

LinkedMusic

LinkedMusic Project

Integrating Online Music Databases



Ichiro Fujinaga & Junjun Cao

Music Technology Area, Schulich School of Music

McGill University

LinkedMusic Project: Key Concepts

- ❖ Be able to search across various online music databases from one website
- ❖ Universal Musical Instrument Lexicon (UMIL)
 - ❖ Create a crowd-sourced website
 - ❖ Images and recordings of musical instruments
 - ❖ Name of the instrument in the local language
- ❖ Funded for 7 years (2022–2029): \$3.2M CAD

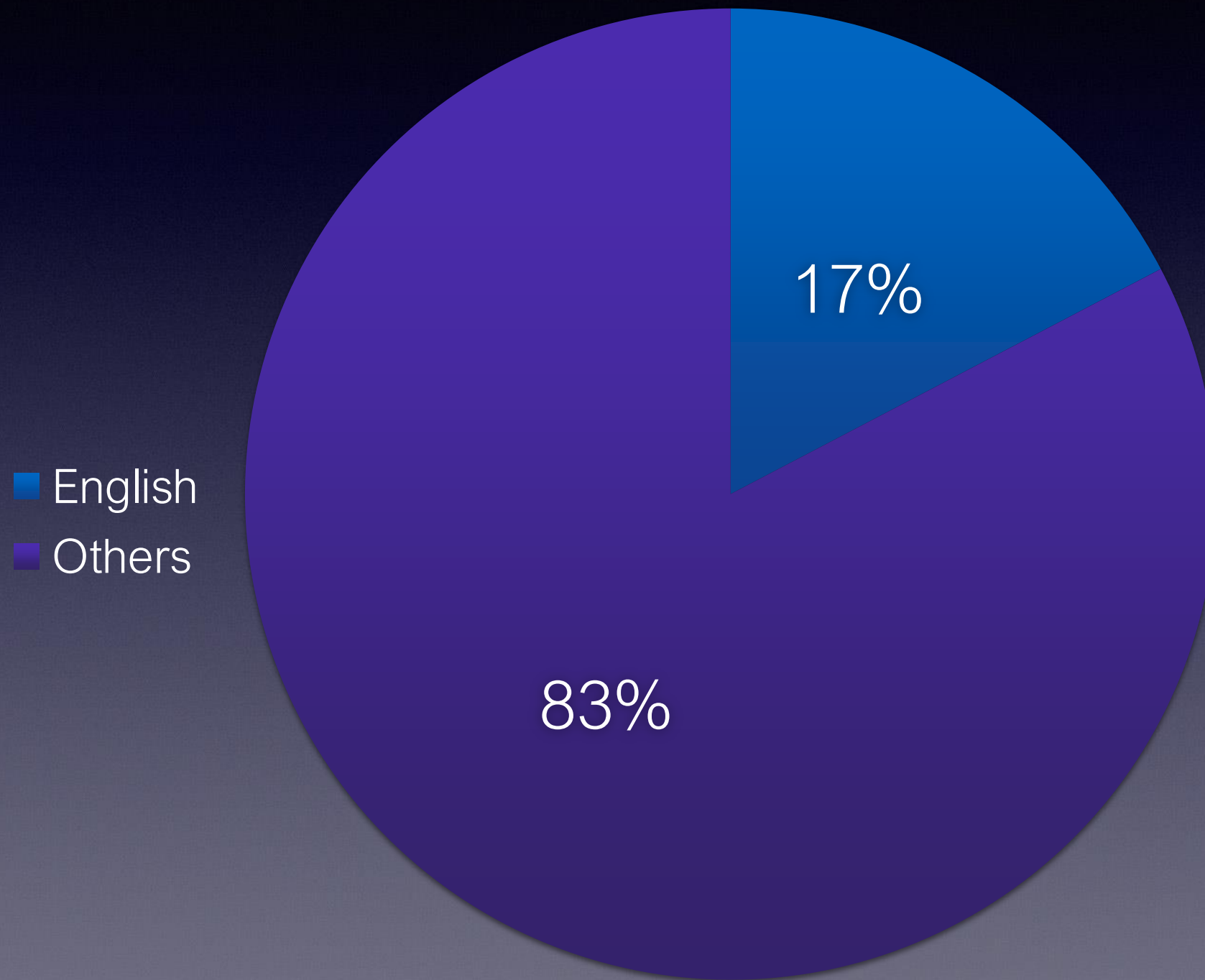
LinkedMusic: Participants

- ❖ 7 Co-investigators: McGill, U of Montreal, Dalhousie, etc.
- ❖ 19 Collaborators: Indiana University, Stanford, Utrecht, Yale, etc.
- ❖ 9 Partners: British Library, MetaBrainz, Oxford, RISM, etc.
- ❖ 18 Graduate students & researchers

