

LinkedMusic

Project Meeting III

26 October 2024

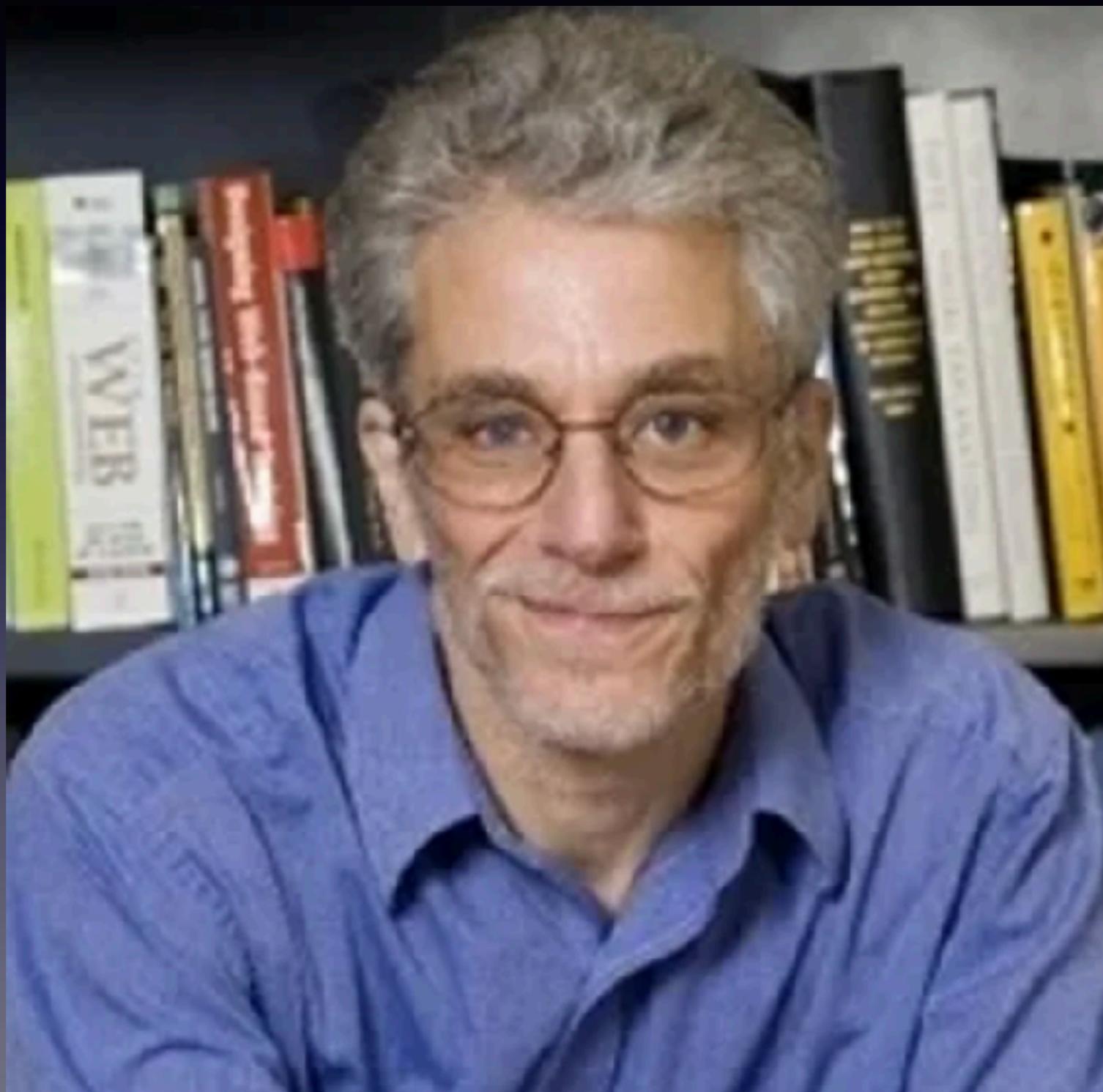


Ichiro Fujinaga

Music Technology Area, Schulich School of Music
McGill University



In Memory of Donald “Don” Byrd



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

- ❖ Barbara Dobbs McKenzie (retired) (RILM: Répertoire International de Littérature Musicale, Tina Frühauf)
- ❖ 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

- ❖ Lucas March
- ❖ Cole Thierrin
- ❖ Liam Pond
- ❖ Yu Chia Kuo
- ❖ Undergraduate Student
 - ❖ Yueqiao Zhang
- ❖ Researchers
 - ❖ Dylan Hillerbrand
 - ❖ Geneviève Gates-Panneton
 - ❖ Yinan Zhou

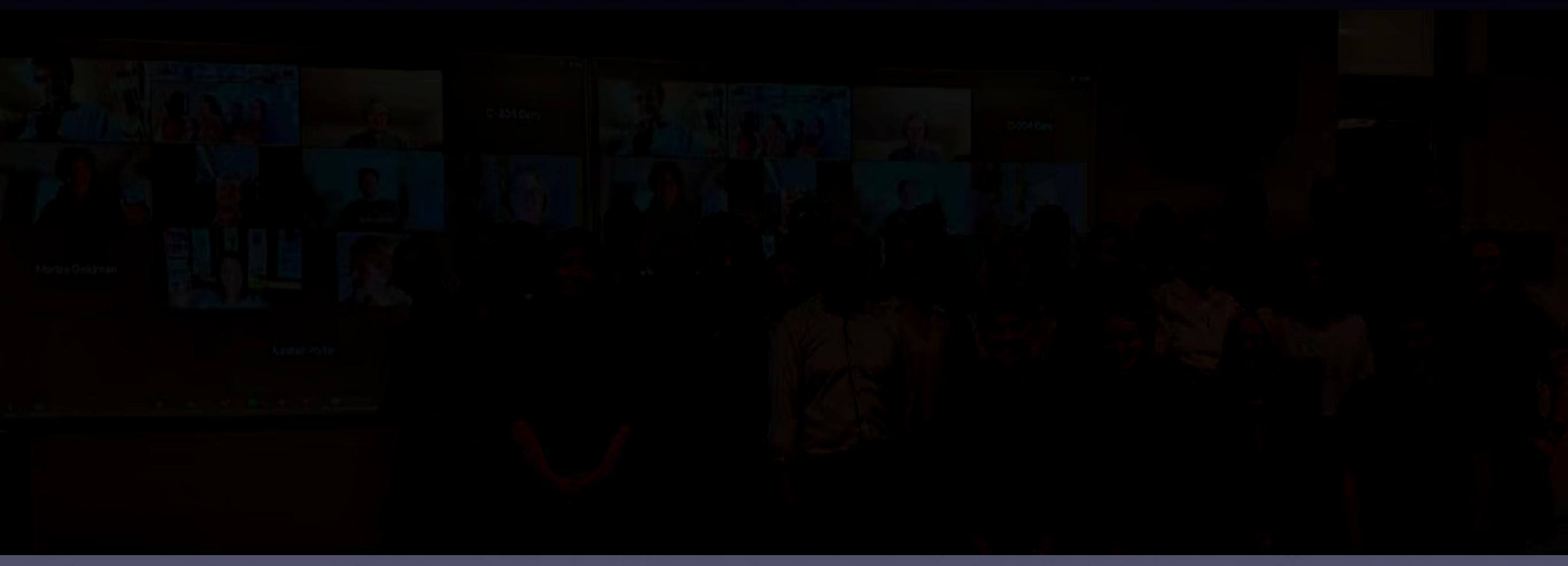
Agenda for Today

- ❖ 09:00–09:30 Introductions (Ichiro Fujinaga)
- ❖ 09:30–10:30 Keynote Talk: Sustaining Digital Musicology (Jennifer Bain)
- ❖ 10:30–11:00 Coffee Break
- ❖ 11:00–11:15 Cantus Ultimus (Dylan Hillerbrand)
- ❖ 11:15–11:30 Reports from Committees
- ❖ 11:30–12:00 Update on the Feast Days Project (Anna de Bakker)
- ❖ 12:00–12:30 Overview on building our Data lake (Ichiro Fujinaga)
- ❖ 12:30–13:30 Lunch (provided)
- ❖ 13:30–15:00 Update on current projects:
 - ❖ Cantus Database (Lucas March)
 - ❖ Liber Usualis (Cole Thierrin)
 - ❖ Rodan (Hanwen Zhang)
 - ❖ UMIL (Universal Musical Instrument Lexicon) (Kun Fang)
 - ❖ Neon (Yinan Zhou)
 - ❖ MS 73 (Kyrie Bouressa)
 - ❖ Annote (Kevin Page)
- ❖ 15:00–15:30 Coffee Break
- ❖ 15:30–16:00 Wikidata Reconciliations (Yueqiao Zhang)
- ❖ 16:00–16:45 Natural Language Query to SPARQL (Junjun Cao)
- ❖ 16:45–17:00 Discussions

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

LinkedMusic Project Meeting II

21 October 2023: McGill University, Montreal



Participants: Julia Craig-McFeely, Andrew Hankinson, Alessandra Ignesti, Jürgen Diet, David Weigl, Jennifer Bain, Martha Thomae, Anna de Bakker, Taz Scott-Talib, Jon Dunn, Wayne Lin, Audrey Laplante, Alastair Porter, Debra Lacoste, Hanwen Zhang, Jenn Reily, Kyrie Bouressa, Jacob deGroot-Maggetti, Yinan Zhou, Cory McKay, Lucas March, Van Pham, Laurent Pugin, Marisa Goldman, Rebecca Mizrahi, Bernhard Lutz, Julie Cumming, Susan Forscher Weiss, Dylan Hillerbrand, David Lewis, Houman Behzadi, Kevin Kishimoto, Jon Manton, Robin Desmeules

LinkedMusic Workshop IV

7 April: Dalhousie University



- ❖ **Guest speaker:** Jan Hajič
- ❖ **Participants:** Jennifer Bain (host), Debra Lacoste, Andrew Hankinson, Anna de Bakker, Dylan Hillerbrand, Julia Craig-McFeely, Tim Expert, Houman Behzadi, Lucero Enríquez Rubio, Julie Cumming, Lionel Li-Xing Hong, Phoebe Durand-McConnell, Marcia Ostashewski

LinkedMusic Workshop V

24 June: Lisbon, Portugal



- ❖ **Guest speaker:** Ichiro Fujinaga
- ❖ **Participants:** Elsa de Luca, (host), Martha Thomae, Alessandra Ignesti, Craig Sapp, Debra Lacoste, Jennifer Bain, Manuel Pedro Ferreira, Francesco Orio, João Pedro d'Alvarenga, Antoine Pham, Hana Vlhová-Wörner

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



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 :

LinkedMusic Committees

❖ Training and Mentorship Committee

- ❖ Chair: Anna de Bakker
- ❖ Jennifer Bain
- ❖ Laurent Pugin
- ❖ Hanwen Zhang

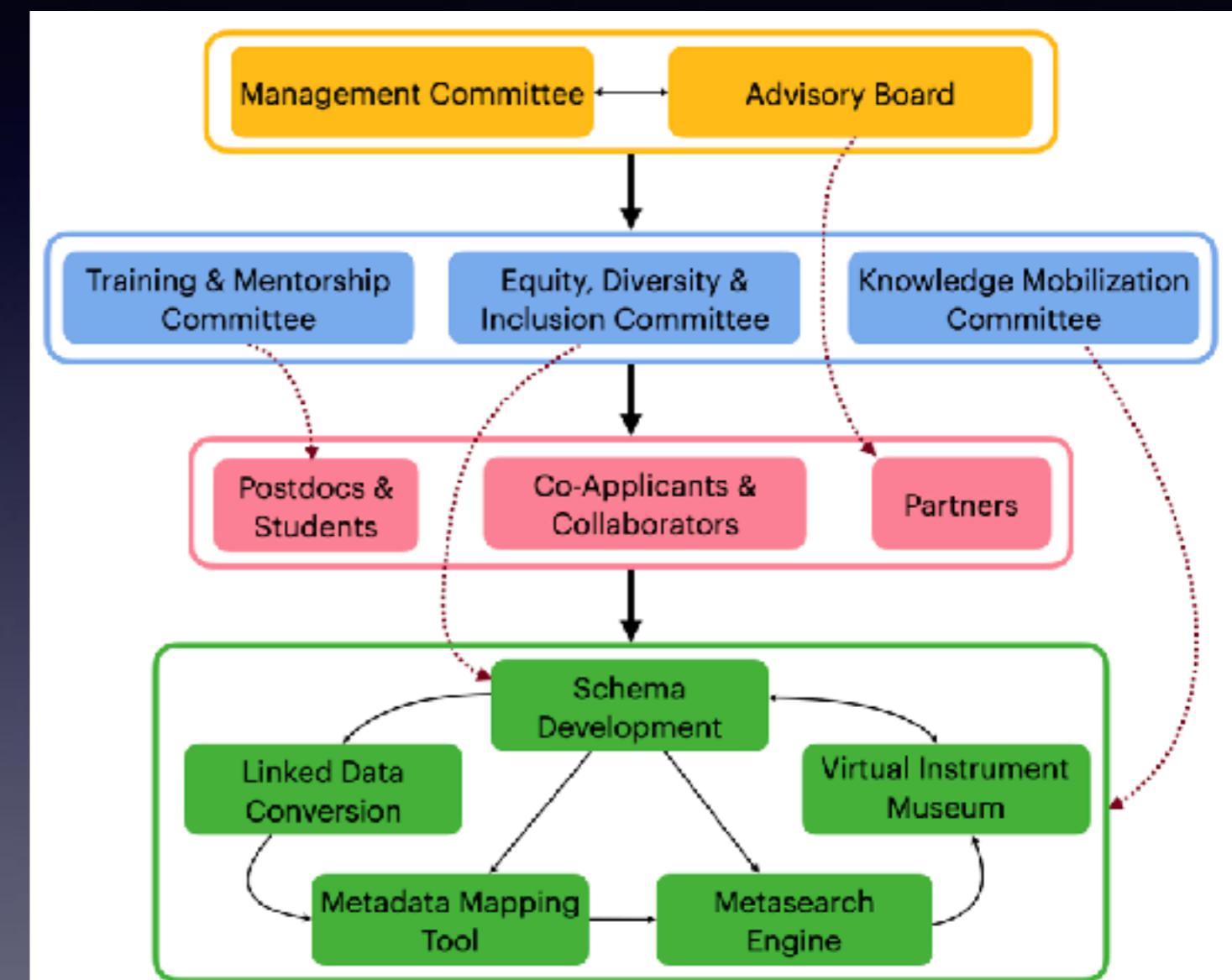
❖ Equity, Diversity, and Inclusion Committee

- ❖ Chair: Julie Cumming
- ❖ Houman Behzadi
- ❖ Robin Desmeules
- ❖ Lucas March

❖ Knowledge Mobilization Committee

- ❖ Debra Lacoste
- ❖ Susan Weiss
- ❖ Julia Craig-McFeely
- ❖ Kyrie Bouressa

❖ Committees will meet at least twice yearly

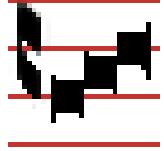


Goals of LinkedMusic

- ❖ Make musical information accessible to more people in the world
- ❖ Make musical queries available in languages other than English

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



Cantus Ultimus and LinkedMusic

Dylan Hillerbrand

Digital Distributed Music Archives and Libraries Lab
McGill University

LinkedMusic Project Meeting – 26 October 2024

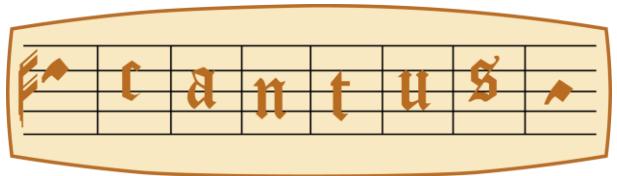
Background

- Developed as part of Single Interface for Music Score Searching and Analysis (SIMSSA) project
 - An end-to-end optical music recognition (OMR) system for chant manuscripts



- Cantus Ultimus provides user interface to the results of OMR process
 - Proof of concept web application for browsing and searching digitized and OMR-ed manuscripts

3 Data Sources



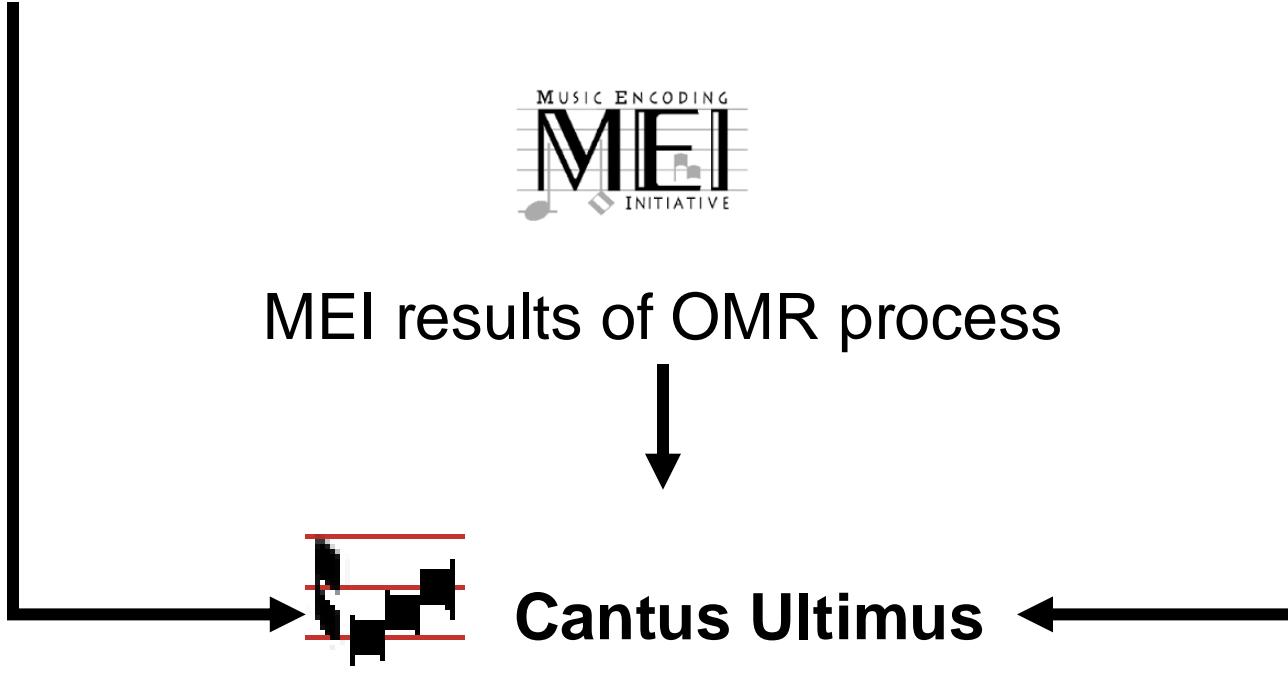
Chant metadata from Cantus Database



Images from host institutions

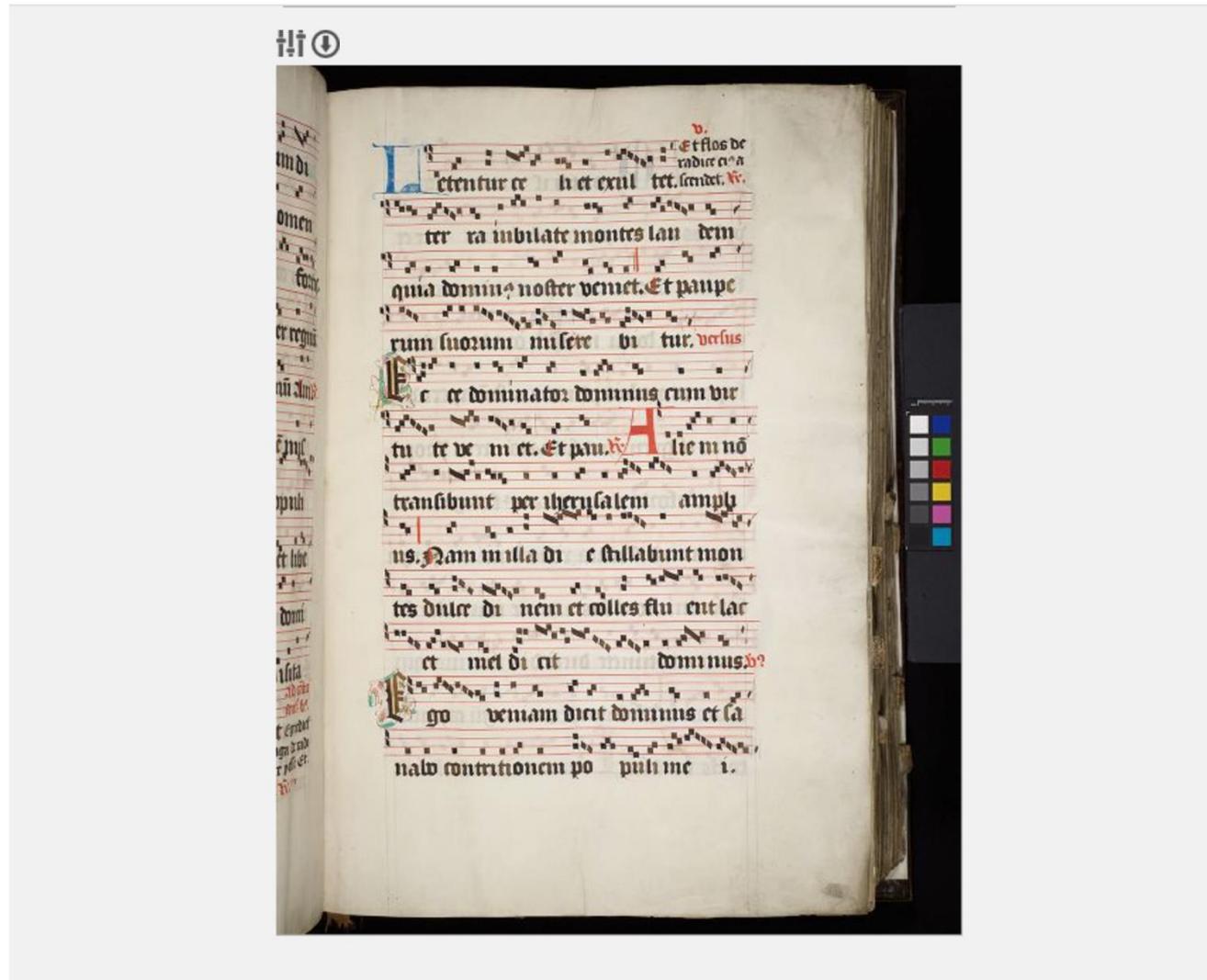


MEI results of OMR process



- + Zoom level: 1.00

Folio 005r (15 of 497)

[Go](#)[Manuscript info](#)[Go to first chant](#)[Previous Folio](#)[Next Folio](#)

Folio 005r

[Search Manuscript](#)

Chants

Laetentur caeli et exsultet terra

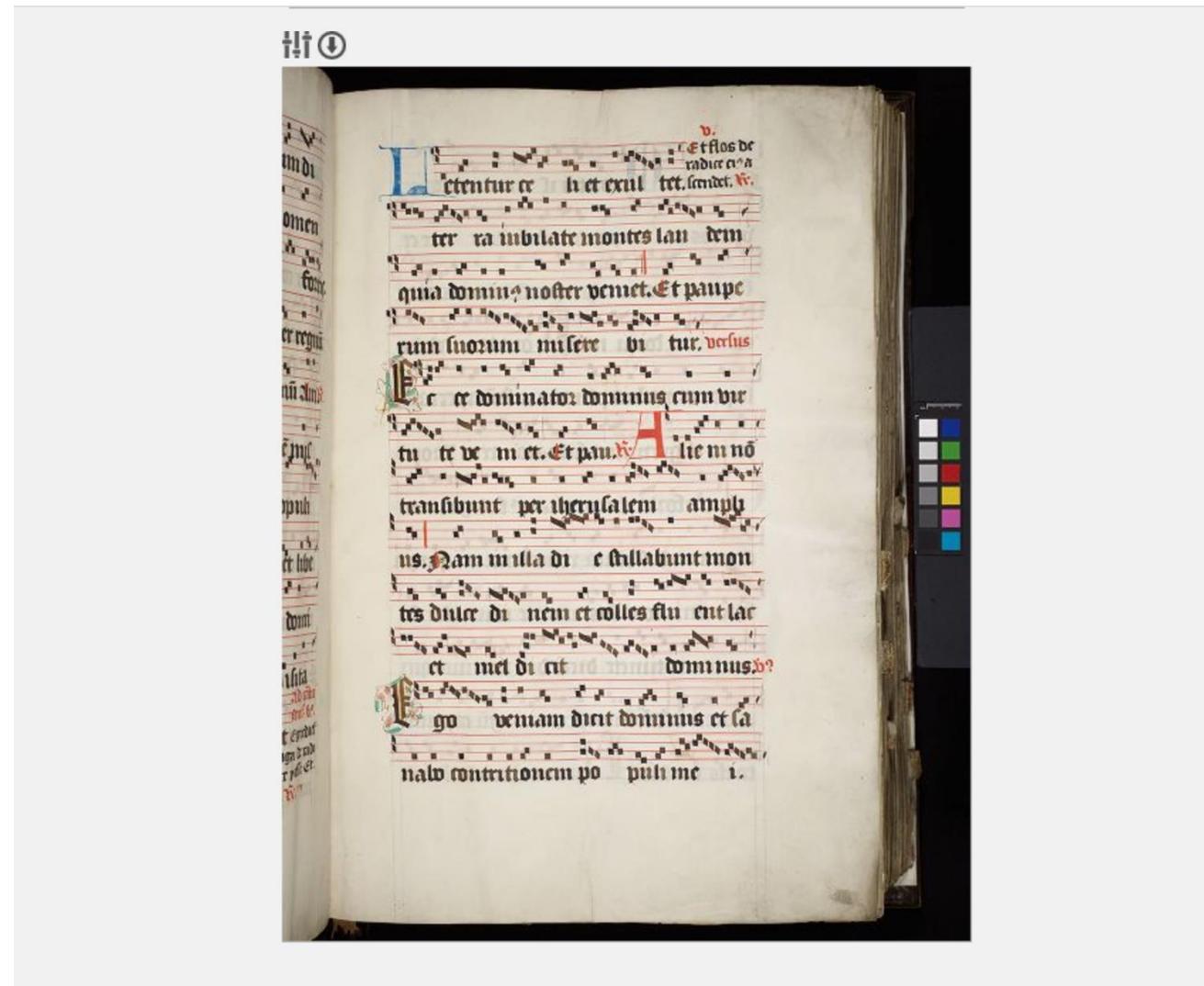
Ecce dominator dominus cum virtute

Alieni non transibunt per Jerusalem

Ego veniam dicit dominus et

- + Zoom level: 1.00

Folio 005r (15 of 497)

[Go](#)[Manuscript info](#)[Go to first chant](#)[Previous Folio](#)[Next Folio](#)

Folio 005r

[Search Manuscript](#)

Ecce dominator dominus cum virtute

Cantus ID: 007068a ([Visit record in Cantus Database](#))**Sequence:** 2**Feast:** Dom. 1 Adventus**Office:** Matins**Genre:** Responsory verse**Mode:** 2[Full Text](#)

Ecce dominator dominus cum virtute veniet

Ec- ce do- mi- na- tor do- mi- nus cum vir- tu- te ve- ni- et Et pauperum

[Play Audio](#) [Stop Audio](#)

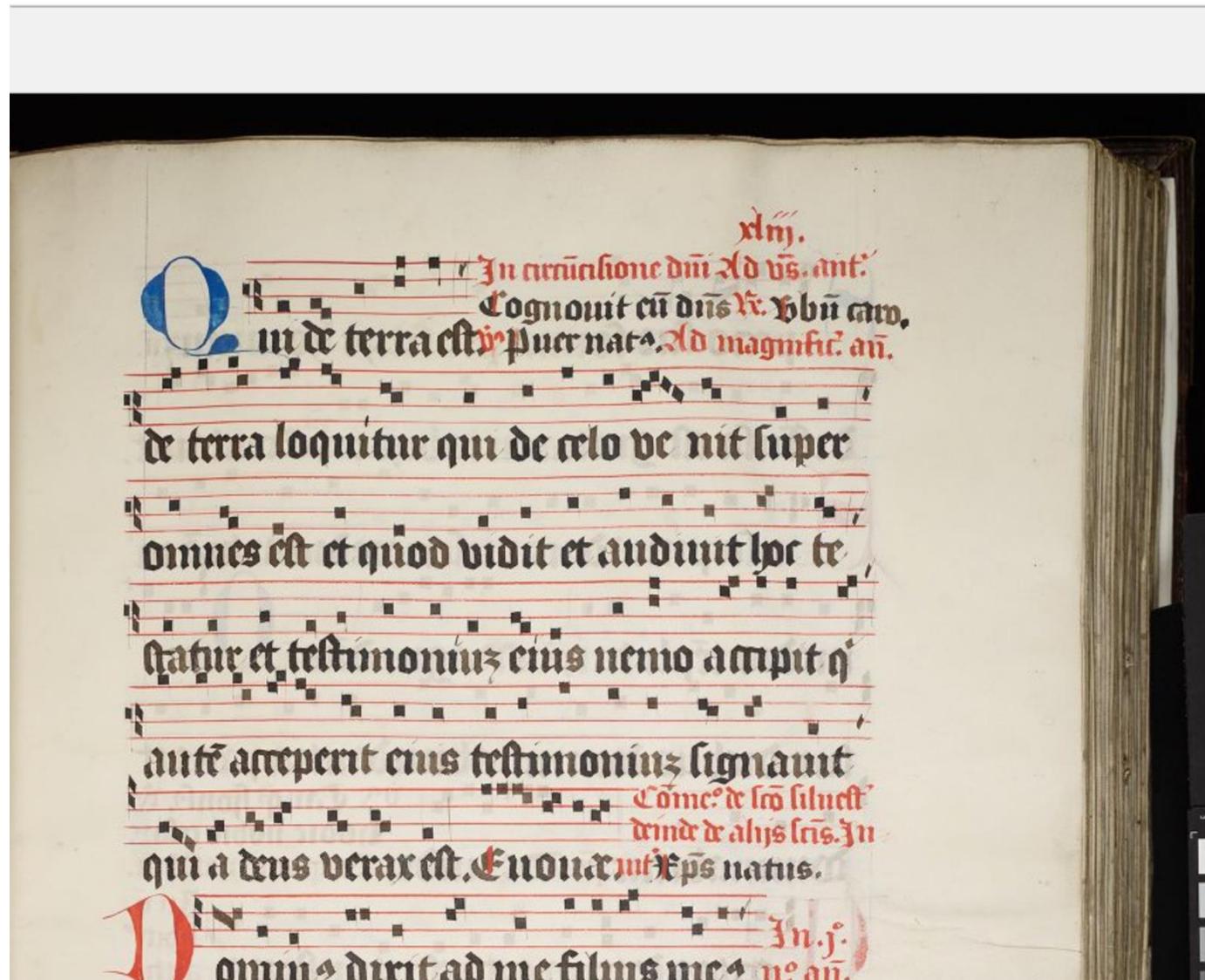
- + Zoom level: 2.00

Folio 043r (93 of 497)

Go



Manuscript info



[Go to first chant](#) [Previous Folio](#) [Next Folio](#)

Folio 043r

Search Manuscript

Full Text

Qui de terra est de terra loquitur qui de caelo venit super omnes est et quod vidit et audivit hoc testatur et testimonium ejus nemo accipit qui autem acceperit ejus testimonium signavit quia deus verax est

Qui de ter-ra est de ter-ra lo-qui-tur qui de ce-lo ve-
nit su-per om-nes est et quod vi-dit et au-di-vit hoc tes-ta-tur et tes-
ti-mo-ni-um e-ius ne-mo ac-ci-pit qui au-tem ac-ce-pe-rit e-ius
tes-ti-mo-ni-um sig-na-vit qui-a de-us ve-rax est E-u-o-u
ae

Play Audio

Stop Audio



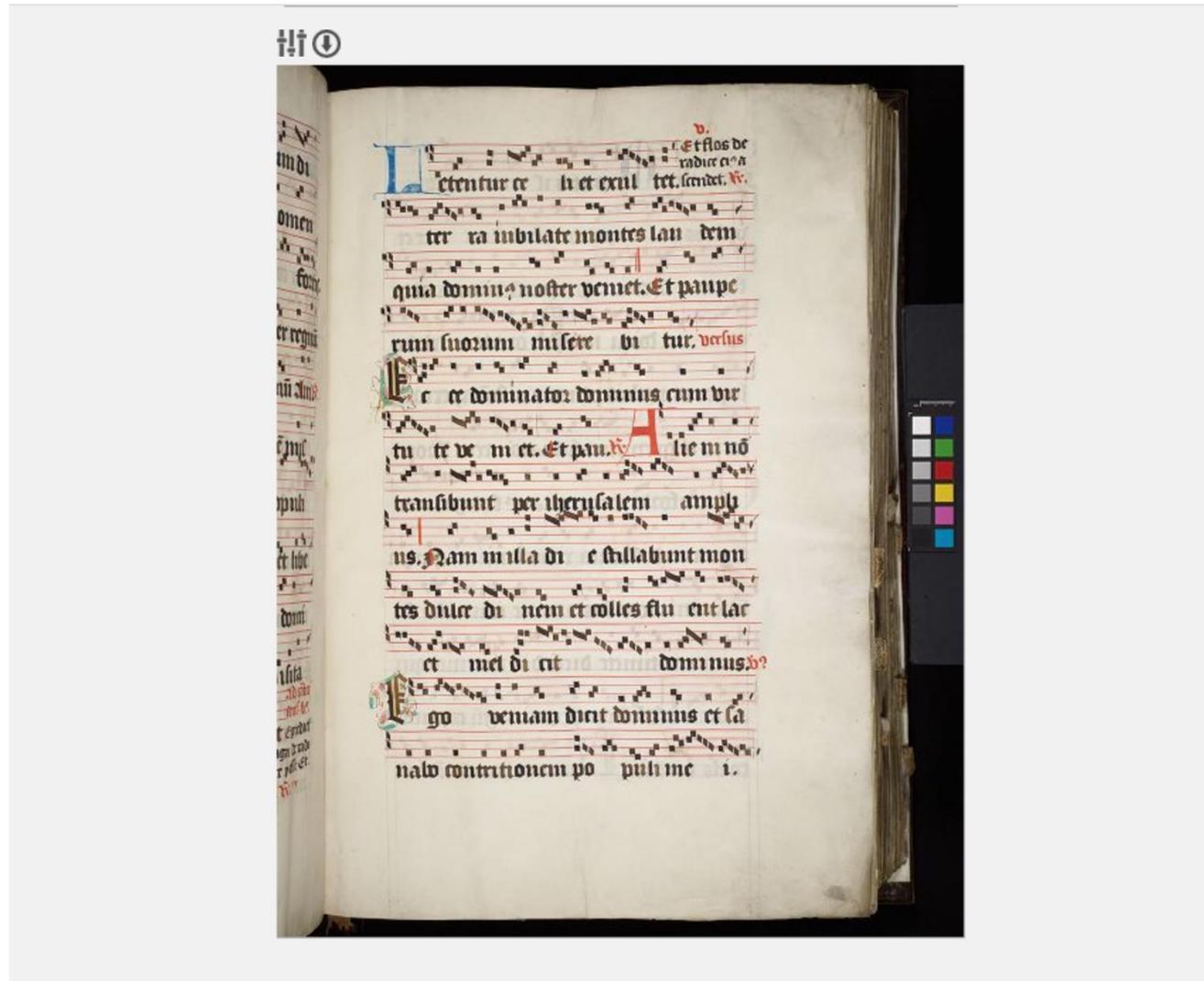
Treble Voice

- + Zoom level: 1.00

Folio 005r (15 of 497)

[Go](#)

Manuscript info

[Go to first chant](#)[Previous Folio](#)[Next Folio](#)

Folio 005r

Search Manuscript

Feast ▾

"Dom. 3 Adventus"

53 results for query: "Dom. 3 Adventus"

Folio	Incipit	Mode	Feast	Genre	Office
014r	Qui venturus* ⚡		Dom. 3 Adventus	Responsory	First Vespers
014r	Rorate caeli desuper* ⚡		Dom. 3 Adventus	Versicle	First Vespers
014r	Deus creator* ⚡		Dom. 3 Adventus	Hymn	First Vespers
014r	Ante me non est formatus ⚡	1 T	Dom. 3 Adventus	Antiphon	First Vespers
014r	Surgite vigilemus venite adoremus quia ⚡	2 T	Dom. 3 Adventus	Invitatory antiphon	Matins
014r	Domine in virtute* ⚡		Dom. 3 Adventus	Antiphon	Matins
014r	Ex Sion species decoris ejus* ⚡		Dom. 3 Adventus	Versicle	Matins
014r	Ecce apparebit dominus super nubem ⚡	1	Dom. 3 Adventus	Responsory	Matins

Cantus Ultimus Today

- 13 manuscripts (~43,000 chants)
 - 2 OMR-ed manuscripts
 - Halifax (Canada), St. Mary's University – Patrick Power Library (CDN-Hsmu) M2149.L4 ("Salzinnes Antiphonal")
 - Einsiedeln, Kloster Einsiedeln – Musikbibliothek (CH-E) 611
 - 1 virtual reconstruction
 - Lambach Abbey, Gottschalk Antiphoner

Cantus Ultimus Today

- Web application (cantus.simssa.ca)
- Debut as a “digital manuscript” kiosk
 - “Centuries of Silence” exhibit at *Musée des Arts anciens*, Namur, Belgium

Cantus Ultimus & LinkedMusic

- Return of OMR-powered search
- Simplifying the Cantus Ultimus – Cantus Database connection
- Linking MEI-encoded images with Cantus Database metadata

OMR Search

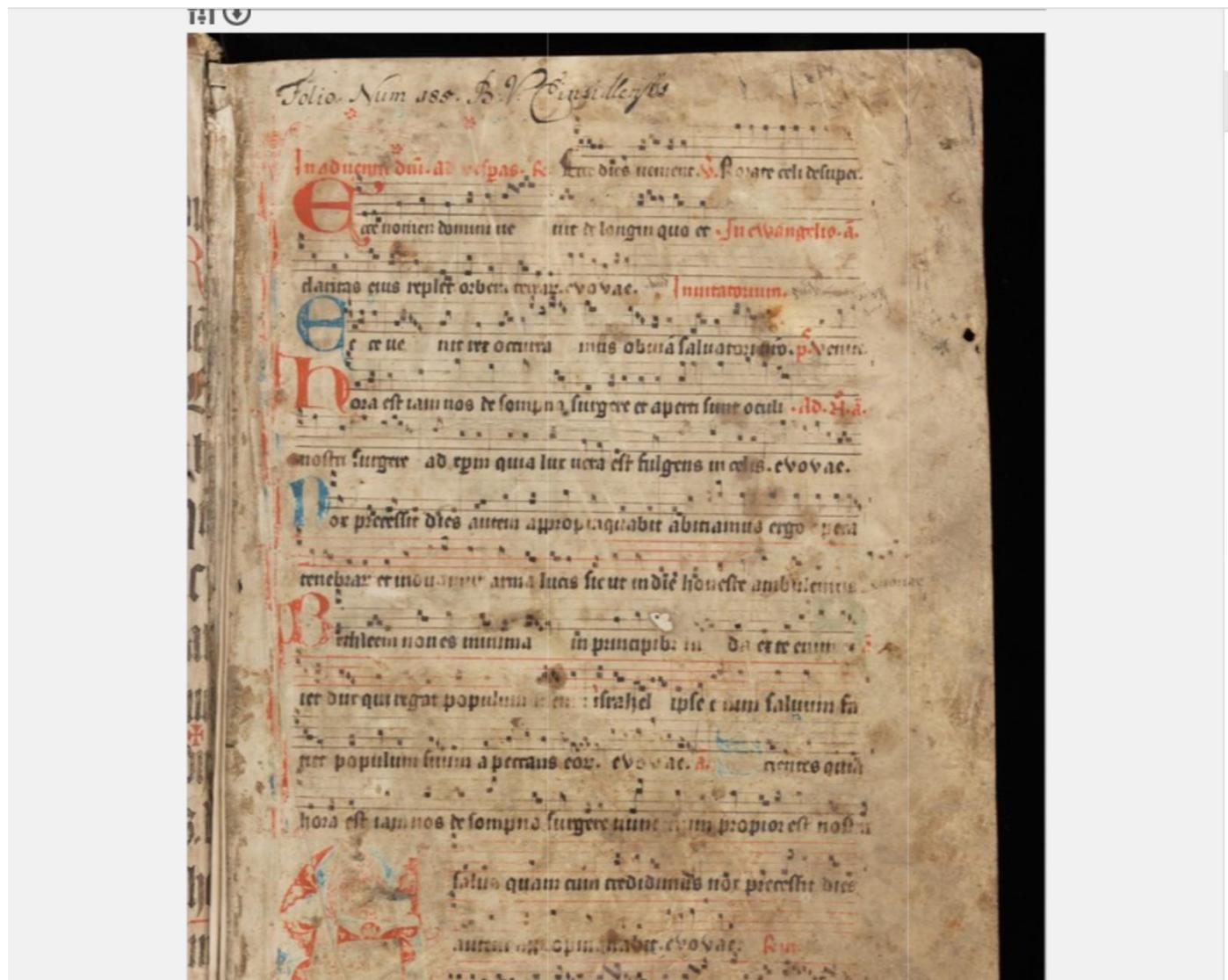
- Old version of Cantus Ultimus had OMR search...what happened?
 - New images
 - MEI has errors
 - MEI grew up (version 5 released)
- Proofread MEI now available:
 - Salzinnes Antiphonal (CDN-Hsmu M2149.L4) MEI completed + proofread
 - Einsiedeln 611 (CH-E 611) completed

