

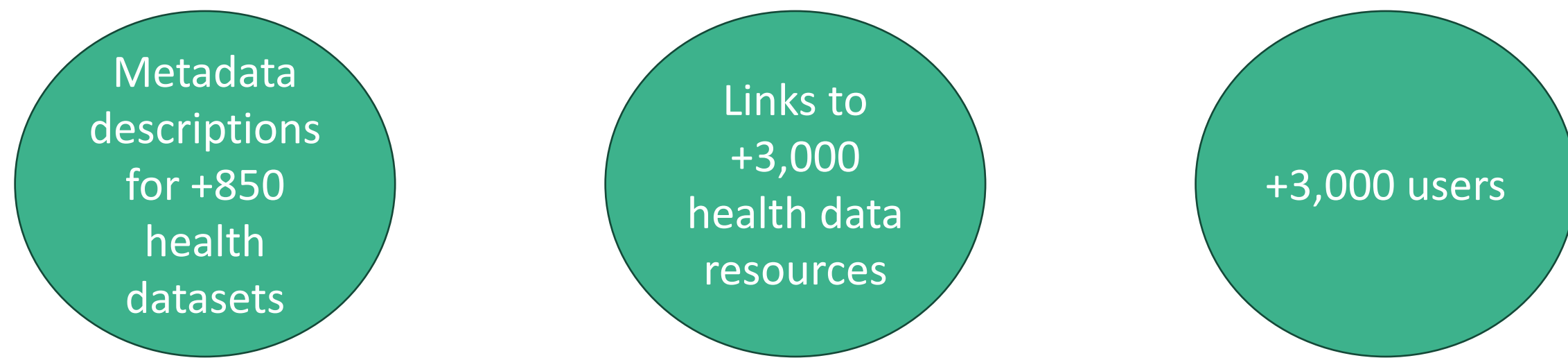
Cohort Discovery and the OMOP Common Data Model

Bradley Kirby, Charlotte Weaver, Tom Giles, Emily Jefferson
Health Data Research (HDR) UK



Introduction

Health Data Research (HDR) UK's mission is to unite the UK's health data to enable discoveries that improve peoples lives. Our vision is for large-scale data and advanced analytics to benefit every patient interaction, clinical trial and biomedical discovery and to enhance public health. The Gateway was established by HDR UK in 2020 and has:

