

# HomeBank: A resource for sharing daylong audio recordings from children and families



Gina M. Pretzer, Anne S. Warlaumont, Mark VanDam, Brian MacWhinney, Christopher J. Reding

## What is HomeBank?

- HomeBank (<http://www.homebank.talkbank.org>) is an NSF-supported project that maintains an online repository which provides a formal system for archiving and sharing daylong audio recordings, and associated transcripts and metadata (see VanDam et al., 2016)
- GitHub repository (<http://www.github.com/HomeBankCode>) hosts computer code for LENA analysis tools written by contributing researchers, as well as documentation for selected scripts.
- 26 members (plus associated RAs and graduate students) and 5 contributors from 4 different countries
- HomeBank is supported by an NSF SBE RIDIR Grants 1539129, 1539133, and 1539010

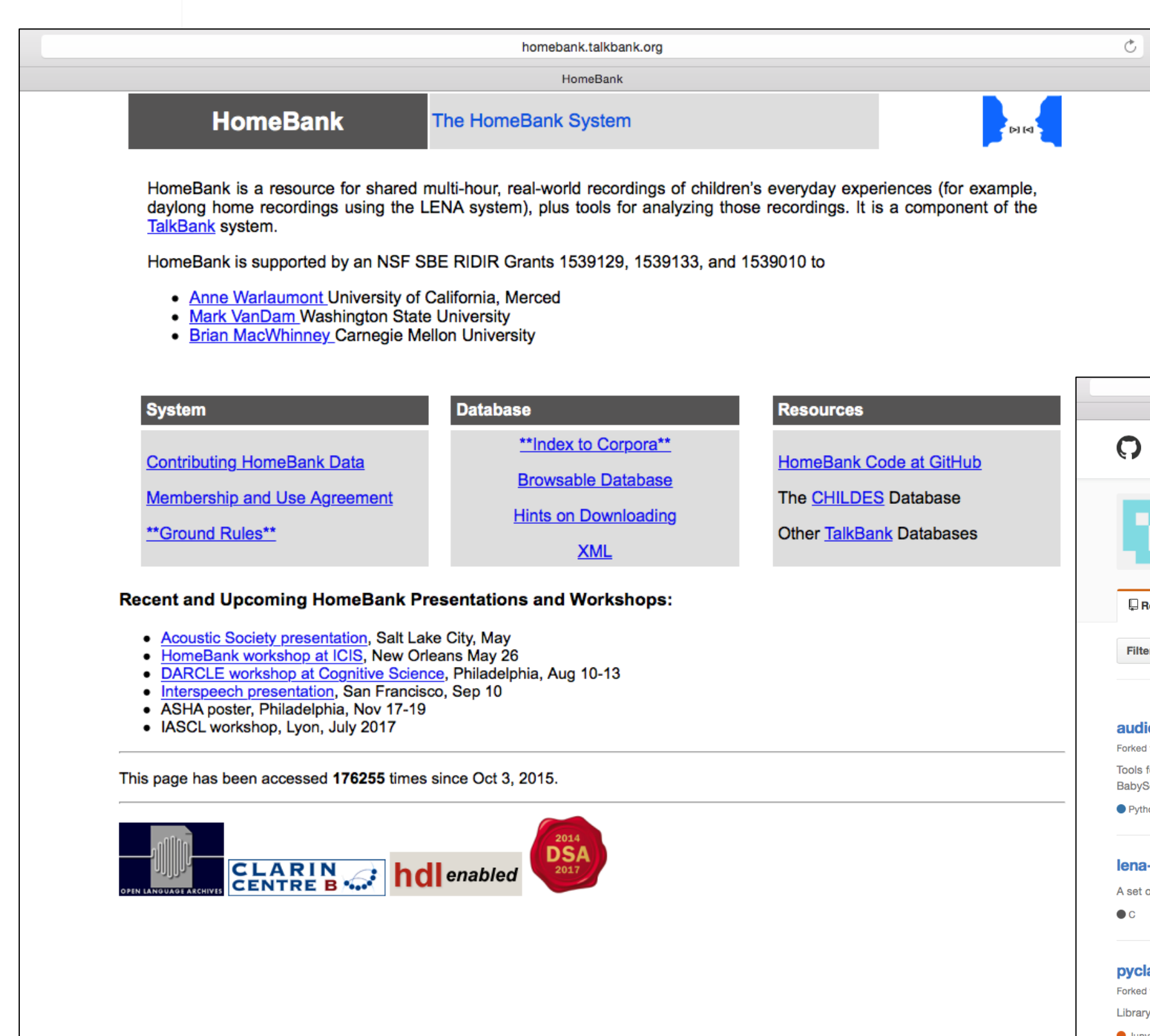
## Online Repository of Daylong Audio Recordings

- HomeBank's online repository is comprised of two levels: Public and Private
- The repository currently holds 8 corpora – 1 Public, 7 Private
- The corpora include audio recordings, related transcripts (in CHAT format), and metadata (e.g., demographic data, SES, family status, disorder, etc.)
- Built in CLAN window to analyze transcripts

- Public Repository
  - Anyone can access the data in this repository
  - All vetted materials – all sensitive and potentially identifying information has been removed
    - Last names, SSN, address, date of birth, etc.
- Private Repository
  - Requires membership for access – users must agree to HomeBank's User Agreement
  - Raw data – sensitive or identifying information may remain

Corpus	Recordings	Description	Contributor
Bergelson Seedings Transcripts Audio	49	Study of exposure to object words	Elika Bergelson
McDuff Transcripts Audio	25	Children born to adolescent and adult mothers	Karmen McDuff
VanDam Public Transcripts Audio	53	159 5-minute transcribed segments from 53 daylong recordings	Mark VanDam
VanDam1 Transcripts Audio	9	Preschoolers with and without hearing loss	Mark VanDam
VanDam2 Transcripts Audio	74	Description not yet ready	Mark VanDam
VanDam3 Transcripts Audio	788	Description not yet ready	Mark VanDam
Warlaumont Transcripts Audio	18	Longitudinal study of children learning English and Spanish	Anne Warlaumont
Whinnery Transcripts Audio	8	Comparison across different child care settings	Melanie Soderstrom

- Access to these data is especially helpful for researchers who are unable to access specific populations (e.g., age, clinical diagnoses, etc.) or who do not have the financial or other resources to collect their own recordings



Pictured Above: HomeBank's website home page

Pictured Right: HomeBank's GitHub Repository

## Interested in gaining access to HomeBank data or contributing data of your own?

Send an email to [4homebank@gmail.com](mailto:4homebank@gmail.com) or visit [www.homebank.talkbank.org](http://www.homebank.talkbank.org) for more information

## GitHub Repository

- Researchers from several labs have contributed original codes for LENA analysis tools.
- Included are scripts for excising audio segments from LENA recordings, analyzing the LENA .its files, and anonymizing .its files written in Python, Matlab, Perl, and Shell.
- Documentation for selected scripts containing step-by-step instructions and pictures detailing how to use these tools

- Researchers can create output files with specific information about the automatically labeled vocalizations, such as frequencies of vocalizations by different speakers, utterance onset/offset times, and frequencies of child-adult and adult-child exchanges
- Some scripts use programs such as PRAAT to glean acoustical information (e.g., fundamental frequency, intensity, etc.) for vocalizations within recordings

