

Integrating Knowledge Graphs for Explainable Artificial Intelligence in Biomedicine*

Marta Contreiras Silva, Daniel Faria, and Catia Pesquita

LASIGE, Dep. de Informática, Faculdade de Ciências da Universidade de Lisboa,
Portugal

The rich panorama of publicly available data and ontologies in the biomedical domain represents an unique opportunity for developing explainable knowledge-enabled systems in biomedical Artificial Intelligence (AI) [1, 3, 4]. Building on decades of work by the semantic web and biomedical ontologies communities, a semi-automated approach for building and maintaining a Knowledge Graph (KG) to support AI-based personalized medicine is within our grasp. However, personalized medicine also poses significant challenges that require advances to the state of the art, such as the diversity and complexity of the domain and underlying data, coupled with the requirements for explainability.

We propose an approach (see Figure 1) to build a KG for personalized medicine to serve as a rich input for the AI system (*ante-hoc*) and incorporate its outcomes to support explanations, by connecting input and output (*post-hoc*).

A preparatory step is **Data and ontology collection and curation**. This includes the selection and curation of relevant public datasets for the domain in question, the identification of ontologies referenced by the datasets, and the selection of other relevant ontologies to ensure adequate coverage of the domain and sufficient semantic richness to support explanations. Additionally, data privacy inherent to patient data should inform the decision to make part of the KG private to its data providers and the data integration process also mostly automatic to reduce the need of human involvement [2].

The first step in our approach is **Ontology Matching**. Key challenges are scalability and complex matching, since building a comprehensive KG requires matching multiple ontologies with hundreds of thousands of concepts, covering different domains, and with different modeling perspectives. Regarding scalability, our solution is to match ontologies iteratively, by matching and merging the largest pair of ontologies into a single one, then mapping and merging this to the third largest ontology, and so on, using complex matching algorithms to uncover rich relations across domains [6, 7]. Before the final integration of ontologies, alignments are partially validated by experts to ensure an accurate KG that can support explanations [5].

The **Semantic Data Annotation** process relies on the development of parsers to interpret each type of dataset, and annotation algorithms to produce an RDF version of the dataset that is semantically integrated into the KG.

Finally, the **Integration with the AI system** ensures that the KG serves as both input to AI methods (directly or through feature generation [8]) and

* Copyright © 2021 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

also encodes the AI outcomes, which supports a shared semantic space for data, scientific context, and predictions capable of supporting KG-based explanations methods, including querying, reasoning and similarity searches [9].

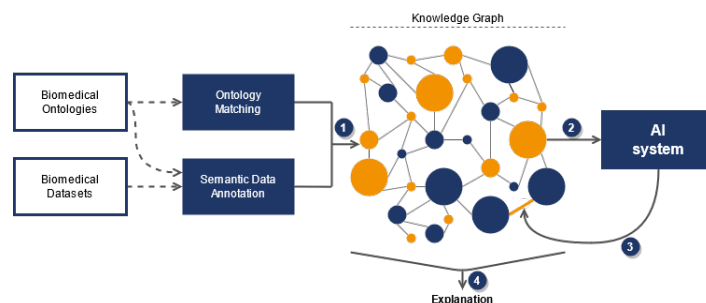


Fig. 1. Overview of the proposed approach to build a KG to support Explainable AI (XAI) in personalized medicine. Ontology Matching and Semantic Data Annotation are used to construct the KG (1), which serves as input for the AI system (2), and incorporates its outcomes (3); explanations will be derived from the KG (4).

Acknowledgments This work was supported by FCT through the LASIGE Research Unit (UIDB/00408/2020 and UIDP/00408/2020). It was also partially supported by the KATY project which has received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement No 101017453.

References

1. Chari, S., Gruen, D.M., Seneviratne, O., McGuinness, D.L.: Directions for explainable knowledge-enabled systems. arXiv preprint arXiv:2003.07523 (2020)
2. Chen, C., Cui, J., Liu, G., Wu, J., Wang, L.: Survey and open problems in privacy preserving knowledge graph: Merging, query, representation, completion and applications. arXiv preprint arXiv:2011.10180 (2020)
3. Holzinger, A., Langs, G., Denk, H., Zatloukal, K., Müller, H.: Causability and explainability of artificial intelligence in medicine. *WIREs Data Mining and Knowledge Discovery* **9**(4), e1312 (2019)
4. Lecue, F.: On the role of knowledge graphs in explainable AI. *Semantic Web* **11**(1), 41–51 (2020)
5. Li, H., Dragisic, Z., Faria, D., Ivanova, V., Jiménez-Ruiz, E., Lambrix, P., Pesquita, C.: User validation in ontology alignment: functional assessment and impact. *The Knowledge Engineering Review* **34** (2019)
6. Lima, B., Faria, D., Pesquita, C.: Pattern-guided association rule mining for complex ontology alignment. In: *ISWC 2021 Poster & Demo Track* (2021)
7. Oliveira, D., Pesquita, C.: Improving the interoperability of biomedical ontologies with compound alignments. *Journal of biomedical semantics* **9**(1), 1–13 (2018)
8. Paulheim, H., Fümkrantz, J.: Unsupervised generation of data mining features from linked open data. In: *Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics - WIMS '12*. p. 1. ACM Press (2012)
9. Pesquita, C.: Towards semantic integration for explainable artificial intelligence in the biomedical domain. In: *BIOSTEC 2021*. vol. 5, pp. 747–753 (2020)