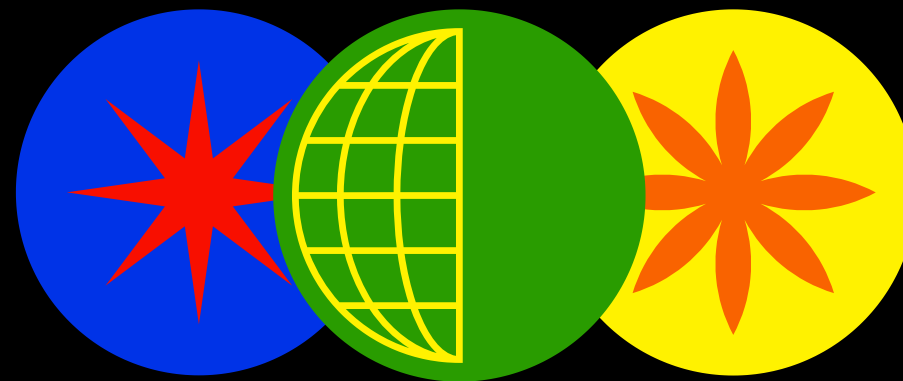


“What in the World?”

Mitigating Western Bias & Maximizing Music Retrieval in
KEXP’s “World Music” Catalog

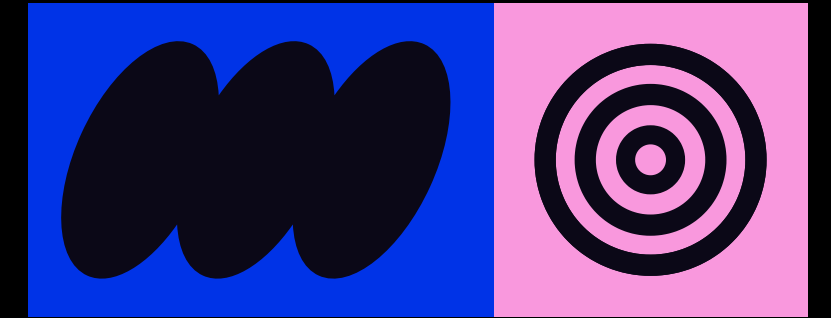


By: Nestor Guerrero, MLIS Graduate

Sponsoring Organization: KEXP

Sponsor: Afsheen Nomai, Media Asset & Archives Manager

Student & Sponsor Introduction



Nestor Guerrero , MLIS Graduate

- 2nd year residential student
- From Los Angeles, CA
- Professional interests: Archives, Metadata Librarianship, Digital Preservation, Cultural Heritage Stewardship



Afsheen Nomai, PhD

- Digital Asset & Archives Manager at KEXP
- Community Archiving Workshop (CAW) Committee Member
- Previously the Technical Director at the Texas Archive of the Moving Image



Information Problem

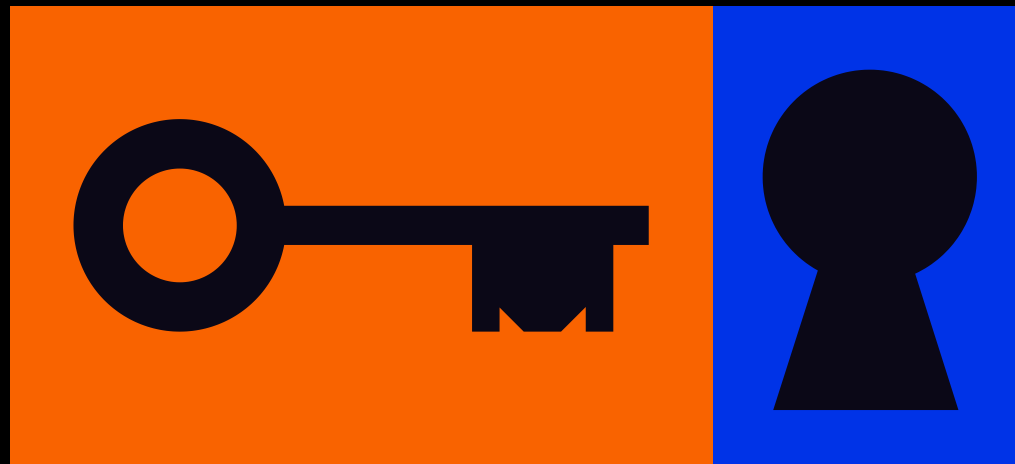
Addressing the limitations of KEXP's metadata classification scheme for their digital music library, particularly concerning their use of the "world" genre label within their music library and Digital Asset Management System, Dalet.

"[World music] is a marketing term created to help sell Third World musics"

--The American Folklore Society (AFS) Ethnographic Thesaurus

- Originally deployed in the 1980s to capitalize on the success of African artists and pander to the exoticized ideations of the "third world" constructed by Western audiences.
- Reinforces the dominance of Western popular culture by perpetuating an "otherness" onto non-Western music.
- Overlooks the diverse and nuanced nature of non-Western music traditions and cultural expressions from across the globe.

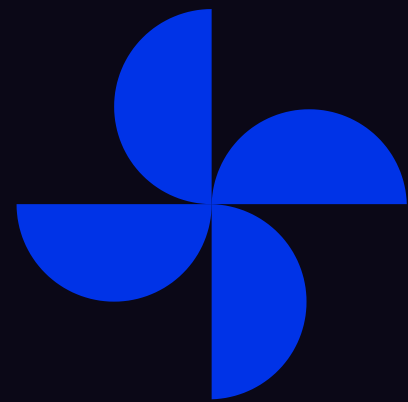
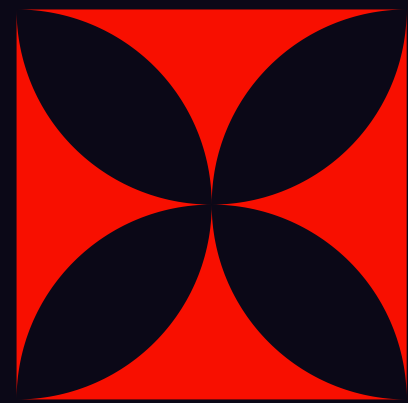
Project Questions



- What is the current state of KEXP's metadata for digital "world music" titles?
- How representative of the non-western world is the station's "world music" collection?
- What are the information needs of DJs when it comes to retrieving music titles for their shows?

