The emrAdapter Tool: A General Purpose Translator for Electronic Clinical Data

Abstract
When assembling research data from multiple and diverse sources, researchers must contend with differing formats and languages of the source data. emrAdapter is a tool that facilitates manipulation and translation of heterogeneous electronic clinical data, allowing for automated analysis of standardized datasets. A graphical user interface allows specification of the structure of input data to be used, the structure of output data to be generated, and the details of transformations to be performed during processing.

Description
Often the source data is not available in a format that is compatible with research tools, or the format and meaning of elements varies across data sources. emrAdapter provides a solution for selecting and transforming diverse source input data, generating standardized records in support of research tasks that process large volumes of heterogeneous data. emrAdapter uses a common file format description language to allow input and output files to be described, including both available Markup Language (XML) and Comma Separated Value (CSV) data formats. Furthermore, because semantics of the data fields within these files may vary considerably, emrAdapter uses Java source code (provided as plug-in functions) to constrain field-mapping with logic that allows standardization of disparate input files into a single shared document type. This is required when performing CER studies that span multiple sites and their heterogeneous EHRs.

Applications using emrAdapter
- The Clinical Research Document (CRD) application is an XML document that aggregates all of the significant medical information from any EHR into a single shared document type. This is required when performing CER studies that span multiple sites and their heterogeneous EHRs.
- The emrAdapter tool is being utilized by large-scale comparative effectiveness research projects to enable a common data format for standardized local processing of clinical data records allowing for the aggregation of study-specific clinical data from diverse EMRs and care organizations.

The CERHUB uses the CRD and emrAdapter Data Processing Pipeline (www.cerhub.org)