Benefits of MEI-powered search

- Previous pitch search powered by volpiano
- MEI contains:
 - Directly connected text and pitches
 - Notation information
 - Encodes attributes of neumes and neume component shapes (for example, puncta and virga)
 - Clef position
 - Location in the image

- + Zoom level: 2.00

Folio 001r (3 of 573) Go   Manuscript info



 Go to first chant  Previous Folio  Next Folio

Folio 001r

Search Manuscript

Neume ▾

Notation

Search

Available neumes



clivis



scandicus



torculus



tristopha

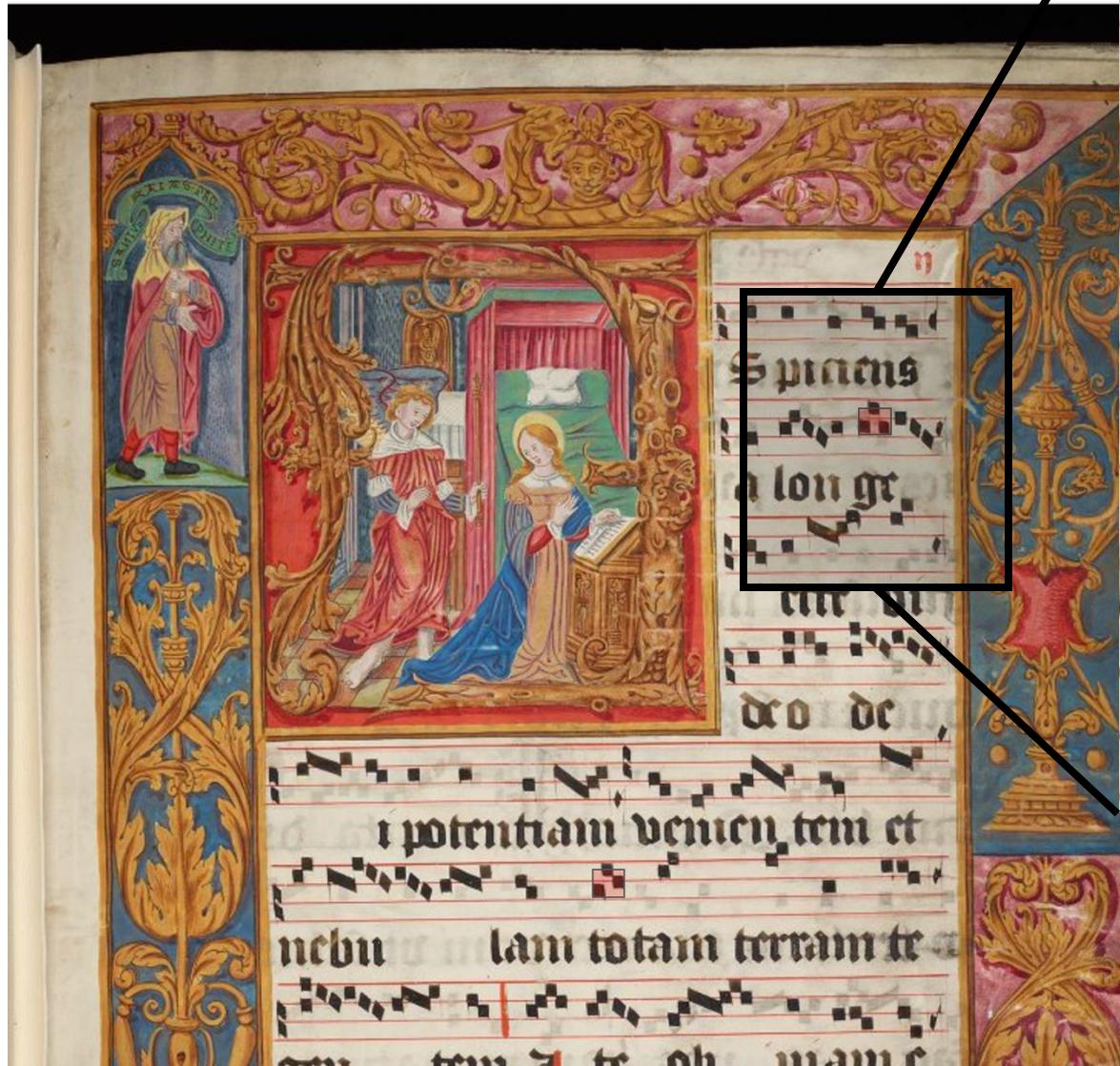


pressus



poi

- + Zoom level: 2.00



manuscript info

Next Folio ❤

Search

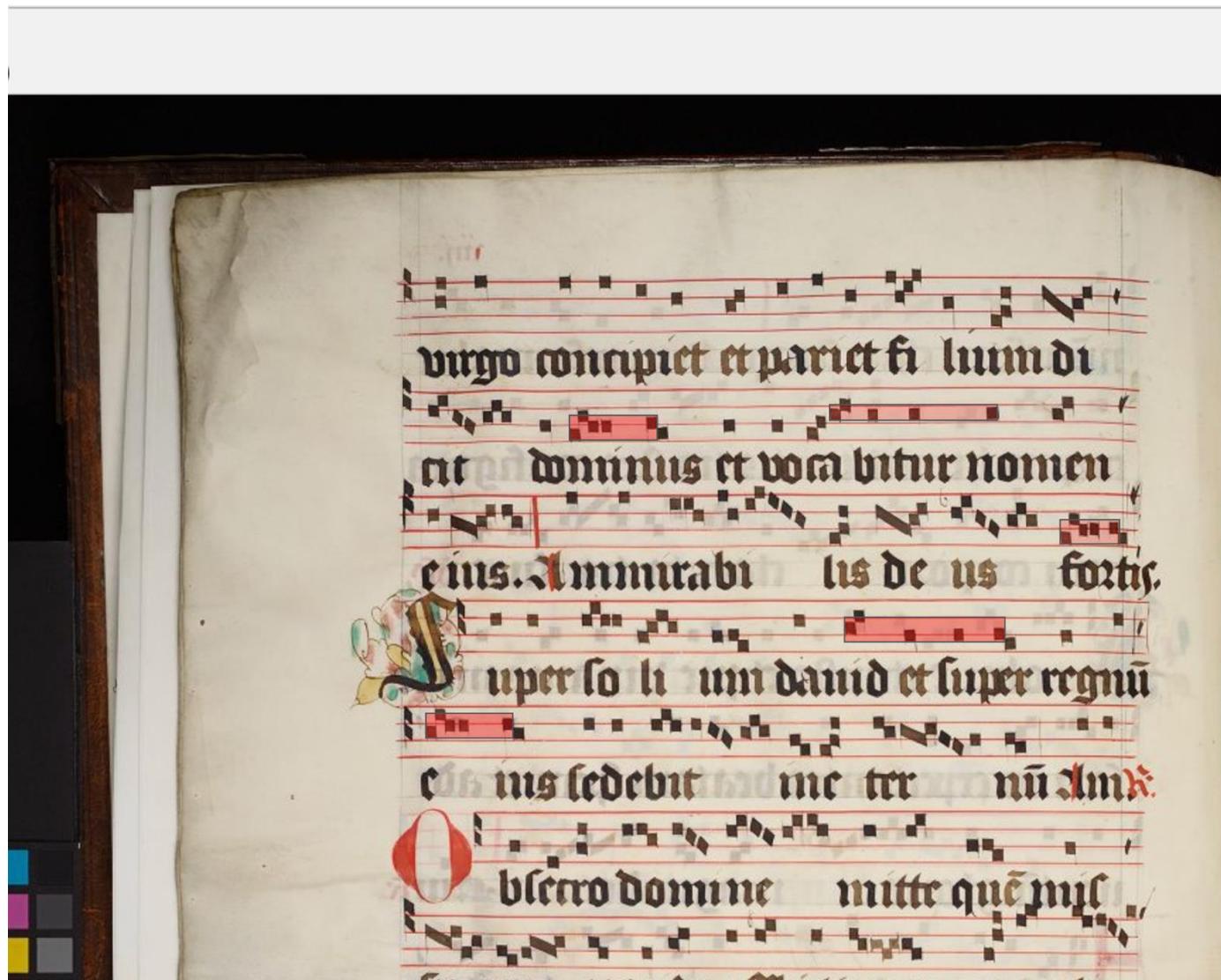
processus

A diagram illustrating the relationship between four specific musical examples (labeled 002r) and their corresponding labels and arrows. The labels are: d f d, d f e, dec, and e f e. Each label is connected by an arrow to a small image of a musical example from folio 002r.

002r	d f d	
002r	d f e	
002r	dec	
002r	e f e	

- + Zoom level: 2.00

Folio 004v (14 of 497) Manuscript info



Go to first chant Previous Folio Next Folio

Folio 004v

Search Manuscript

Contour ▾

u d rr

Search

Up Down Repeat

2070 results for query: u d rr

004r		b d c c c
004r		e g f f f
004r		e g f f f
004r		b d c c c
004r		b d c c c
004v		e g f f f
004v		e g f f f
004v		g a g g g
004v		e g f f f
004v		b c a a

Cantus Ultimus as a frontend to Cantus Database

- Legacy proof of concept designed for a single import of a portion of Cantus Database. Cantus Ultimus misses out on:
 - Improvement in Cantus Database, like new fields
 - Thousands of records created or edited in the past year
- Cantus Ultimus as:
 - Repository of MEI
 - Record the link between images and folios
 - Search engine

Cantus Ultimus as a frontend to Cantus Database

- Simplify (or remove?) the Cantus Ultimus database
- Improve storage of MEI
- Index data directly from Cantus Database
 - Requires new developments on Cantus Database

Linking MEI and Chant Metadata

- MEI search and chant metadata search are siloed
 - MEI encodes the contents of a page, system, and syllable
 - Chant metadata (in Cantus Database) has information on the chant
- “Show me staff where the word ‘caeli’ exists in a chant for Advent and is sung on a up-down contoured neume (torculus) followed by a two-note repeated neume (distropha)?”

Thank you

Dylan Hillerbrand

dylan.hillerbrand@mail.mcgill.ca

Update on Feast Days Project or: “Encoding Time Matters”

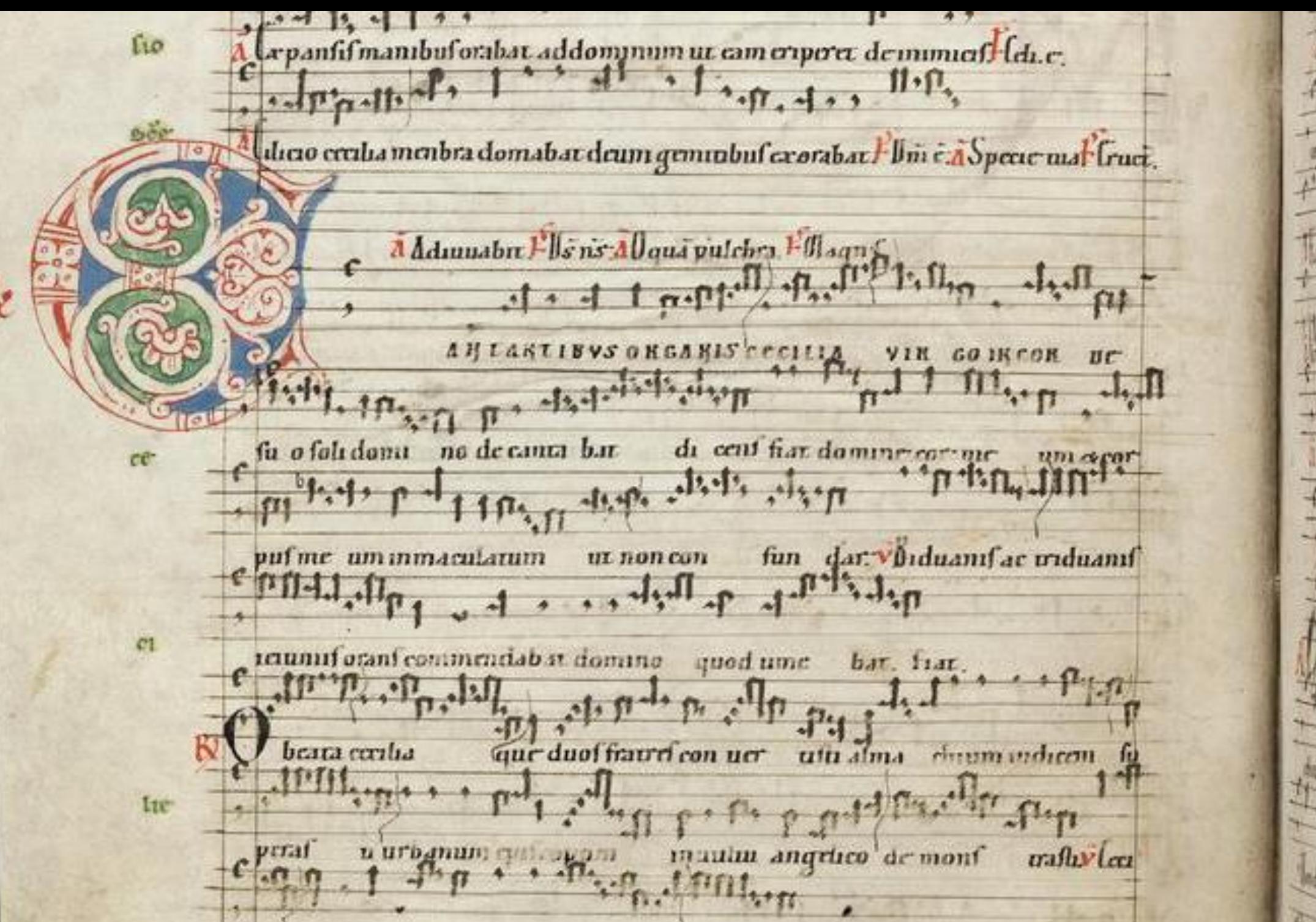
Anna de Bakker, LinkedMusic Project Meeting, 26 October 2024

the general idea:

associating music with a *recurring* point in time

-In Cantus Database these are called “Feasts”

“Caecilae”=Feast of Saint Cecilia, 22 November



Cantus ID: 006267 ([Visit record in Cantus Database](#))

Sequence: 7

Feast: Caeciliae

Office: Matins

Genre: Responsory

Mode: 8

Full Text

Cantantibus organis Caecilia virgo in corde suo soli domino decantabat dicens fiat domine
cor meum et corpus meum immaculatum ut non confundar

CANTANTIBUS ORGANIS CLAVICILIIS LIUTIS

VIRGO IN CORDE

the general idea:

associating music with a *recurring* point in time

-In Cantus Database these are called “Feasts”



Folio 143v

Chants

Praecursor domini*

Cantus ID: 007420 ([Visit record](#))

Sequence: 1

Feast: Joannis Baptistae

Office: First Vespers

Genre: Responsory

Mode: 8

Full Text

Praecursor domini*

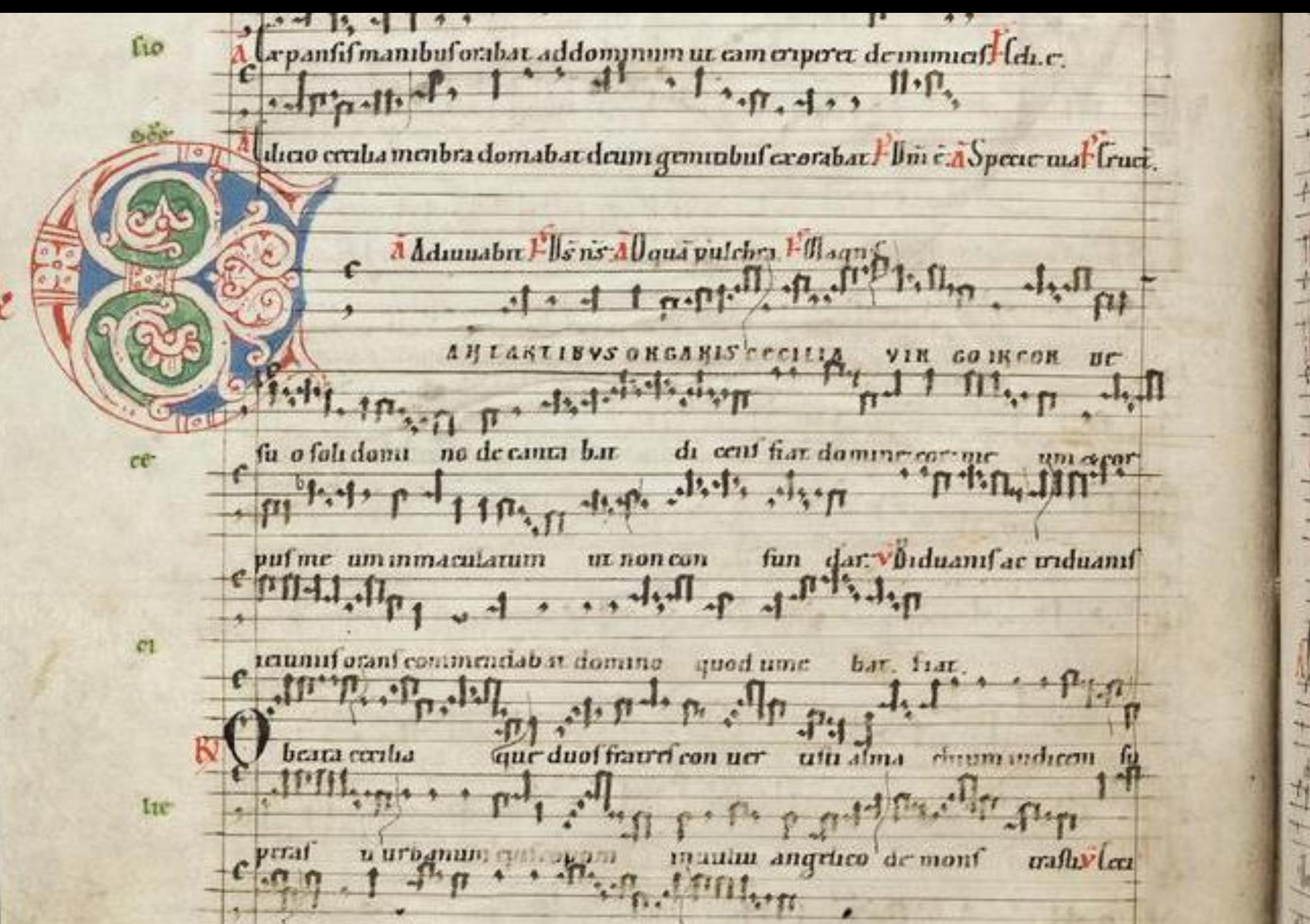
“**Joannis
Baptistae**=Feast of
John the Baptist, 24
June

the general idea:

associating music with a *recurring* point in time

-In Cantus Database these are called “Feasts”

-controlled vocabulary allows comparison across sources



Cantus ID: 006267 ([Visit record in Cantus Database](#))

Sequence: 7

Feast: Caeciliae

Office: Matins

Genre: Responsory

Mode: 8

Full Text

Cantantibus organis Caecilia virgo in corde suo soli domino decantabat dicens fiat domine cor meum et corpus meum immaculatum ut non confundar

4

CANTAN- TI- BUS OR- GA- NIS CE- CI- LI- A VIR- GO IN COR-

Music for saints's days: not just chant...

JOHANN SEBASTIAN BACH
(1685-1750)

CANTATES POUR LA SAINT JEAN-BAPTISTE
CANTATAS FOR THE FEAST OF ST JOHN THE BAPTIST

1 :: 12 **FREUE DICH, ERLÖSTE SCHAR, BWV 30** 34:07

pour soprano, alto, ténor, basse, 2 traversières, 2 hautbois et hautbois d'amour,
cordes et basse continue (Leipzig, 1737-1742)

13 :: 19 **CHRIST UNSER HERR ZUM JORDAN KAM, BWV 7** 23:15

pour soprano, alto, ténor, basse, 2 hautbois d'amour, cordes et basse continue
(Leipzig, 24 juin 1724)

20 :: 24 **IHR MENSCHEN, RÜHMET GOTTES LIEBE, BWV 167** 16:04

pour soprano, alto, ténor, basse, hautbois, hautbois da caccia, cordes et basse continue
(Leipzig, 24 juin 1723)

FESTIVAL
montréal
baroque

ATMA Classique
www.atmaclassique.com

SACD Surround • SACD Stereo • CD Stereo
musicaction
Canada
Société
de développement
des technologies
multimédias
Québec

Edition Eulenburg
No. 1062

PURCELL

ODE FOR ST CECILIA'S DAY 1683

Welcome to all the pleasures

5

ALEXANDER'S FEAST:

OR, THE

POWER of MUSIC.

A N

O

D

E.

Wrote in Honour of

St. C E C I L I A,

By Mr. D R Y D E N.

As performed by the

CASTLE-SOCIETY

A T

HABERDASHERS-HALL.

Set to Music by Mr. HANDEL.

L O N D O N :

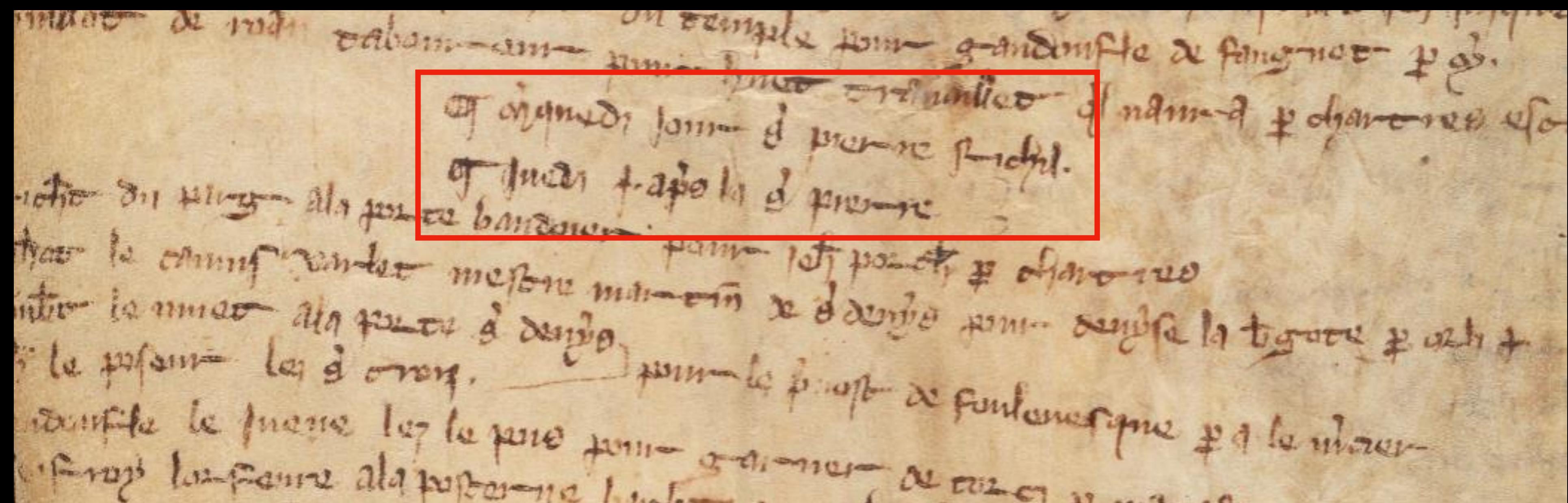
Printed in the YEAR 1760.

1173

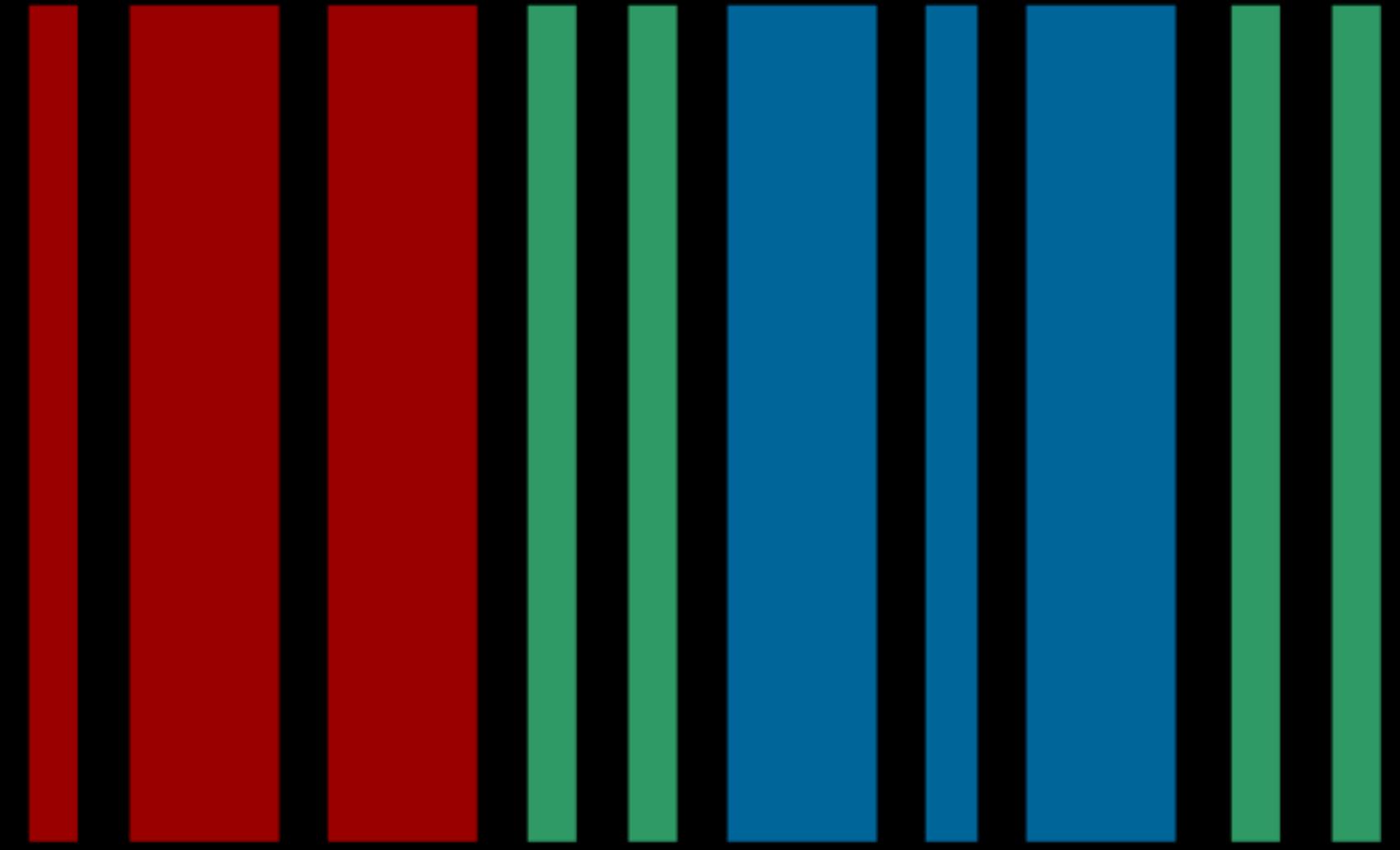
Saints on the calendar: not just in churches...

‘mardi Jour S.Pierre. Nichil.’ =Tuesday, Feast of Saint Peter (June 29)

‘Juedi j. apres la s.pierre.’ =Thursday, Day after Feast of Saint Peter (June 30)



Legal text (arrests and summonses in
Paris, June 1295 (?)



WIKIDATA
WORLD
DOMINATION

types of feasts in Cantus Database

- Sanctorale (~900 items)
- Temporale (~400 items)
- Everything else (~300 items)

types of feasts in Cantus Database

- Sanctorale (~900 items)

- feasts commemorating a holy figure (saint)

- usually on the anniversary of their death

e.g. Remigius: died 13 January 533 CE



Saint Remigius baptizes Clovis, by the Master of Saint Gilles c. 1500. National Gallery, DC

types of feasts in Cantus Database

- Sanctorale (~900 items)
- feasts commemorating a holy figure (saint)
- usually on the anniversary of their death

e.g. St. Piatus: October 1 (ca. 286 CE?)

BnF NAF 16251, f. 84v



types of feasts in Cantus Database

-Temporale (~400 items)

-feasts whose date is calculated by reference to Christmas or Easter

-usually a day of the week + number of weeks from reference point



types of feasts in Cantus Database

Ordo quando rex cum exercitu ad premium egreditur	Rite when the king goes out to battle with his army
Ordo quando sal ante altare ponitur antequam exorcizetur	Rite when the salt is placed before the altar for its exorcism
Ordo super eum qui barbam tangere cupit	Rite for him who desires to anoint his beard
Ordo super sepulcrum quando clamore proclamatur	Rite over the sepulchre when the clamor is proclaimed

-Everything else (~300 items)



types of feasts in Cantus Database

-Sanctorale (~900 items)

-Temporale (~400 items)

~~-Everything else (~300 items)~~

Often describing a ritual action or category, not a date

Fourth Sunday of Easter (Q5256289)

fourth Sunday of the Easter season, being the day that occurs three weeks after the Christian celebration of Easter

 edit

Third Sunday after Easter | Good Shepherd Sunday | Fourth Sunday of Eastertide | Jubilate

▼ In more languages

Configure

Language	Label	Description	Also known as
English	Fourth Sunday of Easter	fourth Sunday of the Easter season, being the day that occurs three weeks after the Christian celebration of Easter	Third Sunday after Easter Good Shepherd Sunday Fourth Sunday of Eastertide Jubilate
French	No label defined	No description defined	
Chinese	No label defined	No description defined	
Cantonese	No label defined	No description defined	

[All entered languages](#)

Statements

subclass of	 Sunday	 edit
	▼ 0 references	+ add reference

part of	 Sundays of Easter	 edit
	follows	Misericordia
	followed by	Fifth Sunday of Easter.

Temporale:
partially extant on
Wikidata already

feast day

Easter + 39 days 

Potential issues X

item-requires-statement constraint Help Discuss

An entity with [feast day](#) should also have a statement [canonization status](#).

subject type constraint Help Discuss

Entities using the [feast day](#) property should be instances of one of the following classes (or of one of their subclasses), but [Feast of the Ascension](#) currently isn't:

- [human](#)
- [group of humans](#)
- [title of Mary](#)
- [human whose existence is disputed](#)
- [legendary figure](#)
- [attributes of God in Christianity](#)
- [Bible story](#)
- [pericope](#)

This result is cached and might be out of date by up to 7 minutes.

named after

followed by

commemorates

Ascension of Jesus

▼ 0 references

Temporale:
partially extant on
Wikidata already

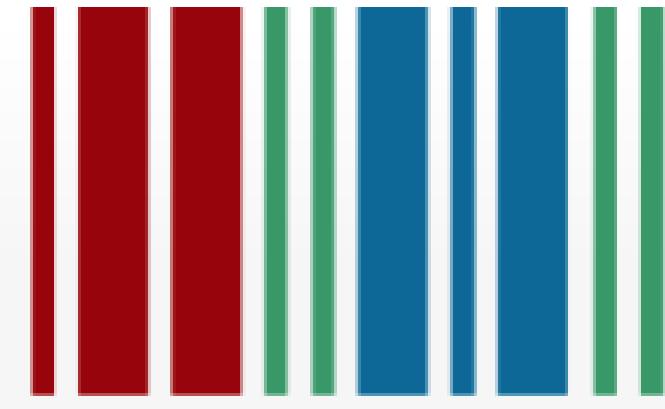
...though not always without issues...

types of feasts in Cantus Database

- Sanctorale (~900 items)
- Temporale (~400 items)
- Everything else (~300 items)

The problem is now easier...right?

What does WikiData have already?



The logo consists of seven vertical bars of varying heights and colors: two red, one green, three blue, and one green. Below the bars, the word "WIKIDATA" is written in a bold, sans-serif font.

Property Discussion

feast day (P841)

saint's principal feast day

-“feast day” property—but this applies to people, not points in time

What does WikiData have already?

Michael Day (Q2048437)

No description defined

► In more languages

Statements

instance of	<ul style="list-style-type: none">holiday<ul style="list-style-type: none">▼ 0 referencesSlavic folk holiday<ul style="list-style-type: none">▼ 0 references
part of	<ul style="list-style-type: none">Slavic folk calendar<ul style="list-style-type: none">▼ 0 references

- surprising number of “Slavic folk holiday”
- some national holidays
- no link to the saints

Saint David's Day (Q2454744)

cultural and religious holiday celebrated on 1 March
St. David's Day | St David's Day | Feast of Saint David of Wales

► In more languages

Statements

instance of	<ul style="list-style-type: none">holiday<ul style="list-style-type: none">▼ 0 referencesnational day<ul style="list-style-type: none">▼ 0 references
-------------	--

What does WikiData have already?

-particular local practices...

-not always with date or saint linked

Feast of St. Anthony of Padua in Palmi (Q3743962)

No description defined

▶ In more languages

Statements

instance of

▷ religious and cultural festive day

▼ 0 references

▷ processional parade

▼ 0 references

Feast of St. Anthony (Boston) (Q5439541)

Festival in Boston, Massachusetts, United States

▶ In more languages

Statements

instance of

▷ religious and cultural festive day

▼ 0 references

dedicated to

▷ Anthony of Padua



▼ 0 references

(“made of”= wood

“height” = 170 cm tall)

Saint Cecilia: the trial entry



<https://www.wikidata.org/wiki/Q118849459>

“Day in year for periodic occurrence” **22 November**

“Commemorates”/“Named after” **Saint Cecilia**
“Instance of”

holiday, Christian holy day, religious and cultural festive day

Reference statements as applicable

Saint Cecilia: the trial entry



-Can now be used to talk about music!

Welcome to all the pleasures (Q2522109)

Composition by Henry Purcell

created for

Feast of Saint Cecilia

▼ 0 references

complications

- some saints have multiple associated feasts!
 - sometimes commemorate life events, e.g. “day they became a bishop”
 - sometimes commemorate events after death!



Remigius: re-buried in the abbey church of Saint-Rémy, Reims

1 October 1099

complications

- some saints have multiple associated feasts!
 - sometimes commemorate life events, e.g. “day they became a bishop”
 - sometimes commemorate events after death!
- sometimes feasts have multiple associated saints
- sometimes the feast defines a period around it
 - e.g. “Sunday after the feast of Remigius”

the breakdown

of ~900 “saints’ days” in Cantus Database:

- 55 commemorate “life events”
- 90 commemorate “after death events”
- 102 commemorate groups of saints
- 163 are part of a mini-holiday season

this leaves ~500 “simple” saints days

Reconciling the “simple” saints

Name	English description	CantusFeastCode	commemorates	CantusFeastCode
Aureae	Aurea, Abbess in Paris	14100410		
Aureae Ostiensis	Aurea (Chryse) of Ostia	14082430	Saint Aurea of Paris Choose new match	14100410
Aureliae Argentoratensis	Aurelia, Virgin (4th century)	14101510	Aurea of Ostia Choose new match	14082430
Austremonii	Austremonius (Stremoine), Bishop, Martyr, Apostle of Auvergne	14110160	Aurelia of Strasbourg Choose new match	14101510
Aviti	Avitus, Abbot and Confessor, born at Aurillac	14061720	Austromoine Choose new match	14110160
Baboleni	Babolenus, Abbot	14062610	Avit du Périgord Choose new match	14061720
Babylae	Babylas, Martyr, 12th Bishop of Antioch	14012420	Babolein Choose new match	14062610
Babylae Pampilonensis	Babylas, Bishop of Pamplona	14103020	Babylas of Antioch Choose new match	14012420

Reconciling the “simple” saints

- found all but 15 saints (eventually)
- 8 of these were incompletely defined in Cantus

Reconciling the “simple” saints

commemorates	CantusFeastCode
Saint Aurea of Paris Choose new match	14100410
Aurea of Ostia Choose new match	14082430
Aurelia of Strasbourg Choose new match	14101510 edit
Austromoine Choose new match	14110160
Avit du Périgord Choose new match	14061720
Babolein Choose new match	14062610
Babylas of Antioch Choose new match	14012420

Saints could now be made into feasts!

