MetaBrainz databases

Alastair Porter (MetaBrainz, MTG-UPF)
CIRMMT LinkedMusic Workshop
18 November 2022

Overview

❖ The MetaBrainz Foundation

❖ MusicBrainz history and structure

❖ CritiqueBrainz

❖ AcousticBrainz

❖ ListenBrainz

❖ BookBrainz

❖ CoverArtArchive

A brief introduction

❖ Alastair Porter

❖ Researcher at Music Technology Group, Universitat Pompeu Fabra, Barcelona

❖ Freelance contractor with MetaBrainz foundation

❖ MA Music Technology (McGill) 2012

❖ Working with music databases for 15+ years

MetaBrainz foundation
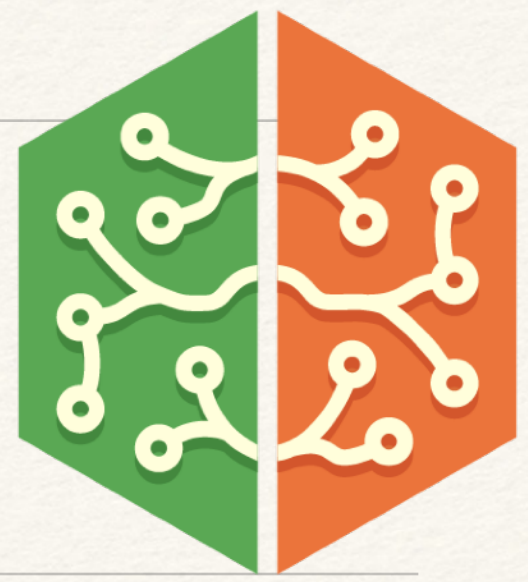
❖ California non-profit, founded in 2004

❖ Umbrella organisation for all of the projects discussed here

❖ Paid team of 12, funded by license fees to the MusicBrainz database

koonisutra photographs
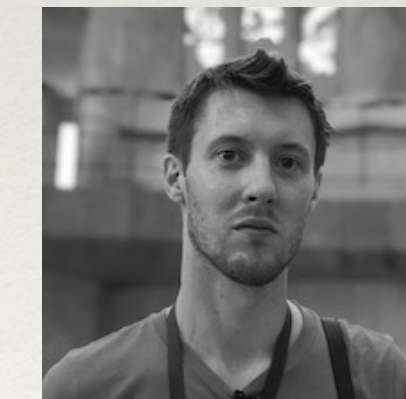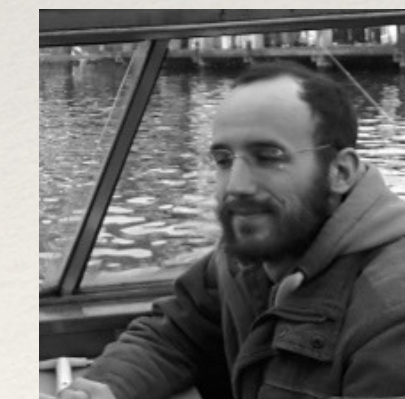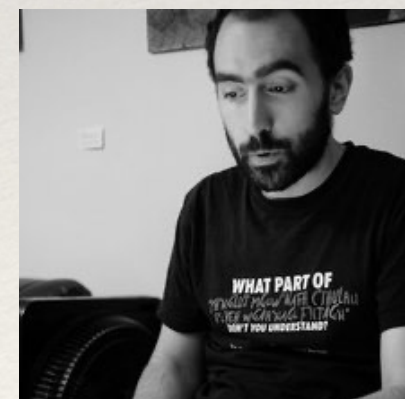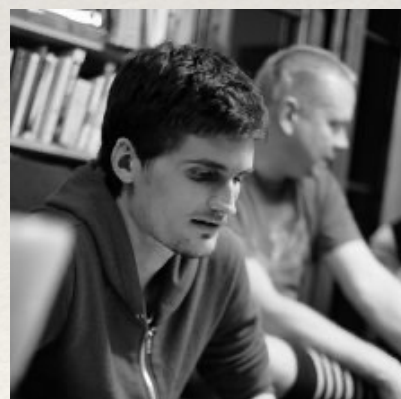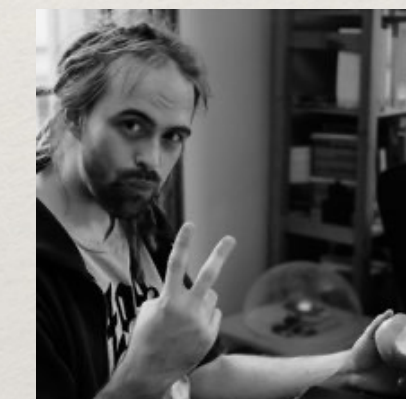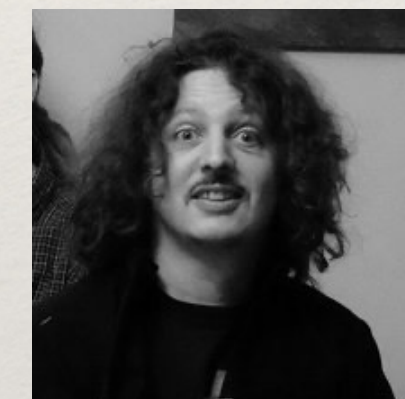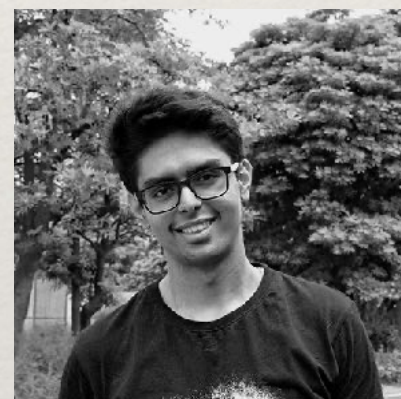
CDDB

