The data provided here is a subset of the data used in the IARPA MATERIAL Program Evaluation[1]. This subset includes the conversational speech (CS) portion of two languages that were tested Somalia (1S) and Bulgarian (2S) and comes in several packages with the following labels:

- BUILD - includes CS audio source files and transcripts for ASR training and development
- ANALYSIS - includes CS audio source files, transcripts, and translations for error analysis in support of MATERIAL's Cross-Lingual Information Retrieval and Summary (CLIR+S) task
- DEV - includes CS audio source files for development in support of CLIR+S task
- EVAL - includes CS audio source files for evaluation in support of CLIR+S task
- QUERY1 - includes English queries used for development
- QUERY2 - includes English queries used for evaluation
- ANALYSIS_ANNOTATION - contains relevance annotations of QUERY1 and QUERY2 that have analysis documents as relevant
- DEV_ANNOTATION - contains relevance annotations of QUERY1 and QUERY2 that have dev documents as relevant
- EVAL_ANNOTATION - contains relevance annotations of QUERY1 and QUERY2 that have eval documents as relevant

An overview of the query types is described in
https://www.nist.gov/system/files/documents/2018/07/12/openclirqueriesandrelevance.pdf

The query syntax is described in
https://www.nist.gov/system/files/documents/2018/07/12/openclirqueryspecification.pdf

The guidelines on how query relevance judgment was made is described in
https://www.nist.gov/system/files/documents/2019/02/13/qrj_guidelines.openclir_version.pdf

[1] https://www.nist.gov/iarpa-material-machine-translation-english-retrieval-information-any-language-program