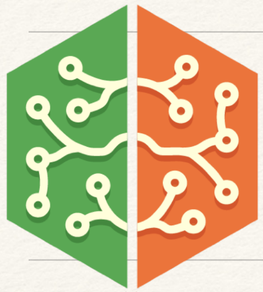
MetaBrainz Database Schemas

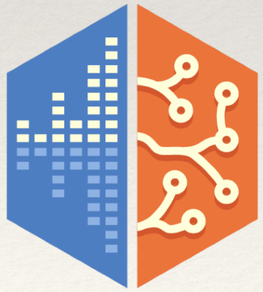
Alastair Porter (MetaBrainz, MTG-UPF)
LinkedMusic Project Meeting
21 October 2023

MetaBrainz projects

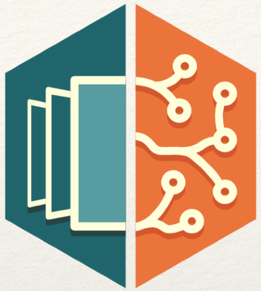


MusicBrainz

CritiqueBrainz



AcousticBrainz



Cover Art Archive

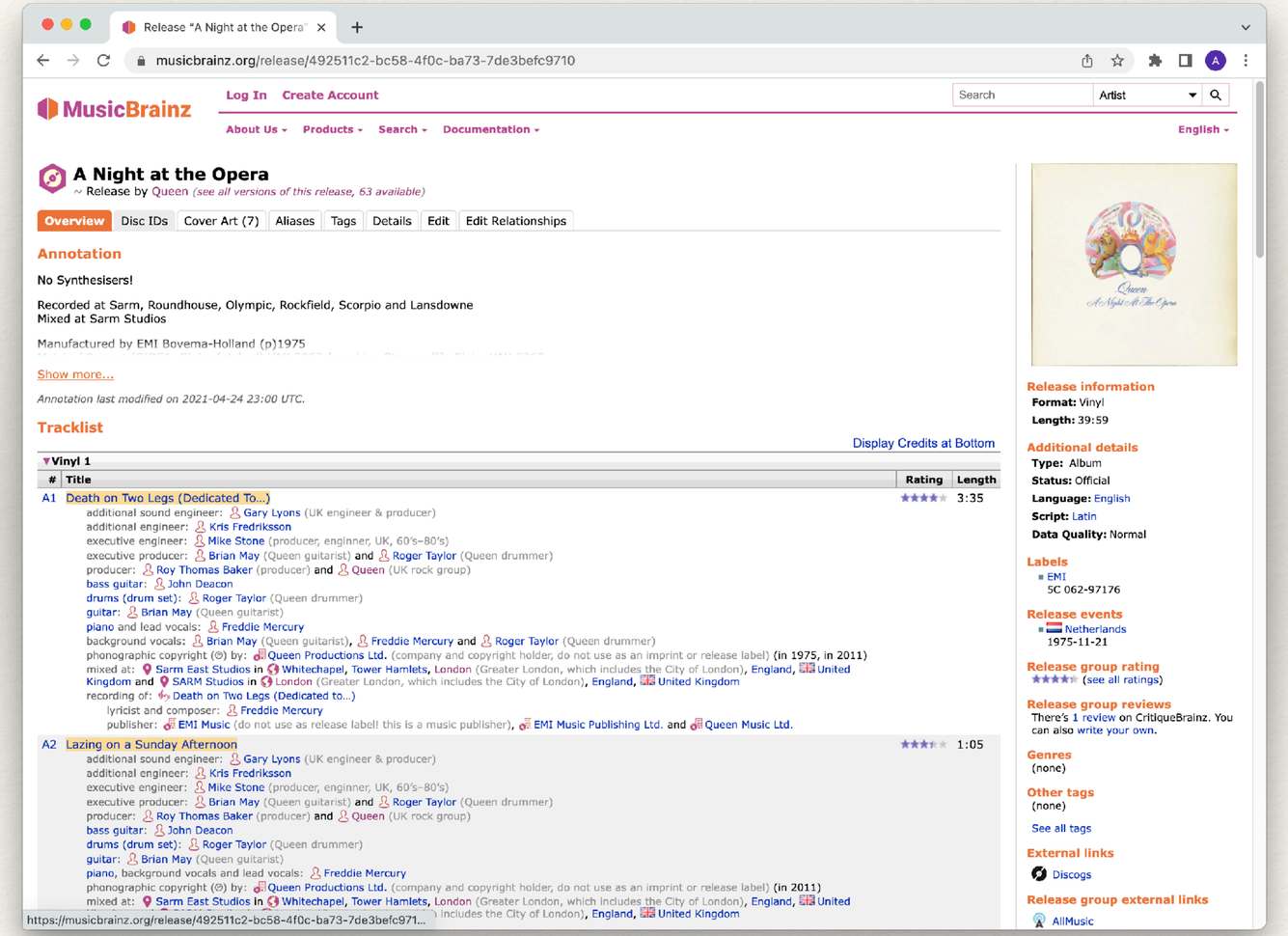
ListenBrainz

BookBrainz

MusicBrainz

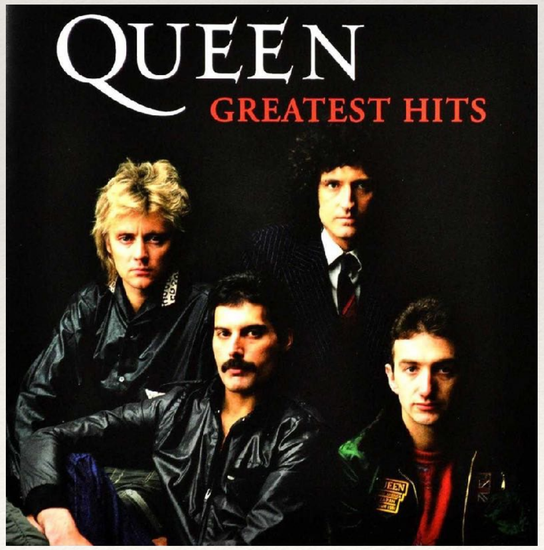
https://musicbrainz.org

- Open database of (recorded) music metadata
- Run by the MetaBrainz Foundation (California non-profit)
- 12 Paid developers/team members
- Community of 20,000+ active volunteers who add data
- Foundation funded by licensing fees for access to the MusicBrainz database

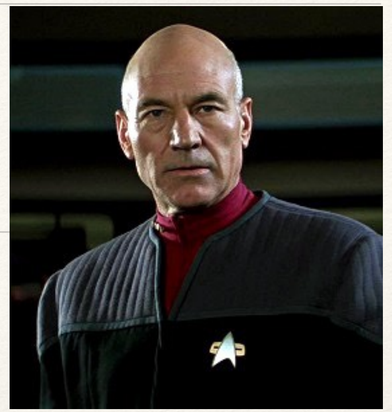


Original schema

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



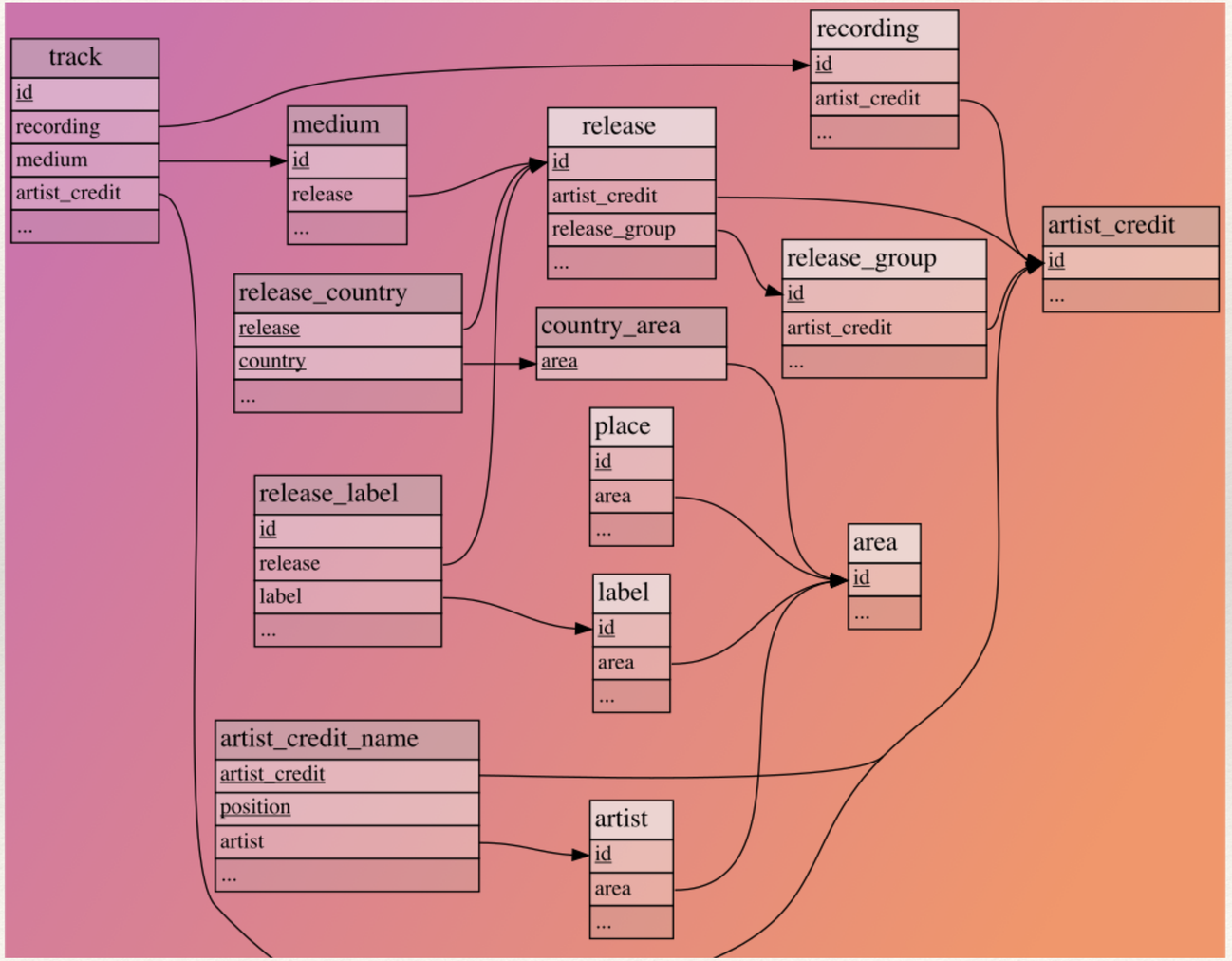
https://wiki.musicbrainz.org/MusicBrainz_Database/Schema

Schema

- https://wiki.musicbrainz.org/MusicBrainz_Database/Schema
- The schema was built from scratch (2011)
  - Solve a specific problem that we had
  - Not specifically based on an existing data schema (FRBR, Schema.org, MO, MARC)
  - We saw that we implemented many features of existing metadata schemas
  - Periodic updates to add new features (up to 2 releases per year, usually fewer)
- All items in the database have a MusicBrainz ID (Random UUID) for public identification
  - In our experience, every time we thought that this wasn't necessary, we eventually realised that it would be useful

