Linked Registries: connecting rare diseases patient registries through a semantic web layer

Pedro Sernadela and José Luís Oliveira (sernadela, jlo)@ua.pt
University of Aveiro, DETI/IEETA, 3810-193 Aveiro, Portugal

Introduction

Patient registries are an essential tool to increase current knowledge regarding rare diseases. Understanding these data is a vital step to improve patient treatments, and to create the most adequate tools for personalized medicine. However, the growing number of disease-specific patient registries brings also new technical challenges. Usually, these systems are developed as closed data silos, with independent formats and models, lacking comprehensive mechanisms to enable data sharing.

To tackle these challenges, we developed a semantic web-based solution that allows connecting distributed and heterogeneous registries, enabling the federation of knowledge between multiple independent environments. This semantic layer creates a holistic view over a set of anonymized registries, supporting semantic data representation, integrated access and querying.

Methods

In this work, we have developed a semantic layer on top of existing patient registries, to allow extracting anonymised data from the original datasets, translate them to a common shared exchange model and make them available to the research community. This approach takes advantage of semantic web technologies to improve how we publish, access, express and share knowledge across the Web. From a technological perspective, the system was built on top of COEUS [1], an application framework that streamlines data integration with semantic representation. As patient registries are shared within this platform, researchers and developers are able to perform federated queries, covering miscellaneous databases, just like they would query a single local dataset (see Fig. 1).

Results

We used the proposed architecture [2–3] for the integration of four patient registries in the neuromuscular and neurodegenerative disease area. These registries collected patient data from ten different countries, and gather information related to four rare diseases: Myotonic Dystrophies, Facioscapulohumeral muscular dystrophy, Fukutin Related Protein related conditions and Huntington’s Disease. The implemented system gave us the opportunity to answer challenging questions across disparate rare disease patient registries.

Our results are significant in at least three major respects:

1) The use of a model agnostic system, which enables the mapping of patient registries’ data from any format to a common shared ontology (see Fig. 2).

2) The creation of an independent system that can be plugged into any existing patient registry without changing it. This enables the extraction of relevant data elements while maintaining patients’ data privacy and security.

3) The adoption of Semantic Web technologies to promote a better translation, interpretation, federation and discovery of new knowledge acquired from linked patient registries datasets.

Conclusions

This strategy empowers a holistic view through connected registries, enabling state-of-the-art semantic data sharing and access (see Fig. 3). The outcome is a unique semantic layer, connecting miscellaneous registries and delivering a lightweight holistic perspective over the wealth of knowledge stemming from linked rare disease patient registries.

References

