LinkedMusic

Project Meeting IV

13 December 2025

Ichiro Fujinaga

Music Technology Area, Schulich School of Music

McGill University

LinkedMusic: Key Concepts

- Be able to search across various music databases from one website
- Universal Music Instrument Lexicon (formerly known as Virtual Instrument Museum)
  - Create a crowd-sourced website
  - Images and recordings of musical instruments
  - Name of the instrument in the local language
  - Basically a front-end for displaying and editing musical instrument data on Wikidata
- Funded for 7 years (2022–2029): $3.2M
  - SSHRC Partnership Grant
  - FRQSC Research Team Support Grant
  - McGill University

LinkedMusic: Co-Investigators (7)

- Jennifer Bain (Dalhousie University)
- Houman Behzadi (McGill)
- Julie Cumming (McGill)
- Debra Lacoste (Dalhousie University)
- Audrey Laplante (Université de Montréal)
- Cory McKay (Marianopolis College)
- Laurent Pugin (RISM-Digital)

LinkedMusic: Collaborators (18)

- Rachelle Chiasson-Taylor (Library and Archives Canada)
- Julia Craig-McFeely (Oxford University)
- Jürgen Diet (Bavarian State Library)
- Robin Desmeules (McGIll)
- Simon Dixon (Queen Mary, University of London)
- Jon Dunn (Indiana University)
- Andrew Hankinson (RISM Digital)
- Johannes Kepper (University of Paderborn)
- Kevin Kishimoto (Stanford University)
- David Lewis (Goldsmiths, University of London)
- Jonathan Manton (Yale University)
- Kevin Page (University of Oxford)
- Alastair Porter (UPF / MetaBrainz)
- Jenn Riley (McGill)
- Patrick Savage (Keio University)
- David Weigl (University of Music and Performing Arts Vienna)
- Susan Weiss (Johns Hopkins University)
- Frans Wiering (University of Utrecht)

LinkedMusic: Partners (9)

- Bavarian State Library (Jürgen Diet / Bernhard Lutz)
- British Library
- Calcul Québec
- Dalhousie University (Jennifer Bain)
- MetaBrainz Foundation (Alastair Porter)
- RISM Digital (Laurent Pugin)
- Université de Montréal (Audrey Laplante)
- University of Oxford (Julia Craig-McFeely)
- University of Waterloo (Debra Lacoste)

Advisory Board Members

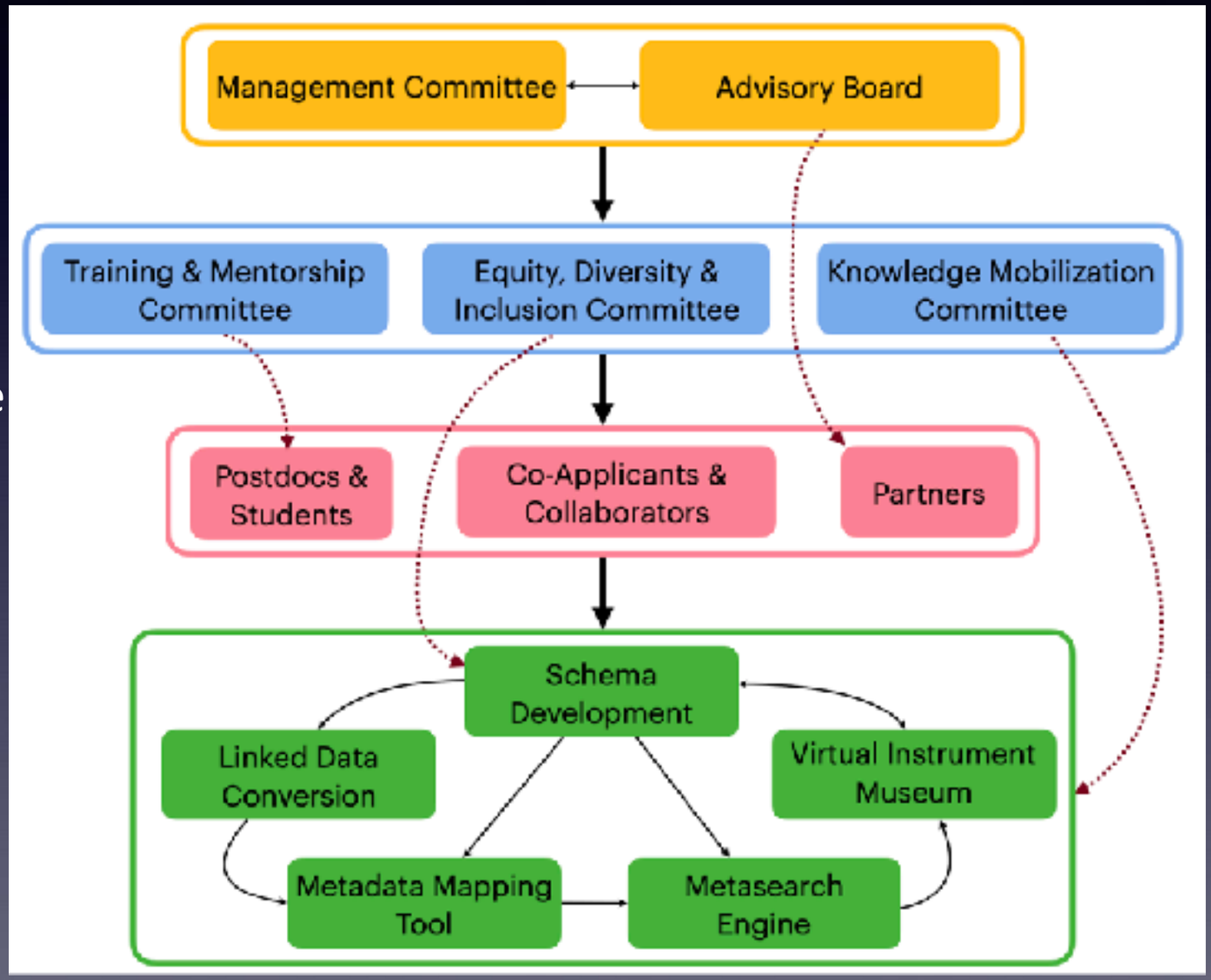
- Tina Frühauf (RILM: Répertoire International de Littérature Musicale) (Barbara Dobbs McKenzie (retired))
- Perry Roland (University of Virginia)
- Rob Sanderson (Yale University)
- Xavier Serra (University of Pompeu Fabra)

People

- Project Manager
  - Vi-An Tran
- Postdoctoral Fellows
  - Anna de Bakker
  - Junjun Cao
- Graduate Students
  - Hanwen Zhang
  - Kyrie Bouressa
  - Kun Fang
  - Zih-Syuan Lin
  - Lucas March
  - Liam Pond
  - Yu-Chia Kuo
  - Mai Lyn Puittinen
  - Caroline Guo
  - Linnea Kirby
  - Pouya Mohseni
- Undergraduate Students
  - Sichen Meng
  - Antoine Phan
  - Sebastien Chow
  - Simon Ngassam
- Researchers
  - Dylan Hillerbrand
  - Geneviève Gates-Panneton
  - Yinan Zhou

LinkedMusic Committees

- Training and Mentorship Committee
  - Chair: Anna de Bakker
  - Jennifer Bain
  - Laurent Pugin
  - Hanwen Zhang
- Equity, Diversity, and Inclusion Committee
  - Chair: Julie Cumming
  - Houman Behzadi
  - Robin Desmeules
  - Lucas March
- Knowledge Mobilization Committee
  - Debra Lacoste
  - Susan Weiss
  - Julia Craig-McFeely
  - Kyrie Bouressa



Agenda for Today

- 09:00–09:45 Introductions (Ichiro Fujinaga)
- 09:45–10:30 LinkedMusic Update (Liam Pond)
- 10:30–11:00 Coffee Break
- 11:00–11:30 Interactive NLQ2SPARQL agent (Junjun Cao)
- 11:30–12:00 Bernard Lutz: The musiconn project
- 12:00–12:30 An update on Linked RISM (Andrew Hankinson and Laurent Pugin)
- 12:30–13:30 Lunch (provided)
- 13:30–14:30 UMIL update (Kun Fang, Mai Lyn Puittinen, and Kyrie Bouressa)
- 14:30–15:30 Integrating RISM API as a Linked Data: Introduction (Kevin Page, David Lewis, and Andrew Hankinson)
- 15:30–16:00 Coffee Break and Committee meetings
- 16:00–17:00 Integrating RISM API as a Linked Data: Continuation

- 19:00 Dinner at Kim Fung (1111 Rue Saint-Urbain, 2nd floor)

LinkedMusic Project Meeting III

26 October 2024: McGill University, Montreal

LinkedMusic Workshop VI

2 June 2025: St George's, University of London

at the Music Encoding Conference

LinkedMusic Workshop VII

6 July 2025 in Salzburg, Austria at the IAML Congress

Papers published in the past year

- Fujinaga, Ichiro. 2025. “Integrating Online Music Databases: The LinkedMusic Project.” *Musical Brainfood, “Musicology in the Age of Artificial Intelligence, Part 2,”* 19 (1).
- Fujinaga, Ichiro. 2025. “Sustainable Archiving of Music Databases through RDF and NLQ2SPARQL.” In *Proceedings of the 12th International Conference on Digital Libraries for Musicology,* 100–104. Seoul, Korea.
- Pond, Liam, and Ichiro Fujinaga. 2025. Teaching LLMs music theory with in-context learning and chain-of-thought prompting: Pedagogical strategies for machines. In *Proceedings of the 17th International Conference on Computer Supported Education,* 1: 671–681.

Presentations (1)

- Sixth Joint Meeting Acoustical Society of America and Acoustical Society of Japan, Honolulu, Hawaii, 3 December 2025. *LinkedMusic Project: A Progress Report.*
- *1st Workshop on Large Language Models for Music & Audio (LLM4MA),* Seoul, Korea, 26 September 2025. ”SESEMMI for LinkedMusic: Democratizing Access to Musical Archives via Large Language Models." By Liam Pond with Linnea Kirby, Sichen Meng, Simon Ngassam, Sebastien Chow, Dylan Hillerbrand, and Ichiro Fujinaga.
- International Society for Music Information Retrieval Conference: Late-breaking Demo Session, Seoul, Korea, 25 September 2025. *Integrating Music Databases with Linked Data* with Junjun Cao, Kyrie Bouressa, Hanwen Zhang, Kun Fang, Liam Pond, Yu-Chia Kuo, Mai Lyn Puittinen, Caroline Guo, Hong Van Pham, Yueqiao Zhang, Antoine Phan, Linnea Kirby, Sebastien Chow, Simon Ngassam, Sichen Meng, Geneviève Gates-Panneton, Yinan Zhou, Dylan Hiller-brand, Andrew Hankinson, and Anna de Bakker. Poster.
- Forum for Information Technology (FIT) 2025, Hokkaido University of Science, Sapporo, Japan, 4 September 2025. *LinkedMusic Project: Integrating Music Databases.*
- International Association of Music Libraries, Archives and Documentation Centres (IAML), Salzberg, Austria, 9 July 2025. *LinkedMusic Project: Integrating Online Music Databases* with Junjun Cao.

Presentations (2)

- Twenty-Third International Conference on New Directions in the Humanities, Hilo, Hawaii, 26 June 2025. *LinkedMusic Project: Integrating Music Databases.*
- Music Symposium 2025, Waseda University, Tokyo, Japan, 13 June 2025. LinkedMusic Project: Integrating Online Music Databases. Poster.
- Digital Humanities Summer Institute, Université de Montréal, 26 May 2025. Plenary Talk: *On the Virtues of Lazy Machines.*
- CIRMMT Workshop: Societal, Creative and Research Impacts of Data-driven Approaches to Music, McGill University, 18 March 2025. *The LinkedMusic Project.*
- Vitrine HN / DH Showcase 2025, McGill University, 17 January 2025. *LinkedMusic Project: Integrating Online Music Databases* with Junjun Cao.
- The 1st Workshop on Utilizing AI/ML to Enhance Information Extraction, Organization, and Retrieval from Large-scale Archival Collections, Hong Kong University, Hong Kong, 20 December 2024. *Sustainable Archiving of Music Databases through RDF and NLQ2SPARQL Frameworks* with Junjun Cao.

Midterm Report to SSHRC: Submitted!
2025-10-31
Anna de Bakker

Kyrie Bouressa and Debra Lacoste

Products	Number Planned (in Milestone Report)	Number Developed	Number Planned (for second half)
Presentations	5	93	50
Interviews (broadcast or text)	0	0	0
Peer-reviewed journal articles (open access)	3	19	10
Peer-reviewed journal articles (subscription based)	2	0	2
Edited journal issues	0	0	0
Books (including edited books)	0	0	0
Book chapters	1	3	1
Entries (dictionary and encyclopedia)	0	0	0
Conference publications	5	34	17

The Persephone Initiative
Anna de Bakker and Kyrie Bouressa

- Digital Research Alliance of Canada (DRAC) Grant
- Special Funding Opportunity: Research Software AI
- Announced 2025-10-06
- Eligibility: Must have a DRAC account
- Submitted 2025-11-14
- Issue Award Notification: 2026-01-15
- If awarded: 2026-04-01 to 2027-03-31 ($121,256)

LinkedMusic.ca

Google Translated to: Chinese (Simplified) Show original Options ▼

链接音乐 关于 參與者 活动 参考 密尔沃基大学 机会 联系我们

Powered by Google Translate

LinkedMusic

我们新成立的 LinkedMusic Partnership 的目标是通过元数据架构（用于组织数据库中存储的信息的结构）链接音乐数据库。这将大大有助于将在线音乐搜索提升到与当前基于文本的资源相同的复杂程度，使我们能够回答有关音乐以及音乐如何与人类创造力、社会、文化和历史互动的基本问题。

我们的目标包括：

- 制定全面的国际音乐元数据模式
- 开发映射工具，将现有的元数据模式映射到我们的新模式上
- 索引聚合元数据，以便同时搜索各个数据库

如需了解更多信息，请联系Ichiro Fujinaga。

L'objectif de notre nouveau partenariat LinkedMusic est de lier des bases de données musicales par des schémas de métadonnées, des structures d'organisation de l'information stockées dans une base de données. Ce projet contribuera grandement à amener la recherche de musique en ligne au même niveau de sophistication que la recherche de ressources textuelles, nous permettant ainsi de répondre à des questions fondamentales sur la musique et comment elle interagit avec la créativité, la société, la culture et l'histoire humaines.

Nos objectifs incluent :

Goals of LinkedMusic Project

- Make musical information accessible to more people in the world
- Be able to search across various music databases from one website
- Make musical queries available in languages other than English

Challenges in Integrating Databases

- Schema (how data is structured) mismatch
  - Different headings, e.g., Songs vs Tracks
- Semantic inconsistencies
  - Artist vs Performer
  - Genre differences: What is “Folk music”?
- Integration complexities
  - Determining matching records
  - “Ludwig van Beethoven” vs “Beethoven, Ludwig van”
  - Who is “Parker”?

UMIL (Universal Musical Instrument Lexicon)

Formerly known as VIM (Virtual Instrument Museum)

- Name of musical instruments may be needed for query
- Music instrument names varies across languages and cultures
- A way to translate musical instrument names in as many language as possible
- User-friendly interface to populate musical instruments in Wikidata

VIM Home Instruments About

Search...

Login

Web page language
Select Language

Hornbostel-Sachs Classification

1 - Idiophones	43
2 - Membranophones	9
3 - Chordophones	61
4 - Aerophones	61
5 - Electrophones	4
Unclassified	84

More facet search coming...

INSTRUMENT LIST English

Instrument name language

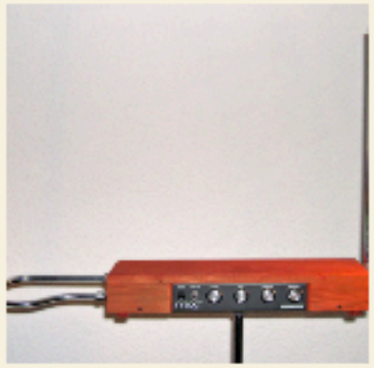
Tambourine

Triangle

Zill

Whistle

Vibraphone

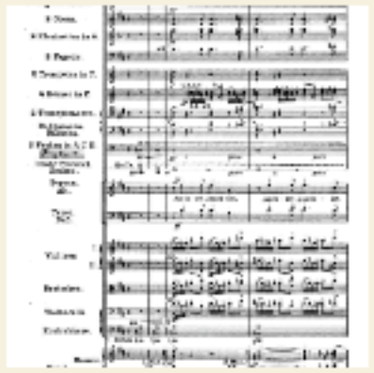


Theremin

Hurdy-Gurdy

Snare Drum

Tabla



Transposing Instrument

Pan Flute

Marimba

Initial 14 Databases

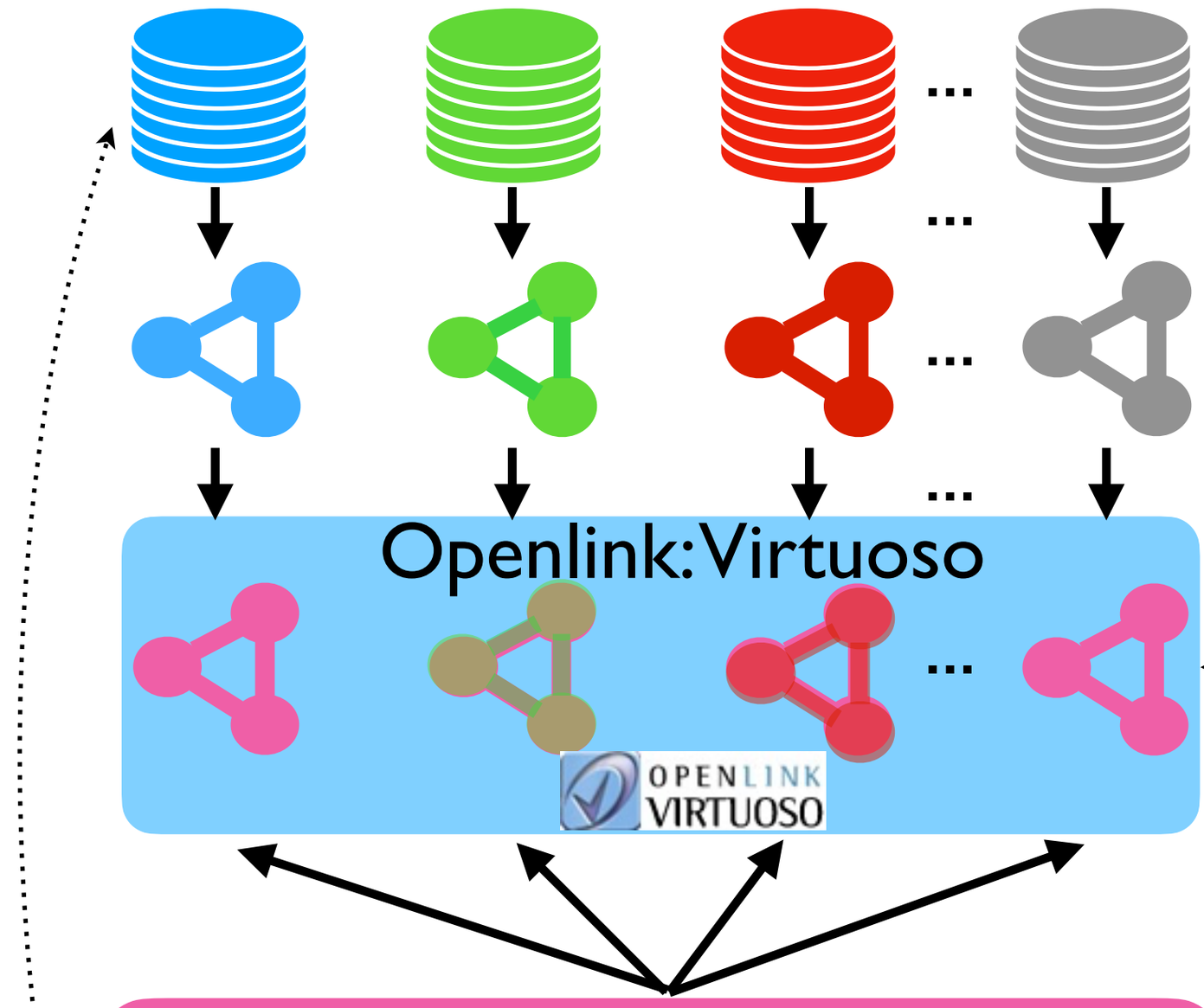
1. SIMSSA DB
2. Cantus Ultimus
3. Cantus Database
4. DIAMM
5. RISM
6. Cantus Index
7. Canadian Chant Database
8. Global Jukebox
9. DTL1000 (Dig That Lick)
10. MusicBrainz
11. AcousticBrainz
12. CritiqueBrainz
13. ListenBrainz
14. MOTET Database (Jennifer Thomas)

15. TheSession.org

Our Overall Process

1. Convert various databases into flat files (CVS or JSON-LD)
2. Reconcile (convert) schema (properties) and entities (name, place, title, etc.) to Wikidata URIs using OpenRefine
3. Store the results in OpenLink Virtuoso graph database (RDF/Linked Data)
4. Use SPARQL for queries

LinkedMusic Overall Process: Virtuoso Version



Different Music Databases (14) with incompatible metadata schemas

Export to simple formats (CSV) but in the original metadata schemas

Universal Music Instrument Lexicon (UMIL)
(provides vocabulary to Wikidata)

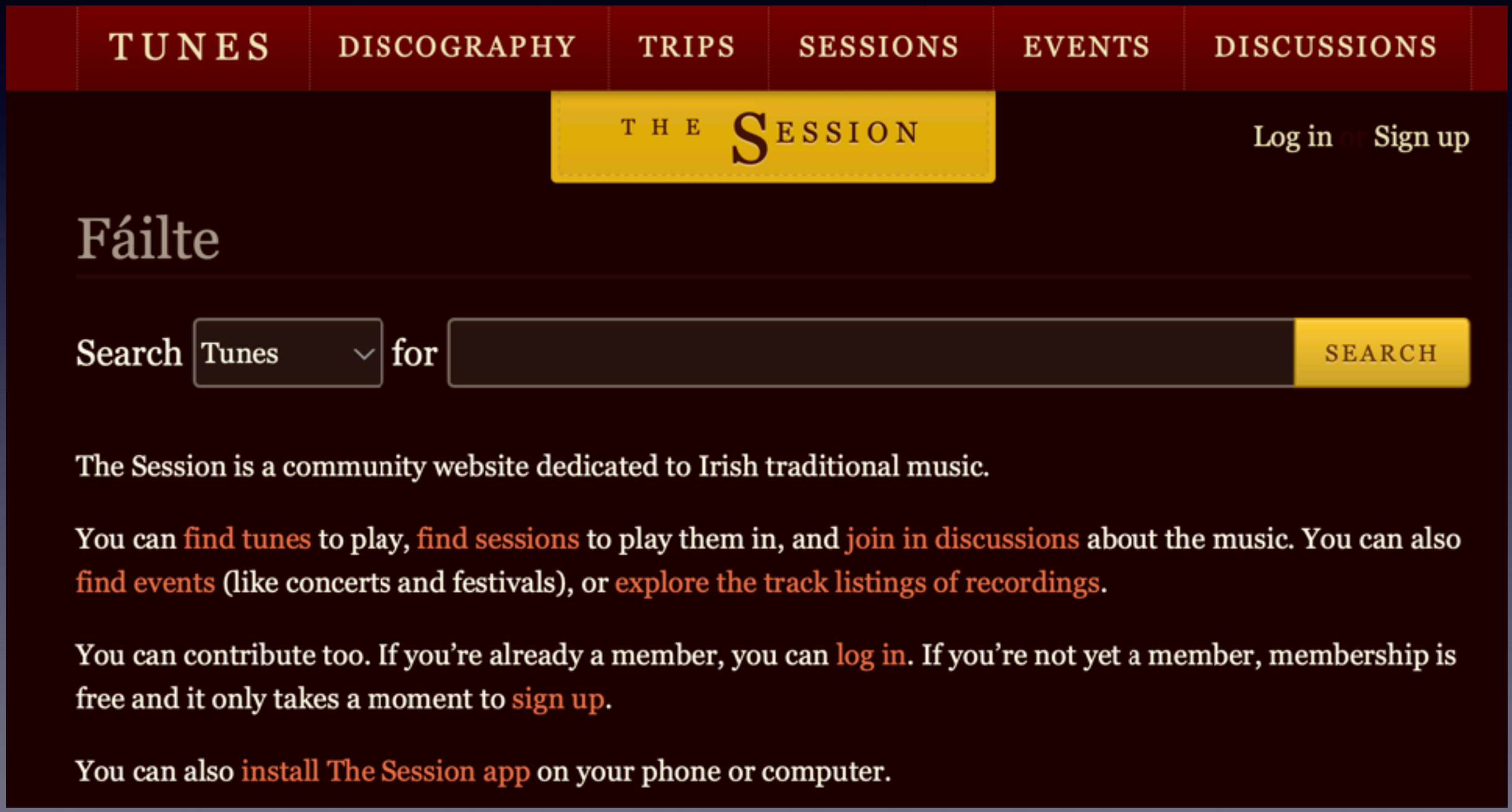
Search for *All Types of Music Information* from SESEMMI website

SESEMMI: Search Engine System to Enhance Music Metadata Interoperability

Once an item is found, the user is guided to the original database for detailed viewing

Example: The Sessions Database

Database of Traditional Irish Music sessions



1. Export the database to text files

Portion of a CSV file of events in Sessions Database

events_id	event	dtstart	dtend	venue	address	town	area	country
https://thesession.org/events/3310	–	1900-01-01 00:00:00	1900-01-01 00:00:00	–		Ober-Kainsbach	Hessen	Germany
https://thesession.org/events/11	Colm Gannon, Sean Mckeon And John Blake	2006-06-07 09:30:00	2006-06-07 12:00:00	The Goalpost	226 Water Street	Quincy	Massachusetts	USA
https://thesession.org/events/5	Brid O'Donohue	2006-06-09 08:00:00	2006-06-09 00:00:00	Glór	Causeway Link	Ennis	Clare	Ireland
https://thesession.org/events/6	National Celtic Festival	2006-06-09 19:00:00	2006-06-12 15:00:00	Various Venues		Portarlington	Victoria	Australia
https://thesession.org/events/7	The Irish Connections Festival	2006-06-09 19:00:00	2006-06-11 00:00:00	Irish Cultural Centre	200 New Boston Drive	Canton	Massachusetts	USA
https://thesession.org/events/19	Louisville Irish Fest	2006-06-10 11:00:00	2006-06-11 06:00:00	Belvedere/River Front Plaza		Louisville	Kentucky	USA
https://thesession.org/events/4	élan Concert (feat Damien Mullane And Sam Proctor)	2006-06-10 20:00:00	2006-06-10 23:30:00	Irish Cultural Centre	5 Black's Road	Hammersmith	London	England
https://thesession.org/events/2	Edel Fox And Ronan O'Flaherty	2006-06-10 21:30:00	2006-06-10	The Crosses Of Annagh		Miltown Malbay	Clare	Ireland

2. Convert to linked data

Assign each item to an URI (Universal Resource Identifier) using OpenRefine and Wikidata

town	area	country
Ober-Kainsbach	Hessen	Germany

town	town_wiki	area	area_wiki	country	country_wiki
Ober-Kainsbach Choose new match	https://www.wikidata.org/wiki/Q2008827	Hesse Choose new match	https://www.wikidata.org/wiki/Q1199	Germany Choose new match	https://www.wikidata.org/wiki/Q183

3. Store as text files (archive)

- Convert the CSV file with URI to flattened RDF (Resource Description Framework), e.g.:
  - Turtle
  - JSON-LD
  - N-Quads
- Also known as RDF serialization
- To be stored in a long-term archive

4. Import into an RDF graph database

We use an open-source software called Virtuoso

OPENLINK®
VIRTUOSO
UNIVERSAL SERVER

Other open-source examples include:

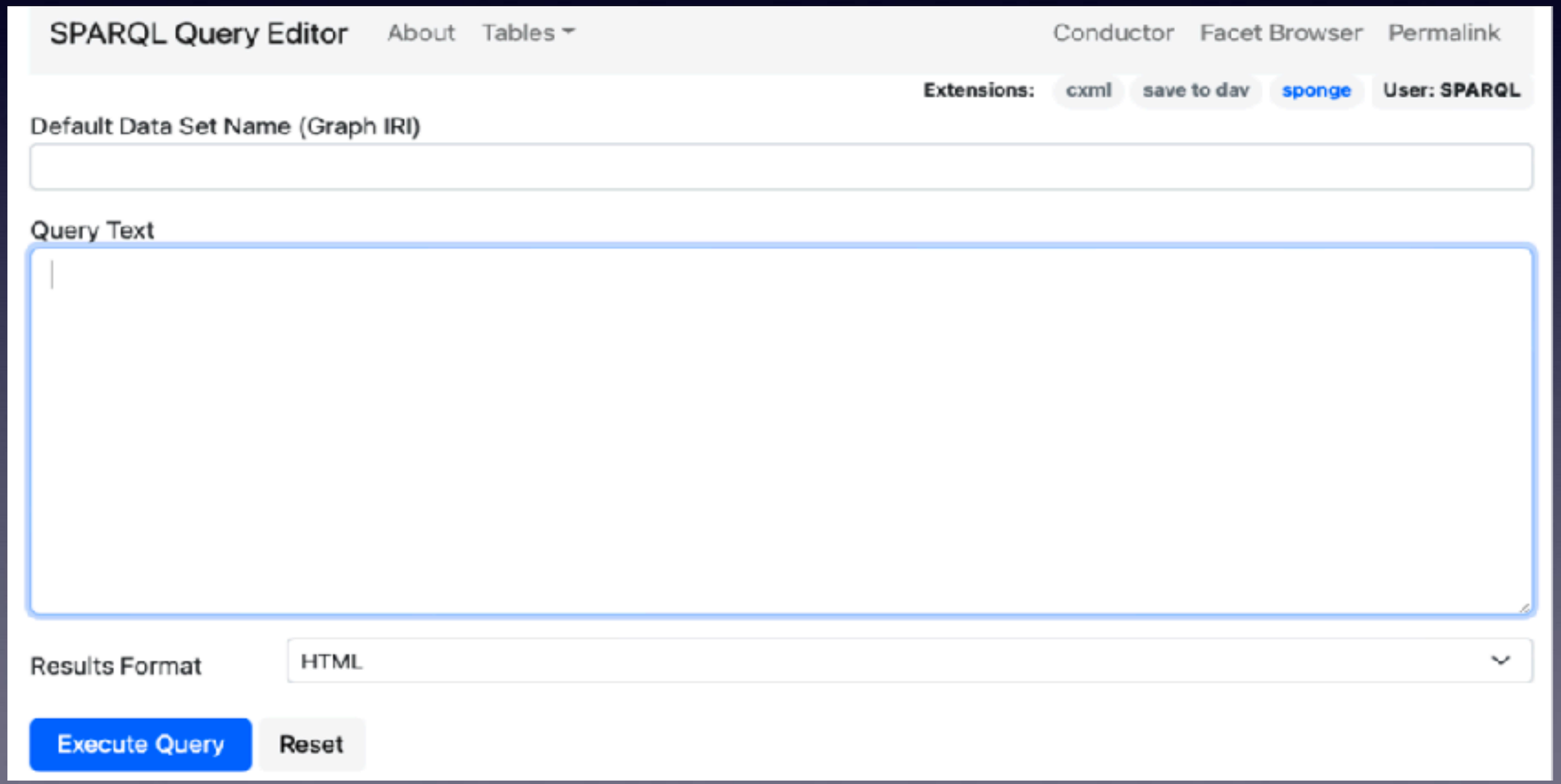
neo4j

Apache Jena

GraphDB

5. Query using natural languages (1)

Usually RDF database is queried using
SPARQL (SPARQL Protocol and RDF Query Language)



5. Query using natural languages (2)

Example query: **"Find sessions that took place in Greece"**

5. Query using natural languages (3)

Example query: **"Find sessions that took place in Greece"**

Equivalent SPARQL code:

```
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>

SELECT ?session
WHERE {
  ?session rdf:type thesession:sessions .
  ?session wdt:P17 wd:Q41 .
}
```

Comment: P17 is the country property and Q41 is Greece in Wikidata

5. Query using natural languages (4)

Inserting the SPARQL query in Virtuoso

SPARQL Query Editor About Tables Conductor Facet Browser Permalink

Extensions: cxml | save to dav | sponge | User: SPARQL

Default Data Set Name (Graph IRI)

http://sample/thesession/reconciled

Query Text

```
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>

SELECT ?session
WHERE {
  ?session rdf:type thesession:sessions .
  ?session wdt:P17 wd:Q41 .
}
```

Results Format: HTML

Execute Query Reset

5. Query using natural languages (5)

Executing the SPARQL query in Virtuoso

SPARQL | HTML5 table

session
https://thesession.org/sessions/7243
https://thesession.org/sessions/8020

7243	The Dubliner	Katouni 16	Thessaloniki
8020	The Lucky Sparrow Irish Pub	Triptolemou 44	Athens

But we cannot ask general users to create SPARQL queries!

5. Query using natural languages (6)

ChatGPT to the rescue!

I have an RDF database reconciled with Wikidata.

