Triangulating evidence in health sciences with Annotated Semantic Queries

Yi Liu[1], Tom R Gaunt[1,2]

[1]MRC Integrative Epidemiology Unit, Bristol Medical School, University of Bristol, Bristol, UK

[2]NIHR Bristol Biomedical Research Centre, University of Bristol, Bristol, UK

Introduction

Annotated Semantic Queries (ASQ; https://asq.epigraphdb.org; Fig.1) is a natural language query interface to the EpiGraphDB platform (https://epigraphdb.org; Liu, et al., Gaunt., 2021 Bioinformatics), which enables users to extract "claims" from unstructured text, and then investigate the evidence that could either support, contradict the claims, or offer additional information to the query.

For example (Fig.2), if a claim triple "Obesity CAUSES Asthma" (*Subject-Predicate-Object*) is identified from the query text, ASQ will retrieve associated biomedical entities and epidemiological evidence from EpiGraphDB in order for the user to assess and triangulate the evidence regarding this claim.

Methods

Claim extraction: From the query text, ASQ extracts claim triples via SemRep, and distinguishes *directional* and *non-directional* claims according to the predicate.

Entity harmonization: Query entities are harmonized with *ontology entities* (EFO) and *evidence entities* via semantic mapping.

Evidence retrieval: Curated evidence in EpiGraphDB are grouped into two evidence groups: *literature and triple* group and *associations* group, and further into evidence types according to how they relate to the claim (*supporting*, *reversal*, etc). ASQ then calculates scores (Fig.3) for evidence items to assist prioritisation of evidence items and groups.
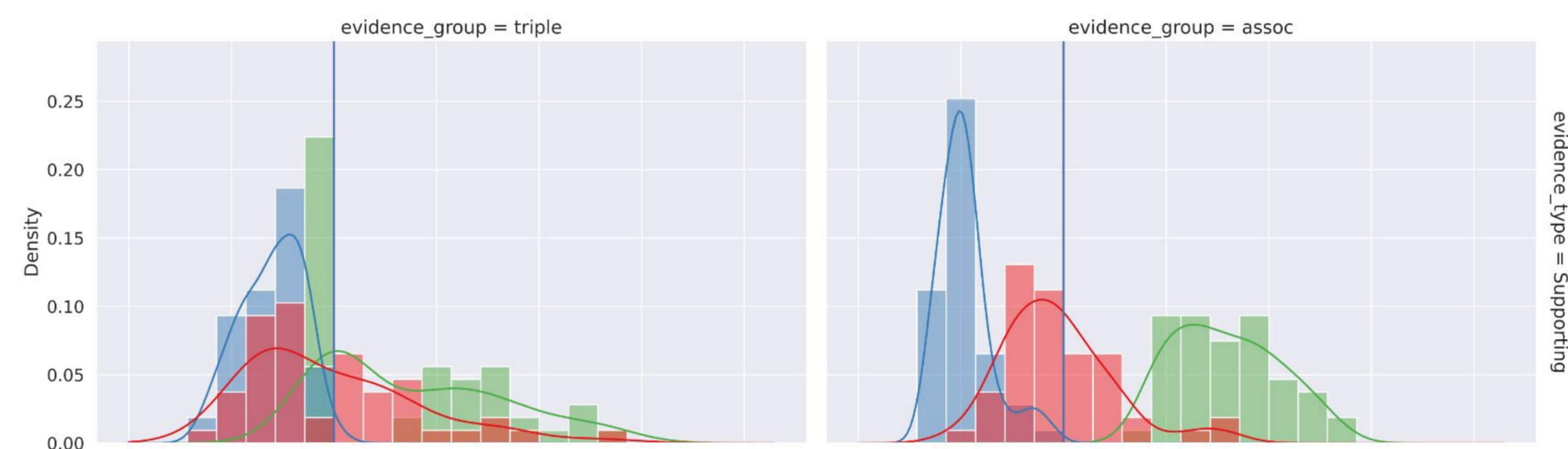

