



# **WIKIDATA AND LINKED OPEN DATA FOR MUSIC VOCABULARIES**

Chris Holden, Library of Congress, Washington, DC, USA

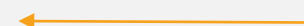
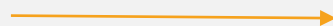
IAML Annual Conference, Prague

26 July 2022

# MEDIUM OF PERFORMANCE VOCABULARIES

- Medium of performance: “the instruments, voices, etc., used in the realization of a musical work” \*
- Libraries have established controlled vocabularies for medium of performance
  - Allows for a granular description of musical works
  - Information can be used for faceted searching
- Medium of performance terms vary by language.
  - Because music transcends language barriers, librarians have an interest in mapping vocabularies in different languages.

PIANO



KLAVIER

# MEDIUM OF PERFORMANCE AND LANGUAGE - LCMPT

- How do medium of performance vocabularies deal with language differences?
- The Library of Congress Medium of Performance (LCMPT) thesaurus primarily uses English
  - “All terms are established in English unless the foreign-language term has been borrowed into English, or the foreign-language term is unique to the language of the work being cataloged and cannot be found in English-language reference sources.” \*
  - “Foreign-language UFs [use-for references] are not made to English-language terms unless the foreign word or phrase is typically found in English-language reference sources.” \*

piano

- UF concert grand (piano)
- fortepiano
- grand piano
- pianoforte
- spinet (piano)
- spinnet (piano)
- upright piano

# MEDIUM OF PERFORMANCE AND LANGUAGE - UNIMARC

- The UNIMARC Medium of Performance vocabulary includes terms in multiple languages.
- There is “preferred” term in each language, as well as alternative terms.

Alternative labels

Preferred labels



Properties			
alternative label	Lyraflügel	German	Published
alternative label	Ditanaklasis	German	Published
alternative label	apolliricon	English	Published
alternative label	apollonion	English	Published
alternative label	apollonium	English	Published
alternative label	cabinet piano	English	Published
alternative label	aliquot piano	English	Published
alternative label	cottage piano	English	Published
alternative label	luthéal	French	Published
alternative label	piano-forte	Italian	Published
alternative label	pforte	Italian	Published
has broader	keyboard		Published
editorial note	English term is preferred globally.	English	Published
preferred label	Klavier	German	Published
preferred label	piano	English	Published
preferred label	piano	Spanish	Published
preferred label	piano	French	Published
preferred label	pianoforte	Italian	Published
preferred label	piano	Portuguese	Published
notation	kpf	English	Published



# WIKIDATA



- Wikidata is “a free and open knowledge base”
- Wikidata acts as the “central storage for the structured data of its Wikimedia sister projects including Wikipedia”
- Wikidata “can be read and edited by both humans and machines”
- Wikidata is multilingual.
  - “Data entered in any language is immediately available in all other languages. Editing in any language is possible and encouraged.”

Language	Label	Language	Label	Language	Label	Language	Label	
English	harpsichord	Chuvash	Клавесин	Croatian	Čembalo	Burmese	ဟာ့ပ်ဆီကော့ဒ်	
German	Cembalo	<a href="#">Library of Congress Medium of Performance Thesaurus ID</a>   <b>mp2013015333</b> subject named as <b>harpsichord</b> ▼ <b>0 references</b>				lo		
Arabic						n		
Asturian	Clavecín					n		
Azerbaijani	klavesin					Portuguese	cravo	
Belarusian	клавесін							
Belarusian (Taraškievica orthography)	клявэсын							
Bulgarian	клавесин					Friulian	cravo	
Breton	Klav-kerdin					Portuguese		
Bosnian	Čembalo					Romanian	clavecin	
Catalan	clavicèmbal					Galician		
Central Kurdish		French	clavecin	Luxembourgish	Cembalo	Russian	клавесин	
Czech	cembalo					Lithuanian	Klavesinas	
						Latvian	klavesīns	
						Macedonian	чембало	
						Malay	Harpsichord	
						Sicilian	Clavicimmalu	
		Western Frisian	Klavesimbel					
		Irish	Cruitchorða					