It contains

country it

Create a c

sions

in Greece.

```
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>

SELECT ?session
WHERE {
  ?session rdf:type thesession:sessions .
  ?session wdt:P17 wd:Q41 .
}
```

This prompt produces the correct SPARQL code!
(most of the time)

Features of LinkedMusic (1)

- By converting a database to an RDF graph database, we can search the database with natural language queries
- Currently we have integrated seven musical databases with 383 million RDF triples
- Because we use ChatGPT, we can make queries in many different natural languages:

  “ギリシャで行われたコンサートを探してください”

Features of LinkedMusic (2)

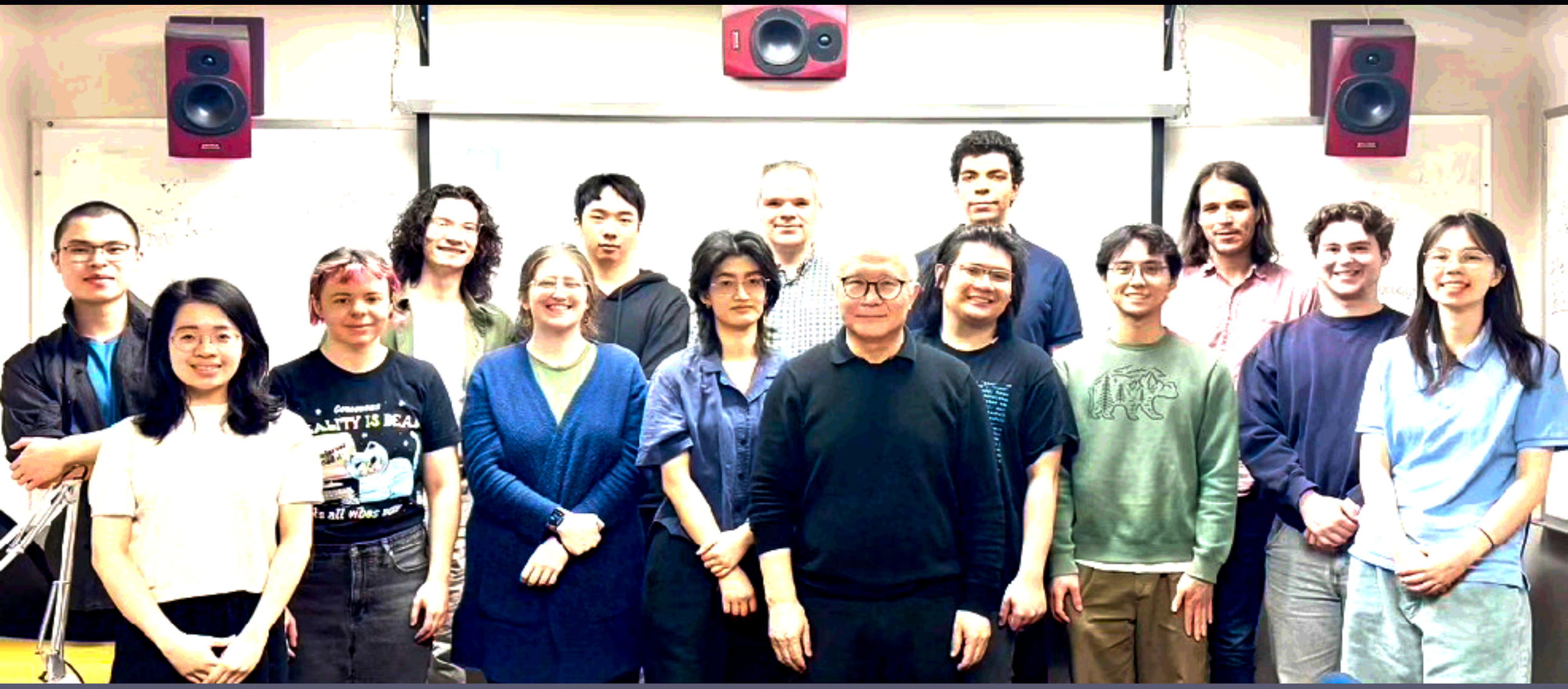
- We can even make queries that were not possible with the original web interface
- Furthermore, because we reconciled with Wikidata, we can query with concepts that were not defined in the original database: “Find women composers”
- Because RDF graphs can be stored as text files, we can easily archive them for long-term preservations

→ “The Persephone Initiative”

Future Directions

- Improve our prompts so that ChatGPT makes fewer errors
- Prompt Engineering
- Ask ChatGPT to create a web interface on the fly (e.g., Val Town)
- Sorting and facets capabilities

LinkedMusic Team @McGill: Summer 2025



Sichen Meng
Yu-Chia Kuo
Linnea Kirby
Liam Pond
Kyrie Bouressa
Yueqiao Zhang

Yinan Zhou
Andrew Hankinson
Ichiro Fujinaga
Antoine Phan
Simon Ngassam
Sebastien Chow

Dylan Hillerbrand
Lucas March
Mai Lyn Puittinen

Not in the photo:
Anna de Bakker
Hanwen Zhang
Kun Fang
Gen Gates-Panneton

Acknowledgements

Social Sciences and Humanities Research Council of Canada

Conseil de recherches en sciences humaines du Canada

Canada

McGill

Schulich School of Music
École de musique Schulich

DDMAL DISTRIBUTED DIGITAL MUSIC ARCHIVES & LIBRARIES LAB

CIRMMT Centre for Interdisciplinary Research in Music Media and Technology

Fonds de recherche Québec

Digital Research Alliance of Canada | Alliance de recherche numérique du Canada

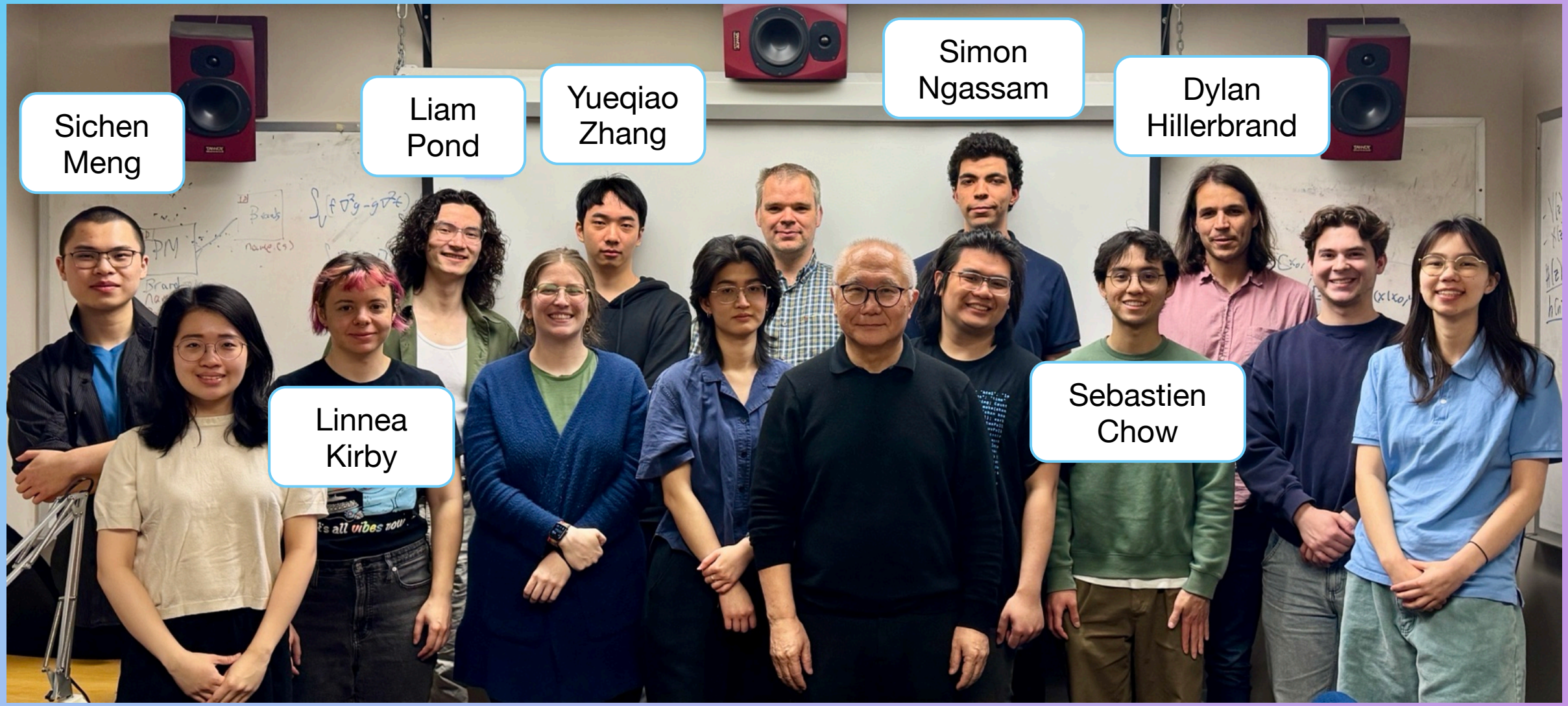
Calcul Québec

These slides were created with the help from ChatGPT, Claude, and Gemini, collectively known as Chappie.

LinkedMusic Update

May 2025 – Present

Presented by Liam Pond



Not pictured: Junjun Cao

Overview

- Status in May 2025
- Standardized data ingestion pipeline
- Database-specific challenges and insights
- NLQ2SPARQL experiments
- Future work — future datasets, NLQ2SPARQL

Status in May 2025

- Fetched and cleaned data
- Partially reconciled, none fully ingested
- MusicBrainz converted to RDF
- Experimented with NLQ2SPARQL in the browser

Status Today

MusicBrainz

DIAMM

RISM
Répertoire International des Sources Musicales

The Global Jukebox

THE SESSION

Dig That Lick 1000

cantus

SIMSSA
Database

AcousticBrainz

CANTUS INDEX

- 10 reconciled databases
- ~383 million RDF triples in Virtuoso
- Standardized ingestion pipeline
- Systematic empirical investigation of NLQ2SPARQL
- 30 Wiki pages, READMEs for each database

Data Ingestion Pipeline

Data Ingestion

- Simple so far (large public databases)
- Occasionally had to scrape websites (DIAMM)
- Could be challenging in the future for niche or dead databases (Persephone Initiative)

Data Cleaning

- Modifying or flattening database structure
- Standardizing formatting, **not** correcting metadata

Examples:

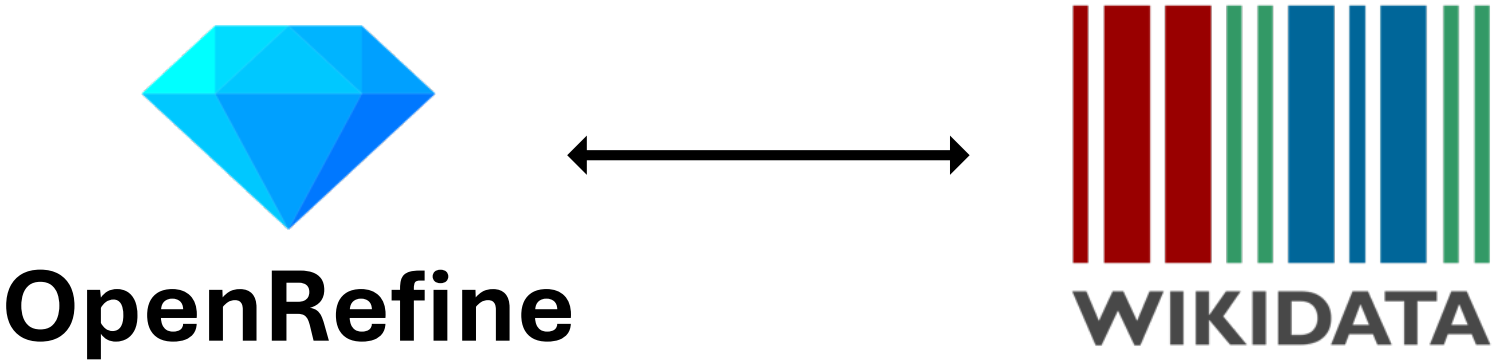
- Standardized dates instead of raw strings
- Splitting fields with multiple entities
- Expanding abbreviated instruments

Data Cleaning

Artists: Horace Silver (p), Joe Calloway (b), Stan Getz (ts), Walter Bolden (dr)

Artist	Instrument
Horace Silver	Piano
Joe Calloway	Bass
Stan Getz	Tenor Saxophone
Walter Bolden	Drums

Reconciliation: OpenRefine



- 99% confidence threshold for automatic reconciliation
- Manual verification otherwise
- P2888: Exact match
- String literals if entities are not in Wikidata

RDF Conversion

Principles

- No blank nodes or qualifiers
- For data with multiple values in multiple languages, specify the language where possible
- Store the most specific information (e.g., venue/city of an event)

RDF Conversion

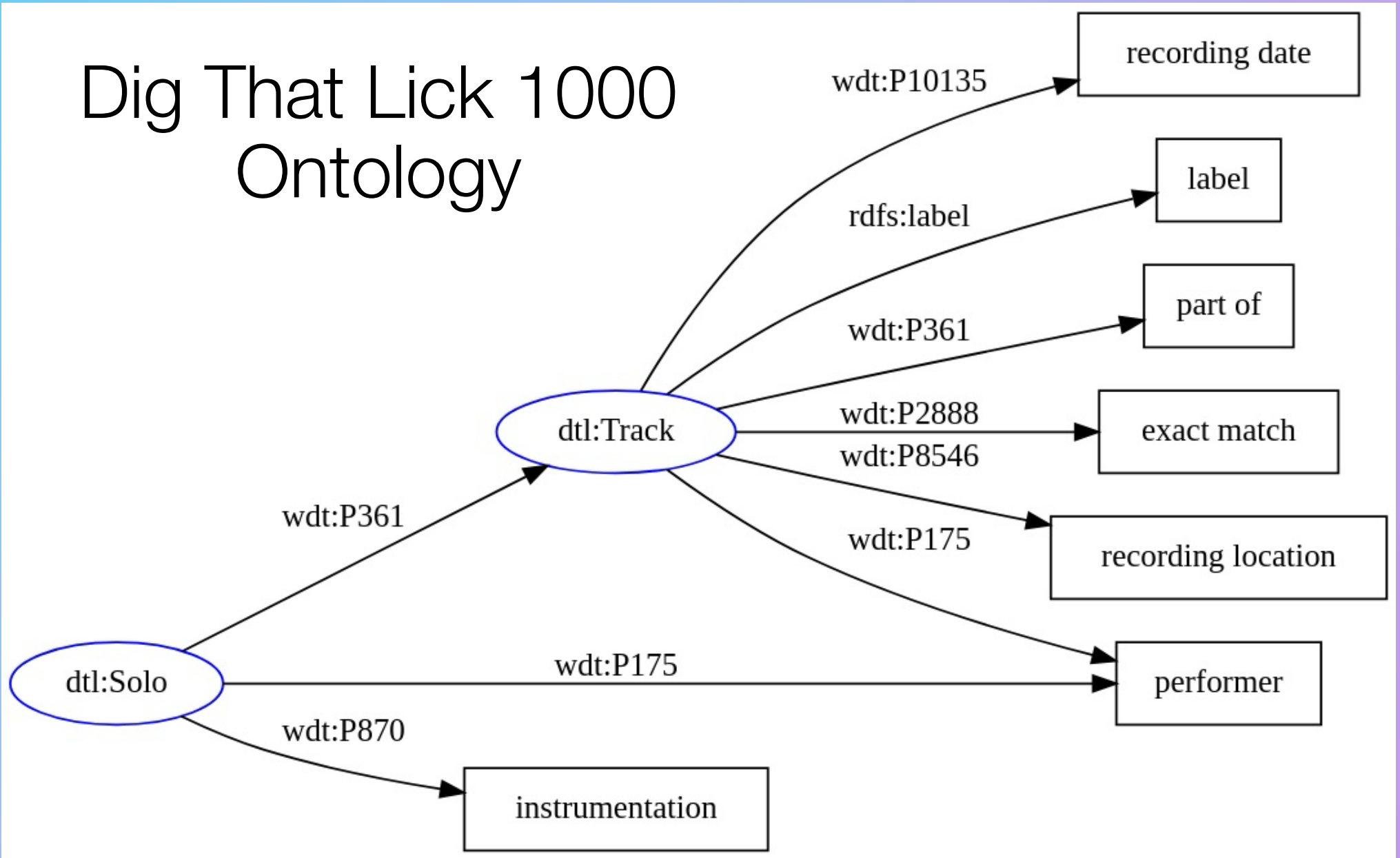
Which property should I use?

- Most precise, unless it is rarely used
- Use the same datatypes as Wikidata for dates, coordinates, etc.
- Respect the domain/range (e.g., P647 “Drafted by”)
- Verify object vs. data properties (point to an object or a literal)
- Direction matters

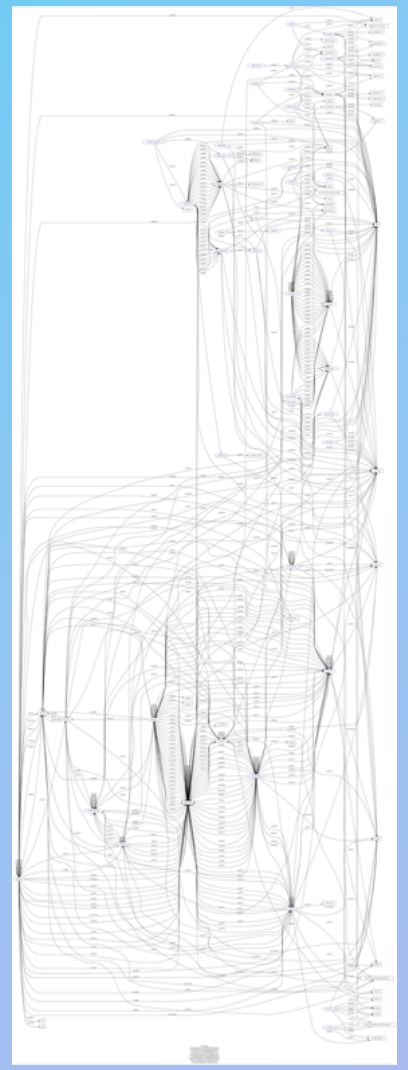
Linked Data Server: OpenLink Virtuoso

- Multi-model virtual database and web server
- SPARQL endpoint
- Two instances on Arbutus Cloud (Digital Research Alliance Canada)
- Production server at virtuoso.simssa.ca
- Staging server at virtuoso.staging.simssa.ca
- Staging runs locally in our lab via a tunnel, requires McGill VPN to access

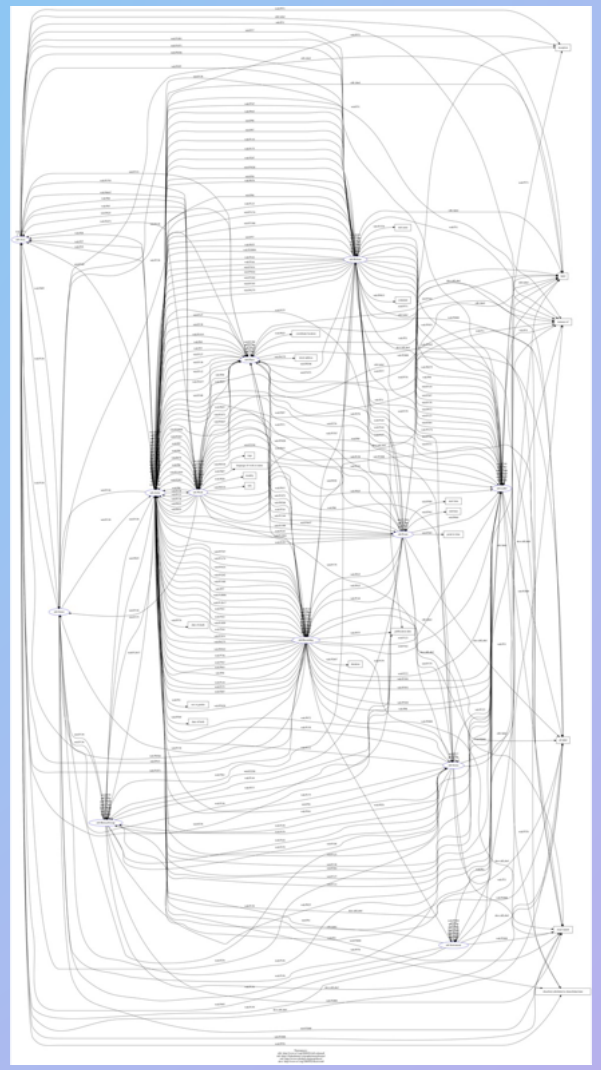
Dig That Lick 1000 Ontology



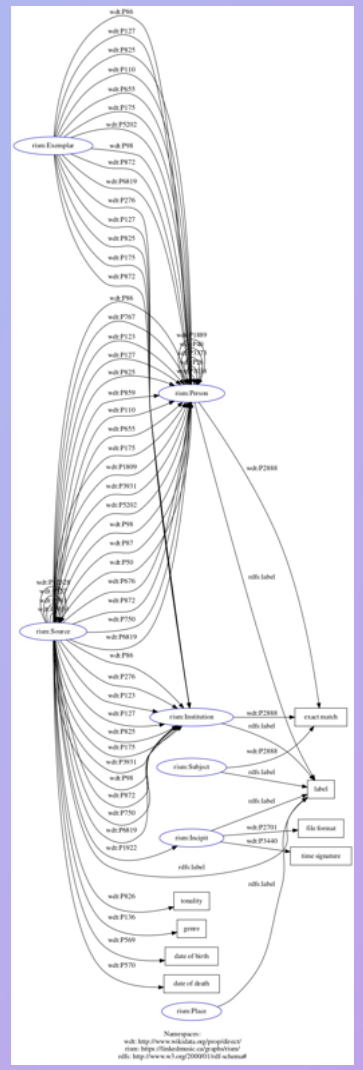
LinkedMusic



MusicBrainz



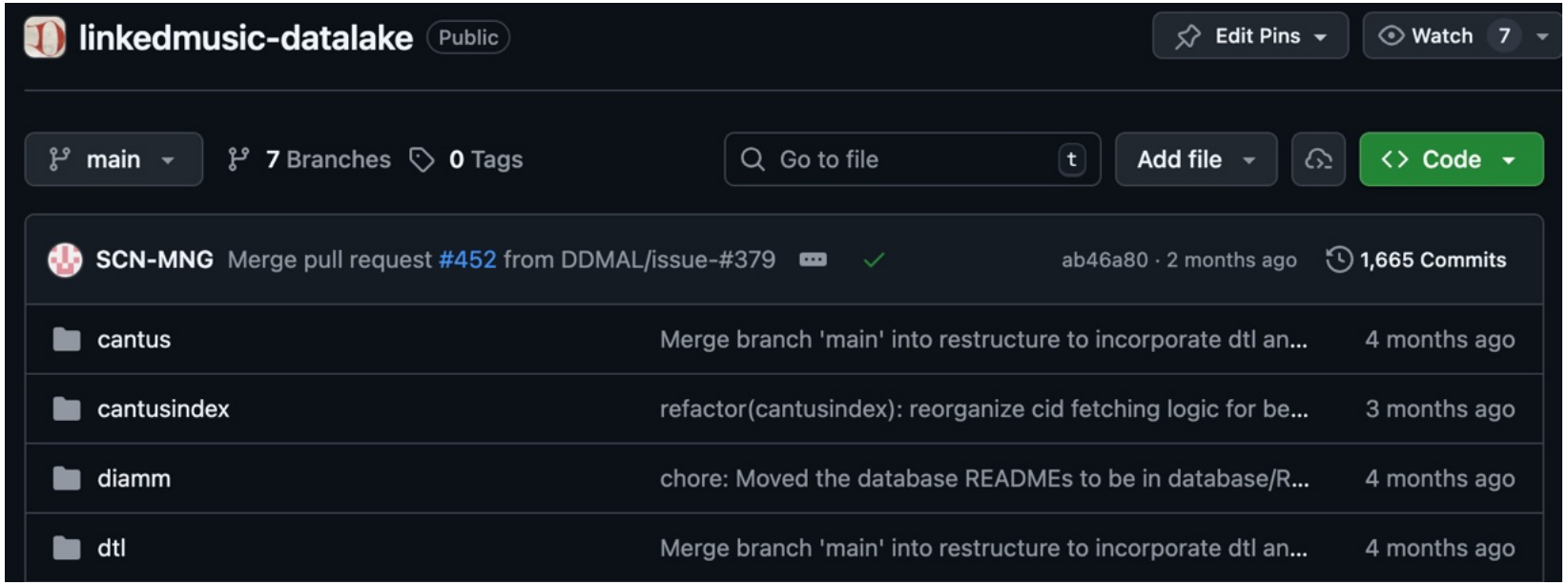
RISM

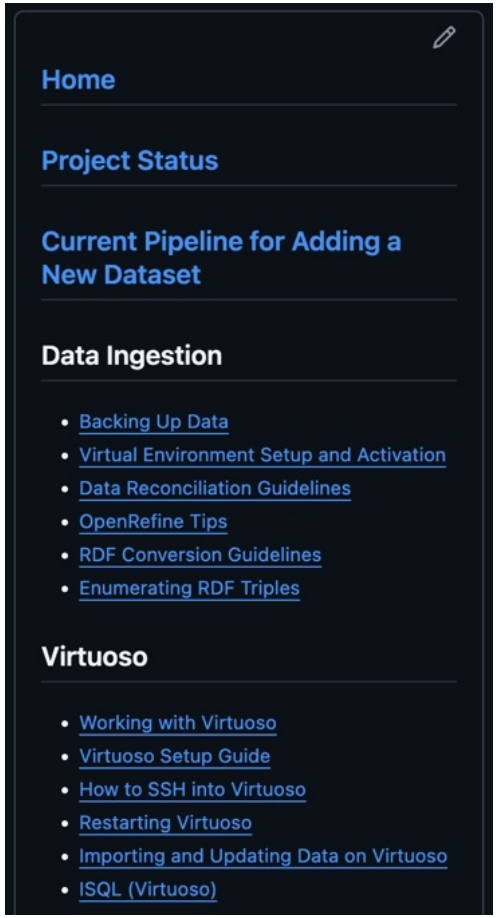


Credit:
Simon Ngassam

Documentation and Repository Structure

- Refactored to a database-centric structure
- README and documentation folder for each database
- 30 Wiki pages, custom sidebar





Credit:
Linnea Kirby

Database-Specific Challenges

The Session

Junjun Cao, Sebastien Chow, Linnea Kirby,
Sichen Meng, Yueqiao Zhang

- Traditional Irish music
- Data fetched from a GitHub repo with CSVs (scraped artist URIs)
- Low-frequency artists kept as string literals
- What to do about “Key: A mixolydian?”
  - “Key: A minor” exists in Wikidata
  - Could reconcile “Tonality: A” and “Mode: Mixolydian”
  - Create “Key: A mixolydian” in Wikidata?
- When should we create Wikidata entities?

DIAMM

Simon Ngassam

- Digital Image Archive of Medieval Music
- Web scraper for data ingestion
- “Saw and described manuscript” → P61 “Discoverer or inventor”
- “Binder” → P170 “Creator”
- “Copied at” → P1071 “Location of creation”
- “Witnessed document” → unmapped
- Wikidata does not distinguished “commissioned by” and “commissioned for”

Dig That Lick 1000

Sebastien Chow

- High ambiguity in performers (e.g., Joe Thomas: American jazz trumpeter)

URI Issues

- Solo IDs in the CSV don’t match IDs on the website
- URIs for DTL solos go to dtl.org not dig-that-lick.hfm-weimar.de

Dig That Lick 1000

Welcome!

Books and eBooks by the Director

DARKNESS TO LIGHT

Christian Ministry

Founded July 1991

Thanking the LORD for over 30 years of ministry

By Gary F. Zeolla

'... to open their eyes [in order] to turn [them] back from **darkness** [i.e., falsehood and unrighteousness] **to light** [i.e., truth and righteousness] and [from] the authority of Satan to God, [in order for] them to receive forgiveness of sins and an inheritance among the ones having been sanctified by faith in Me' (Acts 26:18; g).

Dig That Lick 1000

Sebastien Chow

- High ambiguity in performers (e.g., Joe Thomas: American jazz trumpeter)

URI Issues

- Solo IDs in the CSV don’t match IDs on the website
- URIs for DTL solos go to dtl.org not dig-that-lick.hfm-weimar.de
- Solos are part of tracks, tracks have the metadata
- Tracks don’t have URIs

Dig That Lick 1000

MusicBrainz

Junjun Cao, Simon Ngassam, Yueqiao Zhang

- ~355 million RDF triples, currently 92.7% of LinkedMusic
- High-quality API, mostly reconciled with Wikidata already
- Release status (e.g., album)
  - Q18609099 “Withdrawal” is exclusively related to sports
  - Q2352928 “Expunged” is about destroying criminal records

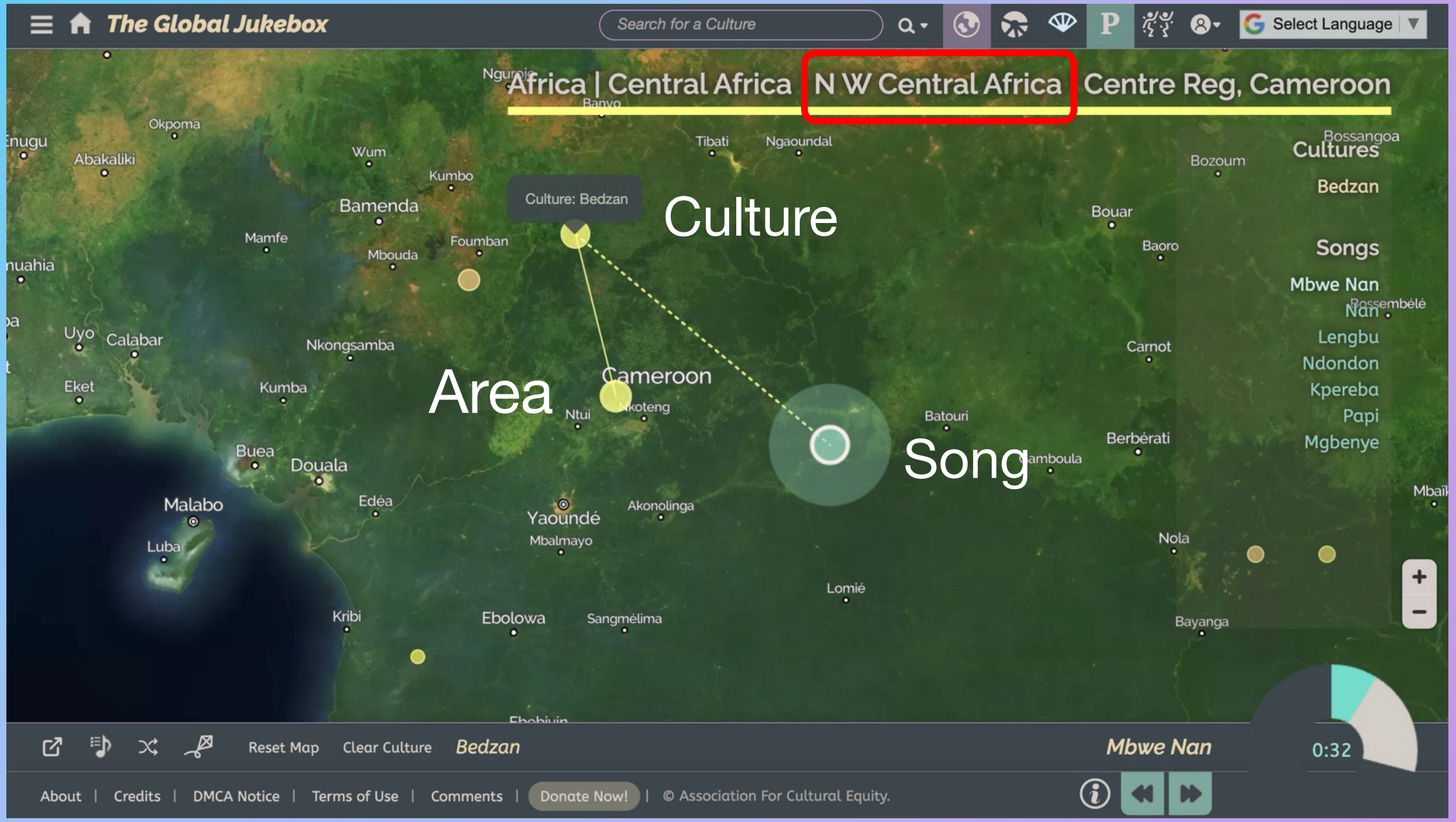
MusicBrainz

MusicBrainz

Junjun Cao, Simon Ngassam, Yueqiao Zhang

- Many packaging types unreconciled (e.g., paper sleeve, Digifile)
- “Slim jewel case” → Q1023101 “Jewel case”
- P767 “Contributor to the creative work or subject” used 33 times
  - Used for photographer, mixer, audio engineer, event booking person, etc.
- If a user asks for photographers, an LLM might naively return all people reconciled to P767

MusicBrainz


The Global Jukebox
Search for a Culture
Select Language
Africa | Central Africa
N W Central Africa
Centre Reg, Cameroon
Culture: Bedzan
Culture
Area
Song
Cultures
Bedzan
Songs
Mbwe Nan
Nan
Lengbu
Ndondon
Kpereba
Papi
Mgbenye
Okpoma
Abakaliki
Wum
Kumbo
Bamenda
Mamfe
Mbouda
Foumban
Tibati
Ngaoundal
Bossangoa
Bozoum
Bouar
Baoro
Rossembélé
Carnot
Uyo
Calabar
Nkongsamba
Eket
Kumba
Cameroon
Ntui
Nkoteng
Batouri
Berbérati
Buea
Douala
Malabo
Edéa
Yaoundé
Mbalmayo
Akonolinga
Luba
Nola
Lomié
Kribi
Ebolowa
Sangmélima
Bayanga
Reset Map
Clear Culture
Bedzan
Mbwe Nan
0:32
About | Credits | DMCA Notice | Terms of Use | Comments | Donate Now! | © Association For Cultural Equity.