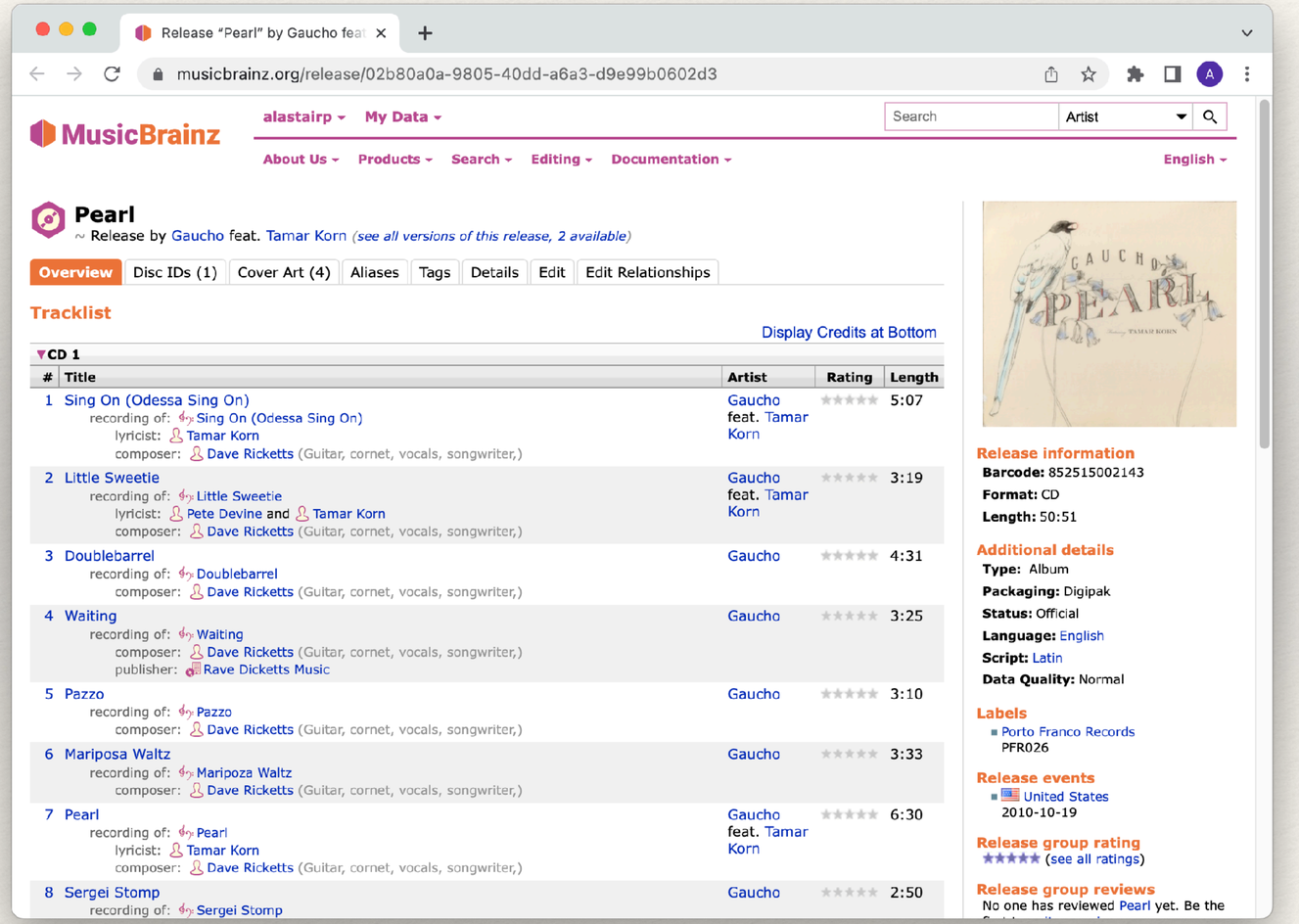
Core entities

- **Release Group**: contains multiple variants of the same **release** (vinyl, cassette, CD, reissue, country variants)
- A release can have 1 or more mediums (CD1, CD2, CD3...)
- **Recordings** can be shared between multiple releases. Link between medium-track-recording. Tracks are unique, but recordings may not be
- **Artist** credits explain who performed



Artist credits

- https://wiki.musicbrainz.org/Artist_Credits
- Who is the release attributed to?
- Recordings can optionally have another artist credit (otherwise it is shared with the release)
- Credits can include additional text (join phrase)
- The name of the artist in the credit can be different to the artist's original name

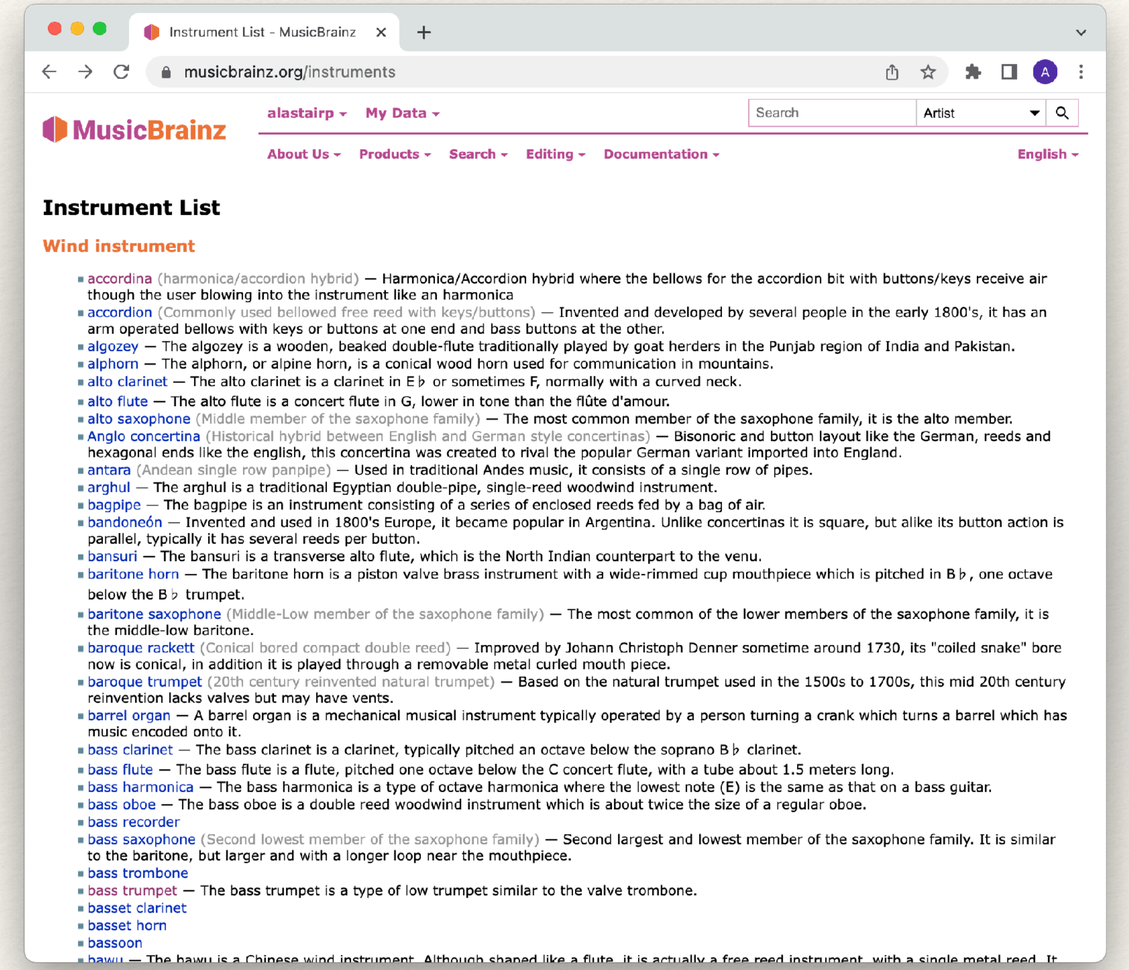


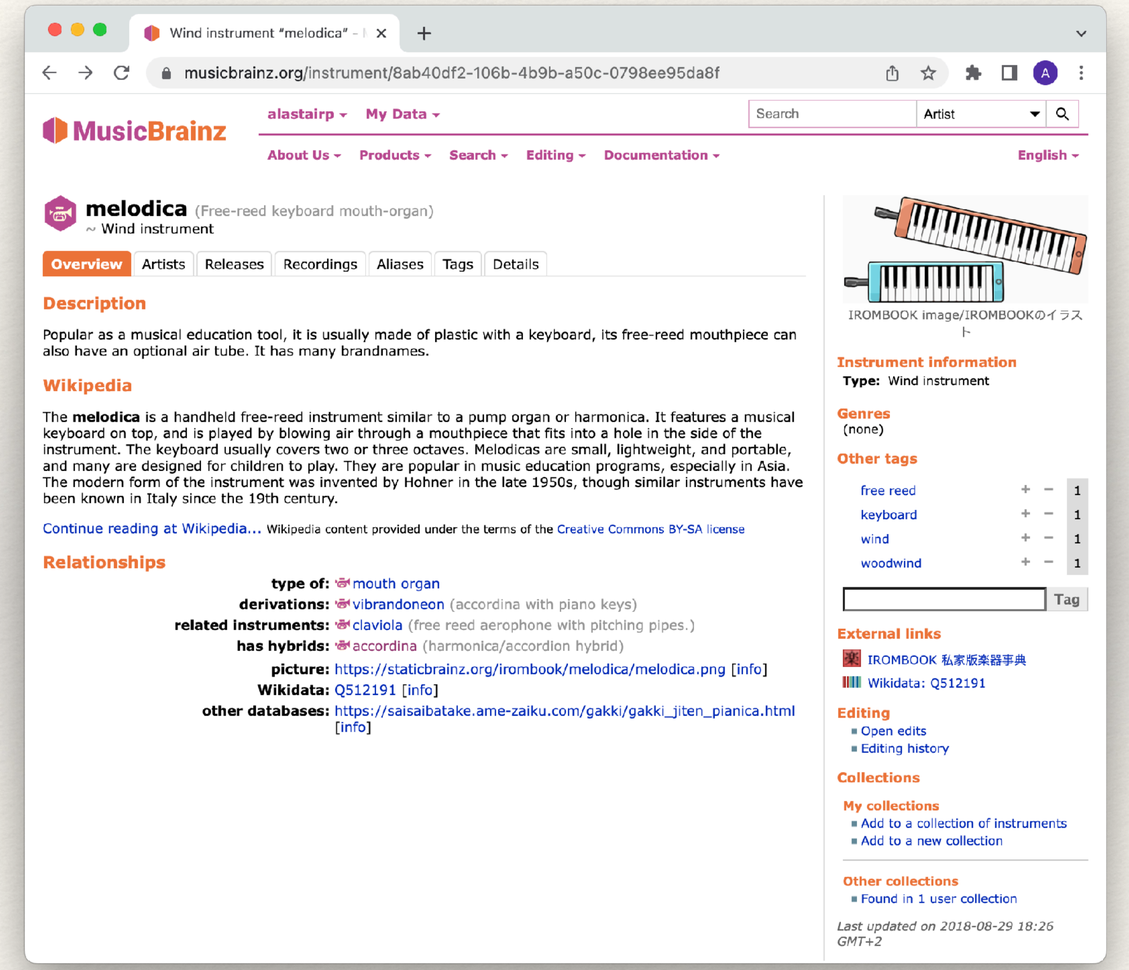
Other entities

- Not an exhaustive list, but these items have a MusicBrainz ID
- https://musicbrainz.org/doc/MusicBrainz_Entity
- Area (geographical)
- Place (Studio, Stadium)
- Event (Concert, Festival)
- Artists (Person or Group)
- Release ("Album")
- Release Group (Collection of releases)
- Recording
- Series
- Work ("Composition", but also movement, symphony)
- Genre
- Instrument

