Interactive Feasibility and Data Explorer for Clinical Data (FEDEX)

Sanketa Hegde[1,2], Harald Wilhelmi[1,2], Christoph Dieterich[1,2,3]

[1]*Department of Internal Medicine III, University Hospital Heidelberg, DE*
[2]*Section of Bioinformatics and Systems Cardiology, Klaus Tschira Institute for Integrative Computational Cardiology, Heidelberg, DE*
[3]*Informatics for Life, Heidelberg, DE*

ABSTRACT:

Dynamic visualization of structured patient data is essential for clinical study conduct and decision making. Physician scientists appreciate interactive graphical interfaces to rapidly explore large structured data volumes originating from clinical studies and/or routine data. We have identified particular needs for hypothesis generation, cohort building and data quality control. The new Feasibility and Data Explorer (FEDEX) is an extension of the DZHK Feasibility Explorer [1], by the integration of Medical Data Explorer (MEDEX) software [2].

The FEDEX software is implemented with a JavaScript front-end with Bootstrap v4, a Python application layer using the Flask framework, and a PostgreSQL database back-end. Our open-source project is available via GitHub URL: https://github.com/dieterich-lab/medex/tree/Fedex. The main layout gives the user several options for visualizing clinical data in various formats: tabular, aggregated statistics or a whole range of plots: Scatter plot, Barchart, Histogram, Box plot and Heatmap. All of the aforementioned features enable hypothesis building from data exploration. Outlier detection is made easy and can be integrated in a quality control step of any study. New features, which can be used for cohort building, include powerful filtering options that allows the user to specify data subsets that meet criteria such as certain age intervals or smoking. Given use and access consent, FEDEX also offers customized data and figure download on demand. Informative error handling ensures that the user fills in all required input fields in order to obtain the desired results.

We have tested the performance of FEDEX on the UK Biobank data set with approximately 500,000 patient health records over a wide range of numerical and categorical data features. Typical response times on standard hardware with an Intel Core i7-7600U CPU at 2.80GHz and 16 GB RAM, fall below 2-3 seconds. FEDEX imports data in CSV format and has been tested on DZHK cohort data and various HiGHmed data sets as well. We currently plan to implement it with MII Broad Consent data at our local Medical Data Intergration Centre (MEDIC). Taken together, FEDEX empowers physician to do data science on clinical data.

[1]   Kindermann, A., Stepanova, E., Hund, H., Geis, N., Malone, B., & Dieterich, C. (2019). MedEx - Data Analytics for Medical Domain Experts in Real-Time. Studies in health technology and informatics, 267, 142–149. https://doi.org/10.3233/SHTI190818

[2]   Scheel, H., Dathe, H., Franke, T., Scharfe, T., & Rottmann, T. (2019, September). A Privacy Preserving Approach to Feasibility Analyses on Distributed Data Sources in Biomedical Research. In *GMDS* (pp. 254-261).