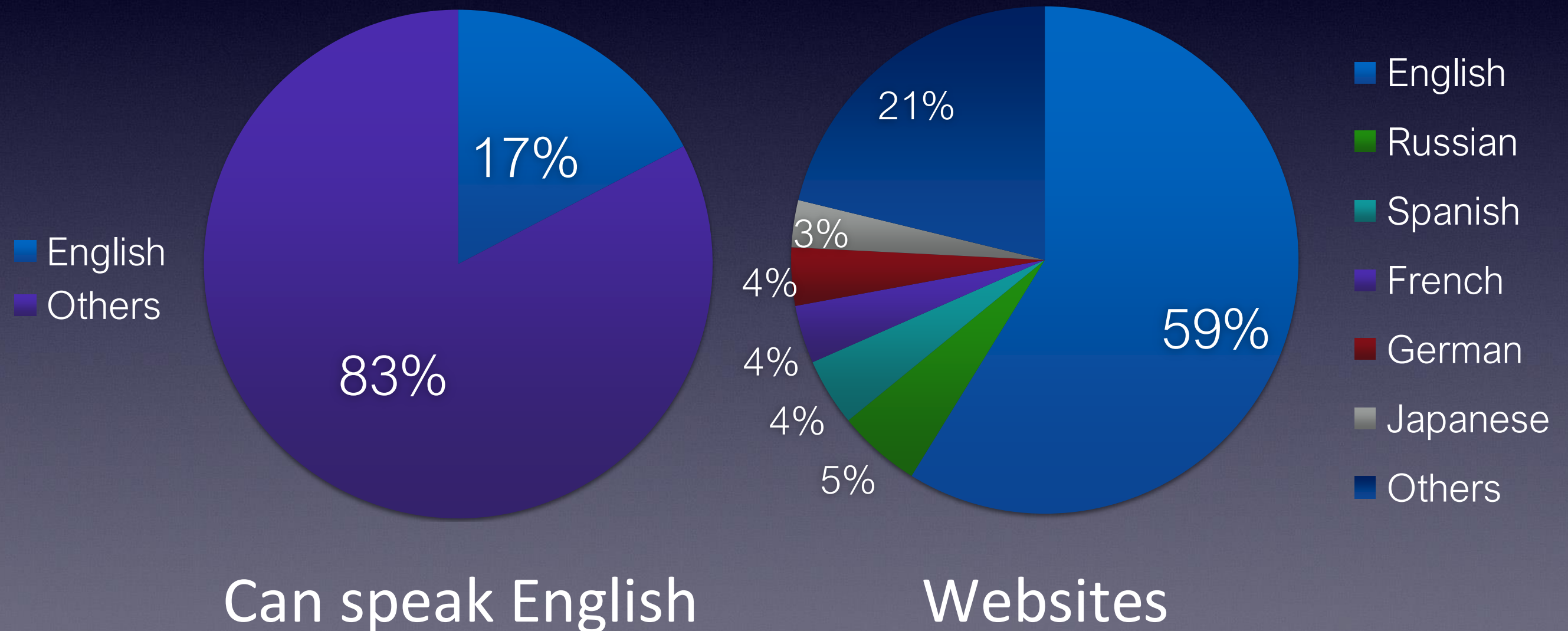
Goals of LinkedMusic Project

- ❖ 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

People who can speak English (www.babbel.com)



People who can speak English compared to websites



LinkedMusic.ca

LinkedMusic

UM

TEILNEHMER

AKTIVITÄTEN ▼

VERWEISE

UMIL

GELEGENHEITEN

KONTAKTIEREN SIE UNS

German ▼

LinkedMusic

Das Ziel unserer neuen LinkedMusic-Partnerschaft ist die Verknüpfung von Musikdatenbanken durch Metadatenschemata: Strukturen zur Organisation von in einer Datenbank gespeicherten Informationen. Dies trägt maßgeblich dazu bei, die Online-Musiksuche auf das Niveau textbasierter Ressourcen zu heben und grundlegende Fragen zur Musik und ihrer Wechselwirkung mit menschlicher Kreativität, Gesellschaft, Kultur und Geschichte zu beantworten.