## HARPSICHORD IN WIKIDATA



- No links to the UNIMARC Medium of Performance vocabulary.
  - There is not even a specific property that can be used for this.
- However, there *is* a general property to link to IFLA vocabularies.
  - One can add a qualifier to indicate the specific IFLA vocabulary is the UNIMARC Medium of Performance thesaurus

IFLA value vocabularies ID



mop#khp



▼ 1 reference

stated in

UNIMARC Medium of performance



# HOW MANY MEDIUMS OF PERFORMANCE ARE IN WIKIDATA?



- We can run specific SPARQL queries in Wikidata to get some results.
- How many Wikidata objects have associated LCMPT terms?
- 161 Wikidata objects are linked to LCMPT terms
  - However, the LCMPT has over 900 terms in its vocabulary, so this is not comprehensive.

#LC	item	itemLabel	object
1			
2	SEL		
3	WHE <a href="#">Q wd:Q543833</a>	uilleann pipes	mp2013015760
4	{		
5	: <a href="#">Q wd:Q587246</a>	contrabass flute	mp2013015177
6	: <a href="#">Q wd:Q685206</a>	alphorn	mp2013015010
7	}		
	<a href="#">Q wd:Q810511</a>	basset clarinet	mp2013015072
	<a href="#">Q wd:Q1095084</a>	tromba marina	mp2013015749
	<a href="#">Q wd:Q1146067</a>	dizi	mp2013015207
	<a href="#">Q wd:Q1494285</a>	garklein recorder	mp2018015019
	<a href="#">Q wd:Q2066397</a>	ching	mp2013015138
	<a href="#">Q wd:Q2059671</a>	diddle bow	mp2013015208





# HOW MANY MEDIUMS OF PERFORMANCE ARE IN WIKIDATA?



- How many Wikidata objects have an associated Hornbasteel-Sachs classification?
- 2340 Wikidata objects have a Hornbasteel-Sachs classification

```
1 #Hornbasteel-Sachs classification
2 SELECT ?item
3 WHERE
4 {
5   ?item wd:Q107894862
6   SERVICE wikibase:label
7 }
```

item	itemLabel	object
<a href="#">Q111824539</a>	soprano trombone	423.22
<a href="#">Q110395066</a>	Trumpets	423
<a href="#">Q107894862</a>	Q107894862	422.112
<a href="#">Q105303913</a>	typewriter zither	314.122-8
<a href="#">Q104996348</a>	kundi	322.11
<a href="#">Q104521646</a>	metal clarinet	422.211.2-71
<a href="#">Q100948993</a>	paidi	421.122.1
<a href="#">Q98654724</a>	octocontra-alto clarinet	422.211.2

```
...[?item]_LANGUAGE],en". }
```



## HOW MANY MEDIUMS OF PERFORMANCE ARE IN WIKIDATA?



- Can we run SPARQL queries looking for specific properties or classes relating to medium of performance?
  - There's no one magic indicator that we can search for to find all the items we're looking for.
  - Instead, there is a complex and inconsistently-applied set of Wikidata relationships.
- There is a Wikidata item for “musical instrument”:

musical instrument (Q34379)

- So can we simply look for all items that are an instance of “musical instrument”?