The Global Jukebox

Linnea Kirby

- Culture rather than country-oriented
- Locations like “N W Central Africa” don’t map nicely with Wikidata
- Did not include encodings
  - Primary function of an ensemble (e.g., ritual, funerary, war)
  - Gender composition (male, female, children, any)
  - Prominence of low back vowels in a song
- Excluded Parlametrics — dataset of conversational recordings

The Global Jukebox

RISM

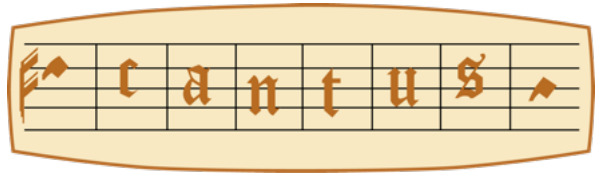
Junjun Cao, Sebastien Chow, Linnea Kirby,
Sichen Meng, Yueqiao Zhang

- RISM is in RDF, but we need CSVs to reconcile with OpenRefine
- Split because files were too big
- Gave blank nodes hashes to fix issues tracking them across files

Cantus Database

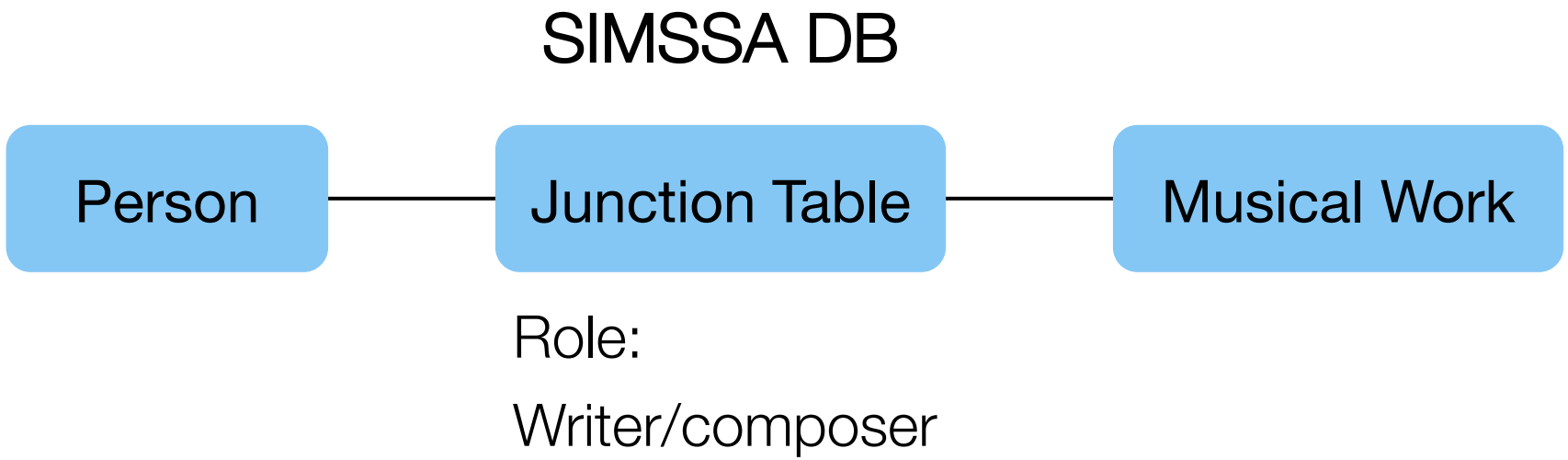
Sichen Meng

- What Wikidata property should link chants to feasts?
- P366: “Has use” is broad
  - Book has use reading
  - 1,1-dimethylhydrazine has use rocket propellant
- P9215: “Set during recurring event” is for fictional works
  - Home Alone (movie) is set during Christmas
- Should we create highly specialized properties?
- Are obscure properties useful if no one else uses them?



SIMSSA DB

Sichen Meng, Simon Ngassam



- SQL view created for each role
- Roles pivoted into columns
- Avoids blank nodes and reification (RDF triple about a triple)

SIMSSA
Database

Cantus Index

Sichen Meng

- List of all Cantus Index chant IDs, mapped to JSON files
- JSON files merged to create a CSV
- Straightforward!

AcousticBrainz

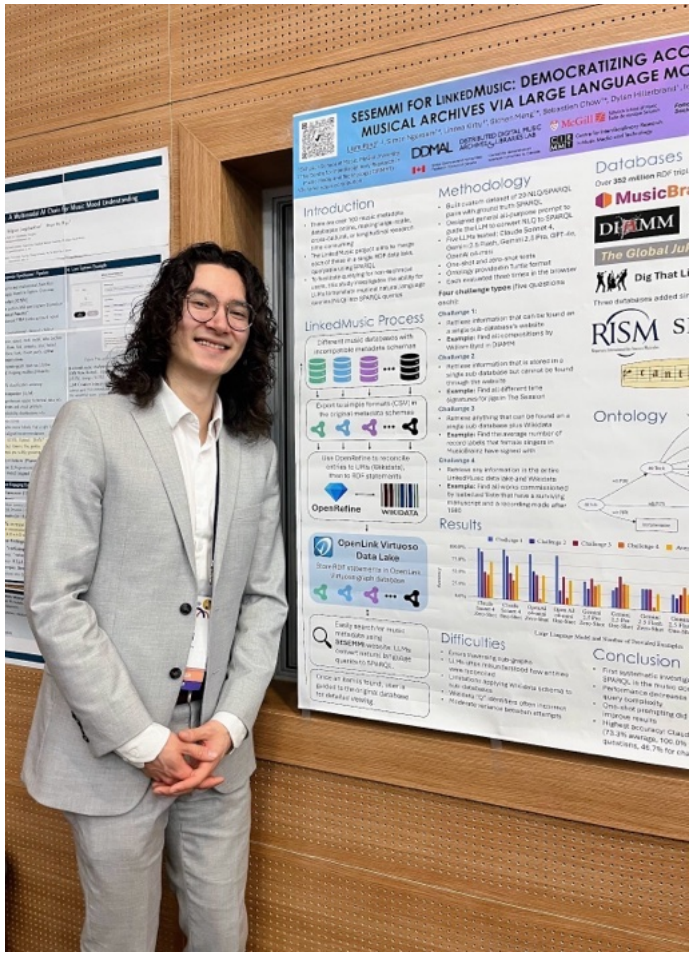
Liam Pond

- Low-level data automatically determined with Essentia for songs in MusicBrainz (e.g., spectral centroid, MFCCs, zero-crossing rate)
- High-level data computed based on the low-level data (e.g., key, beats per minute, danceability)
- Kept full high-level data, partial low-level data (3GB compressed vs. 589GB)

NLQ2SPARQL

NLQ2SPARQL

- Can LLMs turn natural language queries into SPARQL?
- Multilingual, enables conversation, lowers barriers to entry
- Systematic empirical study of work started by Junjun Cao
- Short paper accepted to LLM4MA (Daejeon, South Korea)



Methodology

- Custom dataset of 20 ground truth NLQ/SPARQL pairs
- Four challenge types, five questions each (one per sub-database)
- General prompt to guide conversion between NLQ and SPARQL
- Claude Sonnet 4, Gemini 2.5 Flash, Gemini 2.5 Pro, GPT-4o, OpenAI o4-mini
- Zero- and one-shot (number of examples)
- Full ontology provided in Turtle format
- Evaluated three times each in the browser

Dig That Lick 1000

The Global Jukebox

Question Challenge Types

Challenge 1:

Single sub-database

Challenge 2:

Single sub-database + expanded search

Challenge 3:

Single sub-database + Wikidata

Challenge 4:

Anything in LinkedMusic

Example 1

Find all compositions by William Byrd in DIAMM

Question Challenge Types

Challenge 1:

Single sub-database

Challenge 2:

Single sub-database + expanded search

Challenge 3:

Single sub-database + Wikidata

Challenge 4:

Anything in LinkedMusic

Example 2

Find all different time signatures for jigs in The Session

Question Challenge Types

Challenge 1:

Single sub-database

Challenge 2:

Single sub-database + expanded search

Challenge 3:

Single sub-database + Wikidata

Challenge 4:

Anything in LinkedMusic

Example 3

Find the average number of record labels that female singers in MusicBrainz have signed with

Question Challenge Types

Challenge 1:

Single sub-database

Challenge 2:

Single sub-database + expanded search

Challenge 3:

Single sub-database + Wikidata

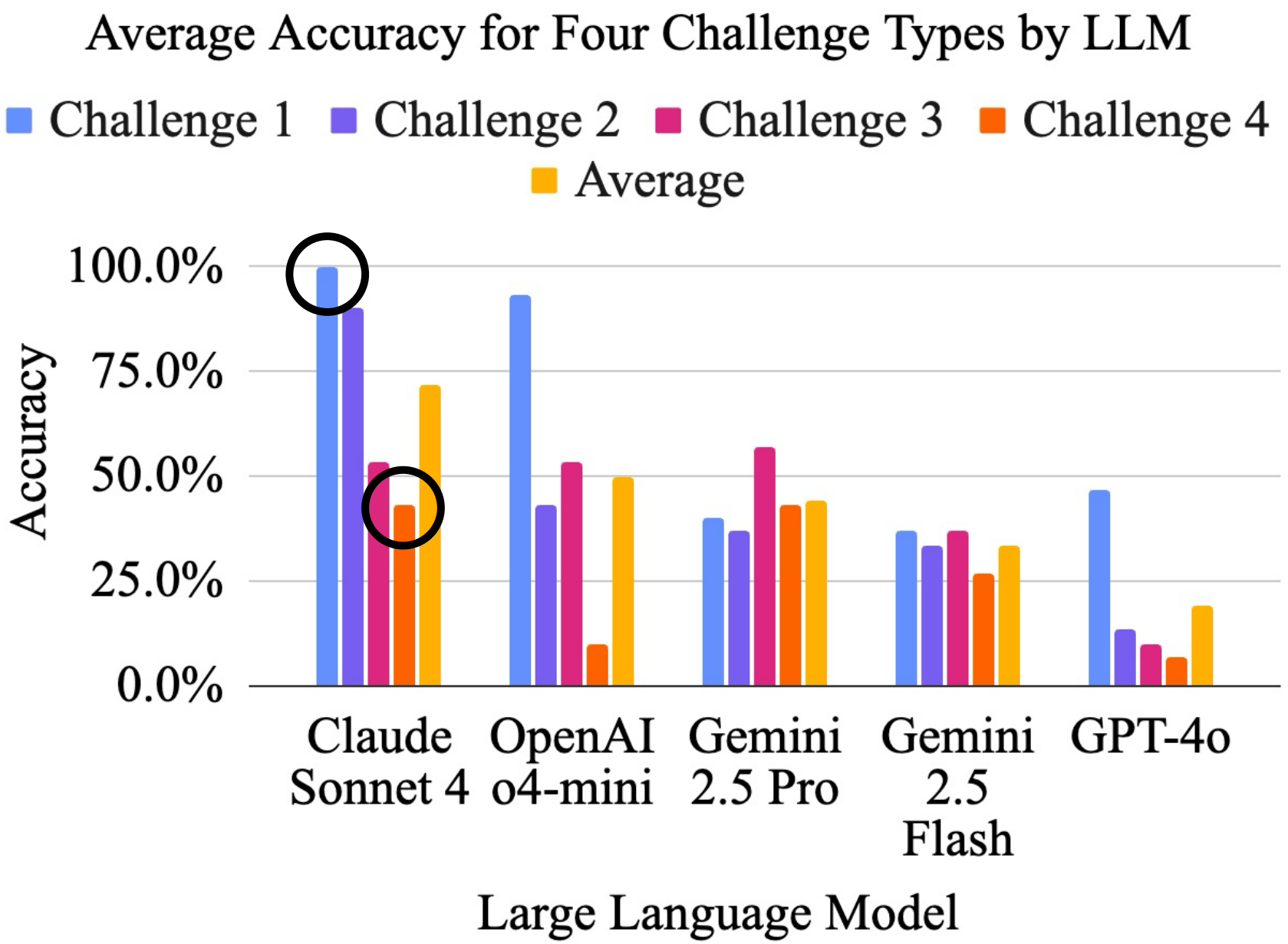
Challenge 4:

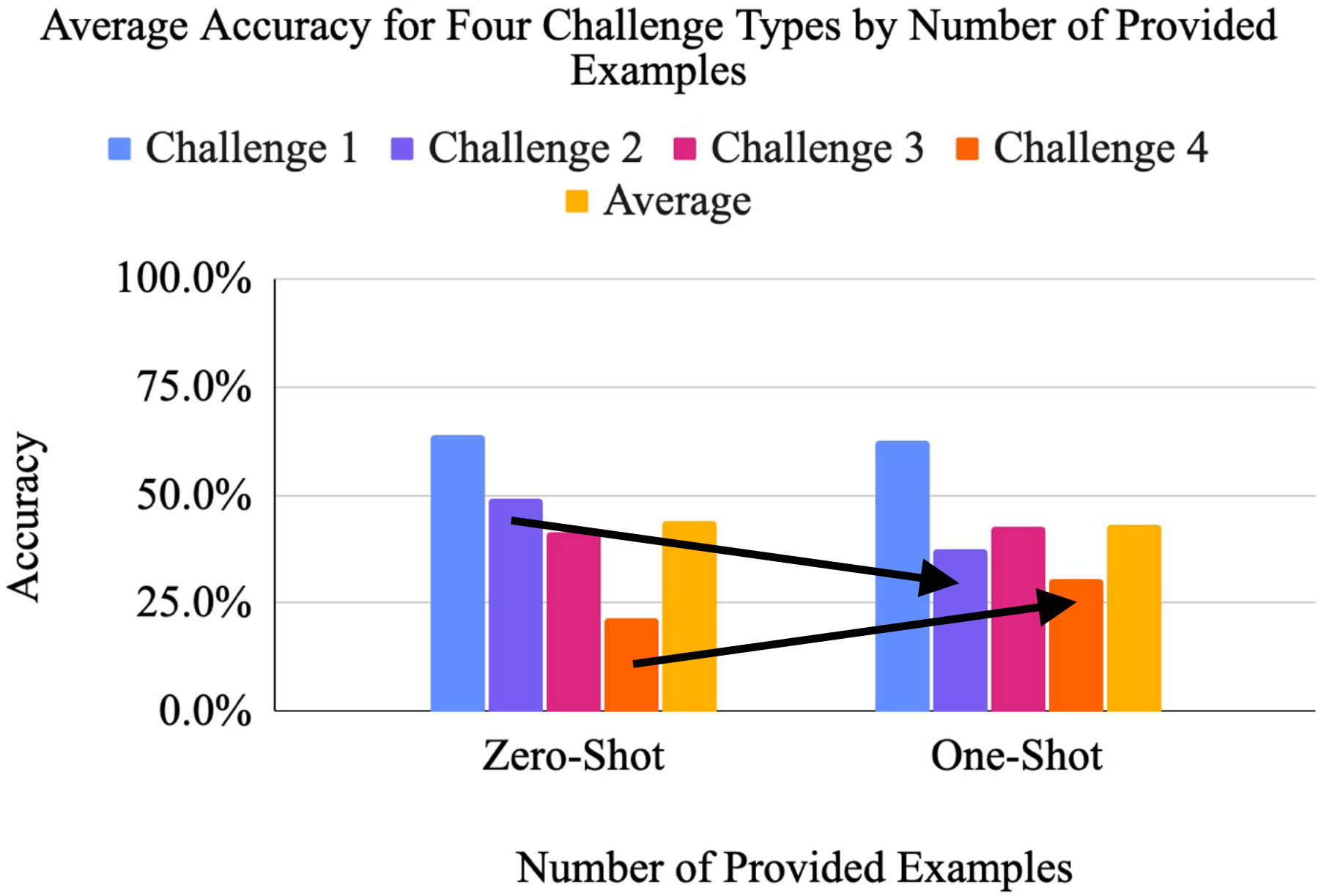
Anything in LinkedMusic

Example 4

Find all works commissioned by Isabella d'Este that have a surviving manuscript and a recording made after 1980

Results

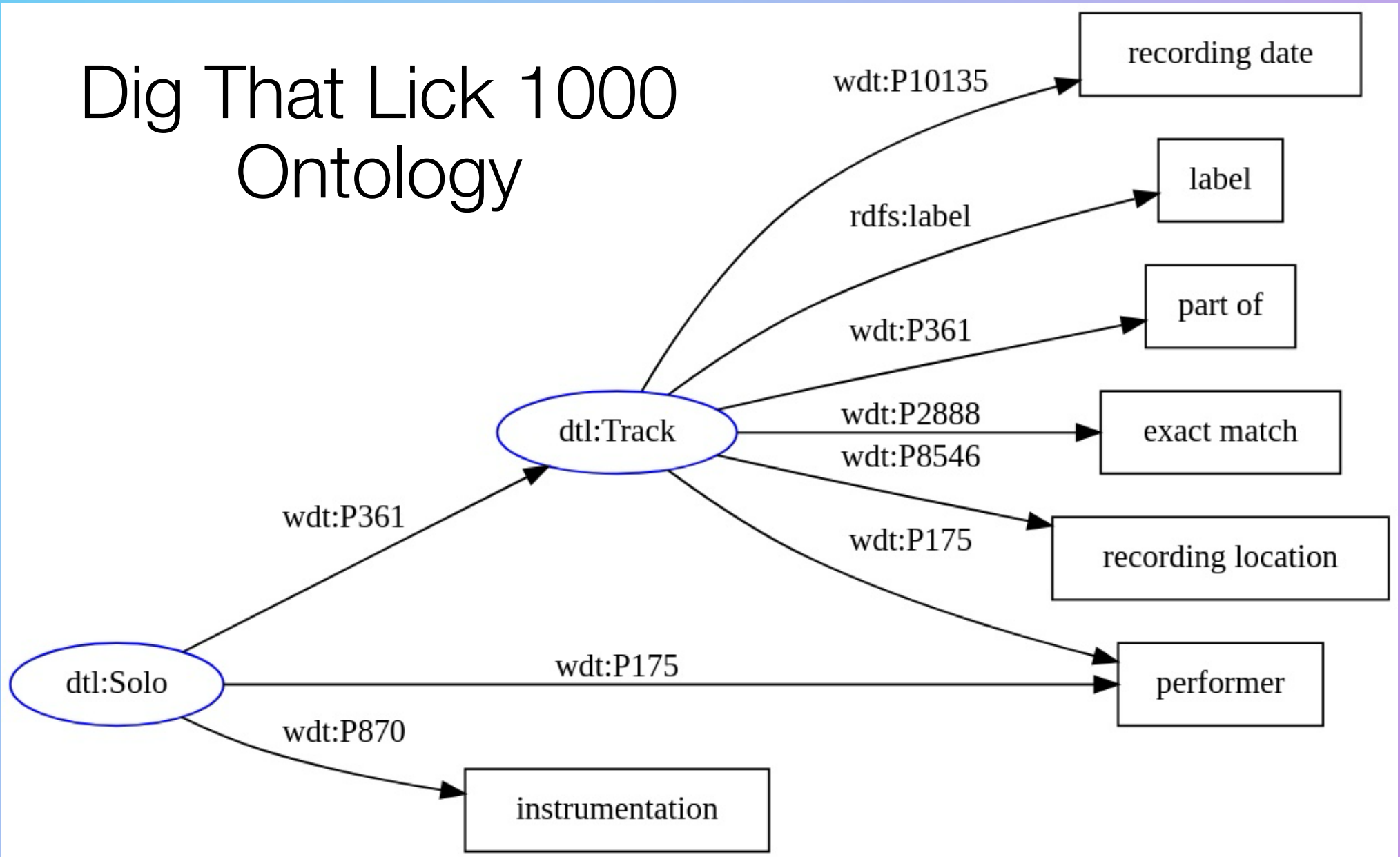

Average Accuracy for Four Challenge Types by LLM
Challenge 1
Challenge 2
Challenge 3
Challenge 4
Average
100.0%
75.0%
50.0%
25.0%
0.0%
Accuracy
Claude Sonnet 4
OpenAI o4-mini
Gemini 2.5 Pro
Gemini 2.5 Flash
GPT-4o
Large Language Model


Average Accuracy for Four Challenge Types by Number of Provided Examples
Challenge 1
Challenge 2
Challenge 3
Challenge 4
Average
100.0%
75.0%
50.0%
25.0%
0.0%
Accuracy
Zero-Shot
One-Shot
Number of Provided Examples

Difficulties

- LLMs often misunderstood how entities were reconciled
- Limitations applying Wikidata schema to sub-databases
- Wikidata “Q” identifiers often incorrectly retrieved
- Moderate variance between attempts
- Errors traversing subgraphs

Dig That Lick 1000 Ontology



NLQ2SPARQL Conclusion

- First systematic investigation of NLQ to SPARQL in music
- Performance decreases significantly with query complexity
- One-shot prompting did not consistently improve results
- Highest accuracy: Claude Sonnet 4 zero-shot
  - 73.3% average
  - 100.0% for Challenge 1 questions
  - 46.7% for Challenge 4 questions

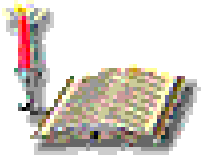
Future Work

- Future databases — please give suggestions!
- Adding entities to Wikidata
- Going over existing databases again
- Implementing NLQ2SPARQL agent
- Front end
- Persephone Initiative

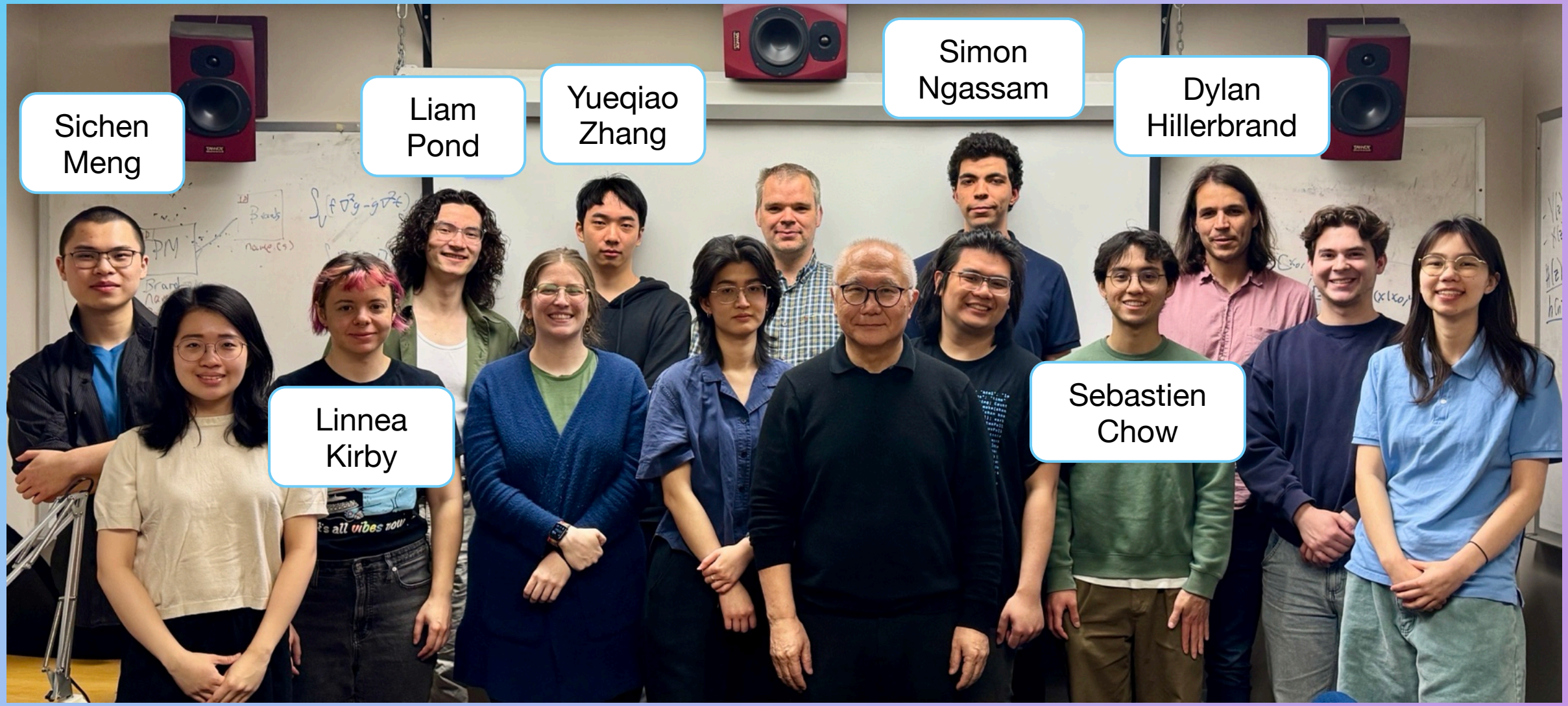
Cover Art Archive

CritiqueBrainz

ListenBrainz



Renaissance Liturgical Imprints: A Census



Not pictured: Junjun Cao

Interactive NLQ2SPARQL Agent:

Using Chinese Traditional Music Knowledge Base as an Example

Junjun Cao (Shaojun Si)

– Previous Postdoctor Researcher (2024–2025) for DDMAL, McGill University

– Music Knowledge Engineering & DH Lab, China Recording Corporation, Beijing

looking back to 2024:

Leveraging ChatGpt for Natural Language Query to SPARQL on Virtuoso

——Using TheSession, Wikidata, MusicBrainz as example

LinkedMusic

Issue: If the ontology (OWL) is (growing) oversized, LLM, more difficult to concentrate for accurate SPARQL generation

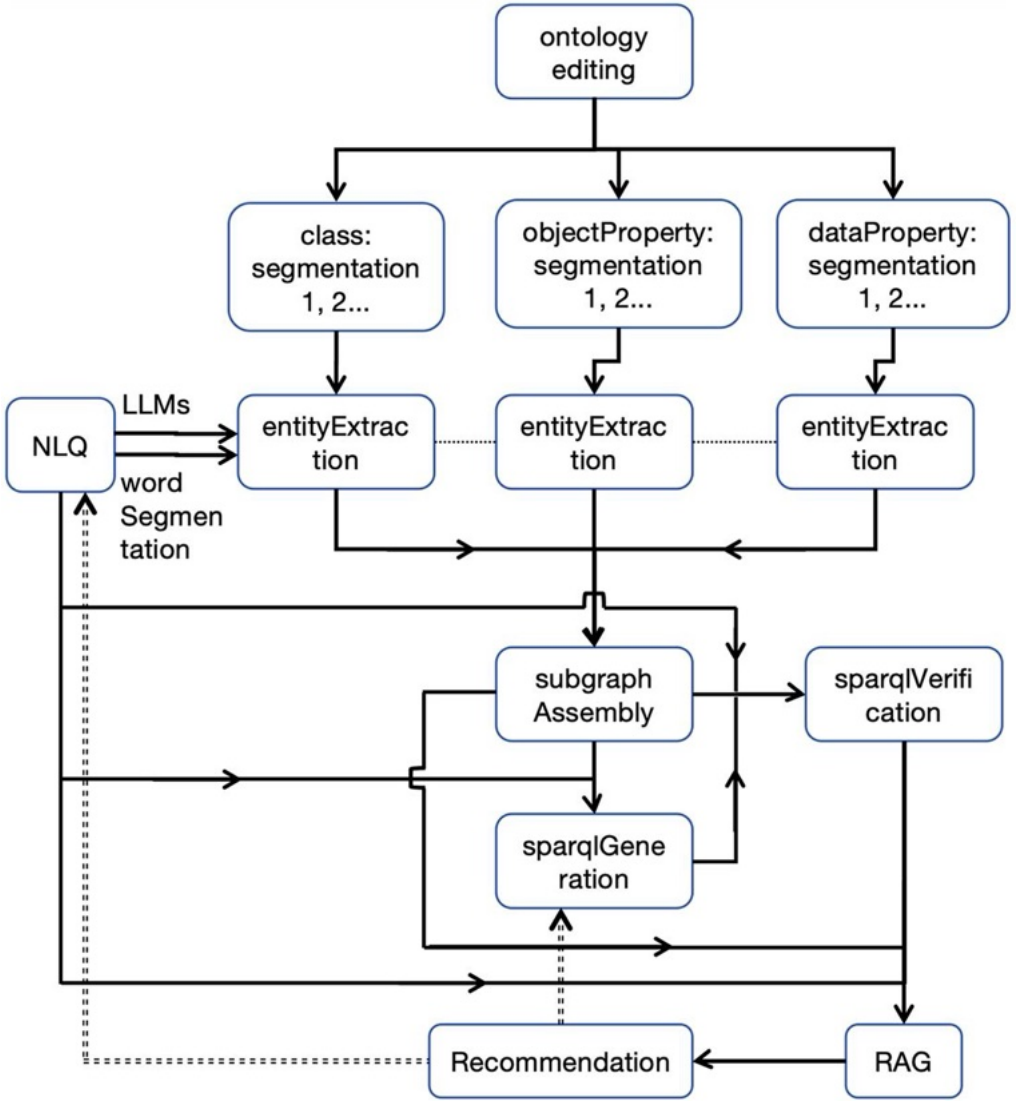
标题默认字号：28号，正文默认字号：22号

Web Ontology Language (OWL) in graph visualization

标题默认字号：28号，正文默认字号：22号

Solution："ontology subgraph" detection/extraction/re-assembly

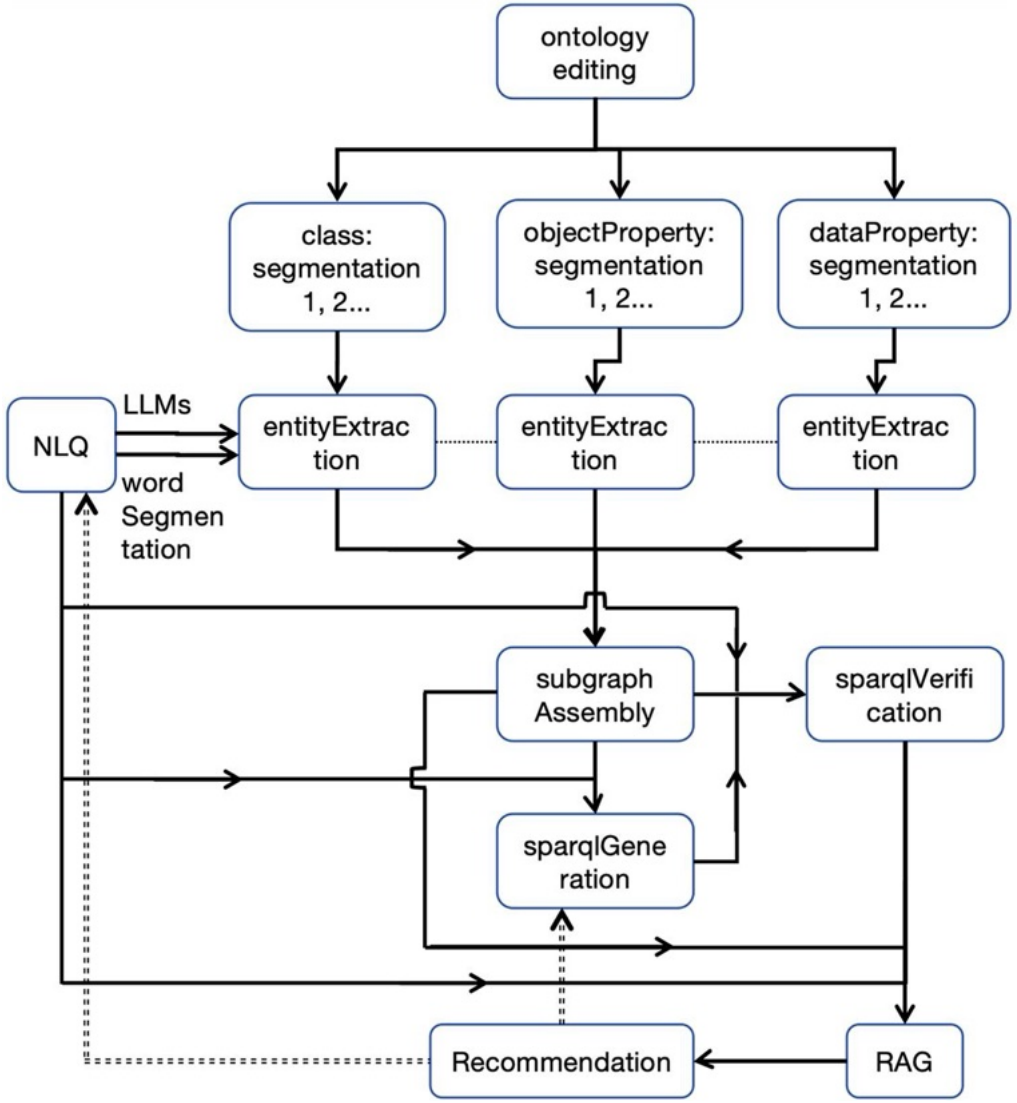
- 1. ontology segmentation
- 2. "entity" extraction
  - find the classes and and properties(relation) in the NLQ and match them in the ontology entities vocabulary list
  - implicit entities detection from NLQ
- 3. subgraph re-assembly



标题默认字号：28号，正文默认字号：22号

Solution："ontology subgraph" detection/extraction/re-assembly

- 4. SPARQL generation and verification
- 5. Retrieval Augmented Generation (RAG) or Generation Augmented Retrieval (GAR)
- 6.ontology subgraph neighborhood based recommendation



标题默认字号：28号，正文默认字号：22号

the global/holistic ontology：background

translation：Chinese traditional music comes from the acient courts, literati circles, religeous groups and folk communities, manifesting in forms as songs, narrative rap (speech—song fusion), singing—dancing integration, instrumental music, which can also be observed from viewpoints as ethnicity, instrumentation, geography and factions。Let's explore various communities within the "knowledge network", reveling in their dazzling multiplicity.

