

A web based tool for Import, Mapping and Harmonization of Tabular Data (applied to clinical data)

Wednesday 11 October 2023 14:06 (3 minutes)

The field of clinical research data is rich in information that is often underutilized due to the complexity resulting from heterogeneous representations of the data and the lack of suitable tooling for its harmonization. Ineffective data preprocessing hinders potential insights and prevents effective reuse and combination of data that could otherwise drive progress in the scientific field by providing additional evidence. To counter this problem, we propose a web application designed to assist users in harmonizing non-standardized tabular data. It enables the seamless import of data in common file formats such as CSV, XLSX, or XLS and allows exporting the harmonized data either as CSV or directly uploading it to REST APIs. Because of its versatility, it can also be useful to address the harmonization problem in other scientific disciplines. Imported data can be mapped and validated against a JSON-Schema. As long as the intended data structure can be articulated in the form of a JSON-Schema, it can be incorporated and utilized in the system. Furthermore, complex transformations on the data can be interactively developed and performed during the import process by utilizing JavaScript. By taking into account the specifics of the research data lifecycle, our tool provides comprehensive support to researchers. In particular, it facilitates essential steps such as pre-processing, validation, and optional data migration. The migration process allows users to map columns of the table to a given JSON schema. Providing these specialized functions not only simplifies data processing but also ensures data longevity that can be effectively adapted to an evolving research environment.

Consequently, this web application is a promising tool for improving data use across a wide range of scientific disciplines. It offers features that serve important functions in a variety of research areas in general. However, it is important to note that while the tool has the potential to meet the use cases and goals outlined, a comprehensive evaluation of its full capabilities in various real-world scenarios has not yet been conducted. Its performance, especially when processing large datasets, the potential security concerns related to JavaScript transformations, and the ability to meet all predefined requirements need further investigation. The next step in the development is to develop use cases for a more detailed evaluation. In its current state, the tool provides a foundation that can help researchers from numerous disciplines to harmonize data and reduce the overhead associated with redundant data collection by leveraging multiple data sources.

Please assign your contribution to one of the following topics

Data interoperability through harmonised metadata and interoperable semantics

Please specify "other" (stakeholder)

In addition please add keywords.

HMC, Tool, Harmonization, Clinical-data

Please assign yourself (presenting author) to one of the stakeholders.

Researchers

Primary author: KULLA, Lucas (DKFZ)

Co-authors: MAIER-HEIN, Klaus (DKFZ); NOLDEN, Marco (DKFZ); SCHADER, Philipp (DKFZ)

Presenter: KULLA, Lucas (DKFZ)

Session Classification: HMC Hands-on Session