Instruments

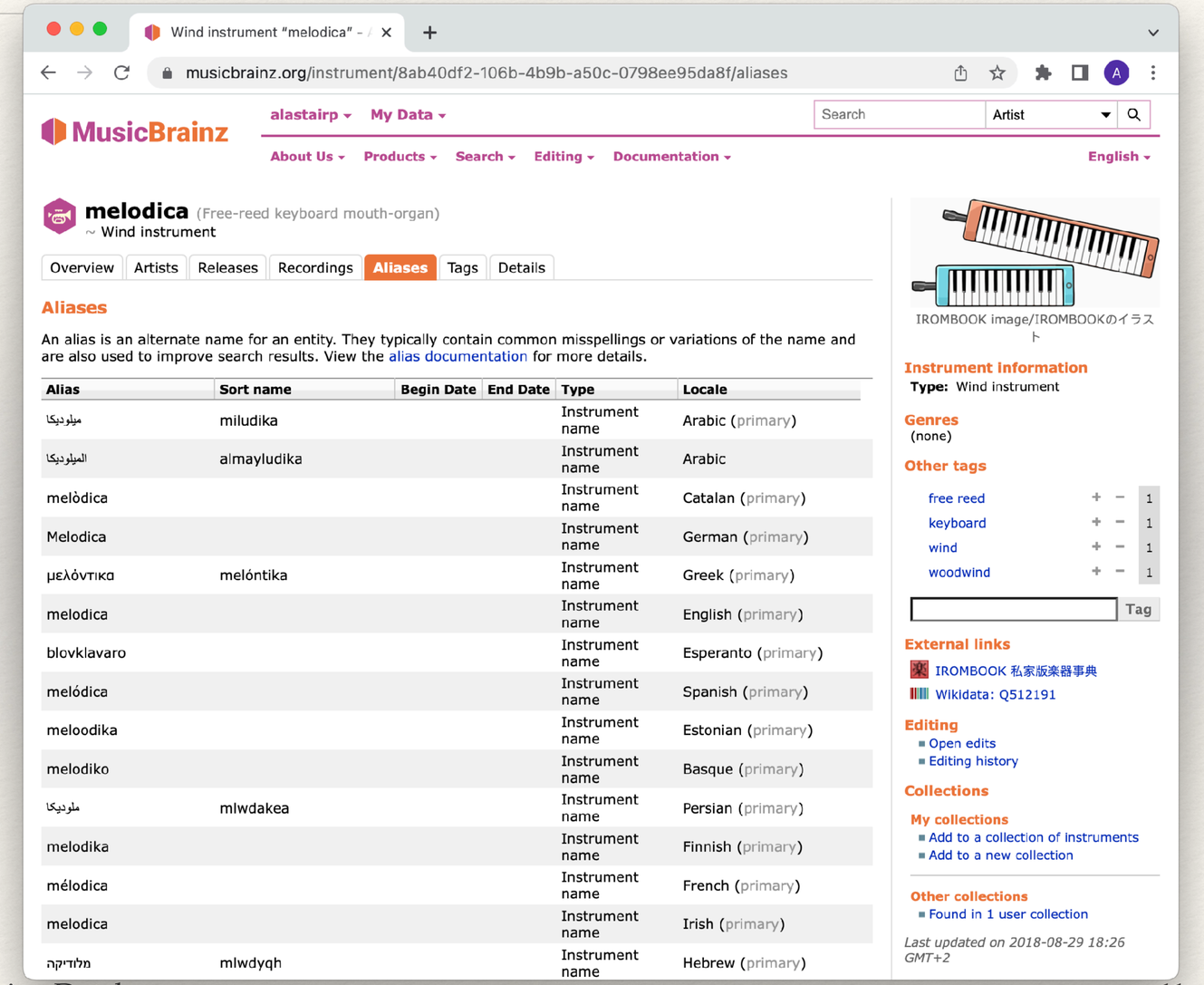
https://musicbrainz.org/instruments





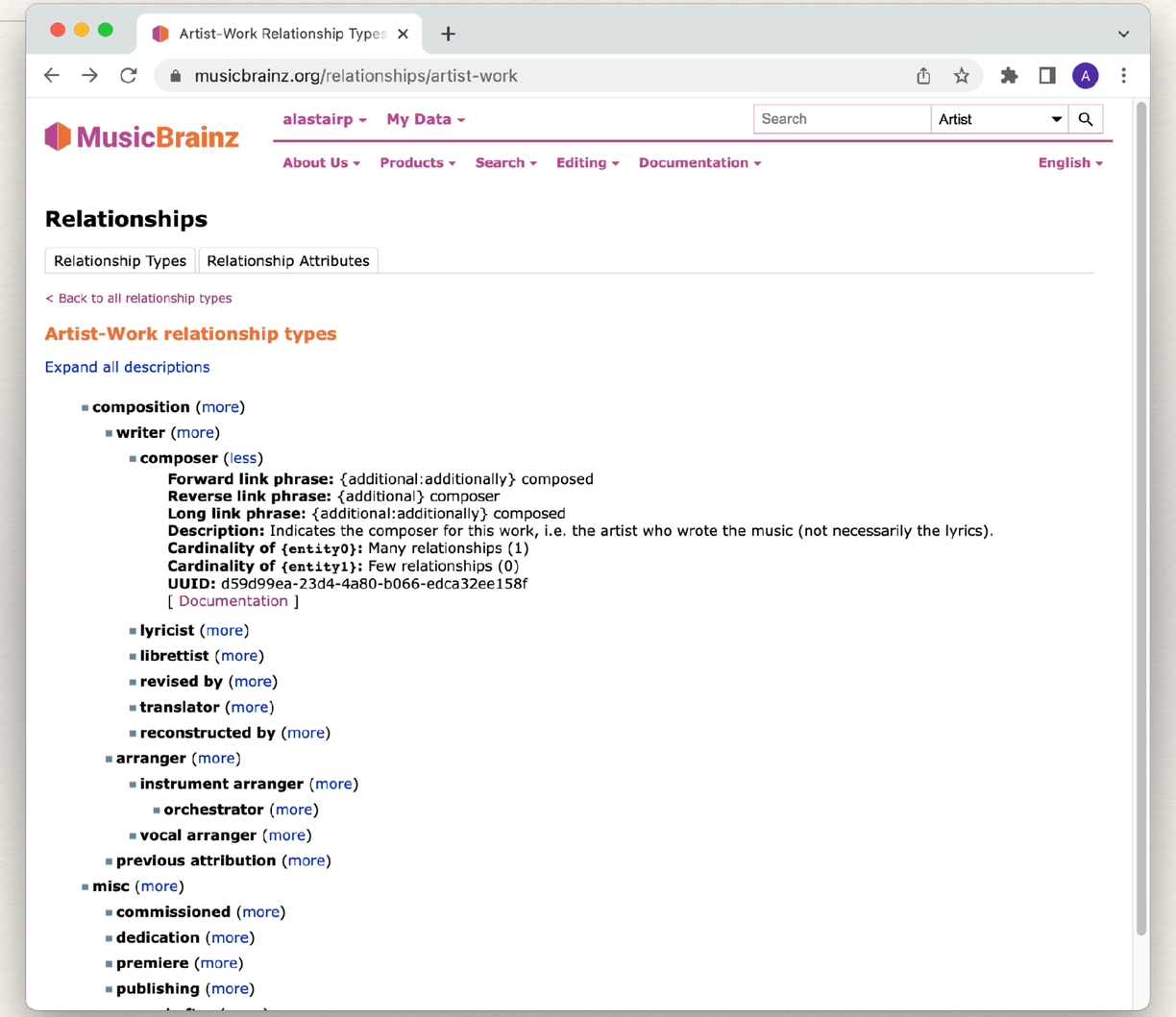
Aliases

- Alternate names for entities
- Most entity types can have an alias
- Per-language variations
  - https://musicbrainz.org/artist/9ddd7abc-9e1b-471d-8031-583bc6bc8be9
  - Name: Пётр Ильич Чайковский
  - "English Primary Alias": Pyotr Ilyich Tchaikovsky
  - "Search hints" (common misspellings)



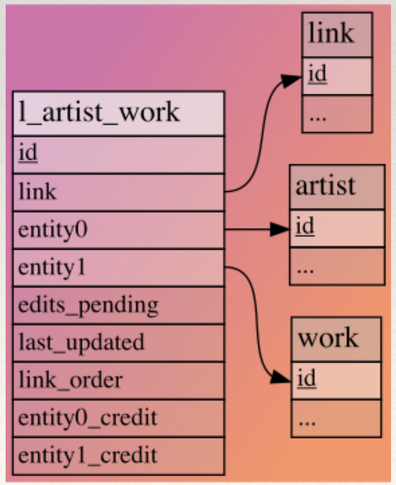
Attributes

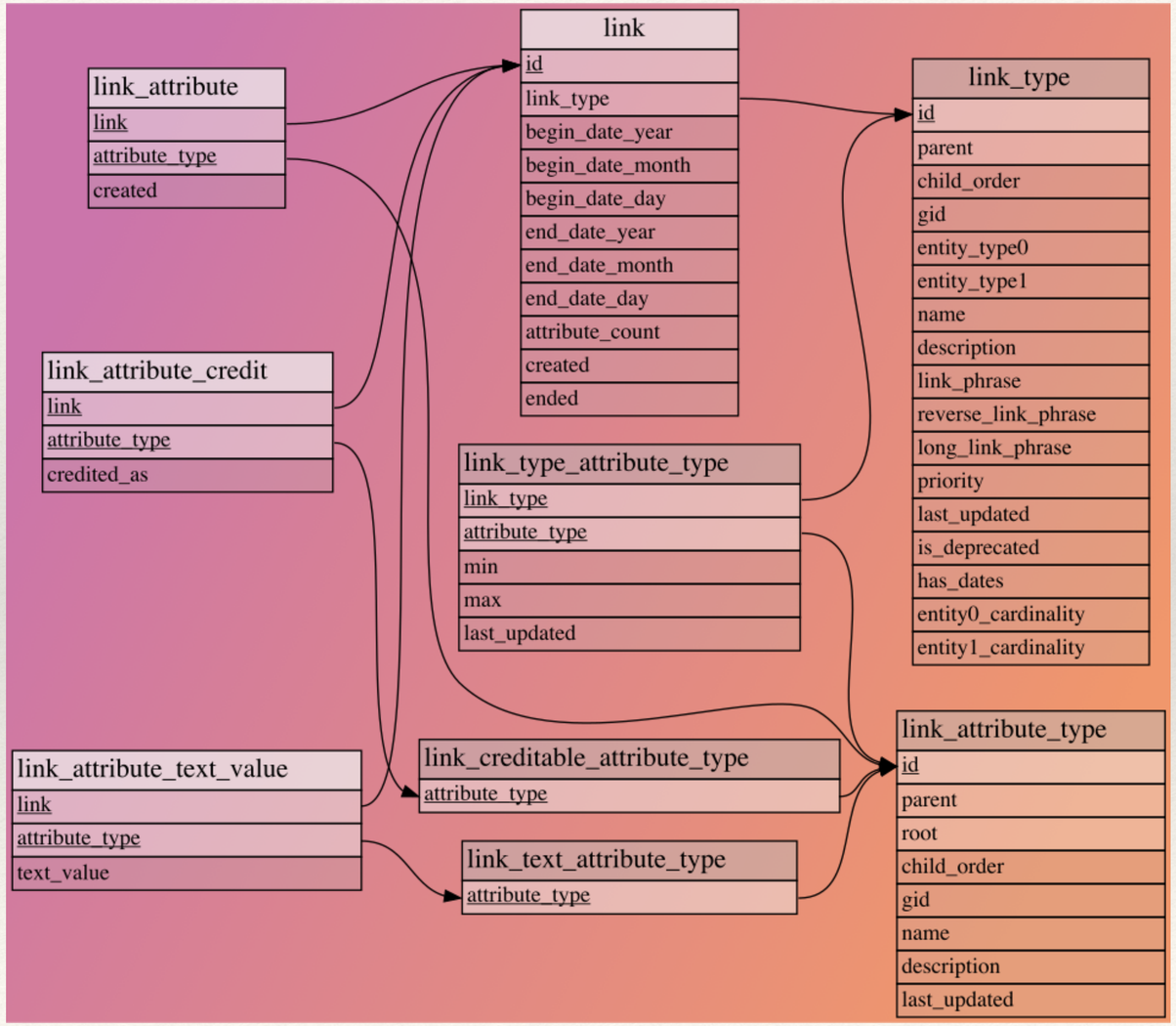
- https://musicbrainz.org/relationships
- Used to create links between entities in the database
  - Artist - composed - a work
  - Person - published - a work
  - Person - performed [instrument] - on a recording
  - Recording - was recorded [on date] - at place