- ❖ Do you remember CDs?

- ❖ Put a CD into your computer and get the metadata!

  - ❖ No CDText

  - ❖ MP3s were new!

  - ❖ Artist, Album, Song

- ❖ Open database, shared to the public

https://en.wikipedia.org/wiki/CDDB

The life and death of CDDB

❖ The originally open CDDB Database gets sold

  ❖ Company gets named Gracenote

  ❖ Eventually bought by Sony and then bounced around a few other companies

❖ Database gets closed and made private, locked behind a license agreement

MusicBrainz starts
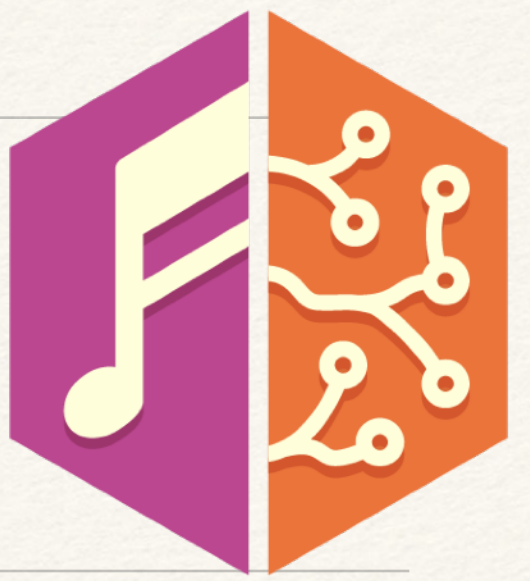
❖ Open source hackers bristle at the prospect of their open database being made closed