Following my analysis, I will propose recommendations that KEXP should consider in implementing to improve the stewardship of music titles within their digital repository, Dalet.

Project Overview



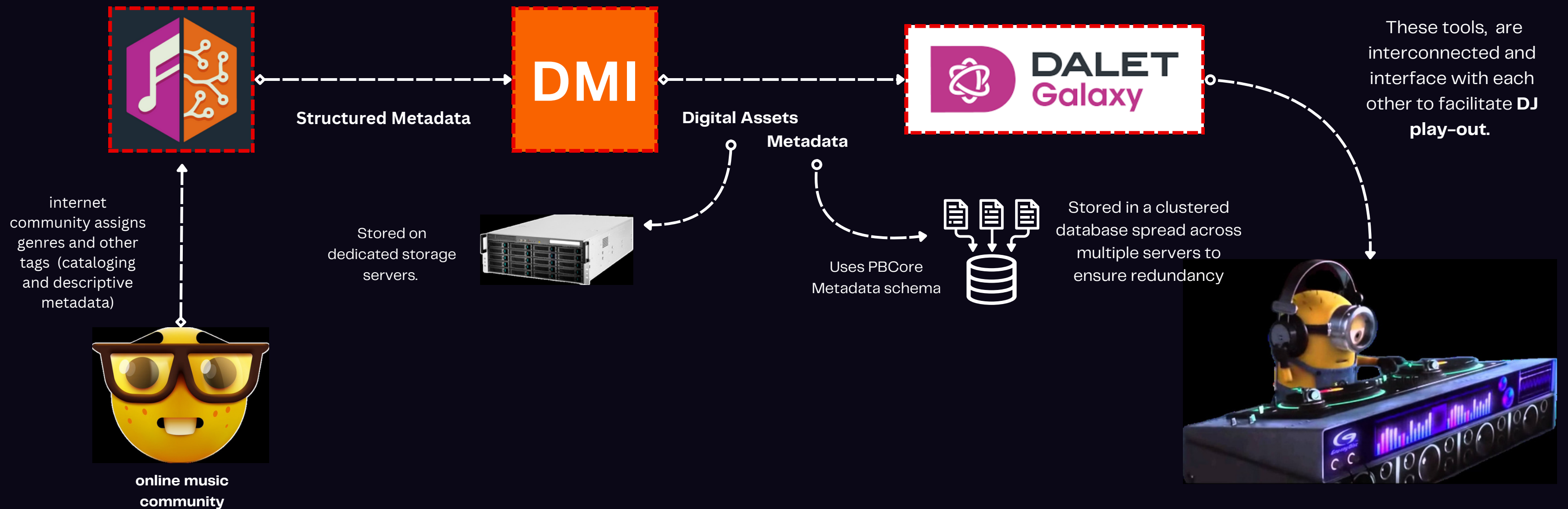
- 01 Initial Assessment**
Broad analysis and observations of KEXP's digital repository environment, descriptive metadata fields, and search + retrieval functionality within Dalet.
- 02 Geographic + Cultural Inventory (Diversity Audit)**
Randomly selecting a subset of digital assets (music) classified as "world"; manual metadata and description work; data analysis using R-Studio. Filling in contextual gaps: country/region of origin, ethnic and cultural context, genre/music style classification, original language of composition, release date, geographic area of the artist, and original script (non-romanized).
- 03 Stakeholder Analysis & Observation**
Conducting semi-structured interviews with KEXP DJs, with a focus on BIPOC radio hosts who center music outside of the western world in their shows. Noting observations volunteering as a DJ assistant for Wo'Pop (Global Music specialty show). Maintaining rapport with Digital Asset & Archives Manager.
- 04 Provide Recommendations**
Offer recommendations and considerations that KEXP's music library and archive can implement in future efforts to improve their current description and cataloging practices.

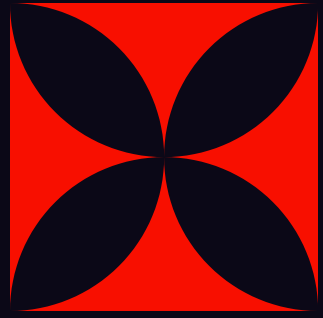
KEXP Music Data Environment

MusicBrainz functions as an open-source, music encyclopedia that relies on volunteers to contribute information about musicians and their releases.

DMI is an Ingest tool that extracts metadata (takes a snapshot) from MusicBrainz, which is then ingested into Dalet. Where KEXP's genre labels are assigned.

Dalet is a proprietary Media Asset Management System (MAMS) that serves as both a repository for digital assets and a media player for DJs to play out of in their shows. It is powered by a SQL database.