Attributes

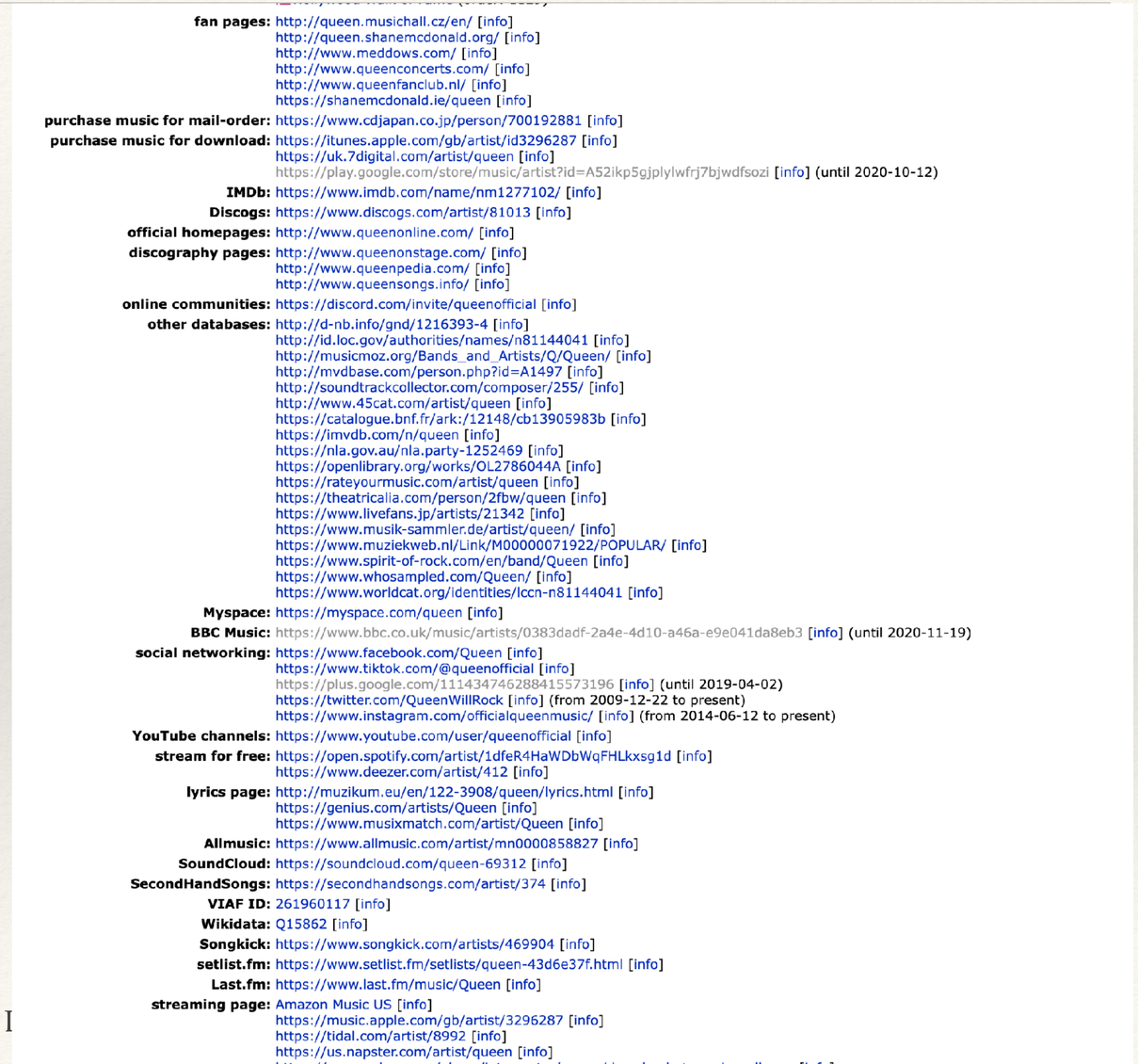
- One table per pair of items that can be related
- `link`: metadata about the link itself
- `link_type`: the nature of the link between the entities
- `attributes`: [lead] perfomer, [assistant] producer





External sites (URLs)

- Point to external locations
- Turns out these are just relationships (URL is an entity)
- Other authority catalogues
- https://musicbrainz.org/artist/0383dadf-2a4e-4d10-a46a-e9e041da8eb3
- Some relationship types are specific to a site
  - Other types are generic (other database)



fan pages: http://queen.musichall.cz/en/ [info]
http://queen.shanemcdonald.org/ [info]
http://www.meddows.com/ [info]
http://www.queenconcerts.com/ [info]
http://www.queenfanclub.nl/ [info]
https://shanemcdonald.ie/queen [info]

purchase music for mail-order: https://www.cdjapan.co.jp/person/700192881 [info]

purchase music for download: https://itunes.apple.com/gb/artist/id3296287 [info]
https://uk.7digital.com/artist/queen [info]
https://play.google.com/store/music/artist?id=A52ikp5gjplylwfrj7bjwdfsozi [info] (until 2020-10-12)

IMDb: https://www.imdb.com/name/nm1277102/ [info]

Discogs: https://www.discogs.com/artist/81013 [info]

official homepages: http://www.queenonline.com/ [info]

discography pages: http://www.queenonstage.com/ [info]
http://www.queenpedia.com/ [info]
http://www.queensongs.info/ [info]

online communities: https://discord.com/invite/queenofficial [info]

other databases: http://d-nb.info/gnd/1216393-4 [info]
http://id.loc.gov/authorities/names/n81144041 [info]
http://musicmoz.org/Bands_and_Artists/Q/Queen/ [info]
http://mvdbase.com/person.php?id=A1497 [info]
http://soundtrackcollector.com/composer/255/ [info]
http://www.45cat.com/artist/queen [info]
https://catalogue.bnf.fr/ark:/12148/cb13905983b [info]
https://imvdb.com/n/queen [info]
https://nla.gov.au/nla.party-1252469 [info]
https://openlibrary.org/works/OL2786044A [info]
https://rateyourmusic.com/artist/queen [info]
https://theatricalia.com/person/2fbw/queen [info]
https://www.livefans.jp/artists/21342 [info]
https://www.musik-sammler.de/artist/queen/ [info]
https://www.muziekweb.nl/Link/M00000071922/POPULAR/ [info]
https://www.spirit-of-rock.com/en/band/Queen [info]
https://www.whosampled.com/Queen/ [info]
https://www.worldcat.org/identities/lccn-n81144041 [info]

Myspace: https://myspace.com/queen [info]

BBC Music: https://www.bbc.co.uk/music/artists/0383dadf-2a4e-4d10-a46a-e9e041da8eb3 [info] (until 2020-11-19)

social networking: https://www.facebook.com/Queen [info]
https://www.tiktok.com/@queenofficial [info]
https://plus.google.com/111434746288415573196 [info] (until 2019-04-02)
https://twitter.com/QueenWillRock [info] (from 2009-12-22 to present)
https://www.instagram.com/officialqueenmusic/ [info] (from 2014-06-12 to present)

YouTube channels: https://www.youtube.com/user/queenofficial [info]

stream for free: https://open.spotify.com/artist/1dfeR4HaWDbWqFHLkxsg1d [info]
https://www.deezer.com/artist/412 [info]

lyrics page: http://muzikum.eu/en/122-3908/queen/lyrics.html [info]
https://genius.com/artists/Queen [info]
https://www.musixmatch.com/artist/Queen [info]

Allmusic: https://www.allmusic.com/artist/mn0000858827 [info]

SoundCloud: https://soundcloud.com/queen-69312 [info]

SecondHandSongs: https://secondhandsongs.com/artist/374 [info]

VIAF ID: 261960117 [info]

Wikidata: Q15862 [info]

Songkick: https://www.songkick.com/artists/469904 [info]

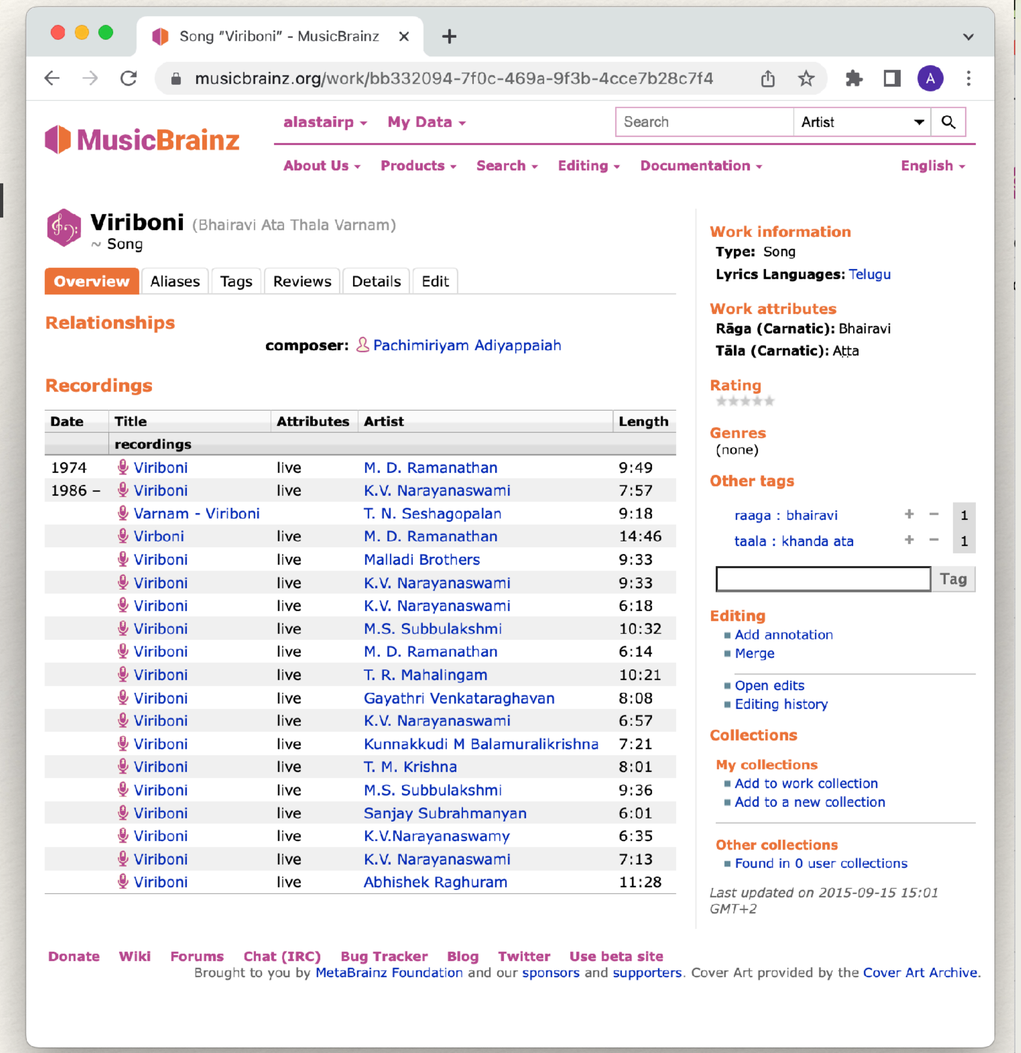
setlist.fm: https://www.setlist.fm/setlists/queen-43d6e37f.html [info]

Last.fm: https://www.last.fm/music/Queen [info]

streaming page: Amazon Music US [info]
https://music.apple.com/gb/artist/3296287 [info]
https://tidal.com/artist/8992 [info]
https://us.napster.com/artist/queen [info]

