

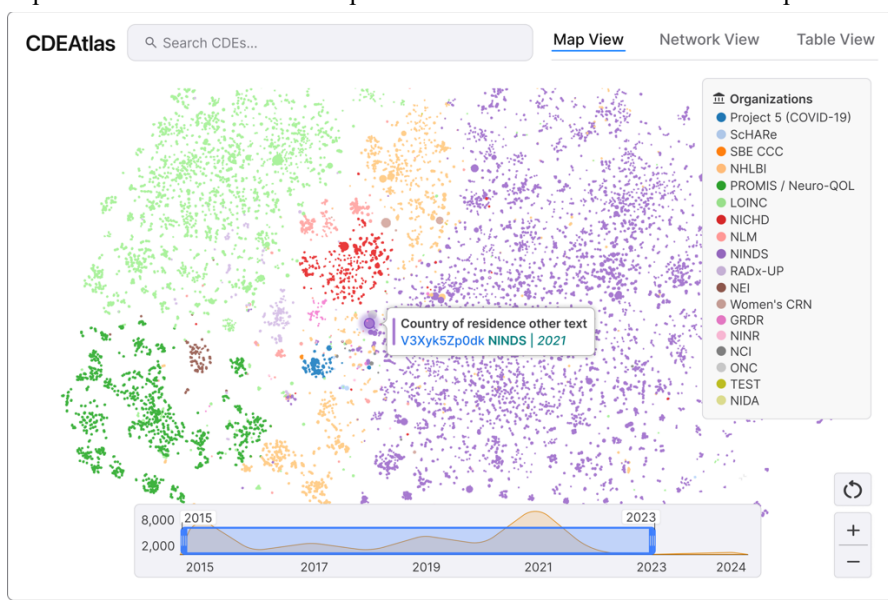
# CDEAtlas: Interactive Visualization and Exploration of Common Data Elements at Scale

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

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

**Introduction.** Despite the growing volume and importance of Common Data Elements (CDEs) in promoting data standardization and interoperability across biomedical research,<sup>1</sup> navigating diverse CDE repositories remains challenging due to their scale, heterogeneity, and limited exploratory tools. Researchers and data curators often struggle to identify relevant CDEs across diverse domains and platforms, gain insights or track CDE evolution and new trends. To fill this gap, we aim to develop CDEAtlas, an interactive visualization tool designed to help researchers, data curators, and policy-makers understand and navigate the entire CDE ecosystem. We identified two key requirements to guide the system design. First, the global overview. The system should provide an at-a-glance understanding of the entire CDE landscape. Second, multi-level insights and exploration. The user interface (UI) should enable users to explore the CDEs across multiple levels, such as clusters, domains, and temporal trends, to reveal semantic patterns and historical evolution. Our tool addresses these needs by offering an intuitive visualization interface that supports integrated exploration of CDEs from multiple sources within a unified interactive platform.

**Methods.** We followed a human-centered and data-driven design process to design the tool, which incorporates three core components: 1) a data pipeline, which preprocess the CDEs and generates CDE embeddings from the question text, CDE definitions and permissible values. These CDE embeddings were further mapped to 2D space; 2) an interactive 2D point-cloud map. We employ Three.js to visualize all the 2D CDEs in the web browser. Users can interact with each point for more details. 3) an in-browser search, which supports keyword search and provides direct link-outs to the source CDEs, enabling users to quickly locate relevant CDEs and transition between global and detailed views.



**Figure 1:** The landscape view visualizes all CDEs as a 2D semantic map colored by publication period, supporting exploration of temporal concept trends with an integrated time-range selector and a search tool that enables direct navigation for further exploration.

**Results.** The interface of CDEAtlas is illustrated in Figure 1. *Exploring the CDE landscape.* We processed 22,743 CDEs from the NIH CDE Repository<sup>2</sup> using large language model based text embeddings, all CDEs are projected into a shared semantic space and visualized as an interactive 2D semantic landscape. Each element is color-coded by organization, enabling users to intuitively observe which domains were most active within the CDE repository. The integrated search bar enables efficient concept retrieval and provides direct navigation for deeper exploration within the CDE semantic space. Each dot directly links to its corresponding NIH CDE pages. *Tracking the temporal evolution of CDEs.* To support longitudinal exploration, a temporal density panel allows users to filter and refine the visualization by specific year ranges. This interactive feature enables the identification of temporal patterns, revealing how CDEs have evolved over time, highlighting periods of increased activity, emerging domains, or shifts in CDE development priorities across the NIH CDE Repository.

**Discussion.** This tool enables researchers and data curators to observe the CDE landscape at a glance, quickly locate specific CDEs, discover similar CDEs, and understand the evolution of CDEs across time and organizations. Next, we will visualize more fine-grained relationships among CDEs and conduct an usability evaluation.

## References

1. Kush RD et al. FAIR data sharing: the roles of common data elements and harmonization. J Biomed Inform. 2020 Jul 1;107:103421.
2. NIH CDE Repository <https://cde.nlm.nih.gov/home>. Accessed Nov. 21, 2025.