Feast of Saint Aurea Choose new match	edit
Feast of Saint Aurea (Chryse) of Ostia Choose new match	
Feast of Saint Aurelia Choose new match	
Feast of Saint Austremonius (Stremoine) Choose new match	
Feast of Saint Avitus Choose new match	
Feast of Saint Babolenus Choose new match	
Feast of Saint Babylas Choose new match	

Reconciling the “simple” saints

commemorates	CantusFeastCode
Saint Aurea of Paris Choose new match	14100410
Aurea of Ostia Choose new match	14082430
Aurelia of Strasbourg Choose new match	14101510 edit
Austromoine Choose new match	14110160
Avit du Périgord Choose new match	14061720
Babolein Choose new match	14062610
Babylas of Antioch Choose new match	14012420

Saints could now be made into feasts!

ongoing:
-secondary feasts related to one saint or saints' day

- [Feast of Saint Aurea](#)
[Choose new match](#)
- [Feast of Saint Aurea \(Chryse\) of Ostia](#)
[Choose new match](#)
- [Feast of Saint Aurelia](#)
[Choose new match](#)
- [Feast of Saint Austremoine \(Stremoine\)](#)
[Choose new match](#)
- [Feast of Saint Avitus](#)
[Choose new match](#)
- [Feast of Saint Babolenus](#)
[Choose new match](#)
- [Feast of Saint Babylas](#)
[Choose new match](#)
- [Feast of Saint Babylas](#)
[Choose new match](#)
- [Feast of Saint Balbina](#)
[Choose new match](#)

last to be solved: multiple saints

Sergius and Bacchus (Q140013)

Roman martyrs and early Christian saints

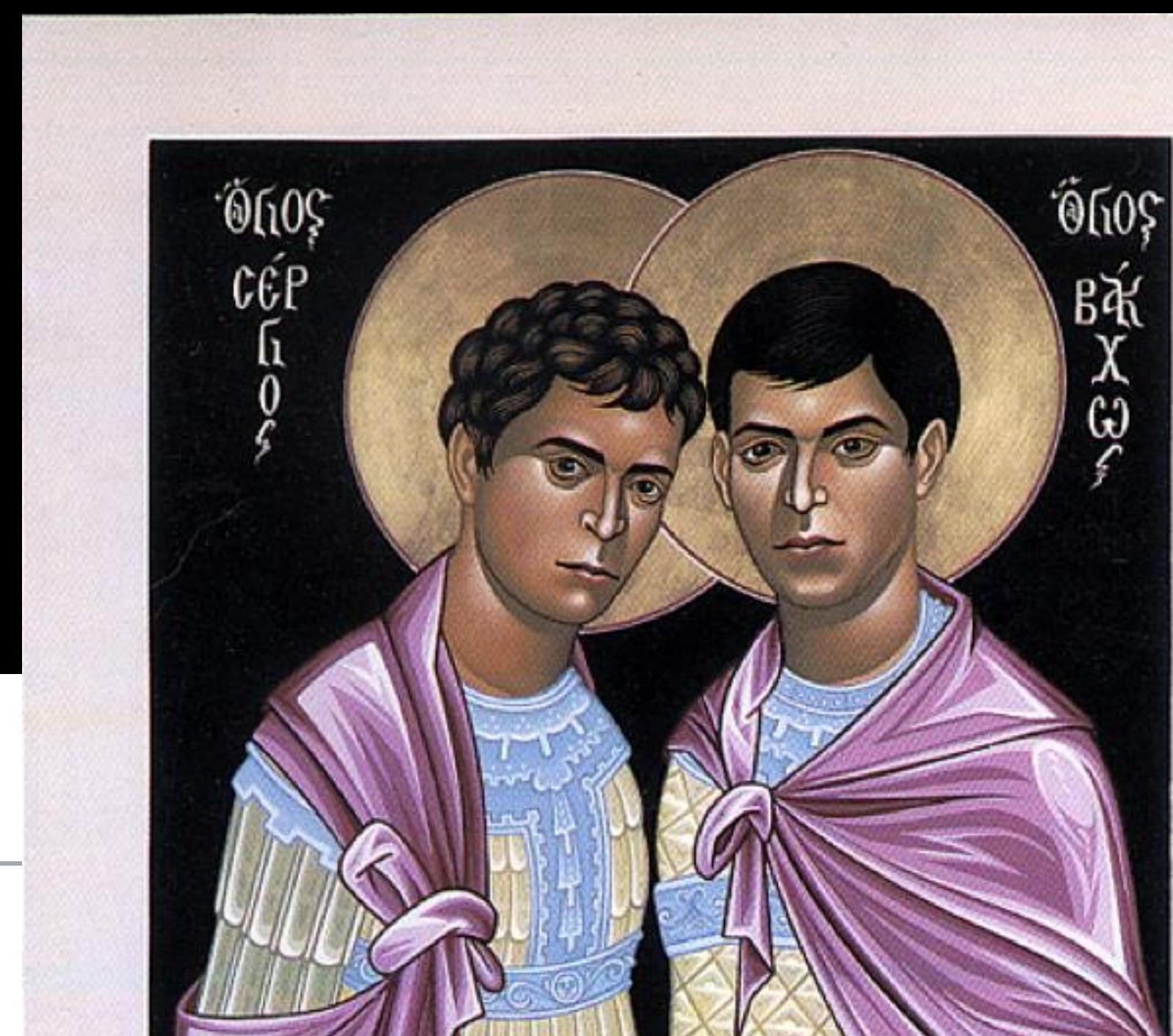
► In more languages

Statements

instance of

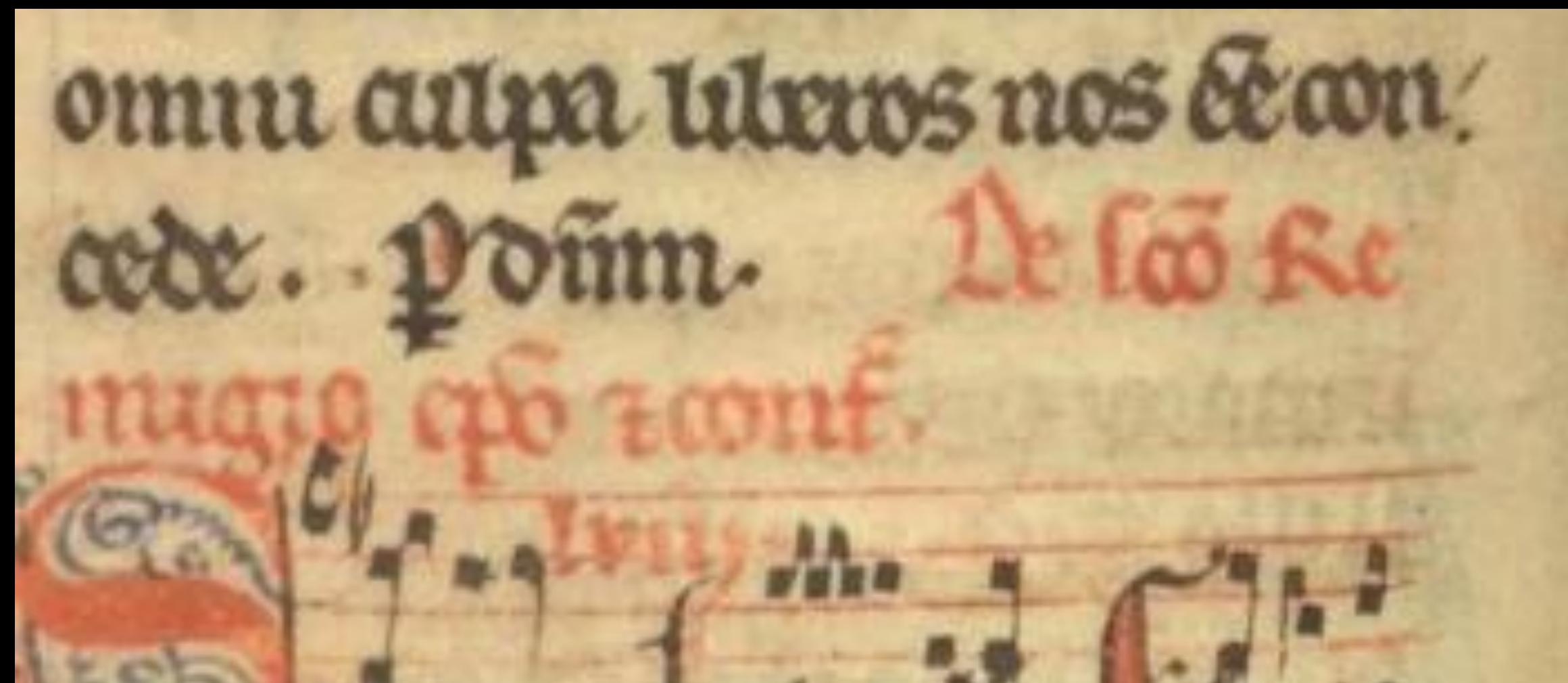
duo

▼ 0 references



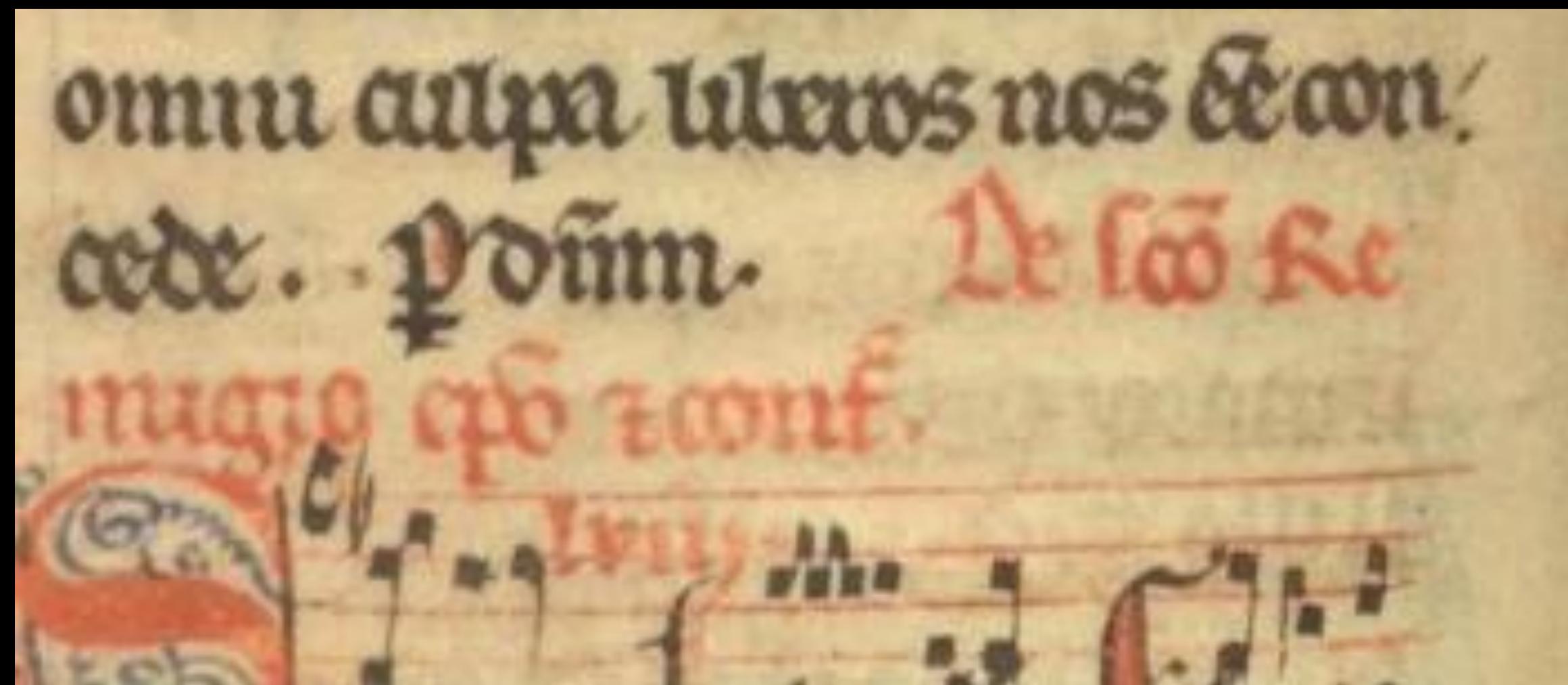
القديسان سرجيوس و باخوس

last to be solved: multiple saints



October 1: Saint Remigius

last to be solved: multiple saints



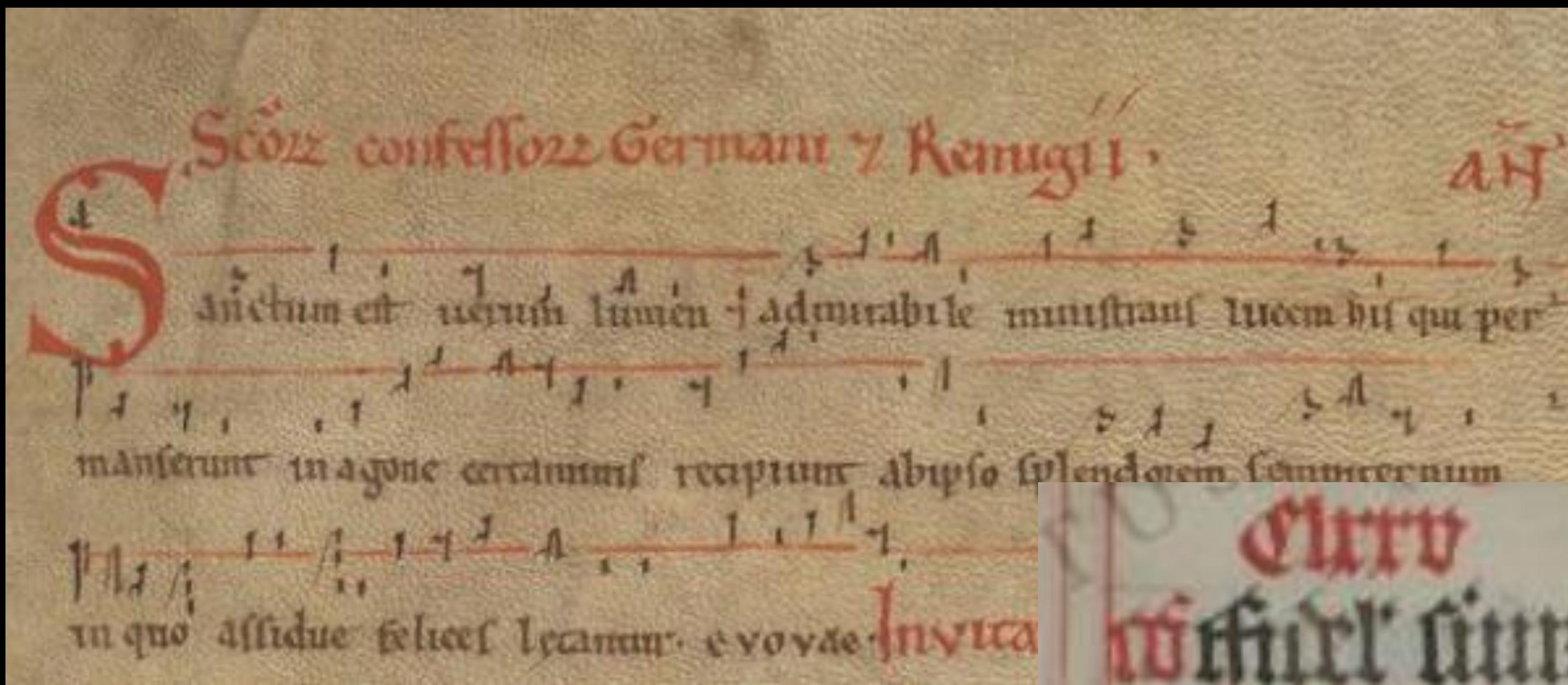
October 1: Feast of Saint Remigius

...and Saint Germanus...

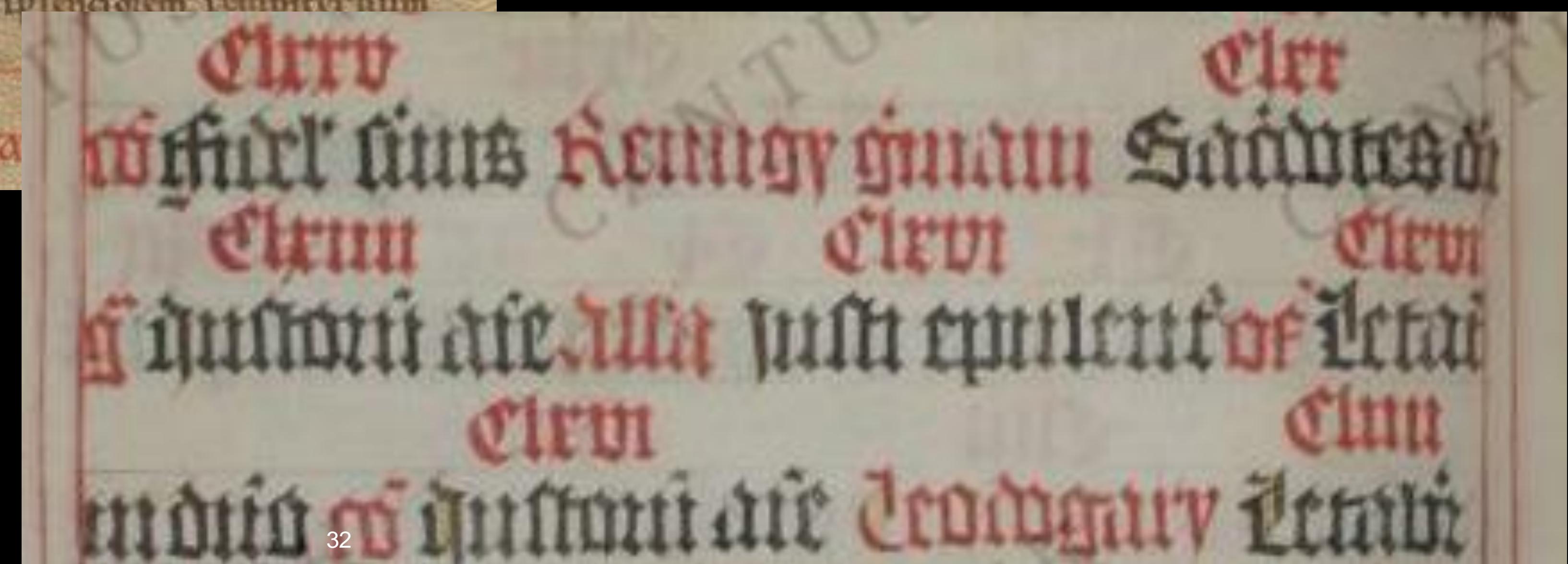


last to be solved: multiple saints

October 1: Feast of Saint Remigius



...sometimes “Germanus
AND Remigius”



last to be solved: multiple saints

October 1: Feast of Saint Remigius

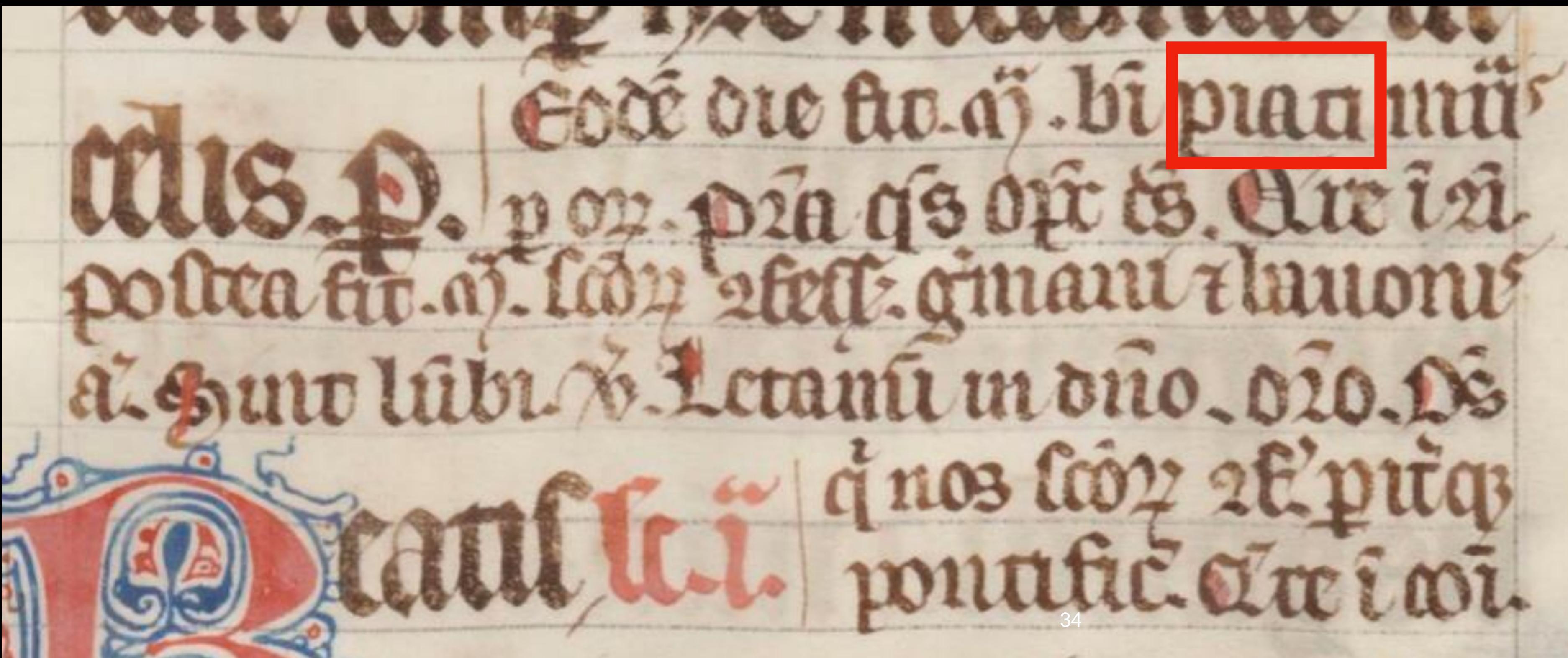
spectu domini. Ieronimi. **Iste** ē q̄ ante dñm **or.** vt de s. lu
sebio. Ab noct. oīā de uno c̄f. q̄ n̄ fuit ep̄c. **Matt.** Unē q̄nq; **W.**
a luge s. **Kemeri.** **W.** lustū dēd. **or.** vt de s. xyxto. **Vedasti.** **ebek**
Alhi. a lnc̄curta tuo. **or.** vt ināt. s. P̄cessi & mākiyimani. fēta
oīā de uno q̄f. q̄ n̄ fuit ep̄c. **Leodegarjo.** abeat. v. q̄ s. **or.** vt de s. Alaxedlo.
R. Iste cogn̄. **S**tolā 10. **E**boronā. **Matt.** **L** me q̄f. **G**oſcarb. Unē q̄nq. **ervō.** a luge ferue. **A**llor. **Matt.**
uvese. **I**cund. **ervō.** a luge dēta. **R**emigis ep̄c.
Sacerdos a p̄ontifice. **G**ramani a **Vedasti.** **A**lnc̄curta tu
Cēt om̄a d̄ uno af̄flore. **ep̄o.** **B**uodecumu N. **Iste** ē de sublimi
a p̄r. ut et. s. p̄montanū monachū p̄n
. **A**bulgebunt iusti. **ii. v̄sl.** lurauit
ope orbi. **M.** abeat us m̄ q̄ suffi

...Vedastus
might be there
too...

Remigis. ad 45. **p.** **Vedasti.** **g.** in eū. a. intercede.

last to be solved: multiple saints

...and don't forget Piatus!



last to be solved: multiple saints

October 1: Feast of Saint Remigius
and Germanus, Vedastus, Piatus, and others...

Should every combination be a different QID?

last to be solved: multiple saints

October 1: Feast of Saint Remigius
and Germanus, Vedastus, Piatus, and others...

Should every combination be a different QID?

Let f be a function Feast Day. Let \mathbb{S} denote the set of Saints.
For $s_1, s_2 \in \mathbb{S}$, is $f(s_1 + s_2) = f(s_1) + f(s_2)$?

Proof. Suppose $s_1 = Remigius$...

summary:

of ~900 “saints days” in Cantus Database:

~500 “simple” saints days now on WikiData

-work in progress on days defined by saints days (e.g. day before/Sunday after)

-solution will be similar for the ~500 Temporale feasts as well

-to do: secondary feasts (~145 items)

-to do: groups of saints (~100 items)

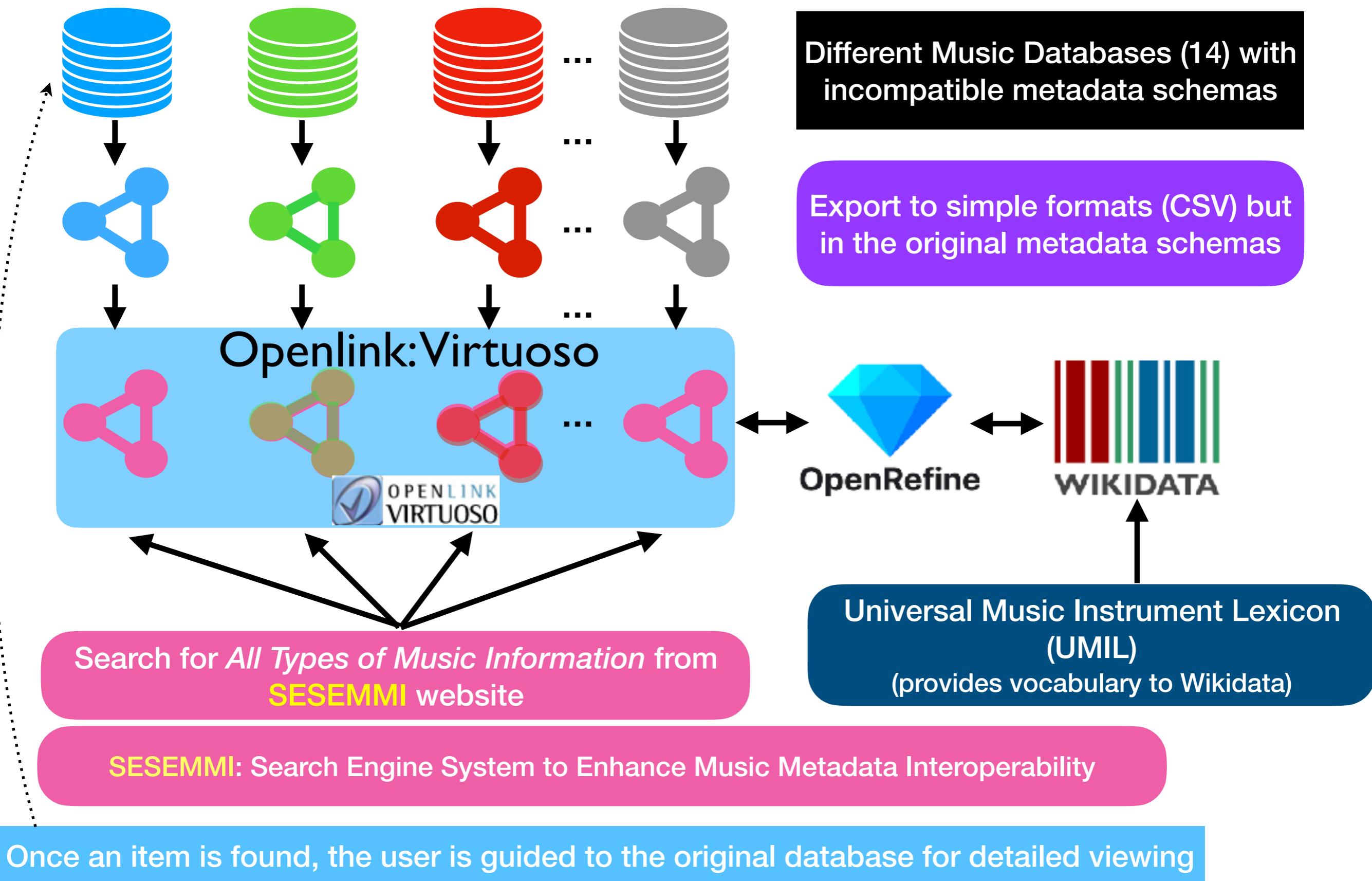
“Lazy Searching”

- ❖ Lazy Learning vs Greedy Learning
 - ❖ Lazy learning: k-nearest neighbour
 - ❖ Greedy learning: artificial neural networks
- ❖ Harvesting vs Federated Search
 - ❖ Harvesting search: Google
 - ❖ Harvested data is indexed
 - ❖ Federated search: Kayak, Google Flights, Expedia
 - ❖ Needs APIs (Application Programming Interface) on the client side (e.g., airlines and hotels)
- ❖ Lazy searching: Harvesting without indexing

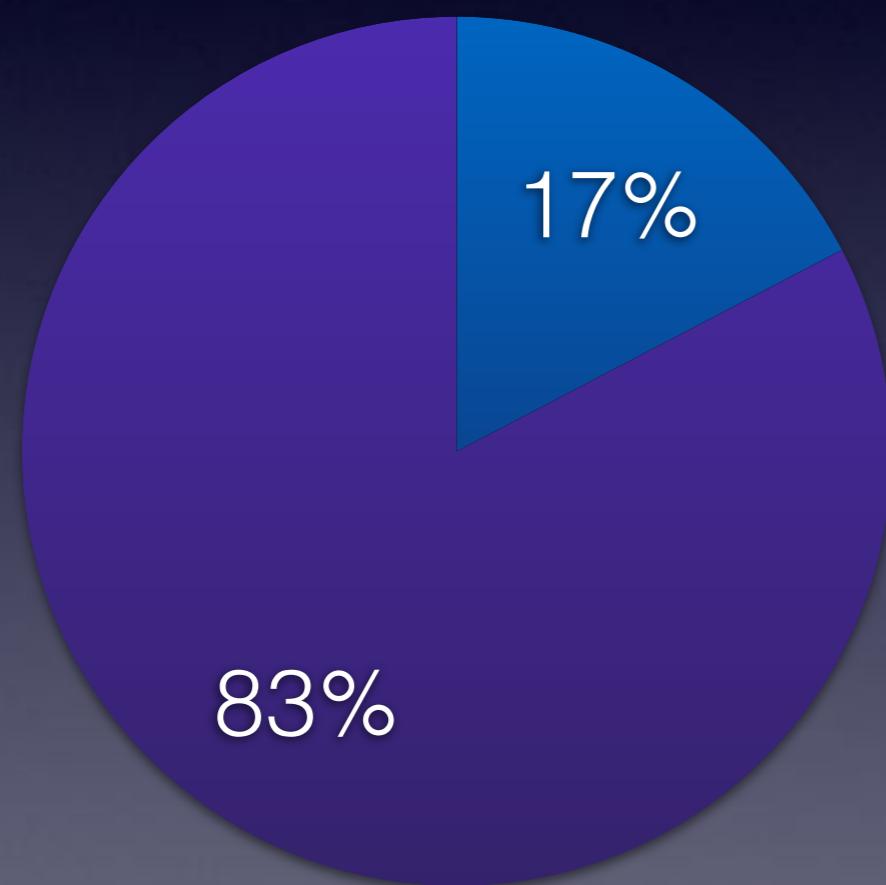
Current Basic Process

- ❖ Combine various sources and dump them into our LinkedMusic Data Lake: OpenLink Virtuoso
- ❖ Reconcile schema (properties) and vocabularies to Wikidata using OpenRefine
- ❖ Use SPARQL and other search engines (e.g., Solr, ElasticSearch) for queries

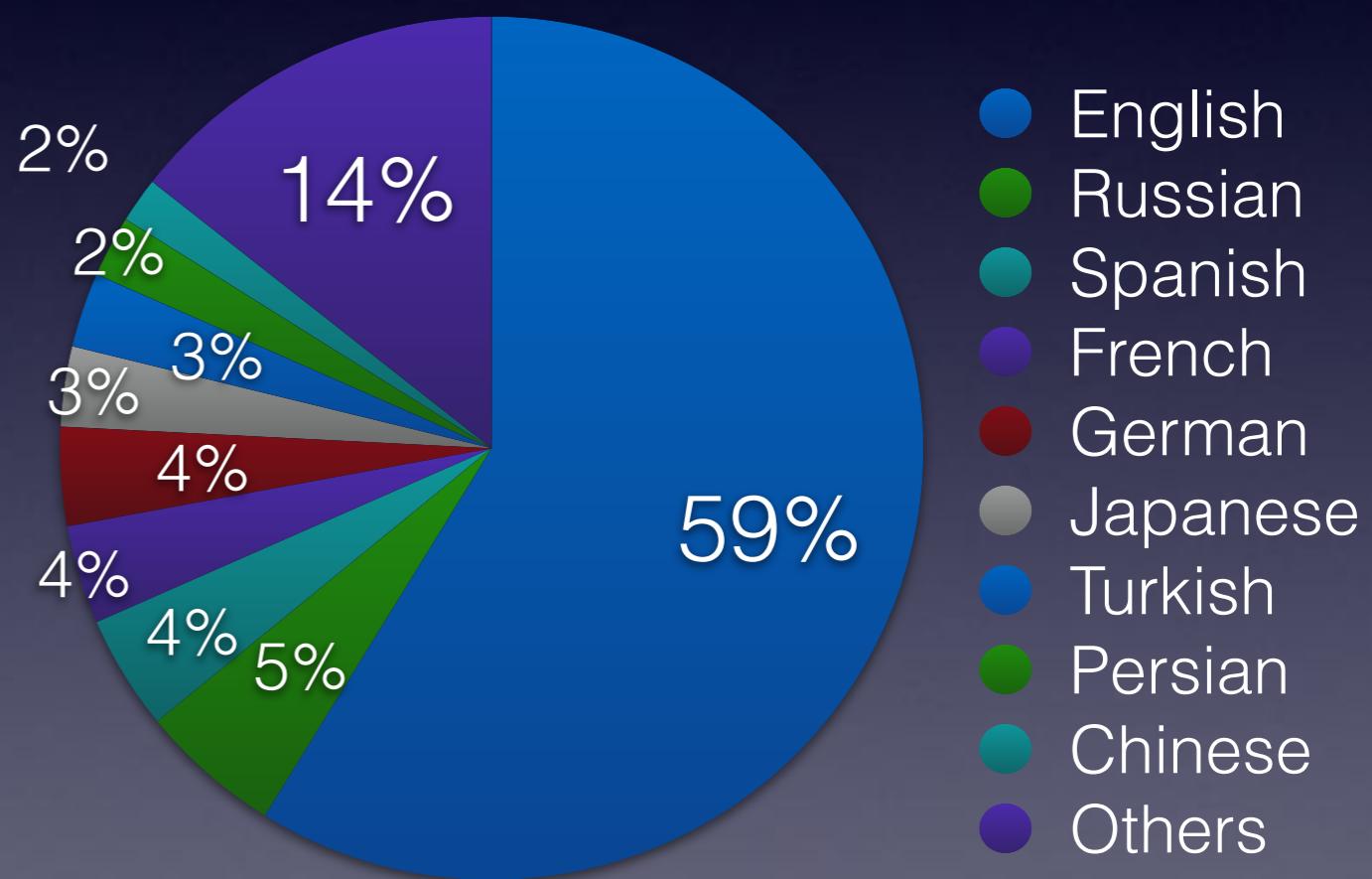
LinkedMusic Overall Process: Virtuoso Version



Who can speak English? Compared to websites



Can speak English



Websites

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

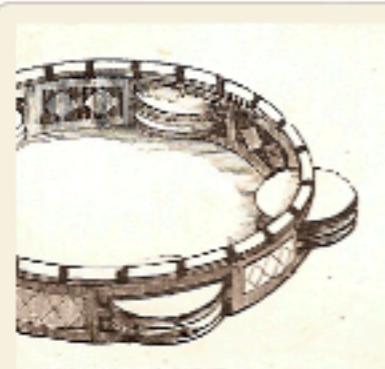
**Hornbostel-Sachs
Classification**

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


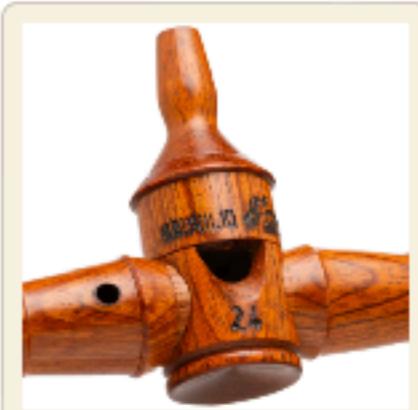
Tambourine



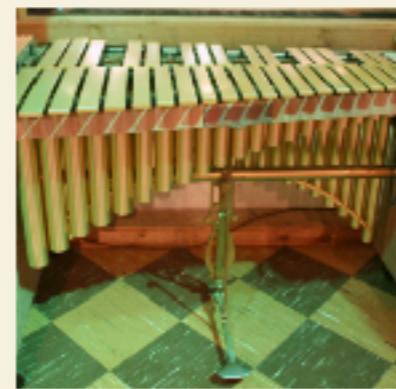
Triangle



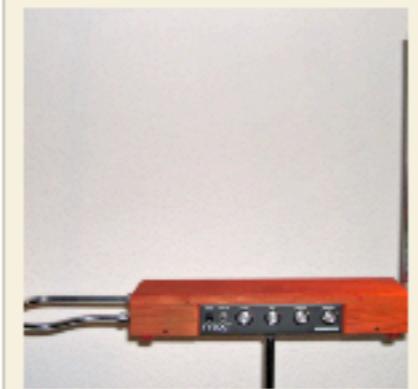
Zill



Whistle



Vibraphone



Theremin



Hurdy-Gurdy



Snare Drum



Tabla


 Transposing
Instrument


Pan Flute



Marimba

Overview of technologies used in LinkedMusic

- ❖ Linked Data
- ❖ Semantic Web
- ❖ Ontologies
- ❖ RDF
- ❖ SPARQL
- ❖ Wikidata
- ❖ OpenRefine

A new task for LinkedMusic Project

- ❖ Improve the quality of musical information in Wikidata
 - ❖ Provide interfaces for editing, e.g., UMIL, Saints' Days
 - ❖ Provide visualization interfaces to identify missing or erroneous data (e.g., “[Transposing Instrument](#)”)

Current Basic Process of Making LinkedMusic Data Lake

- ❖ Combine various data sources and dump them into our LinkedMusic Data Lake
- ❖ Use SPARQL and other search engines (e.g., Solr, ElasticSearch) for queries

NLQ2SPARQL

- ❖ Natural Language Query to SPARQL query conversion
- ❖ SPARQL: SPARQL Protocol and RDF Query Language

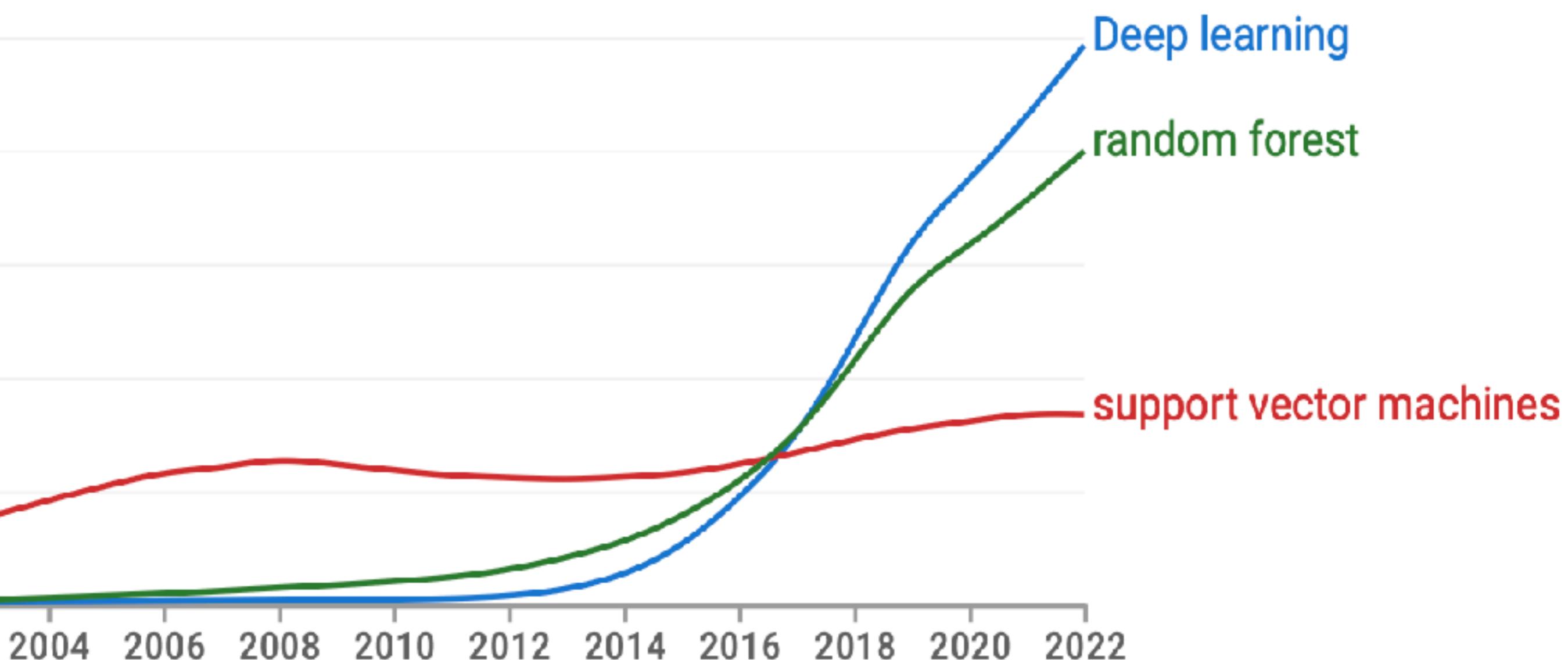
How to Train Your Computer

- ❖ Ground up
- ❖ Fine Tuning
- ❖ In-Context Learning (“In-Context Education”)
- ❖ RAGs retrieval augmented generation

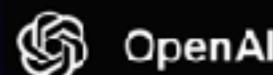
How to Train Your Computer

A Brief History of Machine Learning

Google Books Ngram Viewer



On November 30, 2022 OpenAI announces ChatGPT !



Research

Products

Safety

Company



Andrew Hankinson <andrew.hankinson@rism.digital>

andrew

December 6, 2022 at 5:06 AM

AH

Whoa.

To: Ichiro Fujinaga <ichiro.fujinaga@mcgill.ca>,

Cc: Laurent Pugin <laurent.pugin@rism.digital>

Hide

Hi Ich, Laurent,

Further to the question of LinkedMusic and doing user interfaces for finding music materials...

Have you seen ChatGPT? It's incredibly cool.

We've trained a model called ChatGPT which interacts in a conversational way. The dialogue format makes it possible for ChatGPT to answer followup questions, admit its mistakes, challenge incorrect premises, and reject inappropriate requests.

What is ChatGPT?

- ❖ ChatGPT is a type of Large Language Model (LLM)
- ❖ Designed specifically for conversation by OpenAI

What is a Large Language Model (LLM)?