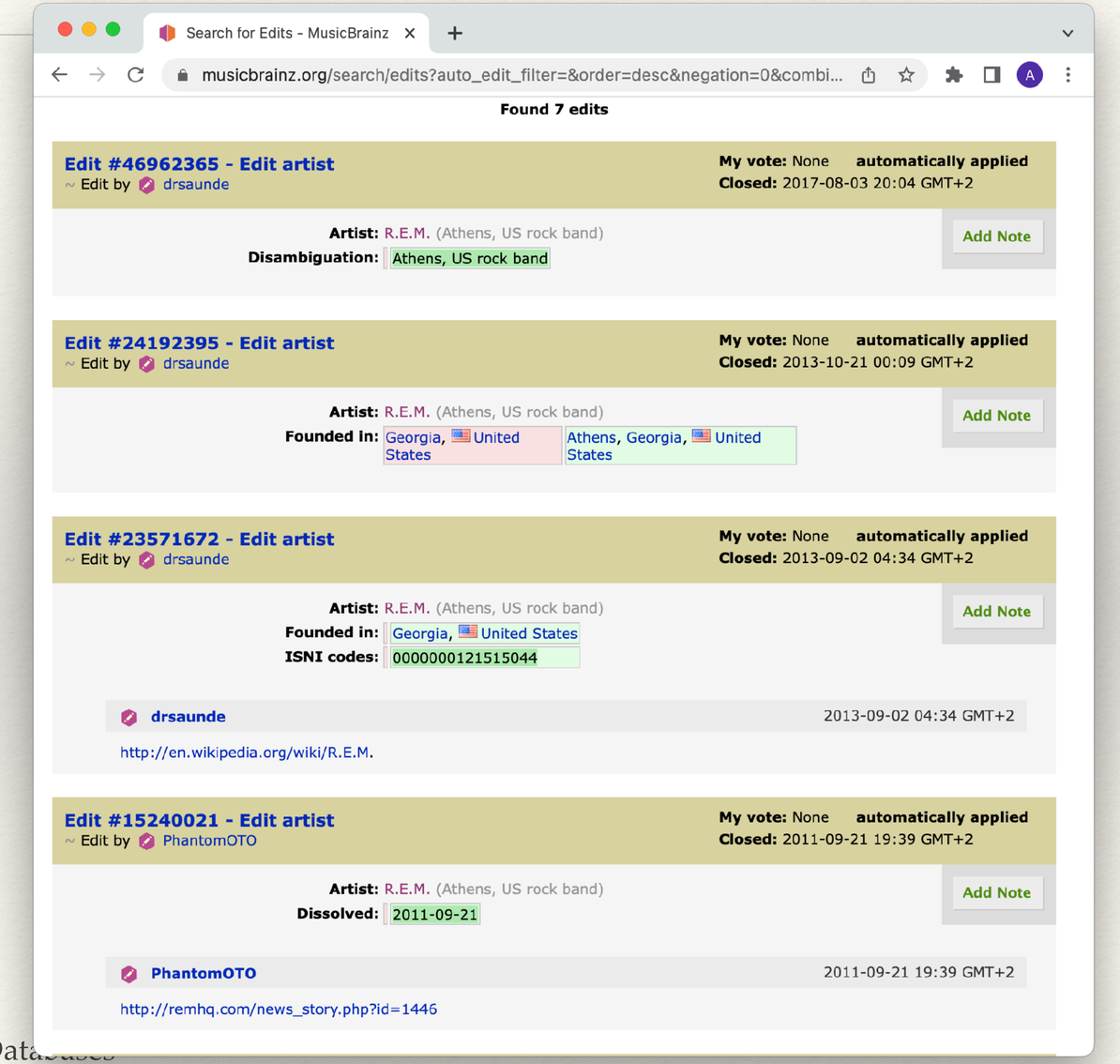
Work attributes

- CompMusic project (2011-2017)
- We decided to add all content to MusicBrainz instead of creating a separate database
- Some information didn't fit in the existing MusicBrainz schema
- We used tags to temporarily store it
- MusicBrainz added additional functionality for our requirements (also subsequently used for other purposes)
- https://musicbrainz.org/work/adcdc472-8b19-4e6f-aa4e-be8c6aea5f8a
- https://musicbrainz.org/work/bb332094-7f0c-469a-9f3b-4cce7b28c7f4



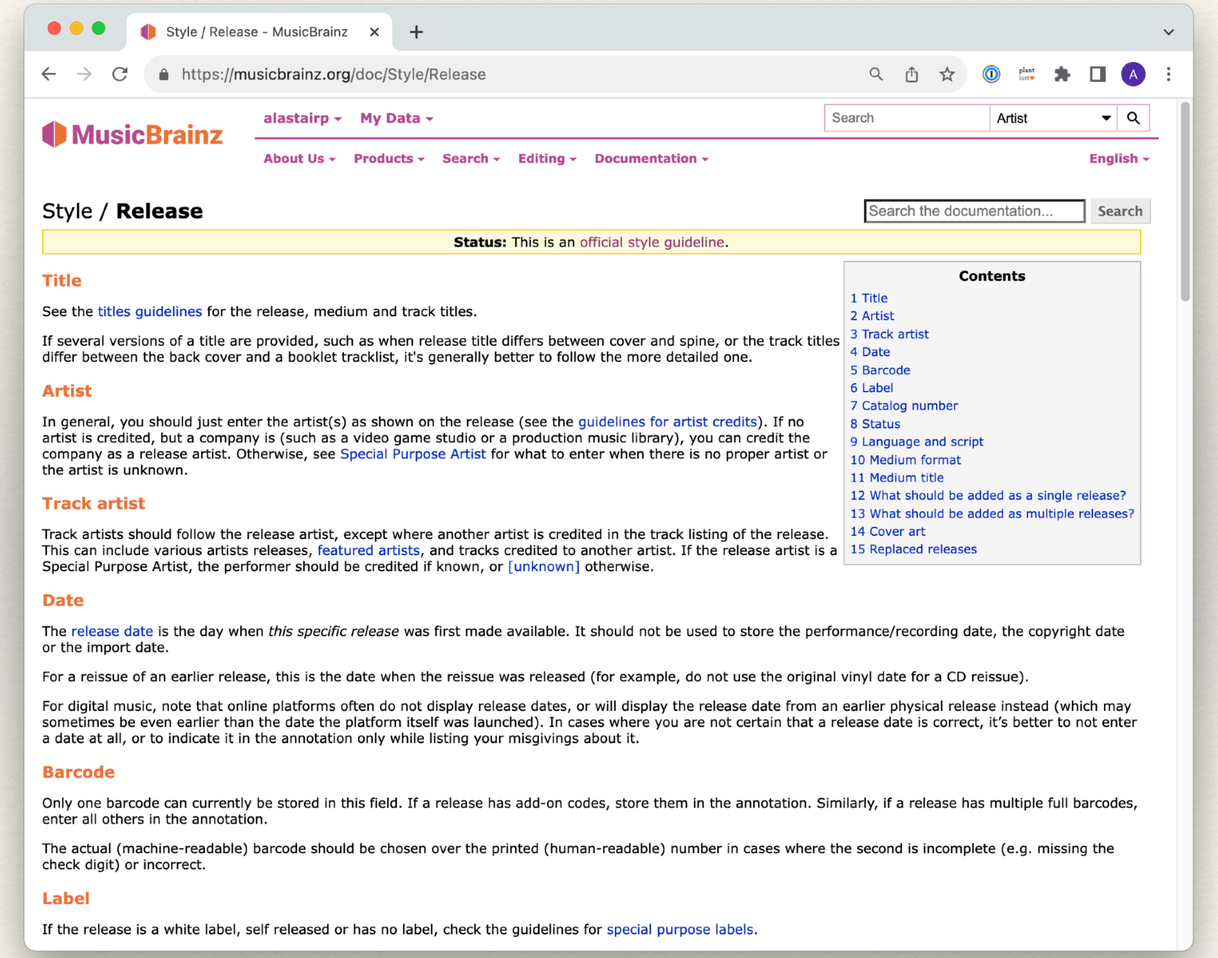
Edits

- All changes to the database are proposed as *edits* to the database
- Edits are voted on by the community to determine if they are valid
- An edit history is stored for all entities
- A list of edits is also used to perform updates to database mirrors
- https://musicbrainz.org/search/edits?auto_edit_filter=&order=desc&negation=0&combinator=and&conditions.0.field=artist&conditions.0.operator=%3D&conditions.0.name=R.E.M.&conditions.0.args.0=10302&conditions.1.field=type&conditions.1.operator=%3D&conditions.1.args=2



Data correctness

- The Style Guide describes how data should be entered in the case of uncertainty
- Rules mostly agreed by consensus, but there is an authoritarian style leader to break deadlocks if required



Data access

- Multiple different formats
- Originally: XML WebService
- Later: JSON API
- Also: JSON-LD (for indexing into search engines)
  - Send `Accept: application/ld-json` to get this

Data formats

- API available via XML (with schema), JSON
- Main pages also available via JSON-LD (used by search engines)


```
{
  "name": "Queen",
  "isnis": [
    "0000000115265309"
  ],
  "gender": null,
  "sort-name": "Queen",
  "type-id": "e431f5f6-b5d2-343d-8b36-72607fffb74b",
  "disambiguation": "UK rock group",
  "ipis": [],
  "area": {
    "id": "8a754a16-0027-3a29-b6d7-2b40ea0481ed",
    "iso-3166-1-codes": [
      "GB"
    ],
    "name": "United Kingdom",
    "disambiguation": "",
    "type": null,
    "sort-name": "United Kingdom",
    "type-id": null
  },
```

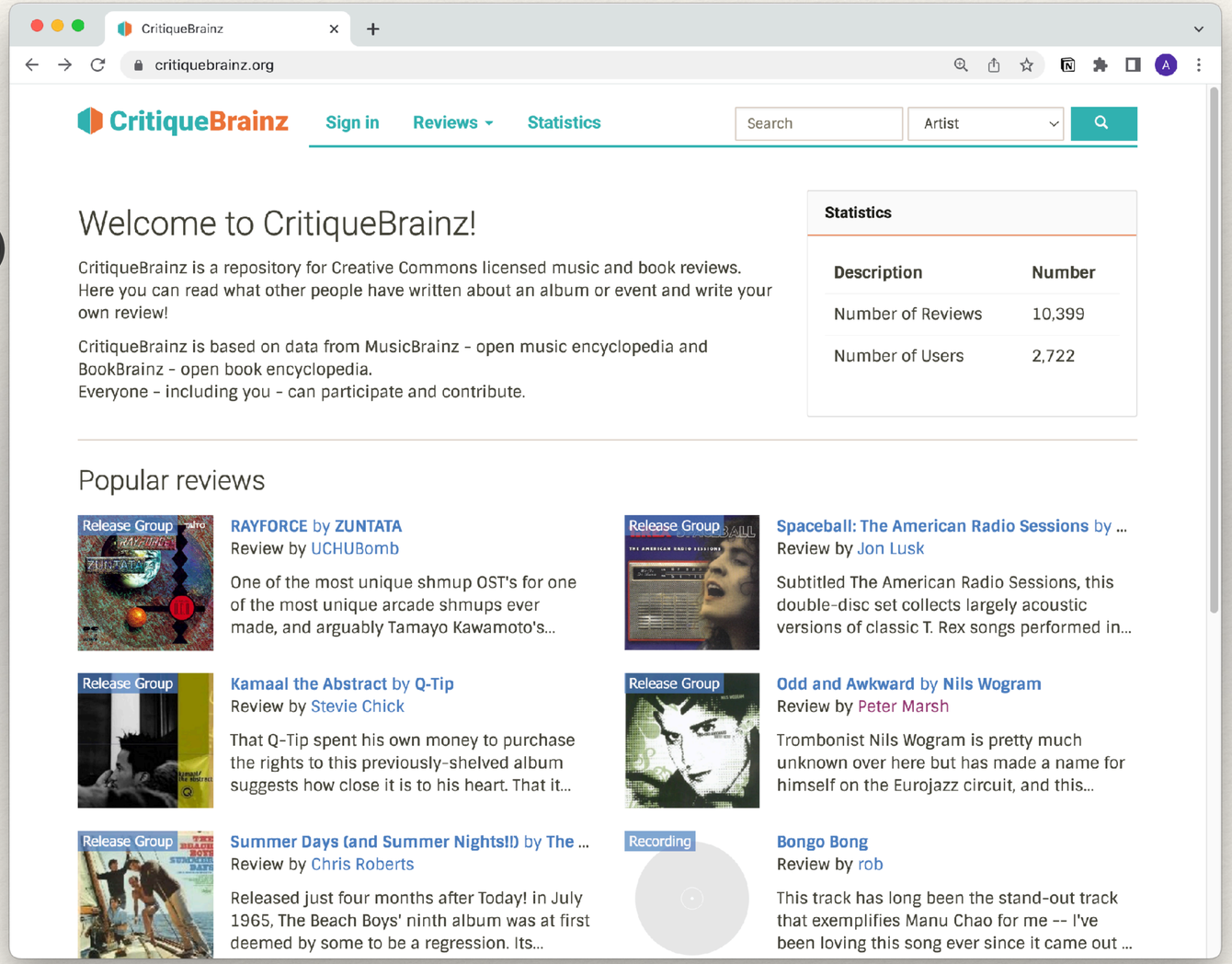
```
{
  "@id": "http://musicbrainz.org/artist/0383dadf-2a4e-4d10-a46a-e9e041da8eb3",
  "@type": "MusicGroup",
  "album": [
    {
      "creditedTo": "Queen",
      "albumProductionType": "http://schema.org/StudioAlbum",
      "byArtist": {
        "name": "Queen",
        "@type": "MusicGroup",
        "@id": "http://musicbrainz.org/artist/0383dadf-2a4e-4d10-a46a-e9e041da8eb3",
        "genre": [
          "http://musicbrainz.org/genre/0e3fc579-2d24-4f20-9dae-736e1ec78798",
          "http://musicbrainz.org/genre/51cb9f91-e6a2-41bf-891f-e78e3f1e52ab",
          "http://musicbrainz.org/genre/54d89e62-5bfb-42bc-a321-9230e6fdcd75",
          "http://musicbrainz.org/genre/b7ef058e-6d83-4ca4-8123-9724bff4648b",
          "http://musicbrainz.org/genre/797e2e85-5ffd-495c-a757-8b4079052f0e"
        ]
      },
      "@type": "MusicAlbum",
      "@id": "http://musicbrainz.org/release-group/810068af-2b3c-3e9c-b2ab-68a3f3e3787d",
      "name": "Queen",
```