Initial Assessment – Dalet's Search Fields & Functionality

When searching for individual tracks, DJs can perform searches using the following fields:

- **Artist Name** [open text] -- #
- **Release Name** [open text] -- #
- **Track Name** [open text] -- #
- **Label** [open text] -- #
- **Gender** [cv / drop down] -- #
- **Genre** [cv / drop down] -- %
- **Sort Name** [open text] -- %
- **Artist Glossary** [cv / drop down] -- #
- **Artist Alias** [open text] -- ?
- **Group Member** [open text] -- #
- **Artist Begin Area (City)** [open text] -- #
- **Artist Country** [cv / drop down] -- #
- **Rotation Status** [conditional] -- %
- **Stars** (review rating) [drop down / range] -- %
- **Duration** [range] -- #
- **Original Release Date** [open text] -- #
- **Track Number** [open text] -- #
- **Obscurity Rating** [cv / drop down] -- %

-- Sourced from MusicBrainz

% -- Determined by KEXP

“Genre Label” Field

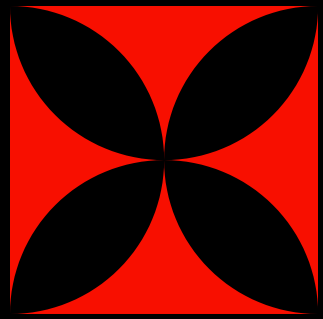
[controlled vocabulary / drop-down]:

Available options include Electronic, Experimental, Hip hop, Holiday, Jazz, Local, Reggae, Rock/Pop, Roots, Soundtrack, World, and “none specified”

“Artist Country” Field

[controlled vocabulary / drop-down]:

Features only 22 country options; 15 of which are European countries



Initial Assessment – Geographical Representation

Used R for data analysis (wrangling & manipulation)

- Provided with a CSV file of KEXP’s world music release glossaries, an export of all records (3,519 music titles) classified as “world” in the genre field of Dalet.
- Counted the number of times a particular country is listed under “**Artist Country**” for a release.
- Collocated all of the releases that have this information missing.
- **2,761 out of 3,519 release titles (78%) did not have “artist country” field filled out**

Artist.Country	total
???	2761
United States	90
Brazil	67
United Kingdom	59
France	41
Nigeria	31
Colombia	30
Argentina	25
Mali	20
Pakistan	20
Senegal	20
Spain	20
India	13
Peru	13
Belgium	12
Cuba	12
Germany	12
Mexico	12
South Africa	11
Canada	9
Chile	8
United States, United States	8
Japan	7
Niger	7
Turkey	7
Netherlands	6
Romania	6
[Worldwide]	6
Algeria	5



Geographic & Cultural Audit

- Engaged in a comprehensive audit/inventory for 200 of the 2,761 music titles missing “artist country” information (random selection) to gauge how much of the non-western world is represented in “world music” KEXP catalog.
- Enriched metadata and description work -- manually entered the following contextual information:
 - country/region of origin
 - ethnic and cultural context
 - genre / style
 - original language,
 - release date
 - geographic area of the artist
 - original script (non-romanized)

- Primary sources for gathering this information: Discogs, Bandcamp, Apple Music, Spotify, Soundcloud and the official websites of artists.
- Excluded “various artists” / compilation releases from sample



Ëda Diaz – Suave Bruta

Genre: **Electronic, Latin**

Style: **Afro-Cuban, Bambuco, Bass Music, Bolero, Vallenato, Leftfield, Broken Beat, Danzon, Cumbia, Downtempo, Latin**

Year: **2024**

Screenshot of release item record for “Sauve Bruta” by Ëda Diaz.

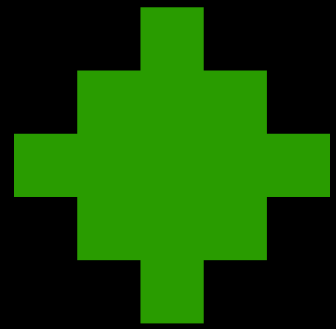
Source: Discogs



Geographic & Cultural Audit

Artist Continent	Artist Region	Artist Country/Territory	Artist Location Based in	Release Ethnic/Cultural Tags	Release Genre /Style Tags	Language of Composition
Asia	Siberia	Russia	Tuva	Russian; Tuvan	Overtone Singing; Russian Folk; Traditional Tuvan; Throat Singing	Tuvan
Africa	Sahara; West Africa	Mali; Algeria		Tuareg; Berber	African Folk; Desert Blues; Rock	Tamasheq
Asia	Central Asia	Uzbekistan	Asaka	Uzbek	Uzbek Folk	Uzbek
Africa		Zimbabwe	Mazowe	Zimbabwean	Zimbabwean Folk; Chimurenga	Shona
Africa	West Africa	Nigeria	Lagos	Nigerian	Hip Hop; Afrobeat; African Folk; Funk; Soul	English
North America	Caribbean	Cuba	London, UK	Cuban; Afro-Cuban	Jazz; Latin; Funk; Soul; Reggaeton	Spanish

Portion of spreadsheet tracking manually entered tags for randomly selected subset of music titles missing geo/cultural metadata. The updated dataset was analyzed using R-Studio to conduct further analysis.



Geographic & Cultural Audit – Countries & Territories

Tracked 67 countries and territories

Top 10

1	USA	25
2	Brazil	16
3	Mexico	8
4	Nigeria	7
5	South Africa	7
6	UK	7
7	Argentina	6
8	France	6
9	India	6
10	Mali	6

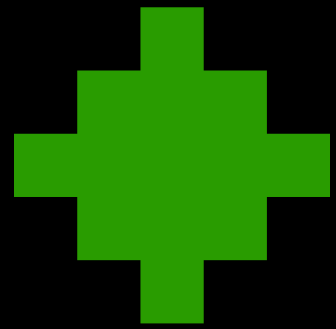
Denmark
Ghana
Guinea–Bissau
Hungary
Italy
Malawi
Mauritania
Palestine
Scotland
Senegal
Democratic Republic of Congo
Greece
Guinea
Israel
Japan
Lebanon

Netherlands
Niger
Pakistan (Punjab)
Romania
Switzerland
Syria
Venezuela
Sweden
Cuba
Colombia
Ethiopia
Indonesia
Algeria
Bulgaria
Egypt

