Towards an Ontology for Sharing Information on Pharmacovigilance Signals

Pantelis Natsiavas[1,2], Magnus Wallberg[3] and Vassilis Koutkias[1,2]

[1]Lab of Computing, Medical Informatics & Biomedical Imaging Technologies, Department of Medicine, Aristotle University of Thessaloniki, Thessaloniki, Greece
[2]Institute of Applied Biosciences, Centre for Research & Technology Hellas, Thermi, Thessaloniki, Greece
[3]Uppsala Monitoring Centre, Uppsala, Sweden

Abstract. We present the rationale and the key design decisions for an ontology currently being built to support linking, sharing and further processing of pharmacovigilance signals. The main goal of the presented ontology is to provide a conceptual model describing such information towards FAIR data principles using the Linked Data paradigm.

Keywords: Pharmacovigilance signals, knowledge engineering, FAIR data.

1 Introduction

Signal detection and management is an important part of pharmacovigilance (PV). A PV signal is defined as "*information that arises from one or multiple sources…, which suggests a new potentially causal association, or a new aspect of a known association, between an intervention and an event or set of related events...*" [1]. Typically, signal information is disseminated via newsletters released by regulatory authorities, e.g. PV centres and the World Health Organization, in free-text format. We currently develop an OWL ontology focusing on the publication of PV signal information in an extendable and reusable manner based on widely-accepted semantic models. The main goals of building such an ontology are: (a) automatic interlinking of data presented in different signal reports, to identify relations between data currently presented in disparate unstructured datasets (i.e. free-text reports), (b) provenance tracking of PV signal data to support the verification of the source, and (c) semantically disambiguating information through reference to widely-accepted thesauri and semantic models. The ultimate goal is to support the semantic enrichment of PV signal information following the FAIR data principles [2], in order to enhance the currently applied PV signal investigation practices, through the reuse and better exploitation of the respective data.

2 Proposed approach

The key design decisions of the ontology focus on the reuse of existing models to semantically enhance the original PV signal information and facilitate its verification, reproducibility and further processing (e.g. through semantic reasoning):

- The micropublications semantic model [3] is used to annotate evidence-based PV signal information, facilitating conclusion reproduction, verification etc.
- PROV-O [4] is used to model the provenance information of PV signal data items, (e.g. patient reports from Spontaneous Reporting Systems), as this can be critical to verify the quality of data for decision making (e.g. duplicate prevention).
- The published PV signal reports can be annotated using the Web Annotation Data Model (WADM) [5], directly relating the information depicted via Linked Data formalisms with the original unstructured information in free-text format.
- Temporal aspects of PV signal data are expressed using the Time Ontology [6]. As such data typically refer to trends and evolution through time, their temporal semantic annotation is important, in order to improve the capabilities of identifying time-related conclusions (e.g. a dependency with a drug release or an epidemic).

The ontology is built according to the NeoN methodology [7] and defines concepts typically shared when presenting PV signal information in free-text format, e.g. spontaneous reports, indications of drug usage, adverse effects, rechallenge or final outcome, disproportionality analysis data etc. These domain specific concepts are semantically related with the concepts of the above-mentioned models using either hierarchy relations (e.g. sub-classing) or via the use of OWL properties. Future work concerns the validation of the ontology with real-world PV reports and its use as information representation formalism in unstructured data sources analysis for public health scenarios as part of the platform presented in [8].

References

1. Council for International Organizations of Medical Sciences (CIOMS): Practical Aspects of Signal Detection in Pharmacovigilance. CIOMS WG VIII report, Geneva (2010).
2. Wilkinson, M.D., et al.: The FAIR Guiding Principles for scientific data management and stewardship. Sci. data. 3, 160018 (2016).
3. Clark, T., et al.: Micropublications: a semantic model for claims, evidence, arguments and annotations in biomedical communications. J. Biomed. Semantics. 5, 28 (2014).
4. Belhajjame, et al.: PROV Model Primer, https://w3.org/TR/prov-primer/.
5. Sanderson, R., et al.: Web Annotation Data Model, http://w3.org/TR/annotation-model/.
6. Cox, S., et al.: Time Ontology in OWL, https://www.w3.org/TR/owl-time/.
7. Suárez-Figueroa, M.C., et al.: The NeOn Methodology for Ontology Engineering. In: Ontology Engineering in a Networked World. pp.9–34, Springer, Berlin, Heidelberg (2012).
8. Natsiavas, P., et al.: A Public Health Surveillance Platform Exploiting Free-Text Sources via Natural Language Processing and Linked Data: Application in Adverse Drug Reaction Signal Detection Using PubMed and Twitter. In: Riaño D. et al. (eds) Knowledge Representation for Health Care. KR4HC 2016, ProHealth 2016. Lecture Notes in Computer Science, vol 10096. Springer, Cham.