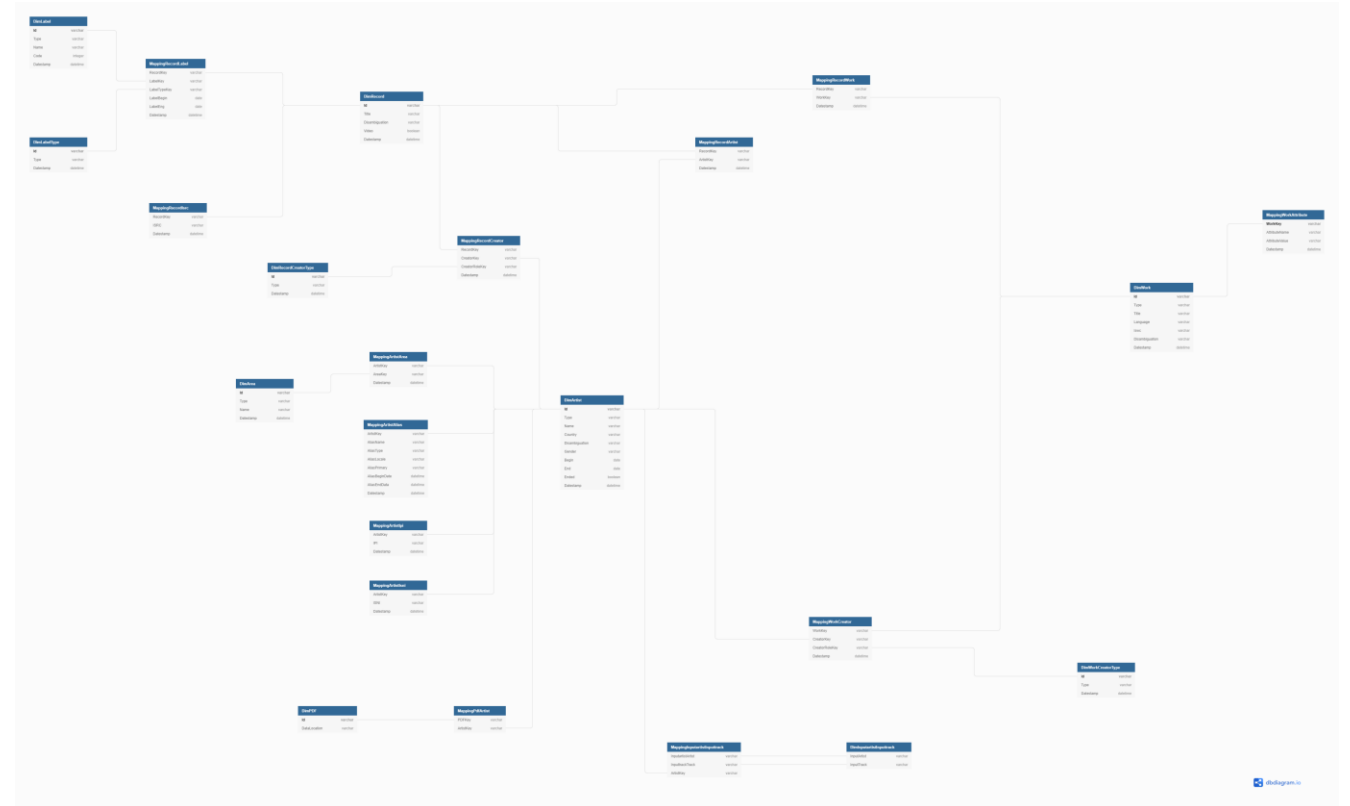# Whole Database Schema
Intro

I have designed a simple schema for structuring the data involved in the task, and have placed it in the same folder as this presentation.

This schema contains the input artist track spreadsheet provided in exercise 3, the corresponding artist, record and work tables, as well as a link to the pdf data obtained in exercise 1.

In my mind, most of the data should be stored in a relational database, such as sql. I'd advocate for this being on the cloud, so using tools such as RDS and Athena (AWS) or Azure SQL Database (Azure).

The exception with this is the information related to each royalty pdf, which I believe should be stored in a document storage solution, such as S3.


Jack Ballinger

Whole Database Schema
Main Tables

In the image to the right, I've highlighted the main tables that make up the database, obtained during exercise 3

Each of the highlighted sections contains one main table, namely DimRecord, DimWork and DimArtist, as well as a number of supplementary tables and mapping tables connecting elements to these main tables.

