Exploring the Capacity of Open, Linked Data Sources to Assess Adverse Drug Reaction Signals

Pantelis Natsiavas[1,2], Vassilis Koutkias[2] and Nicos Maglaveras[1,2]

[1]Lab of Medical Informatics, Medical School, Aristotle University, Thessaloniki, Greece
[2]Institute of Applied Biosciences, Centre for Research & Technology Hellas, Thermi, Thessaloniki, Greece

Abstract. In this work, we explore the capacity of open, linked data sources to assess adverse drug reaction (ADR) signals. Our study is based on a set of drug-related Bio2RDF data sources and three reference datasets, containing both positive and negative ADR signals, which were used for benchmarking. We present the overall approach for this assessment and refer to some early findings based on the analysis performed so far.

Keywords: Adverse Drug Reactions, signal evaluation, linked data, Bio2RDF.

1 Introduction

The task of assessing potential adverse drug reaction (ADR) signals is mostly performed manually by drug safety experts. It includes the review/analysis of scientific literature, clinical trial data, biological properties of drugs, etc., in order to assess causality. Linked data enable the systematic combination of various heterogeneous data sources, and their use for drug safety has been proposed in some studies [1], [2]. In this work, we explore the capacity of relevant Bio2RDF datasets to assess ADR signals, using an evaluation algorithm, and discuss early findings.

2 Material and Methods

Bio2RDF provides access to various life-science data sources, following the conversion of their raw data into RDF [3]. In this study, we employed a set of drug-related data sources, namely, ClinicalTrials.gov, DrugBank, LinkedSPL and SIDER, in order to evaluate their capacity in assessing candidate ADR signals. The evaluation algorithm depicted in Fig. 1 is executed independently for each candidate signal (including relevant drug and condition synonyms) across all data sources. First, we analyzed the respective Bio2RDF data sources, in order to identify RDF properties that semantically imply drug use indications and ADRs. We then constructed the SPARQL queries corresponding to each step of the evaluation algorithm, which are executed through the available Bio2RDF SPARQL endpoints. The assessment relies on three reference datasets [4]-[6], containing both positive and negative signals.

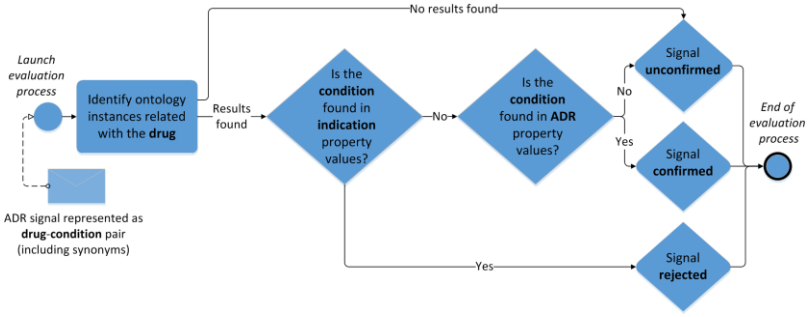

Fig. 1. Evaluation algorithm of candidate ADR signals.

3 Discussion

Our first results indicate that the employed data sources can be used for signal evaluation either independently or in conjunction. The capacity per data source varies, depending on its purpose and the characteristics of its raw data. For example, since SIDER contains explicit information on ADRs, it enables ADR signal confirmation with high sensitivity. On the contrary, DrugBank which provides more general drug information including indications for drug use, is efficient in identifying false ADR signals with high specificity. The biggest challenge faced up to now concerns the load capacity of the public Bio2RDF SPARQL endpoints. We currently setup endpoints similar to Bio2RDF in a cloud infrastructure to address performance issues.

Acknowledgment. This research has been partially funded from the European Union's Horizon 2020 Framework Programme for Research and Innovation Action under Grant Agreement no. 643491 (the PATHway project).

References

1. Koutkias, V.G., Jaulent, M.-C.: Computational approaches for pharmacovigilance signal detection: toward integrated and semantically-enriched frameworks. Drug Saf. 38, 219–232 (2015).
2. Banda, J.M. et al: Provenance-centered dataset of drug-drug interactions. ArXiv150705408 Cs. (2015).
3. Callahan, A. et al: Bio2RDF Release 2: improved coverage, interoperability and provenance of life science linked data. In: Cimiano, P. et al: (eds.) The Semantic Web: Semantics and Big Data. pp. 200–212. Springer Berlin Heidelberg (2013).
4. Coloma, P.M. et al: a reference standard for evaluation of methods for drug safety signal detection using electronic healthcare record databases. Drug Saf. 36, 13–23 (2012).
5. Ryan, P.B. et al: Defining a Reference Set to Support Methodological Research in Drug Safety. Drug Saf. 36, 33–47 (2013).
6. Harpaz, R. et al: A time-indexed reference standard of adverse drug reactions. Sci. Data. 1, 140043, (2014).