Offline data access

- Data dumps made available twice a week
- Mostly CC0 (Factual data)
  - Some data is CC-SA (annotations, tags, ratings)
- PostgreSQL: https://musicbrainz.org/doc/MusicBrainz_Database/Download
- Easy-to-install MusicBrainz server (database, website, search) via docker:
  - https://github.com/metabrainz/musicbrainz-docker
- JSON Dumps: Entire output of the JSON API
  - https://musicbrainz.org/doc/Development/JSON_Data_Dumps

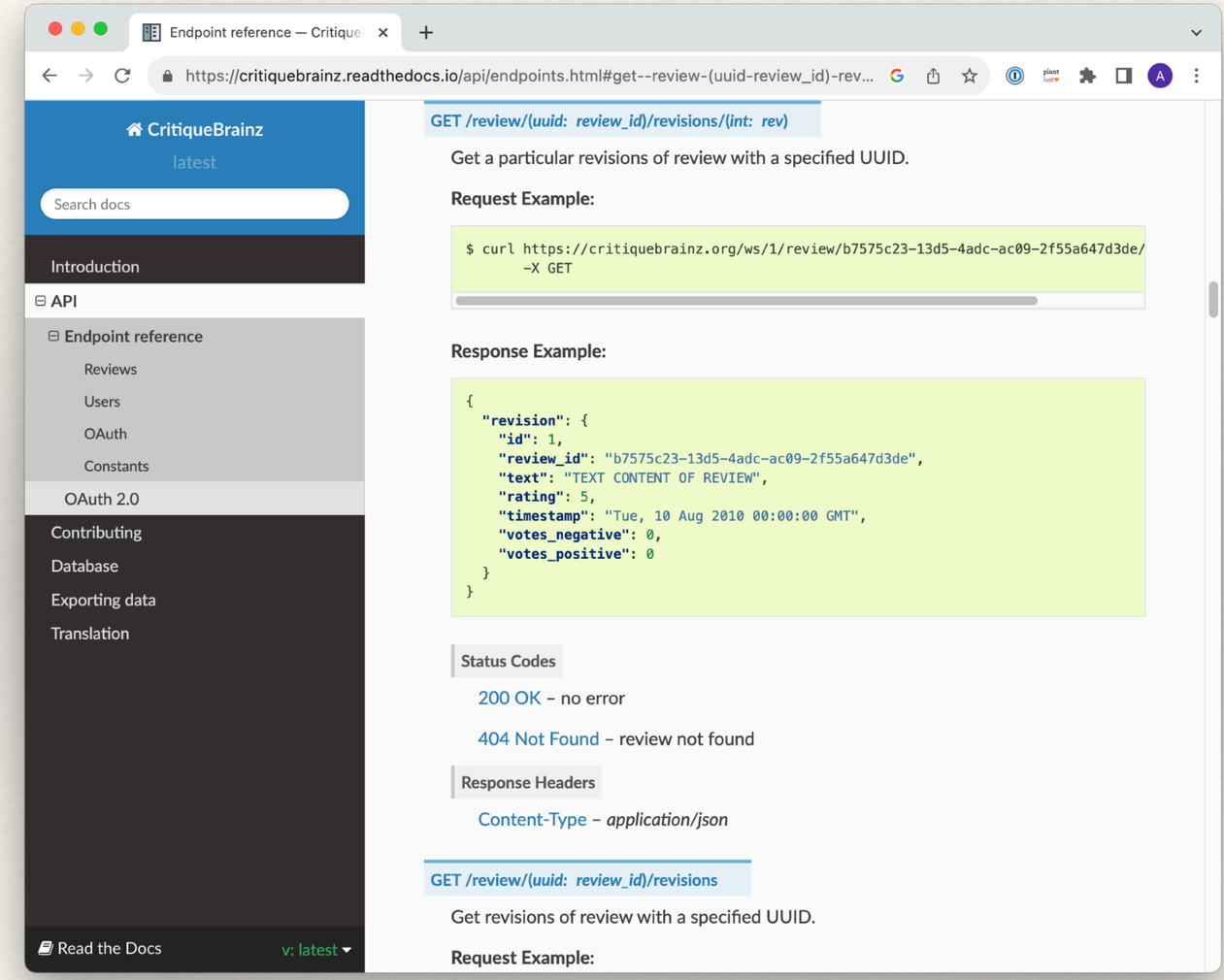
CritiqueBrainz

- https://critiquebrainz.org
- Open database of reviews of music (and books)



CritiqueBrainz API

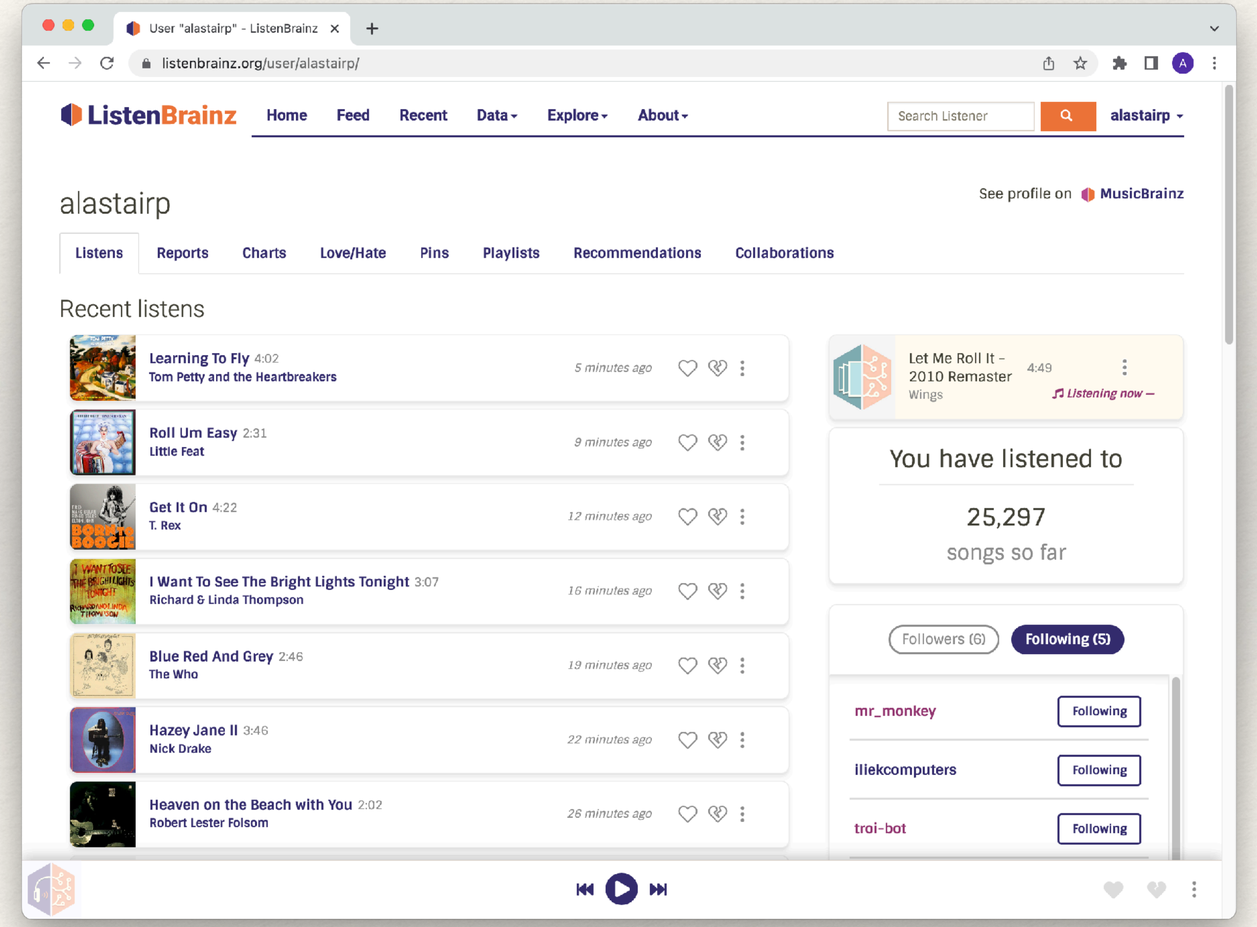
- API
  - https://critiquebrainz.readthedocs.io/
- Database download
  - http://data.musicbrainz.org/pub/musicbrainz/critiquebrainz/dump/
- Main entity type: Review
  - Linked to a specific MusicBrainz/BookBrainz entity
  - Reviews have revisions (public edits)
  - A review can include text, a rating, or both
  - Others can vote on the quality of a review
- Coming soon: Integration into ListenBrainz



ListenBrainz

https://listenbrainz.org

- Open database of user listening history
- import from lastfm, spotify
- Statistics
- Community functionality
- Recommendations



