



Workshop #1

Introduction to Wikidata



Workshop Agenda

Part 1 (slides 4 to 16)

- A knowledge base
- Key concepts
- WikiCommons

Part 2 (slides 17 to 24)

- Search engines
- InfoBox
- Related data
- Wikipedia to Wikidata

Part 3 (slides 25 to 31)

- Header of a Wikidata entity
- Interwiki links
- Declarations
- External IDs

Recap (slides 32 à 37)



Part 1

Knowledge base, triples and glossary of terms



Knowledge base

What is a **knowledge base**? It's a database storing a lot of information represented as triples.



What's a **triple**? It is the smallest factual statement that can be made to record information about anything. For example:

"Nutcracker — composer — Tchaikovsky".

In this workshop, we will take an introductory look at Wikidata, a special knowledge base which was born in the wake of the other Wikimedia Foundation (**WMF**) projects.





Wikidata...



A HUGE knowledge base

- **Accessible** to everyone — both to read information and to make queries and, above all, to edit and contribute to it;
- **Reusable** – all data in Wikidata is published under a free Creative Commons **CC0** license, giving anyone the full ability to make use of the data;
- **Multilingual** – entities can be named and described in any human language.



Key concept — item or entity

- A thing or a concept in Wikidata, whether concrete (the Earth) or abstract (planet), is called an **item** or an entity.
- All items or entities have a unique identifier (ID) beginning with 'Q', such as **Tchaikovsky = Q7315** or **The Nutcracker = Q193705**.
- The unique ID is the actual machine-readable name (and locator) of the item.
- These unique IDs can be used to disambiguate items with the same name:
 - **"Titanic (Q25173)"**, the boat;
 - **"Titanic (Q44578)"**, the film directed by James Cameron;
 - **"Titanic (Q1889132)"**, a Norwegian music band.



Key concept — Property

- In the triple "The Nutcracker has composer Piotr Ilyich Tchaikovsky", **composer is the property**.
- All properties have a unique identifier that begins with 'P', such as "composer (P86)".
- There are far fewer properties than items. Whereas anyone can create an item, the creation of properties is restricted by a collaborative process.



Key concept — Value

- In the triple "The Nutcracker — has composer — Piotr Ilyich Tchaikovsky", **Piotr Ilyich Tchaikovsky is the value**.
- Values can be another **item** (starting with a 'Q' identifier), such as "Piotr Ilyich Tchaikovsky (Q7315)".
- Values can also be a **literal value** (a number, a date, geo coordinates, a URL, a link to an image on Wiki Commons, or any string of text characters).

For a more complete overview of Wikidata's vocabulary, you are strongly advised to read the Wikidata glossary : <https://www.wikidata.org/wiki/Wikidata:Glossary>



Key concept — Statement

- Together, an **item**, a **property** and a **value** forms a triple.
- This is referred to as a **statement** in Wikidata.

For a more complete overview of Wikidata's vocabulary, you are strongly advised to read the Wikidata glossary : <https://www.wikidata.org/wiki/Wikidata:Glossary>

Triples/statements: Different words, same concepts

Triples and **statements** are the same method of representing knowledge. But their terminology is different.

Knowledge representation framework	Referent entity	Link	Value
Resource Description Framework (RDF)	Subject	Predicate	Object
Wikidata statement	Item	Property	Value
Plain language	Thing	Relationship	Thing

Remember: The “value” can either be another **entity/item** or a **literal value** (a number, a string of text characters, a date, geo coordinates, an URL, etc.)



A few sample triples

Take the sentence "The Nutcracker is a ballet composed by Tchaikovsky".

At least three triples or statements can be inferred from this sentence:

1. "The Nutcracker — has composer — Piotr Ilyich Tchaikovsky"
2. "The Nutcracker — is an instance of — a ballet" (which is Wikidata's way of stating that: The Nutcracker is a ballet)
3. "Piotr Ilyich Tchaikovsky has occupation composer"

A few sample statements (with IDs)

