DATA MANAGEMENT AND SHARING PLAN

Element 1. Data Types

A. Types and amount of scientific data expected to be generated in the project:

The proposed research will generate both quantitative and qualitative data.

Quantitative data includes three data sets: two data sets containing questionnaire responses from three questionnaires being administered at different points during this research and one data set containing measurement data for several indoor environmental quality measures. The first quantitative data set will include responses from 1,000 childcare providers and other members of the child serving ecosystem. This first questionnaire will be administered to participants as part of the work described in Aim 2 and will ascertain information about environmental health knowledge, attitudes, and beliefs among those who work in the childcare setting. We are particularly interested in learning more about the levels of environmental health literacy among childcare workers through this questionnaire. The second quantitative data set, which will be generated as part of the work described for Aim 3 will include initial and follow-up responses to questions about provider homes, including the structure, size, and materials, as well as daily activities that take place within the home. We anticipate 60 responses total (two responses from 30 participants over the course of the research). Questionnaire data will be collected via REDCap. The third quantitative data set will include data collected within the homes of the 30 care providers participating in the exposure assessment study in Aim 3. These data will include measurements of indoor environmental quality variables (*e.g.,* temperature, humidity, CO_2), indoor air pollutants (*e.g.,* NO_2 and formaldehyde), water quality measures (*e.g.,* bacteria, lead), and emerging contaminants of concern (*e.g.,* PFOA in house dust). All quantitative data will be stored in tabular format. All quantitative data will be deidentified prior to deposit in the publicly available repository.

Qualitative data include data gathered from the policy and practice scan performed during Aim 1 and from approximately 60 semi-structured interviews with care providers and other members of the child serving ecosystem conducted during Aim 2. Policy and practice scan data will be stored in a tabular format. Semi-structured interviews with childcare providers will focus on provider attitudes and beliefs about children's environmental health, their knowledge about how to promote environmental health in their care settings, and barriers they may face as family-based childcare providers in improving environmental health in their homes. Interviews with members of the child serving ecosystem will focus on opportunities for and barriers to implementing environmental health programming within existing childcare programs. Neither interview is expected to present risks to participants. Interviews will be recorded and transcribed. Transcripts will be stored as text files. Codes generated for these transcripts will be stored in tabular format. Sample quotes from each transcript will be store as text files. To protect participant information, data derived from transcripts will be de-identified prior to deposit in the publicly available repository.

B. Scientific data that will be preserved and shared, and the rationale for doing so:

All deidentified quantitative data will be preserved and shared. This includes deidentified questionnaire responses and indoor environmental quality measurements. To protect participant privacy, transcripts of the semi-structured interviews will not be shared. To accomplish the goals of the NIH data management and sharing policy, we will make available the codes generated for each transcript, the code book, and a set of sample quotes from participants that pertain to our research results.

C. Metadata, other relevant data, and associated documentation:

Metadata to be made available with the data include the original questionnaires, semi-structured interview questions, and instructions made available to participants, data dictionaries, information on where analytic code is stored (*e.g.,* GitHub) and de-identification procedures.

Element 2. Related Tools, Software and/or Code:

Qualitative data analysis will be conducted via Dedoose software, and quantitative data analysis will be conducted via R. Data analysis code will be made available via the PI's GitHub page, with a link included. Tabular data will be stored in comma separated value (.csv) format. Selected quotes pulled from transcripts will be stored in text document (.txt) format. No special software will be needed to access the published data.

Element 3. Standards

All demographic data (*e.g.,* age, race and ethnicity, educational attainment) will be collected and coded in accordance with NIH-endorsed common data elements.

No standardized interview coding scheme exists for this work. All code books and data dictionaries will be included with the metadata for all shared data sets.

Element 4. Data Preservation, Access, and Associated Timelines

A. Repository where scientific data and metadata will be archived:

Deidentified data sets will be made publicly available via the Illinois Data Bank, a data repository managed by the Research Data Service of the University Library, University of Illinois Urbana-Champaign.

B. How scientific data will be findable and identifiable:

The Illinois Data Bank provides curation and long-term access to data sets within their repository. Importantly, the Illinois Data Bank provides each data set with a unique, permanent digital object identified (DOI). This DOI and a link to the data will be included in the acknowledgements section of all manuscripts generated from this work.

C. When and how long the scientific data will be made available:

Consistent with the NIH Data Management and Sharing policy, data will be made available via the Illinois Data Bank at the time of publication in a peer-reviewed journal or at the end of the funding period, whichever comes first. The Illinois Data Bank makes data available for a minimum of five years after deposit and reserves the right to review retention after this five year period.

Element 5. Access, Distribution, or Reuse Considerations

A. Factors affecting subsequent access, distribution, or reuse of scientific data:

The deidentified data sets made available via Illinois Data Bank are openly available to anyone with internet access. All data and metadata will be freely available. Metadata will include information about the deidentification process.

B. Whether access to scientific data will be controlled:

To protect participant privacy, deidentified data sets will include limited data on the location of participants (*e.g.,* county of residence). Researchers in need of more detailed geographic data (*e.g.,* census tracts) will need to contact the PI and request a data set. Sharing of more detailed data sets will be in accordance with a Data Transfer and Use Agreement .

C. Protections for privacy, rights, and confidentiality of human research participants:

Because our work includes human subjects, all data collection protocols will be reviewed by the University's Institutional Review Board. The IRB will approve all protocols and materials prior to data collection. Any researcher with access to the data will be trained in the protection of human subjects and be included on the approved IRB protocol.

Participants in our study will provide informed consent upon enrollment in our study. Part of the consent process will inform participants about future use of their data. Participants will be informed that their deidentified data will be made publicly available. They will be advised that any identifiable information will not be shared in this publicly available dataset and will remain confidential. Because our research involves the collection of data that are place-based (*e.g.,* data collected at a specific geographical location), participants will be informed that we may share limited information (*e.g.,* their census tract) with researchers who request these data. These limited data sets will only be made available to researchers who establish a Data Transfer and Use Agreement with the University of Illinois Urbana-Champaign.

Data will be de-identified following guidance set forth by the university's Privacy Office, which consults with researchers to protect and secure information, and sets policy for identifiable data use and data management.

All data collected as part of this project will be stored in a single directory within a Box Health folder. UIUC has a Business Associate Agreement with Box.com that allows for the storage of protected health information (PHI) in secure directory. Because data will initially be identifiable (i.e., names and addresses are included in the data set), raw, deidentified data will be stored in a password-protected subfolder and access will be restricted to those with approval (*e.g.,* the PI and select project members who are involved with the deidentification of data). Identifiable data and any associated keys will be kept in a separate subdirectory from deidentified data sets. Additionally, "raw" data sets will be stored in separate subdirectories from "clean" or processed data sets.

Element 6. Oversight of Data Management and Sharing:

Compliance with this data management and sharing plan, including all oversight of data collection, management, and storage, will be the responsibility of the project PI (Dr. Sheena Martenies). Data curation and long-term storage will be managed by the Illinois Data Bank.