# HOW MANY MEDIUMS OF PERFORMANCE ARE IN WIKIDATA?



- Here's what happens if we look for all instances of a musical instrument (or all instances of something that is a subclass of a musical instrument)



1	#Musicalinstrumentins			
2	SELECT ?item ?itemLabel	<a href="#">wd:Q65771840</a>	Caixa de música	Item of the collection of the "Museu do Diamante" (Diamond Museum), a Brazilian museum located in the city of Diamantina.
3	WHERE			
4	{	<a href="#">wd:Q79485134</a>	Psaltery	musical instrument by unknown artist (1918.345)
5	?item p:P31/ps:P31/			al instrument
6	SERVICE wikibase:label	<a href="#">wd:Q79485147</a>	Sanxian (Three-Stringed Guitar)	musical instrument by unknown artist (1918.348)
7	}			
		<a href="#">wd:Q79485158</a>	Oud	musical instrument by unknown artist (1918.350)
		<a href="#">wd:Q79485183</a>	Harp	musical instrument by unknown artist (1918.356)
		<a href="#">wd:Q79485198</a>	Sono Koto	musical instrument by unknown artist (1918.359)
		<a href="#">wd:Q79485231</a>	Harp	musical instrument by unknown artist (1918.363)
		<a href="#">wd:Q79485236</a>	Chittarone	musical instrument by unknown artist (1918.364)
		<a href="#">wd:Q79485244</a>	Harp	musical instrument by unknown artist (1918.366)



# HOW MANY MEDIUMS OF PERFORMANCE ARE IN WIKIDATA?



- According to Wikidata guidelines, the only items that we can call an “instance of” a “musical instrument” are “individual specimens.”

 [conflicts-with constraint](#)  [edit](#)

<a href="#">property</a>	<a href="#">instance of</a>
<a href="#">item of property constraint</a>	<a href="#">musical instrument</a>
<a href="#">replacement value</a>	<a href="#">type of musical instrument</a>
<a href="#">constraint clarification</a>	Instances of musical instruments (Q34379) are typically individual specimens and shouldn't be subclasses. If you're working with a proper subclass, try setting P31 = "type of musical instrument" (Q110295396) (English)

[▼ 0 references](#) [+ add reference](#)



## HOW MANY MEDIUMS OF PERFORMANCE ARE IN WIKIDATA?



- Things get confusing quickly!

musical instrument (Q34379)

type of musical instrument (Q110295396)

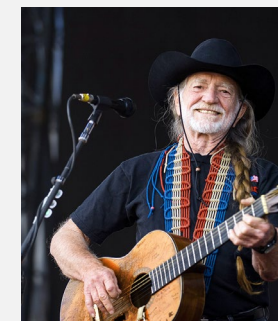
family of musical instruments (Q1254773)

# HOW MANY MEDIUMS OF PERFORMANCE ARE IN WIKIDATA?

musical instrument (Q34379)

type of musical instrument (Q110295396)

family of musical instruments (Q1254773)

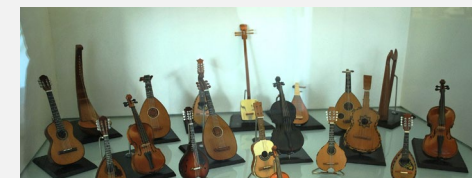


Trigger (guitar) (Q16899451)