- ❖ A type of artificial neural networks called *transformers*
- ❖ Initially trained for language understanding using deep learning techniques
- ❖ LLMs are pre-trained on massive datasets: text, images, music
- ❖ They have generative capabilities: text, images, music
- ❖ GPT stands for Generative Pre-trained Transformer

History of Large Language Models (LLM)

Date	Google	Date	Facebook	Anthropic
2017	Transformer paper	2023-02	LLaMA	
2018	BERT	2023-03	T	Claude
2019	T5	2023-07	LaMA 2	Claude 2
2020		2024-03	-3	Claude 3
2021	LaMDA	2024-04	LLaMA 3	
2022	PaLM	2024-06	(Nov)	Claude 3.5
2023	Bard (Mar)	2024-06	May)	Sonnet
2024	Gemini (Feb)	2024-07	LLaMA 3.2 (May)	
		2024-09	LLaMA 3.3	

Issues in training an LLM

❖ Size

- ❖ Training data

- ❖ Model size

❖ Time

- ❖ Training time

❖ Cost

- ❖ Computing time

- ❖ Environment impact

LLMs	Size	Training time	Cost
ChatGPT	175B	34 days	\$4.6M
LLaMA 3	8B/70B	3/21 Days	
LLaMA 3.1	405B	2 Months	
GPT-4	1,760B	3–4 Months	~\$500M
Gemini	1,560B		~\$1–2B

How to Train Your Computer

- ❖ ~~Ground up~~
- ❖ Fine Tuning
- ❖ In-Context Learning (“In-Context Education”)
- ❖ RAGs retrieval augmented generation

How to Train Your Computer: Fine-tuning

- ❖ A type of transfer learning
 - ❖ Start with a pre-trained model
 - ❖ Retrain with new dataset by adjusting parameters
- ❖ “Catastrophic forgetting”
- ❖ Still requires some computing resources

How to Train Your Computer

- ❖ ~~Ground up~~
- ❖ ~~Fine Tuning~~
- ❖ In-Context Learning (“In-Context Education”)
- ❖ RAGs retrieval augmented generation

How to Train Your Computer: In-Context Learning

- ❖ Prompt engineering
- ❖ Provide the context before the prompt

I'm preparing a slide presentation for academics.

Please make some slides for explaining In-Context Learning.

- ❖ No training necessary! 😊
- ❖ Has no memory, however 😔 “In-Context Education”
- ❖ Restrictions on size of the context window (token limits)
 - ❖ ChatGPT: 2,000 words
 - ❖ ChatGPT 4o: 96,000 words
 - ❖ Claude 3.1 Sonnet: 150,000 words
 - ❖ Google Gemini: 750,000 words

How to Train Your Computer

- ❖ ~~Ground up~~
- ❖ ~~Fine Tuning~~
- ❖ In-Context Learning (“In-Context Education”) 
- ❖ RAGs retrieval augmented generation

How to Train Your Computer: **RAG (Retrieval Augmented Generation)**

- ❖ Combine dynamic search with LLM
- ❖ Documents or other data, including databases, are retrieved then used as an additional context for LLM prompt
- ❖ User preferences can also be retrieved for personalized response
- ❖ Future research avenue (e.g., Knowledge Graph Embeddings)

LLMs and LinkedMusic

- ❖ LLM (ChatGPT) is used for NLQ2PARQL with in-context learning
 - ❖ Demo by Junjun Cao
- ❖ LinkedMusic may become unnecessary as LLMs improve over the next few years

Comparing queries: Google vs SPARQL

- ❖ Google: “17th-century German composers who died in London”
- ❖ SPARQL: “17th-century German composers who died in London”
- ❖ ChatGPT3: 17th-century German composers who died in London
- ❖ ChatGPT: “Generate SPARQL for Wikidata: 17th-century German composers who died in London”
- ❖ With Handel: Wikidata (2023-10-21)
- ❖ With Handel: Wikidata (2024-06-29)
- ❖ ChatGPT4o: 17th-century German composers who died in London
- ❖ Claude: 17th-century German composers who died in London

Screenshot 2024-10-01

Claude 3.5 Sonnet

Claude Q 17th-Century German Composers in London ▾ ☆ ≡

IF • 17th-century German composers who died in London

Here are some notable 17th-century German composers who died in London:

1. Johann Christoph Pepusch (1667-1752)
 - Born in Berlin, died in London
 - Known for arranging music for The Beggar's Opera
2. George Frideric Handel (1685-1759)
 - Born in Halle, died in London
 - One of the most famous Baroque composers
 - Wrote many operas, oratorios including Messiah
3. Johann Christian Bach (1735-1782)
 - Born in Leipzig, died in London
 - Youngest son of Johann Sebastian Bach
 - Known as the "London Bach"

Reply to Claude... edit camera

□ Claude 3.5 Sonnet



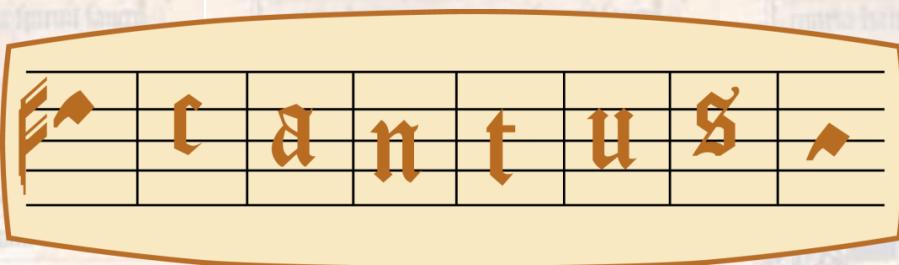
linkedMusic

Project Meeting III 2024

Cantus Database

LinkedMusic Project Meeting – 2024/10/26

Debra Lacoste, Jennifer Bain, Ichiro Fujinaga, Anna de Bakker, Andrew Hankinson, Dylan Hillerbrand, Lucas March



DALHOUSIE
UNIVERSITY



Schulich School of Music
École de musique Schulich



DIGITAL ANALYSIS OF CHANT TRANSMISSION



Social Sciences and Humanities
Research Council of Canada

Conseil de recherches en
sciences humaines du Canada



Digital Research
Alliance of Canada

Alliance de recherche
numérique du Canada

DDMAL

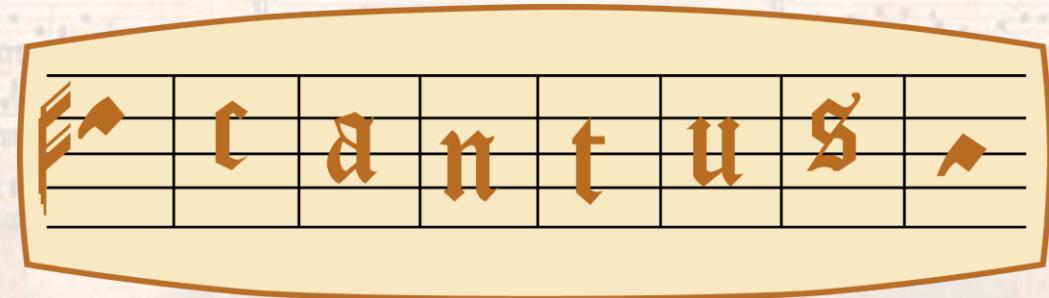
DISTRIBUTED DIGITAL MUSIC
ARCHIVES & LIBRARIES LAB



Canada

Overview

- Ansible migration
- RISM
- Problem description and goals
- Model changes
- Institutions and Source identification
- Display changes
- Future work



Ansible

- Deployment of CantusDB servers are managed by ansible.
- Allows control of the state of VM's hosting CDB ("managed nodes") from local machine ("control node") according to instructions contained in "playbooks".
- Allows you to easily replicate the state of VMs.



What is RISM?

- **RISM (Répertoire International des Sources Musicales)**: Global organization documenting musical sources
- **Purpose**: Catalogs what musical sources exist and where they're located
- Assigns an abbreviation, called a library **siglum**, to institutions worldwide that hold musical sources



Goals

1. Standardize identification fields for clear distinctions among siglum, title, and shelfmark
2. Align with RISM standards for compatibility and future integration with external catalogs
3. Improve database model structure
4. Improve user experience with better filtering and sorting

Old Source Naming Structure

Title: Composed of the city, archive, and shelfmark

Siglum: A combination of the institution's siglum and shelfmark

RismSiglum: Points to a separate RismSiglum model, which is intended to represent the institutional siglum (often referred to as the “RISM” siglum)

Old Source Naming Structure

- **Problems:**

- The **RISM sigla** used in both the source's siglum and the RismSiglum field are often inaccurate.
- The fields **siglum**, **RismSiglum**, and **title** can be modified independently, leading to inconsistencies.
- There is no linkage between sources in CantusDB and external catalogs (e.g., RISM, the institution's own record).

New Source Naming Structure

- Get rid of the RismSiglum table and associated fields -- they are not accurate
- Create an **Institution** model that contains information (name, siglum, city) on the institutions that hold our sources
- Create an **InstitutionIdentifier** model that contains various ID's for the same institution.
 - There might be multiple records in Institution Identifier model that point to the Bodleian (e.g., both [BNF cb11930727f](#) and [RISM institutions/30001627](#))
- Create a **SourcelIdentifier** model: same as InstitutionIdentifier but at the source-level.
 - So that sources can be searched by old identifiers (e.g., “CDN-WatON D-0uhiz”)

Institution Model

- Where (RISM) siglum and institution name are now found
- Source identifier (shelfmark) replaces the former title and siglum fields on the Source model

	NAME	RISM SIGLUM	CITY	COUNTRY	IS PRIVATE COLLECTOR	IS PRIVATE COLLECTOR
<input type="checkbox"/>	State Library of Victoria	AUS-Msl	Melbourne	Australia	✖	✖
<input type="checkbox"/>	State Library New South Wales	AUS-Ssl	Sydney	Australia	✖	✖
<input type="checkbox"/>	University of Sydney - Fisher Library - Rare Books & Special Collections	AUS-Sfl	Sydney	Australia	✖	✖
<input type="checkbox"/>	Universitätsbibliothek	A-Gu	Graz	Austria	✖	✖
<input type="checkbox"/>	Augustiner-Chorherrenstift - Bibliothek	A-KN	Klosterneuburg	Austria	✖	✖
<input type="checkbox"/>	Benediktiner-Stift Kremsmünster - Musikarchiv	A-KR	Kremsmünster	Austria	✖	✖

McGill University - Rare Books and Special Collections - Manuscript Collection (CDN-Mlr)
Montréal, Canada

[View this institution in RISM Online](#) [View this institution in Wikidata](#)

Cantus Database

Sources

[MS Medieval 0092](#)
[MS Medieval 0215](#)
[MS Medieval 0239 early layer \("0239 early layer"\)](#)
[MS Medieval 0212](#)
[MS Medieval 0211](#)

Institution Identifier

- ID's for various occurrences of the same institution in different libraries

Change institution identifier

wkp:Q62535993

HISTORY

Identifier:	RISM Online VIAF <input checked="" type="checkbox"/> Wikidata	
Identifier type:	GND (Gemeinsame Normdatei) Bibliothèque nationale de France Library of Congress Digital Image Archive of Medieval Music	
Institution:	- Rare Books and Special Collections - Manuscript Collection (CDN-Mir)	
Created by:	-	
Last updated by:	-	
Date created:	Oct. 24, 2024, 8:42 p.m. The date this entry was created	
Date updated:	Oct. 24, 2024, 8:42 p.m. The date this entry was updated	

Chant Project

- Chants can be tagged with the Project if their inventories are collected as part of a particular project or initiative.
- Allows for the collection of project-specific chant data.
- Enables filtering by project during search.

Search

Action: ----- Go 0 of 2 selected

PROJECT

Benedicamus Domino

Clavis Sequentiarum

2 projects

Printed Books vs. Manuscripts

- Confusion between the **holding institution** and the **publisher** for printed books.
- Add a field (checkbox or dropdown) to identify whether a source is a **printed book** or a **manuscript**.
- Rename “institution” to “**holding institution**” in the display to clarify the source location.

The screenshot shows a user interface for managing source fragments. At the top, there is a dropdown menu labeled "Complete Source/Fragment" with options "Complete source" and "Incomplete source". Below this is a section titled "Manuscript/Printed" containing two buttons: "Manuscript" (which is checked) and "Printed". At the bottom, there is a field labeled "Cursus" with a dropdown arrow. The background of the interface features a faint watermark of a historical manuscript page.

Complete Source/Fragment:

Complete source

Manuscript/Printed:

✓ Manuscript

Printed

Cursus:

1541"

Source model changes

- New fields:
 - **source_completeness** (replaces full_source boolean):
 - Complete source
 - Fragment
 - Reconstruction
 - **production_method** (notes whether a source is manuscript or printed)
 - **name** (A colloquial or commonly-used name for the source)
- Shelfmark is now a required field and gets a default value of “[No Shelfmark].”
- Heading / short heading property is a “Cantus Siglum” to display source information

The screenshot shows a 'Create Source' form with the following fields and their descriptions:

- Holding Institution:** Halifax (Canada), St. Mary's Univers... (with a blue arrow pointing to it)
- Shelfmark:** [No Shelfmark] (highlighted with a red oval)
- Name:** (highlighted with a red oval)
- Segments:** CANTUS Database, Bower Sequence Database (checkboxes)
- Provenance (origin / history):** If the origin is unknown, select a location where the source was used later in its lifetime and provide details in the "Provenance notes" field.
- Provenance notes:** Complete Source/Fragment: Complete source (highlighted with a red oval)
- Century:** (dropdown menu)
- Manuscript/Printed:** Manuscript (highlighted with a red oval)
- Date:** (text input)
- Cursus:** (dropdown menu)
- Date of the source, if known (e.g. "1541"):** (text input)

Private Collections / Virtual Reconstrucitons

- We don't assign RISM sigla to privately owned fragments.
- Institution model has new fields that identify private collections and collectors.
 - `is_private_collection`
 - `is_private_collector`
- Don't assign RISM sigla to **virtual reconstructions**; instead, refer to the reconstructed versions in the individual source descriptions.

Source Naming

- Heading / short heading property is a “Cantus Siglum” to display source information
- City → holdisnt.name → shelfmark
 - If no holding institution: “Cantus” → shelfmark
- Adds “(fragment)” to display when “is a fragment” is checked in the taxonomy
- Adds “name” for colloquial or commonly used names
- Short Heading uses siglum

Heading

London (ON), University of Western Ontario - Archives and Research Collections Centre (ARCC), M2150 (fragment)

[View](#) | [Edit](#)

Cantus Siglum

CDN-Lu M2150 (fragment) ←

Short heading

Holding Institution

London (ON), University of Western Ontario - Archives and Research Collections Centre (ARCC)
(CDN-Lu)

Manuscript/Printed

Manuscript

Summary

Sixteenth-century “mutilated” antiphoner, possibly from the Burgos monastery, Spain. Square notation on five-line black staves. Secular cursus. 97 extant folios. Nineteen lacunae.(Imperfect: folios 1-26, 32-36, 40-41, 45-46, 63, 67-68, 86, 89, 93-108, 116-117, 130-131, 136, 138-139, 148, 154-162, 170-171, 174, 181-192? wanting.) 55.6 x 38.7 cm.

Source List

- Add “Country” and “City + Holding Institution” columns
- Columns for “country”, “city + institution”, and “source” are now **sortable** on the source list page
- **Filters** for Country and Complete Source/Fragmetn available

Browse Sources

Displaying 1-2 of 2 sources

Segment: CANTUS Database | General search (siglum, city, description ..) | Indexing Notes

Country: Finland | Provenance (origin/history): - Any - | Century: - Any - | Complete Source/Fragmetn: - Any -

Incipit/CantusID:

Country	City + Holding Institution	Cantus Siglum	Summary	Date/Origin	Image Link	Chants / Melodies
Finland	Helsinki, National Library	FIN-Hy F.m.I.26 (fragment)	Fragmented noted missal, 12th century. 24 folios. Unknown origin. The fragments of the manuscript have been preserved a...	12th century (1175-1200)	Images	253 / 160
Finland	Helsinki, National Library	FIN-Hy F.m.I.85 (fragment)	Twenty-two leaves from a notated missal, datable to the last quarter of the thirteenth or first quarter of the fourte...	13th century (1275-1300)	Images	150 / 93

1
Page 1 of 1

Optimization

- Feast Detail with raw SQL queries to improve performance
 - Load time from 3+ seconds to 250ms
 - Reducing the number of queries (from 1,500 to 8)
 - Improving the display of chant genres
- Many other pages similarly optimized

Future Work

- Prints vs. manuscripts
- Fixing institution names
- Create new institutions, only that exist in RISM
- Migrating previous identifiers
- Some other high priority issues:
 - Set up log persistence
 - CSV upload
 - Proofreading changes

Thank you!



linkedMusic

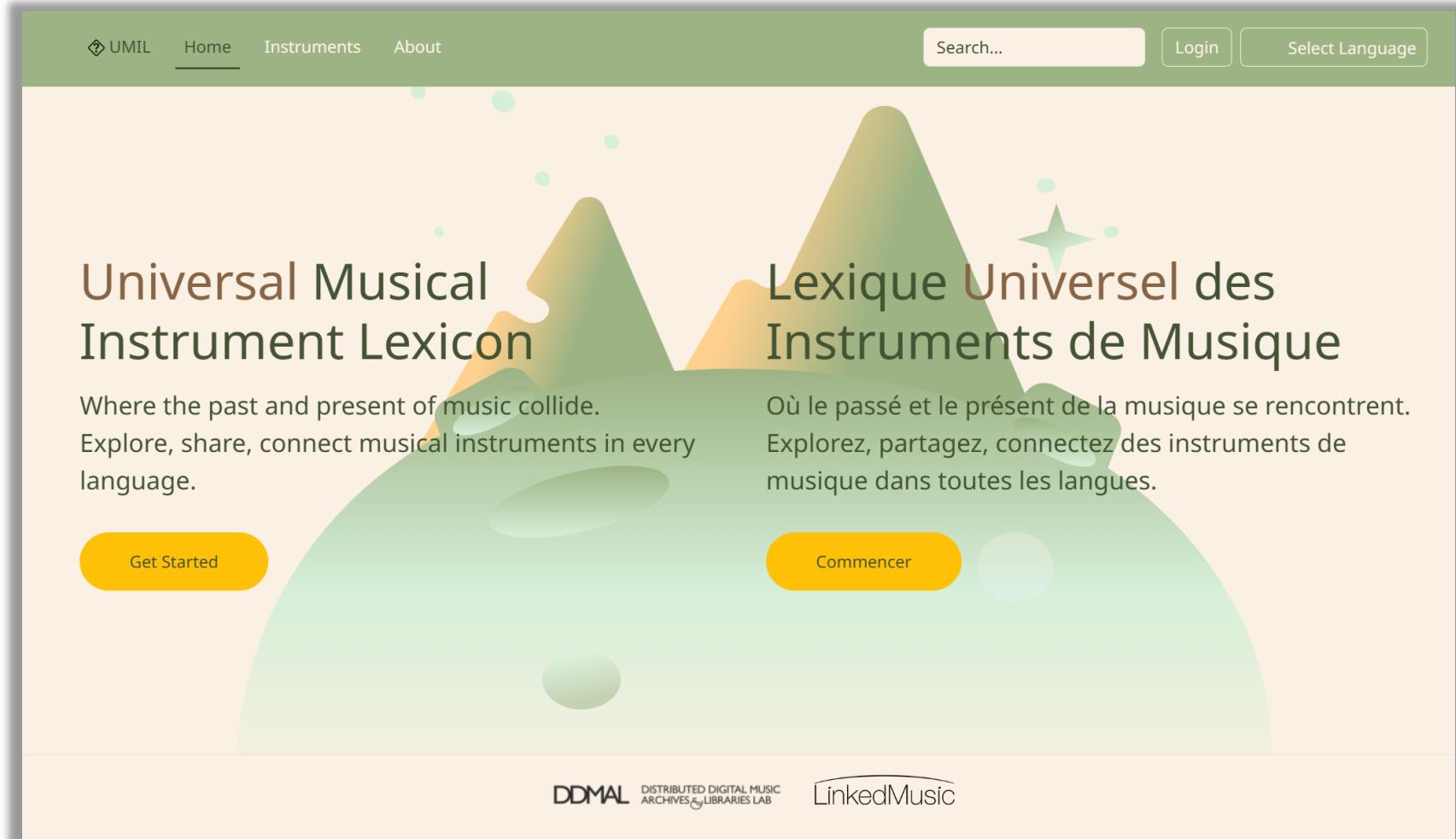
Universal Musical Instrument Lexicon

LinkedMusic Project Meeting

Kun Fang

Overview

- **Crowdsources names and images of instruments**
- **Focus on multilingual interface**
- **More convenient to edit instruments on Wikidata**



Why do we need UMIL?

- **Addressing Issues with Wikidata**
 - **Data Quality:** Incomplete or inaccurate instrument data
 - *UMIL ensures cleaner, well-organized data through community contributions.*
 - **Convenience:** lacks a user-friendly way to view/edit all instrument information at once
 - *UMIL offers a centralized platform where users can easily view, edit, and submit instrument data.*

Why do we need UMIL?

- **UMIL's Key Functions**
 - **Crowdsourcing platform** where users can contribute
 - Contribute new instruments
 - Submit instrument names in different languages
 - Upload instrument images
 - **Interacts with Wikidata**
 - Get information **from** Wikidata
 - Publish information **to** Wikidata

Display Instruments in UMIL

800+ Musical Instruments

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

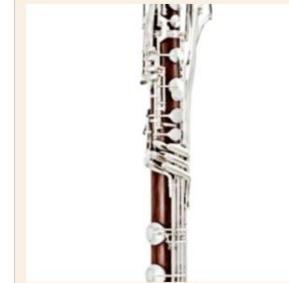
Showing 1 to 20 of 831

English ▾

English
Français

View 20 | 50 | 100 per page

Multi-languages

 Guitar	 Trumpet	 Clarinet	 Bass Clarinet
 Bagpipes	 Trombone	 Violin	 Cello

Hornbostel-Sachs Classification Faceted Search

2024/10

26

Instrument Item: View Mode

- Link to the Wikidata page

INSTRUMENT LIST English ▾

Showing 1 to 20 of 881 entries

Guitar

View on Wikidata

True

English English

WIKIDATA

Main page Community portal Project chat Create a new Item Recent changes Random Item Query Service Nearby Help Donate Lexicographical data Create a new Lexeme Recent changes Random Lexeme Tools What links here Related changes Special pages

Item Discussion

guitar (Q6607)

fretted string instrument

▼ In more languages Configure

Language	Label	Description
English	guitar	fretted string instrument
Chinese	吉他	撥弦樂器
Chinese (Taiwan)	吉他	撥弦樂器
Chinese (China)	吉他	乐器

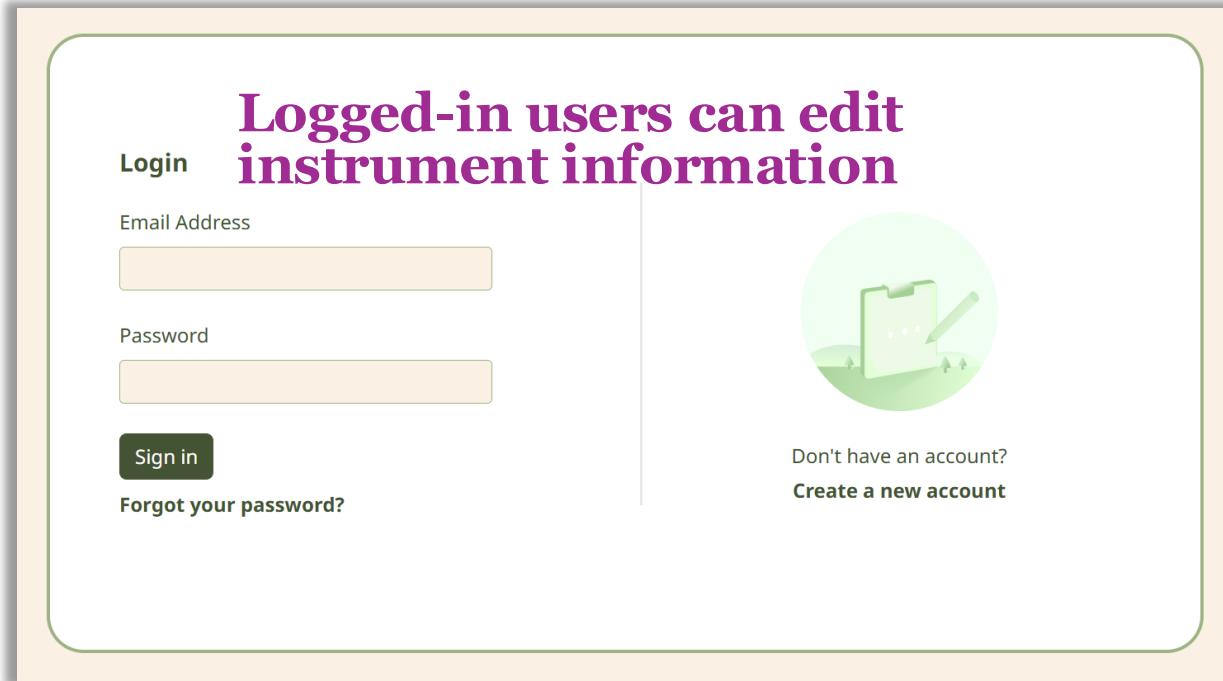
All entered languages

Statements

instance of type of musical instrument

0 references

Instrument Item: Edit Mode



- **Add instrument names**
- **Add image (ongoing)**
- **Create new instruments (ongoing)**

The image shows a list of instruments. At the top, it says "INSTRUMENT LIST" and "English". Below that, it says "Showing 1 to 20 of 881 entries". A pink arrow points from the "Add instrument names" and "Upload new images" buttons on the left to these same buttons in the larger image above. The "Guitar" entry is shown with its image and a "View on Wikidata" button.

Add New Name: Edit Mode

Input fields: Language*, Name*, Source*, Description, Alias

INSTRUMENT LIST English ▾

Add New Name for **Guitar** (wikidata ID:Q6607)

Language	Name	Source	Description	Also known as
Type to search	Enter name	Enter source	Enter description	Enter alias
Type to search	Enter name	Enter source	Enter description	Enter alias

Add another row

Do you want to publish also to Wikidata?

Publish

Add more names at once

Check this box to publish to Wikidata

Add New Name: Language Supported

- Choose a language in the list (600+ languages)
- Identified by Wikidata's language code (e.g., "en", "fr")
- Get the language list by calling Wikidata API

The screenshot shows a user interface for managing item names. A dropdown menu is open, listing various language codes and their corresponding names. The menu is highlighted with a pink border. The list includes:

- aae Arbëresh - Arbëresh
- ab аԥшәа - Abkhazian
- abs bahasa ambon - Ambonese Ma...
- ace Acéh - Acehnese
- acf Kwéyòl Sent Lisi - Saint Lucian...
- acm Iraqi Arabic - عراقي
- ady адыгабзэ - Adyghe
- ady-cyrл адыгабзэ - Adyghe (Cyrillic scri...
- aeb Tùnsi - Tunisian Arabic / تونسي
- aeb-arab Tunisian Arabic (Arabic... - تونسي
- aeb-latn Tùnsi - Tunisian Arabic (Latin sc...
- af Afrikaans - Afrikaans
- agq Aghem - Aghem
- ak Akan - Akan

The main interface shows a table with columns for ID, Description, and Also known as. There are also sections for Description and Also known as with input fields and "Remove" buttons. At the bottom, there are images of a Trombone, Violin, and Cello, along with a "Publish" button.

Add New Name: Interact with Wikidata

- **Where do we use the *Wikidata API* when adding new names?**
 - Get Wikidata's latest *language list*
 - *Get/Set* instrument information
 - Name
 - Description
 - Alias

Add New Image (Ongoing)

- **Objective:** Uploaded images will appear directly on **the instrument's Wikidata page**.
- **How it works:**
 - **Step 1:** Upload image to **Wikimedia Commons**
 - *Wikimedia Commons is a free media repository where images, videos, and audio are stored and shared.*
 - **Step 2:** Retrieve the **title of the uploaded image**.
 - **Step 3:** Link the image title to the instrument's “**image**” **property** (P18).



Challenge 1: Make Edits to Wikidata

- Requires a Wikidata account of a specific type: **Auto-confirmed user** to publish content.
- **Solution**
 - Create a DDMAL account; Contributions made through UMIL will be published using this account.

The screenshot shows a section of the Wikipedia 'User rights' page. It includes two main sections: 'New users' and 'Autoconfirmed users'. The 'Autoconfirmed users' section is highlighted with a pink border. The text in both sections describes the editing permissions and requirements for each user type.

New users [edit]

Any logged-in user may maintain a watchlist at [Special:Watchlist](#), and may email any other user (unless blocked from doing so) by using [Special:EmailUser](#). While they no longer have the editing rate limit, they must still answer a CAPTCHA prompt when adding new external links until they become autoconfirmed or confirmed.

Autoconfirmed users [edit]

An account with at least 4 days of age and at least 50 edits (90 days and 100 edits for editors from Tor nodes; abuse filter can potentially delay or revert automatic promotion) is said to be **autoconfirmed**. Any user account which does not meet these requirements can be granted the **confirmed** permission and receive the same permissions. Autoconfirmed users no longer need to respond to CAPTCHAs for most actions, can edit **semi-protected** pages, and can move most pages (except items and properties, which are simply renamed by changing the appropriate labels instead).

Challenge 2: Upload Images to Wikimedia

- Copyright issues for uploading images.
- **Solution**
 - Follow [Wikimedia's copyright policy](#) by asking users to complete a [questionnaire](#)
 - Review content before publishing
 - **Human review:** Conducted by the UMIL team
 - **Automatic check tools:** Further investigate existing mechanisms for checking/filtering inappropriate content



Call for better solutions!

Future Plans

- **Fuzzy search:** Improve musical instrument search for better user experience.
- **More faceted search:** Allow users to filter instruments based on other classification methods.
- **Expand instrument database:** More instruments and more languages.

Thank you!



linkedMusic

Upgrading the Liber Usualis

F. COLE THIERRIN

Background

- ▶ Collection of 11th century Gregorian chants and hymns
- ▶ Online searchable copy of the Liber Usualis:
<https://liber.simssa.ca/>
- ▶ Version used published in 1961
 - ▶ Editorial decisions
- ▶ Liber Usualis was encoded using MEI 2011

Search the Liber Usualis

What is this? Find out more about what we are trying to do.

defgecd Transposed pitch sequence ▾ Search Clear

previous Result 1 of 9 for defgecd next

Go to page Go
Current page: 466 of 2340

320 1st Sunday of Advent

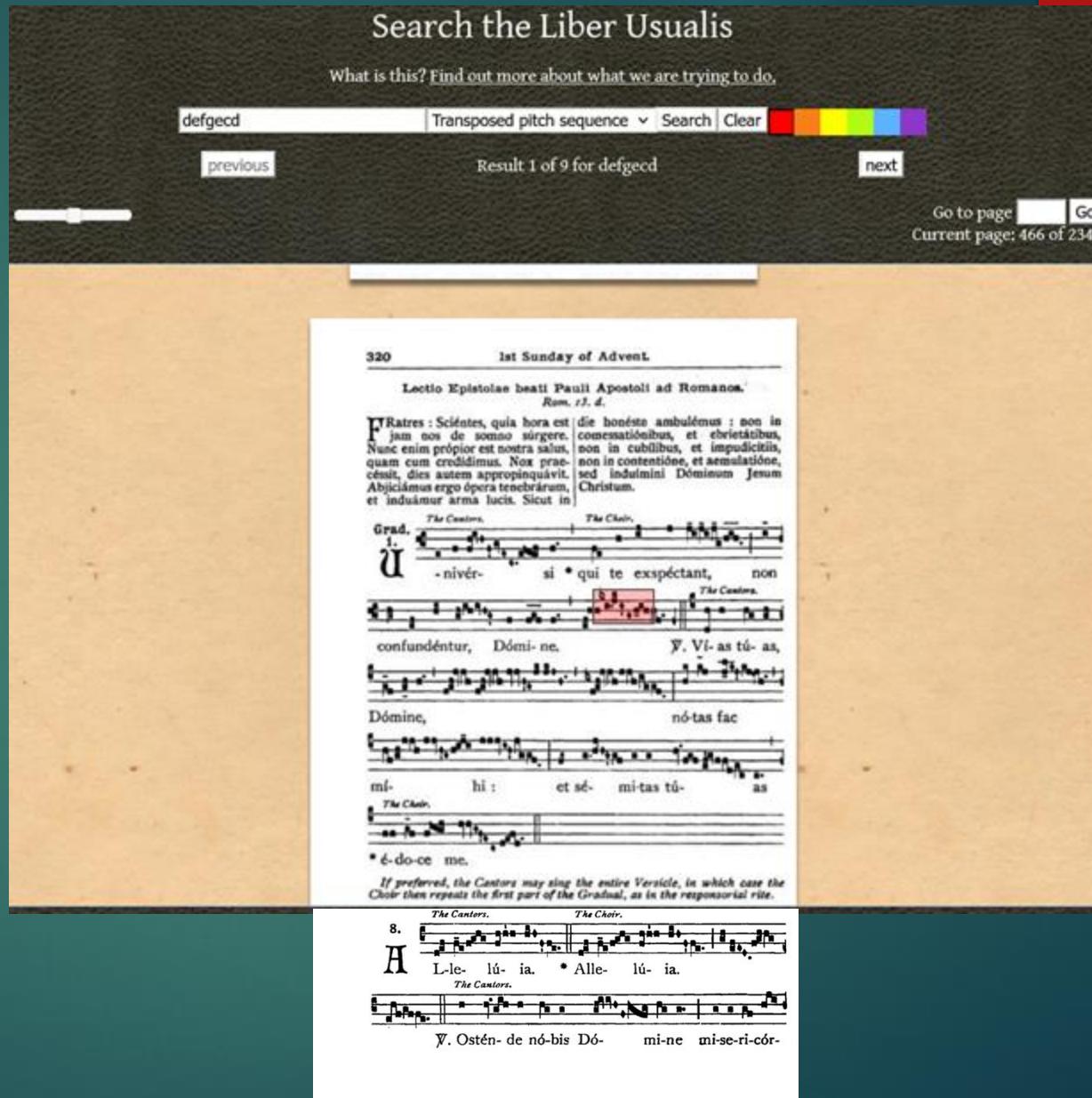
Lectio Epistole beati Pauli Apostoli ad Romanos.
Rom. 13. d.

Fratres : Scientes, quia hora est die honesta ambulamus : non in somno sūrgere, non in cubilibus, et impudicitib⁹, non in contentionē, et aemulacionē, sed induimini Dōminum Iesum Christum. Abiciamus ergo opera tenebrarū, et induamur arma lucis. Sicut in

Grad. *The Cantors.* *The Choir.*
U-nivér-si- qui te exspectant, non
confundéntur, Dómi-ne. *Y. Vi-as tú-as,*
Dómine, nót-as fac
mí-hi : et sé-mi-tas tú-as
The Choir.
* é-do-ce me.

If preferred, the Cantors may sing the entire Versicle, in which case the Choir then repeats the first part of the Gradual, as in the responsorial rite.

8. *The Cantors.* *The Choir.*
A-L-le-lú-ia. * Alle-lú-ia.
The Cantors.
Y. Ostén-de nō-bis Dó-mi-ne mi-se-ri-cór-



Challenges and Issues

MEI File Location



liber.simssa.ca

custom.simssa.ca

neon.simssa.ca

salzinnes.simssa.ca

cloud.simssa.ca

cress.simssa.ca

www.musiclibs.net

rodan2.simssa.ca

db.simssa.ca



Neume Names

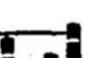
			
Punctum	Virga	Bivirga	Punctum inclinatum <i>(Diagonal)</i>
			
Podatus or Pes	Clivis or Flexa	Epiphonus	Cephalicus
			
Scandicus	Salicus	Climacus	Ancus
			
Torculus	Porrectus	Torculus resupinus	Porrectus flexus
			
Pes subpunctis	Scandicus subpunctis	Scandicus flexus	Climacus resupinus
			
Strophicus	Pes strophicus	Clivis strophica or Clivis with an Oriscus	Torculus strophicus or Torculus with an Oriscus
			Trigon
Pressus	Other Pressus or apposed neums		

Image Size Issues

MEI Editor
Page: 0009
Scale: 1.00

**PREFACE TO THE VATICAN EDITION
OF THE ROMAN CHANT.**

The place of honour in this Solemn Edition of the Vatican Official text is given to the Vatican Preface... Its wise counsels and general Principles of Interpretation are embodied; elucidated and enlarged upon in the Rules given further on.

Holy Mother the Church has received from God the charge of training the souls of her faithful; all honour and for this noble end has ever made a happy use of the help of the sacred Liturgy. Wherein — in order that men's minds may not be Sundered by differences, but that, to the contrary, the unity which gives virtue and beauty to the mystical body of Christ might flourish unimpeded — has been, and still is, to keep the traditions of our forefathers, ever trying frugally to bring our dear body to restore any which might have been forgotten in the course of the ages.

Now among those things which most nearly touch the sacred Liturgy, being as it were interwoven therein and giving it splendour and impressiveness, the first place must be assigned to the Sacred Chant. We have, indeed, learnt from certain Fathers that it gives a singularly great upliftment to the soul, and in wondrous wise, to heavenly things. Wherefore the Church has never ceased to recommend the use of the Chant, and has striven with the greatest assiduity and diligence to prevent its decline from its pristine dignity.

To this end liturgical music must possess those characteristics which make it proper, sacred and adapted to the good of souls. It must surely accomplish above all else the ends of driving out sin and at the same time be able to express pleasantly and truly the sentiments of the Christian soul. It must also be catholic, answering to the needs of every people, country and age, and combine simplicity with artistic perfection.

All these characteristics, however, are nowhere to be found in a higher degree than in Gregorian Chant, — the special Chant of the Roman Church, who has received it alone by inheritance from the Fathers, has kept it carefully throughout the ages in her records, and commends it to the faithful as her own, ordering its exclusive use in certain parts of the Liturgy. (Motu Proprio, Nov. 28, 1903, n. 3.)

Certainly in the course of time the Gregorian Chant incurred no small loss of purity. This was chiefly because the special rules of the Chant, as traditionally received from the Fathers, were either negligently overlooked or allowed to be altogether forgotten. Hence arose an evident decline in the spirit which is spoken of as "Gregorian", and the "spirit of prayer", while at the same time the boldness and grace of the sacred melodies, if they did not wholly disappear, were certainly affected for the worse.

But the Sovereign Pontiff, Pius X. — may his enterprise be crowned with good fortune and success! — emulating herein the zealous endeavours of his predecessors, determined and took measures to prevent any further decadence in the Gregorian Chant. When the Motu Proprio, issued on November 28, 1903, he expressly and clearly laid down the principles of the very first step of reform (whereon the ecclesiastical Chant is based and whereby it is controlled; he gathered together at the same time the principal regulations of the Church against the various abuses which had crept into the Chant in the

Line 1: preface to the satan edition
 Line 2: f t raman chant
 Line 3: th he m***e s**,*,, ea*** th vh**, oeil b
 Line 4: o e w d vhw*, p en m o m*,wn a ee,eh f *or** r
 Line 5: **eph*, q e aba a*oag*a a *,u)m *v** he aau e***
 Line 6: whee
 Line 7: nun vow u cedo ***m fox o u dow a ***g u
 Line 8: u o am w**** m o **u* a he *** fum
 Line 9: a e we u ,*u n*,h) v s**,, shot o r
 Line 10: ray m w**wu b 41***6 f we u wh**v e w **d,
 Line 11: m m d **a, e two a owl ve ***a ***m**
 Line 12: duas hm***** w use e w m *** wwujeo ow w vv
 Line 13: paean b lew g **uw ,*** o, ve we th e)***
 Line 14: o wow u m
 Line 15: x ***m jig m o *** hahn o we ,**u lawn ^v*
 Line 16: b a th t **** w*** we**l oy*d** b qhy**,*g* lot
 Line 17: *p*** we a p *,n* jaw 4*** o* ha eggs u ***
 Line 18: w *hhh** we how joe mee e o th lucent wasp
 Line 19: w wo***d e th caw wuf qn,mww e adeo **d*w
 Line 20: u lh1** v*** a a*** a a mqae auto
 Line 21: t d **m**,* *** who *** o*od*w o a a
 Line 22: y***ab jaw j sum w e we wu n *** how **y***
 Line 23: iuwo a w we u exam ll **e,v md h e m th m u

OCR Work

MEI Editor

Page: 0009

Scale: 1.00

**PREFACE TO THE VATICAN EDITION
OF THE ROMAN CHANT.**

The place of honour in this Solemnes Edition of the Vatican Official text is given to the Vatican Preface. Its wise counsels and general Principles of interpretation are embodied, elucidated and enlarged upon in the Rules given further on.