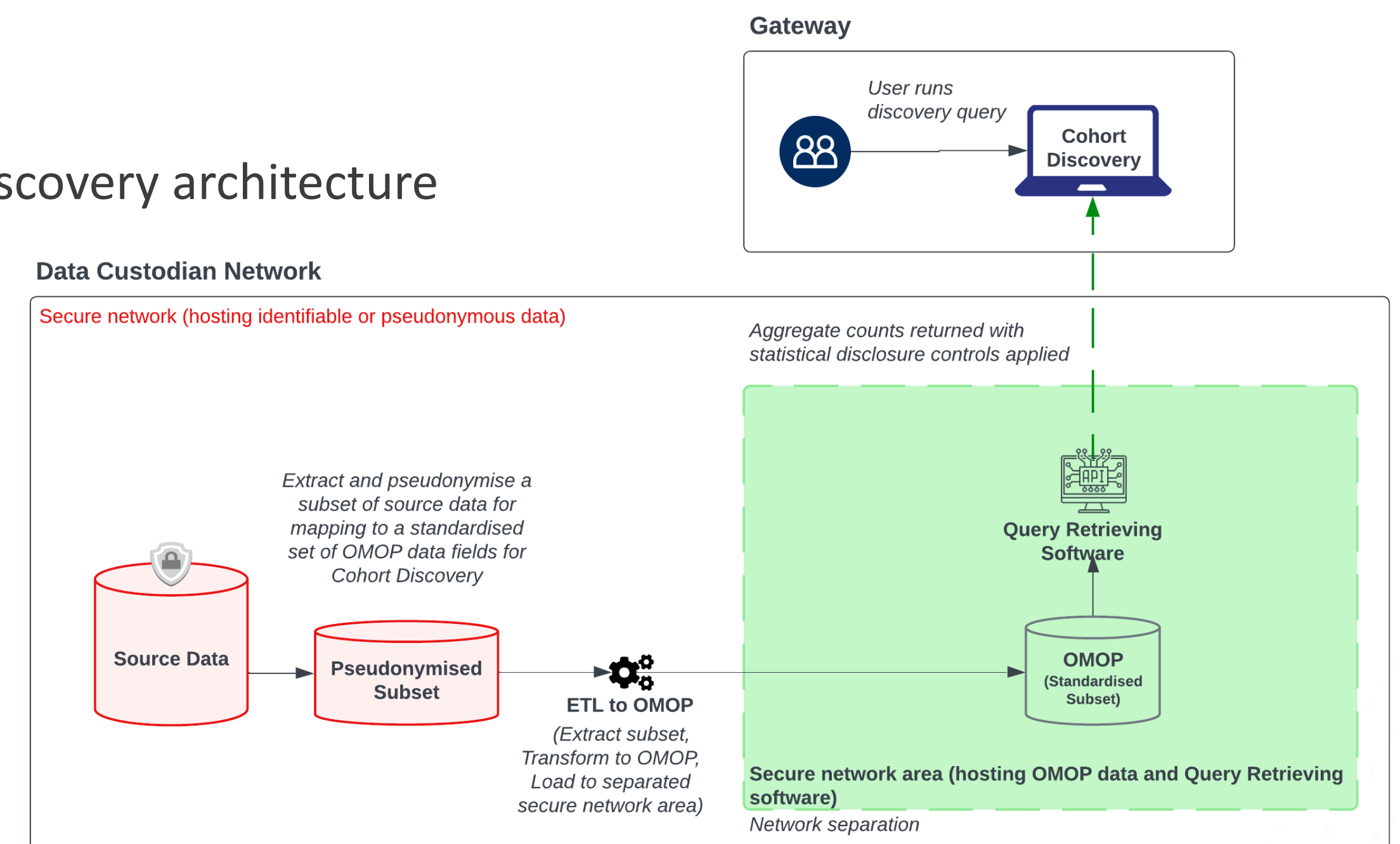


Cohort Discovery is a capability within the Gateway that aims to provide researchers with advanced data discoverability to help them find and use data for their research projects.

Methods

HDR UK are working in collaboration with NHS partners across the UK to operationalise Cohort Discovery capability in new health data domains, using OMOP as a core common data model.

Figure 1.
Cohort Discovery architecture



A key first step has been to refine the number of OMOP data fields required for Cohort Discovery to work (table 1).

Results

The core minimum set of OMOP variables is designed to expedite the transformation of source data to OMOP for Cohort Discovery, thereby making it easier to scale the service being offered by HDR UK.

OMOP Table	OMOP Field Names
Person	person_id , gender_concept_id , year_of_birth , race_concept_id
Condition_Occurrence	condition_occurrence_id , person_id , condition_concept_id , condition_start_date
Procedure_Occurrence	procedure_occurrence_id , person_id , procedure_concept_id , procedure_date
Drug_Exposure	drug_exposure_id , person_id , drug_concept_id , drug_exposure_start_date
Observation	observation_id , person_id , observation_concept_id , observation_date , value_as_number(*) , value_as_concept_id(*)
Measurement	measurement_id , person_id , measurement_concept_id , measurement_date , value_as_number , value_as_concept_id
Death	person_id , death_date , death_type_concept_id , cause_concept_id
Visit_Occurrence	visit_occurrence_id(*) , person_id(*) , visit_concept_id(*) , visit_start_date(*)

Table 1.
Minimum OMOP variables for Cohort Discovery. **Fields in bold** are core content of interest for analysis. Remaining non-bold are for linking tables. Fields with * are optional, but useful.