❖ So they make their own

❖ freedb, gnudb, MusicBrainz (2000)…

❖ https://www.eff.org/deeplinks/2021/06/organizing-public-interest-musicbrainz

Original MusicBrainz Schema

❖ Originally a copy of the CDDB structure: Artist, Album, Track

❖ Give each entity a MBID (MusicBrainz ID), a random UUID

❖ Web interface to add data

❖ API to retrieve data

Shortcomings of the schema

❖ The same recording on different albums is in the database twice

❖ Multi-CD albums get entered as different albums

❖ The "Classical Music Problem"
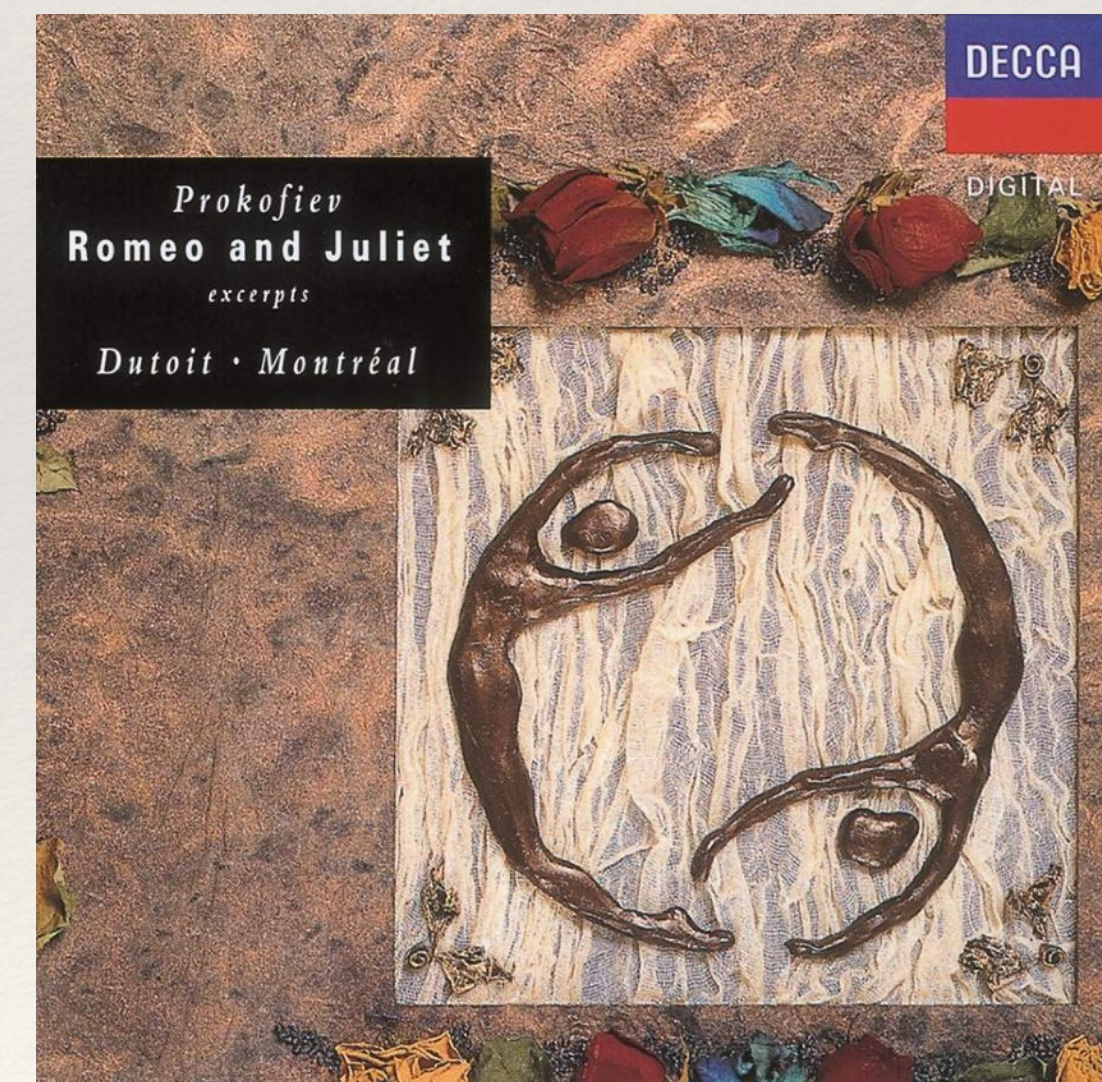
The Next Generation Schema
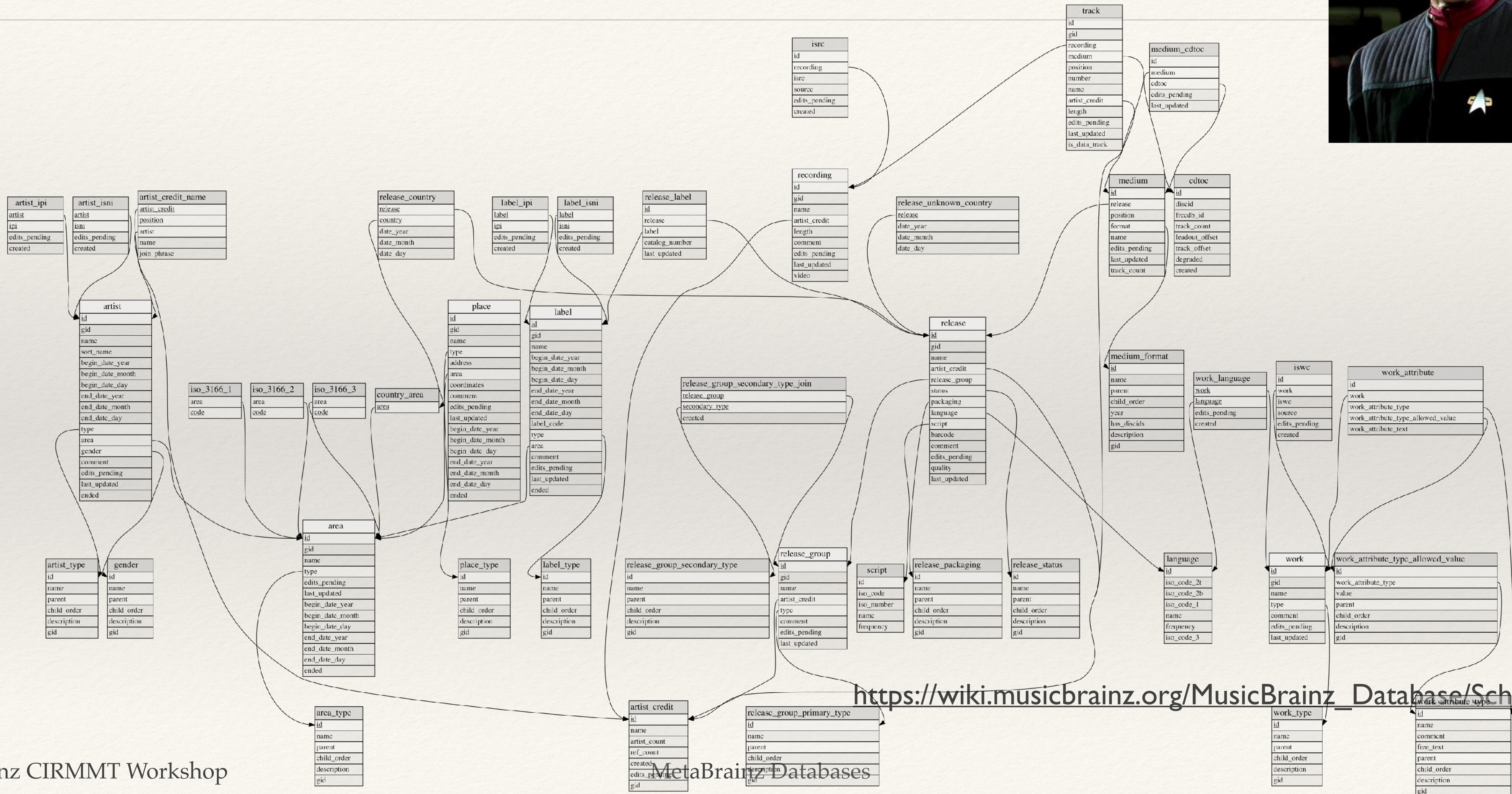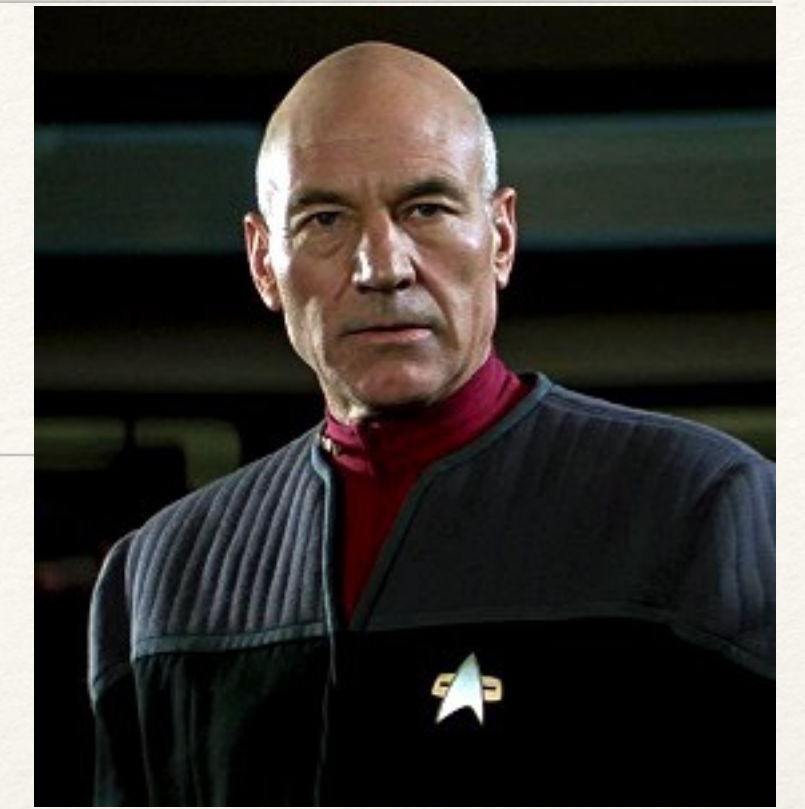
https://wiki.musicbrainz.org/MusicBrainz_Database/Schema