Fig. 3 Distribution of scores (mapping scores: blue; strength scores: green; evidence scores: red) of retrieved evidence items in the systematic analysis, supporting evidence type
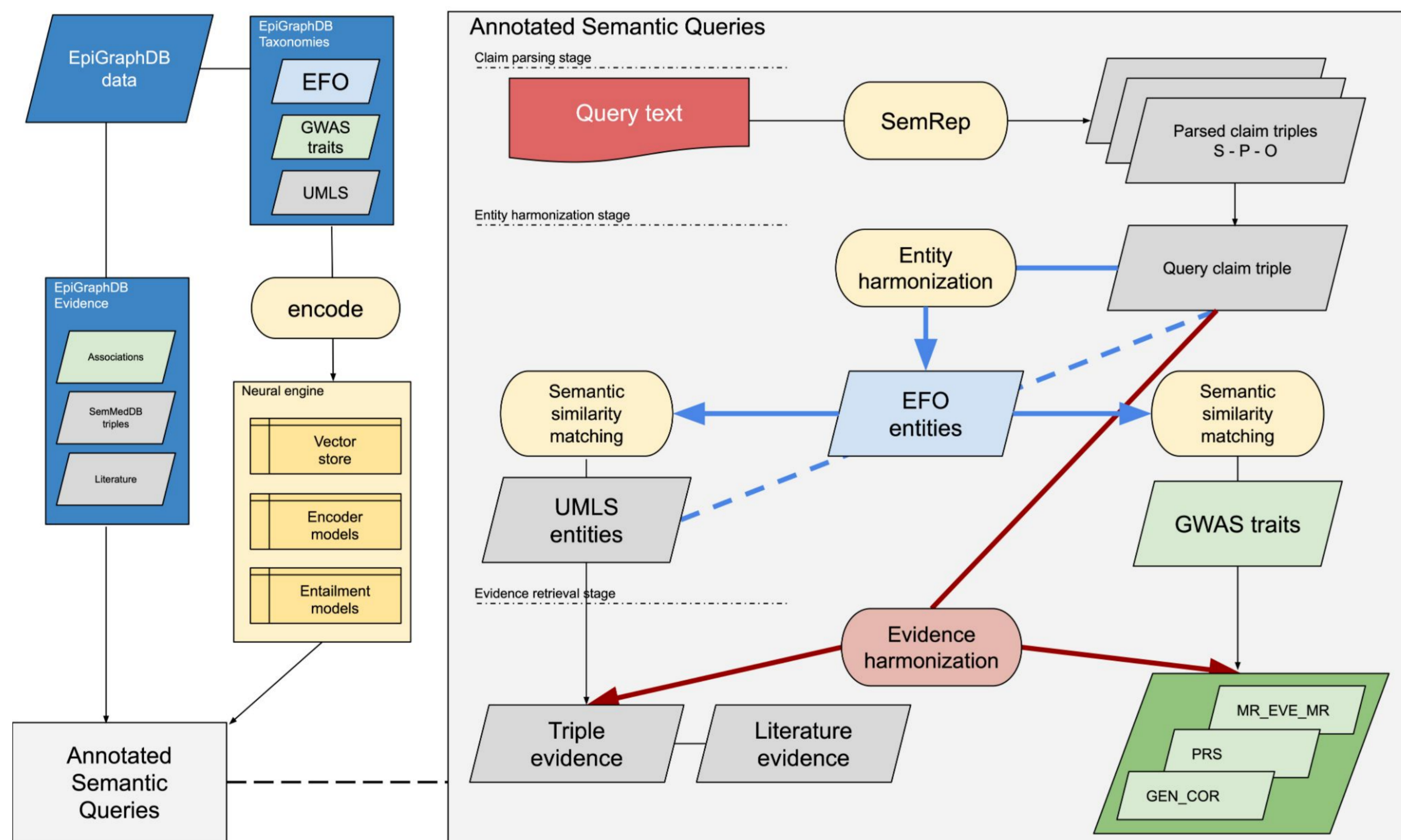

