RESOURCE SHARING AND DATA SHARING PLAN

WHAT DATA WILL BE SHARED:

The PIs will openly share all PLAY data and metadata on the NICHD/NSF-funded Databrary video-sharing library. Permission to share potentially identifiable data (video recordings, birthdate + exact age, Census Block Group) will be obtained using the Databrary release template, which asks participants to grant permission to share their identifiable video data and other data with authorized researchers on Databrary for research and educational purposes. We will share 900 sessions of video data, questionnaire data, non-video data, and metadata from study participants, plus additional videos and metadata from sessions that did not meet stringent quality assurance guidelines but which still have research value.

Primary data include video recordings, time-locked video-coding files, questionnaires, demographic and health data, and a variety of other measures.

Video: One-hour video recordings of children and their mothers during natural everyday activity (play) in their homes; short video segments of children in solitary and dyadic structured play; video home tours accompanied with narration about how rooms are used; and videos of infants' clothing and footgear.

Coding files: Datavyu (datavyu.org) format coding files transcribed for infant and mother speech and scored (with inter-observer reliability on 25% of each session) for foundational communicative acts, gestures, object interactions, locomotion, and emotion during the one-hour video of natural play.

Questionnaires: Selected items from the ECLS-B; parent-report data about infant locomotor milestones and injuries from falling; infant receptive and productive vocabulary (MacArthur-Bates Child Development Inventory, MCDI); infant temperament (very short form of the Early Childhood Behavior Questionnaire, ECBQ); gender labels and socialization; family media time use; and the home environment.

Demographic data: Infants' birth date, due date, birth weight, gender, race, ethnicity, sleep characteristics, breastfeeding, child care arrangements, and health history; family health history, parents' birth dates, race, ethnicity, country of birth, education, employment status, smoking history and status; family structure; home language(s); and the Census Block Group codes for the family home.

Other measures: Measurements of ambient sound levels recorded during the data collection session; laser scan of room dimensions coordinated with the video home tour.

Videos will be transcoded by Databrary into a standard format (currently H.264+AAC in MP4). Some infant and parent demographic data will be stored in Databrary's session-level spreadsheet interface to facilitate searching for and filtering by participant characteristics. The spreadsheet data can be downloaded in the CSV format. All other data will be stored in individual participant-specific session folders in CSV format files, with separate CSV files for ambient sound levels, demographics and health information, MCDI, ECBQ, media time use, and the home environment. We may also store aggregate, within measure but across study, data files in CSV formats to facilitate statistical analysis across platforms.

In addition, a full *web-based protocol* containing details about all study procedures, apparatus, and code definitions will be shared via Databrary and a study-specific *wiki* that has exemplar photos and video clips to augment text-based descriptions.

To facilitate interoperability with other data repositories, we will convert Datavyu files into the CHAT format suitable for sharing with the TalkBank/CHILDES and HomeBank research community, and we will make the videos and CHAT format transcripts available to the child language community via TalkBank/CHILDES and HomeBank. We will make all MCDI vocabulary assessment and participant-level metadata available to the WordBank repository, and we will explore ways to make it possible for researchers using the PLAY project corpus to compare PLAY project data with WordBank-derived norms.

Databrary: We expect that there will be enhancements made to Databrary during the project period that facilitate project activities. Access to Databrary is free, and the Databrary software is freely and openly available to the research community (github.com/databrary) throughout the project period and beyond.

Datavyu: The Datavyu video coding tool, a free, open source (github.com/databrary/datavyu) software tool will undergo revisions and enhancements during the project period that make it even more useful for project-related coding. The tool is and will continue to be freely available to the research community throughout the project period and beyond.

WHO WILL HAVE ACCESS TO THE DATA:

Access to data files will be restricted to Databrary-authorized researchers to protect confidentiality and uphold participants' permission to share. Any researcher who qualifies and whose institution signs the Databrary access agreement (a specialized data use agreement between NYU and a researcher's institution) will be granted access to all PLAY data, including identifiable video recordings and demographic variables considered personally identifying (e.g., any two of birth date, test date or exact age or the Census tract or Census block group codes). Members of the PLAY launch group are authorized on Databrary.

Many data and metadata components that do not contain identifiable information will be shared with the public. These components include the complete protocol wiki and description (minus the identifiable video segments), the analytic reproducibility and transparency guidelines, a subset of the participant-specific metadata, and other materials.

WHERE WILL THE DATA BE AVAILABLE:

All data will be stored and shared via Databrary (databrary.org), a NICHD/NSF-funded data repository whose data access policies and procedures are consistent with NIH and NSF data sharing policies. Databrary requires that researchers secure explicit permission from participants or their parents or guardians to allow identifiable data elements, including video recordings, to be shared with other researchers. Databrary's policies further restrict the use of shared data to scientific, educational, and informational purposes.

WHEN WILL THE DATA BE SHARED:

The draft wiki-based research protocol and code definitions with illustrative video clips are already shared with the public, but only researchers with authorization to access Databrary can view the linked one-hour videos of natural play. Videos and associated data and metadata elements collected as part of the study will be shared with all launch group members in Year 4, after data collection is complete, the files have passed quality assurance, and the transcription and foundational coding passes have been completed. That way all members of the launch group will gain access to the entire dataset of 900 coded sessions at the same time. Videos and associated data and metadata elements will be openly shared with all other authorized Databrary researchers (and when possible, with the public) at the end of the 5-year project period.

HOW WILL RESEARCHERS LOCATE AND ACCESS THE DATA:

Any researcher affiliated with an institution may register for and seek authorization for full access to Databrary. All researchers granted access to Databrary agree to cite materials they use in any publication or presentation they produce. Databrary will provide a standard citation format and a persistent identifier (DOI) for the PLAY project dataset.

WHO IS RESPONSIBLE FOR MANAGING AND SHARING DATA:

The PIs are responsible for establishing data management and sharing practices for the project as a whole. All launch group labs will receive training from project staff in the use of Databrary and Datavyu for data management. Desktop computers with encrypted hard drives will be provided to all launch group members.

Each data collection site will upload video and other data files to Databrary after returning to the laboratory after a session ends. Data collection site staff will also enter session-, family-, and participant-specific metadata into the Databrary spreadsheet. PLAY project staff will monitor data quality and conduct quality assurance reviews. Each data-coding site will download videos and template Datavyu coding files to project-supplied computers for conducting coding passes. The coding files will be uploaded to Databrary for reliability analyses. PLAY project staff will carry out reliability coding, and when a coding pass meets reliability criteria, the session will be flagged as ready to share.

At a time determined by the PIs in mid-Year 4, all data meeting quality assurance and reliability criteria will be released to the launch group for analysis. At the end of the project period, the PIs will share all data with the research community.