Improvements

❖ Recordings can be shared between multiple albums

❖ Artist Credits: More than one person on an album, or with a different name as a special case

❖ Relationships: more detailed information between recordings and artists (and other types of entities)

https://musicbrainz.org/recording/f07a6189-6ae1-4440-8ca9-10cb7f69f6d8

https://musicbrainz.org/recording/32c7e292-14f1-4080-bddf-ef852e0a4c59

Web interface to add/edit data

MusicBrainz learning points

❖ Use UUIDs for everything (<u>cool URIs don't change</u>)

❖ Be able to retrieve all data via APIs (XML, JSON, JSON-LD)

❖ Always use MBIDs when referring to items in the API response

    ❖ This allows us to make changes/restructure the database while keeping the public IDs stable (*Andrew: We reserve the right to change the structure of our DB)

❖ Always keep the data open and freely available

❖ The Community is key to the project (we have a team of 12 going on 20,000)

CritiqueBrainz
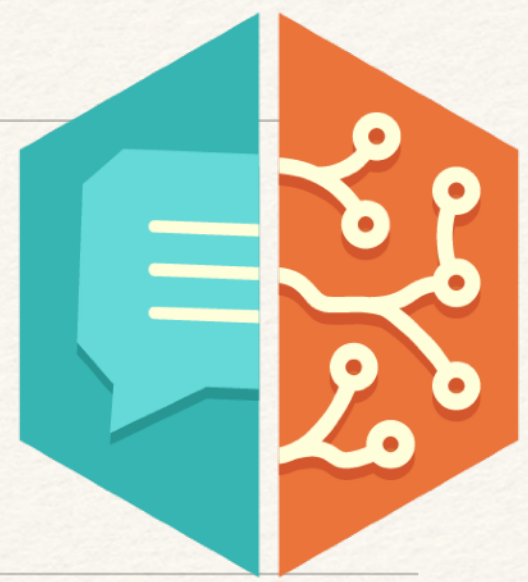
❖ Wikipedia rules say that content has to be from a Neutral point of View

❖ So, how do you write about a new album without being a music critic?

❖ CC licensed, write whatever you want

❖ Initially only albums, but now all "core" MB entities

❖ Refer to items in MusicBrainz, use MBIDs

❖ API to read (and write)

❖ Interface to write reviews directly from MusicBrainz, ListenBrainz, BookBrainz
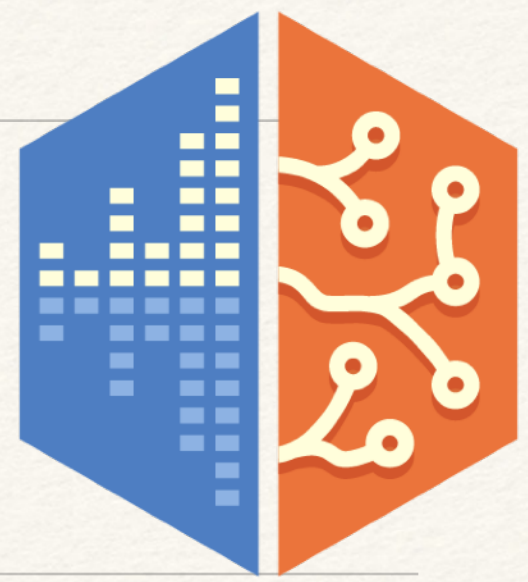
❖ 10,500 reviews (~400-500/year)