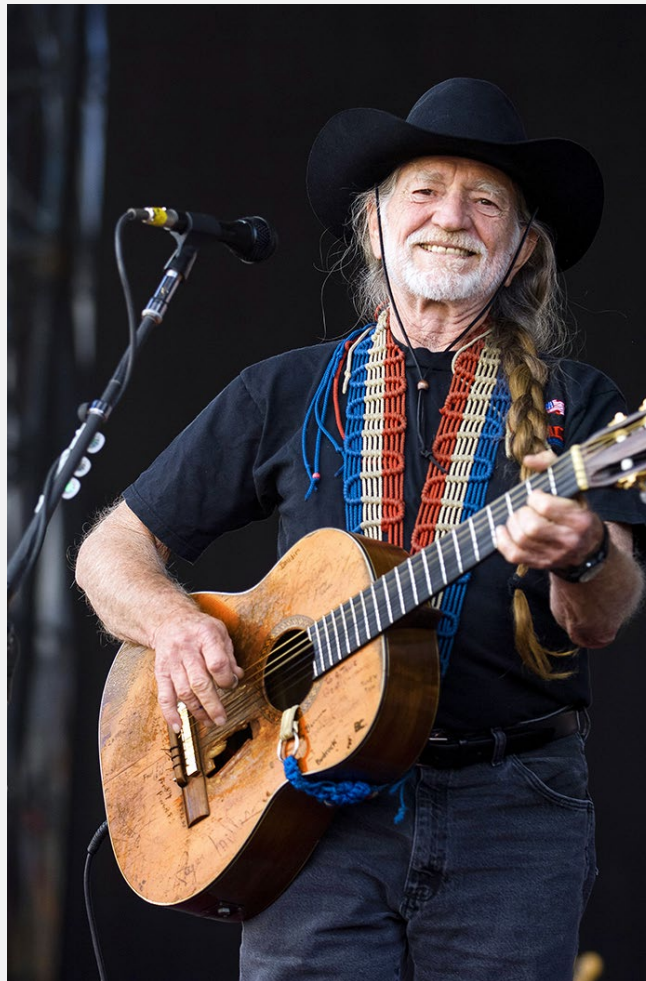
Willie Nelson's guitar



guitar (Q6607)



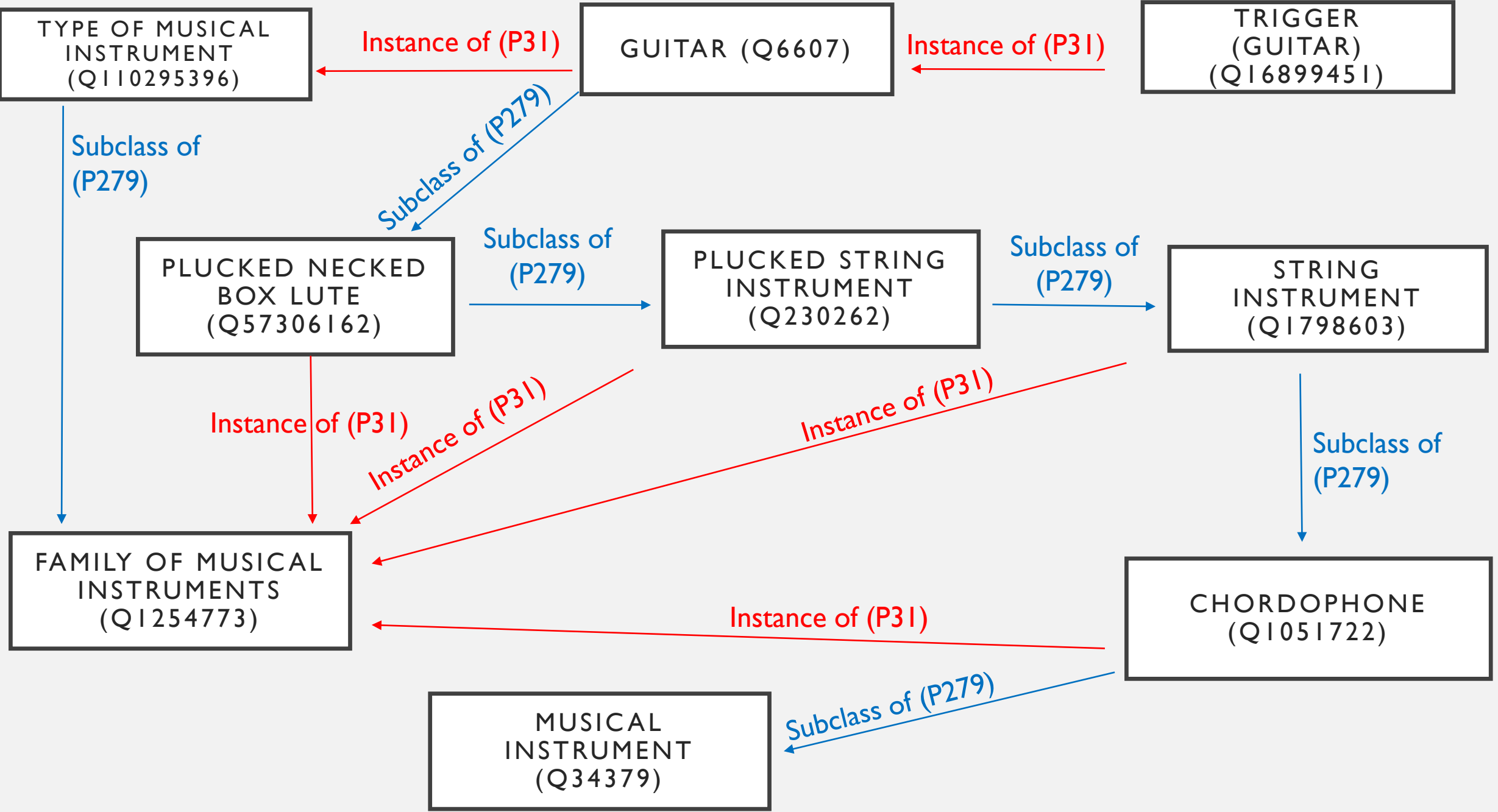
string instrument (Q1798603)



