**Cases** -

| **CaseId** | **CaseName** | **Keywords Loaded** | **Keywords** |
| --- | --- | --- | --- |
| Primary Key (Case UID) | Unique | Options - yes/no  Defaults to no. Can be ‘no’ even if ‘keywords’ column has values as there can be a new upload / recording for a case, and hence we’d recalculate the keywords. Can be ‘yes’ for an empty keywords column if there are no commonalities across witness interviews | List of all keywords obtained for the case. Eg - [‘green shirt’, ‘yellow car’] |

**Recordings -**

| **RecordingId** | **RecordingName** | **CaseId** | **RecordingLink** | **Transcript Link** | **Descriptive phrases** |
| --- | --- | --- | --- | --- | --- |
| Primary Key  (Recording UID) | Unique | Foreign Key to Cases table | Perhaps a Google drive link to the recording | Perhaps a Google drive link to the transcription | If first taking out noun-adj pairs from each case, we can store them as a list here. Then we only need to run the clustering of these phrases if a new recording is uploaded. |