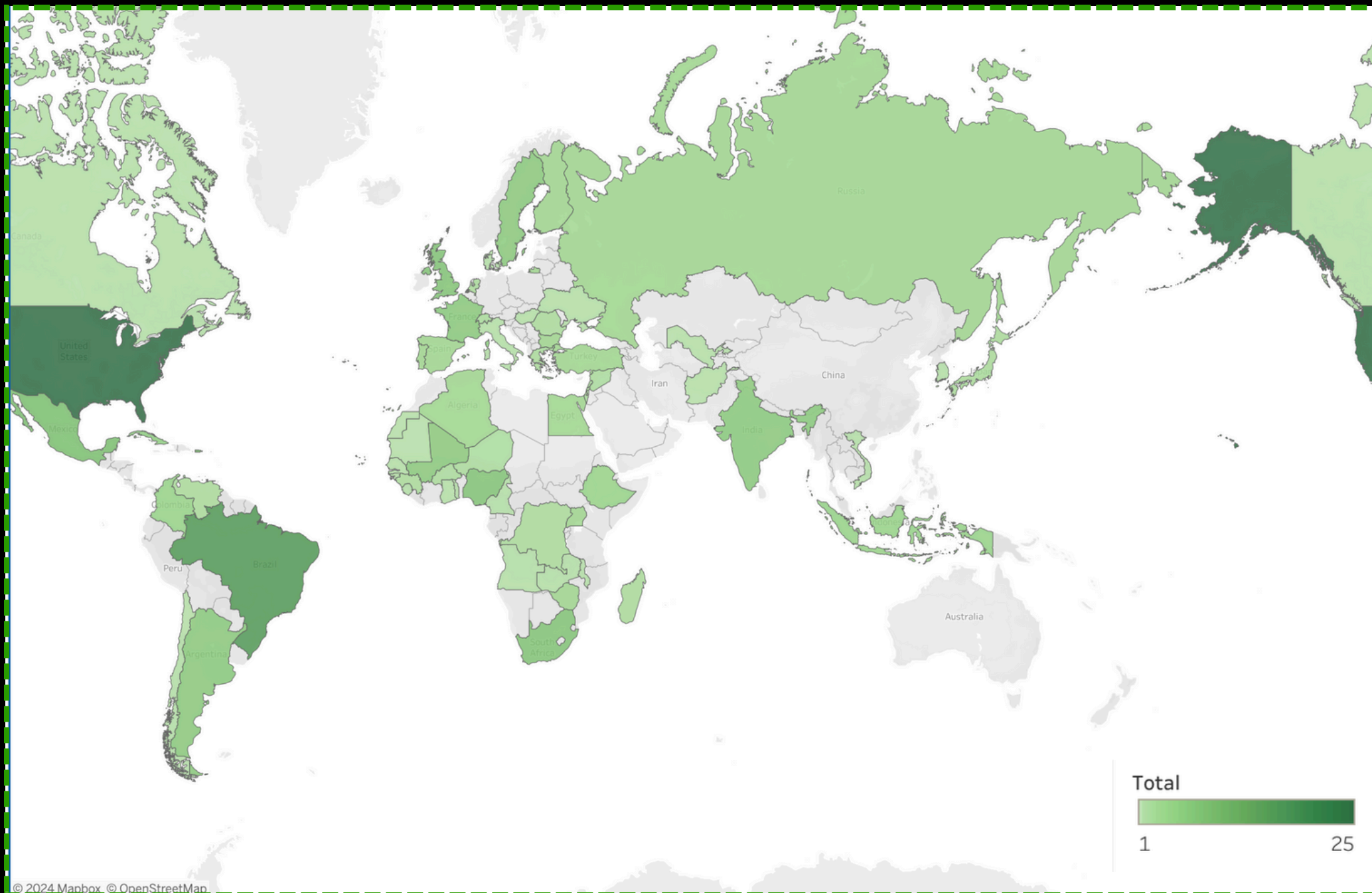
Hungary
Italy
Malawi
Mauritania
Palestine
Scotland
Senegal
Sierra Leone
Afghanistan
Angola
Cameroon
Canada
Cape Verde
Chile
Denmark
Ghana
Guinea–Bissau

Finland
Portugal
Puerto Rico
Russia
Spain
Turkey
Uganda
Zimbabwe
Burkina Faso
Sierra Leone
South Korea
Tibet
Togo
Ukraine
Uzbekistan
Vietnam

Western Sahara
Zambia
Madagascar



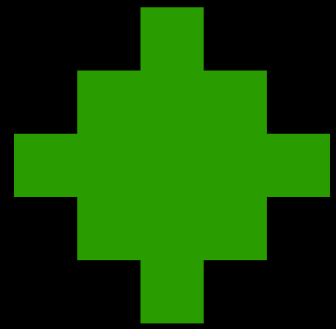
Geographic & Cultural Audit – Mapping the Data



Screenshot of world map data visualization (on Tableau) tracking the frequency of countries/territories represented in the random sample of music titles.

Artist.Continent	n
Africa	56
Europe	48
North America	42
Asia	32
South America	29
Oceania	2

Screenshot of dataframe counting the most represented continents in random world music sample on R-Studio.



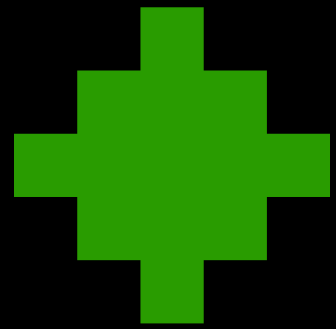
Geographic & Cultural Audit – Ethnic/Cultural Tags

Came up with over 141 unique tags to denote ethnic/cultural groups and communities.