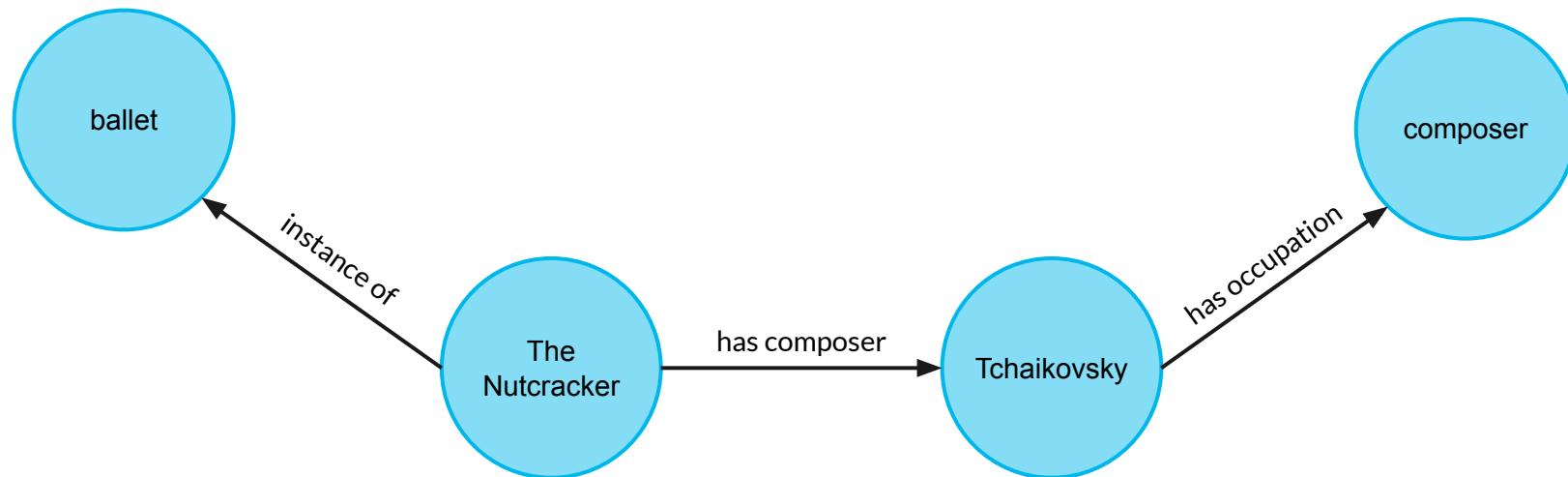
Item (subject)	Property (predicate)	Value (object)
The Nutcracker (Q193705)	composer (P86)	Piotr Ilyich Tchaikovsky (Q7315)
The Nutcracker (Q193705)	instance of (P31)	ballet (Q15079786)
Piotr Ilyich Tchaikovsky (Q7315)	occupation (P106)	composer (Q36834)



A few sample statements (what machines see)

Item (subject)	Property (predicate)	Value (object)
Q193705	P86	Q7315
Q193705	P31	Q15079786
Q7315	P106	Q36834

A few sample statements (graph version)





Wikidata by numbers

- 87,792,976 different entities
- 7705 properties
- 25,783 active users
- 311 robots

- and YOUR contributions !

<https://www.wikidata.org/wiki/Special:Statistics>



Wikidata by numbers (cont'd)

In Wikidata, you will be editing and enriching mostly items – and all manners of them (Q1 = univers...).

- There are currently 87 million items.
- 31% are instances of scientific article.
- 9% are instances of human (Q5) (that's 6 million out of 8 billion human beings)...

More statistics here:

<https://www.wikidata.org/wiki/Wikidata:Statistics>



Part 2

Sister projects within Wikimedia Foundation - Usage of Wikidata



Sister projects within Wikimedia Foundation

The **Wikimedia Foundation (WMF)** is best known for the Wikipedia encyclopedia, but it also oversees several other initiatives or projects — including Wikidata.

Wikidata started as a project to interlink different language editions of **Wikipedia**, so that factual edits could be made once in Wikidata rather than be manually implemented in each language edition of the same article. But Wikidata has evolved to become much more than that.

Another project at the heart of this constellation is **Wikimedia Commons**. It provides storage for all visual, audiovisual and other media file that are shared across all the WMF projects.

In Wikidata, there are no images; there are links to images in Wiki Commons.

These Wikimedia Foundation projects are interdependent.

Wikimedia Commons: the "content" base

Wikimedia Commons is a repository for images, videos and audio files used in other Wikimedia Foundation projects. It is therefore widely used.

- It is also **multilingual by design**, but English is used quite a bit as a reference language.
- All content is **free** of use under a Creative Commons license.
- Beyond image, video and sound files, Wiki Commons also stores **other content** such as PDFs or TAB files that contain data tables in JSON code.





Benefits of Wikidata

Spoiler alert: the following few slides offer a sneak peek at the next workshop

Humans love Wikipedia. Machines love both Wikipedia and Wikidata.

Wikidata is the preferred knowledge base for search/answer algorithms and AI-powered tools such as personal assistants because they need access to an unlimited supply of simple and highly structured information.

How do you think search engines manage to display carousels of an artist's works right on the answer page?