🎵 NLQ to SPARQL Interactive Chat

Chinese Traditional Music Knowledge Base Intelligent Question Answering System v2.0 (Multi-turn Dialogue + Complete Results)

🔄 Reset

👋 Welcome to the interactive SPARQL query system! (Fixed version)

New features :

- 💡Complete **5-part result** : Query results, SPARQL, Ontology, RAG analysis, and recommendations.
- 💬Intelligent **multi-turn dialogue** : Follow-up questions do not require re-execution of the workflow.
- ⚡Fast **Response** : Context-based conversations take only 3-5 seconds

You can :

- 🔍 Raising questions about natural language in traditional Chinese music
- 💬 Provide clarification, comments, or further requests (no need to search again)
- 🔄 Reset the conversation and start a new topic

💡 Example question:

What are some plucked string instruments in traditional Chinese music?

Enter your question or feedback here... (The first question will execute the full workflow; subsequent conversations will be based on context for quick responses)

📤 send

💡 Tip: Press Ctrl+Enter to send quickly | The first question takes about 30-60 seconds, and subsequent conversations take about 3-5 seconds.

In the dialogue (based on the question: "Where is the "Donbla" (...")

Solution："ontology subgraph" detection/extraction/re-assembly

- example
- 1.东不拉这种乐器分布在哪里？
- Where is the instrument dombra distributed?

- 2.东不拉分布在哪里？
- Where is the dombra distributed?

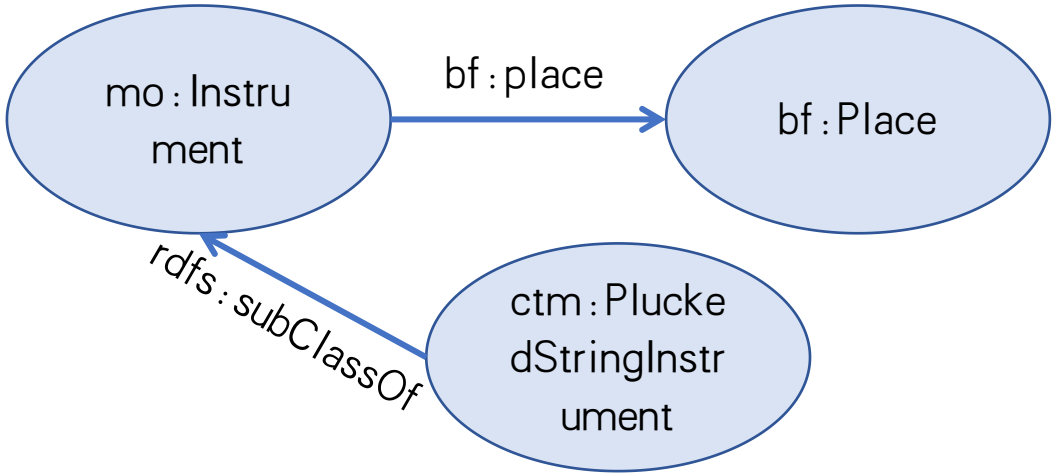
  ——*implicit entities detection*

- 2.1**卡龙**(ka long)分布在哪里？—LLM may not know what *ka long* is
- Where is the *ka long* distributed?

标题默认字号：28号，正文默认字号：22号

Solution：”ontology subgraph” detection/extraction/re-assembly

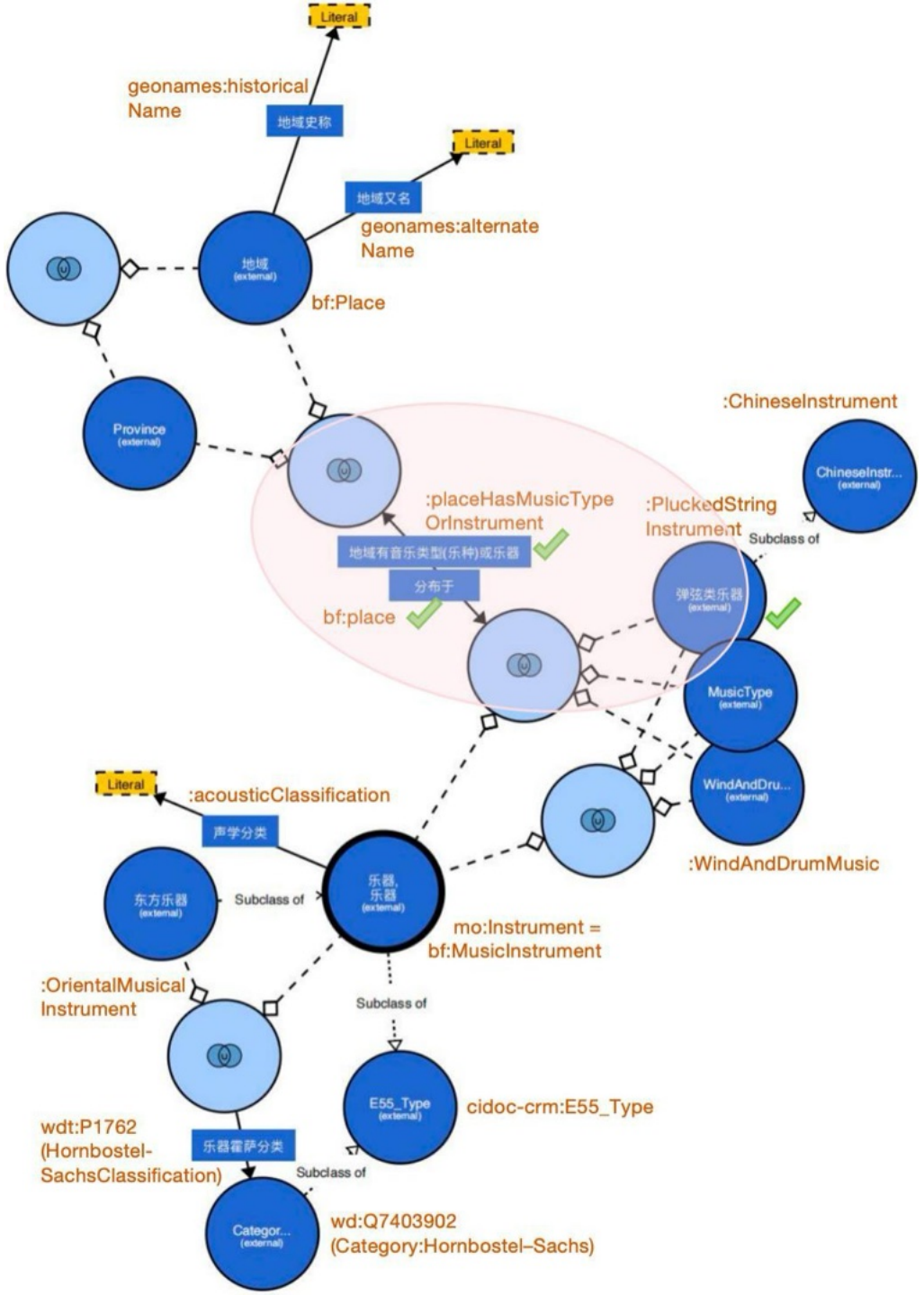
- example
- 3.东不拉分布在哪里；其所分布的地域还有哪些弹弦类乐器？
- Where is the dombra distributed, and what other plucked-string instruments are found in those regions?



You can ask above in one go or through interaction.

agent webpage 1

标题默认字号：28号，正文默认字号：22号




```
define input:inference 'urn:owl.ccmusicrules0913' # activating Reasoning

PREFIX bf: <http://id.loc.gov/ontologies/bibframe/>

PREFIX ctm:
<https://lib.ccmusic.edu.cn/ontologies/chinese_traditional_music#>

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT
DISTINCT ?dombra ?dombraLabel ?place ?placeLabel ?otherInstrument ?otherInstrumentLabel

WHERE { # Find the dombra instrument

?dombra a ctm:PluckedStringInstrument;

        rdfs:label ?dombraLabel .

FILTER(CONTAINS(?dombraLabel, "东不拉"))

?dombra bf:place ?place . # Get the place where dombra is distributed

?place rdfs:label ?placeLabel .

# Find other plucked string instruments in the same place

?otherInstrument a ctm:PluckedStringInstrument;

                 bf:place ?place ; rdfs:label ?otherInstrumentLabel .

# Exclude the dombra itself

FILTER(?otherInstrument != ?dombra)}
```

Distribution：

Gansu Province（甘肃省）

Jiuquan City（酒泉市）

Aksai Kazakh Autonomous County（阿克塞哈萨克自治县）

Qinghai Province（青海省）

Xinjiang Uygur /ˈwiːgər/ Autonomous Region（新疆维吾尔自治区）

Hami City（哈密市）

. . .

Other Plucked String Instruments：

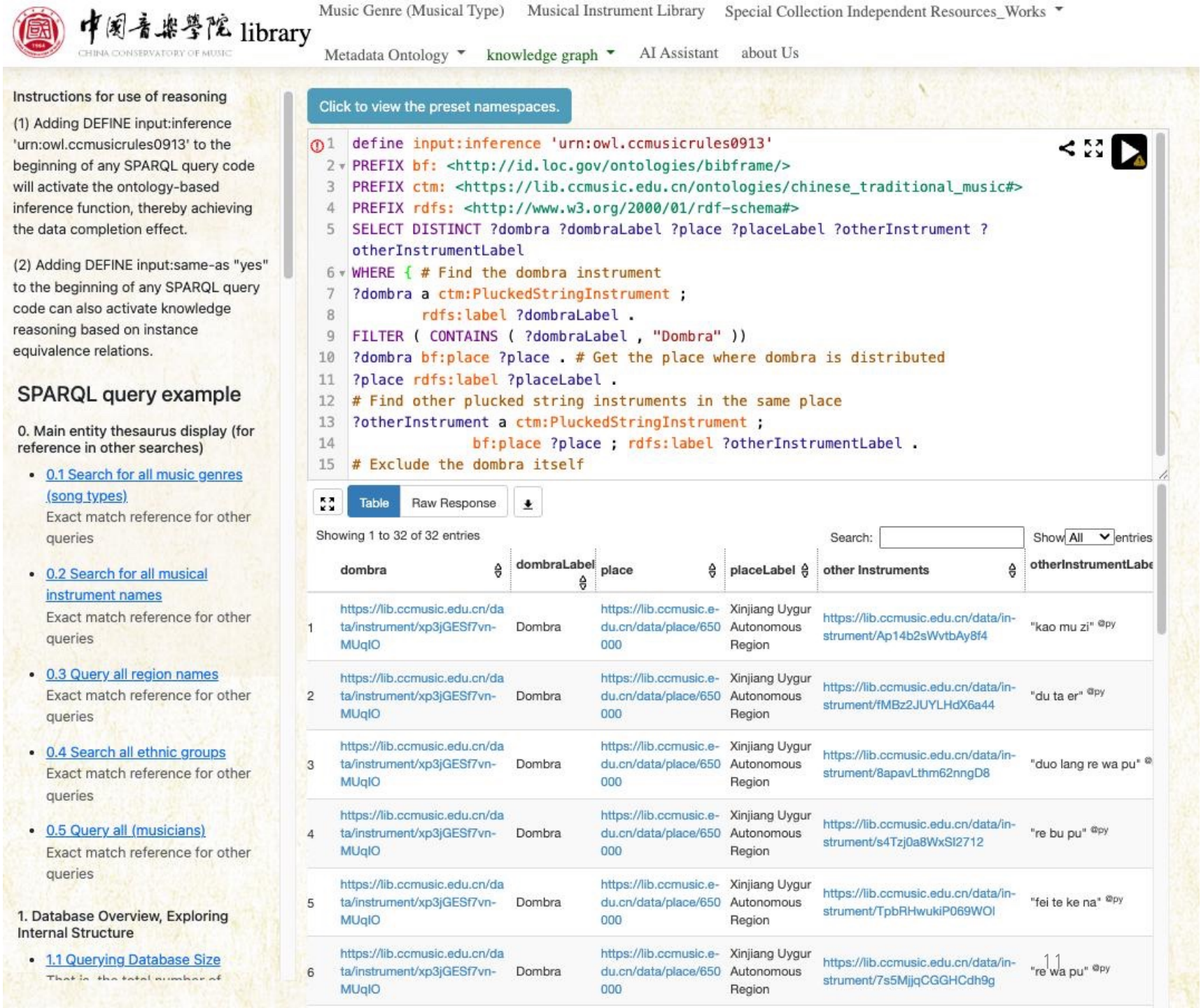
Rawap（热瓦普）

ka long（卡龙）

Dutar（独它尔）

Comz（考姆兹）

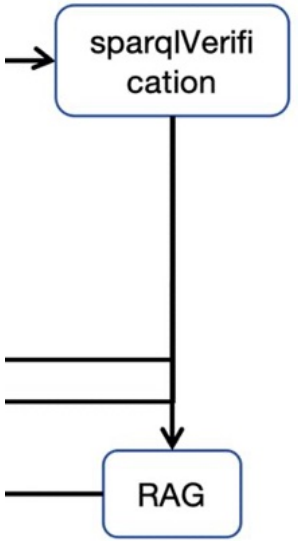
. . .



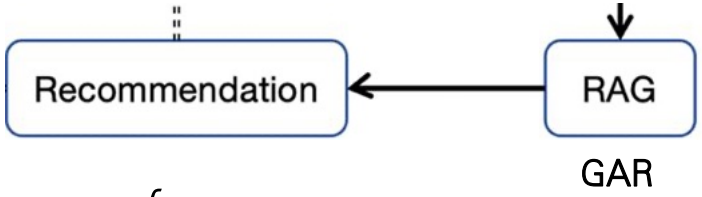
Mutual Complementation of Knowledge Graph and LLM: RAG+GAR

- prompt for RAG + GAR:

'''*Based on a natural language question: {question},*
and the related ontology snippet (subgraph): {turtle_output},
and the subsequent SPARQL query: {sparql_query},
from visiting a SPARQL endpoint we retrieved the result,
part of which is shown as: {sparql_results_for_prompts}.
...
(1) If the result is too large, please conduct statistic analysis for summary
(2) If the result is too small or even empty, please expand the retrieval scope using SPARQL keywords (syntax) OPTIONAL, UNION, | operator... '''



Ontology Subgraph Neighbourhood Based Recommendation



'''*Based on a natural language question: {question} ...*

Please recommend other potential SPARQL query patterns:

These are tips of generating the recommendations, only for your reference:

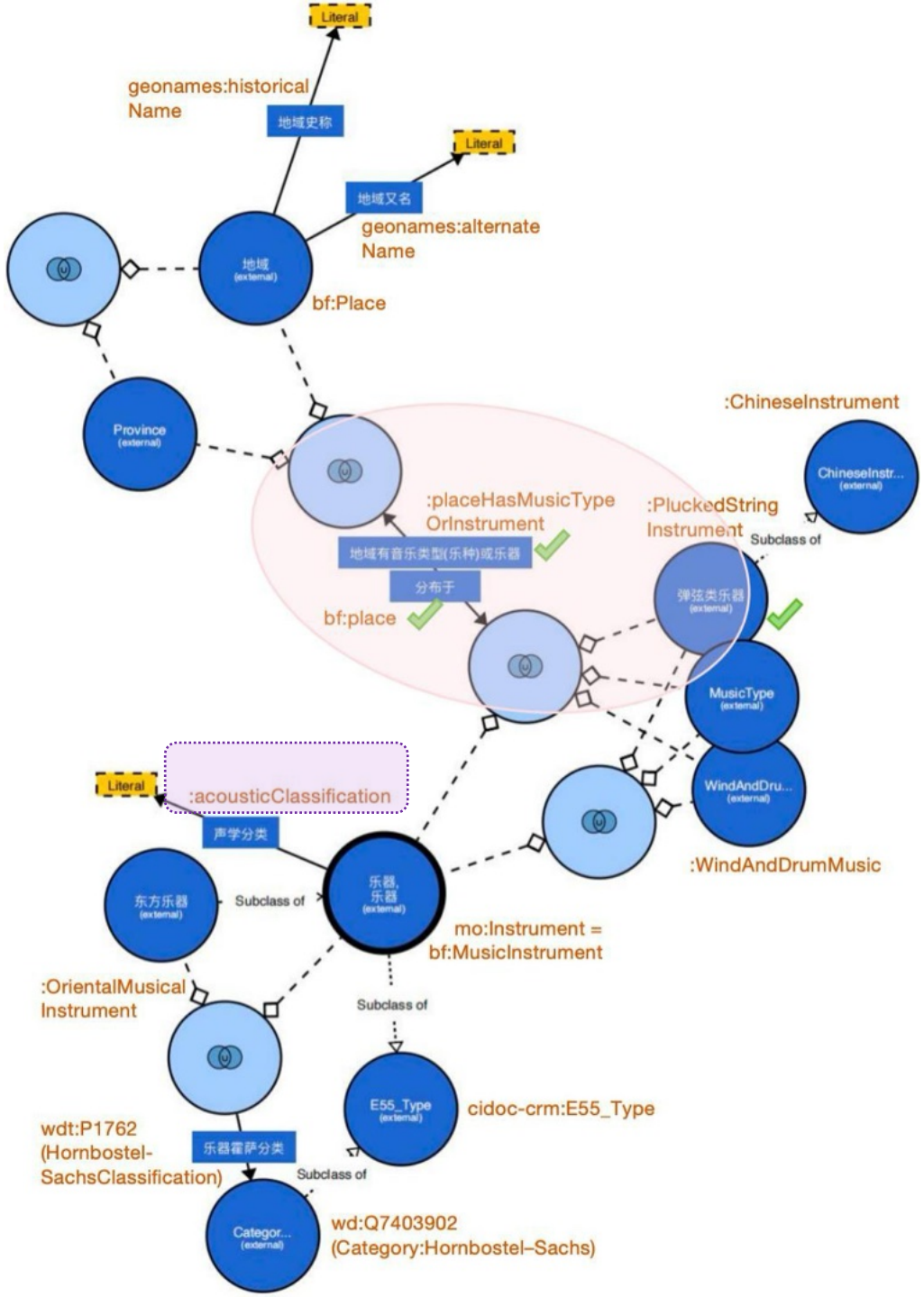
*1. **Identify the classes and properties in the ontology snippet that are used in the existing SPARQL query;*

*2. **Determine their current relationships and position in the ontology snippet;*

*3. **Expand to other adjacent classes or properties in the ontology snippet to recommend other possible query patterns that can yield more results;*

*- **This idea is regarding the ontology as a graph/network, and the recommendation is to explore other nodes (classes or properties) that are connected/adjacent to the ones embodied in the existing SPARQL query.*

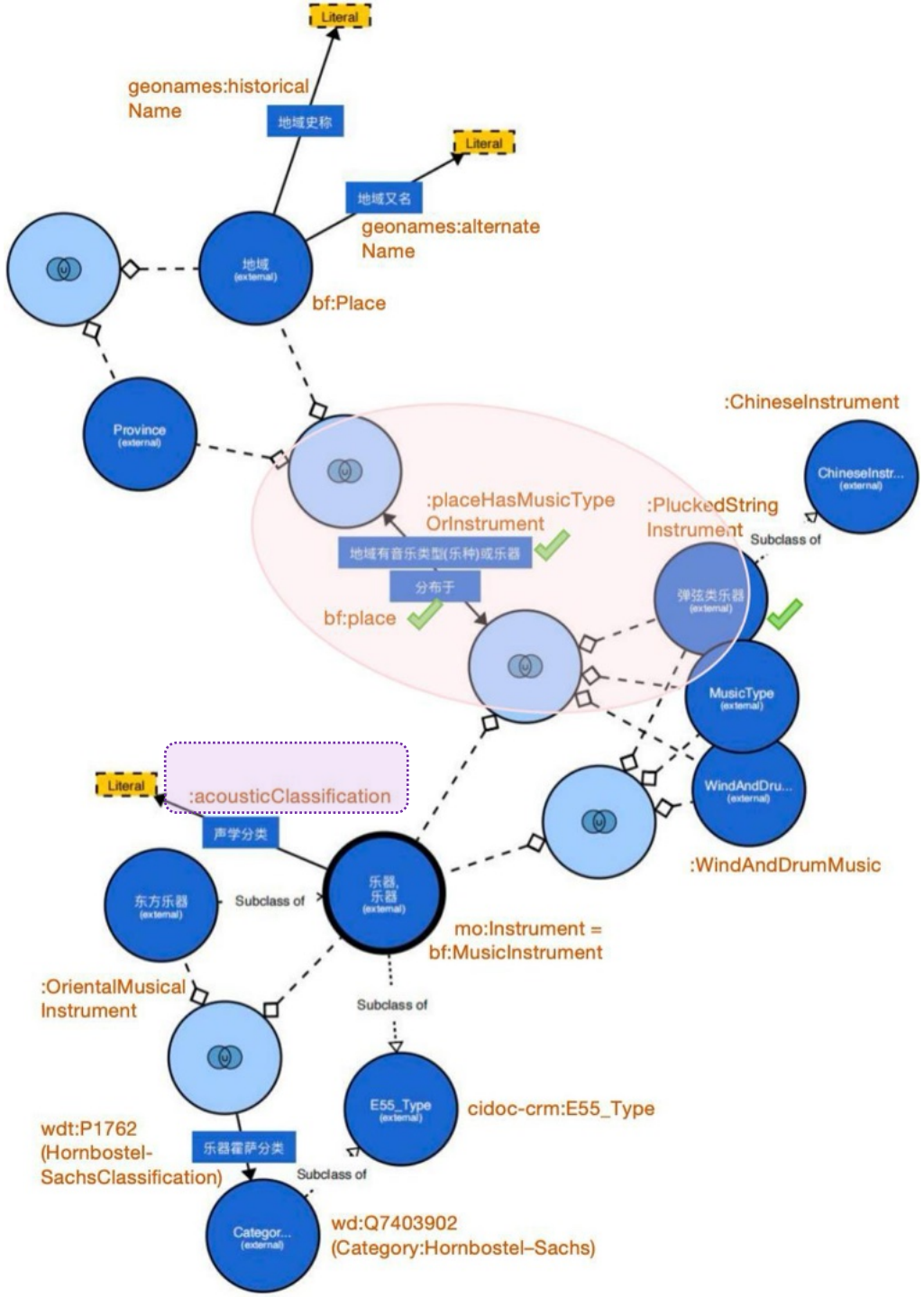
Return several SPARQL query patterns, along with the corresponding natural language questions, in a structured format. '''



- recommended Q&A pairs
- What acoustic classification does 东不拉 (Dombra) belong to, and what other instruments are in the same acoustic classification? Hornbostel—Sachs Classification
- In the areas where Dongbula is distributed, besides plucked string instruments, what other types of instruments are there?
- What music types or genres exist in the regions where Dongbula is distributed?
- What wind and drum music exists in the regions where Dongbula is distributed?

……

The pros and cons:




```
define input:inference 'urn:owl.ccmusicrules0913'
PREFIX bf: <http://id.loc.gov/ontologies/bibframe/>
PREFIX ctm:
<https://lib.ccmusic.edu.cn/ontologies/chinese_traditional_music#>
PREFIX mo: <http://purl.org/ontology/mo/>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT
DISTINCT ?acousticClass ?acousticClassLabel ?otherInstrument ?otherI
nstrumentLabel
WHERE {
  # Identify the instrument Dombra and its Hornbostel-
SachsClassification
  ?dongbula ctm:nameOfMusicTypeOrInstrument "东不拉" ;
            wdt:P1762 ?acousticClass .

  # Retrieve the label(s) of that acoustic classification
  OPTIONAL { ?acousticClass rdfs:label ?acousticClassLabel . }

  # Find other instruments within the same acoustic classification
  ?otherInstrument wdt:P1762 ?acousticClass ;
                   rdfs:label ?otherInstrumentLabel .

  # Exclude the Dombra itself
  FILTER(?otherInstrument != ?dongbula)}
  ORDER BY ?otherInstrumentLabel
```

Instructions for use of reasoning

(1) Adding DEFINE input:inference 'urn:owl.ccmusicrules0913' to the beginning of any SPARQL query code will activate the ontology-based inference function, thereby achieving the data completion effect.

(2) Adding DEFINE input:same-as "yes" to the beginning of any SPARQL query code can also activate knowledge reasoning based on instance equivalence relations.

SPARQL query example

0. Main entity thesaurus display (for reference in other searches)

- 0.1 Search for all music genres (song types)
  Exact match reference for other queries
- 0.2 Search for all musical instrument names
  Exact match reference for other queries
- 0.3 Query all region names
  Exact match reference for other queries
- 0.4 Search all ethnic groups
  Exact match reference for other queries
- 0.5 Query all (musicians)
  Exact match reference for other queries

Click to view the preset namespaces.

```
define input:inference 'urn:owl.ccmusicrules0913'
PREFIX bf: <http://id.loc.gov/ontologies/bibframe/>
PREFIX ctm: <https://lib.ccmusic.edu.cn/ontologies/chinese_traditional_music#>
PREFIX mo: <http://purl.org/ontology/mo/>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT DISTINCT ?acousticClass ?acousticClassLabel ?otherInstrument ?otherInstrumentLabel
WHERE {
  # Identify the instrument Dombra and its Hornbostel-SachsClassification
  ?dongbula ctm:nameOfMusicTypeOrInstrument "东不拉" ;
            wdt:P1762 ?acousticClass .
```

Table | Raw Response

Showing 1 to 143 of 143 entries Search: Show All entries

	acousticClass	acousticClassLabel	other Instruments	otherInstrumentLabel
1	3_String Instruments_6 Pizzicato	3_String Instruments_6 Pizzicato	https://lib.ccmusic.edu.cn/data/instrument/N79jx6849l3PMvO1	Seven Strings
2	3_String Instruments_6 Pizzicato	3_String Instruments_6 Pizzicato	https://lib.ccmusic.edu.cn/data/instrument/xFSxqwH4UYnfHImg	Seven-stringed zither
3	3_String Instruments_6 Pizzicato	3_String Instruments_6 Pizzicato	https://lib.ccmusic.edu.cn/data/instrument/XWv7aDGGqLxMv4Bk	Sanxian
4	3_String Instruments_6 Pizzicato	3_String Instruments_6 Pizzicato	https://lib.ccmusic.edu.cn/data/instrument/8UtXDe6FUC7Ecx1G	Musical bow
5	3_String Instruments_6 Pizzicato	3_String Instruments_6 Pizzicato	https://lib.ccmusic.edu.cn/data/instrument/t5e32p26goTsIARM	Yunhe
6	3_String Instruments_6 Pizzicato	3_String Instruments_6 Pizzicato	https://lib.ccmusic.edu.cn/data/instrument/Ca2w4dnWU5QlZ3TD	Five strings

The Pros and Cons

- 1. **pros**:
- (1) suitable for graph traversal/"chain-hop" query; for RDF datalake with expanding schema
- e.g., 苗族有哪些民间乐人，他们擅长什么乐种？那么这些乐种又会涉及什么乐器，它们的霍萨声学分类情况又如何……
- What folk musicians are there among the *Miao* ethnic group people and what kinds of music genre (MusicType) do they specialize in? Then what instruments are involved in these music genres, and what is the Hornbostel-Sachs acoustic classification of them...
- (2) transparency vs. intransparency/halucination -- see the following "ontology graph traversal based knowledge inference" -- *cliffhanger*

--embodied in fore-mentioned RAG's prompt

- 2. **cons**: (1) only applicable in RDF with OWL... What if there is no ontology in an RDF database? (2) e.g.,RISM: metadata with lots of nested structure -- *cliffhanger*

Workflow_Robustness enhancement—>Interactive NLQ2SPARQL Agent

- Why do we need recommendation?
- Why do we need interaction?
- (1) The question happens to have no answer —— embarassing
- (2) Human's natural language question (NLQ) is sometimes ambiguous
- (3) The necessity of "deeper seek": in—context learning vs. deep learning
- (4*) incompleteness of the ontology (OWL) subgraph

—>e.g....

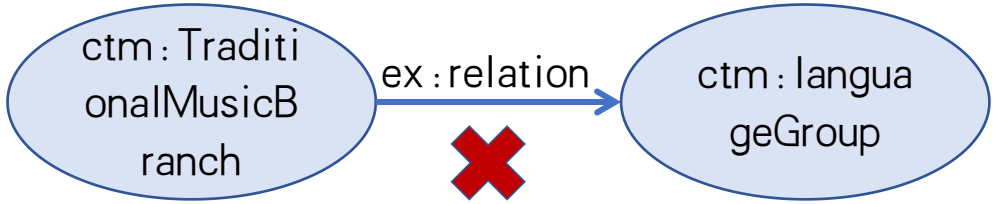
- core mechanism /ˈmekənɪzəm/:
- We prompt copilot as an agent to develop another agent, featuring context accumulation and in—context learning

Workflow_Robustness enhancement—>Interactive NLQ2SPARQL Agent

a lazy and vague query!

- (2)(3)(4)—>What's the relationship of Branch and Language Group?

(translated as:支脉和语族有什么关系？)



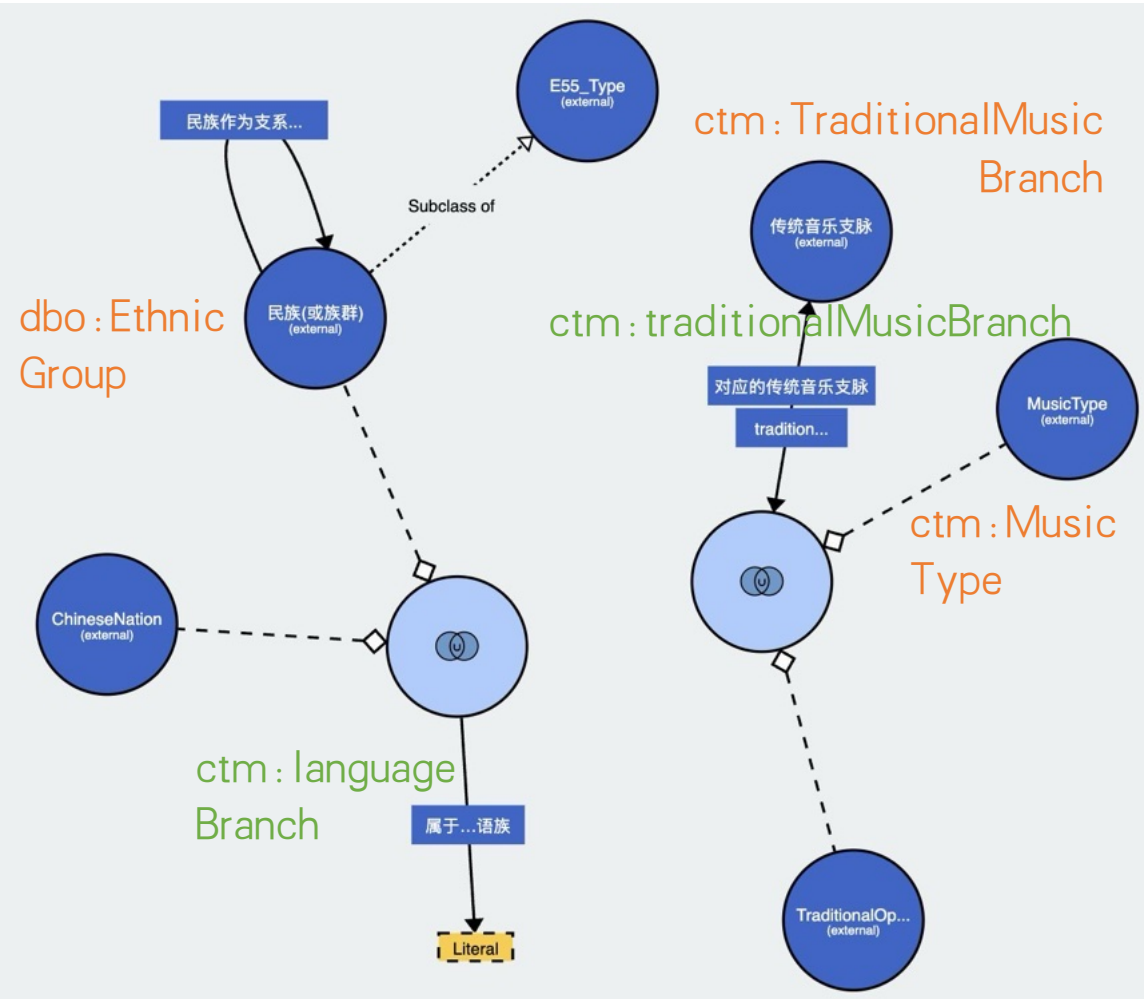
the clear version of the question:

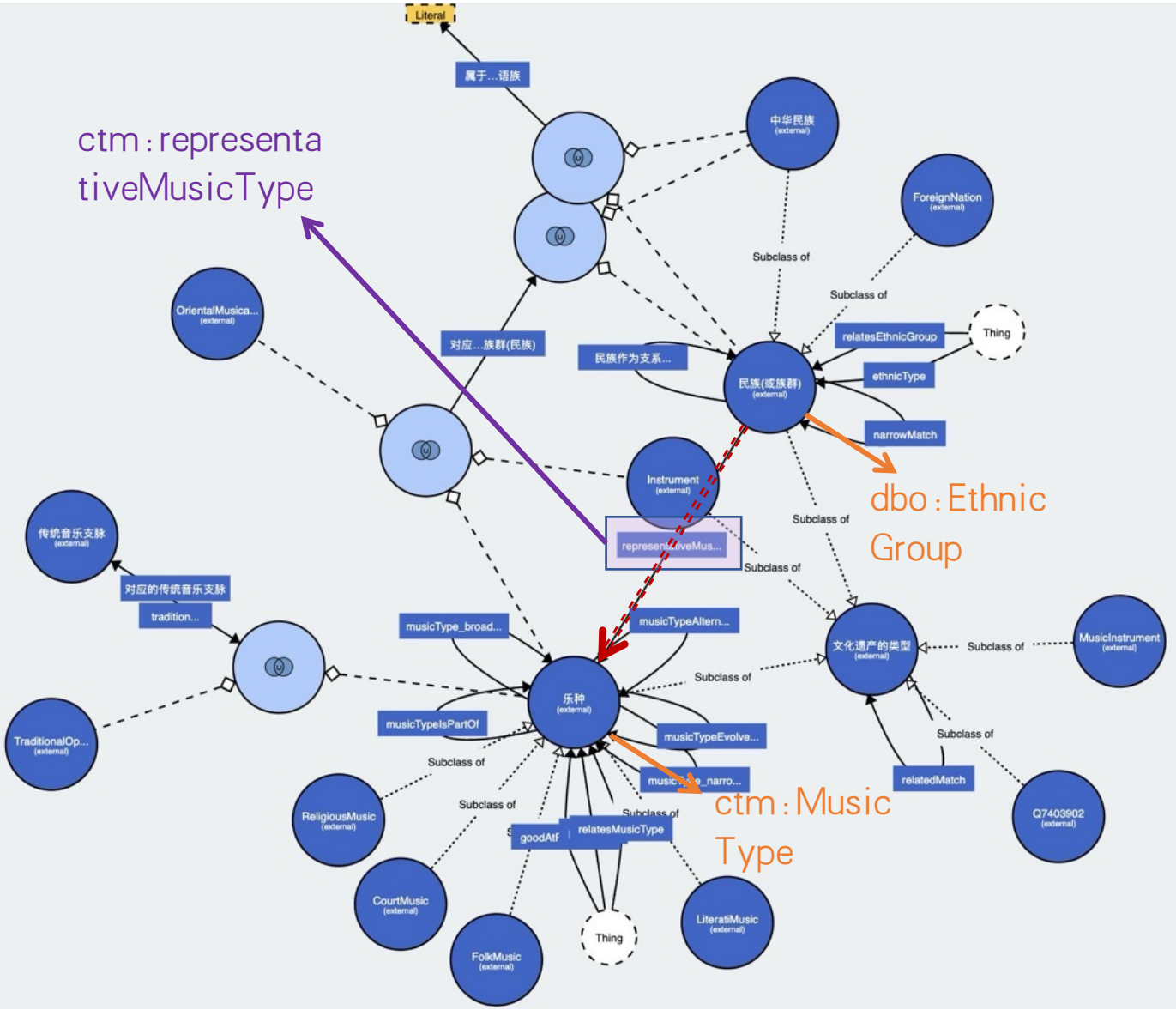
In the context of Chinese traditional musicology, what's the corresponding relationship between a "traditional music branch" and a "language group"(also known as language branch) which an ethnic group belongs to?