Top 12

Brazilian	16
Mexican	9
West African	9
Nigerian	8
Afro-Cuban	6
Swedish	6
Arab	5
Argentinian	5
Cuban	5
French	5
Indian	5
Middle Eastern	5

Russian	Afro-Caribbean	African-American	Hungarian	Ukrainian
South African	American	Afro-Colombian	Iranian-American	Venezuelan
Arabic	British	Algerian	Israeli	Yoruba
Congolese	Bulgarian	Angolan	Japanese	Zimbabwean
Ethiopian	Egyptian	Balinese	Latinx	Abayudaya
Finnish	Guinean	Berber	Lebanese	Afghan
Griot	Latin	Burkinabé	Malagasy	Afro-Latin
Jewish	Portuguese	Cape Verdean	Midwest	Afro-Portuguese
Malian	Puerto Rican	Chilean	North African	Afro-Puerto Rican
Spanish	Romanian	Colombian	Pakistani	Afro-Venezuelan
Turkish	Senegalese	European	Syrian	Akan
Afro-Brazilian	Tuareg	Greek	Ukrainian	Andean

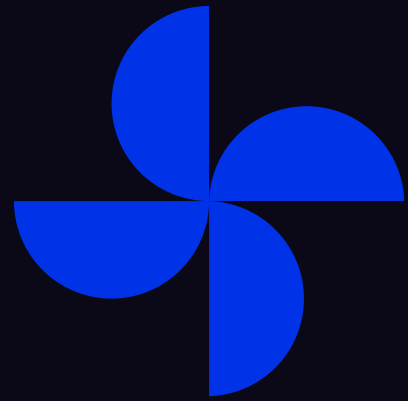


Geographic & Cultural Audit – Genre/Musical Style Tags

Came up with over 292 different genres and musical styles, many of which are unique to specific regions, cultural communities, and ethnic groups across the globe.

	Release.Genre..Style.Tags	n
1	Latin	50
2	Jazz	31
3	Electronic	26
4	Rock	18
5	Funk	15
6	Soul	15
7	Pop	12
8	Afrobeat	11
9	MPB	9
10	Acoustic	7

Axé
Baile Funk
Baka music
Balkan Music
Bambuco
Bass Music
Bebop
Big Band
Brazilian Funk
Breakbeat
Broken Beat
Cape Jazz
Carnatic
Carnavalito
Bossa Nova
Cumbia
Indian Classical
Hip Hop
African
Ambient
Downtempo
Electro
Flamenco
Reggae
Son
Stage & Screen
Bolero
Desert Blues
Folk (Swedish)
Folk (West African)
Folk-Rock
Klezmer
R&B
Bhangra
Bomba
Celtic
Charanga
Chimurenga
Contemporary
Dancehall
Danzon
Dub
Afrobeats
Blues
Classical
Contemporary Jazz
Country
Experimental
Fado
Folk
Folk (African)
Electronic Folk
Folk (Algerian)
Folk (Balinese)
Folk (Finnish)
Folk (Greek)
Folk (Malian)
Folk (Portuguese)
Folk (South African)
Folk (Ugandan)
Free Jazz
Folk (African)
Folk (Brazilian)
Folk (Bulgarian)
Folk (Ethiopian)
Folk (Nigerian)
Folk (Nordic)
Folk (Spanish)
Highlife
Percussion
Psychedelic Rock
Samba
Soul-Jazz
Synth-pop
Acid
African Pop
Afro-Cuban
Afropop
Alternative Rock
Amapiano



Stakeholder Analysis – DJs



Kevin Sur

DJ/Radio host for “Sounds of Survivance,” a show centered on a of Indigenous music and artists from across the globe.



Darek Mazzone

DJ/Radio co-host for “Wo’ Pop,” a speciality show centered on exploring global (“world”) music.



Jyoti B. Fly

DJ/Radio co-host for “Wo’ Pop,” a speciality show centered on exploring global (“world”) music.



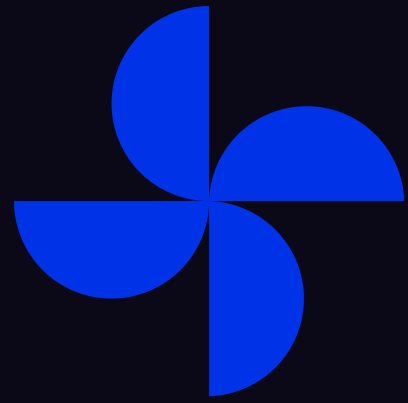
Albina Cabrera

DJ/Radio host for “El Sonido,” a show focused on Latin music and culture from South, Central and North America, as well as Spain and the rest of the globe.



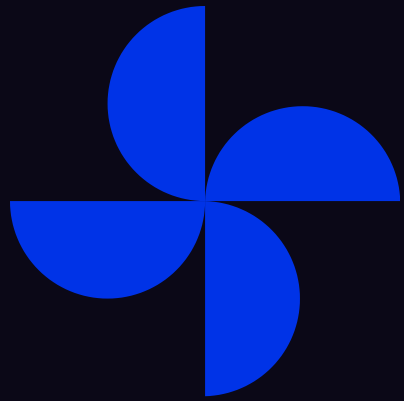
Diana Ratsamee

DJ/Radio host for “Eastern Echoes,” a show focused on the vast realm of music from Asia and the Asian diaspora.



Stakeholder Analysis – Questions for DJs

- How do you organize your personal music library?
- How do you discover music to play on your show?
- Can you describe the typical process when selecting / curating music for your show?
- When in search of new music, what specific information do you look at?
 - (i.e. label, release country of release, country, language, genre, etc)
- What genres do you tend to explore the most?
- How do you stay updated on new music or emerging artists?
 - Are there any specific sources or platforms you rely on for this information?
- When navigating through a large music database like Dalet, what additional information would help streamline your search process?



Stakeholder Analysis - DJ Testimonies

"I use the global map to organize playlists; I separate it by broad region; I do that to make sure that I'm never too heavy handed in one part of the world."