Holy Mother the Church has received from God the charge of training the souls of the faithful in all holiness, and for this noble end has ever made a happy use of the help of the sacred Liturgy. Wherein — in order that men's minds may not be sundered by differences, but that, on the contrary, the unity which gives vigour and beauty to the mystical body of Christ might flourish unimpaired — she has been zealous to keep the traditions of our forefathers, ever trying diligently to discover and boldly to restore any which might have been forgotten in the course of the ages.

Now among those things which most nearly touch the sacred Liturgy, being as it were interwoven therein and giving it splendour and impressiveness, the first place must be assigned to the Sacred Chant. We have, indeed, all learnt from experience that it gives a certain breadth to divine worship and uplifts the mind in wondrous wise to heavenly things. Wherefore the Church has never ceased to recommend the use of the Chant, and has striven with the greatest assiduity and diligence to prevent its decline from its pristine dignity.

To this end liturgical music must possess those characteristics which make it preeminently sacred and adapted to the good of souls. It must surely emphasise above all else the dignity of divine worship, and at the same time be able to express pleasantly and truly the sentiments of the Christian soul. It must also be catholic, answering to the needs of every people, country and age, and combine simplicity with artistic perfection.

All these characteristics, however, are nowhere to be found in a higher degree than in Gregorian Chant — the special Chant of the Roman Church, who has received it alone by inheritance from the Fathers, has kept it carefully throughout the ages in her records, and commands it to the faithful as her own, ordering its exclusive use in certain parts of the Liturgy. (Motu Proprio. Nov. 22. 1903. n. 3.)

Line 1: preface to the satan edition

Line 2: f t raman chant

Line 3: th he m***e s**,*,, ea*** th vh**, oeil b

Line 4: o e w d vhw*, p en m o m*,wn a ee,eh f *or** r

Line 5: **eph*, q e aba a*oaq*a a *,u)m *v** he aau e***

Line 6: whee

Line 7: nun vow u cedo ***m fox o u dow a ***g u

Line 8: u o am w h**** m o **u* a he *** fum

Line 9: a e we u ,,*u n*,h) v s**,, shot o r

Line 10: ray m w**wu b 41***6 f we u wh**v e w **d,

Line 11: m m d **a, e two a owl ve ***a ***m**

Line 12: duas hm***** w use e w m *** wwwjeo ow w ww

Line 13: paean w lew g **uw ,*** **o, ve we th e)***

Line 14: o wow u m

Line 15: x ***m jig m o *** hahn o we,**u l awn *v*1

Line 16: b a th t **** w*** we**1 oy*d** b qhy**,g* lot

Line 17: *p*** we a p *,n* jaw 4***o* ha eqqs u ***

Status

- ▶ Files all but converted*
- ▶ “Lost” MEI files on Github
- ▶ Images resized, uploaded to Arbutus

Future: OCR error correction

46

MEI Editor

Page: 0123

Scale: 1.00

Hi, an OCR got the following result from reading a text: "speeches me,*domim hyssopo, et unda". Can you try to guess what the original text was, given that it's from a book of gregorian chants?

T

Hello! Given the context of a book of Gregorian chants, the OCR result "speeches me,*domim hyssopo, et unda" is likely a distorted version of a Latin text. Here's a possible correction:

"Asperges me, Domine, hyssopo, et mundabor."

bor : lavá- bis me, et super nívem de- albá- bor.

Ps. 50. Mi- se-ré-re mé- i, Dé- us, * se-cúndum mágnam mi-

se-ricórdi- am tú- am. Gló- ri- a Pátri, et Flí-li- o, et

Spi-ri-tu- i Sáncto. * Sic-ut é- rat in princi-pi-o, et nunc,

Line 10: et semper et in saecula saecula rump amend

Line 11: repeat the ant aperies men

Line 12: the frost sunday of the patron re gloria patris a not said ci

Line 13: on the second sunday of the iuuion, or palm today the as

Thank You



linkedMusic

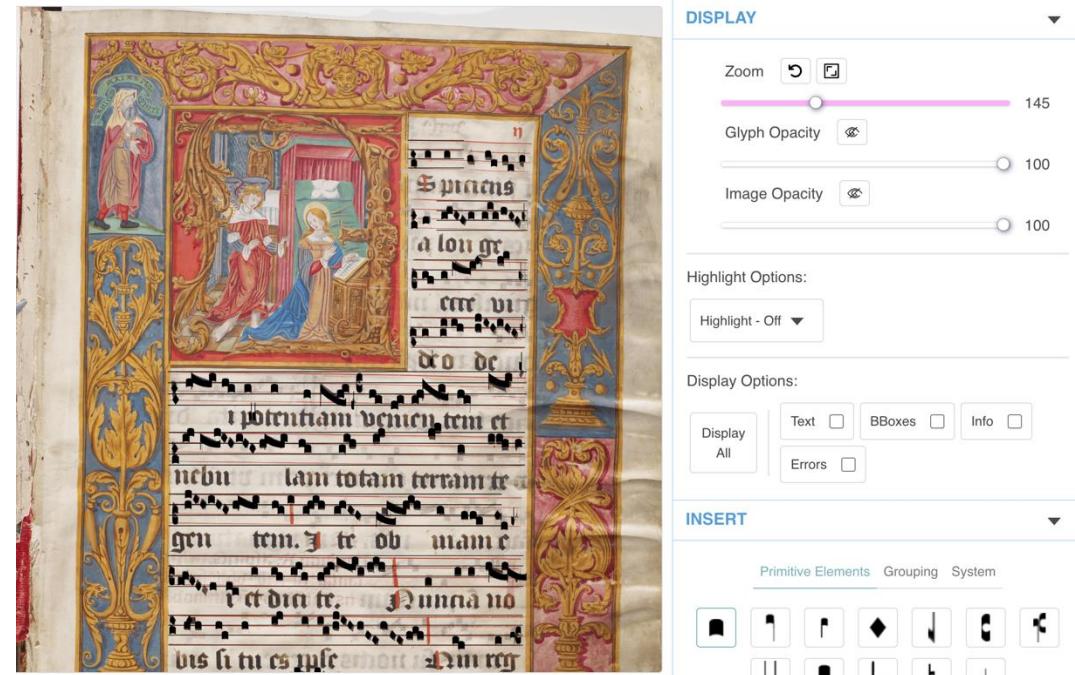
Neon: What's New?

LinkedMusic Project Meeting 2024

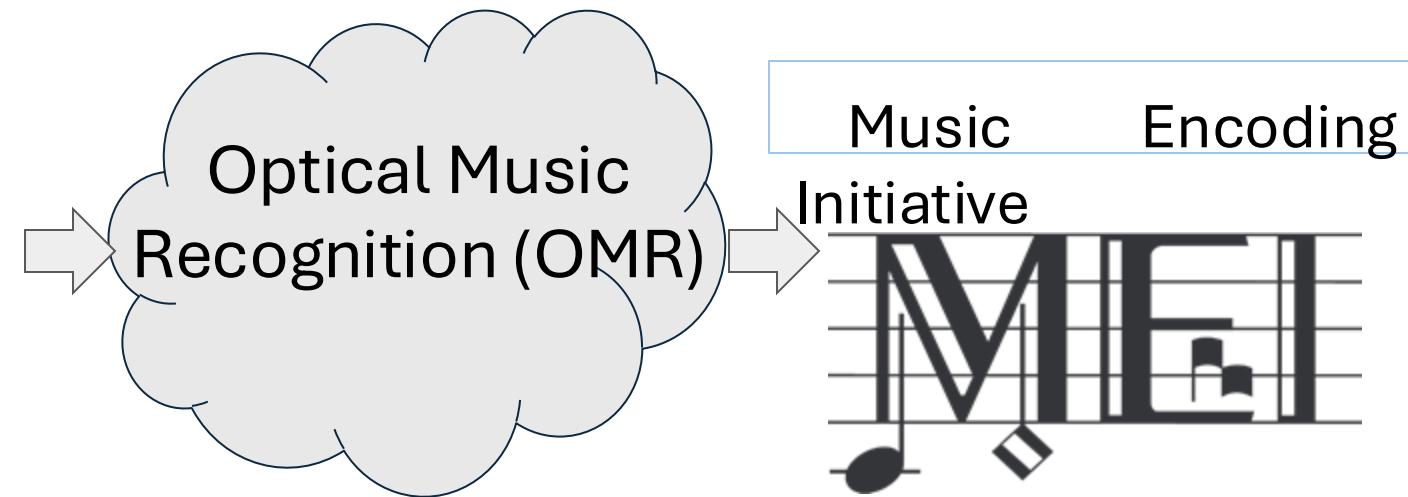
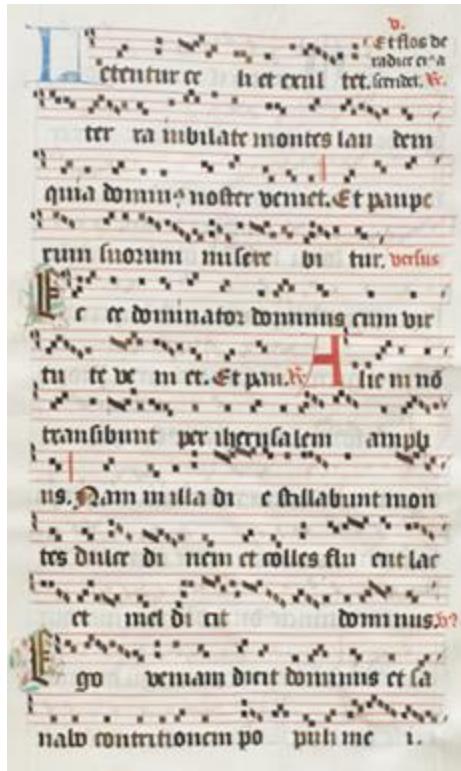
Yinan Zhou
DDMAL, McGill University
October 26th, 2024

What is Neon?

- **Neume Editor Online**
- Square notation editor



Where Neon Fits In



What Neon Does

- Web interface for graphically correcting OMR errors
- Applies changes directly to MEI
- Displays MEI-generated page above source page
- Extra information on MEI encoding (e.g., groups)

What Neon Does

- Web interface for graphically correcting OMR errors
- Applies changes directly to MEI
- Displays MEI-generated page above source page
- Extra information on MEI encoding (e.g., groups)



What Neon Does

- Web interface for graphically correcting OMR errors
- Applies changes directly to MEI
- Displays MEI-generated page above source page
- Extra information on MEI encoding (e.g., groups)

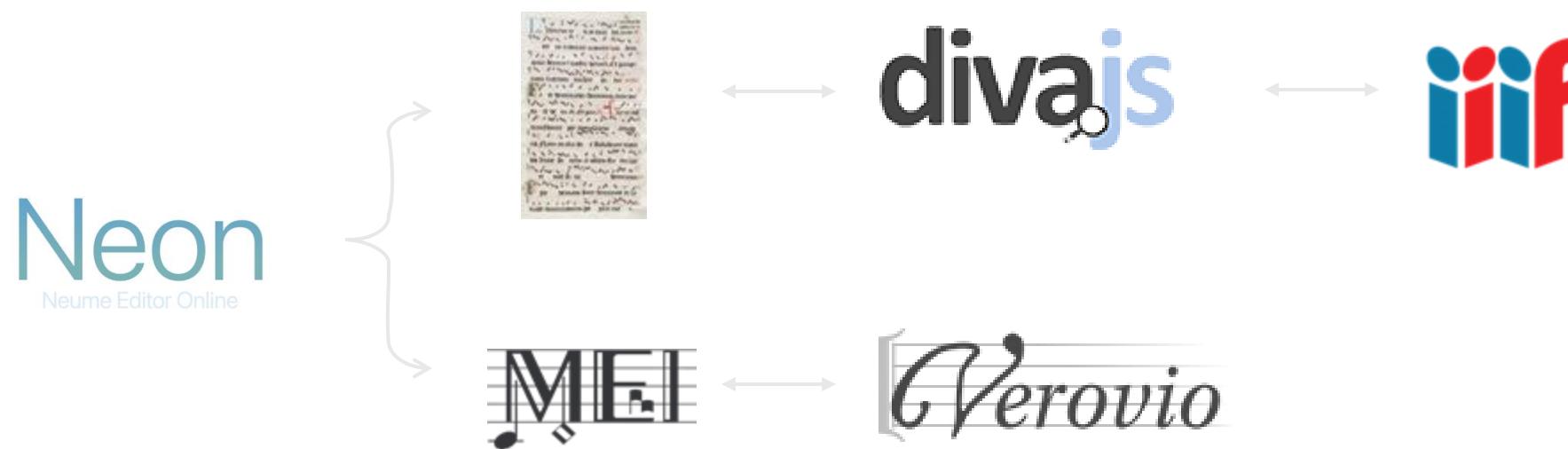


Element Info

Type: neume
Shape: PorrectusFlexus
Pitch(es): A2 G2 A2 G2

How Neon Works

1. Neon downloads the MEI and image files
2. *Diva.js* displays the images using *IIIF*
3. *Verovio* makes an SVG of the MEI
4. Neon coordinates and saves the results



History of Neon

The original Neon.js was started in 2011.

- . Developed by Gregory Burlet and Alastair Porter
- . Stable release in 2017
- . Changes made on a server



Display

Zoom

Glyph Opacity

Image Opacity

Display Text: Display Info:

Insert

Neume Groupin



Edit

Select By: Syllable Neume

Undo

Redo

Delete

Current Neon

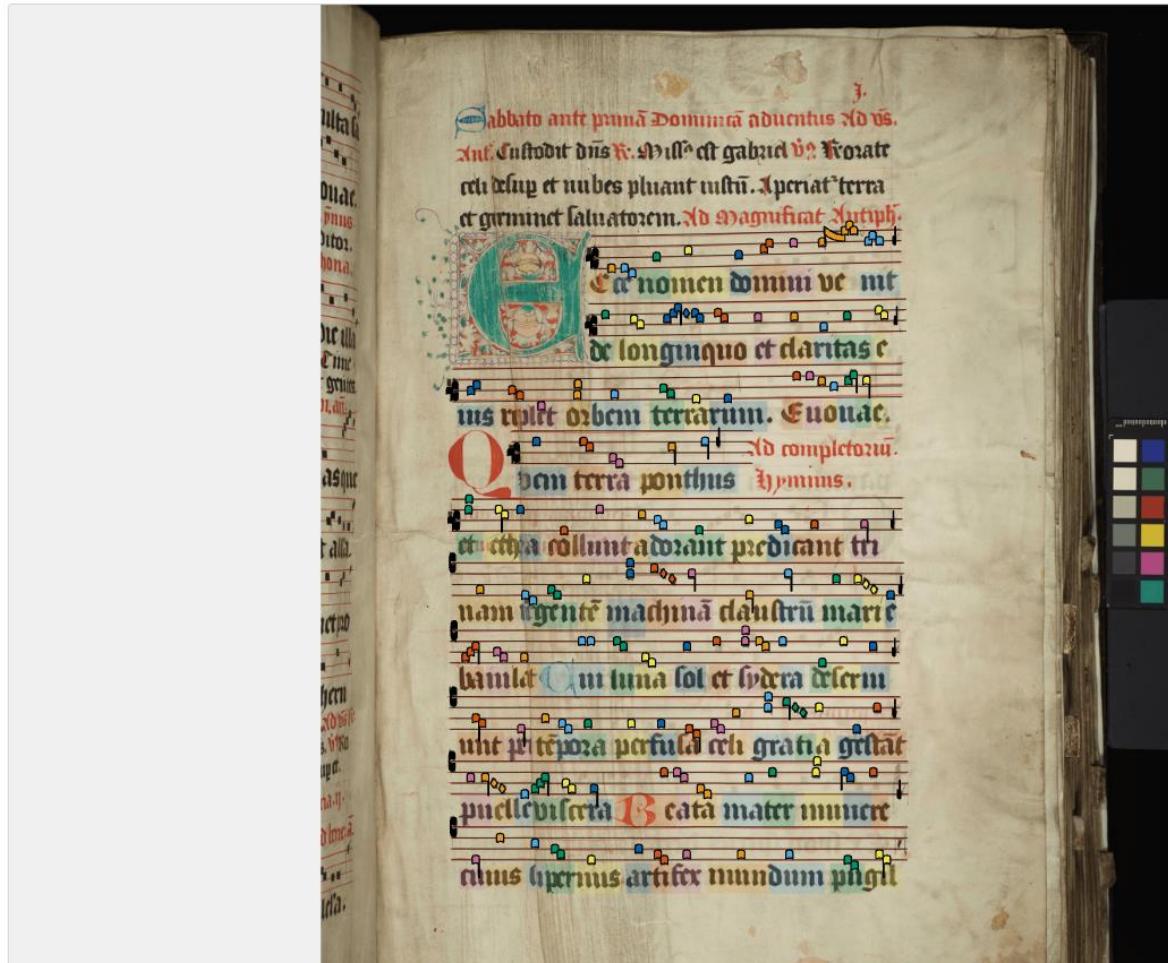
Current Neon started in 2018 by Juliette Regimbal

- . Uses tools that weren't around in 2011
- . Changes made in the browser

So what's new since 2022?

Current Neon

- . UI update in 2022
- . Brain (Verovio) transplant in 2024
 - . Currently available in Neon staging



DISPLAY

Zoom Glyph Opacity Image Opacity BBox Circle

100

100

100

25

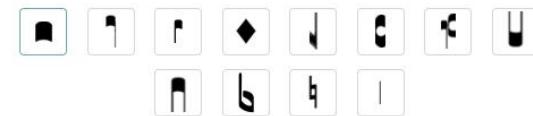
Highlight Options:

Highlight - Selection

Display Options:

INSERT

Primitive Elements Grouping System



New Features

- . New file system
- . Support new glyphs, such as liquecent and divLine
- . Clef and accidentals can be inserted into syllables
- . Octave of a clef can be displaced up/down
- . Introduce debug mode

New Features

- . Heights of all the bounding boxes within a staff can be matched with just one click
- . Adjustable bounding box circle
- . Support multiple columns
- . New MEI files can be created from scratch

Q&A?



linkedMusic

Rodan Updates

Hanwen Zhang

2024 Oct 26 Project Meeting



DISTRIBUTED DIGITAL MUSIC
ARCHIVES & LIBRARIES LAB



Digital Research
Alliance of Canada

Alliance de recherche
numérique du Canada

Rodan

<https://rodan2.simssa.ca>

- A web app that runs jobs for optical music recognition (OMR) and music analysis.
- Users can make, run, and share workflows on the website.
- A typical workflow can take scanned manuscripts as inputs and eventually output OMR results.
- Users can then upload pictures and OMR results to Neon to visualize and further analyze.



V3.0.0

Released 2023 Nov 30

Current production version

01

MEI Encoding

- No empty syllables in encoded MEI file
- Add an extra casting for MEI files to prevent from crashing
- Support multi-column folios in MEI encoding

02

UI Overhaul

- Ctrl + 0 to move the workflow to center of canvas
- Text does not overflow in resource assignment page
- New UI

03

Interactive Classifier

- Use arrow keys to navigate page or classifier glyphs
- Close the edit view when no glyphs are selected
- Hide classifier glyphs by default

04

Other Jobs

- New: Staff Distance
- Updated: Text Alignment, RGB(PNG), Gamera
- Functionality: Password reset, show error messages, etc.

No pending requests

My Projects

test train

Resources

Workflows

Workflow Runs

Run Jobs

issue 1124

Resources

Workflows

Workflow Runs

Run Jobs

[About](#) [Dev](#) [Help](#) [Preferences](#)[Logout](#)

My Projects / test train

Workflow Runs

Run Jobs

Workflows

Resources

[Create new Workflow](#)[Import Workflow from file](#)

TABLE FILTERS

Name	Creator	Created	Updated	Valid
untitled	hanwenzhang	2024-08-28 11:30:34	2024-10-24 23:30:26	true

PROJECT DETAILS

Name: test train

Description:

Creator: hanwenzhang

Created on: 2024-08-28 11:30:03

[Edit Users](#) [Save](#) [Delete](#)

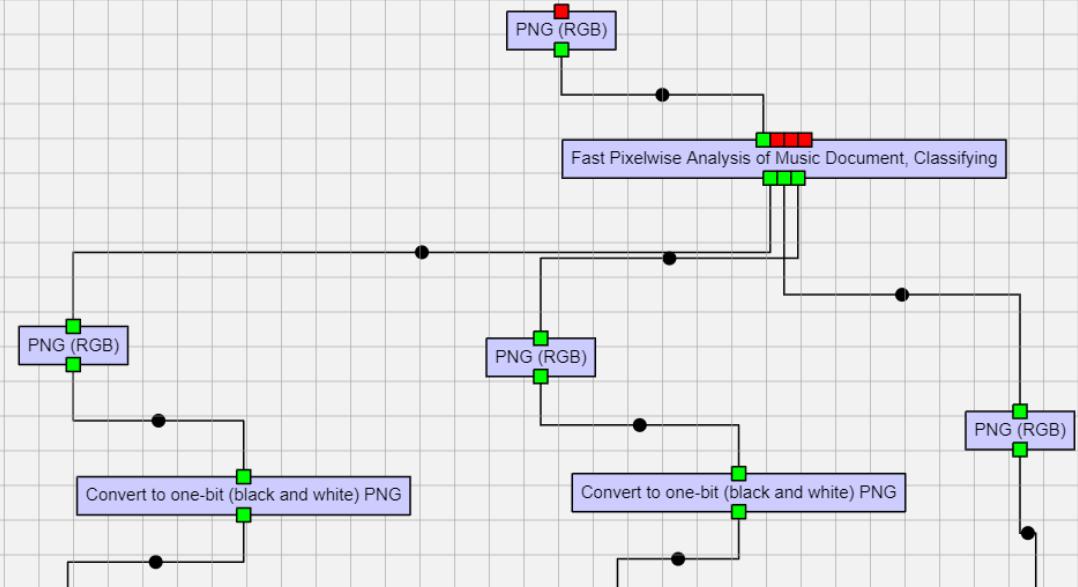
E2E full workflow homemade

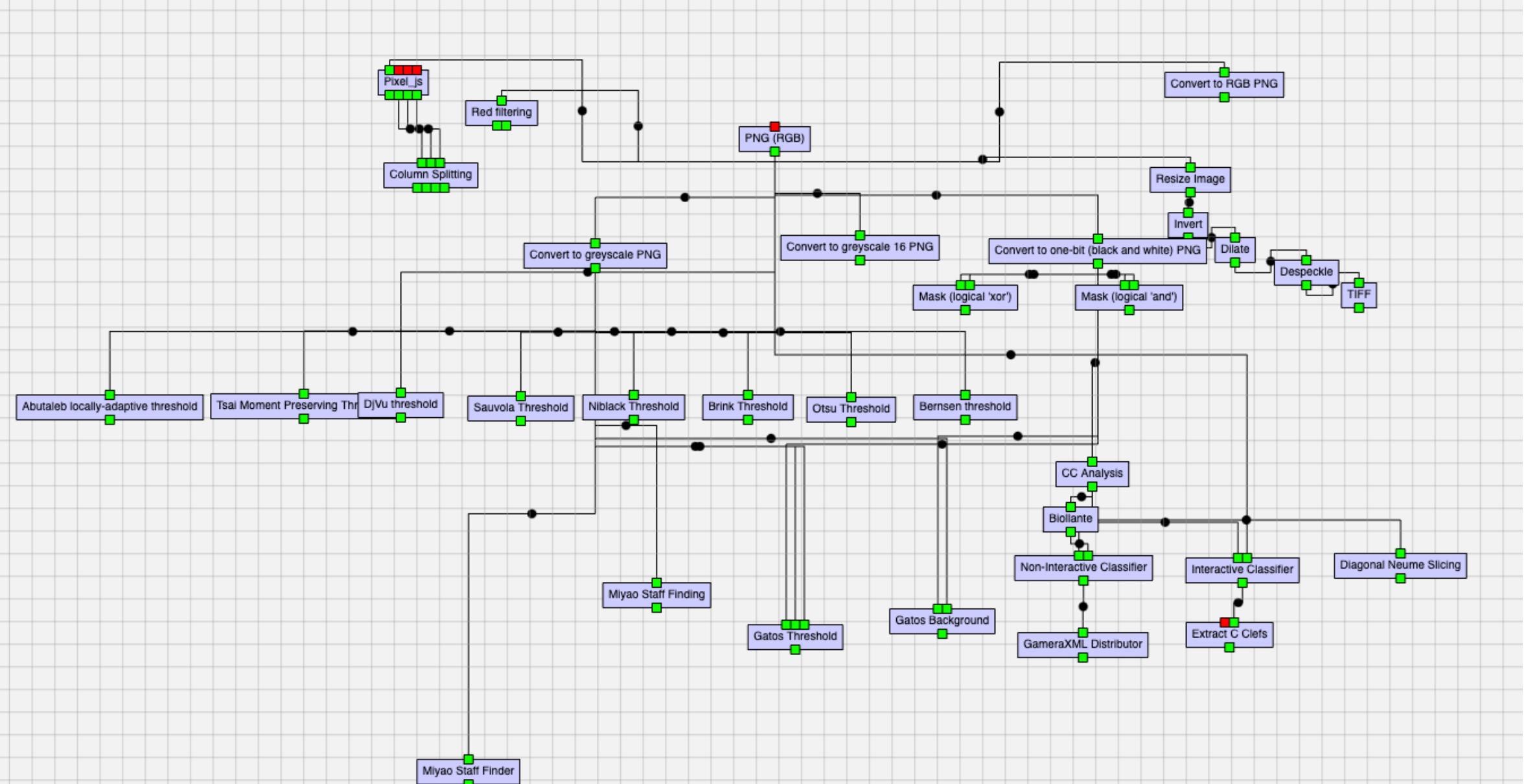
Workflow

View

Settings

Workflow "E2E full workflow homemade" is valid.





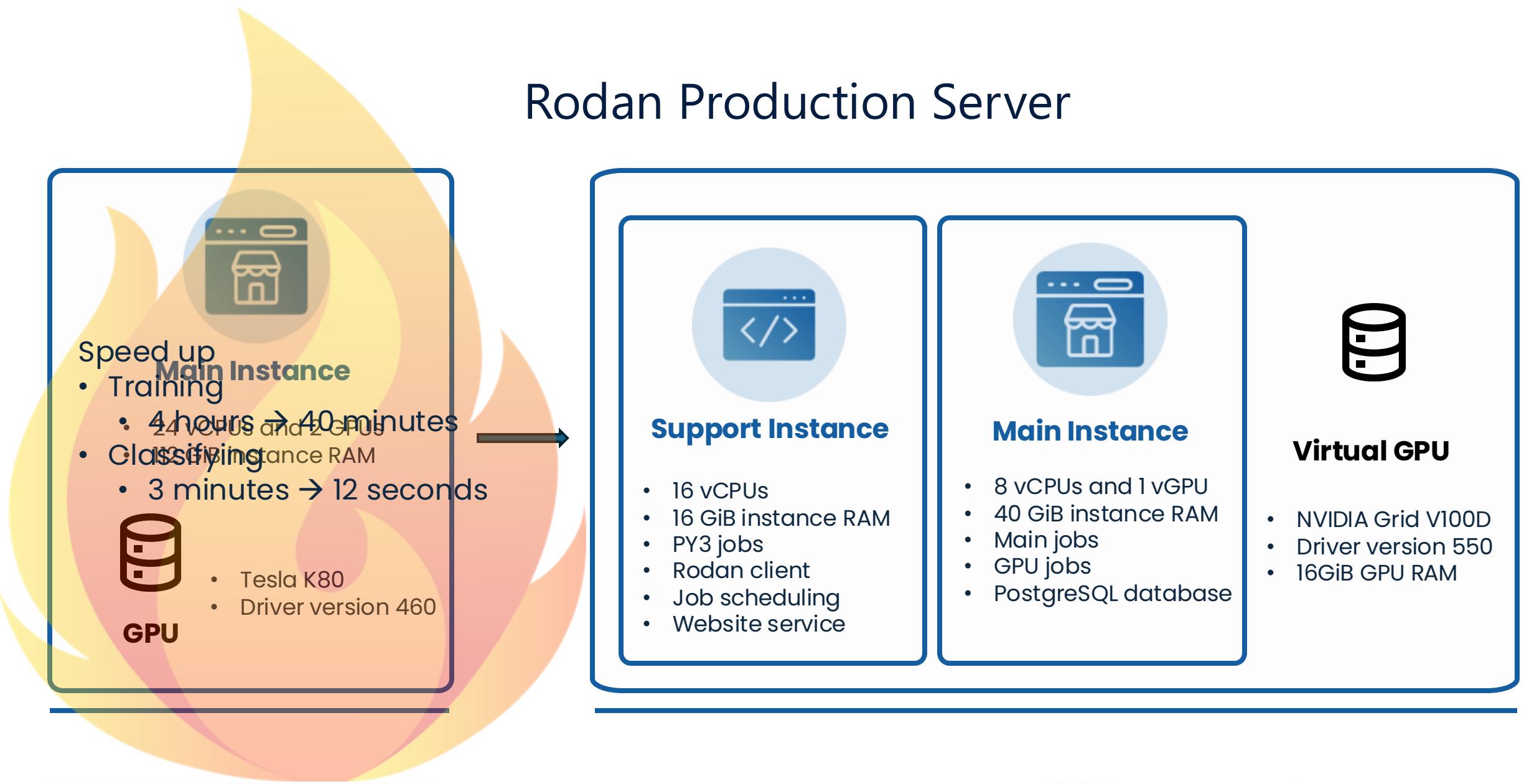
Development since V3.0.0

As of 2024 Oct 26
Current staging version

1. UI & frontend fix
2. Debugging
 - Staff Finding
 - Interactive Classifier
 - Heuristic Pitch Finding
 - Image Layer Separation (PACO) Training
3. Update dependency versions
4. New job
 - Extract C Clef



Rodan Production Server



THANK YOU!



DDMAL

DISTRIBUTED DIGITAL MUSIC
ARCHIVES & LIBRARIES LAB



Digital Research
Alliance of Canada

Alliance de recherche
numérique du Canada



linkedMusic

End-to-end with MS73

[CDN-Mrb 0073]

LinkedMusic Project Meeting III
McGill University, Montréal QC

Kyrie Bouessa

SSHRC = CRSH

C I R
M M T Centre for Interdisciplinary Research
in Music Media and Technology



Schulich School of Music
École de musique Schulich

Rodan: Optical Music Recognition Workflow



RODAN!

```
</neume>
</syllable>
<neume xml:id="m-4974553c-4020-420e-8aec-e6e533d2d9b5">
  <syl xml:id="m-71fdabc-e500-47ee-bbfa-624683a7d600" facs="#m-
  <neume xml:id="m-e999fd3f-3d1f-4263-94cc-60dd7b414461">
    <nc xml:id="m-218f06e8-00ef-479e-8d15-3b567d44cf0e" facs=
    <nc xml:id="m-1ac6df91-4e44-4e1e-b7f7-d12b4be67769" facs=
  </neume>
</syllable>
```

Image: MS73.133v

MEI encoding of a manuscript

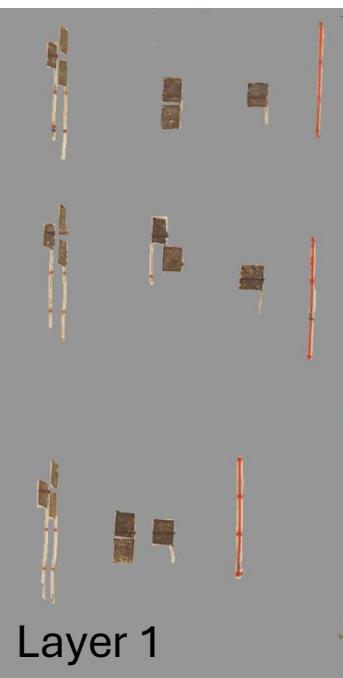
MS 73

- A Dominican choirbook likely produced in Northern Italy.
- Initially thought to have been solely produced in the fifteenth-century, though later research by Alessandra Ignesti established the manuscript to have also been produced in the late thirteenth-century into the early fourteenth-century.
- 1275–1315; 1400s
- 10 distinct sections.



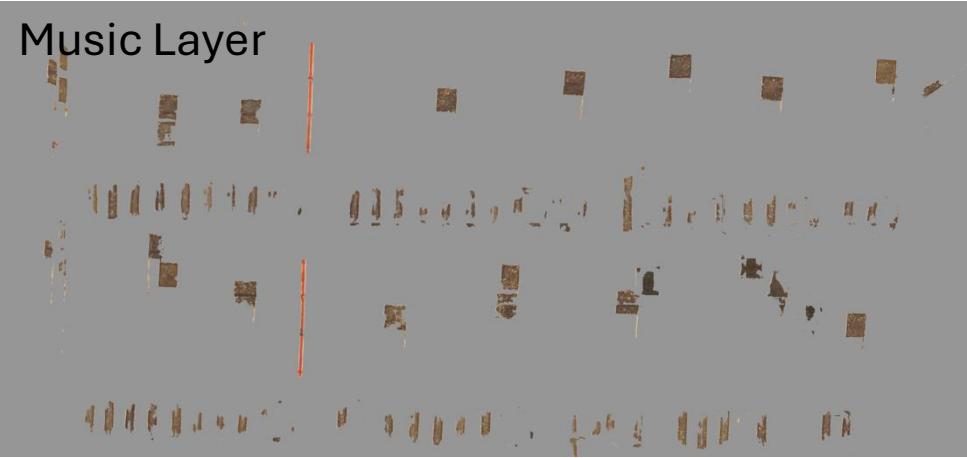
Rodan Phase 1

- Input an image
- Annotate sections of the image
 - No longer need to do the whole thing!

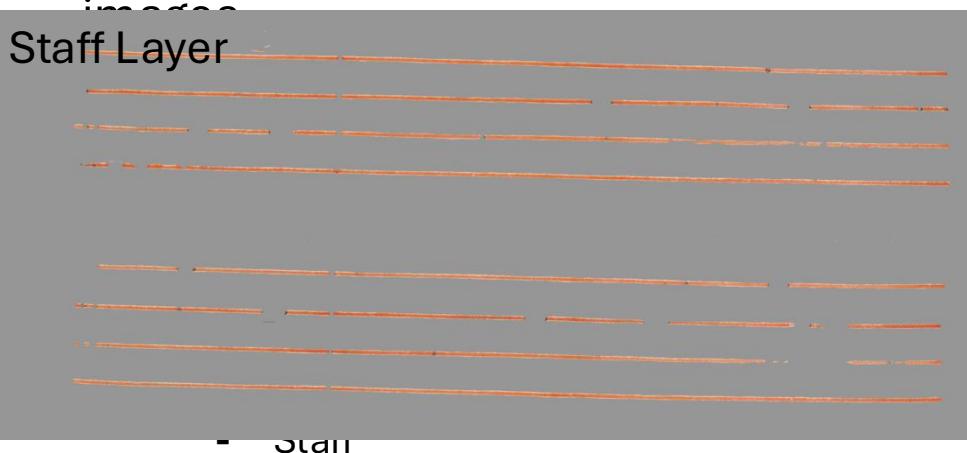


MS73 fol.63

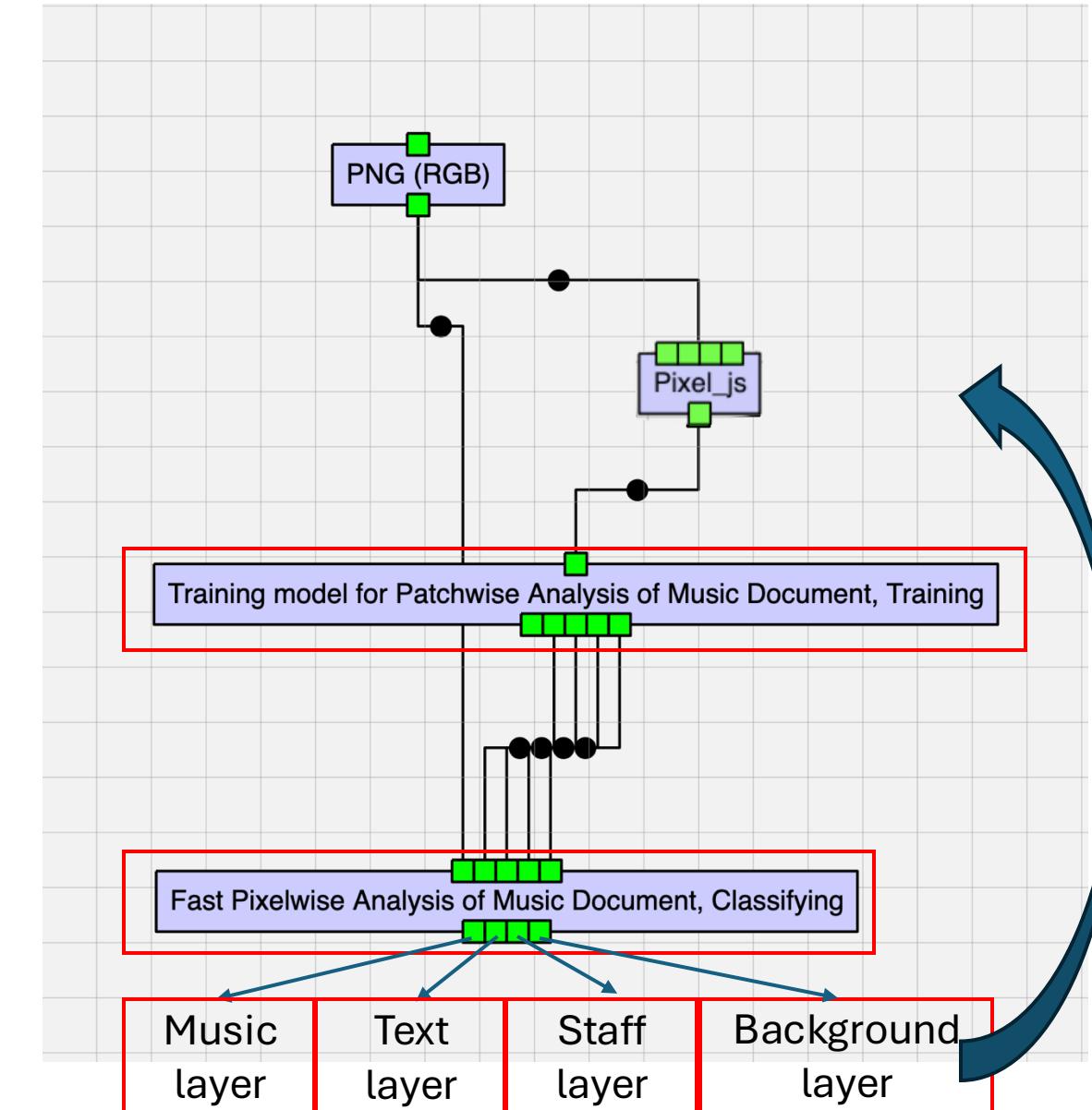
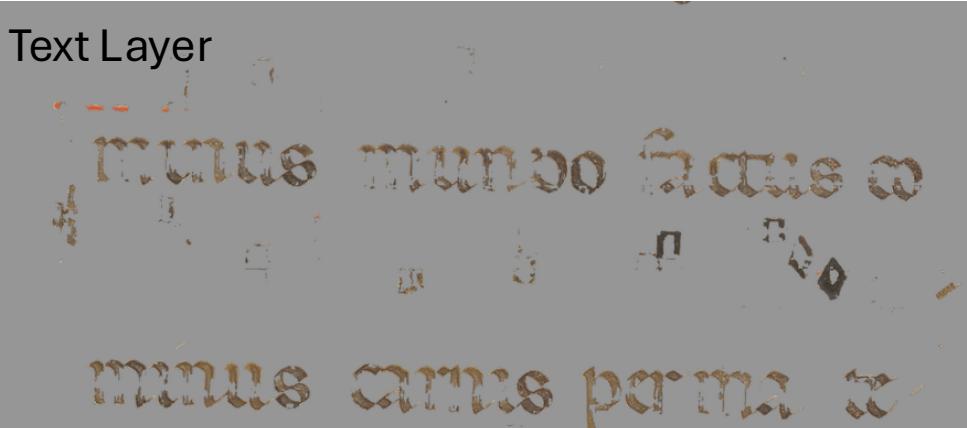
Music Layer