- Music Branch: a geography—cultural perspective for Chinese traditional music classification, e.g., Qinghai—Tibet plateau branch, Pamir branch. [22]
- Language Group: an ethnic—language perspective for classification, e.g.Tibeto—burman language group, Turkic language group (branch)

/plæˈtoʊ/ /ˈbɜːrmən/ /ˈtɜːrkɪk/

LLM： Knowledge Inference Based on OWL Graph Structure

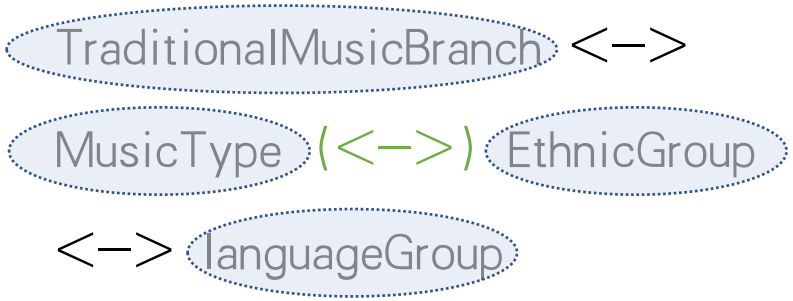




using BF shortest path algorithm—>ontology subgraph connectivity enhancement—>inference：

Interactive NLQ2SPARQL Agent

- 案例：What’s the relationship of Branch and Language Group?
- inference frame：




```
define input:inference 'urn:owl.ccmusicrules0913'
PREFIX ctm: <https://lib.ccmusic.edu.cn/ontologies/chinese_traditional_music#>
PREFIX dbpedia-owl: <https://dbpedia.org/ontology/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
# Query the relationship between traditional music branches (branch) and language families (languageBranch), using musical genres (MusicType) as a bridging entity between the two
SELECT
DISTINCT ?branch ?branchLabel ?languageBranch ?ethnicGroup ?ethnicGroupLabel
WHERE { # 乐种关联到传统音乐支脉
?musicType ctm:traditionalMusicBranch ?branch .
# 乐种关联到民族/族群
?musicType ctm:ethnicGroup ?ethnicGroup .
# 民族/族群具有语族属性
?ethnicGroup ctm:languageBranch ?languageBranch .
# 获取支脉的标签、民族/族群的标签
OPTIONAL { ?branch rdfs:label ?branchLabel . }
OPTIONAL { ?ethnicGroup rdfs:label ?ethnicGroupLabel . }}
```

Interactive NLQ2SPARQL Agent

- 案例：What’s the relationship of Branch and Language Group?
- interaction – deeper seek：
- Could u compile statistics on the frequency of MusicType that are jointly associated with each branch and language group? This would allow me to calculate the strength of the correspondence between them.

The strength of correlation

```
define input:inference 'urn:owl.ccmusicrules0913'
PREFIX ctm:
<https://lib.ccmusic.edu.cn/ontologies/chinese_traditional_music#>
PREFIX dbpedia-owl: <https://dbpedia.org/ontology/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT ?branch ?branchLabel ?languageBranch
       (COUNT(DISTINCT ?musicType) AS ?musicTypeCount)
WHERE { # MusicType -> Branch
  ?musicType ctm:traditionalMusicBranch ?branch .
  # MusicType -> EthnicGroup
  ?musicType ctm:ethnicGroup ?ethnicGroup .
  # EthnicGroup -> languageBranch(languageGroup)
  ?ethnicGroup ctm:languageBranch ?languageBranch .
  # obtaining the labels
  OPTIONAL { ?branch rdfs:label ?branchLabel . }}
GROUP BY ?branch ?branchLabel ?languageBranch
ORDER BY DESC(?musicTypeCount) ?branchLabel ?languageBranch
```

Instructions for use of reasoning

(1) Adding DEFINE input:inference 'urn:owl.ccmusicrules0913' to the beginning of any SPARQL query code will activate the ontology-based inference function, thereby achieving the data completion effect.

(2) Adding DEFINE input:same-as "yes" to the beginning of any SPARQL query code can also activate knowledge reasoning based on instance equivalence relations.

SPARQL query example

0. Main entity thesaurus display (for reference in other searches)

- 0.1 Search for all music genres (song types)
  Exact match reference for other queries
- 0.2 Search for all musical instrument names
  Exact match reference for other queries
- 0.3 Query all region names
  Exact match reference for other queries
- 0.4 Search all ethnic groups
  Exact match reference for other queries
- 0.5 Query all (musicians)
  Exact match reference for other queries

Click to view the preset namespaces.

```
1  define input:inference 'urn:owl.ccmusicrules0913'
2  PREFIX ctm: <https://lib.ccmusic.edu.cn/ontologies/chinese_traditional_music#>
3  PREFIX dbpedia-owl: <https://dbpedia.org/ontology/>
4  PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
5  SELECT ?branch ?branchLabel ?languageBranch
6         ( COUNT ( DISTINCT ?musicType ) AS ?musicTypeCount )
7  WHERE { # MusicType -> Branch
8    ?musicType ctm:traditionalMusicBranch ?branch .
9    # MusicType -> EthnicGroup
10   ?musicType ctm:ethnicGroup ?ethnicGroup .
11   # EthnicGroup -> languageBranch(languageGroup)
12   ?ethnicGroup ctm:languageBranch ?languageBranch .
13   # obtaining the labels
14   OPTIONAL { ?branch rdfs:label ?branchLabel . }}
15 GROUP BY ?branch ?branchLabel ?languageBranch
16 ORDER BY DESC ( ?musicTypeCount ) ?branchLabel ?languageBranch
```

Table | Raw Response

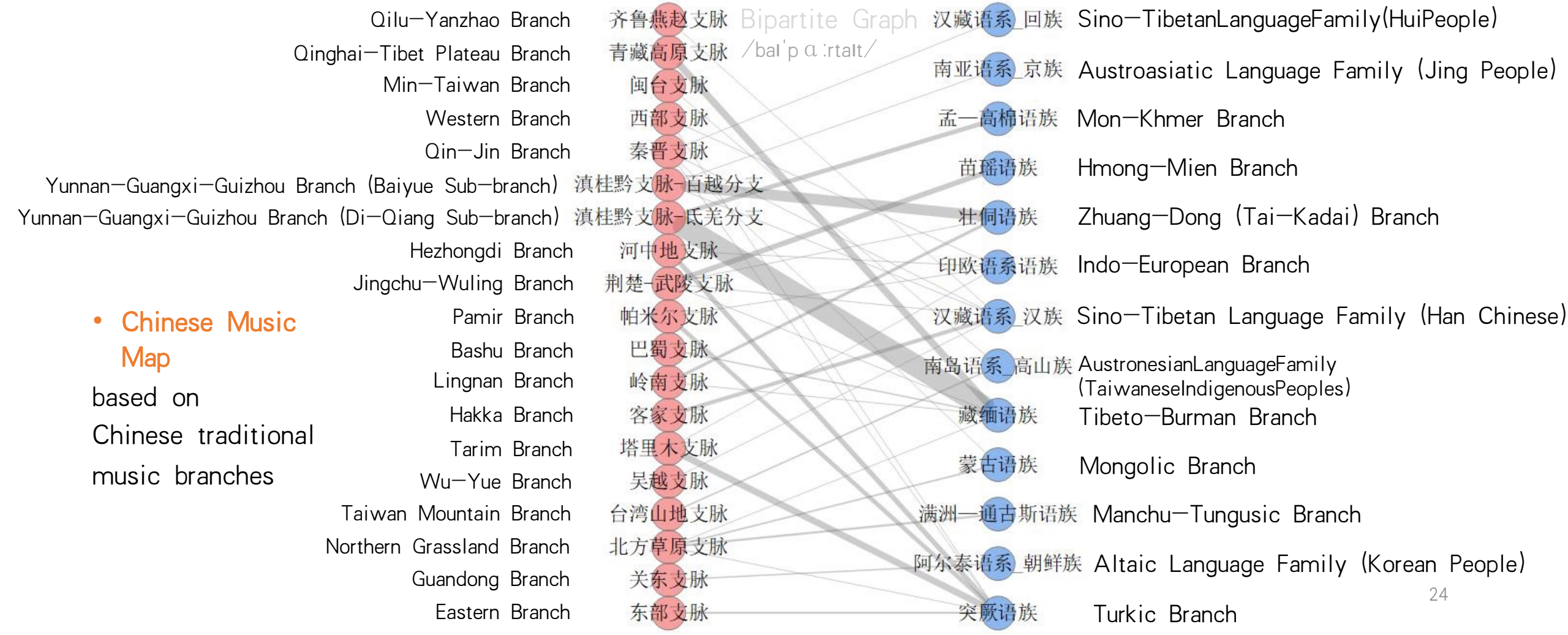
Showing 1 to 30 of 30 entries Search: Show All entries

	branch	branchLabel	languageBranch	musicTypeCou
1	https://lib.ccmusic.edu.cn/ontologies/chinese_traditional_music#Branch_YunnanAndGuangxiAndGuizhou_Dieqiang	Yunnan-Guizhou-Guangxi branch - Di-Qiang branch	Tibeto-Burman language family	"185" ^^xsd:intege
2	https://lib.ccmusic.edu.cn/ontologies/chinese_traditional_music#Branch_YunnanAndGuangxiAndGuizhou_Baiyue	Yunnan-Guizhou-Guangxi branch - Baiyue branch	Zhuang-Dong language family	"87" ^^xsd:integer
3	https://lib.ccmusic.edu.cn/ontologies/chinese_traditional_music#Branch_-Tarim	Tarim branch	Turkic language family	"70" ^^xsd:integer
4	https://lib.ccmusic.edu.cn/ontologies/chinese_traditional_music#Branch_Ti-betPlateau	Branch of the Qinghai-Tibet Plateau	Tibeto-Burman language family	"60" ^^xsd:integer
5	https://lib.ccmusic.edu.cn/ontologies/chinese_traditional_music#Branch_Wuling	Wuling branch	Miao-Yao language family	"34" ^^xsd:integer

Digital Humanities Application: correspondence between geo music branches and language branches / language families

Table | Raw Response

retrieved data can be downloaded as RDF

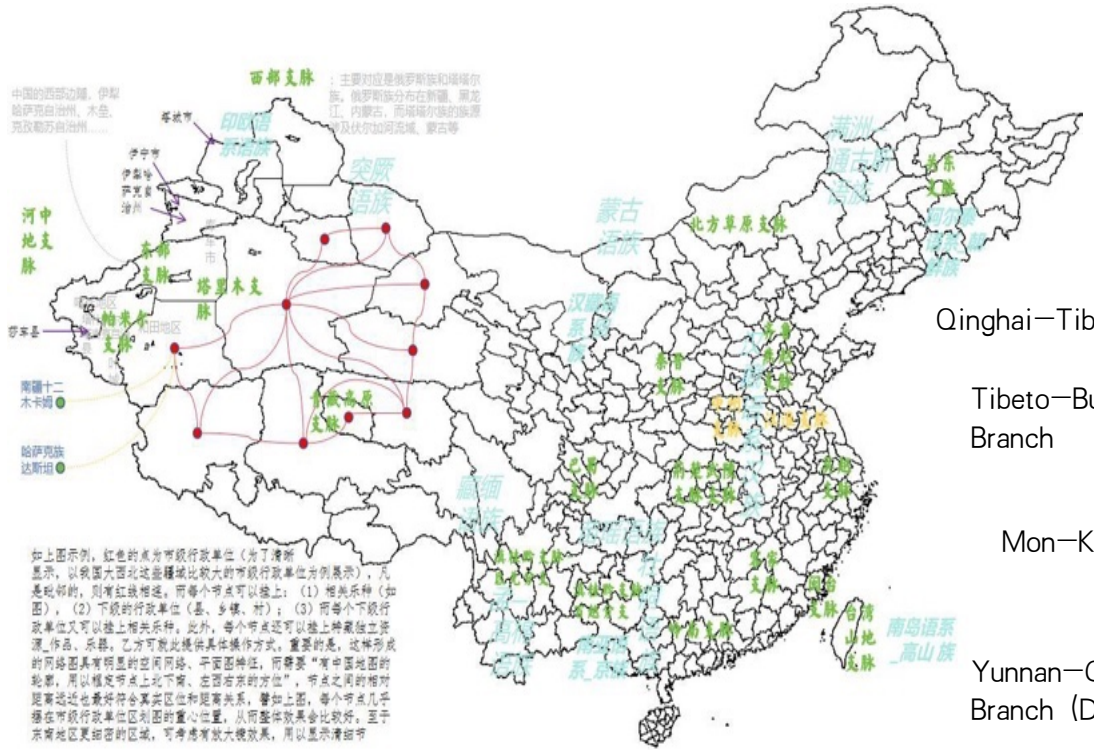


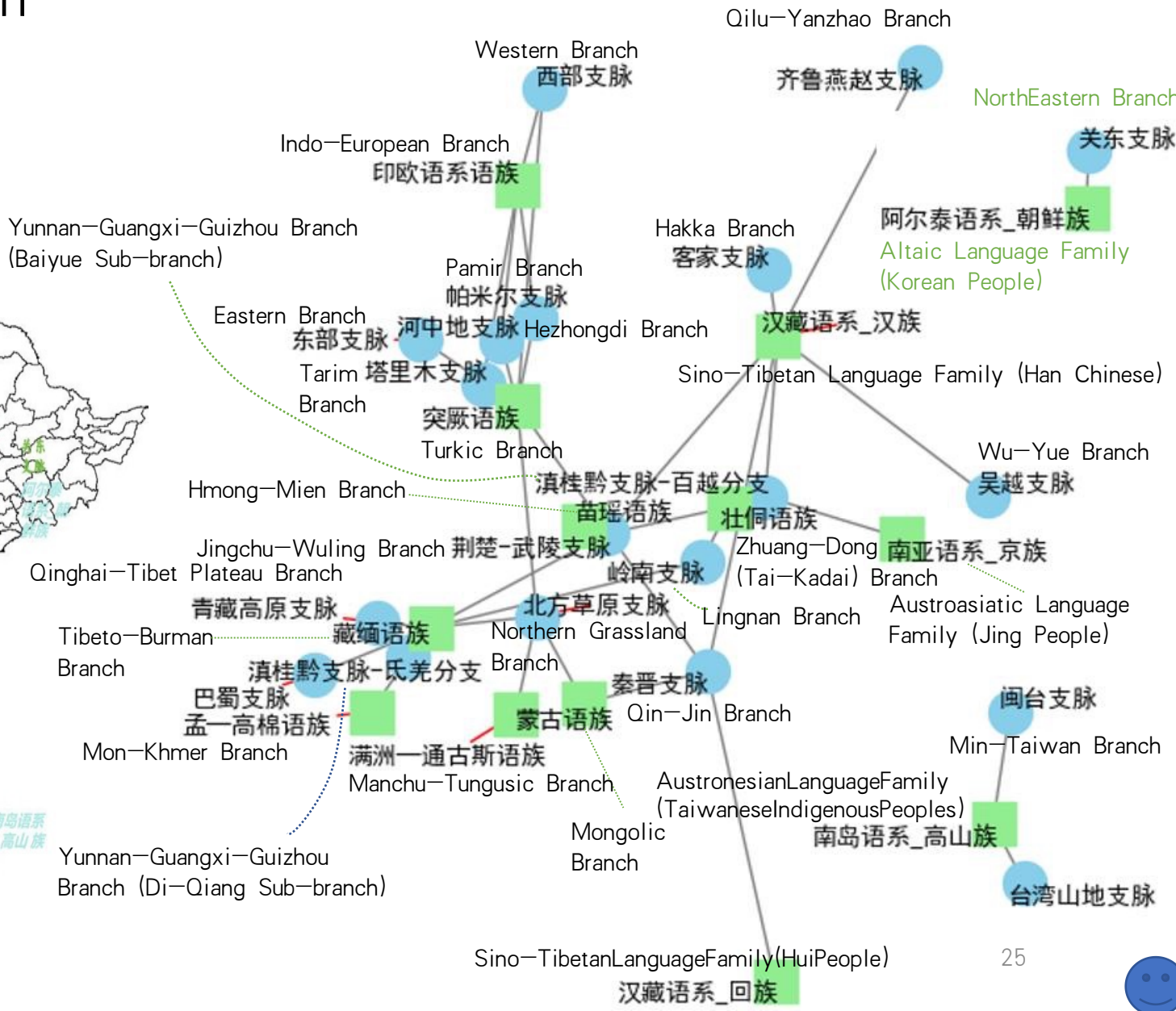
- Chinese Music Map

based on Chinese traditional music branches

Digital Humanities Application

- Chinese music map based on Traditional music branches





Observation

- LLMs are getting more and more concentrated and robust——despite facing a growing ontology
  - ——race between LLMs and sophistication of knowledge base schema/ontology
  - We may overthrow "ontology subgraph extraction" process...
- literature review："for those database without schema"…

prompt types	database type	NO. of shots	technology feature
RDF instance snippets	oversized, unknown schema	1—shot or few—shots	algorithm & computing—strength—oriented
example pairs of NLQ&SPARQL	e.g. XML—featured RISM	1—shot or few—shots	similarity matching between the given NLQ and the existing pairs of NLQ&SPARQL; vector database
"Schema"	with schema or latent schema	0—shots	knowledge—representation—oriented

about LinkedMusic, Future Work or Suggestion

- Datalake from *>14* databases with their respective schemas
- automatically extract ontology for any member database of LinkedMusic, using **VOID** (Vocabulary of Interlinked Datasets), which can be useful for describing the relationships among graphs within a datalake
- We don't have to face a giant schema/ontology for any individual database, so an interactive **federal-query** NLQ2SPARQL agent may be the future approach

THE SESSION

Find a session named "Hurley's Irish Pub" (2024 October 26)

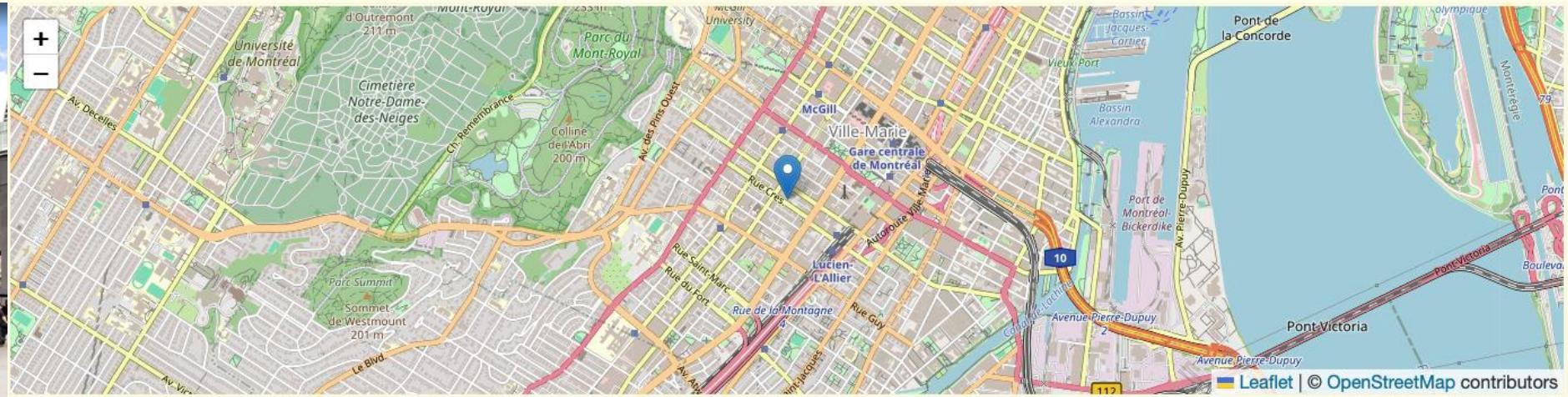
Hurley's Irish Pub

SHARE

1225 Crescent Street, Montreal, Quebec, Canada

Added by Loa 21 years ago.

Last updated with a comment by GW 9 months ago.



nearby members | nearby sessions | nearby events

Schedule: Saturday, Tuesday.

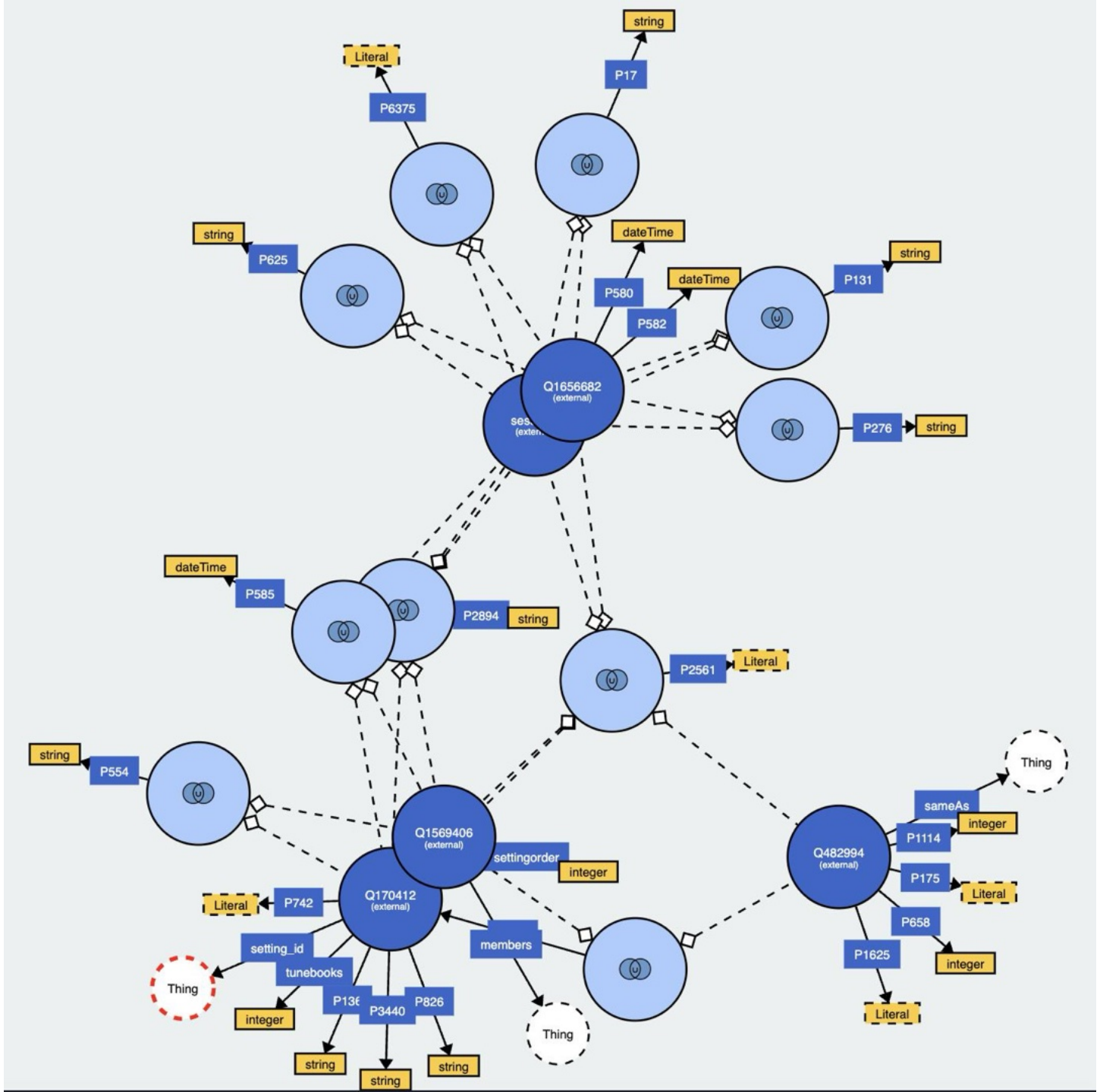
Telephone: 514.861.4111

Website: http://www.hurleysirishpub.com/

Latest: "https://www.siamsa.org/en/sessions"

↳ *Posted by GW 9 months ago.*

Hurley's Irish Pub



- –> after reconciliation with Wikidata, TheOntologyFor TheSession
- unionOf classes
  - light blue circles with "U"
- owl:objectProperty
  - between 2 nodes
- owl:dataProperty
  - yellow squares
- owl:Thing

Thanks and best wishes to members of DDMAL and LinkedMusic!

Specail thanks to Ichiro Fujinaga and Jenn Riley!

ESEA (East-and-Southeast-Asian) Traditional Music Knowledge Base and Its Ontology-Subgraph-Driven NLQ2SPARQL Intelligent Question-Answering System, *Cataloguing and Classification*, 2025

Junjun Cao (Shaojun Si)
Email: alienmusedh@gmail.com

Interactive NLQ2SPARQL Agent (Other NLQ Tests)

1. Which ethnic groups belong to China's musical system while also belonging to other musical systems?

2. Does Fujian Nanyin involve any special collections in our library/museum? If so, which instruments might those resources relate to?

3. In the southeastern region of China, which ethnic minorities have which music types/genres? What are the distribution areas of these genres, and which other instruments are found in those areas?

4. Which musicians specialize in which music types (genres)? Please return no more than 20 records, sorted in descending order by the number of genres each musician specializes in (i.e., most genres at the top).

5. Among instruments, which are associated with the Xinjiang Uyghur (/ˈwiː ɡ ʊər/) Autonomous Region or China's broader Northwest, especially plucked-string instruments (you may also consider this from the perspective of the Hornbostel – Sachs classification)?

6. What acoustic differences (e.g., in terms of Hornbostel – Sachs categories) might exist between instruments of ethnic groups belonging to the Tibeto-Burman branch and those belonging to the Kra – Dai (Zhuang – Dong) branch?

Interactive NLQ2SPARQL Agent（Other NLQ Tests）

1. 有哪些民族既属于中国音乐体系，又属于其他音乐体系？

2. 福建南音是否涉及我馆的某些特藏资源，若有，这些资源有可能涉及什么乐器？

3. 我国东南地区的少数民族有哪些音乐类型_乐种，这些乐种的分布地域各是什么，而这些分布地域又还有哪些其他的乐器？

4. 有哪些乐人擅长哪些音乐类型(乐种)？请返回给我不超过20条数据，并按照擅长乐种的数目做降序列出，即擅长乐种数量最多的排在最前面。

5. 我想看看乐器中，哪些是涉及新疆维吾尔自治区（地域）或中国的大西北地区的，尤其是它的弹拨乐器（也可以考虑从霍萨分类法的角度来考察）。

6. 属于藏缅语族的民族的乐器和属于壮侗语族的民族的乐器在声学特征（如霍萨分类方面）上可能有什么差异？

The musiconn Project

Current Insights and Future Directions

Bernhard Lutz (Bavarian State Library)

What is musiconn? What is FID?

- musiconn is the brand name of the FID Musicology
- FID stands for *Fachinformationsdienst* (= Specialized Information Service)
- The musiconn project has been run jointly by Bavarian State Library (BSB) and the Saxon State and University Library Dresden (SLUB) since 2018

Funded by
DFG Deutsche Forschungsgemeinschaft
German Research Foundation

Milestones in the provision of research literature and information infrastructure for musicology in Germany

Funded by
DFG Deutsche Forschungsgemeinschaft
German Research Foundation

What's new with FIDplus?

- After 12 years (the longest possible duration of funding), the FID format is running out
- The new FIDplus program now offers funding without temporal restrictions, with an evaluation taking place every five years
- The requirement is that the services are solid and mature

Funded by
DFG Deutsche Forschungsgemeinschaft
German Research Foundation

What are the tasks of musiconn?

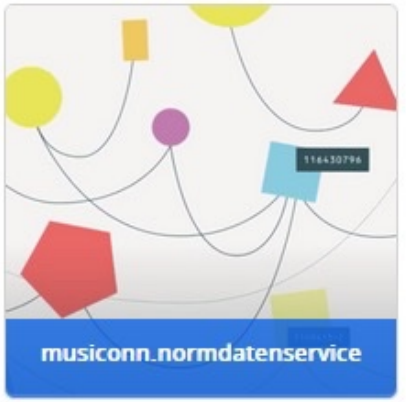
Acquisition of conventional and electronic media by the BSB

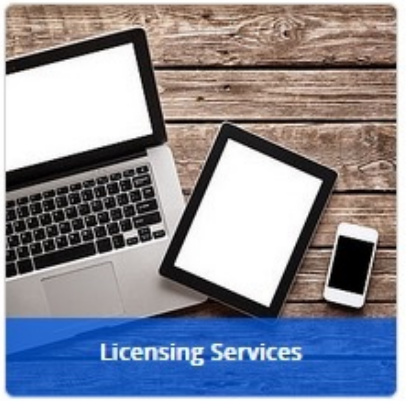
- Books, journals, and sheet music
- Electronic journals and databases

musiconn
MUSICOLOGY CONNECTED

What are the tasks of musiconn?

Provision of a technical information infrastructure for musicological research in Germany

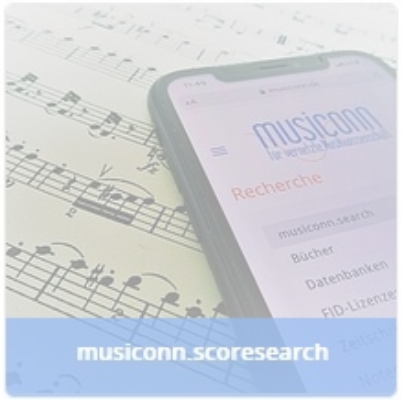


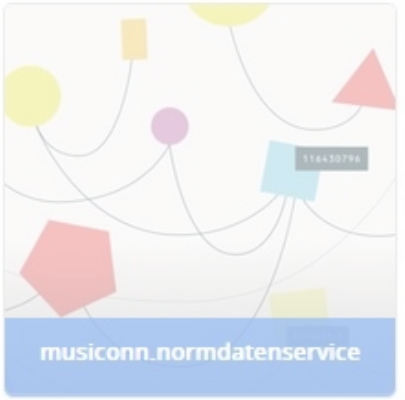


musiconn
MUSICOLOGY CONNECTED

What are the tasks of musiconn?

Provision of a technical information infrastructure for musicological research in Germany





musiconn
MUSICOLOGY CONNECTED

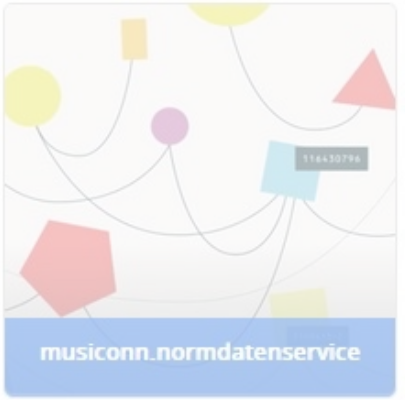
What are the tasks of musiconn?

