

Designing a Visual Analytics Dashboard to Streamline Collaborative Common Data Element Mapping

Ruey-Ling Weng, MS, Huan He, PhD, Vincent J Zhang, MS, Yujia Zhou, MS,
Lingfei Qian, PhD, Jihoon Kim, PhD, Na Hong, PhD, Hua Xu, PhD

Department of Biomedical Informatics & Data Science, Yale School of Medicine, New Haven, CT

Introduction. Common Data Elements (CDEs) are standardized data structures that enable data harmonization and interoperability across studies. Mapping research data to CDEs is essential for achieving these benefits, and several mapping tools have been developed to facilitate this process¹. However, due to the complexity of the CDE mapping workflow, such as extensive manual work and iterative multi-round review, it's still challenging to track the overall mapping progress. Additionally, a lack of understanding of the overall distribution of CDEs and data elements also challenges the decision-making. To address these challenges, we leverage human-centered design principles to design and develop a visual analytics dashboard that aims to facilitate the CDE mapping process.

Method. To understand the design requirements, we conducted several rounds of interviews with biomedical researchers experienced in data normalization with prototype designs. Through this iterative process, we identified key design principles to guide the development. **First, workflow traceability:** During the mapping process, it's crucial to ensure that the progress of every task in every project can be clearly visualized. **Second, multi-level transparency:** It's necessary to obtain the insight from multiple levels, including project, task, and detailed data element, enabling users to easily understand the status. **Lastly, data-driven interactivity.** Domain experts emphasized that the interface should be able to help users get the data behind the charts, which allows users to gain more insights.

Results. Based on the above design principles, we designed the visual analytics dashboard, which consists of three main views: **1) Task heatmap for tracking project-task level progress** (Fig. 1a). This chart provides a granular view across all ongoing projects, enabling users to assess the overall progress at a glance. Mapping tasks (i.e., a list of elements to be mapped) in each project are listed in one row, and each task is represented by a rectangle, colored by its working stage, including *Not Started*, *Under Mapping*, *Under Review*, and *Finalized*.

Additionally, the *Under Review* stage further displays a percentage-based bar to indicate the review progress, allowing for a more detailed understanding of task completion. Clicking on a task reveals a popup tooltip containing additional information and also allows users to access the task directly; **2) Sankey diagram for identifying element level progress** (Fig. 1b). This diagram illustrates the overall distribution of elements in each stage in one project. Hovering over the stage bar or the flow path could present details about the transition, highlighting bottlenecks and overall efficiency. **3) Semantic map for understanding concept coverage** (Fig. 1c). Leveraging large language model-based text embedding, all data elements are embedded into a semantic space and visualized on a 2D map with all the CDEs. Each dot represents a data element, colored by type, i.e., gray dots represent existing CDEs while other colors represent elements to be mapped. Users can quickly identify whether the elements are semantically covered by existing CDEs.

Future work. Initial feedback on the prototype from domain experts has been positive. Our next steps include integrating this dashboard into our tool CDEMapper (<https://cdemapper.clinicalnlp.org/>) and conducting a formal evaluation. The prototype design and online demo are available at <https://clinicalnlp.org/cde-mapper-dashboard/>.

References

1. S. Tao et al., "IMI-CDE: an interactive interface for collaborative mapping of study variables to common data elements," 2022 IEEE 10th International Conference on Healthcare Informatics (ICHI), Rochester, MN, USA, 2022, pp. 465-468

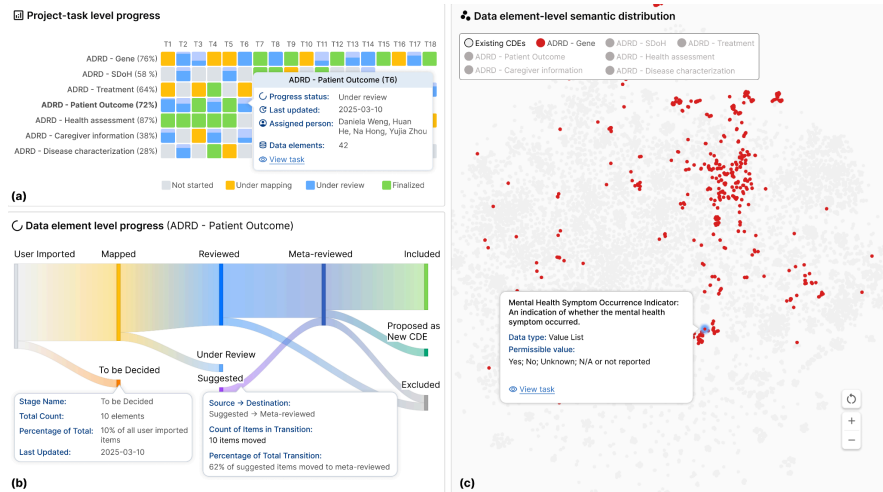


Figure 1: The visualization design, including (a) a task heatmap for project-task level progress tracking; (b) a Sankey diagram showing element level progress; and (c) a scatter plot for showing the element-level semantic distribution of all elements to be mapped on the entire CDEs.