ARES: A Research Exploration System

Frank J DeFalco, Alan Andryc, Anthony Molinaro, Clair Blacketer

Background

The Observational Health Data Sciences and Informatics (OHDSI) community has developed many open source software solutions for data management, transformation, and analysis. The Automated Characterization of Health Information at Large-scale Longitudinal Evidence Systems (ACHILLES) package provides descriptive statistics about an Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) database.[1] The Data Quality Dashboard (DQD) package provides a tool to help improve data quality standards in observational data science.[2] These two distinct and separate packages are valuable to the community individually but the analyses they provide, characterization and quality assessment of observation data, could be more compelling and provide much needed context if presented in combination through A Research Exploration System (ARES).

Methods

A Research Exploration System (ARES) was developed in two parts; AresIndexer and the ARES web interface. AresIndexer is a new data indexing service (AresIndexer)[3] that leverages the outputs of ACHILLES and DQD in order to combine their information into a consolidated web-based user experience (ARES)[4]. The AresIndexer package is delivered as an R package, in accordance with the development methodologies of other OHDSI packages. The ARES web interface is delivered as a simple web interface without the requirement of middleware or other service layers.

Results

An integrated web interface presents the results of the new indexing service in order to combine data quality and characterization into a unified interface.

ARES provides reporting at three distinct levels: Network, Data Source, and Data Source Release. The Network reports provide an overview of all data sources participating in the network. The Data Source reports provide information about an individual data contributor including analyses of changes across that source's historical releases. The Data Source Release reports describe a data source at the release level, providing characterization and data quality results for a point-in-time release of the data.

Data quality results are presented alongside data characterization to ensure visibility to potential issues while performing feasibility analysis. Domain level reports such as condition occurrence can be supplemented with the number of data quality issues identified within that domain as shown at the top of figure 1. Visualization techniques such as preattentive processing are employed to assist in the interpretation of tabular data results as shown in the table in figure 1.
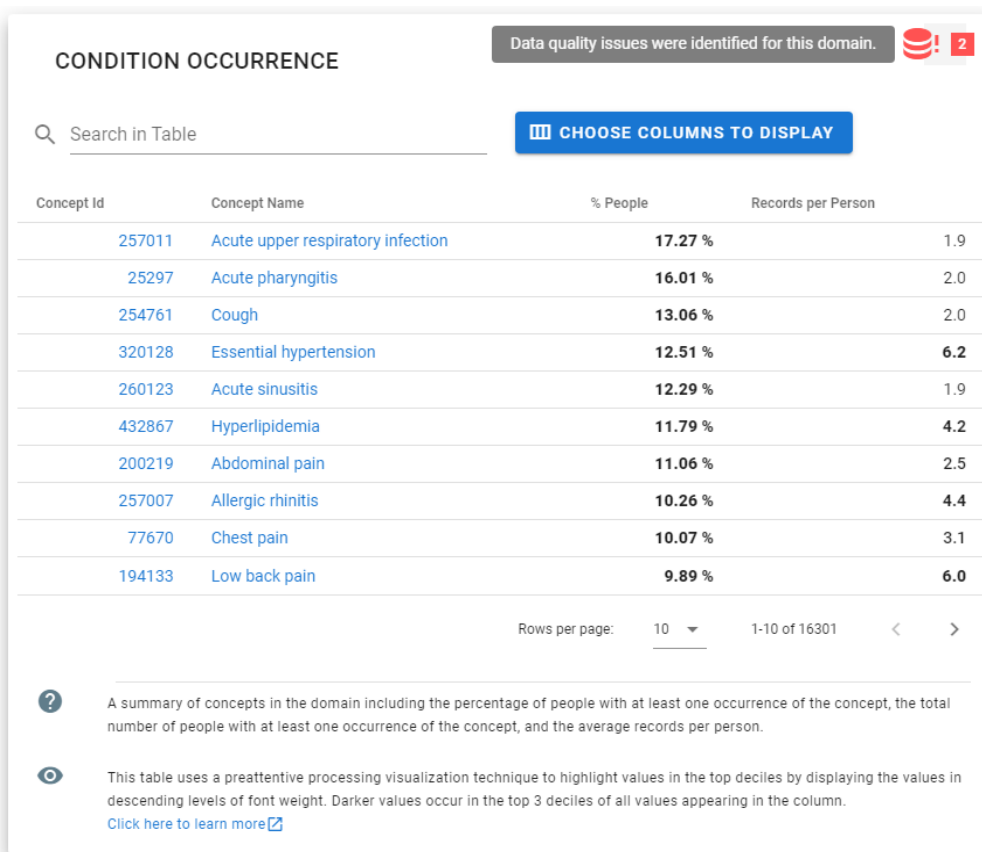
Figure 1. Integrated Characterization and Data Quality