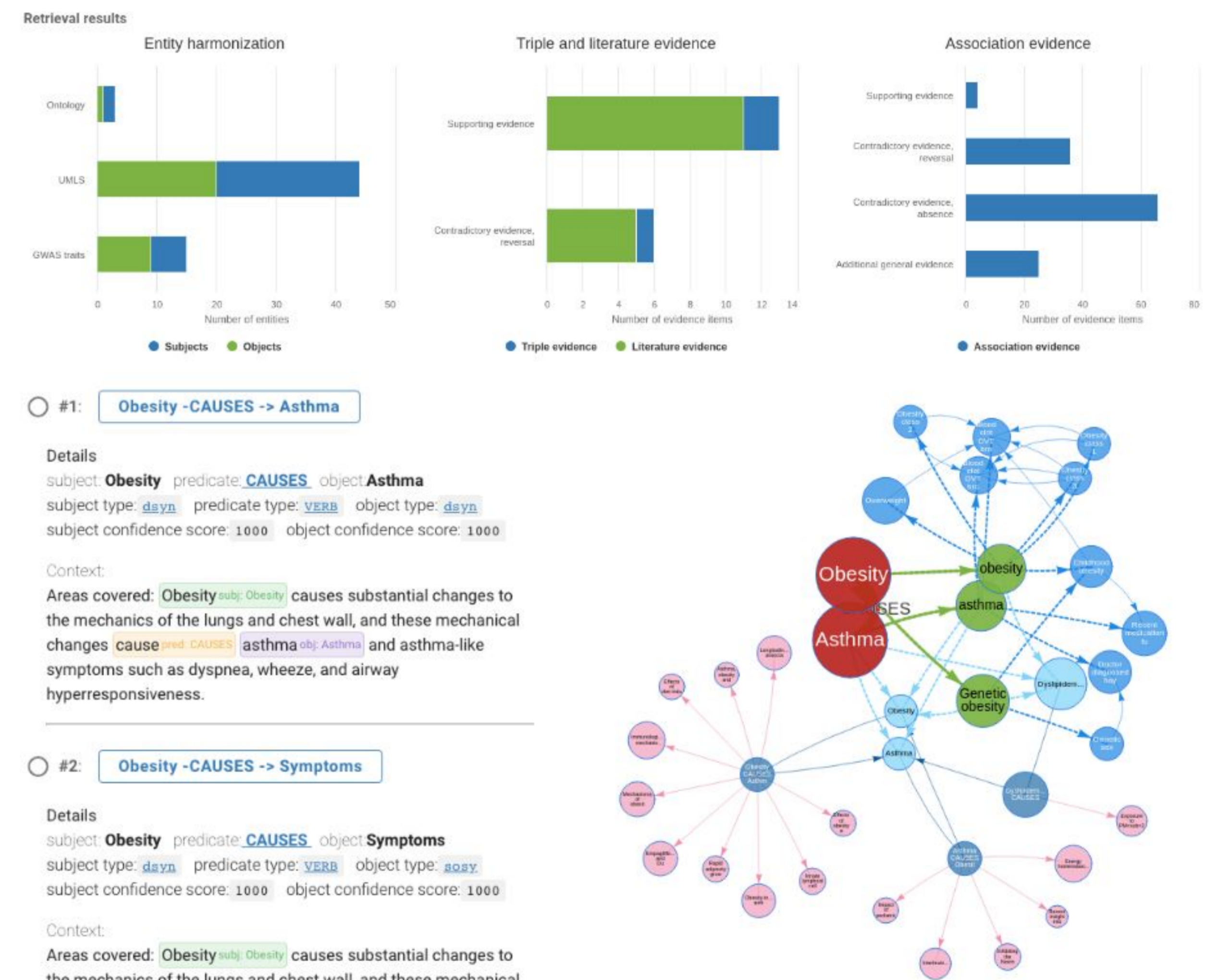*Fig. 1 Architecture of the EpiGraphDB-ASQ platform*

Fig. 2 Triangulation of evidence regarding a claim "Asthma causes Obesity"

medRxiv Analysis

We conducted a systematic analysis of 26,846 preprint abstracts on medRxiv from 2020-2021 and identified 13,295 abstract-triples where the results are available from the ASQ website.