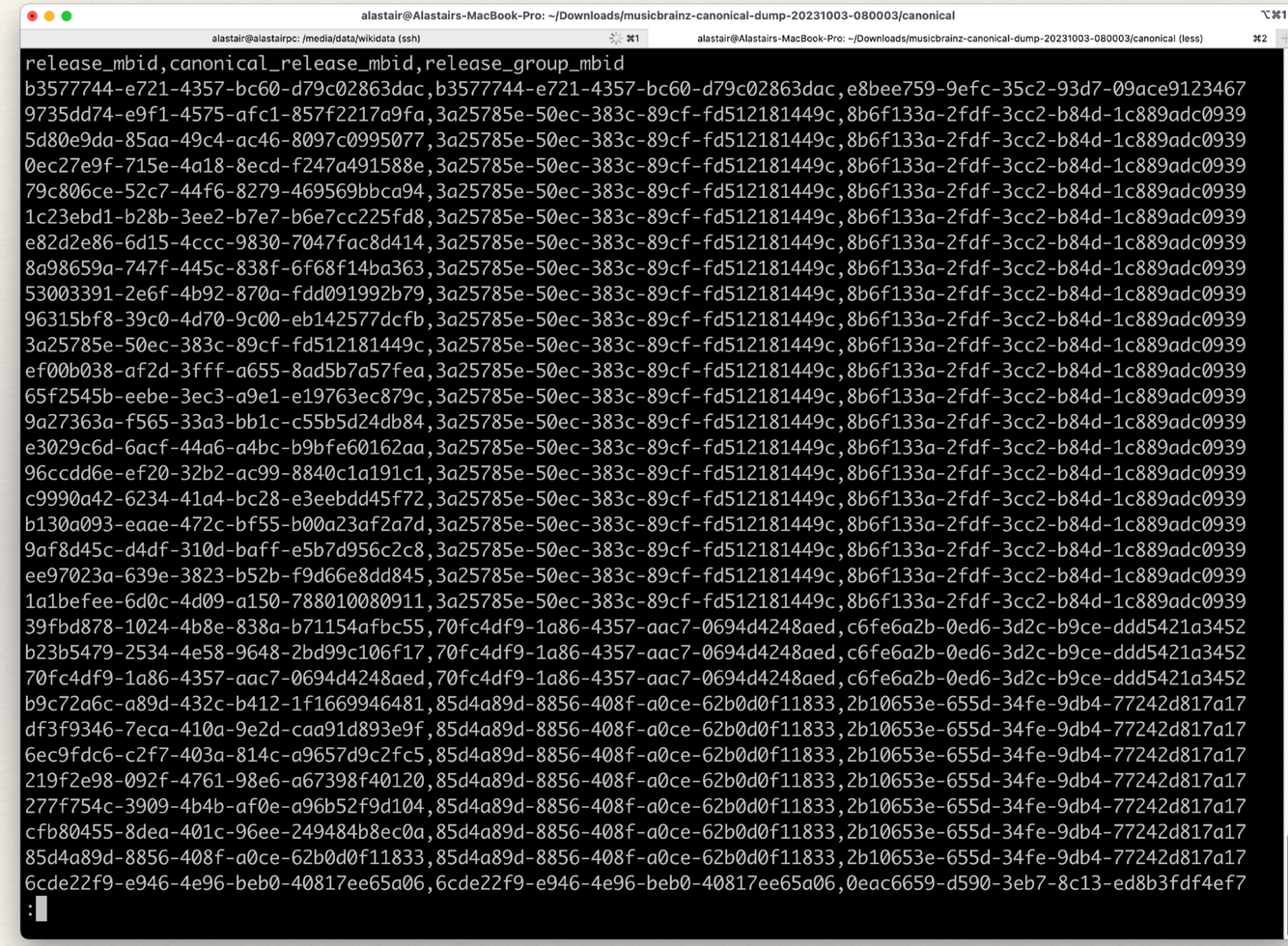
ListenBrainz data access

- Primary data object, a "Listen"
  - user x listened to track y at time z
  - plus additional metadata if you want
  - automatically linked to MusicBrainz when the data is submitted (see Canonical - next)
- https://data.musicbrainz.org/pub/musicbrainz/listenbrainz/


```
- {
    inserted_at: 1697804497,
    listened_at: 1697804053,
    recording_msid: "4c00aaa6-a2ab-42ef-93d5-8c66b1ad22bc",
  - track_metadata: {
    - additional_info: {
      - artist_names: [
            "The Magnetic Fields"
        ],
        discnumber: 1,
        duration_ms: 166853,
        isrc: "USNO11600667",
        music_service: "spotify.com",
        origin_url: "https://open.spotify.com/track/3XSmaFNg87VuralouTs5yf",
        recording_msid: "4c00aaa6-a2ab-42ef-93d5-8c66b1ad22bc",
        release_artist_name: "The Magnetic Fields",
      - release_artist_names: [
            "The Magnetic Fields"
        ],
      - spotify_album_artist_ids: [
            "https://open.spotify.com/artist/6RWjTQqILL7a1tQ0VapyLK"
        ],
        spotify_album_id: "https://open.spotify.com/album/2b3JjOvKw9UN47ND80Tvri",
      - spotify_artist_ids: [
            "https://open.spotify.com/artist/6RWjTQqILL7a1tQ0VapyLK"
        ],
        spotify_id: "https://open.spotify.com/track/3XSmaFNg87VuralouTs5yf",
        submission_client: "listenbrainz",
        tracknumber: 3
      },
      artist_name: "The Magnetic Fields",
    - mbid_mapping: {
      - artist_mbids: [
            "3ff72a59-f39d-411d-9f93-2d4a86413013"
        ],
      - artists: [
          - {
                artist_credit_name: "The Magnetic Fields",
                artist_mbid: "3ff72a59-f39d-411d-9f93-2d4a86413013",
                join_phrase: ""
            }
        ],
        caa_id: 16168564506,
        caa_release_mbid: "9d812089-052f-4f5d-a474-7448a6fd80d1",
        recording_mbid: "786734f0-135b-48e4-ae16-62eec3d88523",
        recording_name: "'68 A Cat Called Dionysus",
        release_mbid: "bd7de75d-3587-49ec-a07f-2e2c825d3538"
      },
      release_name: "50 Song Memoir",
```

ListenBrainz ancillary datasets: Canonical

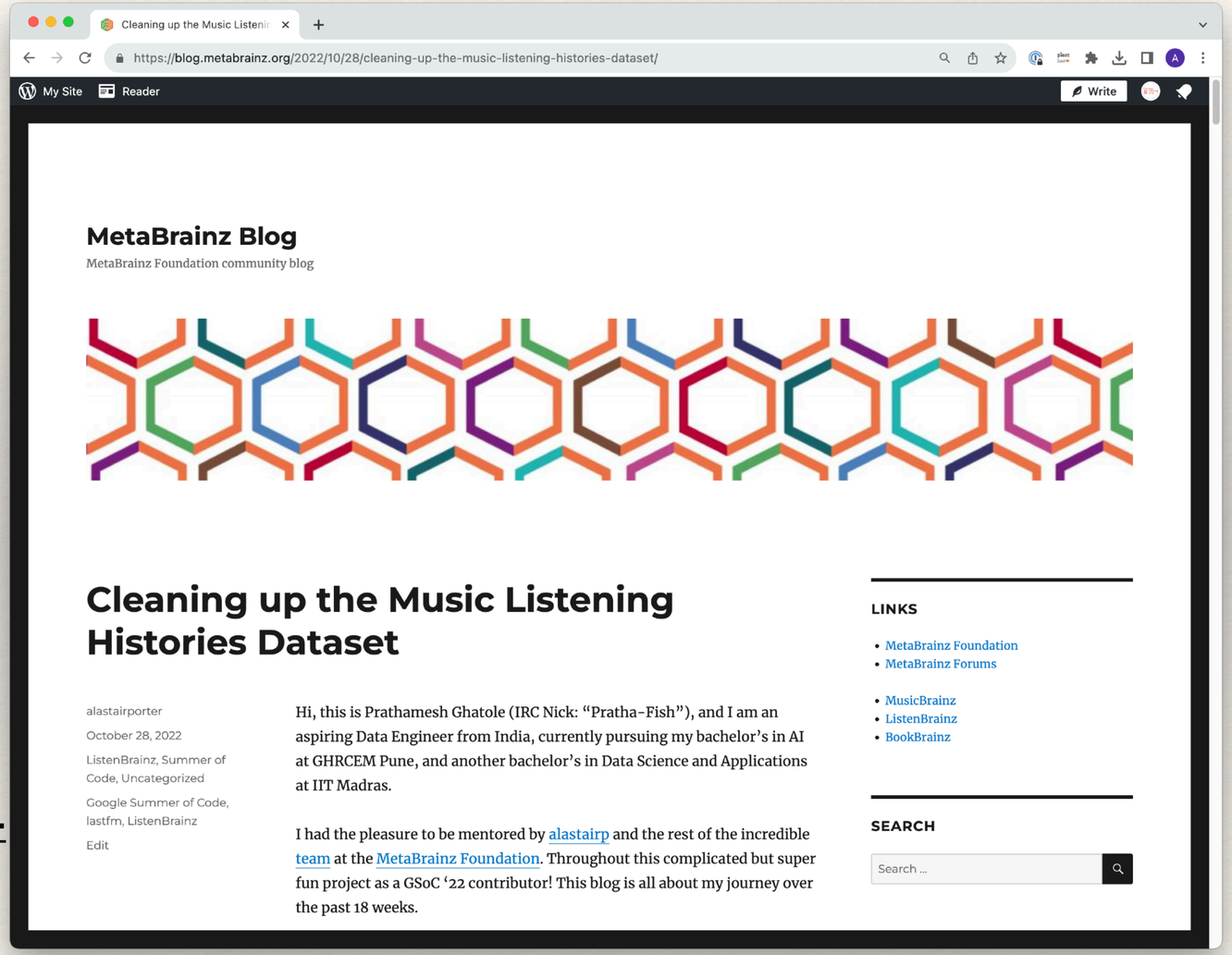
- Because of the nature of the MusicBrainz database it contains many instances where metadata exists for many different versions (formats, reissues) of the same album, or many versions (live recordings, alternate versions) of a recording.
- However for some tasks such as music recommendation, it's often useful to be able to use a single identifier when talking about a recording or a release, both in order to improve the quality of a training dataset for machine learning and when providing results to an end user.
- https://musicbrainz.org/doc/Canonical_MusicBrainz_data



MLHD+

- Improved version of MLHD (DDMAL, 2017)
  - Follow redirects:
  - Identify Canonical Recordings: Allow direct comparison if two people listened to the same recording which was previously reported as 2 items
  - Lookup correct Artist and Releases using Canonical recordings
  - Other data problems: Remove missing data that doesn't match any known MBID
  - Better compression

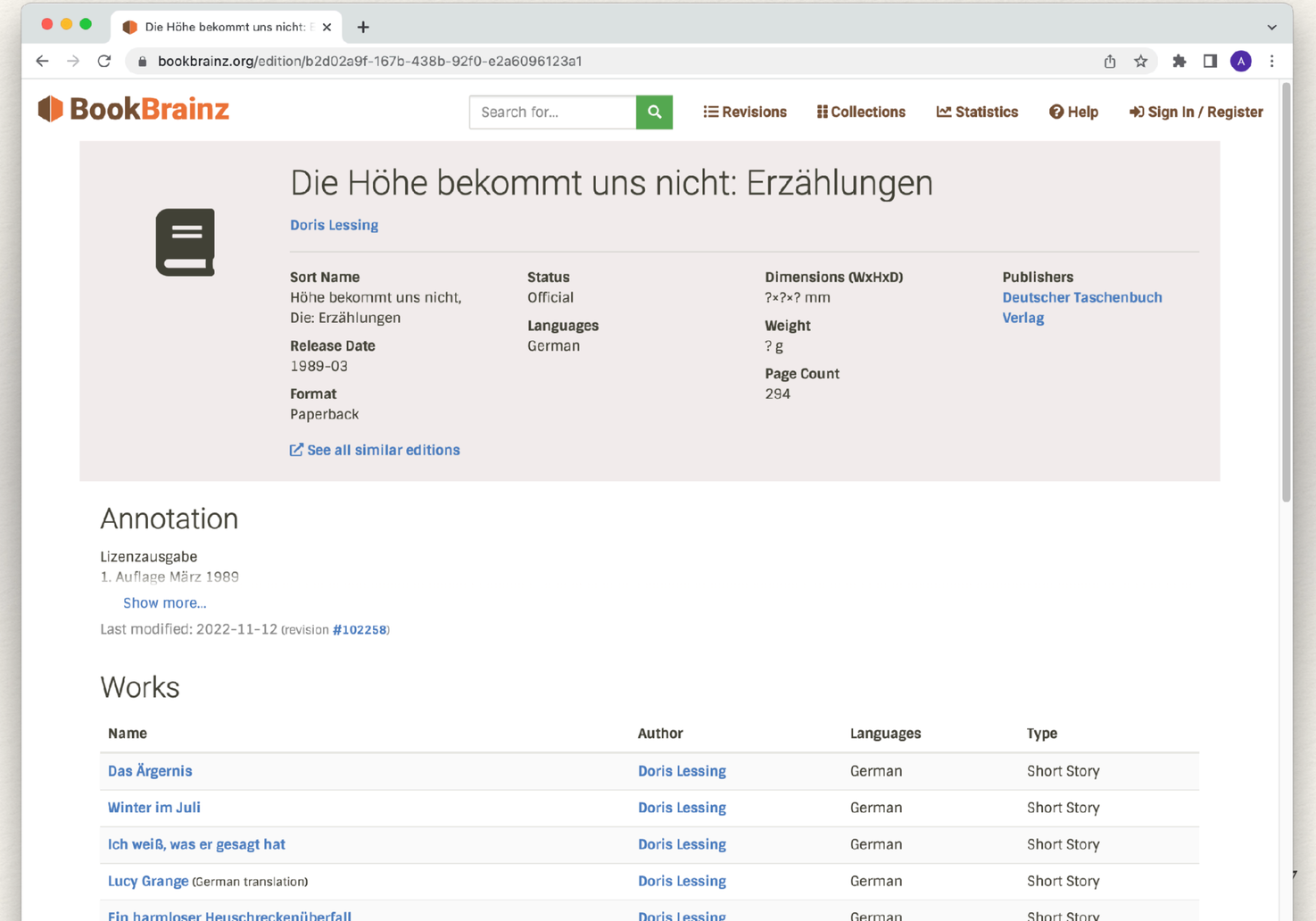
- https://blog.metabrainz.org/2022/10/28/cleaning-up-the-music-listening-histories-dataset/