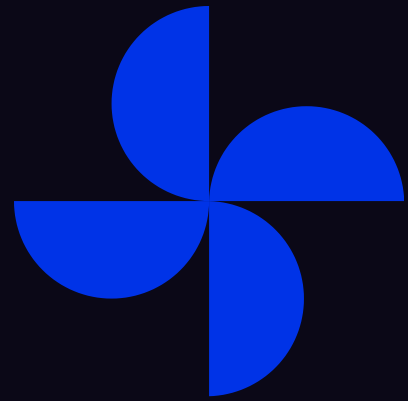
"You can only look for what you know. It would be great if I had a way to find music from non-european languages or Indigenous languages...that would be really dope and something I would use--if it could search by language"

"Back in the day, everyone worked in the same library...it is a bit siloed now because everyone has their own laptops and personal libraries"

"KEXP internally has created its own algorithm...my fear of this being used so heavily is that KEXP can become repetitive"

"[My show's] intent is that people from different regions feel represented (especially the most disenfranchised)...my show should touch all of those (regions) and more, it is important to me. That helps me make sure that I'm not forgetting different parts of the world...Even though I don't agree with the term [world music], I still have that responsibility."

"You have to know the genre and place of where a sound came from to know how to categorize it"



Stakeholder Analysis – Digital Asset & Archives Manager

“We're relying on another database [MusicBrainz] for our metadata...we're saying they are our authority. All the [musician/artist] disambiguation, they're aliases; all that stuff comes from there. We can enter our own aliases and that's fine, but that's the first stop. And since we don't have catalogers, we rely on DJs...and they don't even have the ability to make changes to the database. They can only search and play. That's all they can do.”

“The music directors are really the only ones who are doing any kind of 'real' cataloging through their reviews of the music that goes into the library. Nothing else....Also, because we rely on the MusicBrainz database, that's really where a lot of cataloging should be happening, but it's not....Just to give an example, we gathered a ton of Bay Area music right before the cut over. I begin ingesting the tracks into the library and tried to refer to MusicBrainz but there's so many empty fields.”

“Dalet is a resource intensive system to maintain”

I think awareness is the key—a more nuanced understanding of the library and what's in here and how it could benefit them in whatever department they're in, you know, not just for the DJs....It isn't just music. It's also the live stuff. It's also the photos. It's also the institution”

Release information
Barcode: 031257141529
Format: (unknown)
Length: 52:04

Additional details
Type: Album
Packaging: Jewel Case
Status: Official
Language: English
Script: Latin
Data quality: Normal

Labels
■ Hannibal
HNCD 1415

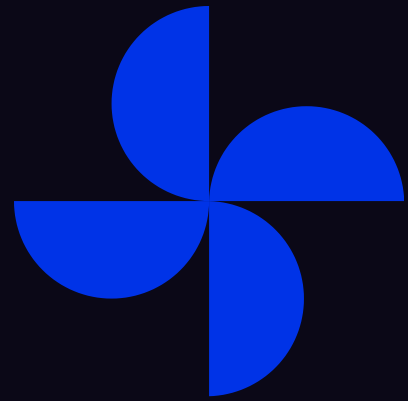
Release events
■ United States
1998-04-28

Tags

Genres
(none)

Other tags
(none)

Screenshots of MusicBrainz metadata for one of the randomly selected KEXP “world music” titles.



Stakeholder Analysis – Key Insights

MusicBrainz Limitations

- As our authority source for contextual information on musicians and records, MusicBrainz has a lot of gaps and empty metadata fields.

DJ Information Needs & Desires

- Interests in exploring AI-assisted search and having a database/MAMs that has a DJ-centered + informed UI/UX
- Desire for Dalet to have mobile functionality
- More access permissions in managing digital assets in Dalet.

Perpetuation of Inaccurate/Inconsistent Metadata

- Metadata for “artist country” reflects location of the label rather than geo-cultural information about the artist themselves.
- Metadata inconsistencies across generations of DJs left unaddressed
- Unclear labeling of alternative versions of songs (e.g. “Live”, “Instrumental”)

Dalet’s Utility and Functionality

- Primarily used for retrieval, rarely for browsing or discovery.
- No mechanism in place for maintaining authority control
- Inability to recognize “foreign”, accented, non-European characters and scripts
- Current music ingest and processing workflow of digital assets is not conducive for collaborative curation or stewardship

Genre Labeling and Catalog Organization

- There exists a “goldilocks” level of genre labels between specificity and broadness when it comes to genre. Depending on the size of the catalog, having parent and child genres can be helpful. But first, KEXP needs to “know” (better describe and document) the metadata of its content



Recommendations + Potential Solutions

01 Metadata

[*] Adding a search functionality for scripts

Adding a new metadata field for “**Script**” to specify the script used to write the release’s track list. MusicBrainz has [a guide](#) listing some common scripts: Latin, Arabic, Cyrillic, Greek, Han (simplified and traditional), Korean, Hebrew, Japanese, Katakana, Thai. Integrating this attribute as a new metadata field, and potentially as a search filter option, can help end-users find and browse music titles based on the original language scripts they were written in.

[*] Adding “Ethnic” or “Cultural” Tags for releases

There exists an abundance of contextual, and useful, information for DJs and listeners that is not being tapped into. The station should always be uplifting and using its platform to uplift artists of different identities, cultures, communities, regions; introducing tags was an open text field (or could also be a structured entry) can help.

[*] Encouraging the cultivation of folksonomies (of structured metadata) to maintain authority control

Use **parenthetical disambiguation** (when appropriate) in the artist country field to denote a specific geographic region of a country.

