

ER Diagram

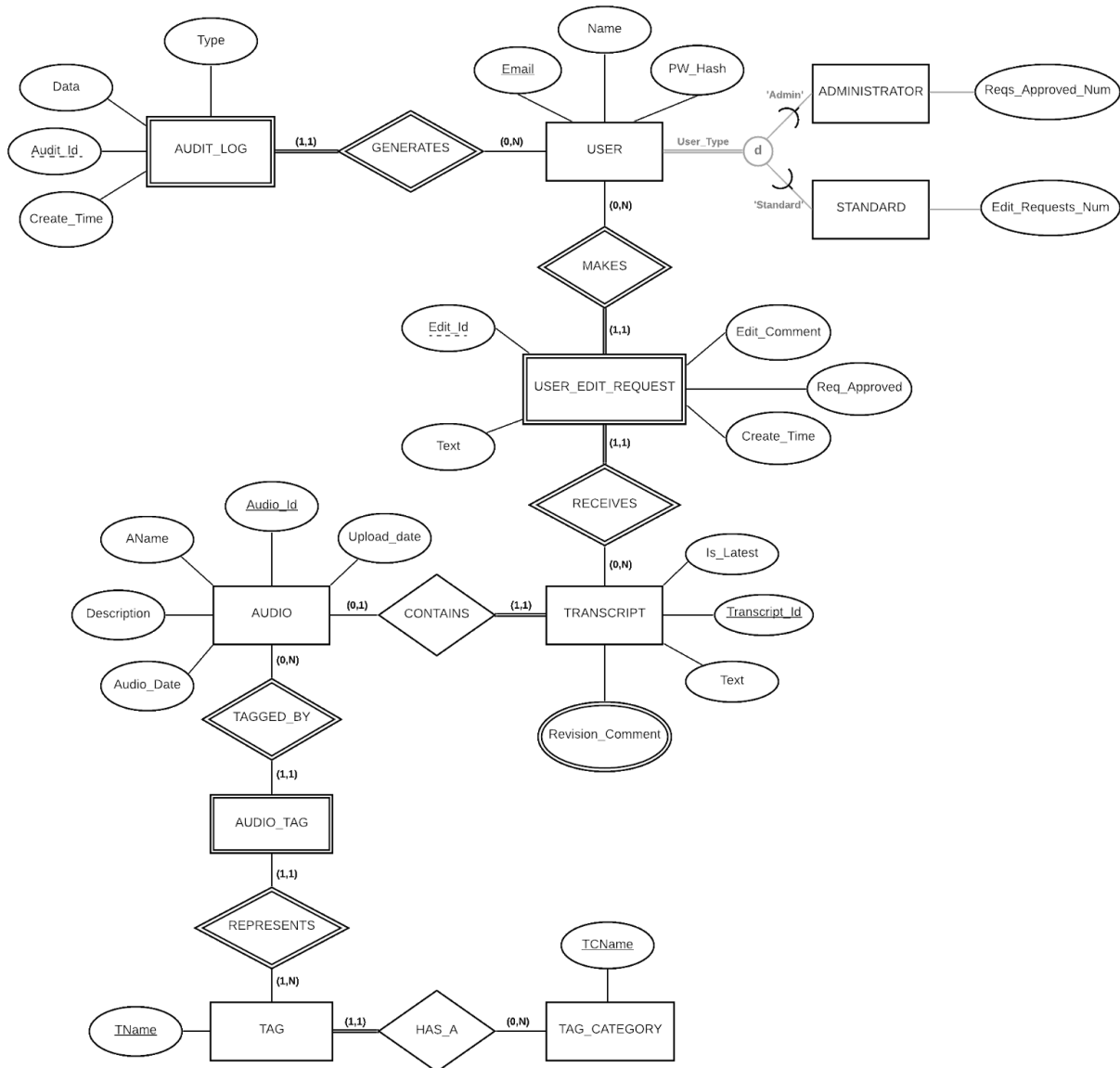


Figure 1. ER Diagram

Relational Schema

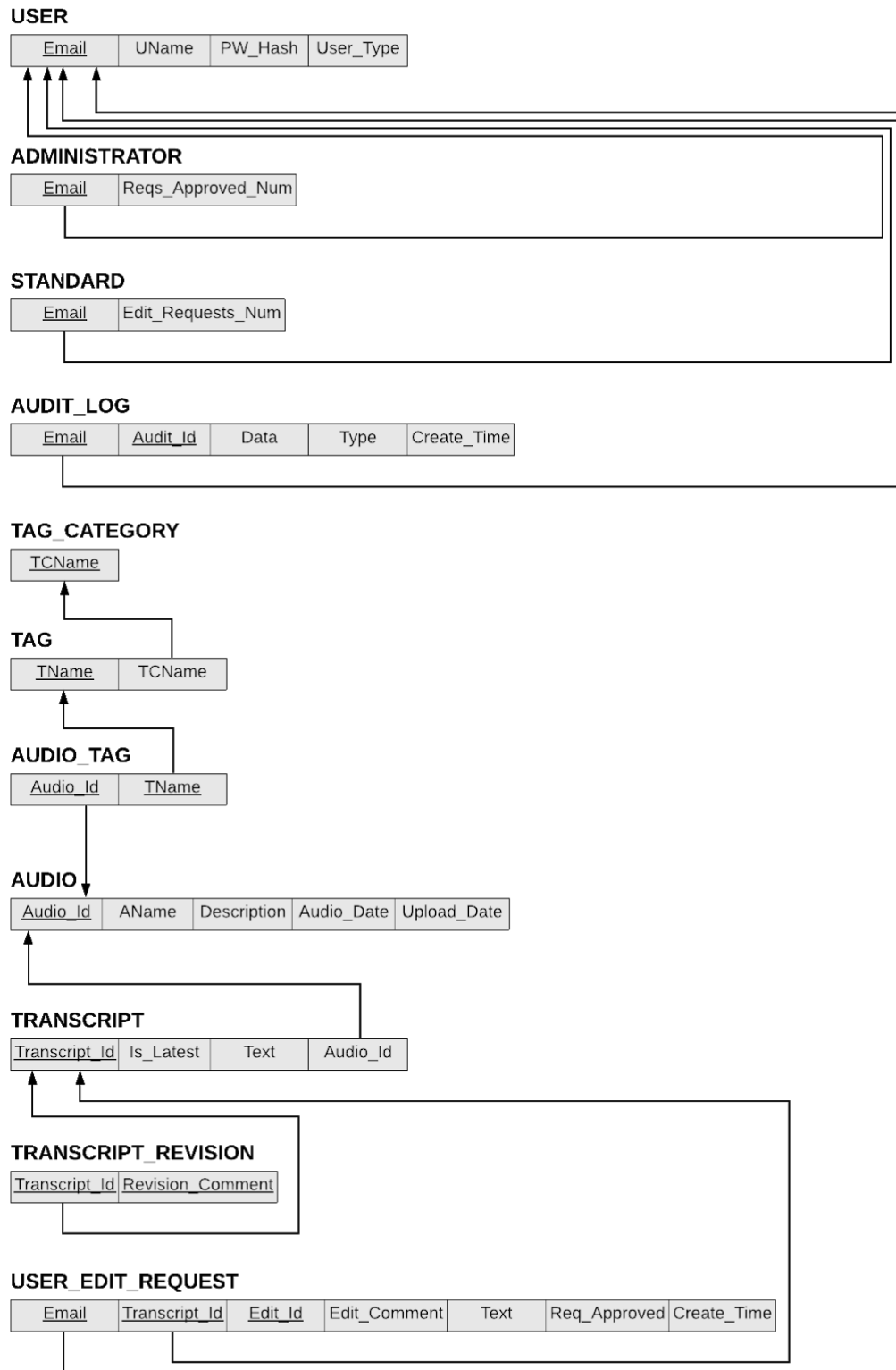


Figure 2. Relational Schema

Database Estimations

- Initial Database Size
 - For the current moment, there will be at most 74 records (for the 74 audio files currently loaded on <https://archive.org/details/trentoniana>).
- Types and Average Number of Searches
 - 74 metadata entries (audio files)
 - On average 3 tags per metadata entry
 - On average 5 tag categories
 - On average 4 tags per category
 - On average $\frac{1}{2}$ of metadata entries will have a transcript ($74/2 = 37$)
 - $(4 * 5) + (3 * 74) + 74 + 37 = 353$ *searches*

Tables

Users:

- int id: primary key (identity)
- varchar name
- varchar email (unique)
- varchar password_hash
- int user_type = 0 (user)

Audio:

- int id: primary key (identity)
- varchar name
- (long string type) description (nullable)
- date audio_date (nullable)
- date upload_date

Transcripts:

- int id: primary key (identity)
- int audio_id (Audio: id)
- (long text) text
- text revision_comment (nullable)
- bool is_latest = true

TagCategory:

- int id: primary key (identity)
- varchar name (unique)

Tags:

- int id: primary key (identity)
- int category_id (TagCategory: id)
- varchar name (unique)

AudioTag (+ unique key with both columns):

- int id: primary key (identity) (**remove?**)
- int audio_id (Audio: id)
- int tag_id (Tags: id)

UserEditRequest:

- int id: primary key (identity)
- int user_id (Users: id)
- (long text) text