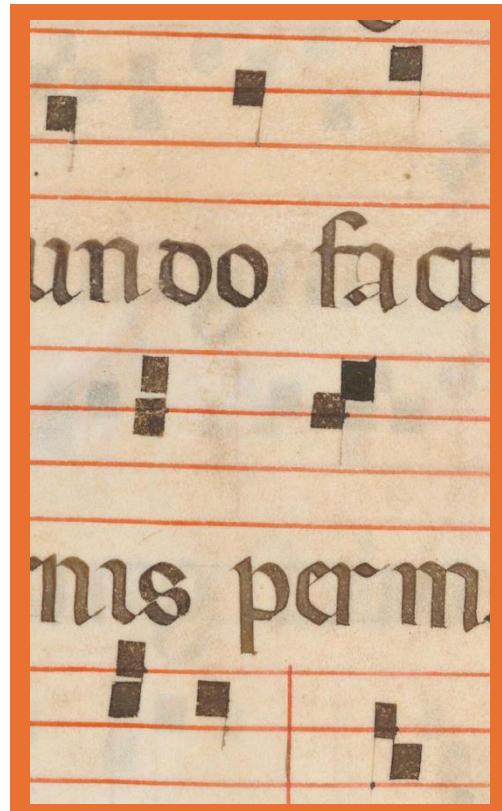
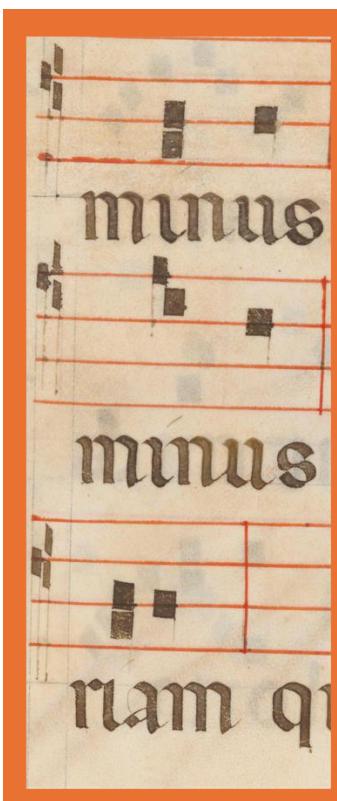


Staff Layer



Text Layer



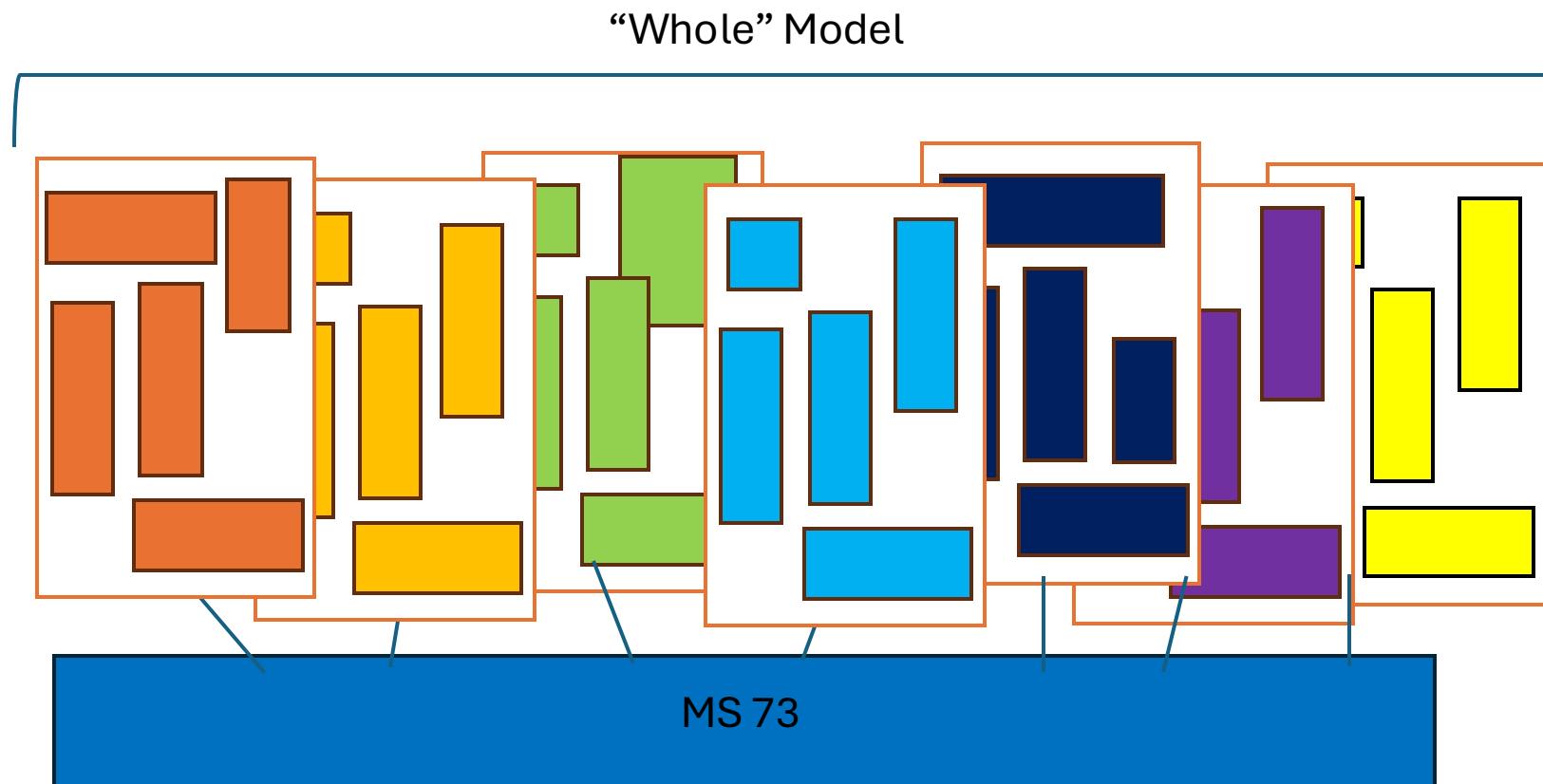


MS73 fol. 63



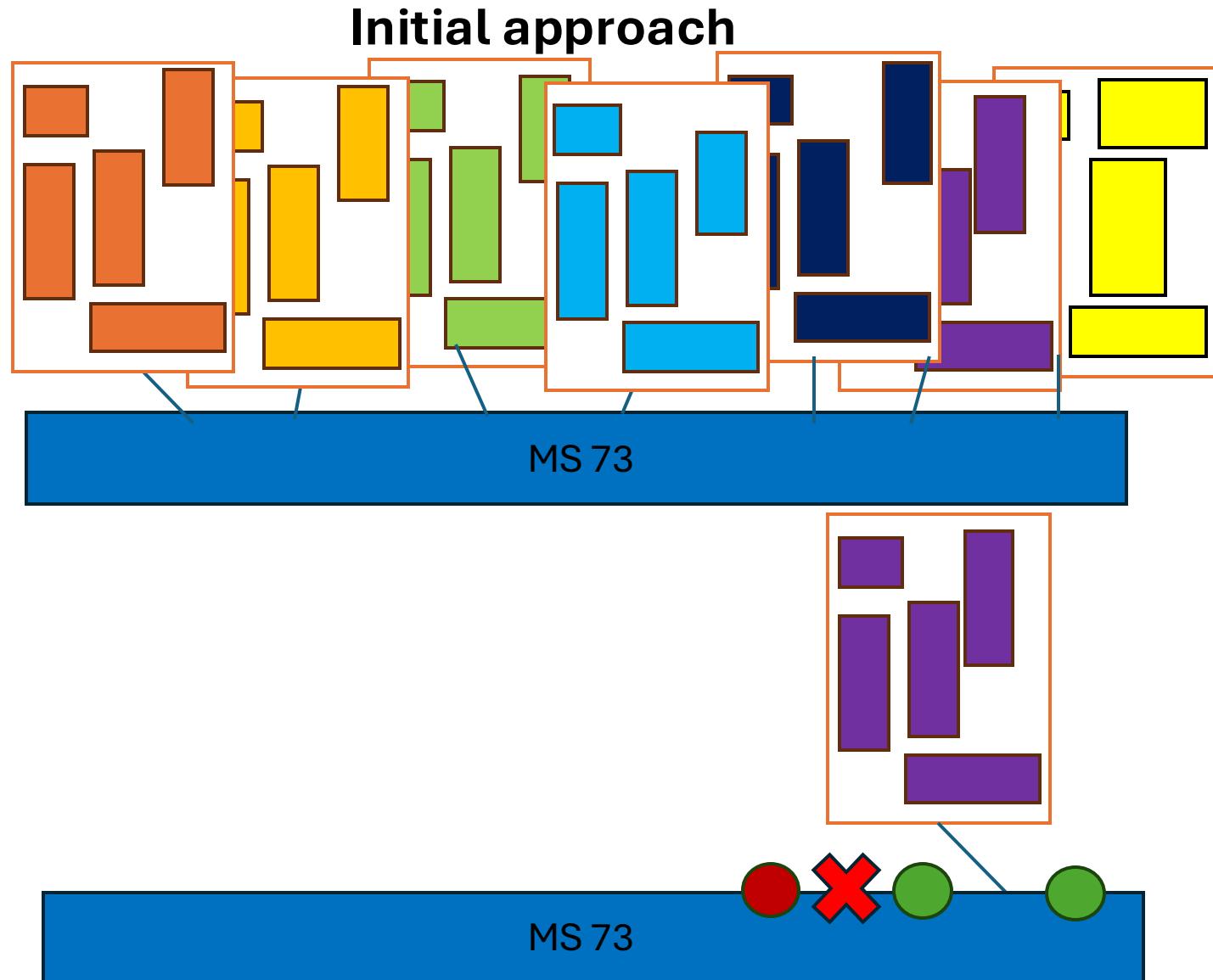
Model Mayhem I

- Initial approaches was to make one model
 - Trained on pages throughout the manuscript
 - Model should work on random images from anywhere in the manuscript
 - This did not happen



Model Mayhem II

- Shrunk our sample groups:
 - “fol.200 model” did well on 230 and 190 but terribly on 170, etc.
- Models began to do significantly better!
 - ... to a point; “leaving the neighborhood”
- How wide can we go until they stop working?
 - Looking a little deeper...



Early



Late



ta da bo mamma abscon
ditum et nomen no
uum alle lu ia. ps
Laudate. Quoniam
v Posuit fines tuos
tuos pacem R Ex adipe
frumenti faciat te.
Ad bñs ant. Ego
sum pa mis uiuus qui

dite cum manebit in ceterū. cā.

Bñs. evocac. Adorāt. Quid

aūt accidit in tiām bonam hū sūt

q in corde bono et optimo ubū re

tinēt et fructū afferūt in pacienti

a. Bñs. evocac.

Hmīca in l. salbo p̄dēti ad uesp. R.
Hū staret. Adorāt. Quidat. In p. n. R.

Middle

Late

pleuerit. **A**braham stabat

corā deo et ait absit a te domi-

ne ut perdas iustūm cū impūis.

Desce. **A**scendens er go de-

us ad abraham pluit i gñē et

sulphur sup sodomā. **A**braham

manc cōsurgēs iste tit et e uer-

deum nostrum In conuersi-
o ne apo stoli pauli. p.

Elemitte. In pmo noctno. x.

SRulus ad hoc spuimis

minarum et cedis in discipu-

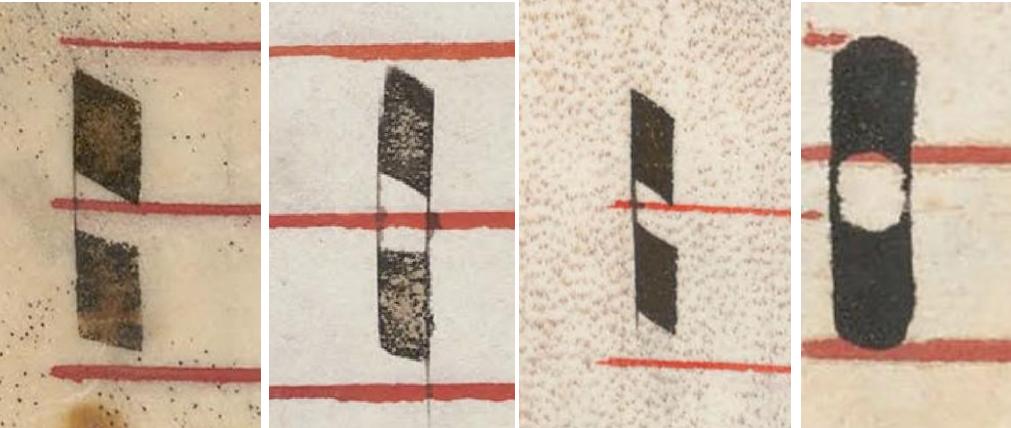
los domini abut ad principē

sacerdotium et petit ab eout

Scribes, Editors, Contributors I

Divisio maxima

C-clefs

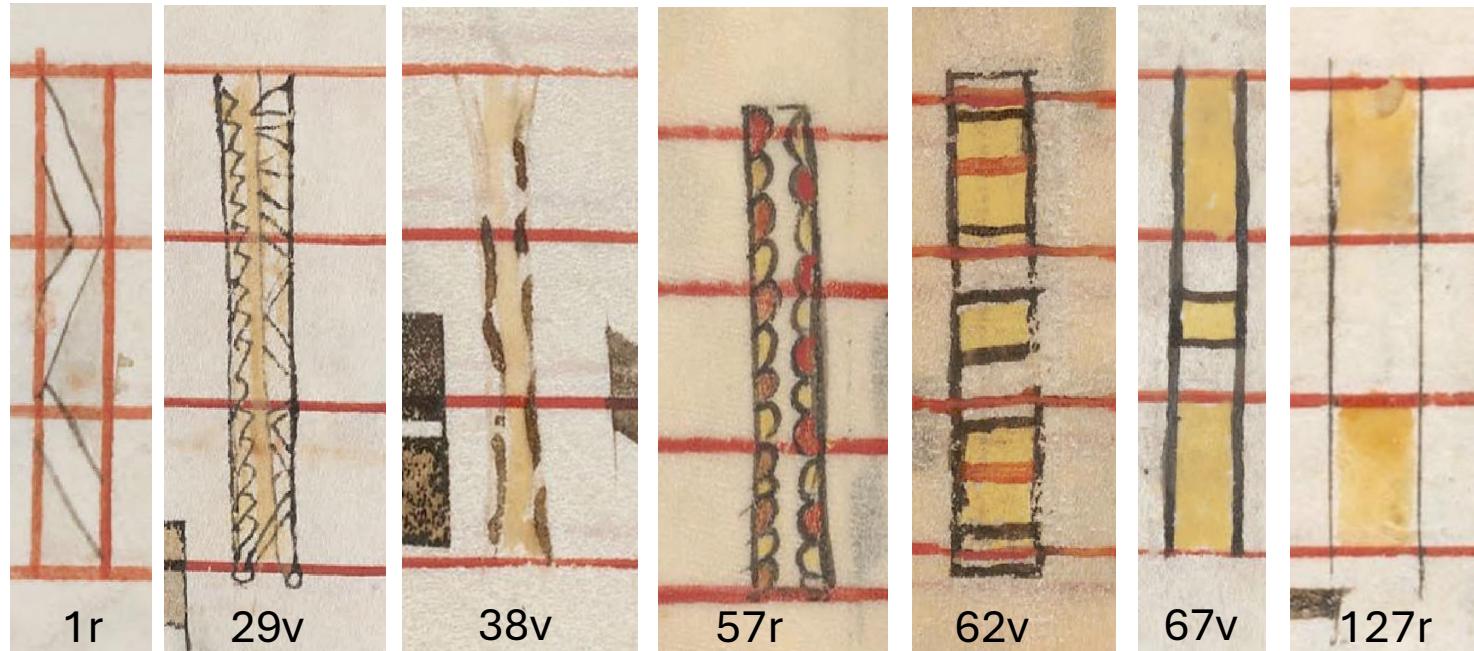


1r

79r

93r

119v



13r

29v

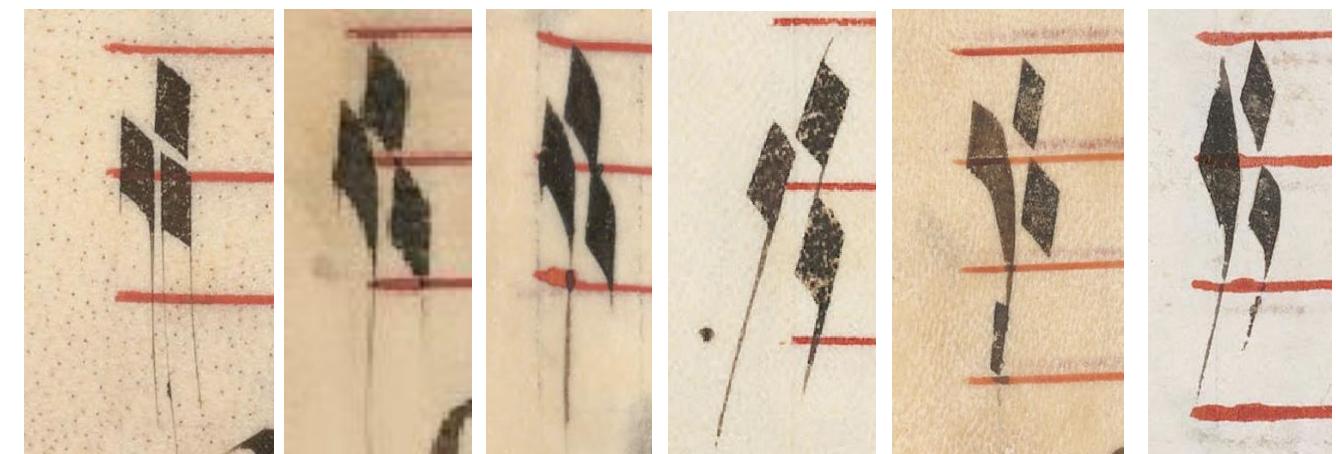
47v

88v

101r

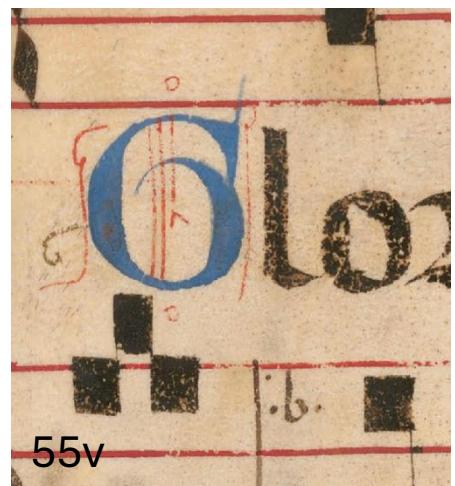
138r

F-clefs



Scribes, editors, contributors II

Initials



Custodes

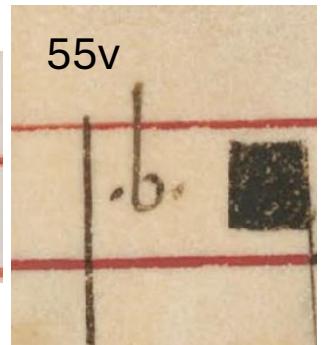
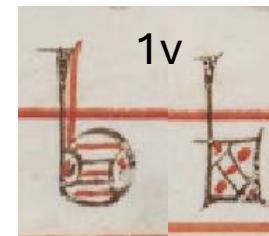
125r



36r

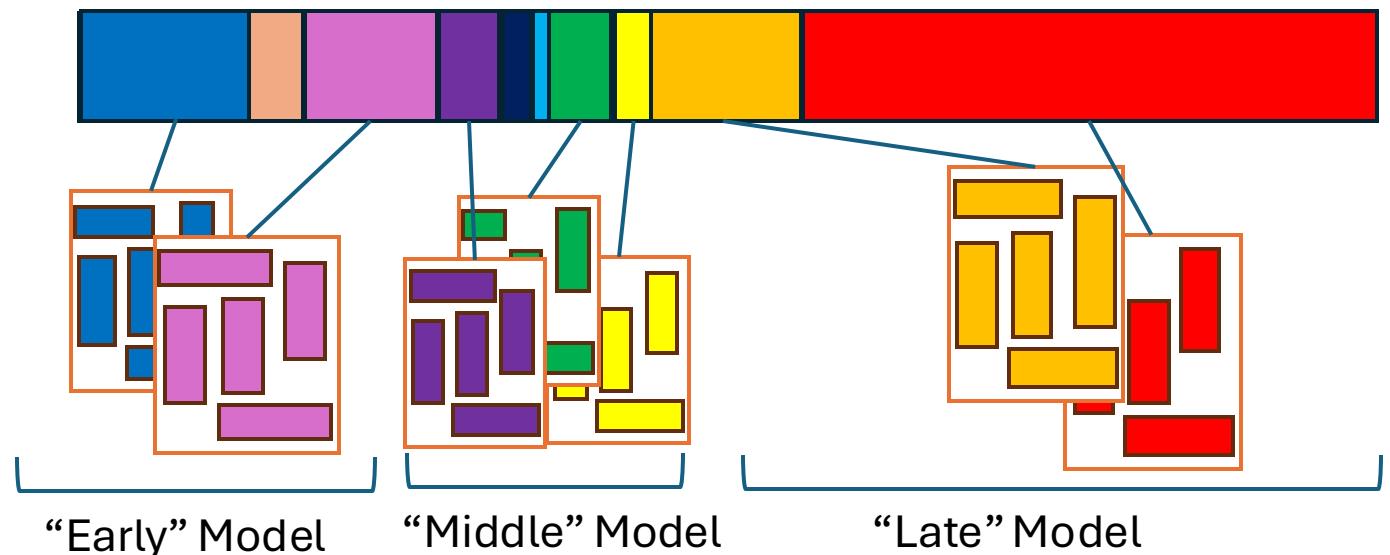


B-flats



Codicological Units

- Looking deeper revealed distinct groupings and similarities.
- Establish further rough boundaries where significant changes between paleography, layout, and style.
 - Some neighbors worked better together than others: bigger groups
 - Trained regional models: early, middle, late.



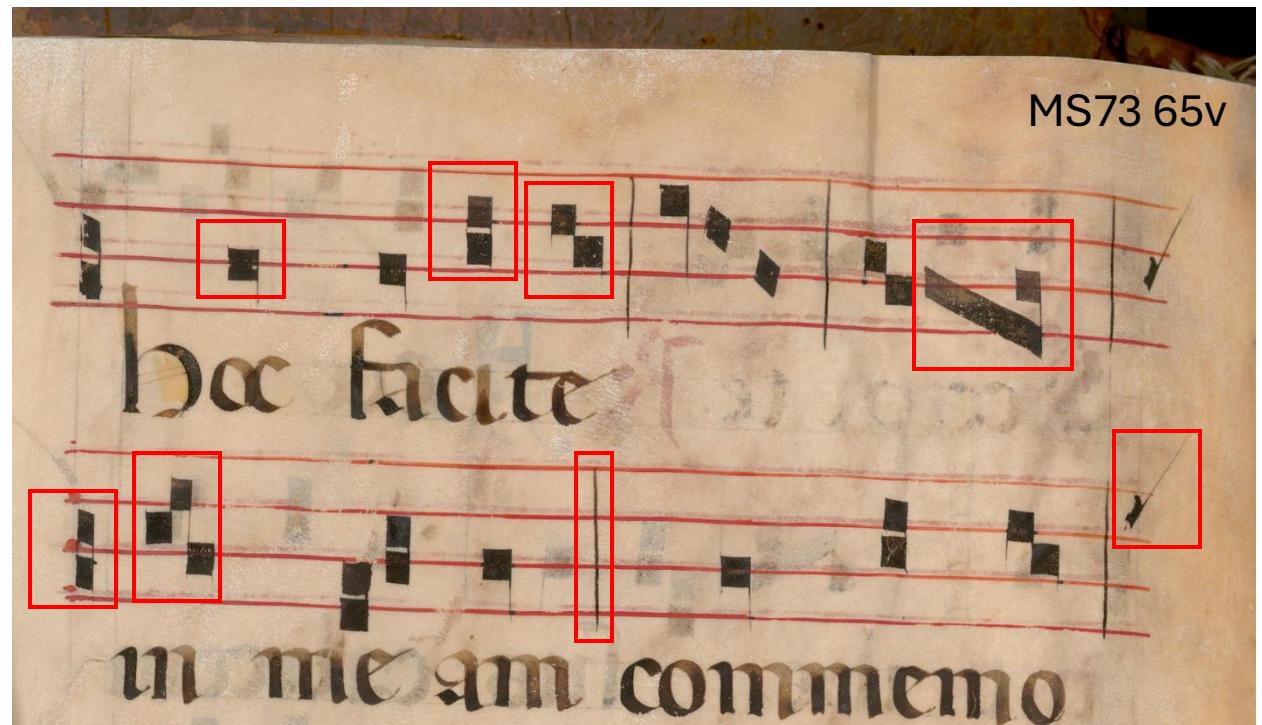
Phase 2: Neume-level Classifying

Stage 1: Interactive (Manual) Classifying

Music layers can now be separated from text and staff—now what?

- Learning types of neumes
- Recognizing different versions of the same neumes, clefs, custodes

We do this manually to train the classifier, until it begins sorting neumes accurately.



neume.virga
neume.podatus2b
neume.clivis
clef.c

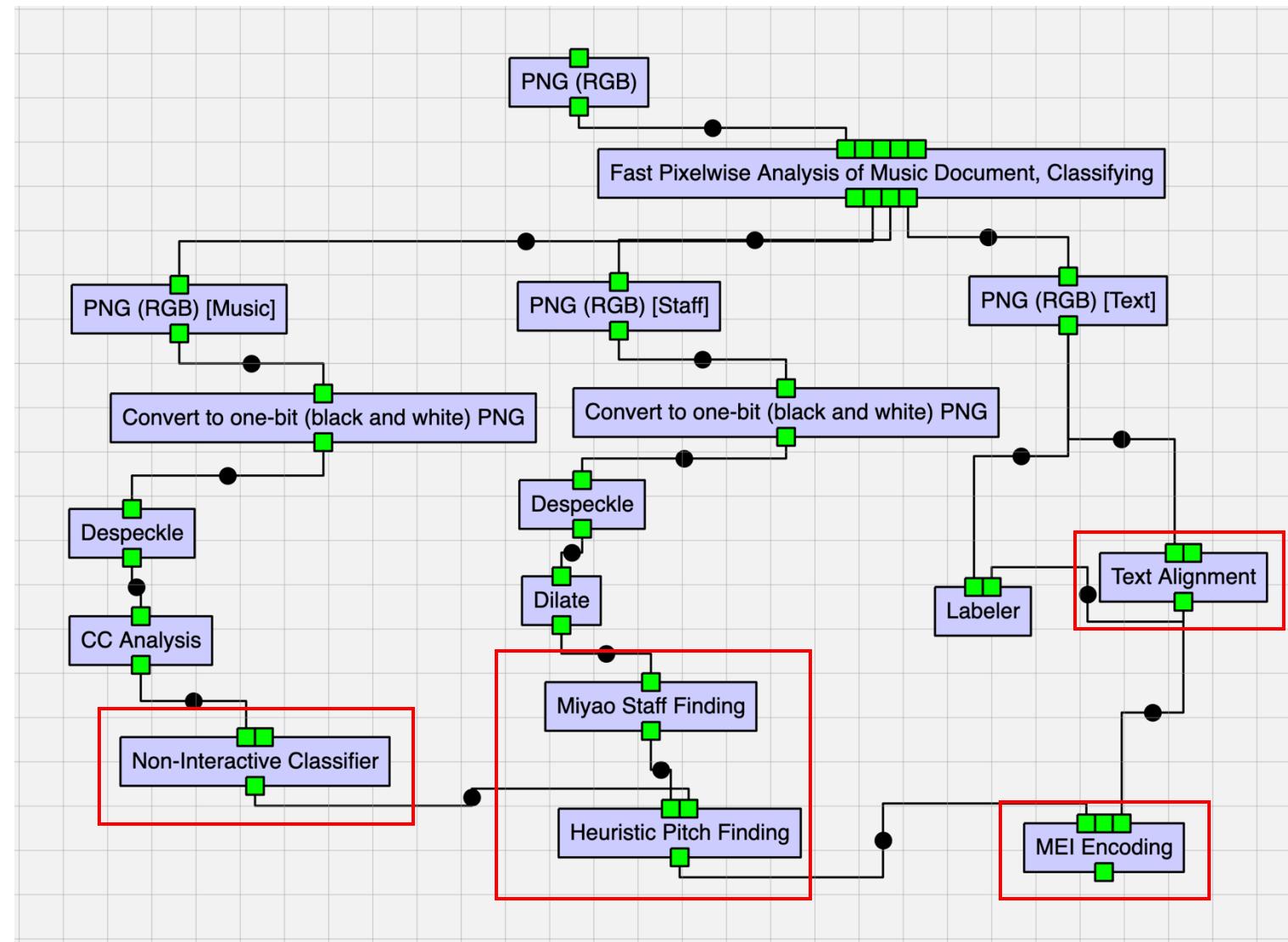
Etc.

Phase 2: Neume-level Classifying

Stage 2: **Non-Interactive** (automatic) Classifying

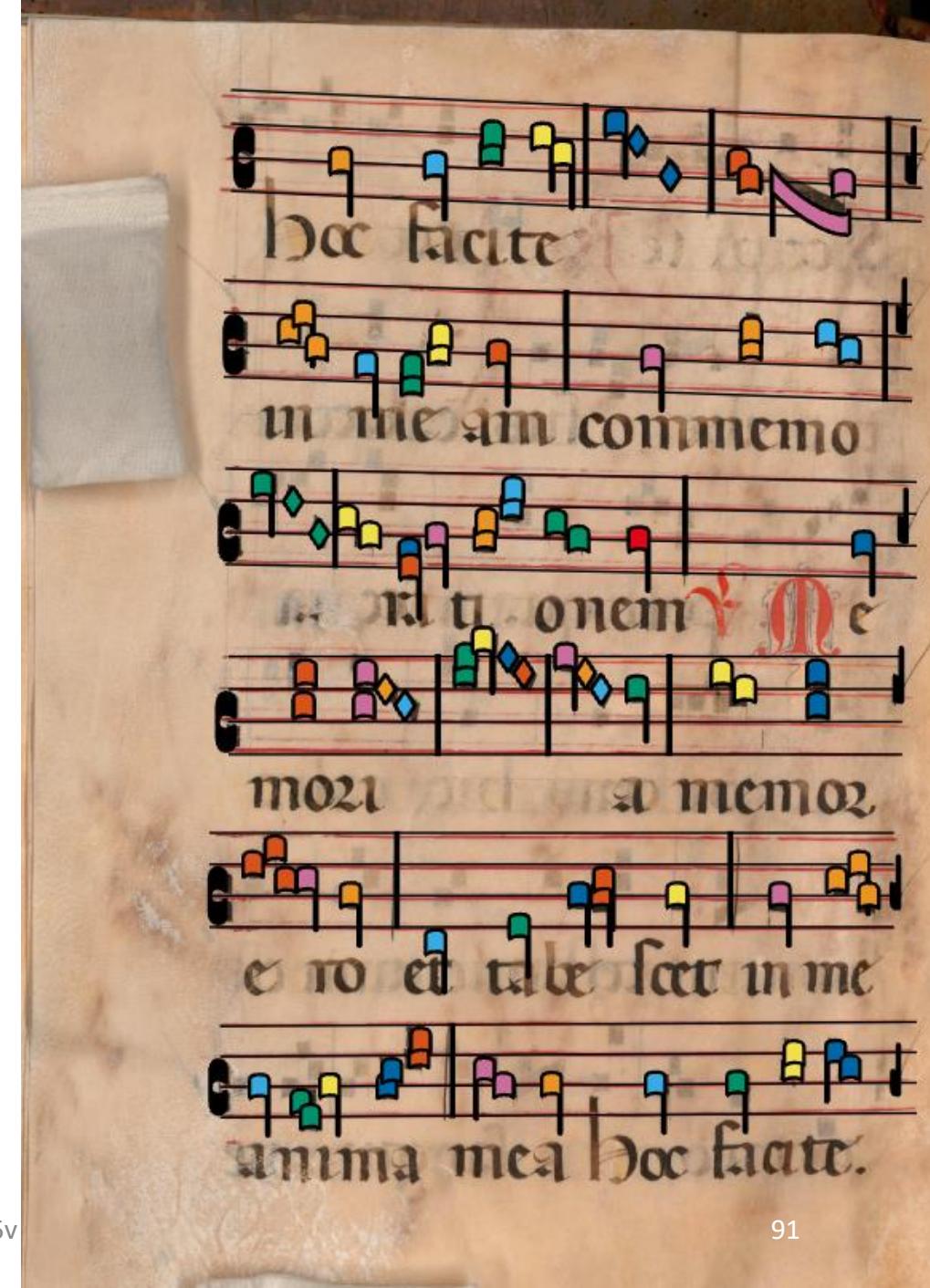
Rodan can now sort neume types even if they look a little different from one another

- Neumes are identified in an MEI format
 - Neume.virga
- Staff layer gets analysed
- Text transcript is applied to the text layer
- An MEI file is produced

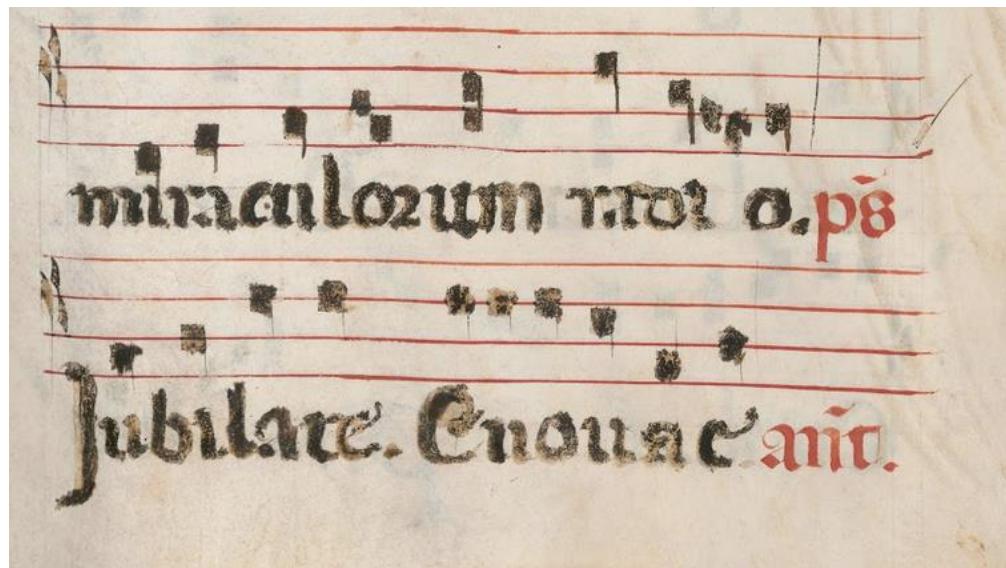
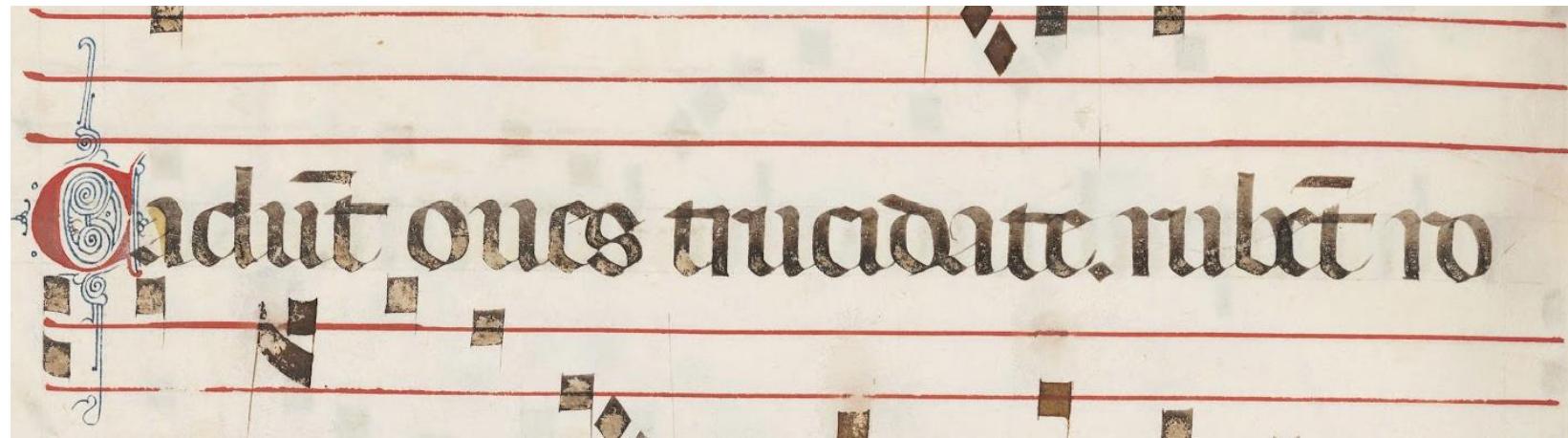


Neon!

- We have an image and an MEI file!
- Correct in Neon



Thank you!



LinkedMusic

• • •

Yueqiao Zhang - McGill University

Self-Introduction

- McGill Undergrad
- Working in DDMAL for the past summer
- Enrolled in Computer Science Major and Music Applied to Technology Minor

Project Overview

*This presentation will use two databases as examples: The Session and MusicBrainz, since the workflow for them are the most completed.

1. Retrieving data dumps
2. CSV and Reconciliation
3. CSV to RDF logic
4. Upload to our Virtuoso graph database

Retrieving Data

Retrieving Raw Data from The Session Database

- fetch_data.py
- There will be 6 available types of data in 6 separated CSV files:
 - aliases.csv, recordings.csv, events.csv, sessions.csv, sets.csv, tunes.csv
- Example on the next page

Details of the CSVs

- aliases: the alternative names for different Irish folk tunes
- recordings: various recordings made based on different Irish folk tunes
- events: the events (e.g. festivals, concerts, etc) related to Irish music and culture
- sessions: music sessions related to Irish music and culture
- sets: the playlists containing tunes made by public users
- tunes: the informations about different tunes

Example: events.csv

```
id,event,dtstart,dtend,venue,address,town,area,country,latitude,longitude
3310,,"1900-01-01 00:00:00","1900-01-01 00:00:00",-,,"Ober-Kainsbach,Hessen",,,
11,"Colm Gannon, Sean McKeon And John Blake","2006-06-07 09:30:00","2006-06-07 10:30:00",,"Glór,Causeway Coast and Glens,Antrim",,,
5,"Brid O'Donohue","2006-06-09 08:00:00","2006-06-09 00:00:00",,"Glór,Causeway Coast and Glens,Antrim",,,
6,"National Celtic Festival","2006-06-09 19:00:00","2006-06-12 15:00:00",,"Vivian,Waterford,Waterford",,,
7,"The Irish Connections Festival","2006-06-09 19:00:00","2006-06-11 00:00:00",,"The Irish Connections Festival,Waterford,Waterford",,,
19,"Louisville Irish Fest","2006-06-10 11:00:00","2006-06-11 06:00:00",,"Belmont Park,Louisville,Kentucky",,,
```

events.csv cont.

The raw events.csv contains its ID in the first column, followed by all of its attributes.

▼ id	▼ event	▼ dtstart	▼ dtend	▼ venue	▼ address
3310	–	1900-01-01 00:00:00	1900-01-01 00:00:00	–	
11	Colm Gannon, Sean McKeon And John Blake	2006-06-07 09:30:00	2006-06-07 12:00:00	edit The Goalpost	226 Water Street
5	Brid O'Donohue	2006-06-09 08:00:00	2006-06-09 00:00:00	Glór	Causeway Link
6	National Celtic Festival	2006-06-09 19:00:00	2006-06-12 15:00:00	Various Venues	
7	The Irish Connections Festival	2006-06-09 19:00:00	2006-06-11 00:00:00	Irish Cultural Centre	200 New Boston Drive

`find_artist.py`

- A special CSV made specifically for recordings.csv
- Since the URLs for the artist are not in the data dumps but they still appear on The Session website, we need the `find_artist.py` script to retrieve those data.