Fig. 4 Exemplary view of obesity related disease and risk factors identified in the analysis
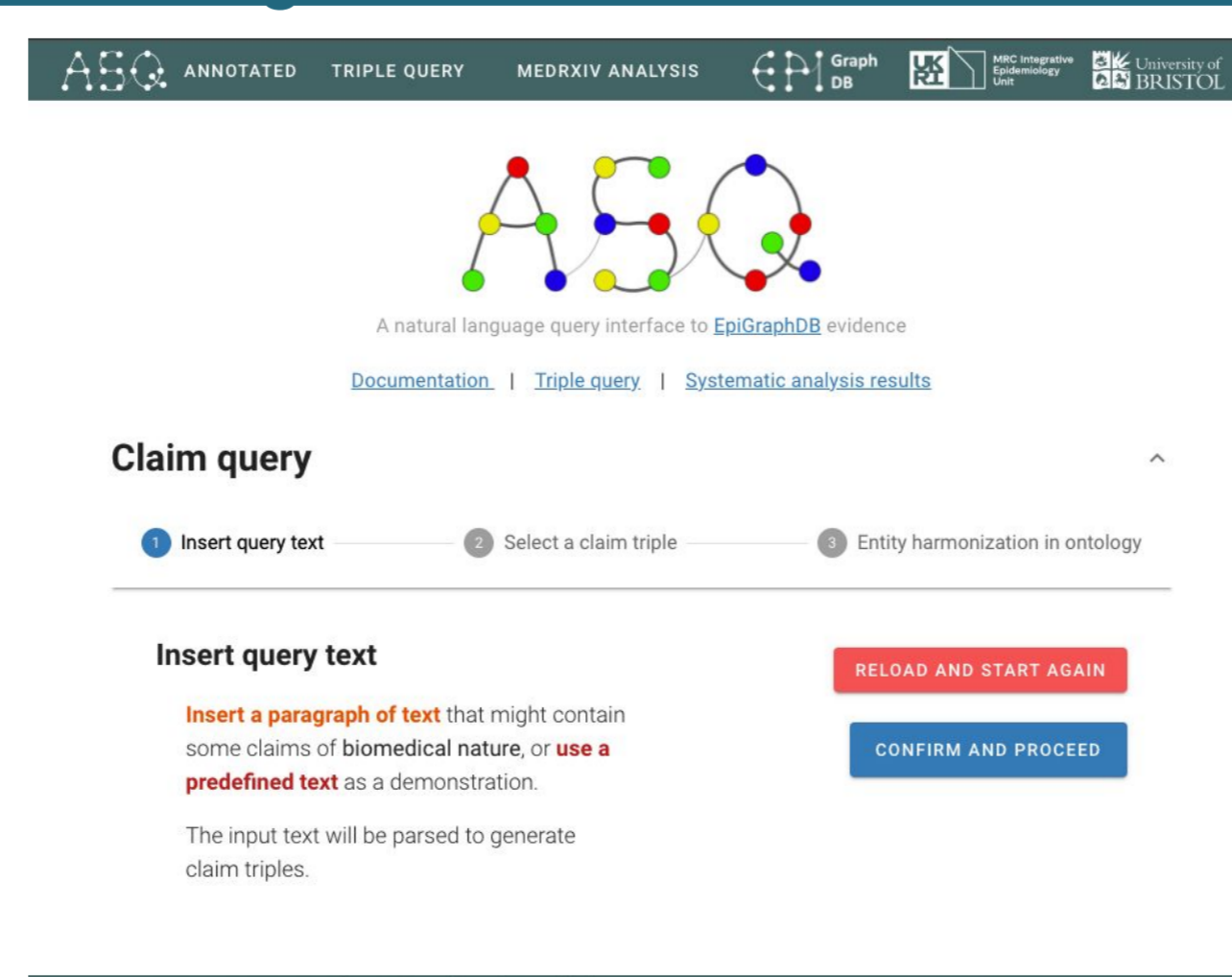
Access

Web interface: https://asq.epigraphdb.org
Source Code (GPL-3):
https://github.com/MRCIEU/epigraphdb-asq
Preprint article:
https://doi.org/10.1101/2022.04.12.22273803

Fig. 5 Web interface of the EpiGraphDB-ASQ platform

Acknowledgements

ASQ me anything