

Patricia Kovatch, Deborah L. McGuinness, Chris Gennings, Susan L. Teitelbaum. Accelerating Child Health Exposure Research with the CHEAR Data Center. In International Society of Exposure Science (ISES), Utrecht, Netherlands, October 2016.

The United States National Institute of Environmental Health Sciences has established an infrastructure, the Childrens Health Exposure Analysis Resource (CHEAR), to provide the extramural research community access to laboratory and statistical analyses aimed at adding or expanding the inclusion of environmental exposures in their research. The CHEAR project is a \$50 million multiunit infrastructure composed of a coordinating center, a network of laboratories and a data center tasked to provide researchers access to comprehensive exposure assessment for NIH funded studies of childrens health. The CHEAR Data Repository, Analysis and Science Center (Data Center) is located at the Icahn School of Medicine at Mount Sinai in collaboration with Rensselaer Polytechnic Institute in New York. The goal of the data center is to catalyze new scientific insight from the co-location, integration and advanced statistical and data science analysis of multimodal data sets. The data center provides the intellectual and logistical support for the validation, interpretation, curation, and maximum reuse of data generated by the laboratory network. We aim to provide access to tools and services that incorporate and extend exposure analysis on an exposome scale (i.e., to study complex environmental influences on health) by providing a strong data, knowledge, and analytic infrastructure. We are developing semantic infrastructure for support in consistent modeling, unambiguous interpretation, and enhanced integration. For the investigators that utilize CHEAR for studies of childrens environmental health using the data generated within and outside the network, the Data Center provides: 1) data repository and management; 2) statistical consultation and analysis services; 3) collaborative research support; 4) statistical and analytical methods development; and 5) data science resources, including semantic infrastructure and services powered by a family of child health exposure ontologies. In this presentation, we will discuss the opportunities for advancing the study of early life environmental exposures and later life health consequences with advanced statistical and data science approaches including the use of knowledge graphs and ontologies. We will also review our initial lessons learned from building the data repository and developing its accompanying policies for data sharing.