Zu unseren Zielen gehören:

- Entwicklung eines umfassenden internationalen Metadatenschemas für Musik
- Entwicklung von Mapping-Tools zur Abbildung bestehender Metadatenschemata auf unser neues Schema
- Indexieren Sie aggregierte Metadaten, um die gleichzeitige Suche in einzelnen Datenbanken zu ermöglichen

Für weitere Informationen wenden Sie sich bitte an [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.

UMIL (Universal Musical Instrument Lexicon)

[UMIL](#)
[Home](#)
[Instruments](#)
[About](#)
[Login](#)
[Select Language](#)

Hornbostel-Sachs Classification

- 1 - Idiophones (131)
- 2 - Membranophones (62)
- 3 - Chordophones (275)
- 4 - Aerophones (278)
- 5 - Electrophones (29)
- Unclassified

More facet search coming...

INSTRUMENT LIST

[English](#)


Showing 1 to 20 of 859 entries

View 20 | 50 | 100 per page



Guitar



Trumpet



Clarinet



Bagpipes



Trombone



Why UMIL (Universal Music Instrument Lexicon)?

- ❖ 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
 - ❖ Create a pictorial interface to populate musical instrument names in Wikidata

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. The Session

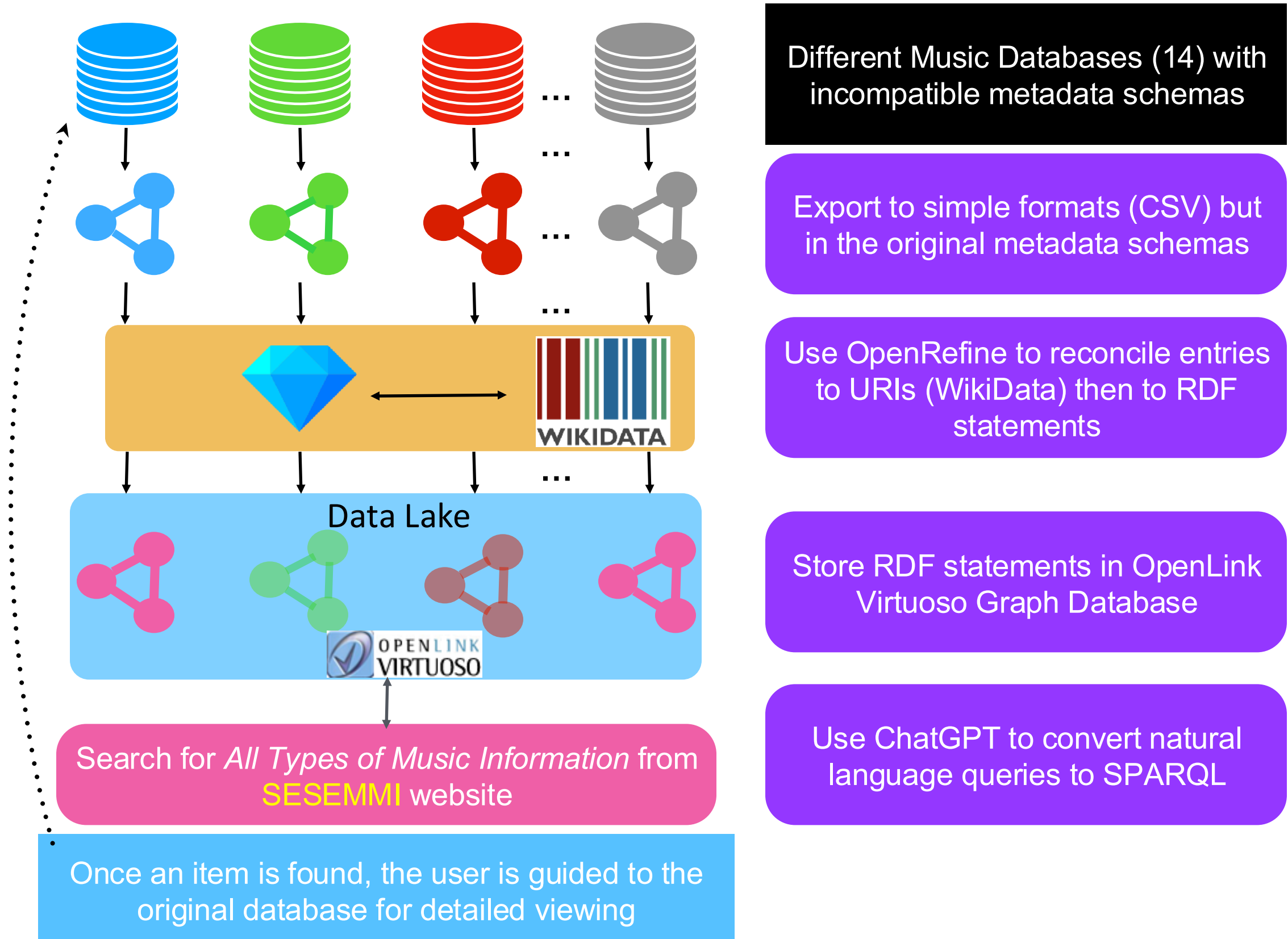
Challenges in Integrating Databases

- ❖ 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
 - ❖ “Wolfgang Amadeus Mozart” vs “Mozart, W. A.”

Our Overall Process

1. Convert various databases into flat files (CSV 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 Project: Overall Process



Example: The Session Database

Database of Traditional Irish Music sessions

TUNES

DISCOGRAPHY

TRIPS

SESSIONS

EVENTS

DISCUSSIONS

THE SESSION

Log in or Sign up

Fáilte

Search

Tunes

for

SEARCH

The Session is a community website dedicated to Irish traditional music.

You can **find tunes** to play, **find sessions** to play them in, and **join in discussions** about the music. You can also **find events** (like concerts and festivals), or **explore the track listings of recordings**.

You can contribute too. If you're already a member, you can **log in**. If you're not yet a member, membership is free and it only takes a moment to **sign up**.

You can also **install The Session app** on your phone or computer.