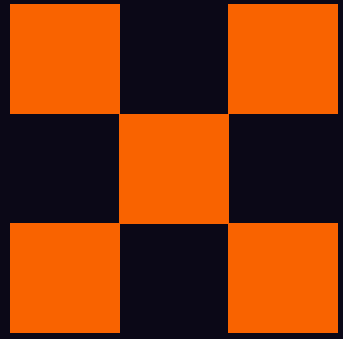
- Ex. Pakistan (Punjab)
 - Scope Note: Punjab a state in northern India. It forms part of the larger Punjab region of the Indian subcontinent.

Incorporate **modifiers** to provide more specificity when using broader genre terms like, “Folk”, “Classical”, “Rock”, “Traditional”, “Contemporary”, “Alternative”

- Ex. Folk (Malian)

[*] Adding a “reference/citation” field

Where users who ingest and catalog the file to add a link of the source where they gleaned information relating to an artists’ cultural, ethnic, and musical identity for transparency.



Recommendations + Potential Solutions

02 Digital Asset Management System

[*] Implement Nuanced Access/Management Permissions & Controls

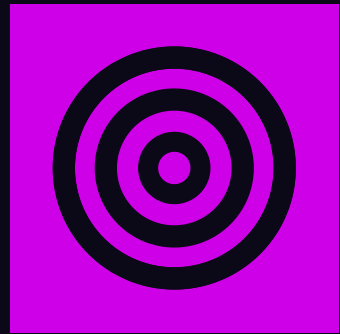
Providing DJs with remote access to upload music to Dalet using the DMI. This would encourage the practice of direct music ingest into KEXP's internal music database, further allowing it to expand. It is not practical for DJs to only have access in stations to be able to do this process.

[*] Migrating to a digital asset management system that can provide a linked data environment and the designation of authority records for artists/musicians and descriptive tags

Identity-based management systems that utilize linked data can serve as a solution for many of the challenges and difficulties associated with authority control for names in non-Latin scripts.

[*] Creating an API that can harvest metadata from the DJ playlist tool and interface with Dalet.

There is a treasure trove of contextually rich metadata for music logged by DJs and their assistants using the DJ playlist tool (specifically the open text description field). It could be advantageous to develop an API that can relate/connect the open text/description field metadata to KEXP's music database (at the track level).



Future Questions & Considerations

How are we preventing the replication of colonial narratives / notions of people and places in our metadata?

What are considerations for cataloging and developing metadata for aggregates (compilation albums or “various artists” releases)?

How can we adequately capture the documentation of contextual geo-cultural metadata for DJs and artists who produce edits and remixes of songs created by other artists in the database/digital asset management system?

How can we find a balance between reparative metadata work while also establishing a robust foundation for creating/out-sourcing new metadata for music that is being ingested now/in the future?

What processes can be automated? What processes should be done manually?

How can we better connect the ongoing efforts and work being done by KEXP’s cultural community liaisons and relational managers to the stewardship of the station’s digital music library/catalog ?



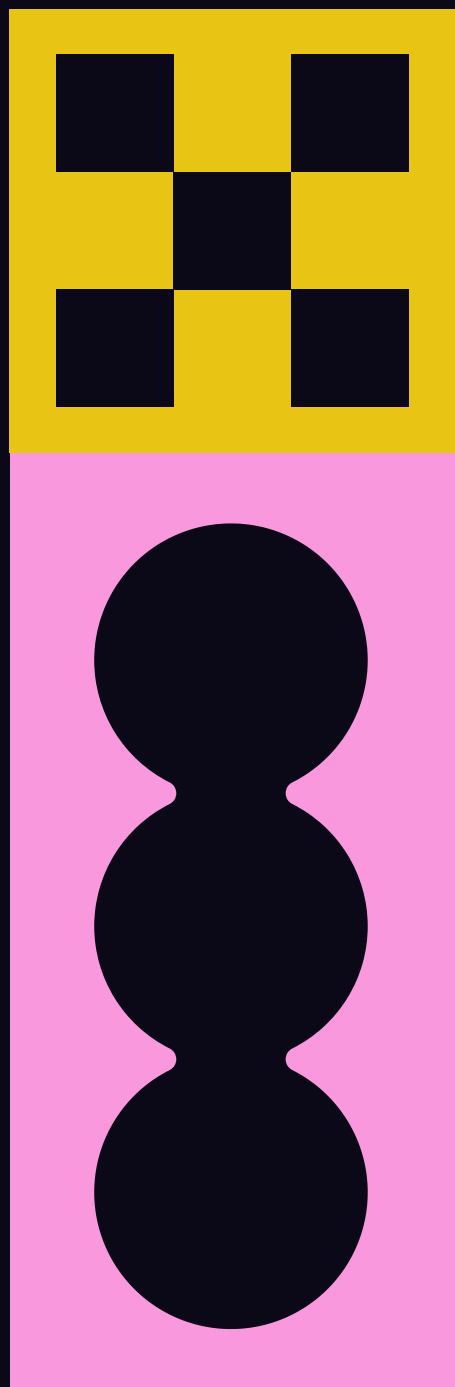
Project Evaluation – Objectives & Limitations

Objectives Met:

- Enriched metadata for 200 titles in KEXP’s “world music” collection
- Gathered preliminary insights through data analysis of sample
- Interviewed/consulted with 5 KEXP DJs/Radio hosts
- Grew familiarity with information retrieval process and data repository environment through my role as a DJ assistant + metadata consultant
- Raised awareness of issue and drafted recommendations to stakeholder community

Limitations:

- Constraints of time limited more in-depth analysis of rich collection
- Absence of other stakeholder perspectives: DJ assistants, music directors, and other radio hosts
- Small sample size of world music titles
 - (200/3,519) is 5.6% of collection