## Trigger (guitar) (Q16899451)

---

Willie Nelson's guitar







## HOW MANY MEDIUMS OF PERFORMANCE ARE IN WIKIDATA?



- These large and sprawling relationship diagrams are the side effect of Wikidata trying to be a database for everything.

musical instrument (Q34379)

→ 17,554 items in Wikidata

type of musical instrument (Q110295396)

→ 16 items in Wikidata  
(This class is not deployed as often as it should be!)

family of musical instruments (Q1254773)

→ 766 items in Wikidata



## HOW MANY MEDIUMS OF PERFORMANCE ARE IN WIKIDATA?



- How do we find a count for “types of musical instruments” if no one is using this term in Wikidata?
- This query looks for all items that are a subclass of a musical instrument, or a subclass that falls under “musical instrument” on some lower level of the hierarchy

```
1 #subclass
2 SELECT ?item ?itemLabel
3 WHERE
4 {
5   ?item wdt:P279+ wd:Q34379.
6   SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en". }
7 }
```



# HOW MANY MEDIUMS OF PERFORMANCE ARE IN WIKIDATA?



- This returns 4071 results, which is the closest thing we'll get to a definitive list of instrumental mediums of performance in Wikidata.

item	itemLabel
<a href="#">wd:Q12639</a>	idiophone
<a href="#">wd:Q52954</a>	keyboard instrument
<a href="#">wd:Q105738</a>	electrophone
<a href="#">wd:Q133163</a>	percussion instrument
<a href="#">wd:Q212892</a>	membranophones
<a href="#">wd:Q217306</a>	transposing instrument
<a href="#">wd:Q230062</a>	Q230062
<a href="#">wd:Q245519</a>	6-plus-6-instruments
<a href="#">wd:Q258681</a>	bladder pipe
<a href="#">wd:Q334606</a>	practice instrument
<a href="#">wd:Q356923</a>	Adiaphone
<a href="#">wd:Q365289</a>	pyrophone
<a href="#">wd:Q386382</a>	tympanum
<a href="#">wd:Q424111</a>	acoustic musical instrument
<a href="#">wd:Q431585</a>	rawap
<a href="#">wd:Q432358</a>	shak-shak
<a href="#">wd:Q471918</a>	pong lang
<a href="#">wd:Q487818</a>	shruti box



# HOW MANY MEDIUMS OF PERFORMANCE ARE IN WIKIDATA?



- What about mediums of performance that are not instruments?

type of musical ensemble/group (Q105756498)

→ 346 items in Wikidata

voice type (Q1063547)

→ 30 items in Wikidata

Other miscellaneous mediums of performance:

narrator (Q58363586)

rapper (Q2252262)

computer (Q68)

basso continuo (Q724321)