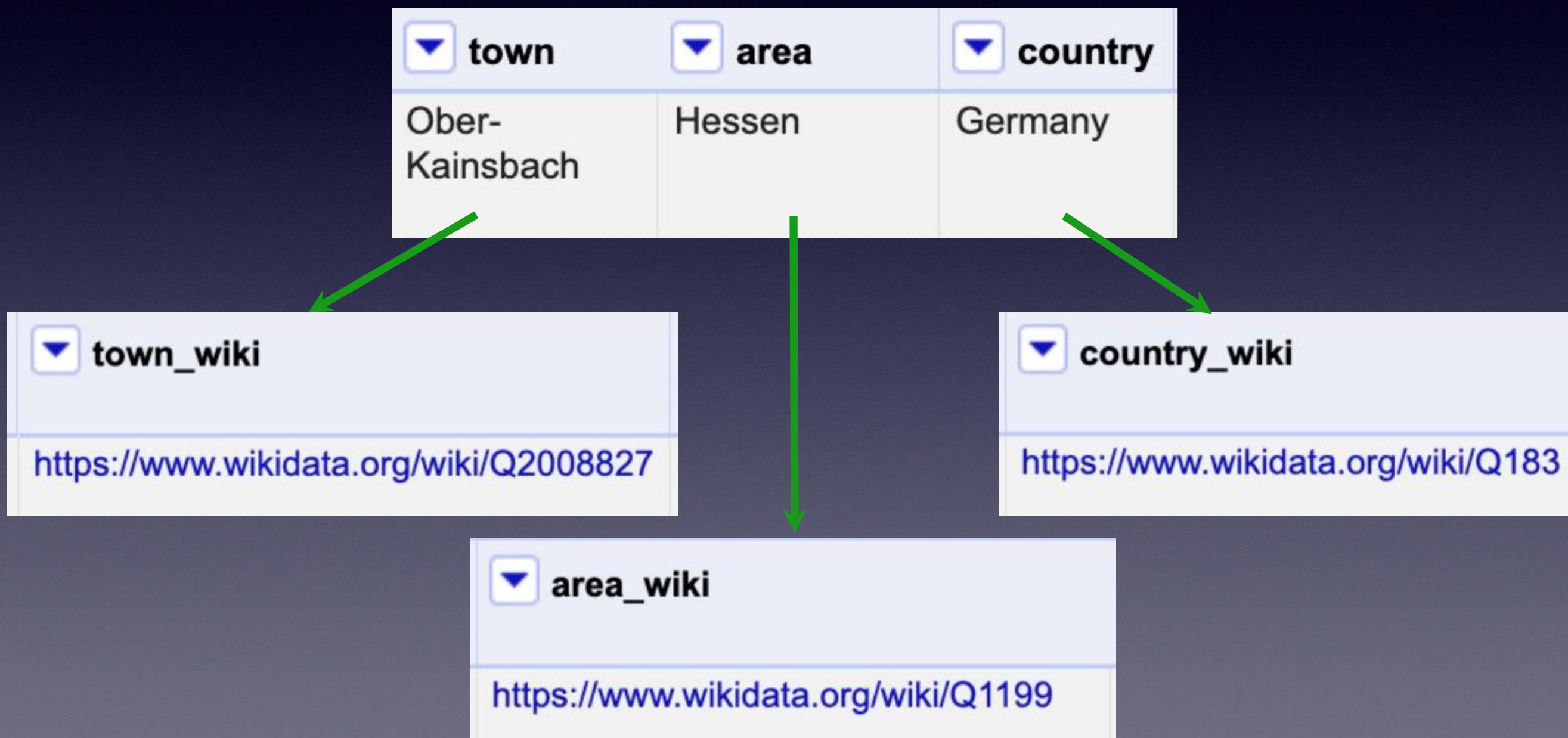
1. Export the database to text files

Portion of a CSV file of events in Session's 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		Portarlinton	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



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



Other open-source examples include:



5. Query using natural languages (1)

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

SPARQL Query Editor
About
Tables ▼
Conductor
Facet Browser
Permalink

Extensions:
cxml
save to dav
sponge
User: SPARQL

Default Data Set Name (Graph IRI)

Query Text

|

Results Format