Provision of a technical information infrastructure for musicological research in Germany

musiconn.scoresearch

musiconn.publish

musiconn.performance


musiconn.normdatenservice

musiconn.libretto

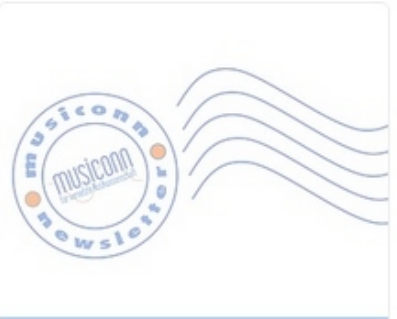
RISM Catalog

RIdIM Germany

Licensing Services

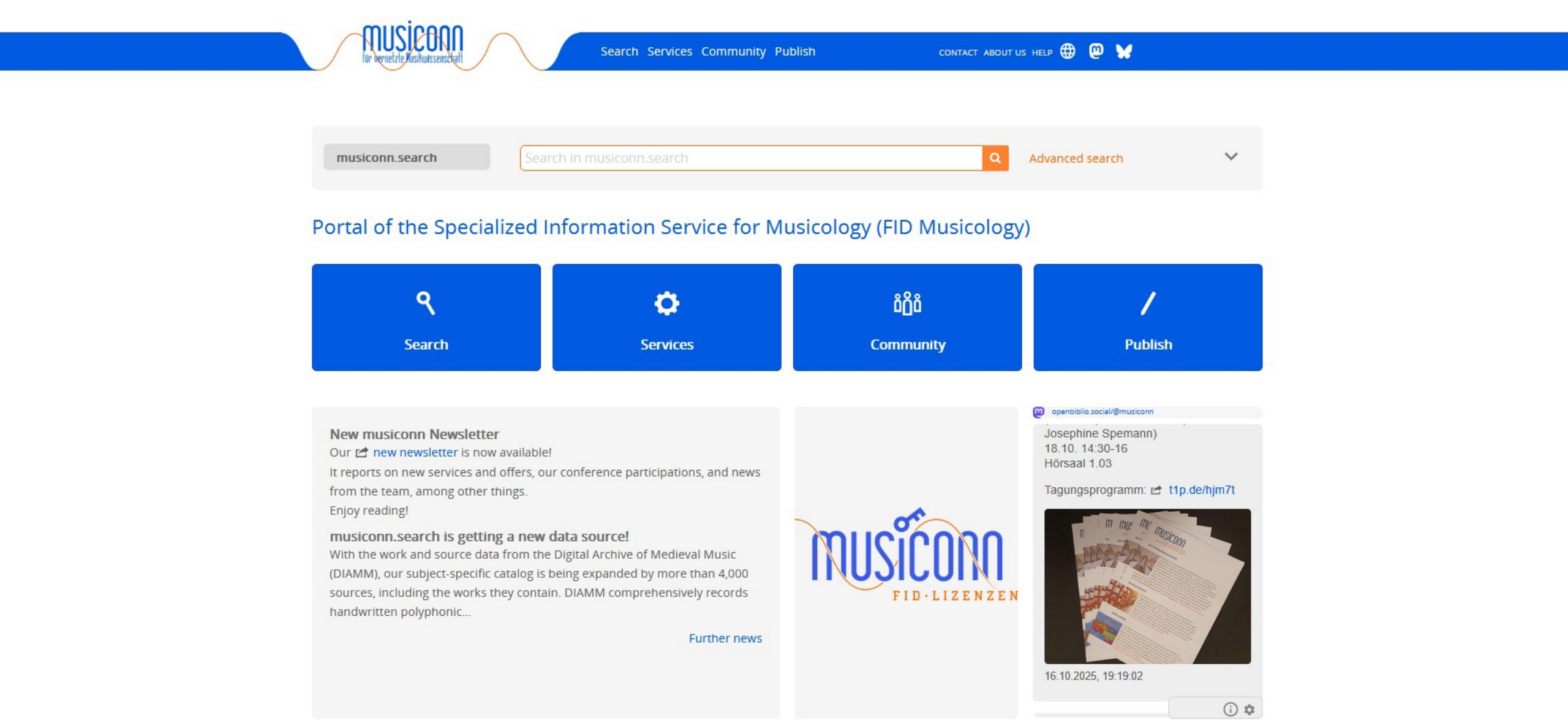
Web Archiving

Wish List



Newsletter

Recent Acquisitions

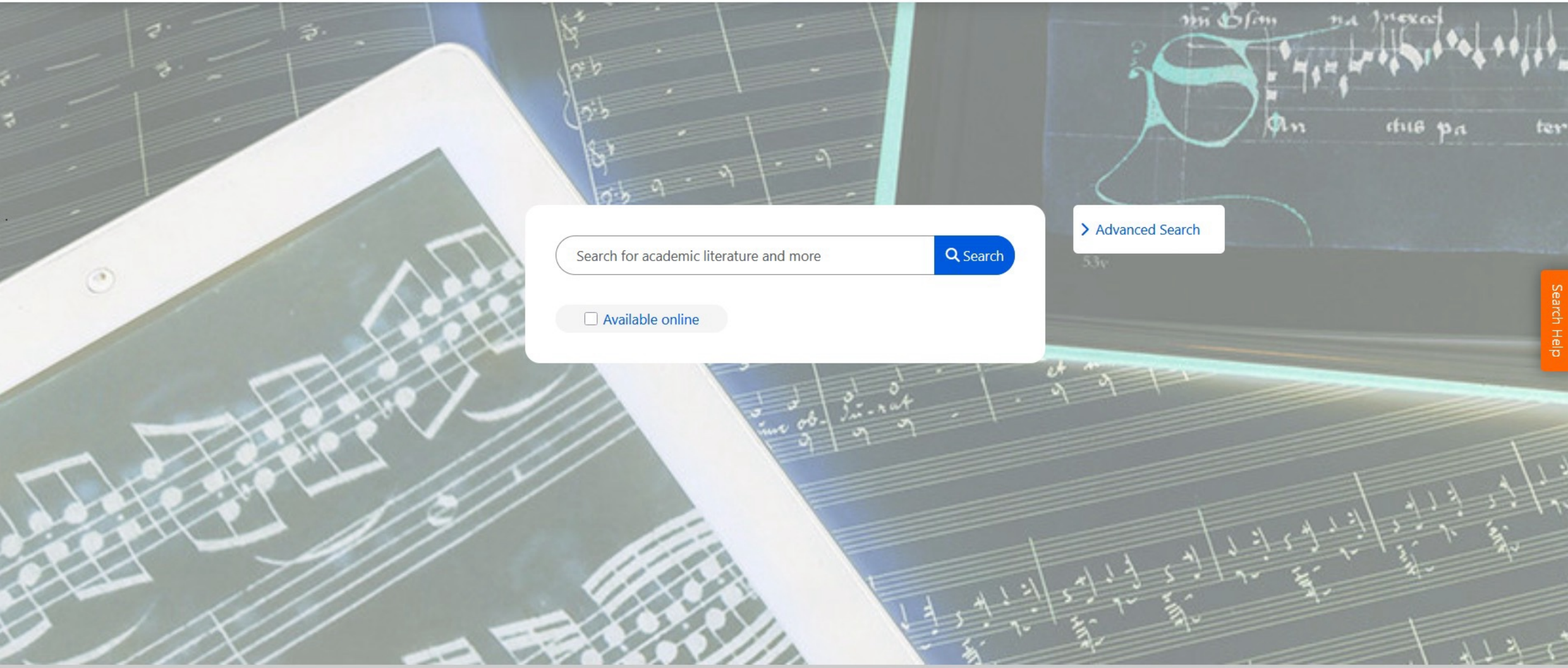

musiconn
für vernetzte Musikwissenschaft
Search Services Community Publish
CONTACT ABOUT US HELP
musiconn.search
Search in musiconn.search
Advanced search
Portal of the Specialized Information Service for Musicology (FID Musicology)
Search
Services
Community
Publish
New musiconn Newsletter
Our new newsletter is now available!
It reports on new services and offers, our conference participations, and news from the team, among other things.
Enjoy reading!
musiconn.search is getting a new data source!
With the work and source data from the Digital Archive of Medieval Music (DIAMM), our subject-specific catalog is being expanded by more than 4,000 sources, including the works they contain. DIAMM comprehensively records handwritten polyphonic...
Further news
musiconn
FID·LIZENZEN
openbiblio.social/@musiconn
Josephine Spemann)
18.10. 14:30-16
Hörsaal 1.03
Tagungsprogramm: t1p.de/hjm7t
16.10.2025, 19:19:02

musiconn
SEARCH

musiconn.search

- Simultaneous search in currently 18 databases on music and musicology
- Solr index, web application via the vufind discovery system
- Mapping of data using a specific FID-XML
- Integrated full-text search
- Integrated Mirador Viewer (IIIF)
- Display of authority data via LDBrowser

- [x] **All Data Sources**
- [ ] Austrian National Library, Music Catalogue (336386)
- [ ] BASE: Search Engine for Academic Resources (5979)
- [ ] Bavarian Musicians, Dictionary (27693)
- [ ] Berlin State Library, Music catalogue (578803)
- [ ] British Library: Music Prints (1097998)
- [ ] DBIS (Database Information System) (265)
- [ ] DIAMM (Digital Image Archive of Medieval Music) (4619)
- [ ] EZB (Electronic Journals Library) (1099)
- [ ] German Music Archive, Music Prints and Sound Carriers (797530)
- [ ] Historical Scores of the Bavarian State Opera (9765)
- [ ] Music Literature on-line, Bibliography (475755)
- [ ] Musical sources of the Library of Congress (270000)
- [ ] MusicOPAC at BSB (912255)
- [ ] OLC-SSG: Journals, Tables of Content (212646)
- [ ] Petrucci Music Library, IMSLP (101430)
- [ ] RIdIM (Music iconography) (24771)
- [ ] RISM (International Inventory of Musical Sources) (1599714)
- [ ] Web Resources of musiconn (3284)


musiconn
SEARCH
Book Request
My Searches
0 Favorites
Login
Language
Search for academic literature and more
Search
Available online
Advanced Search
Search Help
DFG Deutsche Forschungsgemeinschaft
BSB Bayerische StaatsBibliothek
Information in erster Linie
SLUB
Wir führen Wissen.

musiconn.search – future directions

- Enhancement of musiconn.search in the area of music notation-based search entries
- Potential integration of additional data sources based on our modifiable FID-XML
- AI-based further developments in the areas of data curation, index optimization, and user guidance

musiconn
PERFORMANCE

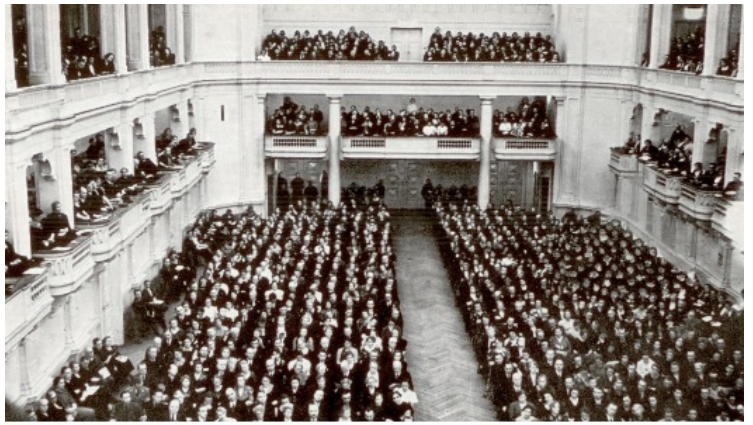
musiconn.performance

- Repository for musical events and related data on works, people, places, and institutions
- musiconn.performance currently contains almost 160,000 records on source-documented musical performances from over twenty different research projects

KIK - KOMPONISTINNEN IM KONZERT

How is the data entered?

- Transfer of existing data sets from completed research projects
- Direct entry of concert data into musiconn.performance by project partners of the scientific community



FRANKFURTER MUSEUMSGESELLSCHAFT

musiconn PERFORMANCE

Recherche | Projekte | Sammlungen | FAQ | Über uns | API

musiconn.performance, ein Angebot des Fachinformationsdienstes Musikwissenschaft, dient der komfortablen Recherche nach musikalischen Aufführungen und vernetzt die Forschungsdaten und -ergebnisse verschiedenster Partner.

Suche

suchen

Unsere Datenbank beinhaltet gegenwärtig mehr als **161.755 Datensätze** zu musikalischen Aufführungen und assoziierten Personen, Werken und Aufführungsstätten.

PERSONEN

KÖRPERSCHAFTEN

ORTE

WERKE

EREIGNISSE

QUELLEN

musiconn.performance - further directions

- Provision of a *Datalab*
  - Experimental environment for data retrieval and data analysis
  - Implementation of a musiconn.performance API

- Provision of a *Toolkit*
  - Compilation of suitable digital humanities tools: Data cleaning, Data visualization, Network analysis, Data extraction and structuring

musiconn
SCORESEARCH

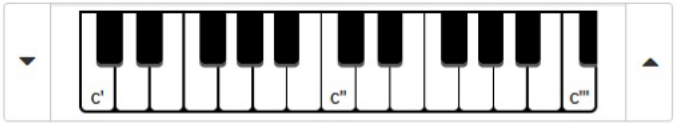
musiconn.scoresearch

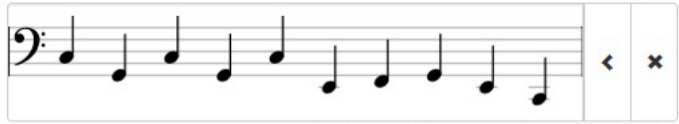
- musiconn.scoresearch offers you a program to search for melodies in selected digitized music prints
- Over 159,000 individual scans as searchable OMR files
- Implementation of music recognition using proprietary SmartScore software
- Repertoire: Works from the 18th & 19th centuries

musiconn
SCORESEARCH
Search
About this project
Help
Contact
DE | EN
c'
c''
c'''
Find transposed results
Exact search only
Please give at least 3 notes
Imprint | Privacy policy | Accessibility
musiconn
MUSICOLOGY CONNECTED
BSB Bayerische StaatsBibliothek
Information for Excellence
Funded by
DFG Deutsche Forschungsgemeinschaft
German Research Foundation
Your visit to this website is currently being monitored by Matomo Web Analytics. The tracking is completely anonymous and serves the continuous improvement of this website.
Click this link to disable tracking (opt out)

musiconn SCORESEARCH

Search | About this project | Help | Contact | DE | EN





Find transposed results | Exact search only

Search

Collection

Franz Liszt (1)

Ludwig van Beetho-ven (1)

Composer

Beethoven, Ludwig van, (1)

Liszt, Franz, (1)

Year of publication

1862 (1)

1926 (1)

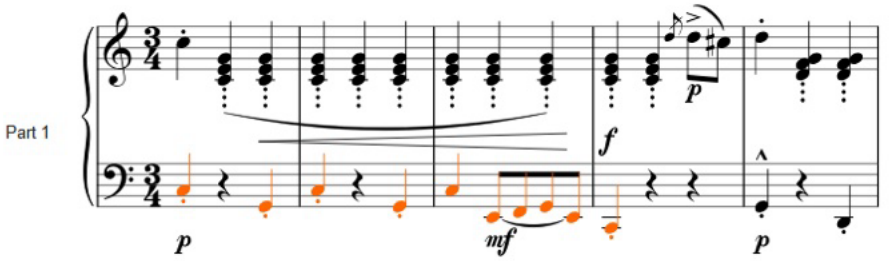
Time range

1860 - 1879 (1)

1920 - 1939 (1)

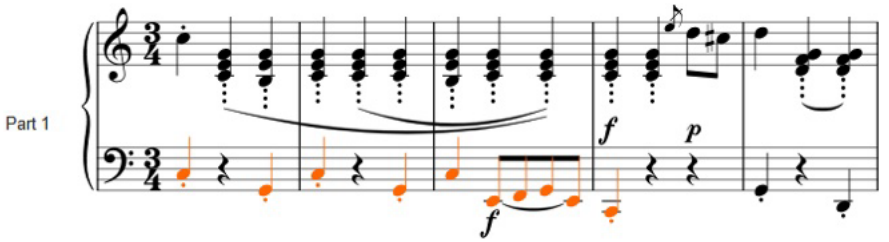
Results (2) Searched pitches: Searched intervals:

1. Musikalische Werke, 2,7: Verschiedene Werke für Pianoforte zu zwei Händen, Pg.11
Liszt, Franz
Leipzig : Breitkopf & Härtel



Read Online | Preview | MusicXML | Play

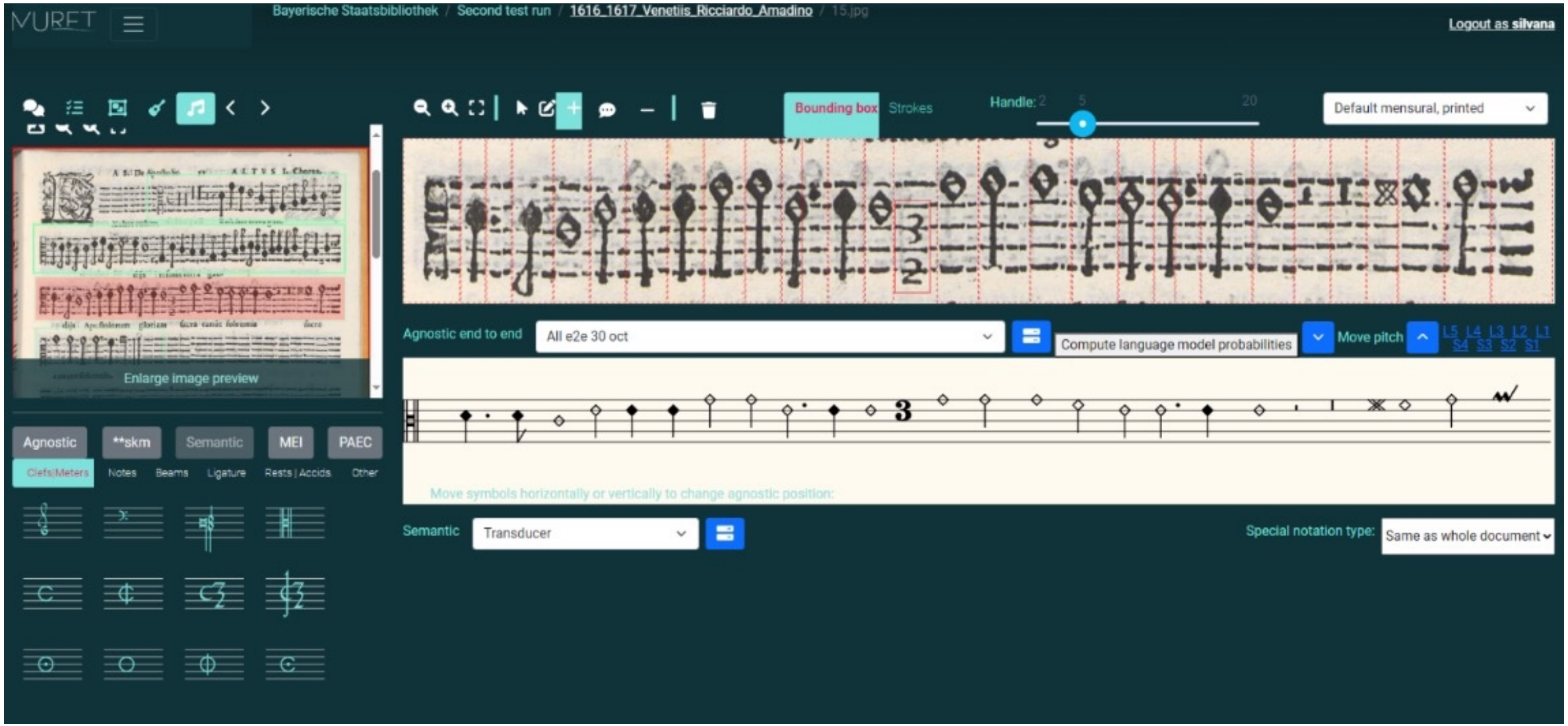
2. Beethoven's Werke, 165 = Serie 17: Variationen für das Pianoforte: 33 Veränderungen über einen Walzer von A. Diabelli : op. 120, Pg.3
Beethoven, Ludwig <<van>>, Beethoven, Ludwig van
Leipzig : Breitkopf und Härtel



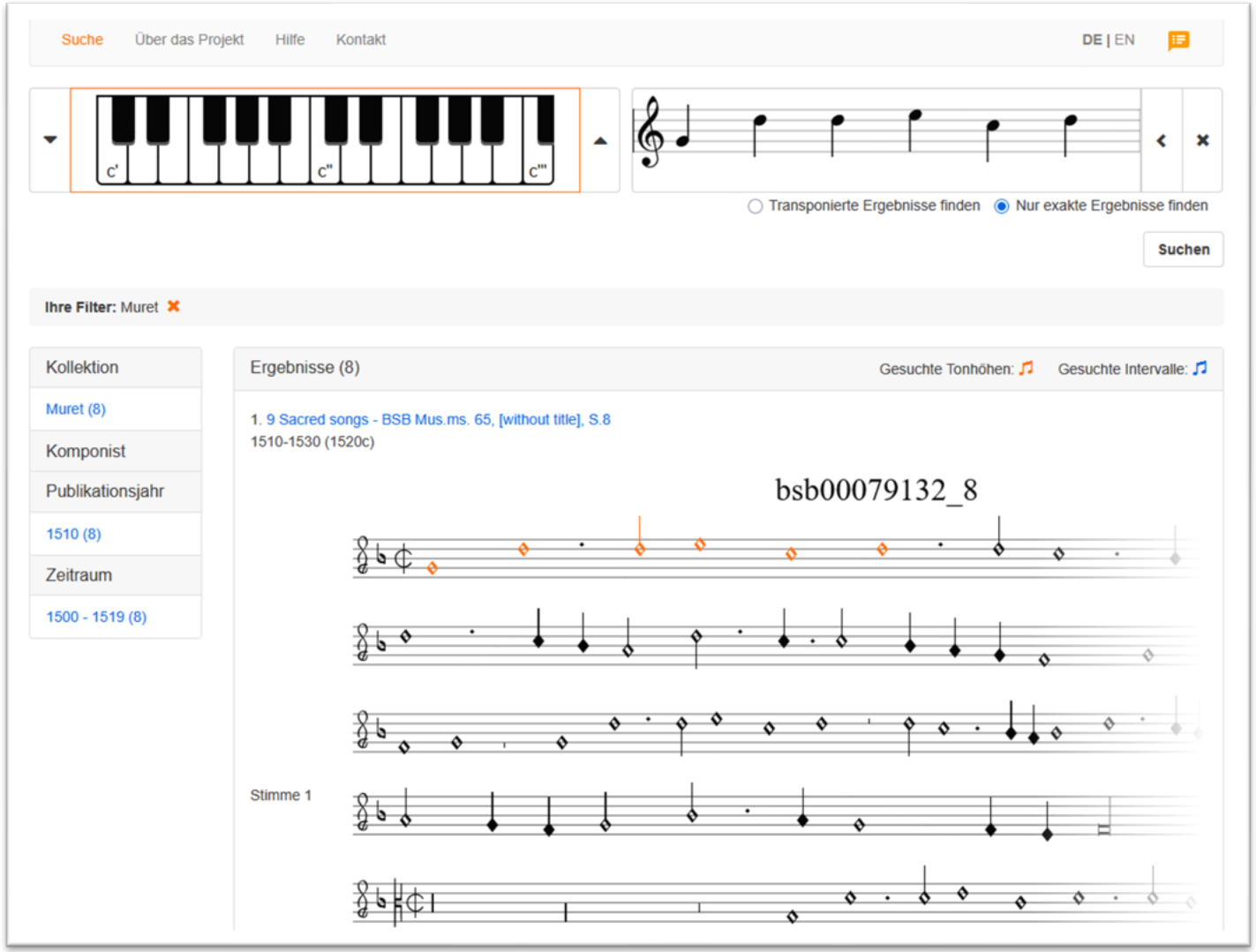
Current developments

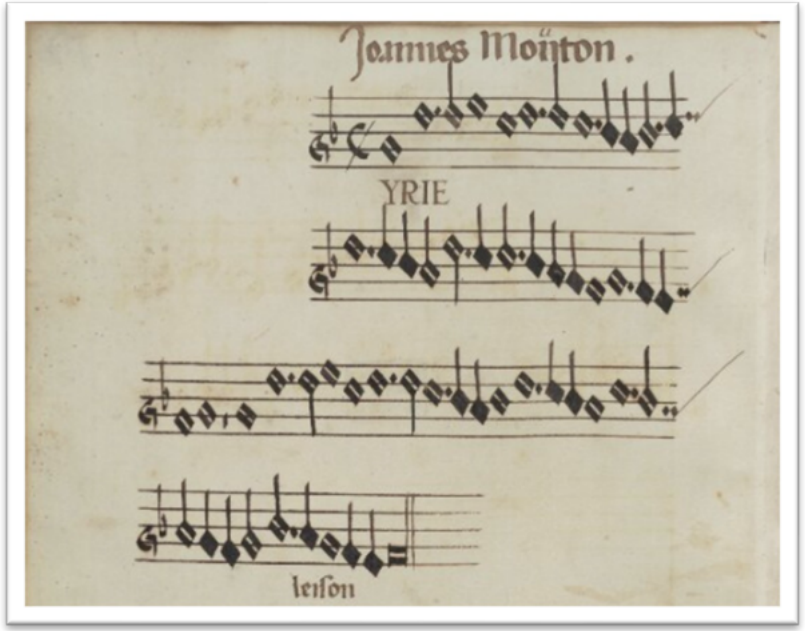
- Transition from MusicXML to MEI (in progress)
- Expansion of the repertoire to include music sources from the 16th & 17th centuries
- Approximately 1,800 music sources from the Bavarian State Library
- Use of the OMR software MuRET developed at the University of Alicante as part of a cooperation with the group of Prof. David Rizo
- MuRET training by the musiconn team

musiconn
MUSICOLOGY CONNECTED



musiconn
MUSICOLOGY CONNECTED





D-Mbs Mus.ms. 65

musiconn.scoresearch – future directions

- Integration of OMR data into digitization workflows
- Repertoire expansion with data from other libraries and academic edition projects
- Publication of our MEI parser
- Improvement of search and filter options

Further information on our website
www.musiconn.de

Get in touch with us
info@musiconn.de

Bernhard Lutz (Bavarian State Library)

UMIL: Universal Musical Instrument Lexicon

December 13
LinkedMusic Project Meeting 2025

Presented by:

Geneviève Gates-Panneton, Kun Fang, Kyrie Bouressa,

Mai Lyn Puittinen, Pouya Mohseni, Yu-Chia Kuo, Zih-Syuan Lin

Part One: Overview

Presenter: Kun Fang

What is UMIL (Universal Musical Instrument Lexicon)?

- We collect **musical instrument data**.
- We specially focus on **multilingual** support.
- We provide a simple, user-friendly way to **contribute** instrument data to Wikidata.

UMIL

921	625	31,017	19
Instruments	Languages	Names	Editors

Thanks to all the contributors for developing, testing, or providing suggestions and ideas！

Yinan Zhou, Dylan Hillerbrand, Kun Fang, Mai Lyn Puittinen, Antoine Phan, Geneviève Gates-Panneton, Kyrie Bouressa, Anna de Bakker, Yu-Chia Kuo, Pouya Mohseni, Caroline Guo, Zih-Syuan Lin, Andrew Hankinson

Why we need UMIL?

- Wikidata has huge but fragmented musical instrument data.
- It’s hard to browse all instruments at once.
- For new contributors, it may require steep learning curve.
- Without an account, editing privileges are limited.

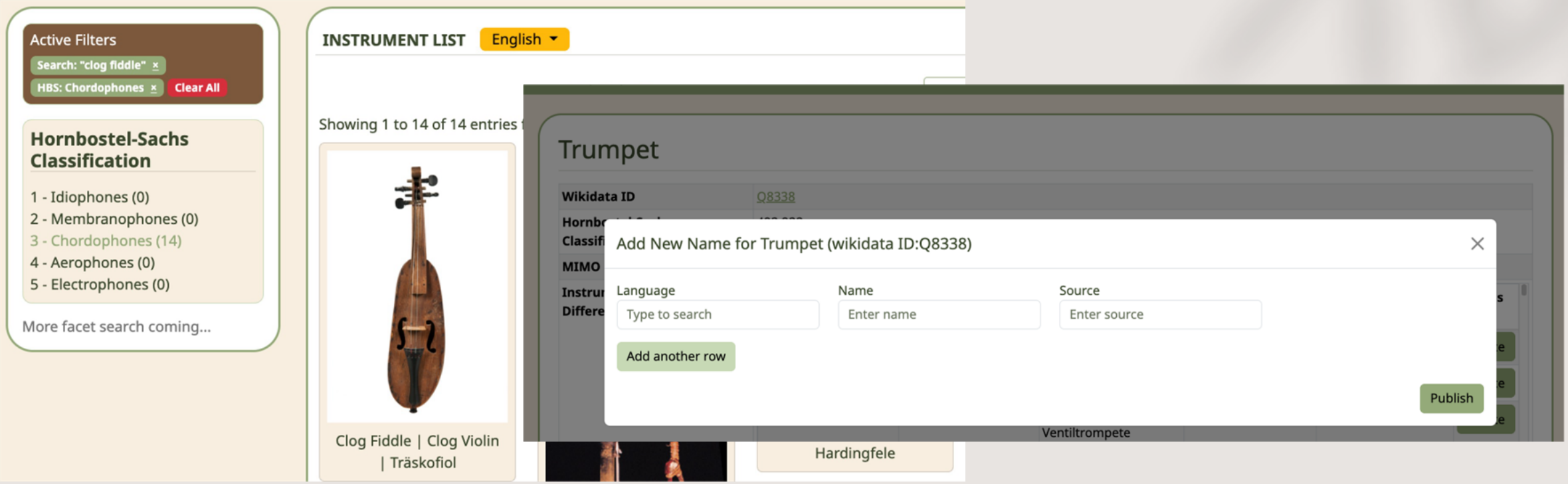
However, UMIL is designed to solve these issues…

UMIL

- UMIL is a one-stop platform for browsing all instruments at once.
- UMIL supports instrument view in any language.
- Most importantly, UMIL makes it easier and more convenient for everyone to contribute — No Wikidata account needed

What we have in UMIL so far…

- Browsing all instruments in any language.
- Support HBS classification, and fuzzy search…
- Main feature: users can add new names/aliases to existing instruments.