AcousticBrainz

- ❖ The Echo Nest (2005) builds a huge database of music metadata, semantic graph, and automatically extracted features (MFCCs, beats, danceability, popularity)

- ❖ 2014: Spotify buys the Echo Nest

- ❖ Open source hackers bristle at the prospect of a database being closed down

  - ❖ Rob comes to my office and says "hey, so..."

  - ❖ Prototype built in an afternoon

AcousticBrainz

❖ Collaboration with UPF

❖ Uses Essentia feature extractor (https://essentia.upf.edu/)

❖ Audio descriptors (MFCCs), Tonal features (key, chords), Rhythmic Features (bpm)

❖ Support Vector Machines to compute features using common MIR tasks

  ❖ e.g. Using MFCC, rhythmic summaries or other computed features

  ❖ Genre, Mood, Instrumentation

AcousticBrainz

- Some problems with the data

    - Uncertainty about the accuracy of some features (30m submissions, which ones are correct?)

    - Lower quality ML models (Deep Learning got big about the time we released, summary MFCCs became less useful)

    - We hoped that we would encourage developement and collaboration, but it didn't eventuate

- Shut down at the beginning of 2022

- Data still available for download

ListenBrainz

❖ last.fm (2002) pioneered a way to make a record of the songs you are listening to, and a community of users passionate about discussiing their music

❖ In 2007 it was bought by CBS

❖ Open source hackers bristle at the prospect of their database disappearing

❖ Over a weekend in September 2015 a bunch of us get together and built a prototype