HTML
▼

Execute Query
Reset

5. Query using natural languages (2)

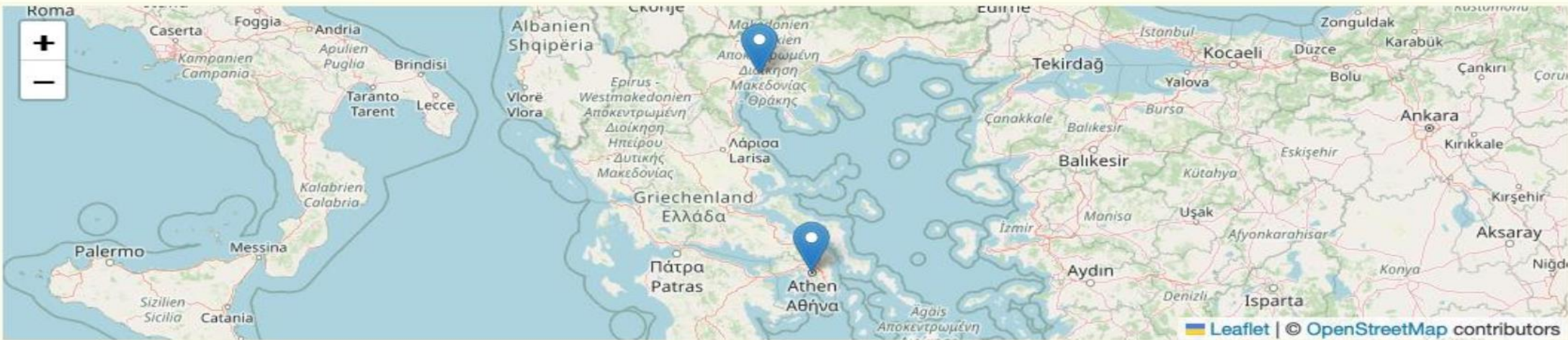
Example query: **“Find sessions that took place in Greece”**

TUNES
DISCOGRAPHY
TRIPS
SESSIONS
EVENTS
DISCUSSIONS

Sessions

Search for sessions on in

Searching for sessions in “Greece”



- The Lucky Sparrow Irish Pub**, Athens, Athens, Greece
Added by Athens Irish Festival 9 months ago.
- The Dubliner**, Thessaloniki, Makedonia, Greece
Added by lukegarry 3 years ago. Updated 2 months ago.

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 rescue!

I have an RDF database reconciled with Wikidata.