Novel visualizations such as the Data Strand report shown in figure 2, provide a high level overview of the domain contents of the data network exposing variability of content across data sources in the network in a readily interpretable format.
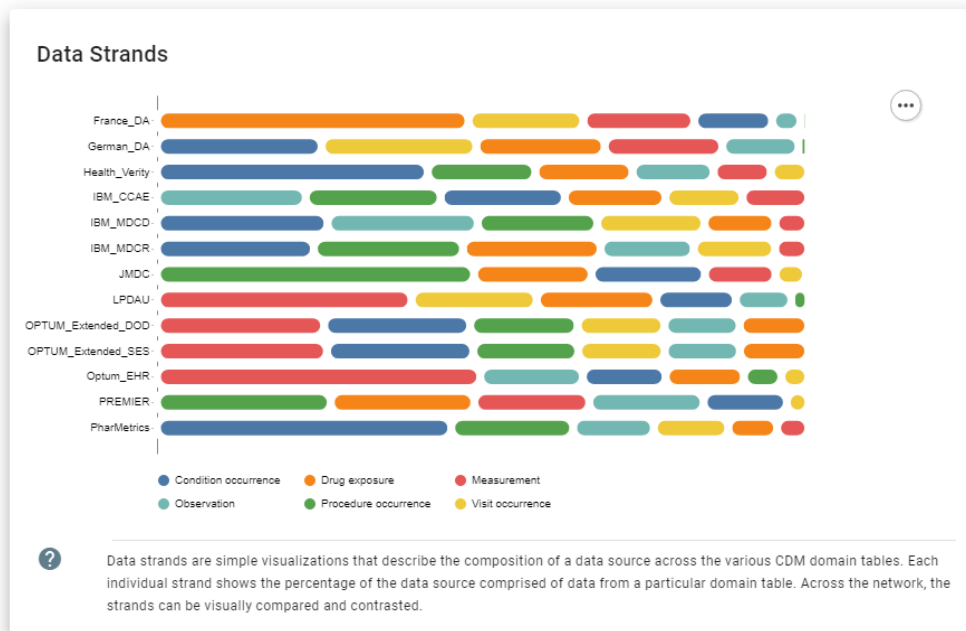
Figure 2. Network Domain Data Strand Visualization

Data source release reporting allows for historical trends in data quality to be analyzed for the first time within the

OHDSI tools as shown in figure 3. The results from both ACHILLES and DQD are indexed to allow for analyses to be performed not only within a particular release but also across releases of the same data source.
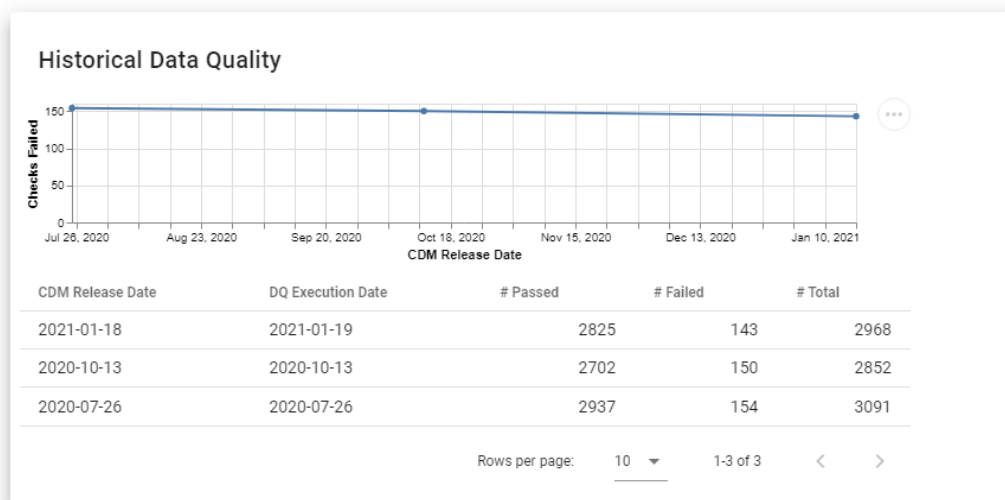

Figure 3. Data Source - Data Quality Trend Reporting

Conclusion

ARES provides a unique combination of reports and visualizations leveraging existing OHDSI tools and integrating information that was previously disparate. New reports for data network visualization and characterization are provided to enable new insights into domain data content. ARES sets a new standard for characterization and data quality reporting for OHDSI data networks.

References/Citations

1. Automated Characterization of Health Information at Large-scale Longitudinal Evidence Systems (ACHILLES) [web page]. GitHub. Observational Health Data Sciences and Informatics; [Available from: https://github.com/ohdsi/achilles]

2. Data Quality Dashboard (DQD) [web page]. GitHub. Observational Health Data Sciences and Informatics; [Available from: https://github.com/ohdsi/dataqualitydashboard]
3. ARES Indexing Service (AresIndexer) [web page]. GitHub. Observational Health Data Sciences and Informatics; [Available from: https://github.com/ohdsi/aresindexer]
4. A Research Exploration System (ARES) [web page]. GitHub. Observational Health Data Sciences and Informatics; [Available from: https://github.com/ohdsi/ares]