Conclusion

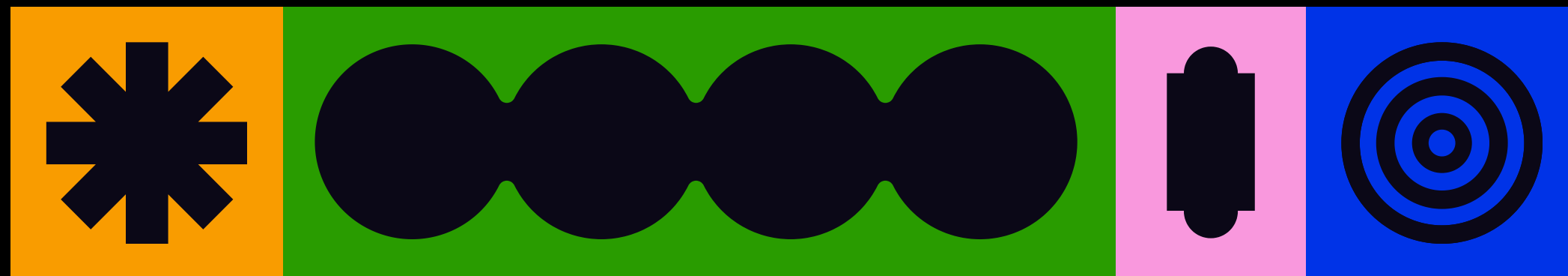
This effort is more about finding innovative ways to expand linkages and webs of relationships within KEXP's digital assets (particularly music files) rather than creating rigid distinctions and classifications for what specific genre/label a song should fall under.

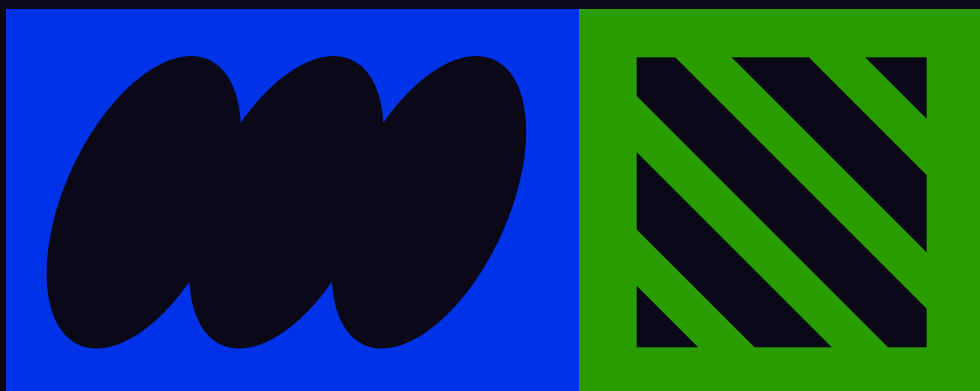
Having robust and consistent metadata standards and practices when it comes to contextualizing and cataloging the music that KEXP plays on its airwaves, and ultimately ingests into its digital asset management system, can bring the station closer towards being in alignment with, and acting on, their mission, vision, and values.

Next Steps

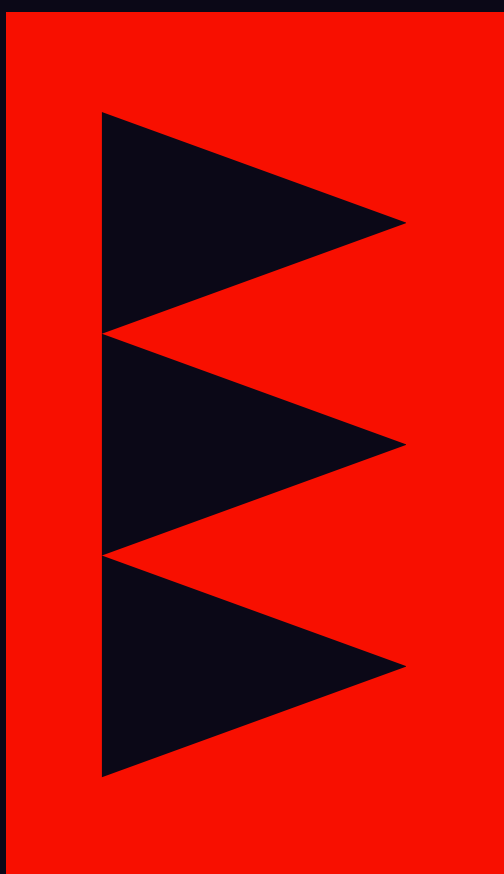
I'll be raising awareness about this issue by presenting my findings and recommendations to the stakeholder community.

I'll be sharing my project insights on June 20th with KEXP managerial staff and DJs during their monthly "Lunch & Learn" series at the station. The presentation will also be recorded and live-streamed via Zoom.





Project Experience



I extend my gratitude to KEXP, with special thanks to Afsheen, for making this project possible and for their invaluable guidance throughout. This experience has allowed me to push my boundaries as an emerging digital archivist and information professional. Working on a topic that I'm deeply passionate about has been truly enriching. I've gained valuable insights in the realms of cataloging, digital repositories, media asset management, contextual metadata, and data stewardship. Most importantly, I've had the privilege of connecting with remarkable individuals within this local community that are tied together by their shared love for music and commitment to uplifting artists and creatives.

Acknowledgements

A BIG thank you to KEXP and everyone whose support and guidance have been instrumental throughout this project.

Afsheen Nomai

Jyoti B. Fly

Darek Mazzone

Kevin Sur

Albina Cabrera

Diana Ratsamee

Mathilde Xiao

Franny Thomas

Katie Atkins

Ishita Chordia

John Vallier

Lilium Rajan

Melanie Walsh

Christian Wright

Rita Tesfai

Cedric Duyck

Jessie M. Garcia Guttierrez

Nela

Maura Guerrero

Mario Guerrero

Melina Guerrero

