8. The VigiBase™ Web site: <http://www.unc-products.com/>, last access: September 30, 2014.
9. The FDA Adverse Event Reporting System (FAERS) Web site: <http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/default.htmAdverseDrugEffects/>, last access: September 30, 2014.
10. Goldman, S.A.: Limitations and strengths of spontaneous reports data. *Clin Ther.* 20(Suppl C) (1998) C40–C44
11. Shang, N., Xu, H., Rindfleisch, T.C., Cohen, T.: Identifying plausible adverse drug reactions using knowledge extracted from the literature. *J. Biomed. Inform.* DOI: 10.1016/j.jbi.2014.07.011 [ahead of print]
12. Studying the Science of Observational Research: Empirical findings from the Observational Medical Outcomes Partnership. *Drug Saf.* 36(Suppl 1) (2013)
13. Freifeld, C.C., et al.: Digital drug safety surveillance: monitoring pharmaceutical products in Twitter. *Drug Saf.* 37, 343–350 (2014)
14. Hauben, M., Bate, A.: Decision support methods for the detection of adverse events in post-marketing data. *Drug Discov. Today* 14 (2009) 343–357
15. Suling, M., Pigeot, I.: Signal detection and monitoring based on longitudinal healthcare data. *Pharmaceutics* 4 (2012) 607–640
16. Harpaz, R. et al.: Text mining for adverse drug events: the promise, challenges, and state of the art. *Drug Saf.* 37 (2014) 777–790
17. van Holle, L., Bauchau, V.: Signal detection on spontaneous reports of adverse events following immunisation: a comparison of the performance of a disproportionality-based algorithm and a time-to-onset-based algorithm. *Pharmacoepidemiol. Drug Saf.* 23 (2014) 178–185
18. Ryan, P.B., Madigan, D., Stang, P.E., Overhage, J.M., Racoosin, J.A., Hartzema, A.G.: Empirical assessment of methods for risk identification in healthcare data: results from the experiments of the observational medical outcomes partnership. *Stat. Med.* 31 (2012), 4401–4415
19. Harpaz, R. et al.: Combing signals from spontaneous reports and electronic health records for detection of adverse drug reactions. *JAMIA* 20 (2013) 413–419
20. Koutkias, V., Jaulent, M.-C.: An agent-based approach for integrated pharmacovigilance signal detection. In *Proc. of the Multi-Agent Systems for Healthcare (MASH) Workshop, 13th Int. Conf. on Autonomous Agents & Multiagent Systems (AAMAS)*, Paris, France, May 6, 2014.
21. Ahmed, I., Poncet, A.: PhViD: an R package for Pharmacovigilance signal Detection. R package version 1.0.6. (2013). Available at: <http://cran.r-project.org/web/packages/PhViD/>, last access: September 30, 2014.
22. The Observational Medical Outcomes Partnership (OMOP) Methods Library: <http://omop.org/MethodsLibrary>, last access: September 30, 2014.
23. Gagne, J.J., et al.: Taxonomy for Monitoring Methods within a Medical Product Safety Surveillance System: Year Two Report of the Mini-Sentinel Taxonomy Project Workgroup. Available at: http://www.mini-sentinel.org/work_products/Statistical_Methods/Mini-Sentinel_Methods_Taxonomy-Year-2-Report.pdf, last access: September 30, 2014.
24. Rector, A., Iannone, L.: Lexically suggest, logically define: quality assurance of the use of qualifiers and expected results of post-coordination in SNOMED CT. *J. Biomed. Inform.* 45 (2012) 199–209
25. W3C Working Group Note 30 April 2013: PROV Overview: An Overview of the PROV Family of Documents”. Groth, P., Moreau, L. (Eds.). Available at: <http://www.w3.org/TR/prov-overview/>, last access: September 30, 2014
26. The Protégé knowledge modeling tool: Available at: <http://protege.stanford.edu/>, last access: September 30, 2014

27. Brank, J., Grobelnik, M., Mladenić, D.: A survey of ontology evaluation techniques. In: Proc. of Conf. on Data Mining and Data Warehouses (SiKDD), 2005, Ljubljana, Slovenia.
28. Caster, O., Juhlin, K., Watson, S., Norén, G.N.: Improved statistical signal detection in pharmacovigilance by combining multiple strength-of-evidence aspects in vigiRank. *Drug Saf.* 37 (2014) 617–628
29. Whetzel, P.L., et al.: BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. *Nucleic Acids Res.* 39(Web Server issue) (2011) W541–W545
30. The Bio2RDF Web site: <http://bio2rdf.org/>, last access: September 30, 2014
31. Jupp, S., et al.: The EBI RDF platform: linked open data for the life sciences. *Bioinformatics* 30 (2014) 1338–1339.
32. The openFDA Web site: <https://open.fda.gov/>, last access: September 30, 2014