It contains
country it was

Create a c
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 .
}
```

sions

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

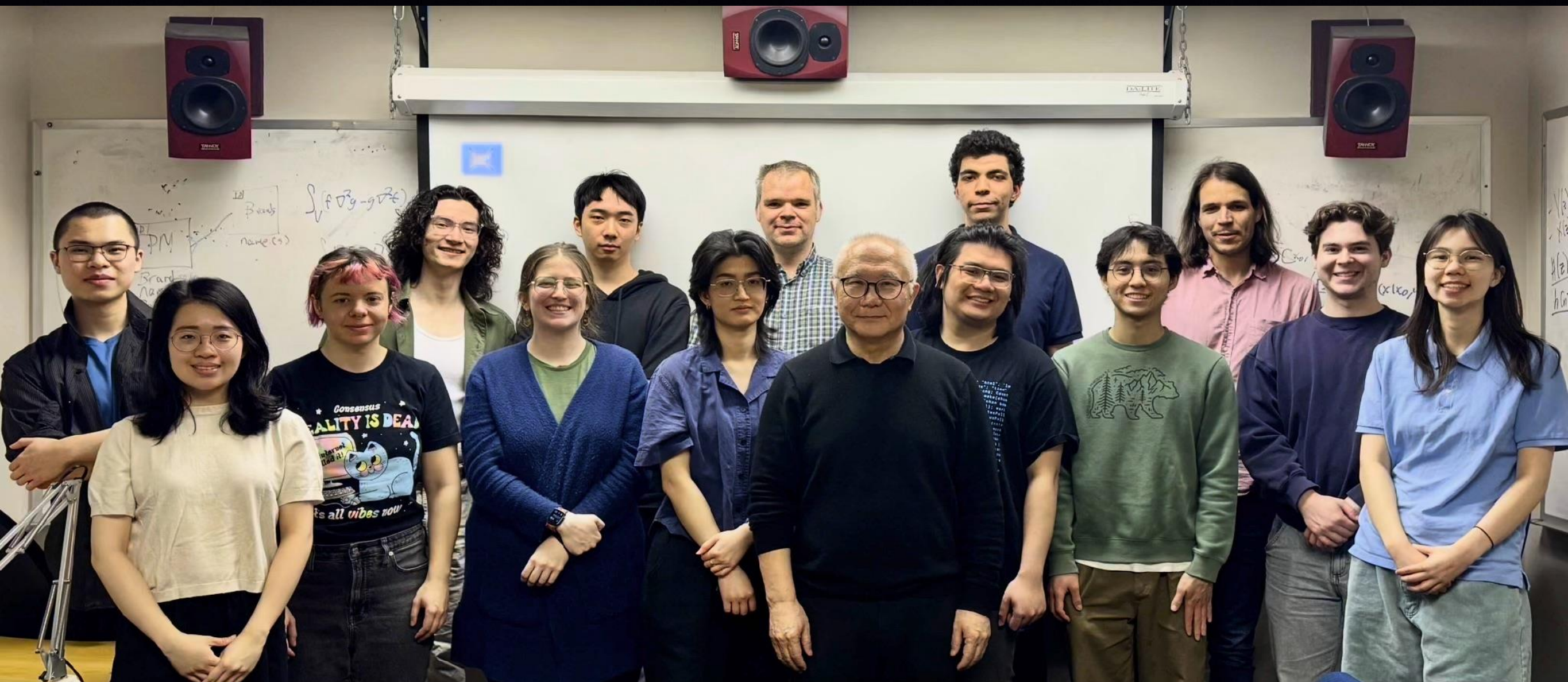
Features of LinkedMusic

- ❖ By converting a database to an RDF graph database, we can search the database with natural language queries
- ❖ Because we use ChatGPT, we can make queries in many different natural languages: “找到係希臘嘅演奏會”
- ❖ 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

Future Directions

- ❖ Improve our prompts so that ChatGPT makes fewer errors
 - ❖ Prompt Engineering
 - ❖ In-context Learning
- ❖ 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
Sebastian 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



Schulich School of Music
École de musique Schulich

DDMAL DISTRIBUTED DIGITAL MUSIC
ARCHIVES & LIBRARIES LAB



Centre for Interdisciplinary Research
in Music Media and Technology

Fonds de recherche
Société et culture

Québec

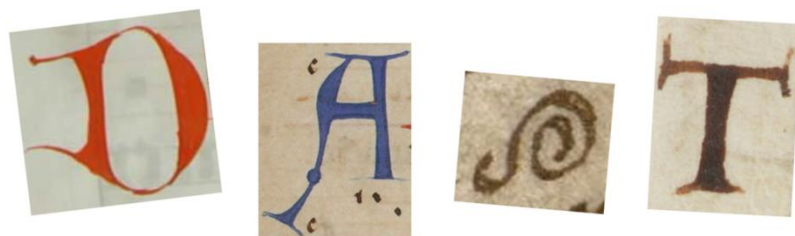


Digital Research
Alliance of Canada

Alliance de recherche
numérique du Canada



Calcul Québec



DIGITAL ANALYSIS OF CHANT TRANSMISSION



DALHOUSIE
UNIVERSITY