- int transcript_id (Transcripts: id)
- varchar edit_comment (nullable)
- bool request_approved = false
- timestamp create_time

AuditLogs:

- int id: primary key (identity)
- int user_id (Users: id)
- varchar type (type of log: e.g. approval of edit request, ban user(?), etc.)
- jsonb data
- timestamp create_time

Entity Types

Strong:

- User, Audio, Transcript, TagCategory, Tags

Weak:

- AudioTag, UserEditRequest, AuditLogs

Relationships

Transcripts → Audio (audio_id)

- 1:1
 - one transcript can have at most one audio file associated
 - an audio file can have at most one transcript associated with it

Tags → TagsCategory (category_id)

- N:1
 - a tag category can apply to any number of tags
 - a tag can have at most one tag category

UserEditRequest → User (user_id)

- N:1
 - a user can make any number of user edit requests
 - a user edit request is only made by a single user

UserEditRequest → Transcript (transcript_id)

- N:1
 - a transcript can have any number of user edit requests
 - a user edit request can only be applied to one transcript at a time

AudioTag → Audio (audio_id)

- 1:N
 - a single audio file can have many audio tags
 - an audio file tag is unique to the audio file associated with it

AudioTag → Tags (tag_id)

- 1:N
 - an audio tag has a single tag associated with it
 - a tag can be applied to many different audio file tags

AuditLogs → User (user_id)

- 1:N
 - a user could have any number of audit logs associated with them
 - an audit log is usually associated with a single user