BookBrainz

https://bookbrainz.org

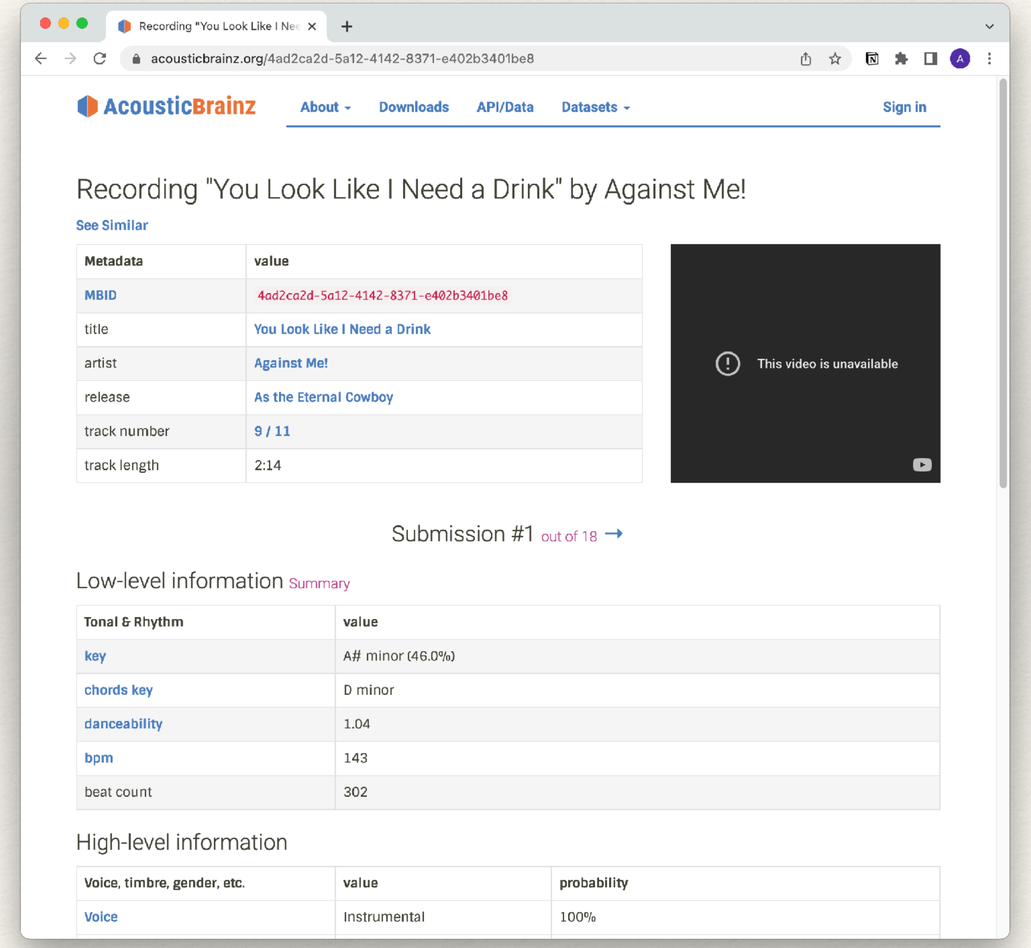
- Open database of books
- API (in testing still)
  - https://api.test.bookbrainz.org/1/docs/
- Database download
  - https://data.musicbrainz.org/pub/musicbrainz/bookbrainz/



AcousticBrainz

https://acousticbrainz.org

- Public archive of features automatically extracted from music
- Uses Essentia (https://essentia.upf.edu/)
- Audio descriptors (MFCCs), Tonal features (key, chords), Rhythmic Features (bpm)
- Support Vector Machines to compute features using common MIR tasks
- Ran from 2015-2022



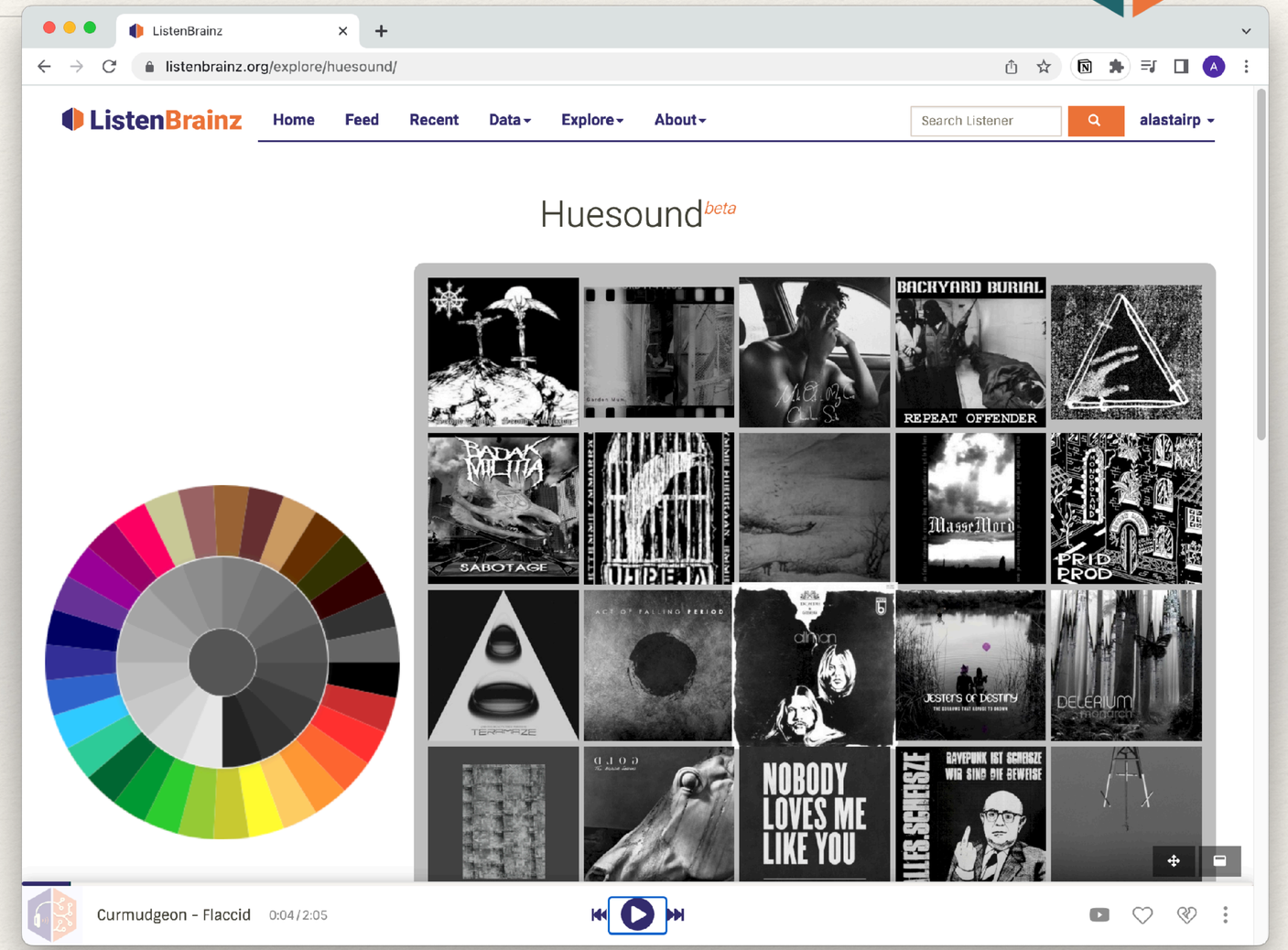
AcousticBrainz data

- API to download data for a single recording
  - https://acousticbrainz.readthedocs.io/api.html
- Download a copy of the entire database
  - https://acousticbrainz.org/download
- Data format is based on the Essentia data extractor
  - https://essentia.upf.edu/streaming_extractor_music.html#music-descriptors


```
            5.51501289081e-10
        ]
    },
    "erbbands_crest": {
        "dmean": 3.08244657516,
        "dmean2": 4.69690227509,
        "dvar": 10.7596349716,
        "dvar2": 22.9963569641,
        "max": 33.8548240662,
        "mean": 13.2115125656,
        "median": 11.8160438538,
        "min": 2.76867485046,
        "var": 38.8526344299
    },
    "erbbands_flatness_db": {
        "dmean": 0.0205260887742,
        "dmean2": 0.0296374950558,
        "dvar": 0.000480096263345,
        "dvar2": 0.00105449254625,
        "max": 0.445085823536,
        "mean": 0.255228340626,
        "median": 0.2532556355,
        "min": 0.0331353694201,
        "var": 0.00267568021081
    },
    "erbbands_kurtosis": {
        "dmean": 1.92492461205,
        "dmean2": 2.92350172997,
        "dvar": 19.9219417572,
        "dvar2": 44.4127502441,
        "max": 76.6245803833,
        "mean": 2.09969305992,
        "median": 0.450933337212,
        "min": -1.78218781948,
        "var": 29.8014659882
    },
    "erbbands_skewness": {
        "dmean": 0.455330520868,
        "dmean2": 0.656669139862,
        "dvar": 0.281475663185
```

CAA

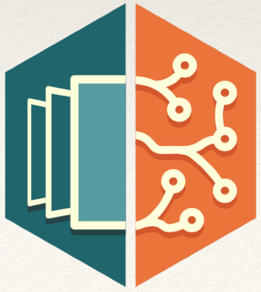
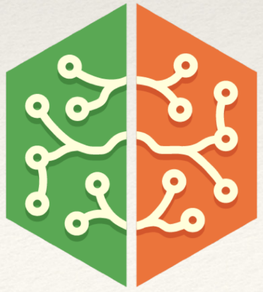
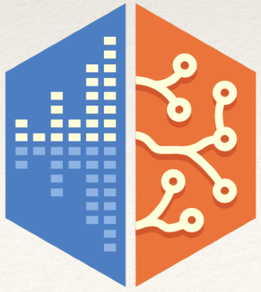
- https://coverartarchive.org/
- Open database of image cover art
- Collaboration between metabrainz and the internet archive (which stores the images)



Cover art archive data

- JSON API Available: https://musicbrainz.org/doc/Cover_Art_Archive/API
  - Given a release, get all associated images
  - Download the front image for a release
- Can also download the raw data as part of the main musicbrainz database
  - cover_art_archive.cover_art table

	id	release	comment	edit	ordering	date_uploaded
1	829358878	268158		17462125	1	2012-04-30 21:48:55.085782 +00:00
2	829428054	1152196		17462303	1	2012-04-30 22:00:14.293091 +00:00
3	829450710	1146044		17462335	1	2012-04-30 22:04:16.986456 +00:00
4	829455661	1146044		17462338	2	2012-04-30 22:04:43.137076 +00:00
5	829469212	1085462		17462380	1	2012-04-30 22:07:06.911130 +00:00
6	829475704	1068029		17462391	1	2012-04-30 22:08:20.872703 +00:00
7	829482774	765857		17462404	1	2012-04-30 22:09:09.357813 +00:00
8	829483276	486978		17462408	1	2012-04-30 22:09:43.573213 +00:00
9	829489979	1145207		17462420	1	2012-04-30 22:10:35.240668 +00:00
10	829493562	769730		17462442	1	2012-04-30 22:11:06.816147 +00:00
11	829493142	1078515		17462459	1	2012-04-30 22:11:41.062497 +00:00
12	829513566	657279		17462505	1	2012-04-30 22:15:12.024470 +00:00
13	829520403	1112876		17462511	1	2012-04-30 22:15:53.731375 +00:00
14	829524889	822030		17462531	1	2012-04-30 22:16:29.401351 +00:00
15	829525823	1112876		17462533	2	2012-04-30 22:16:34.306956 +00:00
16	829529534	775749		17462542	1	2012-04-30 22:17:11.007889 +00:00
17	829531636	822368		17462549	1	2012-04-30 22:17:30.342034 +00:00
18	2041988331	1061541		19062386	1	2012-09-18 06:14:27.655893 +00:00
19	829535118	1151956		17462560	1	2012-04-30 22:18:08.998556 +00:00
20	829537263	812067		17462561	1	2012-04-30 22:18:30.737246 +00:00
21	829539032	383024		17462563	1	2012-04-30 22:18:41.181773 +00:00
22	829540179	800665		17462589	1	2012-04-30 22:18:57.540274 +00:00
23	829541317	1151956		17462591	2	2012-04-30 22:19:03.490767 +00:00

Thanks

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