Core aspects of ListenBrainz

❖ A "listen", your record of listening to a song

❖ Submit to the API

❖ Give us a Recording MusicBrainz ID if you have one

❖ If you don't, we generate a "MessyBrainz" id, based on the metadata you send

   ❖ And then match it to a MusicBrianz Recording

❖ Import from Spotify, or send data from a music player on your computer

❖ Incoming stream of songs
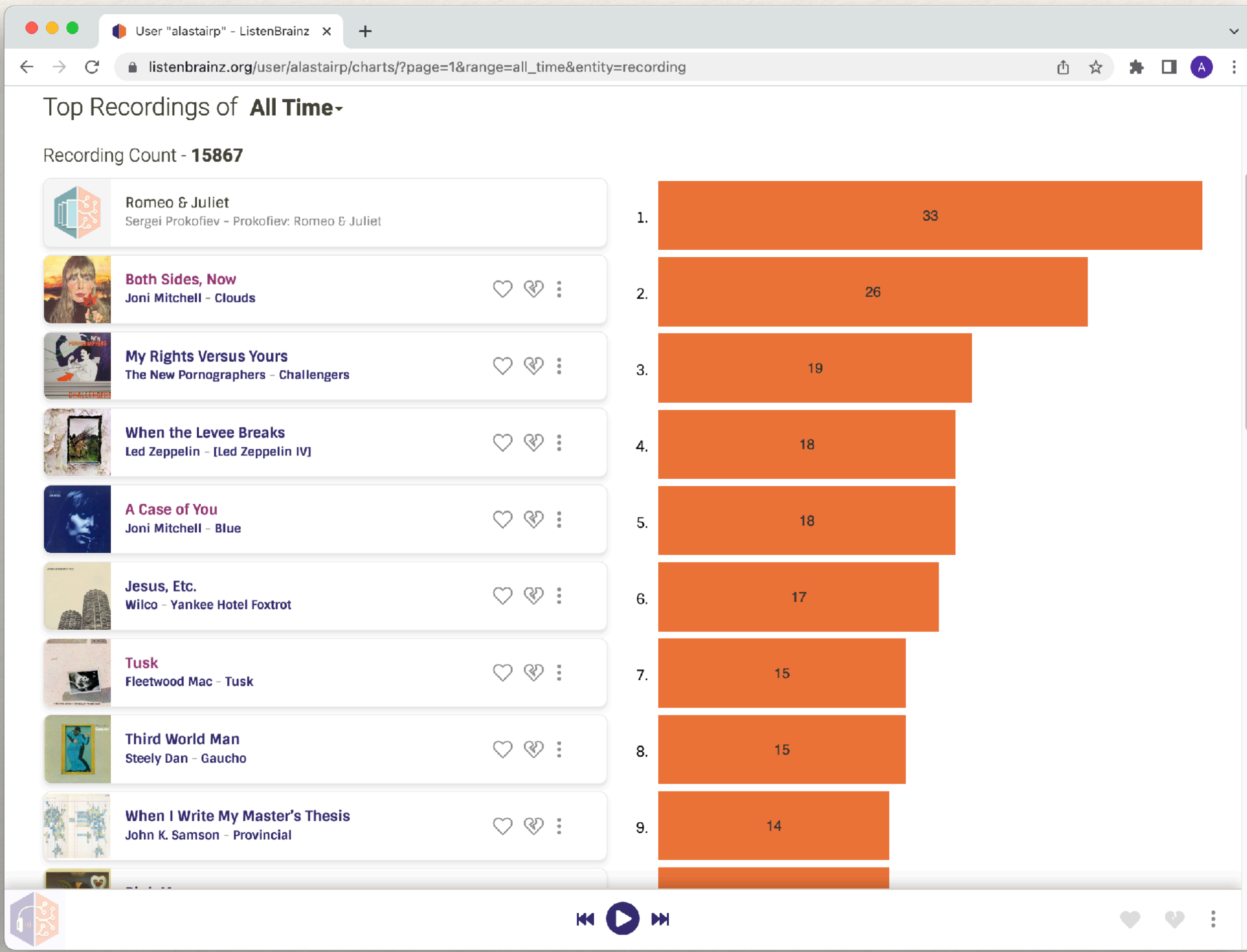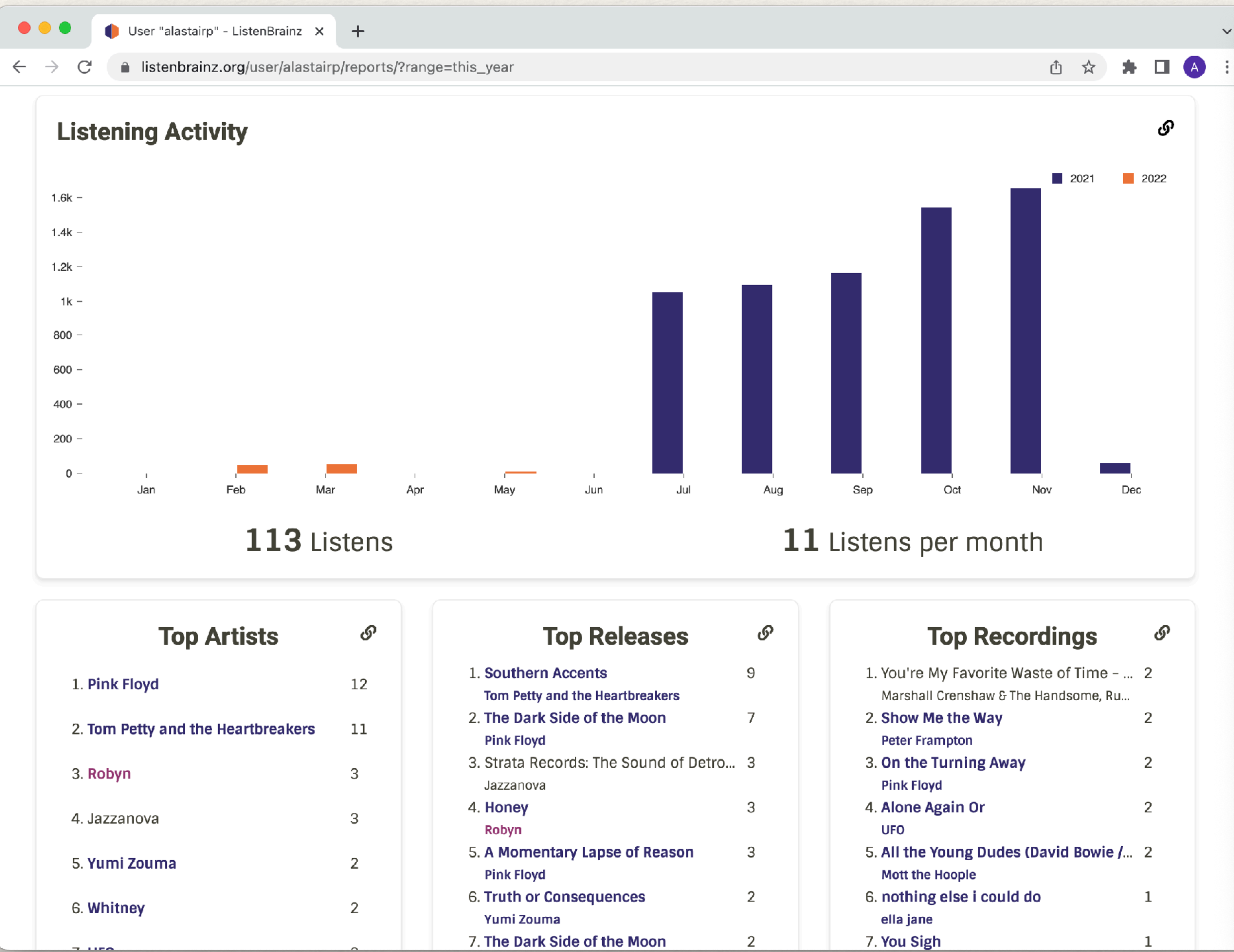
- Statistics
  - What you've been listening to
  - When
  - How much
  - Where it's from