Retrieving Raw Data from MusicBrainz

- fetch.py & untar.py
- The newest data will be downloaded and parsed into separated JSON files: areas, artists, events, instruments, labels, places, recordings, releases, release-groups, series, works, genres

Brief Overview of Data Size

- release.json has
 - 12.94GB of pure text in JSON format
 - 1 char takes 1B
 - $1.3*(10^{10})$ ~Ten billion characters

get_genre.py

- genre.json could only be retrieved by API calls
- get_genre.py is a specific script made for this.

```
▼<metadata xmlns="http://musicbrainz.org/ns/mmd-2.0#">
  ▼<genre-list count="1983">
    ▼<genre id="18797864-41b0-4602-8241-adbee761774c">
      <name>2 tone</name>
      </genre>
    ▼<genre id="db325bd7-ae64-40bd-966a-a3af3cef8bb9">
      <name>2-step</name>
      </genre>
```

Advantages for the Retrieving Scripts

1. Facilitate the following work by retrieving the data dumps to local spaces
2. Avoid manual downloading
3. Easy updating

CSV and Reconciliation

Reconciliation

- Data Reconciliation:
 - This is the process of identifying and resolving inconsistencies or discrepancies between different data sources. (Wikipedia)
 - To allow cross-database querying, we use reconciliation in our LinkedMusic project to link multiple databases together.
 - In our project, Reconciliation means matching the values to Wikidata.

OpenRefine

- OpenRefine is a powerful software for graph-editing.
- It has a built-in Reconciliation function that facilitates our researching works.



DEMO

Why is Reconciliation beneficial?

1. Allow cross-database querying
2. Allow multi-language searching
3. Making the logic (relational schema) behind linked data more complete.
4. More to explore...

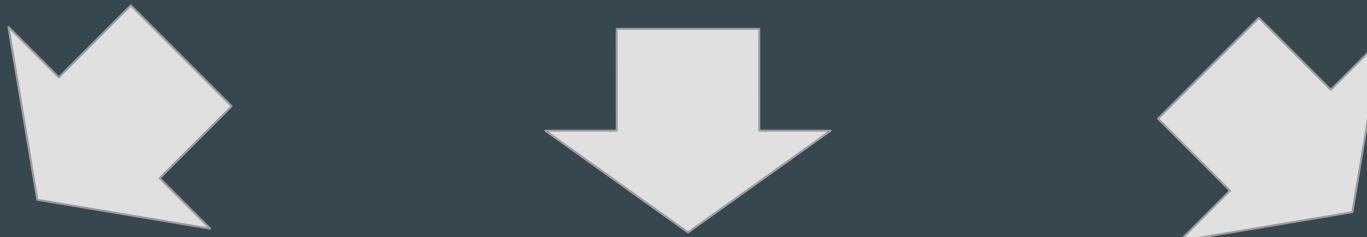
Reconciliation for The Session Database

- Since the raw data is in CSV format, we can import it into OpenRefine directly.
- Follow the step-by-step reconciliation guide in our GitHub site
- More reconciliation steps and decisions should be applied according to updates of The Session.

Example: events.csv

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

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

Reconciliation for MusicBrainz

- To convert from JSON file to CSV file, use [convert_to_csv.py](#)
- Similar to The Session Database, we need to further reconcile some attributes for the CSVs. We import the CSVs into OpenRefine and perform specific actions as shown in the demo.
- MusicBrainz's main entities are already reconciled and they are contained in the raw data dumps. We keep them during our conversion process to facilitate our reconciliation steps.

Example: area.csv

▼ area_id	▼ sort-name	▼ relations_wiki	▼ type	▼ type_wiki_url
https://musicbrainz.org/area/b0f5611f-a6ad-470f-9da9-21ff76e5686d	Kashima	https://www.wikidata.org/wiki/Q684494	city Choose new match	https://www.wikidata.org/wiki/Q15253706
https://musicbrainz.org/area/f7aa751ca4ac-4a1b-b23e-9778e8da3063	Tosa	https://www.wikidata.org/wiki/Q851163	city Choose new match	https://www.wikidata.org/wiki/Q15253706
https://musicbrainz.org/area/28ab5a3e-8327-4d31-8106-361cf1283b4d	Isahaya	https://www.wikidata.org/wiki/Q596469	city Choose new match	https://www.wikidata.org/wiki/Q15253706

Wikidata Entities Explanation

- Q-values:
 - Their URLs has the format <http://www.wikidata.org/entity/Q16>
 - Canada
 - They represent an “entity”
- P-values:
 - Their URLs has the format <http://www.wikidata.org/prop/direct/P17>
 - Country
 - They represent a “property”

Difficulties Encountered

1. Many very unpopular or detailed entities are not present in Wikidata such as “event names”, “addresses”, etc. We need to enrich Wikidata to solve this problem.
2. Some predicates are vague in their definitions. It becomes difficult to determine which Wikidata properties to use. For example, P50 is “author” and P175 is “performer”. An artist in MusicBrainz could be either one of them. We had some trouble identifying properties similar to these two.

CSV to RDF

What is RDF?

- **Resource Description Framework** is a standard model for representing information on the web
- **Triples** : RDF data is built using "triples", which consist of three parts: subject, predicate, and object.
- For example, an RDF triple might look like this: John (Subject), has_age (predicate), 25 (object)
 - This represent that “John has age of 25.”

Application of RDF

URIs (Uniform Resource Identifiers) :

RDF uses URIs to uniquely identify resources (subjects, predicates, and some objects). This allows data to be linked and shared across different datasets.

In our project, URIs are usually represented by hyperlinks.

Linked Data :

RDF is central to the concept of linked data, enabling datasets to be connected by referencing common resources via URIs. This allows for richer data relationships and more powerful queries across databases.

RDF example

```
<https://thesession.org/events/3631>;  
ns2:P17 <http://www.wikidata.org/entity/Q16>,
```

- Subject: <https://thesession.org/events/3631>
- Predicate: <http://www.wikidata.org/prop/direct/P17>
 - The Wikidata property for “country”
- Object: <http://www.wikidata.org/entity/Q16>
 - This is Wikidata URL to “Canada”

The Script

- csv2rdf_single_subject.py
- Two approaches to conversion:
 - Merge all CSV to one huge RDF:
 - Advantage: Only one command, easy workflow, easy update
 - Disadvantage: Large file might not upload to Virtuoso
 - One separate RDF file for each CSV:
 - Advantage: Smaller files, easy to correct if something went wrong, no error when uploading
 - Disadvantage: Much more complicated workflow, difficult to update
- Mapping.json NEEDED!

Conversion Logic

1. In a row of CSV, we identify the first column of that row to be the subject.
2. Each column after the first will represent a triple in RDF.
3. The header of the column is the predicate, the cell in that column x row is the object.
4. The subject must be an entity type URI and the predicate must be a property type URI. The object can be a Literal of any type or an entity type URI.

Example conversion

https://thesession.org/events/5	Brid O'Donohue	< https://thesession.org/events/5 >; ns2:P2561 "Brid O'Donohue" ;
https://thesession.org/events/6	National Celtic Festival	< https://thesession.org/events/6 > ; ns2:P2561 "National Celtic Festival" ;

mapping.json

- This file must include the JSON dictionary for all the predicates. Run the script `get_relations.py` to retrieve a new JSON mapping.json file for all input CSV.
- In the file that contains a dictionary, all column headers of all input CSVs will be as the keys. The values starts to be empty, the operator will have to fill in the correct Wikidata property URI for all the corresponding header predicates.

Difficulties Encountered

- If multiple subject exists in one single CSV, the script is not able to recognize it and cannot convert the data corresponding to the correct schema. This issue is now resolved by making separate CSVs. More efficient fixes will involve further discussion.

Virtuoso

What is Virtuoso?

- Virtuoso is a multi-model database and application server, and it supports a wide range of features, including
 - RDF Data Management
 - SPARQL Endpoint
 - Relational and Graph Database
 - Data Integration
 - Linked Data and Web Services
- For my project, uploading the final RDF for a database into Virtuoso is the goal.
- After uploading several database onto Virtuoso, its SPARQL Endpoint feature will allow cross-database search.

Virtuoso operation example

The screenshot shows the Virtuoso Web Application Server interface. The top navigation bar includes links for Home, System Admin, Database, Replication, Web Application Server, XML, Web Services, and Linked Data. Below this is a secondary navigation bar with links for NNTP, SPARQL, Sponger, Statistics, Graphs, Schemas, Namespaces, Views, R2RML, and Quad Store Upload. The 'Quad Store Upload' link is highlighted with a red arrow pointing from the title above. The main content area is titled 'Quad Store Upload'. On the left, there are two radio button options: 'File*' (selected) and 'Resource URL*'. A file input field labeled '选择文件' (Choose file) is shown with the message '未选择任何文件' (No file selected). Below this is a text input field for 'Resource URL*'. Underneath these fields is a checkbox labeled 'Create graph explicitly'. At the bottom are two buttons: 'Cancel' and 'Upload'. Red annotations in the screenshot highlight the 'File*' radio button, the '选择文件' input field, the 'Resource URL*' input field, the 'Create graph explicitly' checkbox, and the 'Upload' button.

Quad Store Upload

● File* 选择文件 未选择任何文件 Choose file to upload

○ Resource URL*

Create graph explicitly http://thesession/reconciled

Named Graph IRI* http://localhost:8890/DAV
http://musicbrainz/reconciled

Cancel Upload

Further Improvements

Further Improvements

1. Live and efficient data dump updates
2. More automatic and universal reconciliation process
3. Allow more complex relational schema, such as multiple subject included in one CSV
4. More specific data type recognition
 - a. Such as automatic language detection, datetime format recognition, coordinate location recognition, etc.

Thank You for Listening!

...

Yueqiao Zhang

Leveraging ChatGpt for Natural Language Query to SPARQL on Virtuoso

--Using TheSession, Wikidata, MusicBrainz as example



Junjun Cao
Postdoctor of DDMAL
Email: junjun.cao@mail.mcgill.ca

- **Section 1** TheSession Webpage Queries vs. Using SPARQL on Virtuoso
- **Section 2** Queries Only Executable Using SPARQL on Virtuoso but not on TheSession Webpage
- **Section 3** Queries across TheSession and Wikidata
- **Section 4** Queries across TheSession, MusicBrainz and Wikidata

TheSession Webpage Queries vs. Using SPARQL on Virtuoso



The “session”: mostly informal gatherings at which people play Irish traditional music.

- Q1-1: Find all the sessions that took place in Greece
- Q1-2: Find the tunes of genre “jig,” in “D major,” having “Malcolm” in its name
- Q1-3: Find albums containing “Tola” in its name or its artist’s name
- Q1-4: Find albums related to “Mirella”

Searching on TheSession webpage (Q1-1)

<https://thesession.org/sessions>

The screenshot shows the homepage of thethesession.org. At the top, there is a navigation bar with links for 'Log in or Sign up', 'TUNES', 'DISCOGRAPHY', 'TRIPS', 'SESSIONS' (which is highlighted in yellow), 'EVENTS', and 'DISCUSSIONS'. Below the navigation bar, there is a large yellow sidebar with the word 'SESSION' in white. The main content area has a title 'Sessions' and a search bar with the placeholder 'Search for sessions on any day in Greece'. A yellow 'SEARCH' button is to the right of the search bar.



1. **The Lucky Sparrow Irish Pub**, Athens, Athens, Greece

Added by Athens Irish Festival 9 months ago.

2. **The Dubliner**, Thessaloniki, Makedonia, Greece

Added by lukegarry 3 years ago. Updated 2 months ago.

Prompting ChatGPT to generate SPARQL query (Q1-1)

SPARQL: SPARQL Protocol and RDF Query Language

Enter the following in ChatGPT (**the text in red will be provided to users**):

Given a database reconciled with Wikidata, knowing `?session a
<<https://thesession.org/sessions>>`, for question “Find the sessions that took place
(wdt:P17) in Greece”, please generate a SPARQL query. **Don’t forget namespace.**

Generated SPARQL by ChatGPT (Q1-1)

Find the sessions that took place in Greece

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

PREFIX wd: <<http://www.wikidata.org/entity/>>

PREFIX rdf: <<http://www.w3.org/1999/02/22-rdf-syntax-ns#>>

SELECT DISTINCT ?session

WHERE {

?session rdf:type <<https://thesession.org/sessions>> .

?session wdt:P17 wd:Q41 . # P17 is the 'country' property, and Q41 is the Wikidata entity for Greece

}

Virtuoso SPARQL query interface (Q1-1)

The screenshot shows the Virtuoso SPARQL Query Editor interface. At the top, there is a browser-style header with icons for window control (red, yellow, green), a shield icon, a URL field containing "virtuoso.staging.simssa.ca", and a refresh/circular arrow icon. To the right of the URL field are a download icon, a plus sign, and a search/magnifying glass icon.

The main menu bar includes "SPARQL Query Editor", "About", "Tables", "Conductor", "Facet Browser", and "Permalink". Below the menu, there are "Extensions" buttons for "cxml", "save to dav", "sponge" (which is highlighted in blue), and "User: SPARQL".

A "Default Data Set Name (Graph IRI)" input field is present. The central feature is a large "Query Text" input area with a blue border, currently empty except for a cursor.

At the bottom, there is a "Results Format" dropdown set to "HTML" and two buttons: "Execute Query" (blue background) and "Reset".

Copy SPARQL code then Execute (Q1-1)

<https://virtuoso.staging.simssa.ca/sparql>

Default Data Set Name (Graph IRI)

http://sample/thesession/reconciled

Query Text

```
PREFIX wdt: <http://www.wikidata.org/prop/direct/>
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
SELECT ?session
WHERE {
    ?session rdf:type <https://thesession.org/sessions> .
    ?session wdt:P17 wd:Q41 . # P17 is the 'country' property, and Q41 is the Wikidata entity for Greece
}
```

Results Format

HTML



Execute Query

Reset

The result from Virtuoso (Q1-1)

SPARQL | HTML5 table

session

<https://thesession.org/sessions/7243>
<https://thesession.org/sessions/8020>

This is why Linked Data is associated with
“Semantic Network”?

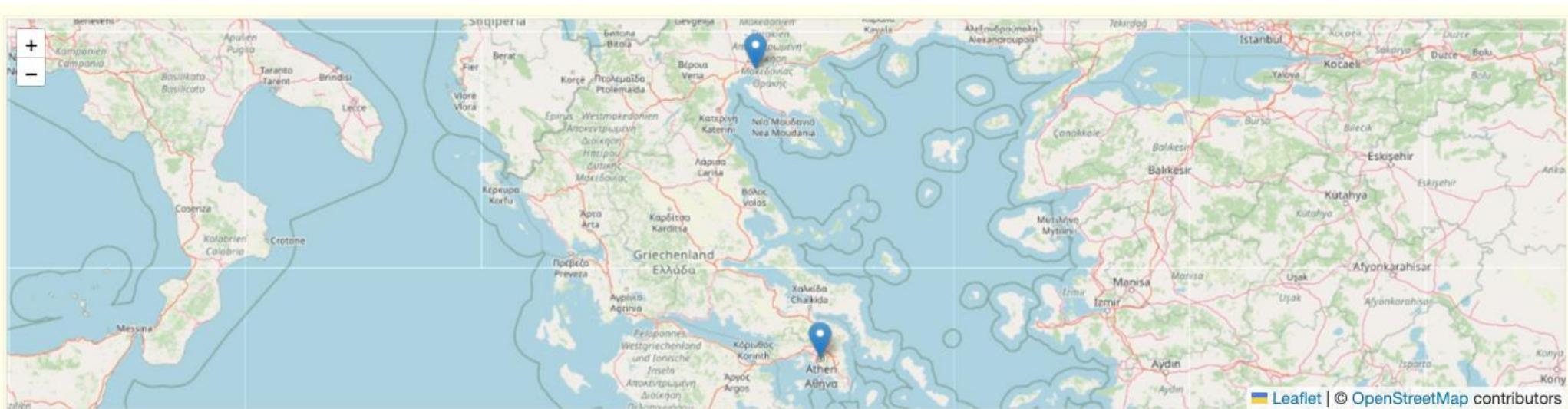
Comparison with The Session result (Q1-1)

SPARQL | HTML5 table

session

<https://thesession.org/sessions/7243>

<https://thesession.org/sessions/8020>



1. The Lucky Sparrow Irish Pub, Athens, Athens, Greece

Added by [Athens Irish Festival](#) 9 months ago.

2. The Dubliner, Thessaloniki, Makedonia, Greece

Added by [lukegarry](#) 3 years ago. Updated 1 month ago.

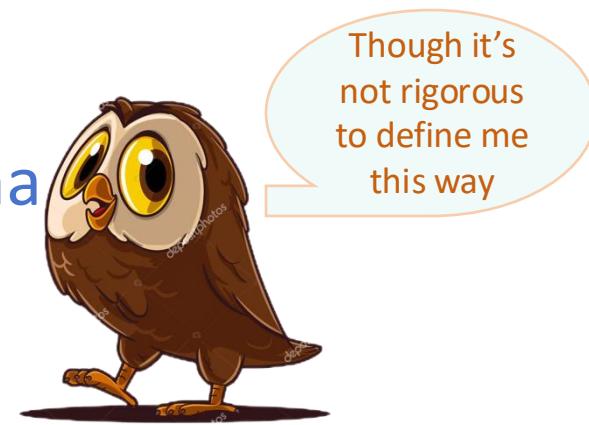
Search for sessions on in

SEARCH

Prompting ChatGPT to generate SPARQL query (Q1-1)

Using Ontology (OWL, Web Ontology Language) as a schema

Enter the following in ChatGPT (**the text in red will be provided to users**):



Given that the properties of an RDF database are reconciled with Wikidata,
and ` `

```
@prefix wd: <http://www.wikidata.org/entity/> .  
@prefix wdt: <http://www.wikidata.org/prop/direct/> .  
<https://thesession.org/sessions> rdf:type owl:Class .  
wdt:P2561 rdfs:label "name"; rdfs:domain <https://thesession.org/sessions> .  
wdt:P17 rdfs:domain <https://thesession.org/sessions>.  
` `
```

Make sure to generate a correct SPARQL for the following question:

Find all the sessions that took place in Greece

Generated SPARQL by ChatGPT (Q1-1)

Find all the sessions that took place in Greece

PREFIX wd: <<http://www.wikidata.org/entity/>>

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

SELECT DISTINCT ?session ?name

WHERE {

?session rdf:type <<https://thesession.org/sessions>> ;

 wdt:P17 wd:Q41 ; # Assuming wd:Q41 is the Wikidata entity for Greece

 wdt:P2561 ?name .

}

Searching on TheSession webpage (Q1-2)

<https://thesession.org/tunes>

The screenshot shows a web browser window with the URL <https://thesession.org/tunes> in the address bar. The page has a dark header with navigation links: Log in or Sign up, TUNES, DISCOGRAPHY, TRIPS, SESSIONS, EVENTS, and DISCUSSIONS. On the left, there's a yellow sidebar with 'THE SESSION' logo. The main content area is titled 'Tunes'. Below it is a search bar with dropdown menus: 'Search jigs' (selected), 'in Dmajor', and 'for Malcolm'. A yellow 'SEARCH' button is to the right. The main content below the search bar displays the results of the search: 'Searching jigs in Dmajor for “Malcolm”'. Two items are listed: '1. Malcolm’s New Fiddle (Jerry Holland’s) JIG' and '2. Malcolm Johnstone JIG'.

Log in or Sign up TUNES DISCOGRAPHY TRIPS SESSIONS EVENTS DISCUSSIONS

THE
SESSION

Tunes

Search in for

Searching jigs in Dmajor for “Malcolm”

1. [Malcolm’s New Fiddle \(Jerry Holland’s\) JIG](#)

2. [Malcolm Johnstone JIG](#)

Prompting ChatGPT to generate SPARQL query (Q1-2)

Enter the following in ChatGPT (**the text in red will be provided to users**):

Given an RDF database reconciled with Wikidata, and

...

wd:Q170412 rdf:type owl:Class ; rdfs:label "tune" .

wdt:P136 rdfs:type owl:DatatypeProperty ; rdfs:label "genre" .

wdt:P826 rdfs:type owl:DatatypeProperty ; rdfs:comment "This property is also used about the tonality of a tune etc" .

wdt:P2561 rdfs:label "name" .

...

Generate SPARQL for the following [Note: SPARQL be syntactically right such as

`filter(contains(lcase(str(?x)))` . For string values, please add @en]:

Find the tunes of genre “jig”, in “D major”, having “Malcolm” in its name

Generated SPARQL by ChatGPT (Q1-2)

Find the tunes of genre “jig”, in “D major”, having “Malcolm” in its name

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

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

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

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
```

```
SELECT ?tune ?name
```

```
WHERE {
```

```
  ?tune rdf:type wd:Q170412 ;
```

```
    wdt:P136 "jig"@en ;
```

```
    wdt:P826 "D major"@en ;
```

```
    wdt:P2561 ?name .
```

```
  FILTER (CONTAINS(LCASE(STR(?name)), "malcolm")) }
```

Copy SPARQL code then Execute (Q1-2)

Query Text

```
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

SELECT ?tune ?name
WHERE {
    ?tune rdf:type wd:Q170412 ;
        wdt:P136 "jig" ;
        wdt:P826 "D major" ;
        wdt:P2561 ?name .
    FILTER (CONTAINS(LCASE(STR(?name)), "malcolm"))
}
```

Results Format

HTML



Execute Query

Reset

The result from Virtuoso (Q1-2)

SPARQL | HTML5 table

tune	name
https://thesession.org/tunes/9347	"Malcolm's New Fiddle"@en
https://thesession.org/tunes/14626	"Malcolm Johnstone"@en

Comparison with The Session result (Q1-2)

SPARQL | HTML5 table

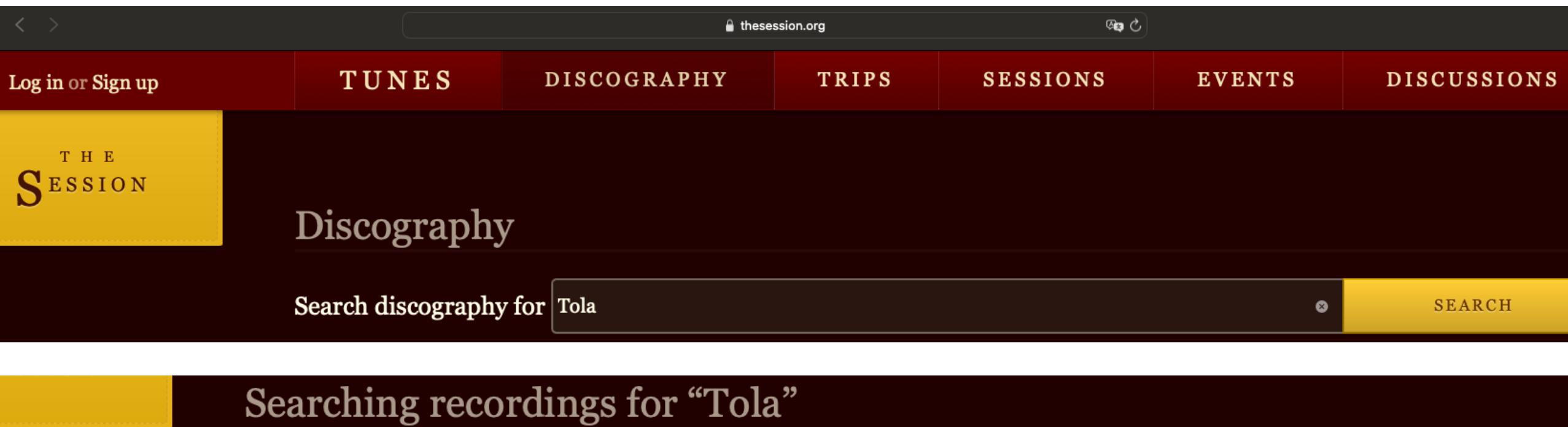
tune	name
https://thesession.org/tunes/9347	"Malcolm's New Fiddle"@en
https://thesession.org/tunes/14626	"Malcolm Johnstone"@en



Searching jigs in Dmajor for “Malcolm”

1. **Malcolm’s New Fiddle** (Jerry Holland’s)
JIG
2. **Malcolm Johnstone** JIG

Searching on TheSession webpage (Q1-3)



The screenshot shows the TheSession.org website. At the top, there is a navigation bar with links for 'Log in or Sign up', 'TUNES', 'DISCOGRAPHY' (which is the active tab), 'TRIPS', 'SESSIONS', 'EVENTS', and 'DISCUSSIONS'. Below the navigation bar, the page title 'Discography' is displayed. A search bar contains the text 'Search discography for Tola'. The main content area features the heading 'Searching recordings for “Tola”' followed by a list of three items:

- 1. Setting Free by Tola Custy And Cyril O'Donoghue
- 2. Three Sunsets by Mirella Murray And Tola Custy
- 3. Dísirt Tola : Traditional Music From Clare And Dublin by Various Artists

Searching recordings for “Tola”

1. Setting Free by Tola Custy And Cyril O'Donoghue
2. Three Sunsets by Mirella Murray And Tola Custy
3. Dísirt Tola : Traditional Music From Clare And Dublin by Various Artists

Search discography for Tola

SEARCH

Prompting ChatGPT to generate SPARQL query (Q1-3)

Enter the following in ChatGPT (**the text in red will be provided to users**):

Given an RDF database partially reconciled with Wikidata, and:

wd:Q482994 a owl:Class . wdt:P175 a owl:DatatypeProperty .

wdt:P2561.

Generate SPARQL for the following question:

[Note: (1)The value of wdt:P175 is only rdf:literal instead of a URI, so it mustn't be subject of other property.

(2)The subject of wdt:P2561 can only be instances of wd:Q482994.

(3)Use filter(contains(lcase(str(?x))). Any parts joined by UNION should be in {}.

(4)Don't forget namespace]

Question: Find albums containing “Tola” in its name or its artist’s name

Generated SPARQL by ChatGPT (Q1-3)

Find albums containing “Tola” in its name or its artist’s name

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

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

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

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
```

```
SELECT ?album ?name ?artistName WHERE { {
```

```
    ?album rdf:type wd:Q482994 ;
```

```
        wdt:P2561 ?name . FILTER(contains(lcase(str(?name)), "tola"))
```

```
} UNION {
```

```
    ?album rdf:type wd:Q482994 ;
```

```
        wdt:P175 ?artistName .
```

```
FILTER(contains(lcase(str(?artistName)), "tola"))}}
```

Copy SPARQL code then Execute (Q1-3)

Query Text

```
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
SELECT ?album ?name ?artistName WHERE {
    ?album rdf:type wd:Q482994 ;
        wdt:P2561 ?name . FILTER(contains(lcase(str(?name)), "tola"))
} UNION {
    ?album rdf:type wd:Q482994 ;
        wdt:P175 ?artistName .
    FILTER(contains(lcase(str(?artistName)), "tola"))}
```

Results Format

HTML



Execute Query

Reset

The result from Virtuoso (Q1-3)

SPARQL | HTML5 table

album	name	artistName
https://thesession.org/recording/4016	"Disirt Tola : Traditional Music From Clare And Dublin"@en	
https://thesession.org/recording/448		"Tola Custy And Cyril O'Donoghue"@en
https://thesession.org/recording/1525		"Mirella Murray And Tola Custy"@en

Comparison with The Session result (Q1-3)

SPARQL | HTML5 table

album	name	artistName
https://thesession.org/recording/4016	"Dísirt Tola : Traditional Music From Clare And Dublin"@en	
https://thesession.org/recording/448		"Tola Custy And Cyril O'Donoghue"@en
https://thesession.org/recording/1525		"Mirella Murray And Tola Custy"@en

Searching recordings for “Tola”

1. [Setting Free](#) by Tola Custy And Cyril O'Donoghue
2. [Three Sunsets](#) by Mirella Murray And Tola Custy
3. [Dísirt Tola : Traditional Music From Clare And Dublin](#) by Various Artists

Search discography for **Tola**



SEARCH

Searching on TheSession webpage (Q1-4)

The screenshot shows a web browser window for the website thesession.org. The page title is "Discography". A search bar at the top contains the text "Search discography for Mirella". The main content area displays two search results:

- 1. Three Sunsets by Mirella Murray And Tola Custy
- 2. Celtic Breeze by Celtic Breeze

Searching recordings for “Mirella”

1. [Three Sunsets](#) by [Mirella Murray And Tola Custy](#)

2. [Celtic Breeze](#) by [Celtic Breeze](#)

Prompting ChatGPT to generate SPARQL query (Q1-4)

Enter the following in ChatGPT (**the text in red will be provided to users**):

Given an RDF database partially reconciled with Wikidata, and:

...

wd:Q482994 rdf:type owl:Class .

ex:Shape a sh:NodeShape ; sh:targetClass wd:Q482994 ; sh:not [sh:property [sh:path rdfs:label]].

...

Generate SPARQL for the following question:

[Caution: Use filter(contains(lcase(str(?x))); “related to” means some of one’s attributes contains “Mirella”)]

Find albums related to “Mirella”

Generated SPARQL by ChatGPT (Q1-4)

Find albums related to “Mirella”

PREFIX wd: <<http://www.wikidata.org/entity/>>

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

PREFIX rdf: <<http://www.w3.org/1999/02/22-rdf-syntax-ns#>>

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

SELECT DISTINCT ?album ?attribute ?value WHERE {

?album rdf:type wd:Q482994 .

?album ?attribute ?value .

FILTER (contains(lcase(str(?value)), "mirella")).

}

Copy SPARQL code then Execute (Q1-4)

Query Text

```
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX sh: <http://www.w3.org/ns/shacl#>
```

```
SELECT DISTINCT ?album ?attribute ?value WHERE {
    ?album rdf:type wd:Q482994 .
    ?album ?attribute ?value .
    FILTER (contains(lcase(str(?value)), "mirella")).
```

```
}
```

Results Format

HTML



Execute Query

Reset

The result from Virtuoso (Q1-4)

SPARQL | HTML5 table

album	attribute	value
https://thesession.org/recording/3355	http://www.wikidata.org/prop/direct/P1625	"Mirella Murray's"@en
https://thesession.org/recording/1525	http://www.wikidata.org/prop/direct/P175	"Mirella Murray And Tola Custy"@en

Comparison with The Session result(Q1-4)

SPARQL | HTML5 table

album	attribute	value
https://thesession.org/recording/3355	http://www.wikidata.org/prop/direct/P1625	"Mirella Murray's"@en
https://thesession.org/recording/1525	http://www.wikidata.org/prop/direct/P175	"Mirella Murray And Tola Custy"@en

Searching recordings for “Mirella”

1. Three Sunsets by Mirella Murray And Tola Custy

2. Celtic Breeze by Celtic Breeze

A comprehensive context for several questions (from Q1-1 to Q1-4)

- We can merge the different contexts into a single prompt for ChatGPT to generate SPARQL queries for various questions, such as those previously mentioned:
- Q1-1: Find all the sessions that took place in Greece
- Q1-2: Find the tunes of genre “jig,” in “D major,” having “Malcolm” in its name
- Q1-3: Find albums containing “Tola” in its name or its artist’s name
- Q1-4: Find albums related to “Mirella”

The entire context for generating various SPARQL queries (Sec1)

Given that the entities of an RDF database are reconciled with Wikidata, and

```
``` @prefix wd: <http://www.wikidata.org/entity/> .  
 @prefix wdt: <http://www.wikidata.org/prop/direct/> .
 @prefix thesession: <https://thesession.org/> .
 thesession:sessions rdf:type owl:Class . wdt:P17
 wdt:P136 a owl:DatatypeProperty ; rdfs:label "genre" .
 wdt:P175 a owl:DatatypeProperty .
 wdt:P826 a owl:DatatypeProperty ; rdfs:label "mode" ;
 rdfs:comment "also used for the tonality of a tune" .
 wdt:P2561 rdfs:label "name" .
 rdfs:domain thesession:sessions .
 wd:Q170412 a owl:Class ;
 rdfs:label "tune" .
 wd:Q482994 a owl:Class .
 ex:Shape a sh:NodeShape ; sh:targetClass wd:Q482994 ; sh:not [sh:property [sh:path rdfs:label, wdt:P1476]].
 ````
```

Note:

1. In terms of SPARQL generation, it must be syntactically right such as:
 - (1) `filter(contains(lcase(str(?x)))`--Don't forget str() function
 - (2) Any parts joined by `UNION` should be in {}
 - (3) All the namespace prefixes should be defined for each SPARQL query
2. SPARQL generation should strictly adhere to the given context associated with OWL or SHACL, etc.
For example, distinguish between ObjectProperty and DataProperty!
As above, pay attention: wdt:P136, wdt:P175 and wdt:P826 are DataProperties! So, for example, the value of wdt:P175 are only strings in our context!
3. “related to” means some of one’s attributes contains “Mirella”
Make sure to generate correct SPARQL for the following
4. For string values, please add @en question:

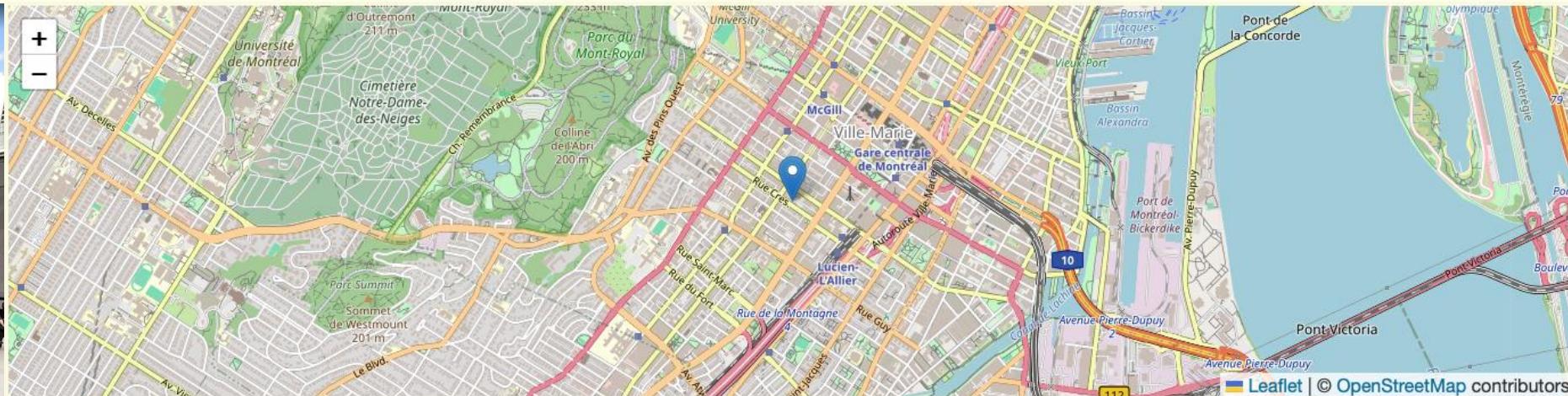
2. Queries only Executable Using
SPARQL on Virtuoso but not on
TheSession Webpage



- Q2-1: Find a session named "[Hurley's Irish Pub](#)"
- Q2-2: What genres appear in the recording whose artist is "Adam Agee & Jon Sousa"?
- Q2-3: Find the session that took place within a 20-kilometer radius centered on the session "[Hurley's Irish Pub](#)"
- Q2-4: 找到在希腊的演奏会 (mandarin / Chinese, translated as “Find sessions in Greece”)

[Log in or Sign up](#)[TUNES](#)[DISCOGRAPHY](#)[TRIPS](#)[SESSIONS](#)[EVENTS](#)[DISCUSSIONS](#)THE
SESSION

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](#)

Hurley's Irish Pub

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.](#)

Prompting ChatGPT to generate SPARQL query (Q2-1)

Enter the following in ChatGPT (**the text in red will be provided to users**):

Given an RDF database with properties(wdt:P2561) reconciled with Wikidata, generate SPARQL for the following (Use wdt:P2561 instead of rdfs:label. For string values, please add @en):

Find a session named "**Hurley's Irish Pub**"

Generated SPARQL by ChatGPT (Q2-1)

Find a session named "Hurley's Irish Pub"

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

SELECT ?session

WHERE {

?session wdt:P2561 "Hurley's Irish Pub"@en .

} Subject predicate object
 (property) (property value)

The result from Virtuoso (Q2-1)

SPARQL | HTML5 table

session

<https://thesession.org/sessions/411>

<https://thesession.org/sessions/411>

Prompting ChatGPT to generate SPARQL query (Q2-2)

Enter the following in ChatGPT (the text in red will be provided to users):

Given an RDF database reconciled with Wikidata(such as Q482994)

and ...

@PREFIX thesession: <<https://thesession.org/>> .

wd:Q170412 rdfs:label "tune" .

wd:Q482994 rdfs:label "album" .

wd:Q483394 rdfs:label "genre" .

wdt:P175 rdf:type owl:DatatypeProperty . #even if it's an

ObjectProperty in Wikidata, here we take it as DataProperty

thesession:tunes rdfs:domain wd:Q482994 ;

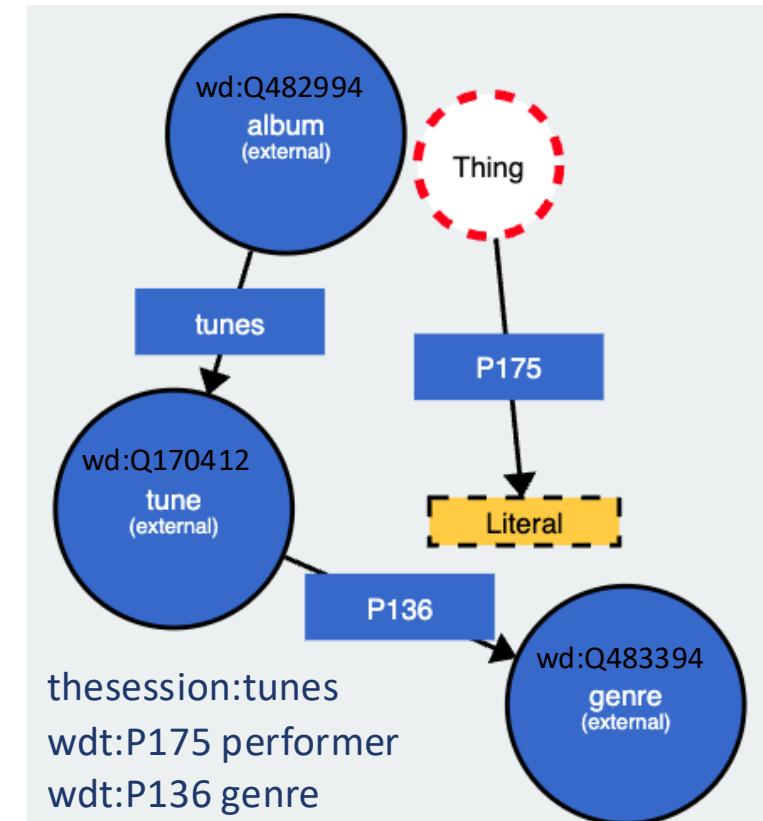
 rdfs:range wd:Q170412 .

wdt:P136 rdfs:domain wd:Q170412 ;

 rdfs:range wd:Q483394

Generate SPARQL for the following [note: No need to return
the label of an entity. For string property values, please add @en]:

What genres appear in the album whose artist is "Adam Agee & Jon Sousa"?



Generated SPARQL by ChatGPT (Q2-2)

What genres appear in the album whose artist is "Adam Agee & Jon Sousa"?

PREFIX thesession: <<https://thesession.org/>>

PREFIX wd: <<http://www.wikidata.org/entity/>>

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

SELECT DISTINCT ?genre

WHERE {

Match the album of the artist "Adam Agee & Jon Sousa"

?album wdt:P175 "Adam Agee & Jon Sousa"@en .

Find all tunes in the matched albums

?album thesession:tunes ?tune .

Find genres of the matched tunes

?tune wdt:P136 ?genre . }

The result from Virtuoso (Q2-2)

SPARQL | HTML5 table

genre

<http://www.wikidata.org/entity/Q118120>

"reel"

<http://www.wikidata.org/entity/Q1079270>

"jig"

<http://www.wikidata.org/entity/Q211025>

"march"

<http://www.wikidata.org/entity/Q8701407>

"waltz"

Prompting ChatGPT to generate SPARQL query (Q2-3)

Enter the following in ChatGPT (**the text in red will be provided to users**):

Given an RDF database reconciled with Wikidata and

...

wdt:P625 rdfs:domain <<https://thesession.org/sessions>> .

wdt:P2561 rdfs:label “**sessionName**” ; rdfs:domain <<https://thesession.org/sessions>> .