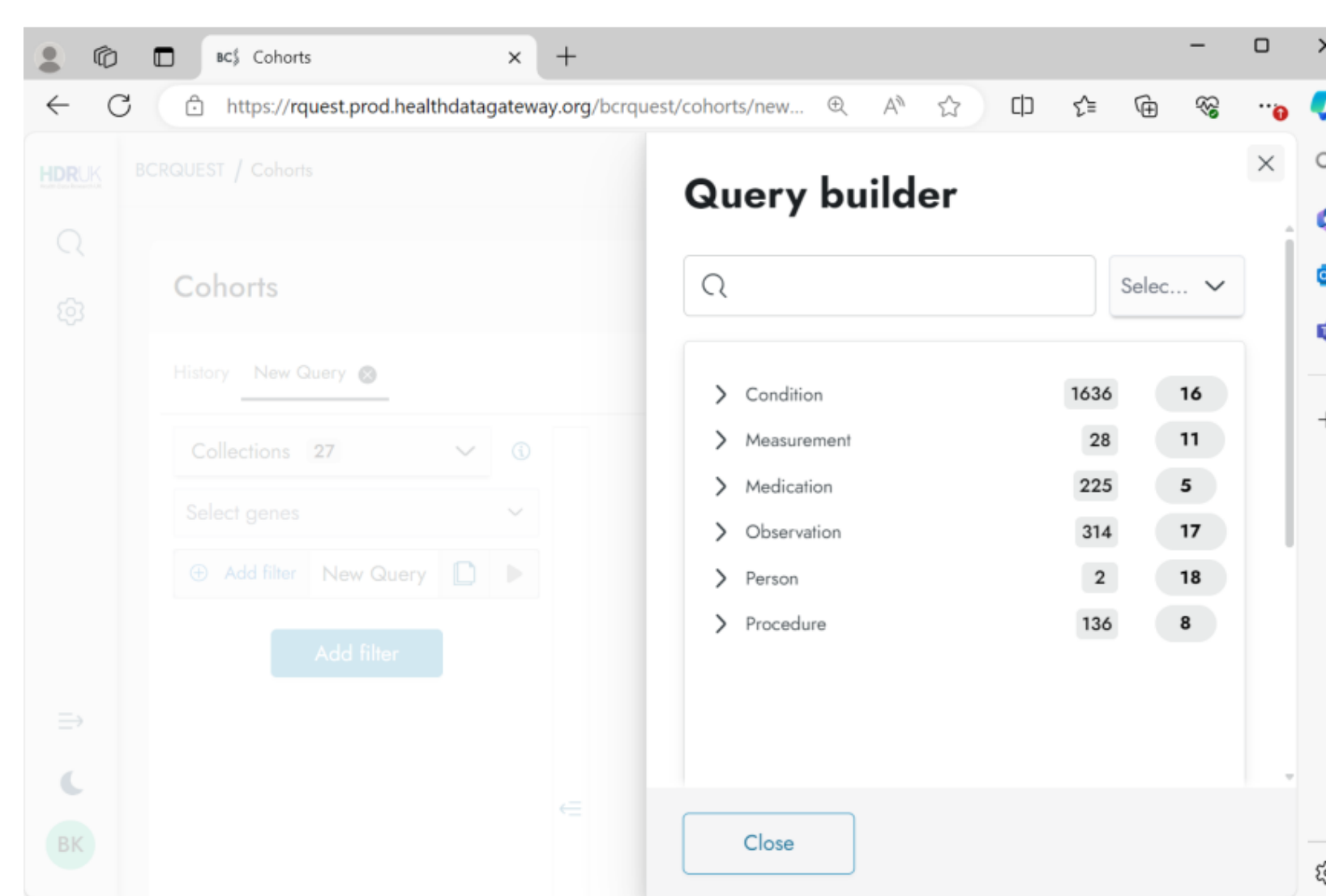


Figure 2.
Cohort Discovery query builder using the OMOP data space.

Researchers construct their query using predefined fields (figure 2), ensuring that only data that has been authorised can be queried.

Query results (figure 3) are controlled for disclosure by low number suppression and rounding; and individual identifiers cannot be queried meaning differencing attacks are not possible.

The system is designed to return granular enough results to inform a researcher where data exists that is relevant to their study, but not granular enough to allow individuals to be identified.