## FUTURE PROJECTS



- There are 392 UNIMARC Medium of Performance terms.
- A recent IAML project mapped these terms to 279 different LCMPT terms.
- There have been complications in finding a good home to publish the mapping between these two vocabularies.
- Would Wikidata work as a multilingual hub to link these vocabularies to terms in different languages?
- Benefits include:
  - The capacity for labels in multiple languages
  - The ability to link different vocabularies to the same Wikidata items
  - Low barrier of entry in terms of training and technical expertise



# FUTURE PROJECTS



Possible issues to be addressed:

- 1) Does Wikidata need a specific property for UNIMARC Medium of Performance?

A screenshot of the Wikidata interface. On the left, a grey box contains the text "IFLA value vocabularies ID". To its right, the value "mop#khp" is displayed in blue. Below this, a blue triangle icon indicates a dropdown menu with the text "1 reference". A light blue box highlights the reference, showing the property "stated in" and the value "UNIMARC Medium of performance".

IFLA value vocabularies ID mop#khp


▼ 1 reference


stated in UNIMARC Medium of performance

# FUTURE PROJECTS

IFLA value vocabularies ID



mop#pgo 

 edit

subject named as

gong

mapping relation type

exact match


▼ 1 reference

stated in

UNIMARC Medium of performance

+ add reference



mop#ptg 

 edit

subject named as

tuned gong

mapping relation type

narrow match

▼ 1 reference

stated in

UNIMARC Medium of performance

+ add reference

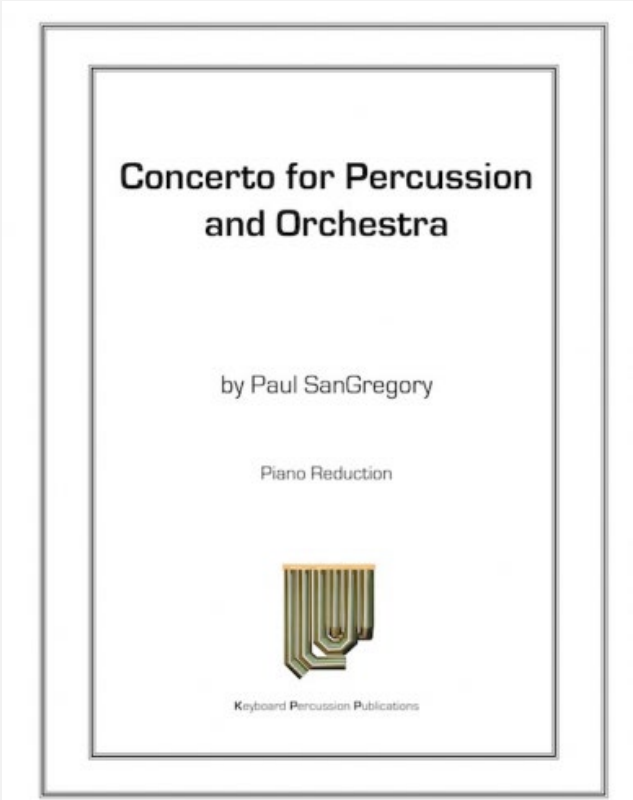
# FUTURE PROJECTS

Possible issues to be addressed:

- 3) How do we deal with terms for instruments vs. terms for performers?
- A lot of library medium of performance vocabularies name an instrument, but they are really referring to a performer.

percussion instrument (Q133163)

type of musical instrument that produces a sound by directly hitting it



Concerto for Percussion  
and Orchestra

by Paul SanGregory

Piano Reduction



Keyboard Percussion Publications

percussionist (Q4351403)

musician who plays percussion





## FUTURE PROJECTS



- 4) Anything else?



**THANK YOU!**

Chris Holden, Library of Congress, Washington, DC, USA

IAML Annual Conference, Prague

26 July 2022

[chold@loc.gov](mailto:chold@loc.gov)