...

Use **bif:st_distance()**. Remember to confine the type for entities .

Generate SPARQL for the following (For string property values, please add @en):

Find other sessions that took place within a 20-kilometer radius centered on the session "[Hurley's Irish Pub](#)"

Generated SPARQL by ChatGPT (Q2-3)

Find other sessions that took place within a 20-kilometer radius centered on the session "Hurley's Irish Pub"

```
PREFIX wdt: <http://www.wikidata.org/prop/direct/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT ?session ?sessionName ?location ?distance
WHERE {
  # Find the location of "Hurley's Irish Pub"
  ?centerSession wdt:P2561 "Hurley's Irish Pub"@en ;
    wdt:P625 ?centerLocation .
  # Find other sessions and their locations
  ?session a <https://thesession.org/sessions> ;
    wdt:P2561 ?sessionName ;
    wdt:P625 ?location .
  # Calculate distance between "Hurley's Irish Pub" and other sessions
  BIND(bif:st_distance(?centerLocation, ?location) AS ?distance)
  # Filter sessions within a 20-kilometer radius
  FILTER(?distance <= 20)} ORDER BY ?distance
```

The result from Virtuoso (Q2-3)

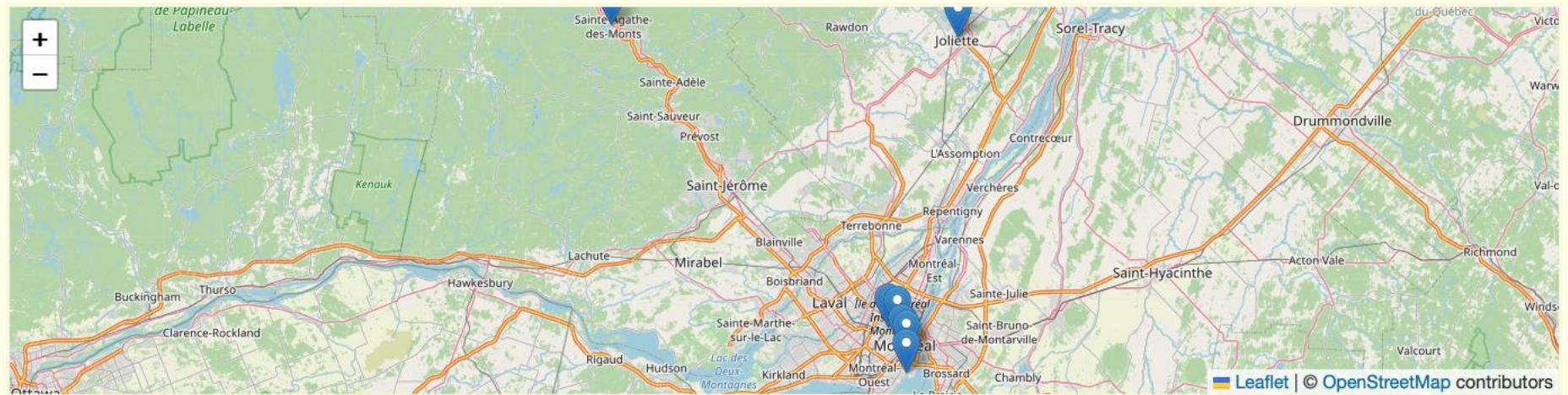
SPARQL | HTML5 table

| session | sessionName | location | distance |
|---|-------------------------|--|----------|
| https://thesession.org/sessions/411 | "Hurley's Irish Pub"@en | "POINT(45.4969101 -73.57475281)"^^< http://www.openlinksw.com/schemas/virtrdf#Geometry > | 0.0 |
| https://thesession.org/sessions/410 | "McKibbin's"@en | "POINT(45.49675751 -73.57733154)"^^< http://www.openlinksw.com/schemas/virtrdf#Geometry > | 0.286621 |
| https://thesession.org/sessions/8805 | "Lord William Pub"@en | "POINT(45.48695374 -73.5670166)"^^< http://www.openlinksw.com/schemas/virtrdf#Geometry > | 0.91493 |
| https://thesession.org/sessions/2779 | "Les Îles En Ville"@en | "POINT(45.45341873 -73.56800842)"^^< http://www.openlinksw.com/schemas/virtrdf#Geometry > | 1.55896 |
| https://thesession.org/sessions/7202 | "La Petite Marche"@en | "POINT(45.52675629 -73.58866119)"^^< http://www.openlinksw.com/schemas/virtrdf#Geometry > | 1.80777 |
| https://thesession.org/sessions/2780 | "Aux Vices & Versa"@en | "POINT(45.53131866 -73.61063385)"^^< http://www.openlinksw.com/schemas/virtrdf#Geometry > | 4.13126 |

geographical coordinates

[Log in or Sign up](#)[TUNES](#)[DISCOGRAPHY](#)[TRIPS](#)[SESSIONS](#)[EVENTS](#)[DISCUSSIONS](#)THE
SESSION

Sessions near Montreal, Quebec



[nearby members](#)

[nearby sessions](#)

[nearby events](#)

A similiar result from
TheSession

[https://thesession.org/
sessions/411](https://thesession.org/sessions/411)

Hurley's Irish Pub, Montreal, Quebec, Canada

Added by Loa 21 years ago. Updated 9 months ago.

McKibbin's, Montreal, Quebec, Canada

Added by Loa 21 years ago. Updated 2 months ago.

Lord William Pub, Montreal, Quebec, Canada

Added by Jeremy 6 months ago.

La Petite Marche, Montreal, Quebec, Canada

Added by Patrice Bouchertere 3 years ago.

Aux Vices & Versa, Montreal, Quebec, Canada

Added by cpbmtl 14 years ago. Updated 2 years ago.

Les Îles En Ville, Montreal, Quebec, Canada

Added by cpbmtl 14 years ago. Updated 2 years ago.

Brasserie Artisanale Albion, Joliette, Quebec, Canada

Added by Dana Whittle 11 years ago. Updated 6 years ago.

Le P'tit Refrain Du Nord, Sainte-Agathe-des-Monts, Quebec, Canada

Added by Geneviève Poulin 2 years ago.

Prompting ChatGPT to generate SPARQL query (Q2-4)

Enter the following in ChatGPT (**the text in red will be provided to users**):

Given an RDF database reconciled with Wikidata, and

```

```
@prefix wd: <http://www.wikidata.org/entity/> .
@prefix wdt: <http://www.wikidata.org/prop/direct/> .
<https://thesession.org/sessions> a owl:Class .
wdt:P2561 rdfs:label "name";
 rdfs:domain <https://thesession.org/sessions> .
wdt:P17 rdfs:domain <https://thesession.org/sessions>.
``` (use wdt:P2561 instead of rdfs:label)
```

Make sure to generate a correct SPARQL for the following question:

找到在希腊的演奏会 (mandarin / Chinese, translated as “Find sessions in Greece”)

Generated SPARQL by ChatGPT (Q2-4)

找到在希腊的演奏会 (mandarin / Chinese, translated as “Find sessions in Greece”)

PREFIX wd: <<http://www.wikidata.org/entity/>>

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

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

SELECT ?session ?name

WHERE {

?session a <<https://thesession.org/sessions>> ;

 wdt:P17 wd:Q41 ; # Q41 is the Wikidata entity ID for Greece

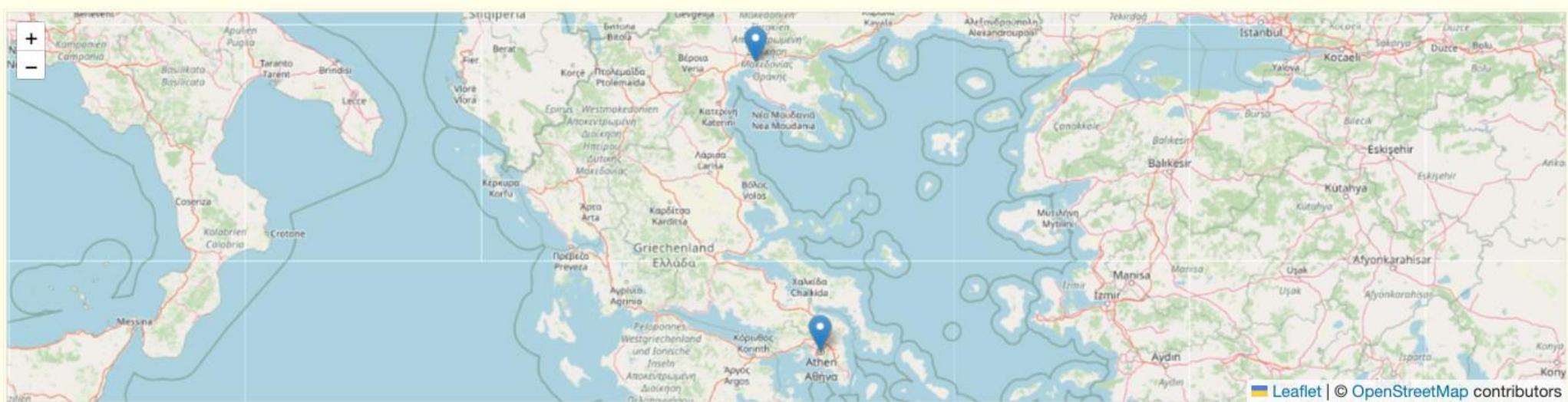
 wdt:P2561 ?name .

}

Comparison with The Session result (Q2-4)

SPARQL | HTML5 table

session	name
https://thesession.org/sessions/7243	"The Dubliner"
https://thesession.org/sessions/8020	"The Lucky Sparrow Irish Pub"



1. The Lucky Sparrow Irish Pub, Athens, Athens, Greece

Added by [Athens Irish Festival](#) 9 months ago.

2. The Dubliner, Thessaloniki, Makedonia, Greece

Added by [lukegarry](#) 3 years ago. Updated 1 month ago.

Search for sessions on in

SEARCH

A comprehensive context for several questions (from Q2-1 to Q2-4)

- We can merge the different contexts into a single prompt for ChatGPT to generate SPARQL queries for various questions, such as those previously mentioned:
- Q2-1: Find a session named "Hurley's Irish Pub"
- Q2-2: What genres appear in the recording whose artist is "Adam Agee & Jon Sousa"?
- Q2-3: Find the session that took place within a 20-kilometer radius centered on the session "Hurley's Irish Pub"
- Q2-4: 找到在希腊的演奏会 (mandarin / Chinese, translated as Find sessions in Greece)

The entire context for generating various SPARQL queries (Sec2)

Given an RDF database with properties reconciled with Wikidata, and

```
@PREFIX thesession: <https://thesession.org/> .  
@PREFIX wd: <http://www.wikidata.org/entity/> .  
@PREFIX wdt: <http://www.wikidata.org/prop/direct/> .  
thesession:sessions a owl:Class .  
thesession:tunes rdfs:domain wd:Q482994 ;  
rdfs:range wd:Q170412 ;  
rdfs:comment "For albums have tunes" .  
wd:Q170412 rdfs:label "tune" .  
wd:Q482994 rdfs:label "album" .  
wd:Q483394 rdfs:label "genre" .  
wdt:P17 rdfs:domain thesession:sessions .  
wdt:P175 rdf:type owl:DatatypeProperty .  
wdt:P136 rdfs:domain wd:Q170412 ;  
rdfs:range wd:Q483394 .  
wdt:P625 rdfs:domain thesession:sessions .  
wdt:P2561 rdfs:label "sessionName" ; rdfs:domain thesession:sessions .
```

- Before generating SPARQL, note:
1. No need to return the label of an entity
 2. Use `bif:st_distance()`, the unit in which is kilometer and `bif:` is not a namespace prefix
 3. Remember to confine the type for entities
 4. Don't forget the namespace prefix for each SPARQL query!!
 5. Don't use property that's not provided above
 6. Albums don't relate genres directly, but Albums have tunes which relate genres
 7. Even if `wdt:P175` is an ObjectProperty in Wikidata, here we take it as DataProperty
 8. For string property values, please add `@en`

Generate correct SPARQL for the following question:

3. Queries across TheSession and Wikidata

Some information is not available solely through TheSession. Given that we have already done reconciliation with Wikidata, we can extend our query to external data sources (e.g., Wikidata) for more specific questions such as...



- Q3-1: Which sessions took place **in Eastern Europe?**
- Q3-2: From TheSession, please find recordings by Altan, which were known to be **released between 2007 and 2012**
- Q3-3: Find in TheSession performers who **are Canadians**. And find the recordings they performed in TheSession

The information about the green part comes from Wikidata.

Prompting ChatGPT to generate SPARQL query (Q3-1)

Enter the following in ChatGPT (**the text in red will be provided to users**):

Given an RDF graph <<http://sample/thesession/reconciled>>, the properties and classes of instances are reconciled with Wikidata as much as possible, so regarding SPARQL queries across different databases, we use `graph <...>` and `service <...>`. and Given:

@prefix thesession: <<https://thesession.org/>>. thesession:sessions a owl:Class.

wdt:P17 rdfs:domain <<https://thesession.org/sessions>>.

wdt:P2561 rdfs:domain <<https://thesession.org/sessions>>.

--**for graph** <<http://sample/thesession/reconciled>>

wdt:P361 rdfs:domain <[ex:country](#)>. wd:Q27468 rdfs:label “Eastern Europe”.

--**for service** <<https://query.wikidata.org/sparql>>

Make sure to generate a correct SPARQL for the following question (Note: Use the properties or classes as above primarily; traverse wikidata before local graph):

Which sessions took place in Eastern Europe?

Generated SPARQL by ChatGPT (Q3-1)

Which sessions took place in Eastern Europe?

PREFIX thesession: <<https://thesession.org/>>

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

PREFIX wd: <<http://www.wikidata.org/entity/>>

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

SELECT ?session ?location WHERE {

```
SERVICE <https://query.wikidata.org/sparql> {  
    ?location wdt:P361 wd:Q27468 . # Q27468 is the Wikidata item for Eastern Europe  
}
```

```
GRAPH <http://sample/thesession/reconciled> {  
    ?session wdt:P17 ?location .  
    ?session a thesession:sessions . }}}
```



The result from Virtuoso (Q3-1)

session	location		
https://thesession.org/sessions/1218	http://www.wikidata.org/entity/Q36	https://thesession.org/sessions/6721	http://www.wikidata.org/entity/Q159
https://thesession.org/sessions/1896	http://www.wikidata.org/entity/Q213	https://thesession.org/sessions/6918	http://www.wikidata.org/entity/Q36
https://thesession.org/sessions/1915	http://www.wikidata.org/entity/Q212	https://thesession.org/sessions/7549	http://www.wikidata.org/entity/Q36
https://thesession.org/sessions/2206	http://www.wikidata.org/entity/Q159	https://thesession.org/sessions/8013	http://www.wikidata.org/entity/Q213
https://thesession.org/sessions/2521	http://www.wikidata.org/entity/Q36	https://thesession.org/sessions/8014	http://www.wikidata.org/entity/Q213
https://thesession.org/sessions/2760	http://www.wikidata.org/entity/Q213	https://thesession.org/sessions/8731	http://www.wikidata.org/entity/Q159
https://thesession.org/sessions/3241	http://www.wikidata.org/entity/Q36	https://thesession.org/sessions/8775	http://www.wikidata.org/entity/Q219
https://thesession.org/sessions/3452	http://www.wikidata.org/entity/Q213	https://thesession.org/sessions/8806	http://www.wikidata.org/entity/Q213
https://thesession.org/sessions/3631	http://www.wikidata.org/entity/Q184	https://thesession.org/sessions/8823	http://www.wikidata.org/entity/Q159
https://thesession.org/sessions/5941	http://www.wikidata.org/entity/Q36	https://thesession.org/sessions/8824	http://www.wikidata.org/entity/Q159
		https://thesession.org/sessions/8905	http://www.wikidata.org/entity/Q213

Q36: Poland, Q159: Russia, Q184: Belarus, Q213: Czech Republic, Q219: Bulgaria

Prompting ChatGPT to generate SPARQL query (Q3-2)

Enter the following in ChatGPT (**the text in red will be provided to users**):

Given an RDF graph <<http://sample/thesession/reconciled>>, the properties and classes of instances are reconciled with Wikidata as much as possible, so regarding SPARQL queries across different databases, we use `graph <...>` and `service <...>`. and Given:

For graph <<http://sample/thesession/reconciled>> :

wd:Q482994 a owl:Class. wdt:P175 rdfs:domain wd:Q482994; a owl:DataProperty.

For service <<https://query.wikidata.org/sparql>> :

wdt:P577 rdfs:range xsd:dateTime.

Make sure to generate a correct SPARQL for the following question

(Note: Use the properties or classes as above primarily; Even if on Wikidata, wdt:P175 is an ObjectProperty, we see it as a DataProperty here; Use such as `?sessionRecording wdt:P2888 ?wikidataRecording .`; For string property values, please add @en):

From TheSession, please find recordings by Altan, which were known to be released between 2007 and 2012

Generated SPARQL by ChatGPT (Q3-2)

From TheSession, please find recordings by Altan, which were known to be released between 2007 and 2012

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

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

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

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
```

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
```

```
SELECT ?sessionRecording ?wikidataRecording ?releaseDate WHERE {
```

```
GRAPH <http://sample/thesession/reconciled> { ?sessionRecording a wd:Q482994 ;
```

```
        wdt:P175 "Altan"@en .
```

```
        ?sessionRecording wdt:P2888 ?wikidataRecording . }
```

```
SERVICE <https://query.wikidata.org/sparql> { ?wikidataRecording wdt:P577 ?releaseDate .
```

```
FILTER (?releaseDate >= "2007-01-01T00:00:00Z"^^xsd:dateTime && ?releaseDate <= "2012-12-31T23:59:59Z"^^xsd:dateTime) }
```

The result from Virtuoso (Q3-2)

SPARQL | HTML5 table (faceted browsing links)

sessionRecording	wikidataRecording	releaseDate
https://thesession.org/recordings/3560	http://www.wikidata.org/entity/Q4632227	2010-01-01T00:00:00Z
https://thesession.org/recordings/4142	http://www.wikidata.org/entity/Q5567392	2012-01-01T00:00:00Z

Prompting ChatGPT to generate SPARQL query (Q3-3-0)

Enter the following in ChatGPT (**the text in red will be provided to users**):

Given an RDF graph <<http://sample/thesession/reconciled>>, the properties and classes of instances are reconciled with Wikidata as much as possible, so regarding SPARQL queries across different databases, we use `graph <...>` and `service <...>`. and Given:

The value of wdt:P175 is performer

--for graph <<http://sample/thesession/reconciled>>

wdt:P175 performer

wdt:P27

wdt:P27 country of citizenship

--for service <<https://query.wikidata.org/sparql>>

Make sure to generate a correct SPARQL for the following question:

Find in TheSession performers who are Canadians ...

Expected SPARQL (Q3-3-0)

Find in TheSession performers who are Canadians

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

PREFIX wd: <<http://www.wikidata.org/entity/>>

SELECT ?performer

WHERE {

GRAPH <<http://sample/thesession/reconciled>> {

?x a wd:Q482994 ;

wdt:P175 ?performer .

FILTER isIRI(?performer)

}

SERVICE <<https://query.wikidata.org/sparql>> {

?performer wdt:P27 wd:Q16

}

performer

<http://www.wikidata.org/entity/Q725953>

<http://www.wikidata.org/entity/Q6968206>

<http://www.wikidata.org/entity/Q7417356>

<http://www.wikidata.org/entity/Q6968206>

<http://www.wikidata.org/entity/Q7417356>

<http://www.wikidata.org/entity/Q6183761>

<http://www.wikidata.org/entity/Q645562>

<http://www.wikidata.org/entity/Q4960771>

<http://www.wikidata.org/entity/Q6183761>

<http://www.wikidata.org/entity/Q6183761>

<http://www.wikidata.org/entity/Q6968206>

<http://www.wikidata.org/entity/Q6968206>

<http://www.wikidata.org/entity/Q6968206>

<http://www.wikidata.org/entity/Q6968206>

<http://www.wikidata.org/entity/Q6183761>

<http://www.wikidata.org/entity/Q725953>

<http://www.wikidata.org/entity/Q7417356>

<http://www.wikidata.org/entity/Q6183761>

<http://www.wikidata.org/entity/Q6183761>

<http://www.wikidata.org/entity/Q4960771>

<http://www.wikidata.org/entity/Q6968206>

<http://www.wikidata.org/entity/Q6183761>

Prompting ChatGPT to generate SPARQL query (Q3-3)

Enter the following in ChatGPT (**the text in red will be provided to users**):

Given an RDF graph <<http://sample/thesession/reconciled>>, the properties and classes of instances are reconciled with Wikidata as much as possible, so regarding SPARQL queries across different databases, we use

(1) `graph <<http://sample/thesession/reconciled>>` with:

wd:Q482994 rdfs:label “recording”. wdt:P175 # the value of it can be either rdfs:Literal or URI

(2) `service <<https://query.wikidata.org/sparql>>` with:

wdt:P27

[Note: Use isURI(?performer); Don’t use `rdfs:label` as a predicate in SPARQL anyway!]

Make sure to generate a correct SPARQL for the following question:

Find in TheSession performers who are Canadians. And find the recordings they performed in TheSession

Generated SPARQL by ChatGPT (Q3-3)

Find in TheSession performers who are Canadians. And find the recordings they performed in TheSession

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

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

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

```
SELECT ?performer ?recording WHERE {
```

```
GRAPH <http://sample/thesession/reconciled> {
```

```
?recording wdt:P175 ?performer .
```

```
}
```

```
SERVICE <https://query.wikidata.org/sparql> {
```

```
?performer wdt:P27 wd:Q16 .
```



```
}
```

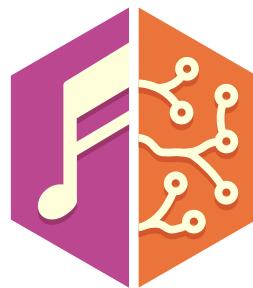
```
FILTER(isIRI(?performer)) }
```

Note: It probably takes a long time to get the result from running the code on localhost; sometimes, it reports timeout.
Recently, there is timeout on staging

The result from Virtuoso (Q3-3)

performer	recording
http://www.wikidata.org/entity/Q725953	https://thesession.org/recording/5709
http://www.wikidata.org/entity/Q6968206	https://thesession.org/recording/662
http://www.wikidata.org/entity/Q7417356	https://thesession.org/recording/6621
http://www.wikidata.org/entity/Q6968206	https://thesession.org/recording/6747
http://www.wikidata.org/entity/Q7417356	https://thesession.org/recording/7463
http://www.wikidata.org/entity/Q6183761	https://thesession.org/recording/7677
http://www.wikidata.org/entity/Q645562	https://thesession.org/recording/3369
http://www.wikidata.org/entity/Q4960771	https://thesession.org/recording/3617
http://www.wikidata.org/entity/Q6183761	https://thesession.org/recording/4075
http://www.wikidata.org/entity/Q6183761	https://thesession.org/recording/4190
http://www.wikidata.org/entity/Q6968206	https://thesession.org/recording/427
http://www.wikidata.org/entity/Q6968206	https://thesession.org/recording/430
http://www.wikidata.org/entity/Q6968206	https://thesession.org/recording/434
http://www.wikidata.org/entity/Q6968206	https://thesession.org/recording/5059
http://www.wikidata.org/entity/Q6183761	https://thesession.org/recording/560
http://www.wikidata.org/entity/Q725953	https://thesession.org/recording/1215
http://www.wikidata.org/entity/Q7417356	https://thesession.org/recording/1606
http://www.wikidata.org/entity/Q6183761	https://thesession.org/recording/1664
http://www.wikidata.org/entity/Q6183761	https://thesession.org/recording/1704

4. Queries across TheSession, MusicBrainz and Wikidata



- Q4-1: How many / Which recording performers on TheSession (who) are **not** on MusicBrainz?
- Q4-2: List **top five most popular record labels** used by musicians on TheSession?
- Q2-3: List all albums **recorded by bands** in TheSession, that was **released in 2003 in America**

Prompting ChatGPT to generate SPARQL query (Q4-1-1)

Enter the following in ChatGPT (**the text in red will be provided to users**):

Given RDF graphs <<http://sample/thesession/reconciled>> and <<http://musicbrainz/reconciled>>, the properties and classes of instances are reconciled with Wikidata as much as possible, so regarding SPARQL queries across different databases, we use

(1) `graph <<http://sample/thesession/reconciled>>` with:

wdt:P175 rdfs:range ex:performer.

(2) `graph <<http://musicbrainz/reconciled>>` with:

wdt:P2888 rdfs:range ex:performer.

[note: DISTINCT; The value of performer must be IRI, that should be embodied earlier in SPARQL, please use FILTER isIRI())]

Make sure to generate a correct SPARQL for the following question:

How many recording performers on TheSession are **not** on MusicBrainz?

Generated SPARQL by ChatGPT (Q4-1-1)

How many recording performers on TheSession are not on MusicBrainz?

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

```
SELECT (COUNT(DISTINCT ?performer)
```

```
AS ?performerCount) WHERE {
```

```
GRAPH <http://sample/thesession/reconciled>
```

```
{?recording wdt:P175 ?performer .
```

```
FILTER isIRI(?performer)
```

```
}
```

```
FILTER NOT EXISTS {
```

```
GRAPH <http://musicbrainz/reconciled> {
```

```
?otherRecording wdt:P2888 ?performer .
```

```
FILTER isIRI(?performer)
```

```
}}}
```

The result from Virtuoso (Q4-1-1)

SPARQL | HTML5 table

performerCount

15

Note: Why is the result such a small number? This is because only a small fraction of artists/performers from TheSession have been reconciled with Wikidata

Prompting ChatGPT to generate SPARQL query (Q4-1-2)

Enter the following in ChatGPT (**the text in red will be provided to users**):

Given RDF graphs <<http://sample/thesession/reconciled>> and <<http://musicbrainz/reconciled>>, the properties and classes of instances are reconciled with Wikidata as much as possible, so regarding SPARQL queries across different databases, we use

(1) `graph <<http://sample/thesession/reconciled>>` with:

wdt:P175 rdfs:range ex:performer.

(2) `graph <<http://musicbrainz/reconciled>>` with:

wdt:P2888 rdfs:range ex:performer.

[note: DISTINCT; The value of ex:performer must be IRI, that should be embodied in SPARQL]

Make sure to generate a correct SPARQL for the following question:

Which recording performers on TheSession **are not on MusicBrainz**?

Generated SPARQL by ChatGPT (Q4-1-2)

Which recording performers on TheSession are not on MusicBrainz?

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

```
SELECT DISTINCT ?performer
```

```
WHERE {
```

```
GRAPH <http://sample/thesession/reconciled> {
```

```
?recording wdt:P175 ?performer .
```

```
}
```

```
MINUS {
```

```
GRAPH <http://musicbrainz/reconciled> {
```

```
?otherRecording wdt:P2888 ?performer .
```

```
}
```

```
FILTER(isIRI(?performer)) }
```

The result from Virtuoso (Q4-1-2)

SPARQL | HTML5 table

<http://www.wikidata.org/entity/Q101325026>
<http://www.wikidata.org/entity/Q121384360>
<http://www.wikidata.org/entity/Q2198713>
<http://www.wikidata.org/entity/Q2286403>
<http://www.wikidata.org/entity/Q2784150>
<http://www.wikidata.org/entity/Q3418793>
<http://www.wikidata.org/entity/Q35489573>
<http://www.wikidata.org/entity/Q3860276>
<http://www.wikidata.org/entity/Q539531>
<http://www.wikidata.org/entity/Q594693>
<http://www.wikidata.org/entity/Q7171905>
<http://www.wikidata.org/entity/Q7417356>
<http://www.wikidata.org/entity/Q7490578>
<http://www.wikidata.org/entity/Q7557090>
<http://www.wikidata.org/entity/Q93448268>

...

Prompting ChatGPT to generate SPARQL query (Q4-2)

Enter the following in ChatGPT (the text in red will be provided to users):

Given RDF graphs <<http://sample/thesession/reconciled>> and <<http://musicbrainz/reconciled>>, the properties and classes of instances are reconciled with Wikidata as much as possible.

1. For graph <<http://sample/thesession/reconciled/noLangLabel>>:

wd:Q482994 a owl:Class; rdfs:label “Record”. wdt:P2561 rdfs:label “name”.

2. For graph <<http://musicbrainz/reconciled>>

wd:Q3972943 a owl:Class; rdfs:label “Record”.

wdt:P1476 rdfs:label “title”; rdfs:domain wd:Q3972943.

wdt:P264 rdfs:label “record label”; rdfs:domain wd:Q3972943; rdfs:range wd:Q18127.

wdt:P2561 rdfs:domain wd:Q18127; rdfs:comment “It’s the name of ‘record label’. Don’t use rdfs:label as a predicate in the triple pattern of the generated SPARQL!”!!!

wd:Q18127 rdfs:label “Label”.

Make sure to generate an absolutely correct SPARQL for the following question

(Note: Primarily use the properties or classes mentioned above for each graph. You can use wdt:P2561 of wd:Q482994 to match wdt:P1476 of wd:Q3972943. Return both the URI of the record labels and the name of record labels): *List top five most popular record labels used by musicians on TheSession? ->deceptive, Modification:*

List the top five most popular labels from MusicBrainz whose records are also on TheSession database

Generated SPARQL by ChatGPT (Q4-2)

List the top five most popular labels from MusicBrainz whose records are also on TheSession database.

```
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT ?labelURI ?labelName (COUNT(?record) AS ?recordCount) WHERE {
    # Records from TheSession
    GRAPH <http://sample/thesession/reconciled/noLangLabel> { ?record a wd:Q482994.
        ?record wdt:P2561 ?recordName. }
    # Matching Records from MusicBrainz
    GRAPH <http://musicbrainz/reconciled> { ?musicBrainzRecord a wd:Q3972943.
        ?musicBrainzRecord wdt:P1476 ?recordName.
    # Link records to their labels
    ?musicBrainzRecord wdt:P264 ?labelURI. ?labelURI a wd:Q18127. ?labelURI wdt:P2561 ?labelName. }
    GROUP BY ?labelURI ?labelName ORDER BY DESC(?recordCount) LIMIT 5
```



List top five most popular record labels used by musicians on TheSession?

The result from Virtuoso (Q4-2)



SPARQL | HTML5 table

labelURI	labelName	recordCount
https://musicbrainz.org/label/011d1192-6f65-45bd-85c4-0400dd45693e	"Columbia"	62
https://musicbrainz.org/label/8f638ddb-131a-4cc3-b3d4-7ebdac201b55	"Epic"	49
https://musicbrainz.org/label/1b3b52a5-ef89-4f8b-8d5f-e15a7a58863b	"Vertigo"	33
https://musicbrainz.org/label/29d7c88f-5200-4418-a683-5c94ea032e38	"BMG"	29
https://musicbrainz.org/label/da631842-b2b2-4b65-a526-592f1b776c18	"Supersonic Records"	28

Prompting ChatGPT to generate SPARQL query (Q4-3)

Enter the following in ChatGPT (**the text in red will be provided to users**):

Given 2 RDF graphs, the properties, instances, classes of which are reconciled with Wikidata as much as possible, regarding SPARQL queries across different databases, refer to:

album **wd:Q482994** owl:equivalentClass **wd:Q3972943**. The values of wdt:P175 have property wdt:P31 to indicate whether the performer is a band.

release

1. For `GRAPH <<http://sample/thesession/reconciled/noLangLabel>>`:

wd:Q482994 a owl:Class; rdfs:label “Album”.

wdt:P175 rdfs:label “performer”; rdfs:domain wd:Q482994. wdt:P2561 rdfs:label “name”.

2. For `GRAPH <<http://musicbrainz/reconciled>>`:

wd:Q3972943 a owl:Class; rdfs:label “Release”.

wdt:P17 rdfs:label “country”. wdt:P585 rdfs:label “date”. wdt:P1476 rdfs:label “title”.

3. For `SERVICE <<https://query.wikidata.org/sparql>>`:

wd:Q215380 rdfs:label “bands”. wdt:P31

Make sure to generate a correct SPARQL for the following question [note: match value of P2561 and of P1476 in order to match the instances of wd:Q482994 and of wd:Q3972943 (Please use `FILTER(?date=2003)` in SPARQL)]:

List all albums **recorded by bands** in TheSession, that was **released in 2003** in America

Generated SPARQL by ChatGPT (Q4-3)

List all albums recorded by bands in TheSession, that was released in 2003 in America

PREFIX owl: <<http://www.w3.org/2002/07/owl#>> PREFIX rdfs: <<http://www.w3.org/2000/01/rdf-schema#>>

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

SELECT ?albumName ?album ?performer

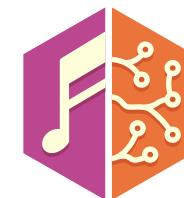
WHERE { GRAPH <<http://sample/thesession/reconciled/noLangLabel>> {

?album a wd:Q482994 ; #recording of TheSession
wdt:P2561 ?albumName ;
wdt:P175 ?performer . }



GRAPH <<http://musicbrainz/reconciled>> {

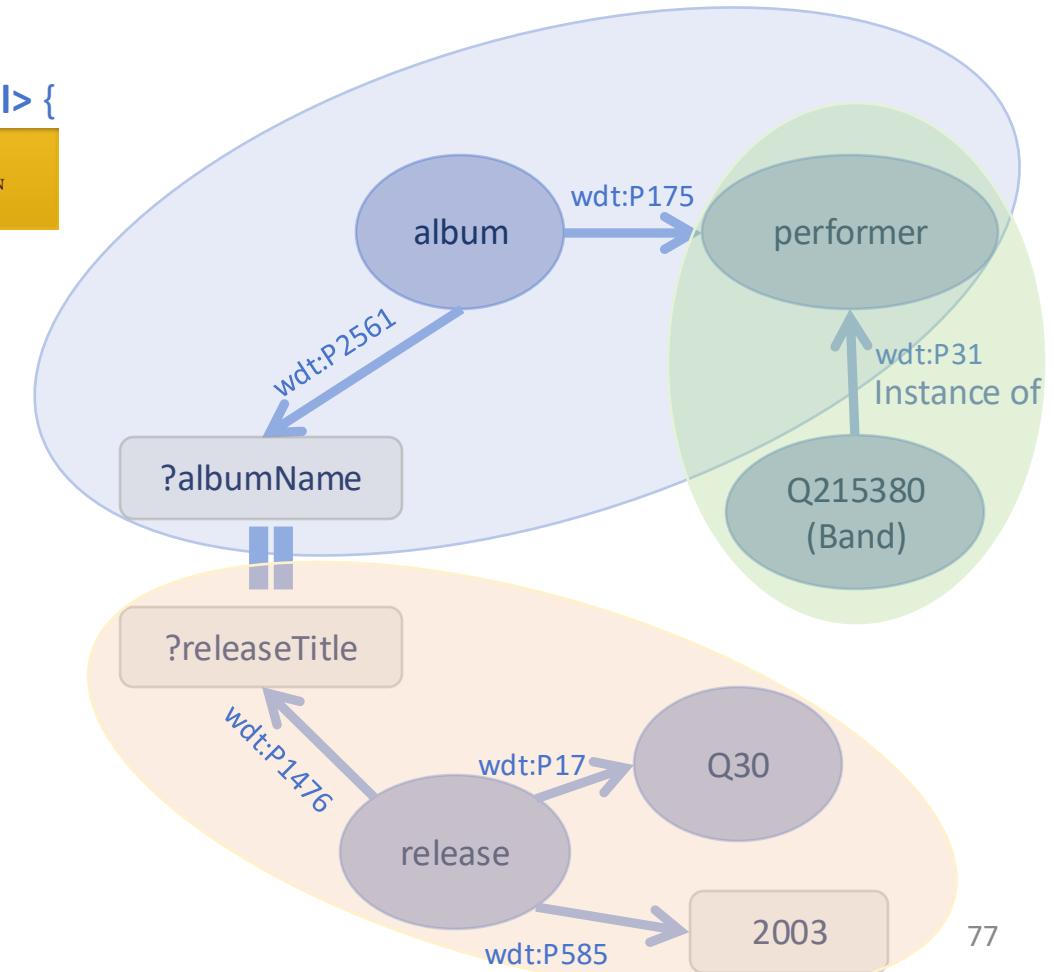
?release a wd:Q3972943 ;
wdt:P1476 ?releaseTitle ;
wdt:P17 wd:Q30 ; #
wdt:P585 ?date .



FILTER(?date = 2003) FILTER(?albumName = ?releaseTitle)

SERVICE <<https://query.wikidata.org/sparql>> {

?performer wdt:P31 wd:Q215380 . }



The result from Virtuoso (Q4-3)

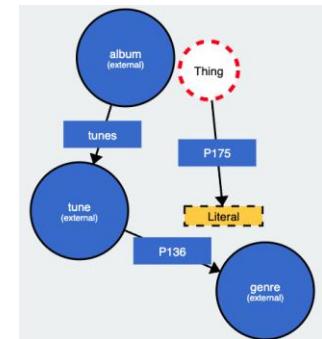
SPARQL | HTML5 table

albumName	album	performer
"The Best Of"	https://thesession.org/recordings/245	http://www.wikidata.org/entity/Q1410132

<https://thesession.org/recordings/245>

Discussion or Issue

- Should we or how to understand the ontology?
 - Ontology can be automatically generated during CSV2RDF in Virtuoso
 - Understanding OWL (Web Ontology Language) isn't essential right now
 - It can be substantially seen as a graph structure, serving as a framework for the DB
 - A familiarity of ontology will contribute to a more professional way of asking questions. Refer to:
 - Q1-1 "Find the sessions that took place in Greece." (Interoperable)
 - Q2-2 What genres appear in the album whose artist is "Adam Agee...?"?
 - Q4-2 List top 5 most popular record labels used by *musicians* on TheSession?
 - Editing or visualization tools for ontology: (1) protégé (2)WebVOWL



- Issue: The accuracy of SPARQL generation
 - What if there is a complex context (e.g., an oversized ontology)?

Future Work

- Further reconciliation Q4-3
- Knowing more about Wiki-Project Music
(https://www.wikidata.org/wiki/Wikidata:WikiProject_Music)
- E.g. “Find albums containing “Tola” in its name or its artist’s name” wdt:P175 performer
- Prompt engineering or In-context learning:
 - Intelligent agents designing and RAGs
 - E.g., ChatGPT can even read the visualization of OWL
- Knowledge-Reasoning, Recommendation, Visualization, Network Analysis... Q4-2 if ... a lazy searcher, asking natural language queries in a ...

Thank you!

Junjun Cao
Postdoctor of DDMAL
Email: junjun.cao@mail.mcgill.ca

Future projects

- ❖ Public UMIL: Crowd sourcing
- ❖ NLQ2SQL for Cantus Database (Lucas March)
- ❖ Continuous Integration for NLQ2SPARQL
 - ❖ Integrate database reconciliation as part of the in-context learning
- ❖ Investigate the use of RAG (Retrieval-Augmented Generation) for queries
- ❖ Use LLMs to generate frontend (e.g., sortable table)
- ❖ Continue to update Wikidata: e.g., Saints and Feasts
- ❖ How to keep Wikidata updated and correct

Future Meetings

- ❖ LinkedMusic Workshops:
 - ❖ Music Encoding Conference in London, UK
 - ❖ Monday 2 June 2025: 2 pm
 - ❖ IAML in Salzburg
 - ❖ Sunday 6 July 2025: 3 pm
- ❖ Keynote speaker suggestions
 - ❖ MIMO
 - ❖ DOREMUS
 - ❖ Mapping Manuscript Migrations
 - ❖ CIMCIM

Possible Future Workshop Sites

- ❖ 2026: MEC (Tokyo, Japan)
- ❖ 2026: IAML (Thessaloniki, Greece)
- ❖ 2027: MLA (TBA)
- ❖ 2027: IMS Congress (University of Stavanger, Norway)

SIMSSA / LinkedMusic Team @McGill

Summer 2024



Hanwen Zhang

Lucas March

Kyrie Bouressa

Anna de Bakker

Cole Thierrin

Ichiro Fujinaga

Yueqiao Zhang

Dylan Hillerbrand

Yinan Zhou

Junjun Cao

Kun Fang

Geneviève Gates-Panneton (MIA)

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*

Québec



Digital Research
Alliance of Canada

Alliance de recherche
numérique du Canada



Calcul Québec

These slides were created with the assistants from ChatGPT, Claude, and Gemini.