- 600,000,000 total submissions
- 400,000 per day (mostly from Spotify)
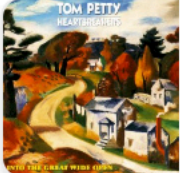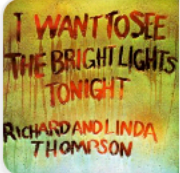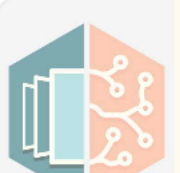
Recommendations
- ### Collaborative filtering based on listening history in LB
- ### Make a Playlist
- ### Listen on LB or export to Spotify
- ### API to create playlists, others can contribute

BookBrainz

❖ What if you had MusicBrainz, but for books instead?

❖ Originally a community-driven project, now an official part of MetaBrainz

❖ Deals with entities such as authors, works, editions, series, publisher

❖ Like MusicBrainz, all entities have a BBID

❖ Discogs used to have a partner project *Bookogs*, but it was closed at the end of 2020

    ❖ Much of the community moved to BookBrainz

❖ Write reviews about these things on CritiqueBrainz!

❖ Work: 37220, Author: 14635, Edition: 7286

CAA

- ❖ https://coverartarchive.org/
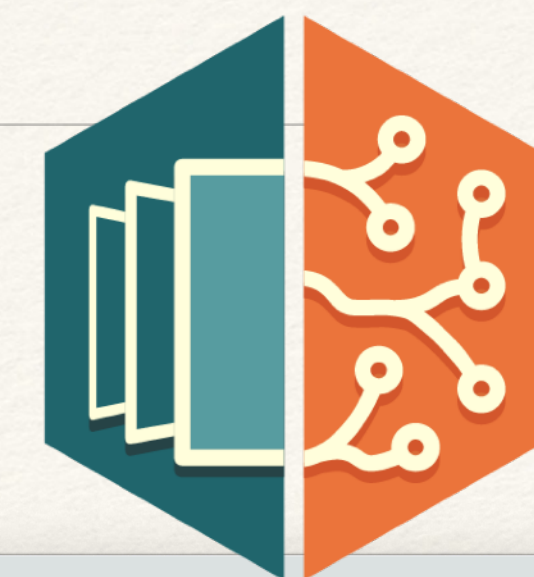
- ❖ Store images at the Internet Archive

- ❖ Direct integration to MusicBrainz to add images

- ❖ API gives direct URLs to download

Thanks

https://metabrainz.org
https://musicbrainz.org
https://twitter.com/musicbrainz