As mentioned on the first slide, I see these being stored in a relational database, ensuring quick and easy interrogation of the data.

Jack Ballinger
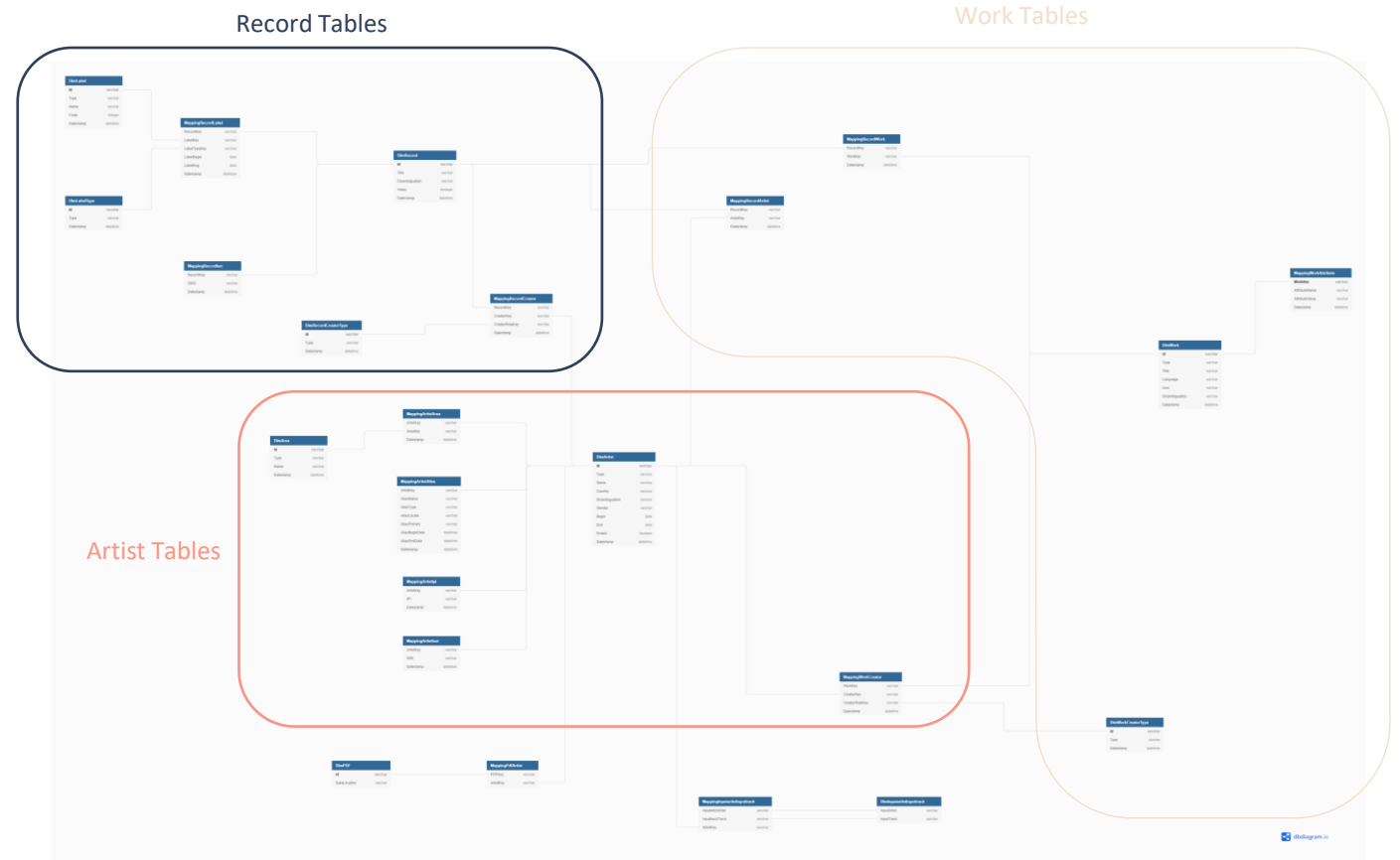
Whole Database Schema:
Linking Tables

PDF Mapping

Billboard Mapping

In the image to the left, I've highlighted the linking tables contained in the database

The Billboard Mapping tables contain the artists to match input, as well as the mapping to the musicbrainz artist entity.

The PDF Mapping tables contain the links to the storage location of the respective PDF data tables, as well as the mapping of the artists related to the royalty statements to the PDFs.

These links will be links to document storage locations, where the music royalty data should live.

Jack Ballinger