**54. FAIRVASC: A semantic web approach to rare disease registry integration**

Kris Mcglinn1, Matthew Rutherford2, Karl Gisslander, Lucy Hederman1, Mark A. Little1, Declan O'Sullivan1 *1ADAPT Centre Trinity College Dublin, Ireland,*, *1Institute of Infection, Immunity and Inflammation, University of Glasgow, Glasgow, United Kingdom, 3Division of Rheumatology, Department of Clinical Sciences, Lund University, Lund, Sweden*

**Background:** Rare disease data is often fragmented within multiple heterogeneous siloed regional disease registries, each containing a small number of cases. These data are particularly sensitive, as low subject counts make the identification of patients more likely, meaning registries are not inclined to share subject level data outside their registries. At the same time access to multiple rare disease datasets is important as it will lead to new research opportunities and analysis over larger cohorts.

**Methods:** To enable this, two major challenges must therefore be overcome. The first is to integrate data at a semantic level, so that it is possible to query over registries and return results which are comparable. The second is to enable queries which do not take subject level data from the registries.

**Results**: To meet the first challenge, this paper we present the FAIRVASC ontology to manage data related to the rare disease anti-neutrophil cytoplasmic antibody (ANCA) associated vasculitis (AAV), which is based on the harmonisation of terms in seven European data registries. It has been built upon a set of key clinical questions developed by a team of experts in vasculitis selected from the registry sites and makes use of several standard classifications, such as Systematized Nomenclature of Medicine - Clinical Terms (SNOMED-CT) and Orphacode. It also presents the method for adding semantic meaning to AAV data across the registries using the declarative Relational to Resource Description Framework Mapping Language (R2RML). To meet the second challenge, a federated querying approach is presented for accessing aggregated and pseudonymized data, and which supports analysis of AAV data in a manner that protects patient privacy. For additional security, the federated querying approach is augmented with a method for auditing queries (and the uplift process) using the provenance ontology (PROV-O) to track when queries and changes occur and by whom.

**Conclusions:** The main contribution of this work is the successful application of semantic web technologies and federated queries to provide a novel infrastructure that can readily incorporate additional registries, thus providing simultaneous access to harmonised data relating to unprecedented numbers of patients with rare disease from nearly 10,000 patients with vasculitis, while also meeting data privacy and security concerns.

**Disclosures:** None