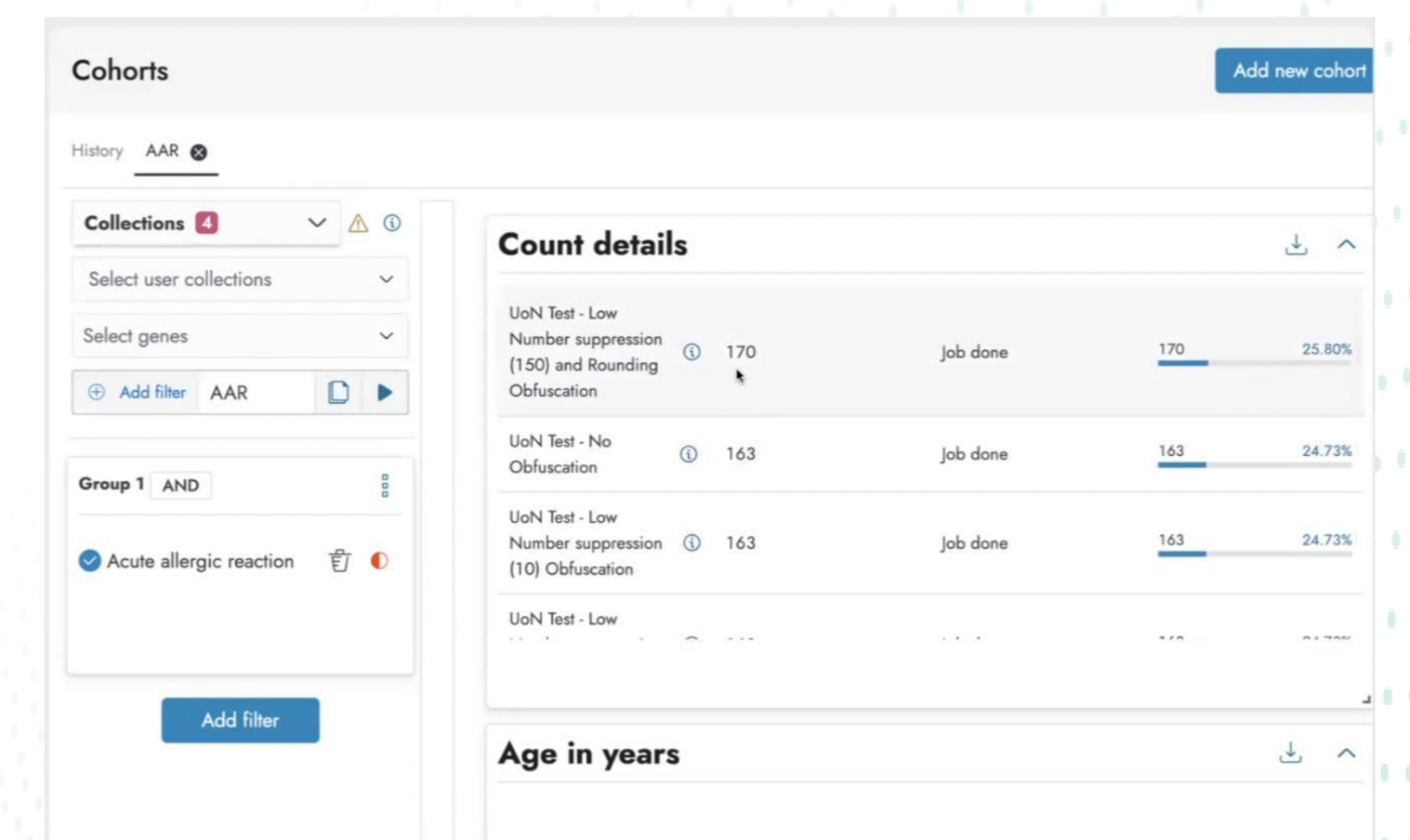


Figure 3.
Results pane for a search of 'acute allergic reaction' run against 4 OMOP data collections.

Once relevant data is identified, researchers can then submit a data access request to individual data custodians in order to carry out more detailed analysis for their research.

Discussion

Cohort Discovery currently has data collections from a range of Data Partners including the Universities of Nottingham, Bristol, Cambridge, Dundee, Edinburgh and UCL Barts; NHS GOSH; Public Health Scotland and NIHR BioResource. Work is underway to onboard many more Data Partners such as the NHS Secure Data Environment (SDE) Network and Clinical Practice Research Datalink (CPRD), as well as expanding the number of data collections available via current Data Partners. Federated querying of data conforming to the OMOP common data model streamlines the discovery phases of research by enabling researchers to rapidly find out what data is available without having to directly contact individual organisations that hold the data. The specificity of subsequent data access requests is improved, thus saving time, effort and costs for Data Partners by reducing the burden of responding to individual researcher queries about their data. A key challenge is encouraging data custodians to map source data to OMOP, due to the complexities of data transformation and slow ROI.

References

- Health Data Research UK. Health Data Research UK launches transformational new search functionality [Internet]. UK: HDR UK; 2021 [cited 2024 Jul 4].
- Jefferson E, Cole C, Mumtaz S, Cox S, Giles T, Adejumo S, et al. A Hybrid Architecture (CO-CONNECT) to Facilitate Rapid Discovery and Access to Data Across the United Kingdom in Response to the COVID-19 Pandemic: Development Study. J Med Internet Res 2022 [cited 2024 Jul 8];24(12):e40035.
- Health Data Research UK. NHS Research SDE Network agrees to adopt common data model [Internet]. UK: HDR UK; 2023 [cited 2024 Jul 4].
- Observational Health Data Sciences and Informatics (OHDSI). Standardized Data: The OMOP Common Data Model [Internet]. [place unknown; cited 2024 Jul 4].