What we're going to talk about today…

- **Part Two: Add-Instrument-Name Workflow** *(Presenter: Mai Lyn Puittinen)*
- **Part Three: Issues with Instrument Taxonomy and Classification in UMIL** *(Presenter: Kyrie Bouressa)*
- **Part Four: Challenges & Open Discussion** *(Presenter: Pouya Mohseni, Zih-Syuan Lin, Yu-Chia Kuo, Mai Lyn Puittinen, Geneviève Gates-Panneton)*

Part Two: Add-Instrument-Name Feature

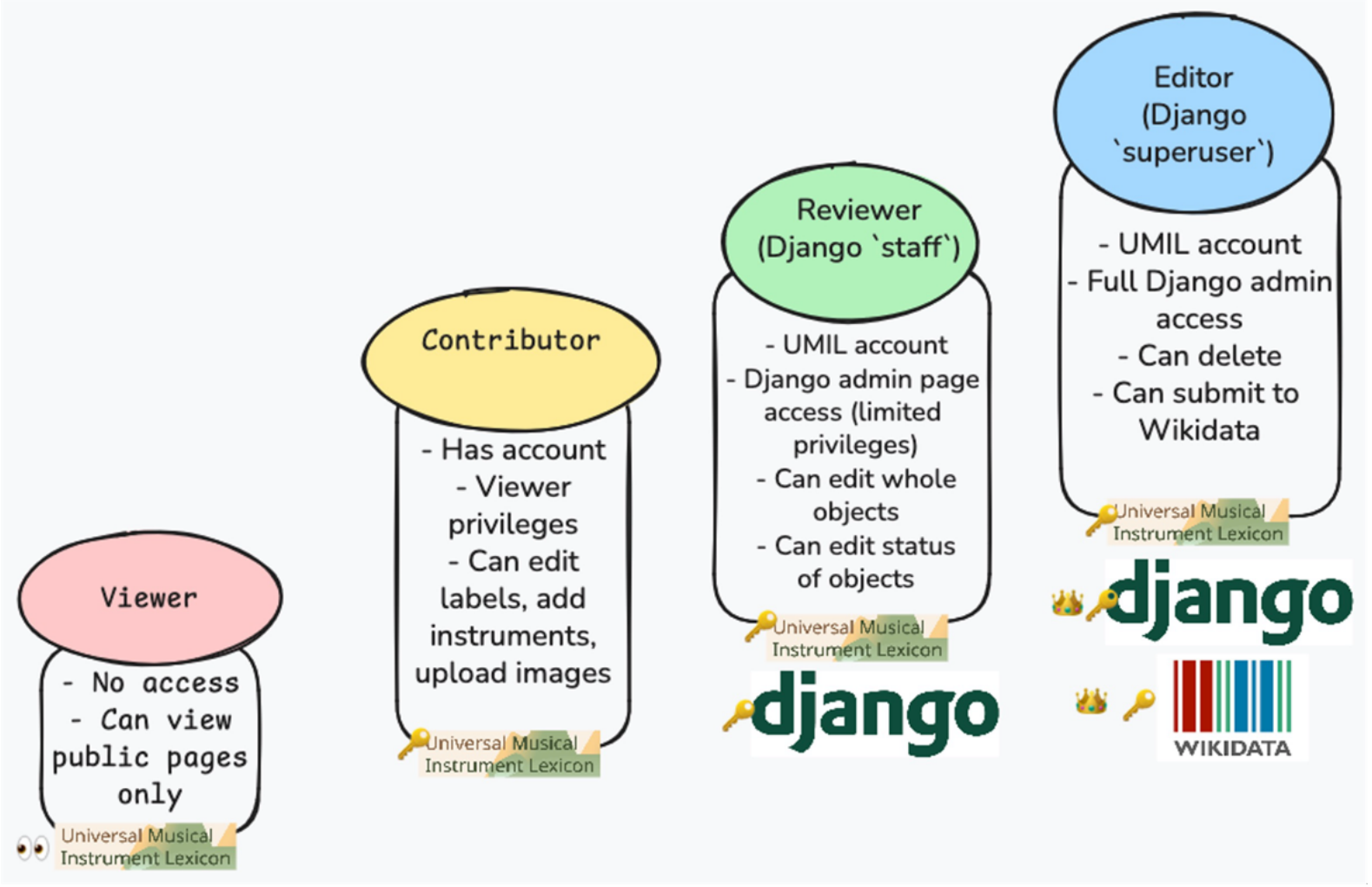
Presenter: Mai Lyn Puittinen

UMIL: Add-Instrument Name Feature

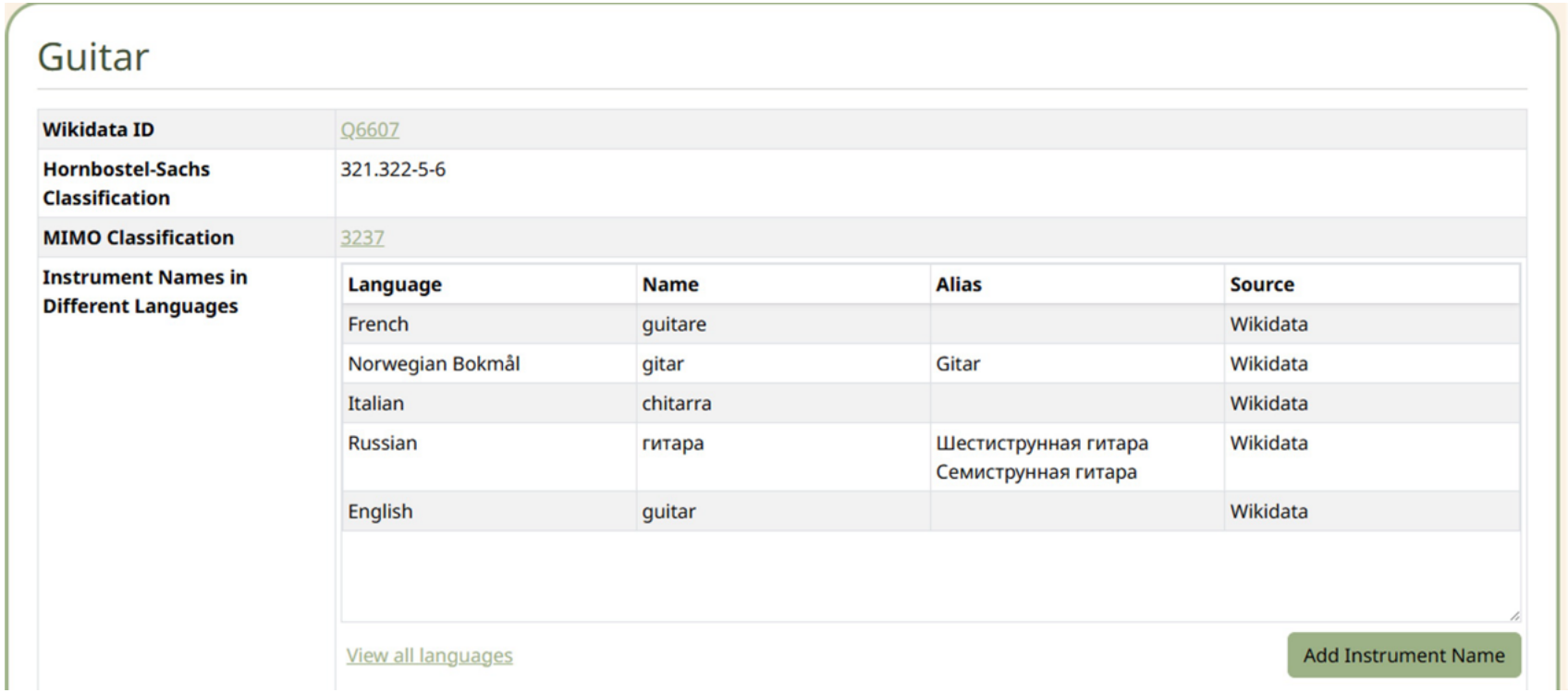
Development of Backend Name Verification Workflow

13 December 2025
LinkedMusic Project Meeting 2025
Mai Lyn Puittinen

Defining UMIL's Users



Viewer



Guitar

Wikidata ID	Q6607
Hornbostel-Sachs Classification	321.322-5-6
MIMO Classification	3237

Instrument Names in Different Languages

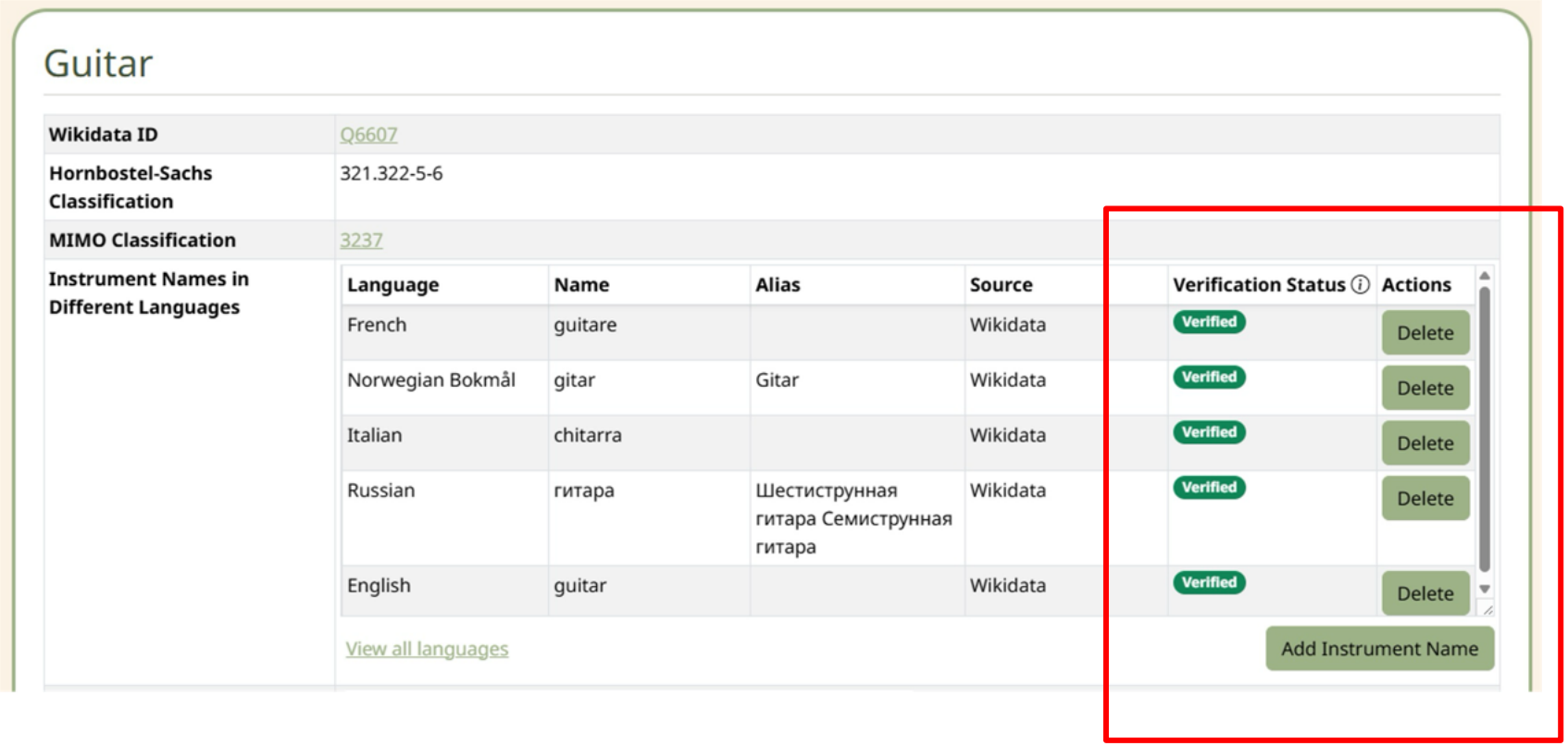
Language	Name	Alias	Source
French	guitare		Wikidata
Norwegian Bokmål	gitar	Gitar	Wikidata
Italian	chitarra		Wikidata
Russian	гитара	Шестиструнная гитара Семиструнная гитара	Wikidata
English	guitar		Wikidata

View all languages

Add Instrument Name

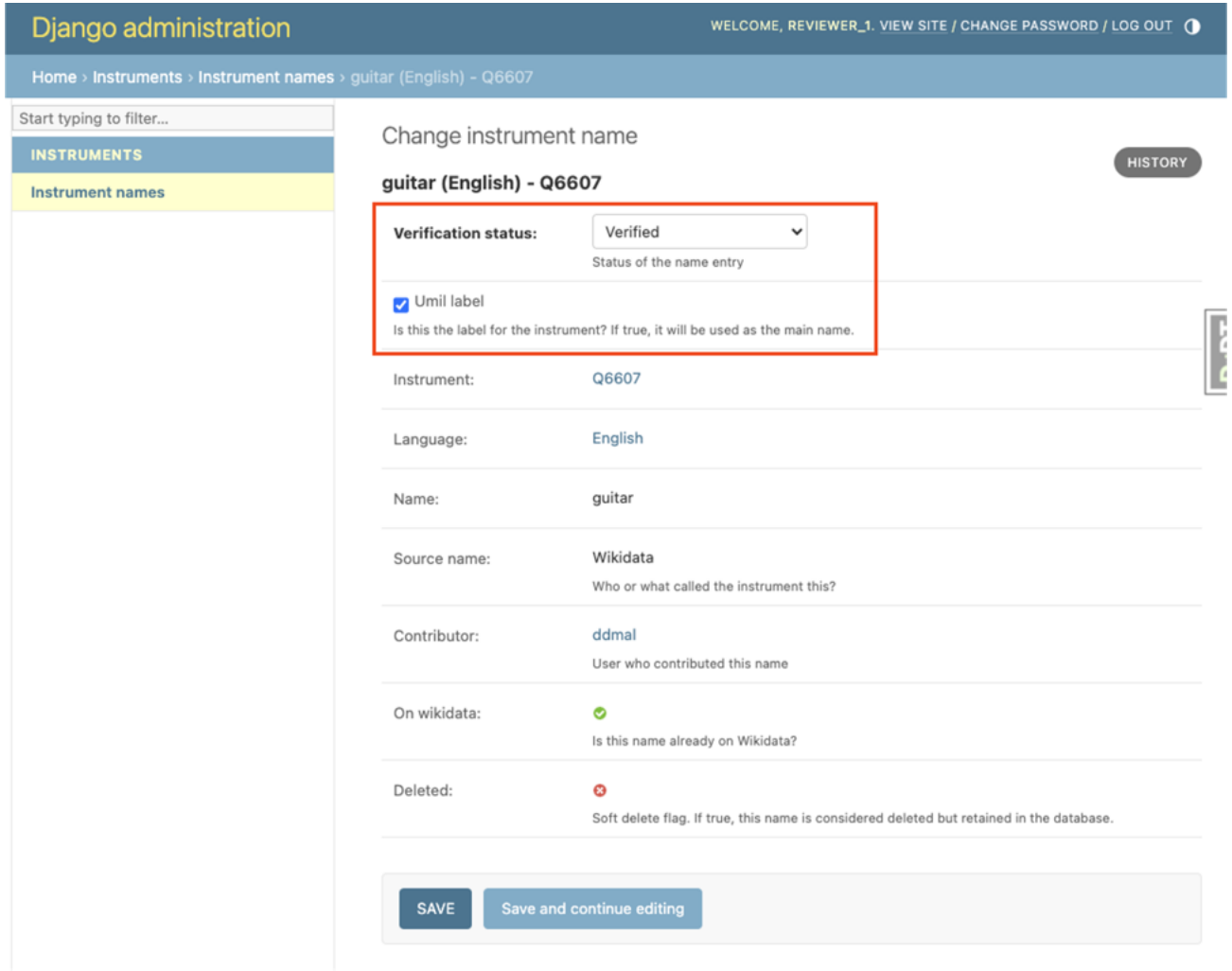
- Most basic
- View capability only
- Can only view public pages
- Can only view verified names

Contributor



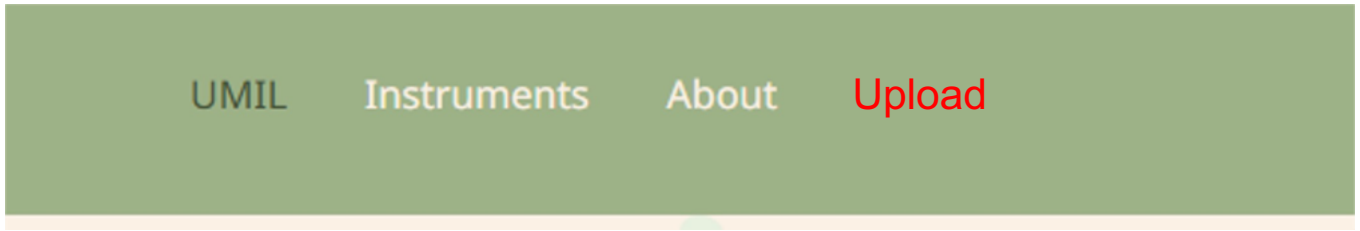
- Has UMIL account
- Can submit new instrument names
- Can submit new images
- Can submit new instruments
- Cannot edit/delete items outside of their own
- Can only view public pages
- Can view all names and their verification status

Reviewer



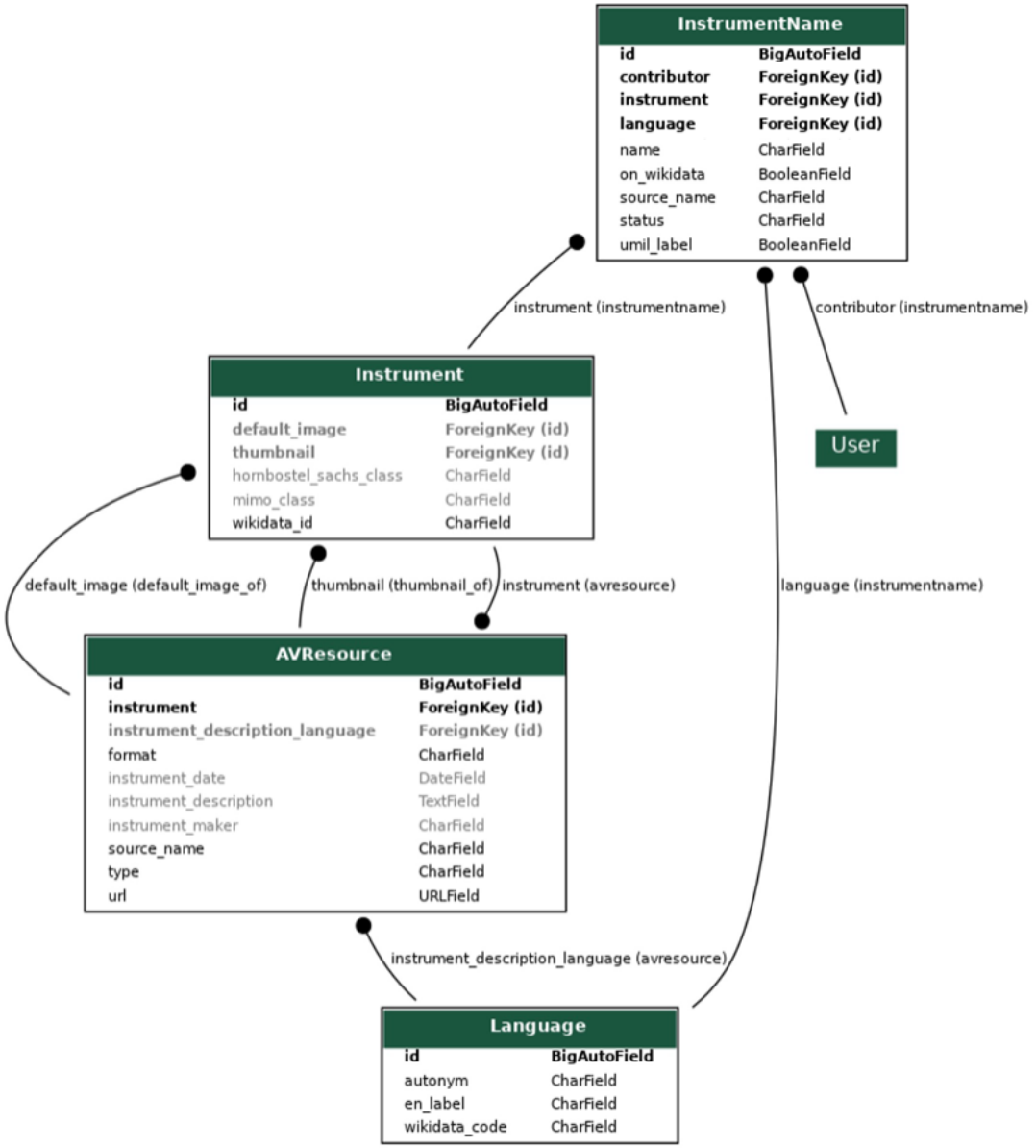
- Has UMIL account
- **Has Django access (staff)**
- Has Contributor powers
- Can edit the **verification_status** of an inject
- Can edit the **umil_label** of an inject
- Sets verification status of and investigates contributed items

Editor

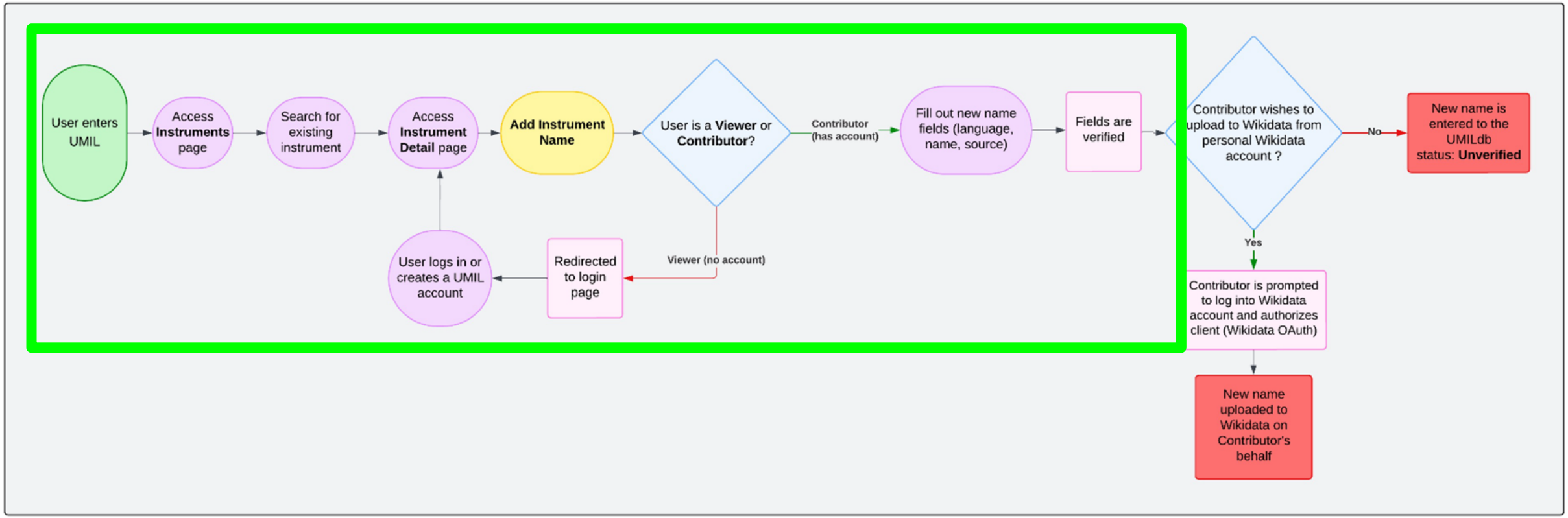


- Has a UMIL account
- **Has Django access (full)**
- Has all Reviewer powers
- Can delete any item
- Can submit new names to Wikidata

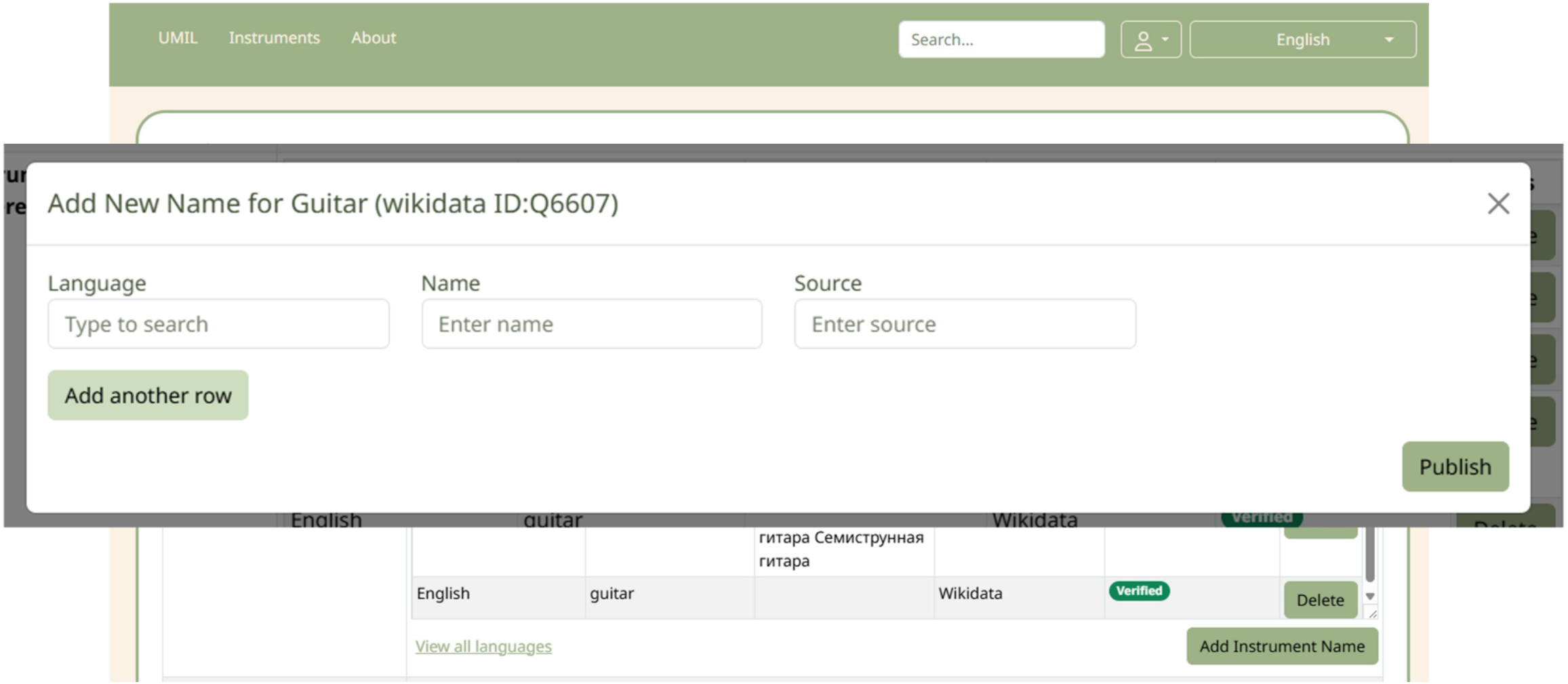
UMIL Database



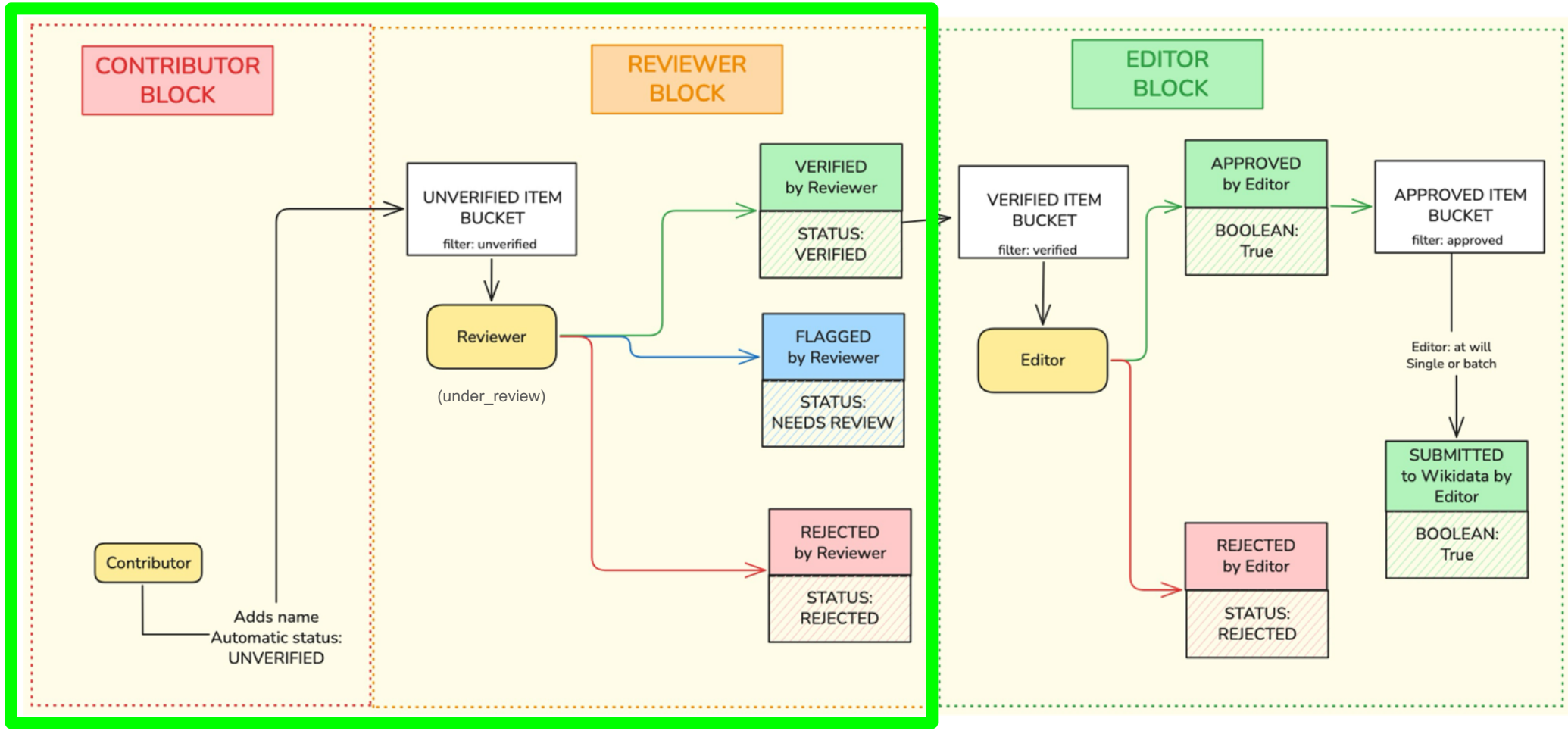
Adding New Name: Contributor Experience



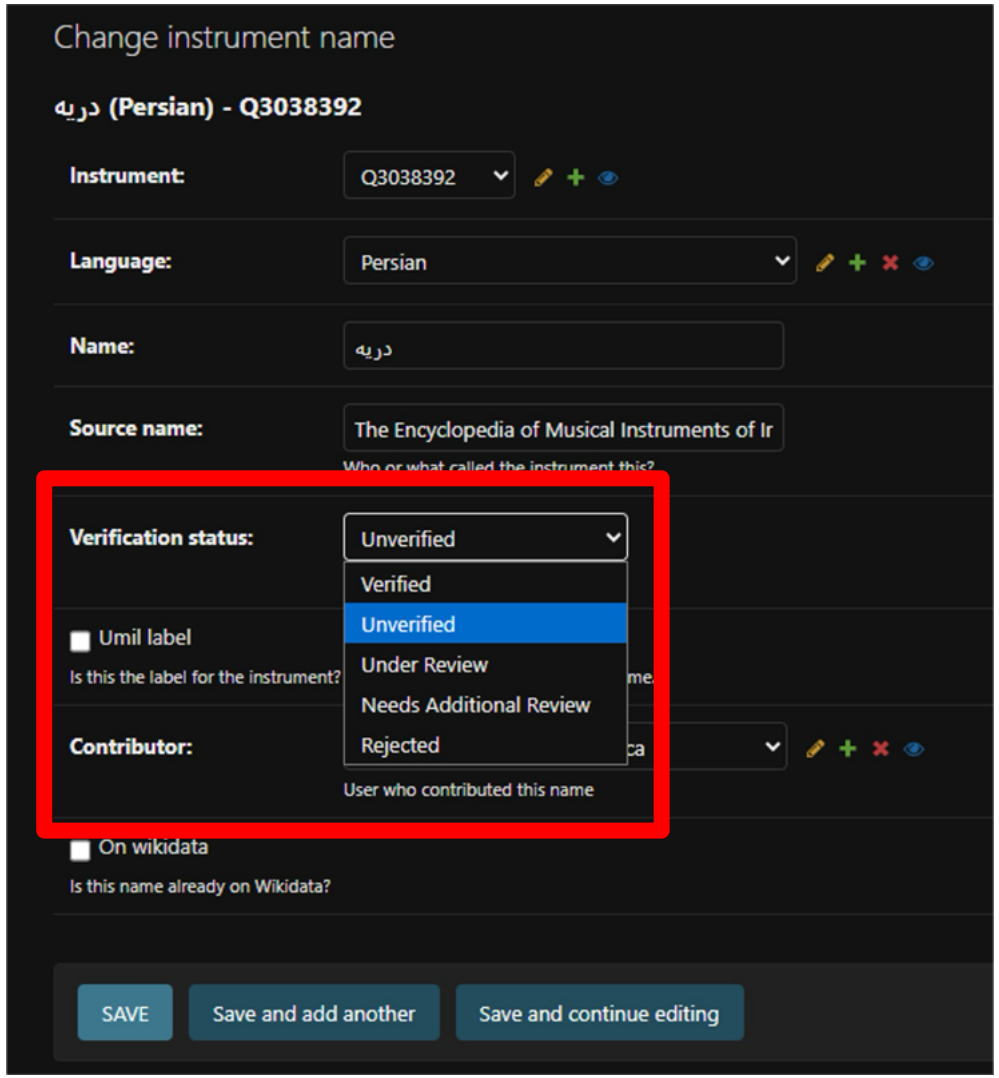
Adding New Name: Contributor Experience



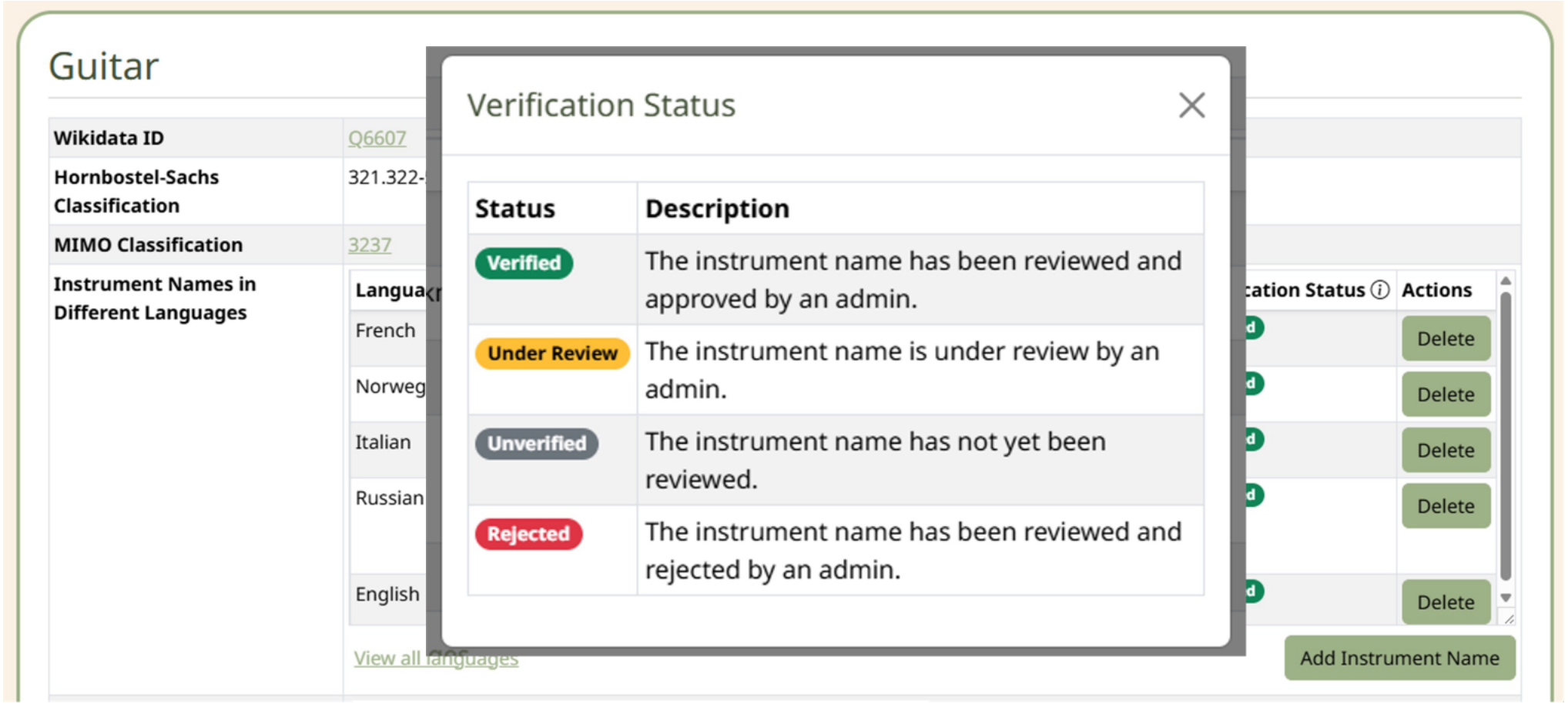
Backend Verification Workflow



Backend Verification Workflow



Frontend Verification Status Column



Part Three: Instrument Taxonomy and Classification Issues

Presenter: Kyrie Bouressa

Issues with Instrument Taxonomy and Classification in UMIL

LinkedMusic Project Meeting IV 2025
Kyrie Bouressa

It’s hard to figure out instrument labels versus aliases. These are our thoughts and current steps.

Violin

Fiddle

Hardanger fiddle

Double Bass

Contrabass

Instrument Taxonomy in UMIL

How do we determine what is a unique instance of an instrument, versus an alternative or additional name for an instrument?

What kind of open, broad guidance can we establish for ourselves and for our users?

Note: all examples in this presentation are pulled from lab discussion and issues discussed on GitHub, if you're curious about our larger discussions on these topics

We don’t want:

- Brands
  - Gibson-vs-Fender
  - YAMAHA PSR-EW425, etc
- Individuals’ instruments; discovered instruments
  - JS Bach’s harpsichord
  - Tutankhamun’s trumpets (Q2539852)
- Synthesizers, samplers (etc) variants
  - Roland SP-404; Geddy Lee Minimoog Model D
  - Pinned for later; beyond the scope of where UMIL is currently
    - Requires deep modification to suit Hornbostel-Sachs (HBS) and other hierarchies/classification:
      - HBS: Moog Satellite 523.1 = 511.11 + 512.121 + 512.13*1 + 514.11 1 + 514.122 + 514.21 *1
      - “The more information added, the more accurate the description will be. Even just mentioning the number of oscillators, noise generators, and filters of the Synthi 100 will drastically reduce the number of matching devices.” (Weisser & Quanten 2019) **Do we want that level of detail?**

We *do* want:

Relationships

- Using an instrument classification system like Hornbostel-Sachs (HBS) helps group and find instruments
  - Classified by what vibrates to produce sound (body, membrane, cord, air, electric signal)
- Tags
  - Material (wood, metal, plastic)
  - Method of playing (plucked, bowed, blown)
  - Event (wedding, funeral, birthday)
  - Genres or types (jazz, khyal)
  - Place (continent, country, group)
- “In close connection to” or “related to” field
  - Ex.: Connects an instrument deployed solely for weddings to its generic instance

Thumb piano

Nyonganyonga

Instrument	Alias	MIMO Classification (Musical Instrument Museum Online)	HBS	In close connection to ("different from"?)
Nyonganyonga	Nyonga Nyonga	128	122.1 + 112	thumb piano (mbira)

Things to consider

Hornbostel-Sachs (HBS) classifies instruments by their method of sound production (what is vibrating to produce sound), and their shape*

- Eurocentric
  - Does not account for **function**, **role**, **material**, or **genre/type** which, dependent on context, may indicate an entirely separate instrument concept in one culture than another; e.g., an instrument used only in weddings, versus a very similar if not nearly identical version of that instrument used elsewhere; or an instrument made of metal is wholly different from one made of wood

We want to avoid prescription, duplication, and exclusion.

Where's the line?

Guidance for us, guidance for users

- Range, register, morphology, method of playing, method of sound production.
  - ⅗ met to be considered a different instrument > alias?
- Association with specific functions in a genre or type
- Affiliated with specific events
  - Weddings, birthdays, funerals, religious events
- Flowcharts to help users make upper level HBS classifications

Alignment with Wikidata

- A Wikidata user may be making what appears to be a duplicate entry of a musical instrument because they feel the current entry does not reflect their context
    - UMIL is potentially uniquely positioned to help with this issue?
- Gen will talk shortly about additional Wikidata oddness

Does anyone have any ideas, questions, or comments?

Thank you for your time!

LinkedMusic DDMAL

Part Four: Challenges (1)

Presenter: Pouya Mohseni

RTL (Right-To-Left) Display

UMIL is multilingual, but several UI components were originally designed assuming LTR (left-to-right) scripts. When using RTL (right-to-left) languages, inconsistencies appear.

Language-specific content includes:

- Original entries (instrument name, aliases, sources)
- Translated components of the website

Current Issues

1. Interaction with Google translate
2. Displaying mixed-direction tables
3. Multiple scripts for one language
4. Direction for instrument sources

Figure 1. Instrument detail interface

Figure 3. Interface for adding new name

Figure 2. Different scripts for one language

Part Four: Challenges (2)

Presenter: Zih-Syuan Lin

Language Disambiguation

- Wikidata's language system is comprehensive but flat
- When adding names, the interface offers no guidance for choosing among variants
- Do we assume the contributor's choice is appropriate, or do we want clearer conventions to guide these decisions?

Han-script variants

Wikidata Code	Language Name
zh	Chinese
lzh	Literary Chinese
zh-hans	Simplified Chinese
zh-hant	Traditional Chinese
zh-classical	Classical Chinese Literary Chinese
zh-cn	Chinese (China)
zh-hk	Chinese (Hong Kong)
zh-mo	Chinese (Macau)
zh-my	Chinese (Malaysia)
yue	Cantonese
nan-hant	Minnan (Traditional Han script)
…	…

Romanization & dialect variants

Wikidata Code	Language Name
hak	Hakka Chinese
nan	Minnan
nan-latn-pehoeji	Minnan (Peh-oe-ji Romanization)
nan-latn-tailo	Minnan (Tai-lo Romanization)
zh-min-nan	Minnan (Chinese Min Nan)
…	…

An example for Minnan
Han script: 因為當初是
Pe̍h-ōe-jī: In-uī tong-tshoo sī
Tai-lo Romanization: In-uī tong-tshoo sī

Part Four: Challenges (3)

Presenter: Yu-Chia Kuo

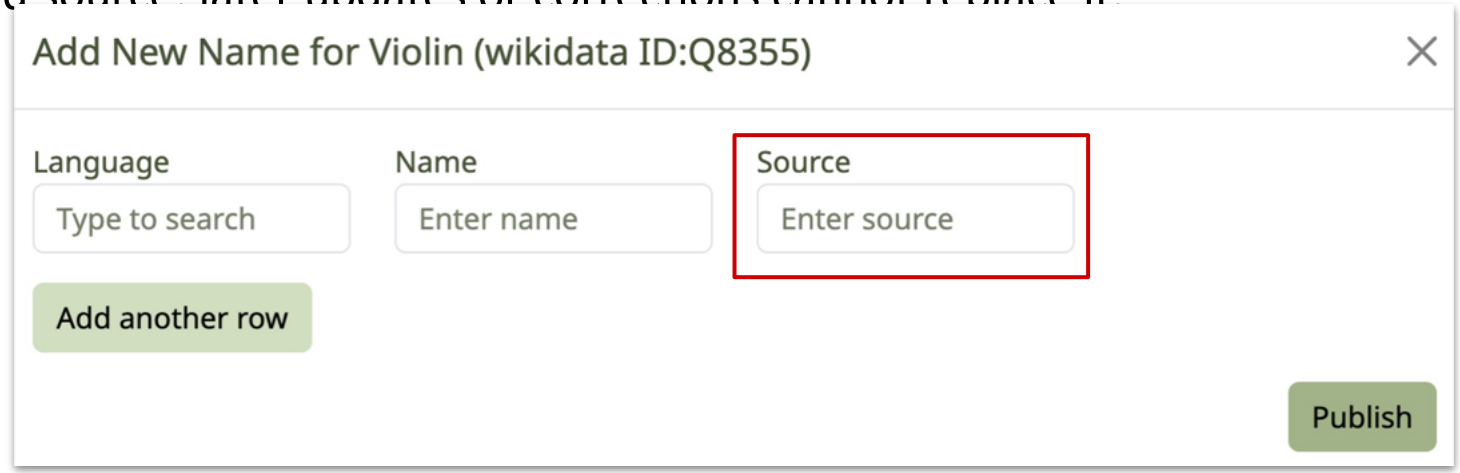
Difficulties When Adding New Instrument Names

1. Repeated manual entry for translations with the same source

Languages that share the same reference (e.g., Chinese varieties) still require separate submissions, resulting in unnecessary repetitive work.

2. Source update limitations

The system retains only the earliest submitted source; later updates or corrections cannot replace it, leading to the loss of more reliable sources.



Future Work

1. Multi-language entry support

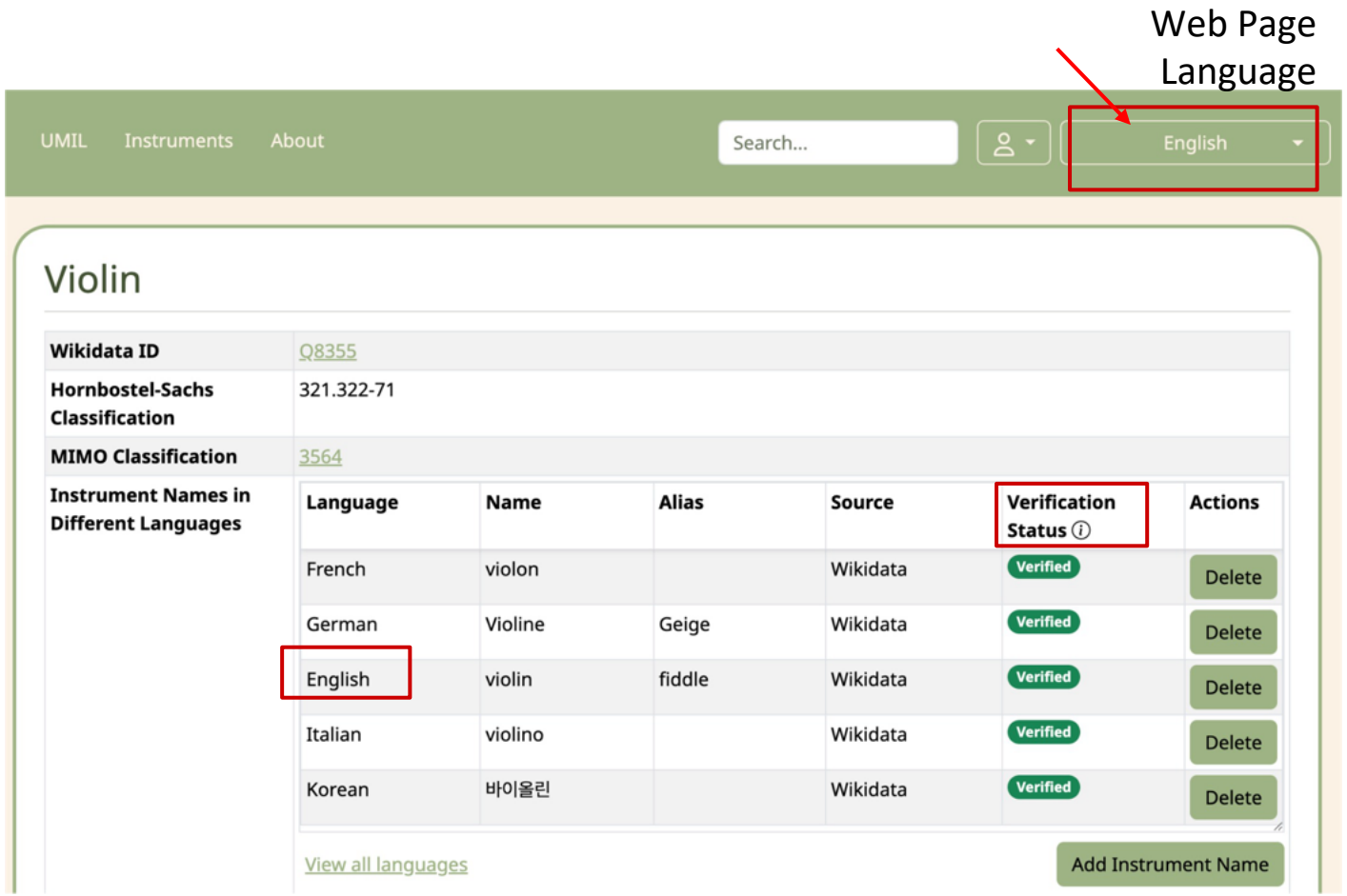
- Allow adding several related languages in a single submission when they share the same source.

2. Better visibility and verification

- Highlight or pin the active language row for easier confirmation.
- Extend verification tags to aliases for clearer provenance tracking.

3. Improve source preservation

- Enable updating multiple sources for the same name.



Part Four: Challenges (4)

Presenter: Mai Lyn Puittinen

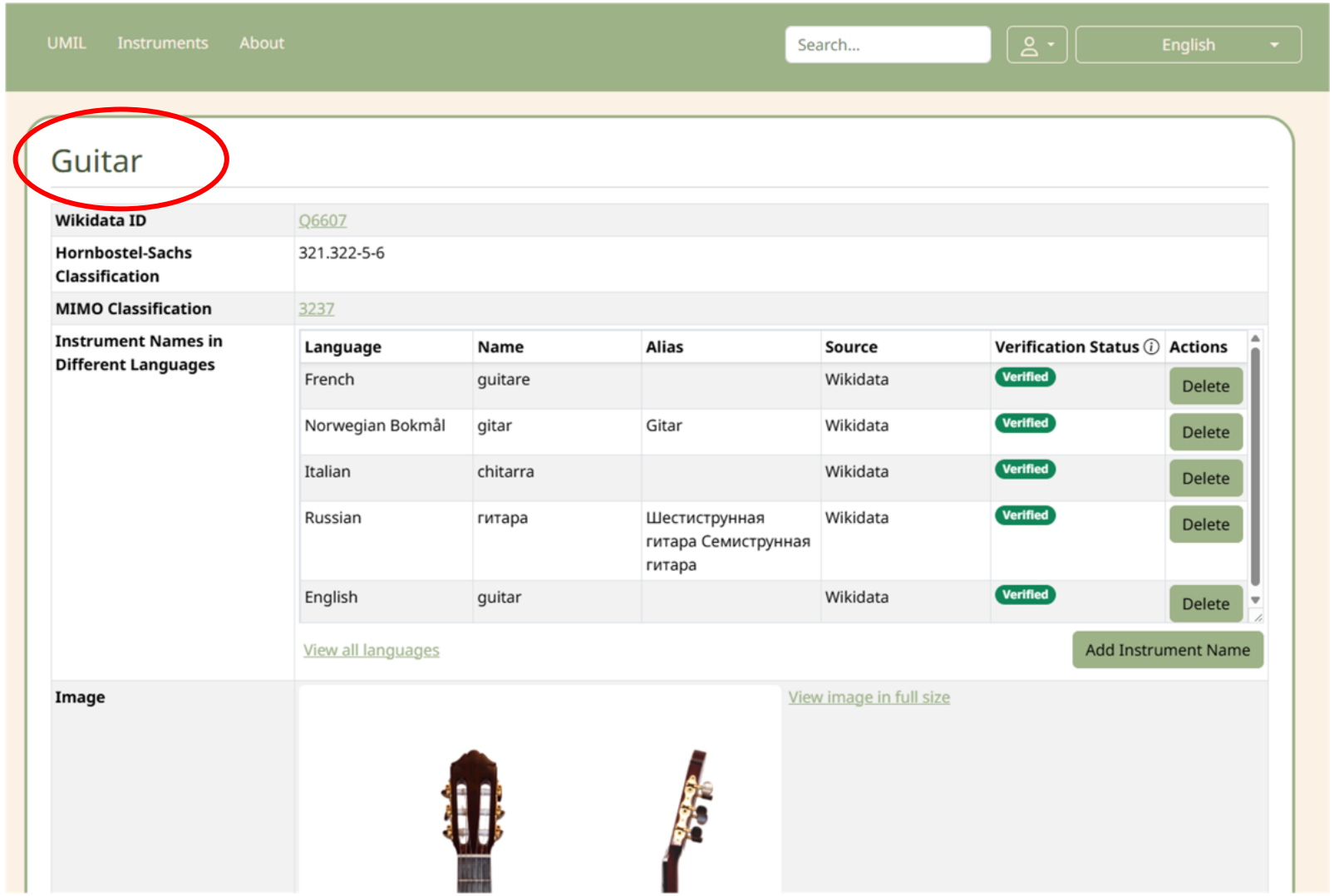
Deleting Names

- Entries should be preserved in the database even if they are deleted from view on UMIL.
- **Solution:** create Deleted boolean flag attribute for each Instrument Name.
- A Deleted Instrument Name will remain in the UMIL database but is no longer displayed on the UMIL frontend (regardless of status).
- Deleted Instrument Names will not be considered for Wikidata Upload.

Deleting Names: Considering UMIL labels

- On Wikidata, there may be more than one instrument name in a language.
  - The most common name/first name added is considered the instrument's LABEL.
  - Additional names are considered as ALIAS.
- UMIL labels are displayed as the title of the Instrument Detail page.
- **Problem:** if a contributor chooses to delete their name that is a umil_label, there must be some sort of automation that reassigns umil_label to the earliest alias (umil_label = false) in that language.
- If there is no other verified names available, a placeholder label must be displayed on the detail page.
- Complex problem to consider changes freely done on the backend.

Deleting Names: Considering UMIL labels



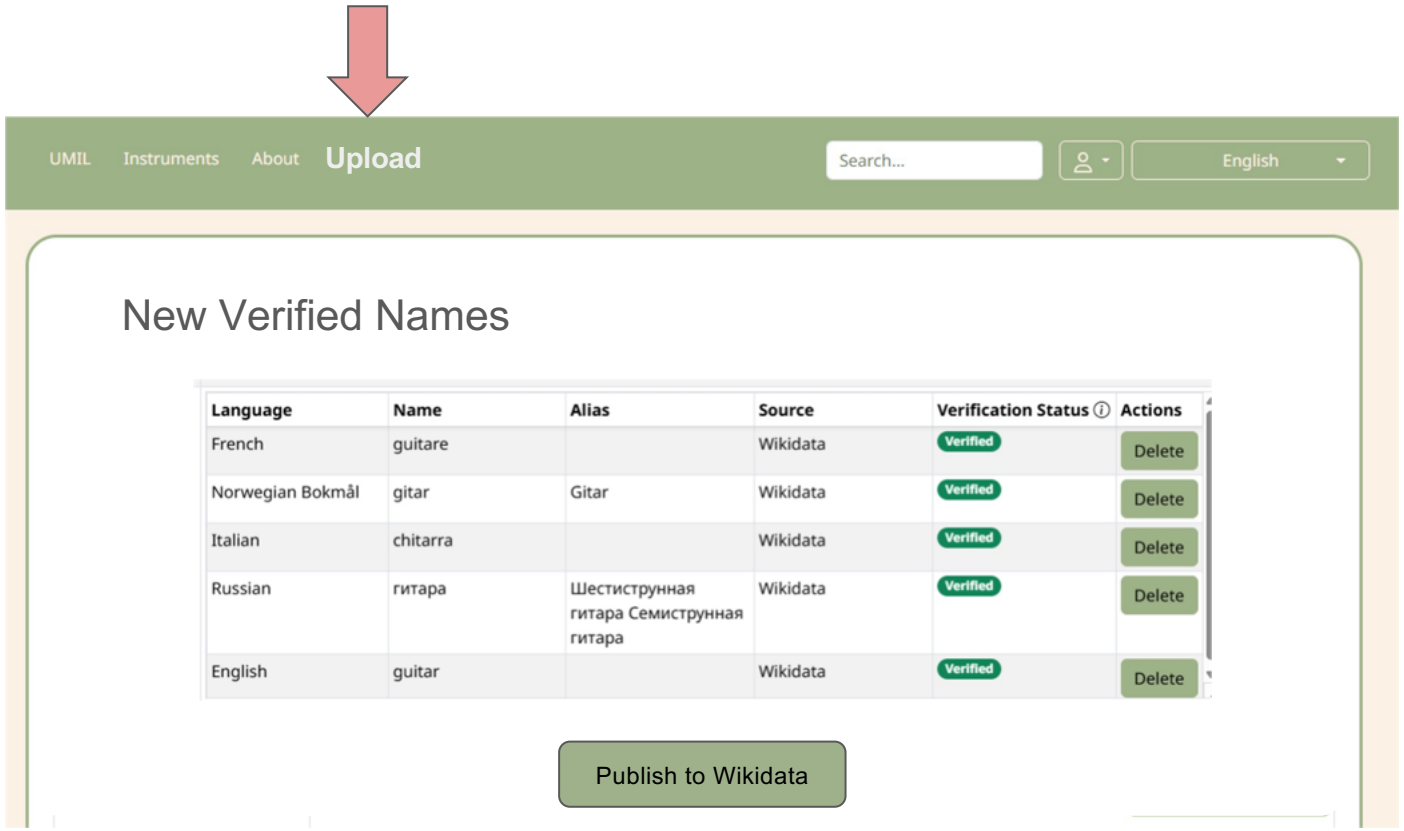
Rejected: Inappropriate

- Although we want to let contributors know what names have been rejected to prevent repeated attempts, it is not ideal to continue displaying “names” that contain inappropriate words.
- **Solution:** create another verification status of rejected: inappropriate or some adjacent naming, and only display these names to the contributor that added them.
    - Alternatively, never display these names and routinely delete them from the database

Wikidata Interaction: Upload Page

The workflow for uploading verified NEW names to wikidata has been outlined, however is not implemented.

- Only accessible to Editors
- Displays all Verified names NOT on_wikidata
- Allows Editor to edit verification status or delete entry
- Publish to Wikidata prompts for Editor to log into their personal Wikidata account



What is a source?

- When a Contributor adds a name, they must include the source from where the name comes from.
- Currently, **source** is a text box where the contributor can type whatever they want.
- **Consider:** specifying source categories so that UMIL data maintainers can better understand where our names are coming from and whether these sources are valid.
- **Problem:** as we aim to collect names that are not particularly well documented, how can we categorize specific sources that are from family (“mom”, “grandpa”)?
- Should we have categories and then allow a contributor to further specify?

Part Four: Challenges (5)

Presenter: Geneviève Gates-Panneton

Instrument entries on Wikidata that we do not want

Two relevant issues:

When transferring entries from Wikidata to UMIL - Avoid transferring entries we don’t want

When inputting new instruments in UMIL - Avoid repeating these same mistakes

Duplicate entries

Two entries for the same instrument

ngombi (Q3339324)
central African arched harp
7 statements, 7 sitelinks - 20:41, 25 October 2025

Edmond **Ngombi** (Q28718176)
7 statements, 1 sitelink - 12:19, 31 March 2024

Ngombi (arched Harp) (Q116413065)
ngombi (arched Harp) at the Metropolitan Museum of Art (MET, 89.4.3527)
7 statements, 0 sitelinks - 01:55, 23 July 2024

Lius Alombah **Ngombi** (Q131490093)
lawyer in Cameroon
4 statements, 0 sitelinks - 23:06, 7 February 2025

Ngômbi (Q105072195)
Musical Instrument
4 statements, 0 sitelinks - 16:21, 25 January 2021

ngombi (Q3339324)

Item Discussion

central African arched harp edit

In more languages
Configure

Language	Label	Description	Also known as
default for all languages	No label defined	–	
English	ngombi	central African arched harp	
Māori	No label defined	No description defined	
Catalan	ngombi	instrument de corda	
Persian	نگومبی	چنگ قوس‌دار آفریقایی	
French	ngombi	harpe arquée d'Afrique centrale	
Hebrew	No label defined	כלי פריטה	
Hungarian	ngombi	közép-afrikai ívhárfa	
Norwegian Nynorsk	Ngombi	No description defined	
Russian	Нгомби	Музыкальный инструмент	

Fewer languages

Ngômbi (Q105072195)

Item Discussion

Musical Instrument edit

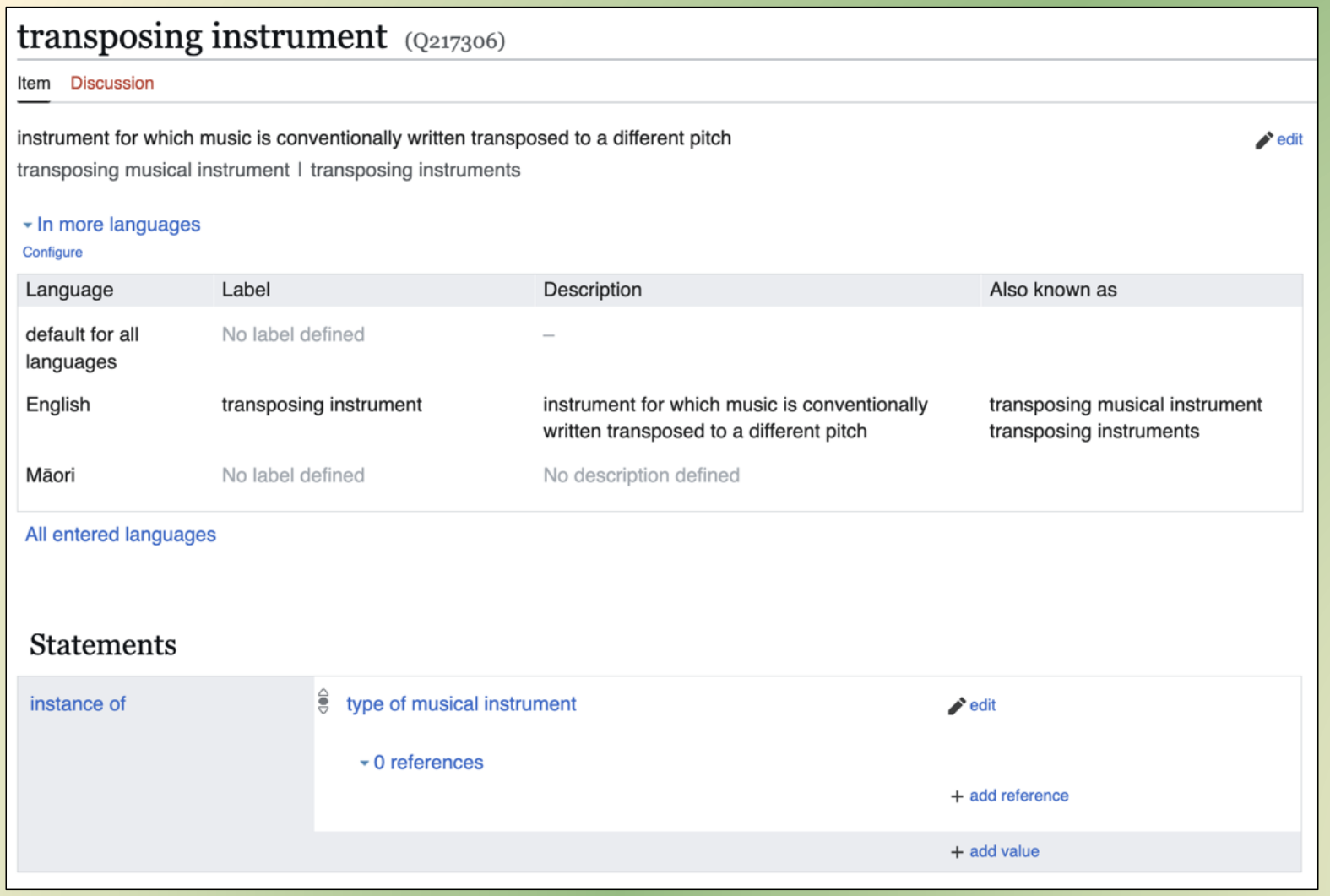
In more languages
Configure

Language	Label	Description	Also known as
default for all languages	No label defined	–	
English	Ngômbi	Musical Instrument	
Māori	No label defined	No description defined	
German	Ngômbi	Musikinstrument	Harmonica, Accordéon Duala, Littoral Cameroun
French	Ngômbi	Instrument de musique	

Fewer languages

Overly vague entries

Entries that are more of an instrument description or characteristic than an instrument itself



transposing instrument (Q217306)

Item Discussion

instrument for which music is conventionally written transposed to a different pitch

edit

transposing musical instrument | transposing instruments

In more languages

Configure

Language	Label	Description	Also known as
default for all languages	No label defined	–	
English	transposing instrument	instrument for which music is conventionally written transposed to a different pitch	transposing musical instrument transposing instruments
Māori	No label defined	No description defined	

All entered languages

Statements

instance of — type of musical instrument — edit

0 references

+ add reference

+ add value

Dubious electrophones

Electrophones that aren't actually played, or that are a component of a larger instrument

sound module (Q715223)

Item Discussion

electronic musical instrument without a human-playable interface, operated using an externally connected device, e.g. a MIDI controller edit

▾ In more languages

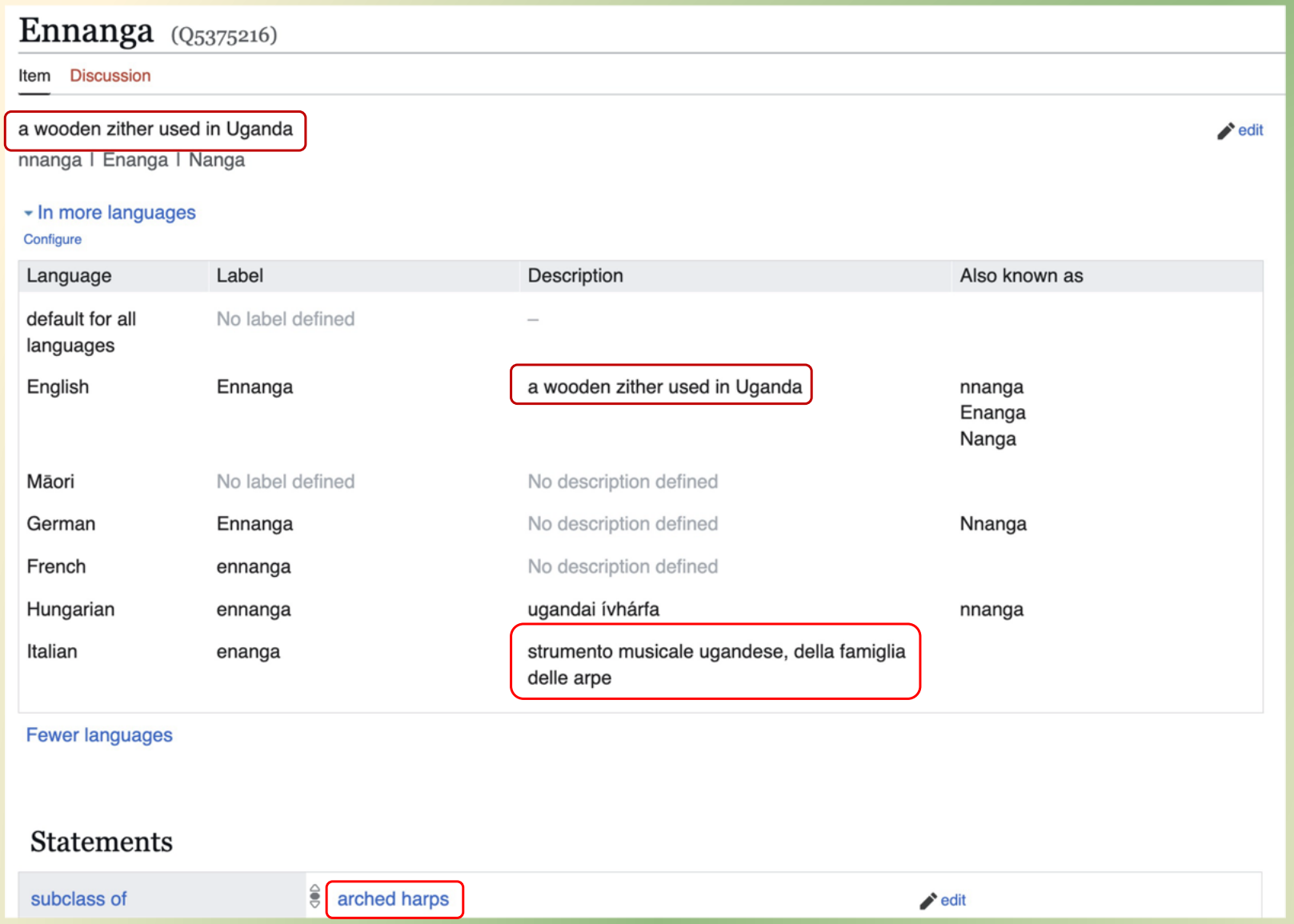
Configure

Language	Label	Description	Also known as
default for all languages	No label defined	–	
English	sound module	electronic musical instrument without a human-playable interface, operated using an externally connected device, e.g. a MIDI controller	
Māori	No label defined	No description defined	

All entered languages

Factual errors

Entries that contain a mistake in the instrument description or aliases



Ennanga (Q5375216)

Item Discussion

a wooden zither used in Uganda

edit

nnanga | Enanga | Nanga

In more languages

Configure

Language	Label	Description	Also known as
default for all languages	No label defined	–	
English	Ennanga	a wooden zither used in Uganda	nnanga Enanga Nanga
Māori	No label defined	No description defined	
German	Ennanga	No description defined	Nnanga
French	ennanga	No description defined	
Hungarian	ennanga	ugandai ívhárfa	nnanga
Italian	enanga	strumento musicale ugandese, della famiglia delle arpe	

Fewer languages

Statements

subclass of arched harps edit

Thank you!