The Google Knowledge Card

The screenshot shows a Google search for "Piotr Ilitch Tchaïkovski". The search bar at the top contains the name. Below the search bar, there are tabs for "Tous", "Images", "Vidéos", "Actualités", "Maps", "Plus", "Paramètres", and "Outils". The search results show "Environ 138 000 résultats (0,60 secondes)". The first result is from Wikipedia, titled "Piotr Ilitch Tchaïkovski — Wikipédia". Below this, there are several other search results from various websites, including Olyrix, Musicologie.org, and Symphozik.info. A red dotted line highlights a "Knowledge Card" for "Piotr Ilitch Tchaïkovski". The card features a grid of images, including a large portrait of Tchaikovsky and several smaller portraits. Below the images, the name "Piotr Ilitch Tchaïkovski" is displayed, followed by the title "Compositeur". Underneath, there is a section titled "Proposé sur" with links to YouTube, Spotify, and Play Musique. At the bottom of the card, there is a brief biography: "Piotr Ilitch Tchaïkovski ou Tchaïkovsky est un compositeur russe de l'époque romantique né le 25 avril 1840 à Votkinsk et mort le 25 octobre 1893 à Saint-Petersbourg. Wikipédia".

This is what you get when you
Google "Piotr Ilyich Tchaikovsky"...

The Google Knowledge Card,
named « InfoBox »
in the Wikimedia world

Google Results for “Piotr Ilitch Tchaïkovski”



If linked data about an artist and their works is available, Google can populate the knowledge card with lists of popular works or with platforms where those works are available.

In addition, basic biographical information automatically added at the bottom of the card may help convey a sense of notoriety about the artist.

Piotr Ilitch Tchaïkovski ou Tchaïkovsky est un compositeur russe de l'époque romantique né le 25 avril 1840 à Votkinsk et mort le 25 octobre 1893 à Saint-Petersbourg. Wikipédia

Date et lieu de naissance : 7 mai 1840, Votkinsk, Russie

Date et lieu de décès : 6 novembre 1893, Saint-Petersbourg, Russie

Nom complet : Pyotr Ilyich Tchaikovsky

Style : Musique romantique Opéra, ballet, musique

<https://www.google.com/search?q=Piotr+Ilitch+Tcha%C3%AFkovski&og=Piotr+Ilitch+Tcha%C3%AFkovski>



Google loves open knowledge bases

Google relies on many from many open knowledge bases and content repositories to generate rich results and to populate knowledge cards.

In the previous example, Google:

- Displays a short summary from Wikipedia
- Pulls basic factual information from Wikidata
- Retrieves photos from Wikimedia Commons

Wikimedia projects are by no means the only sources of information and content for Google. But they are accessible, easy and powerful platforms for anyone wishing to participate in them.

The link between Wikipedia and Wikidata

Article Talk

Wiki Loves Earth: An international photographic contest where you can share photos of nature and potentially win a prize. [Ever](#)

Pyotr Ilyich Tchaikovsky

From Wikipedia, the free encyclopedia

"Tchaikovsky" redirects here. For other uses, see Tchaikovsky (disambiguation). This name uses Eastern Slavic naming customs; the patronymic is Ilyich and the family name is Tchaikovsky.

Pyotr Ilyich Tchaikovsky^[a 2] (English: /tʃɑːˈkɒfski/ *chy-KOF-skee*;^[1] Russian: Пётр Ильич Чайковский [ˈkɐj.ɫɨˈskʲɪj] (listen); 7 May 1840 – 6 November 1893^[a 4]) was a Russian composer of the Romantic era whose music made a lasting impression internationally. He was honored in 1884 by the Russian government with a lifetime pension.

Although musically precocious, Tchaikovsky was educated for a career as a civil servant. The Russian system at the time and no system of public music education. When an opportunity to enter the nascent Saint Petersburg Conservatory, from which he graduated in 1865. The fact that he received there set him apart from composers of the contemporary nationalist movement and the Five with whom his professional relationship was mixed.

Tchaikovsky's training set him on a path to reconcile what he had learned with the native music he had exposed to from childhood. From that reconciliation, he forged a personal but unmistakably Russian musical style. European music, which seemed to defeat the potential for using Russian music in large-scale composite style, and it caused personal antipathies that dented Tchaikovsky's self-confidence. Despite his many popular successes, Tchaikovsky's life was punctuated by personal crises and included his early separation from his mother for boarding school followed by his mother's death.

Every Wikipedia page has a Wikidata item which is language agnostic.

This is how you can switch from Wikipedia EN to Wikipedia FR without any problems.

Click on the "wikidata item" link to discover the related wikidata item.



Part 3

Components of a Wikidata item :
multilingual header, statements, external identifiers, interwiki links

Label, description and alias of an item

Pyotr Ilyich Tchaikovsky (Q7315)

Russian composer edit

Pyotr Ilyich Tchaikovsky | Pyotr Tchaikovsky | Peter Illich Tchaikovsky

↪ In more languages

Language	Label	Description	Also known as
English	Pyotr Ilyich Tchaikovsky	Russian composer	Pyotr Ilyich Tchaikovsky Pyotr Tchaikovsky Peter Illich Tchaikovsky
French	Piotr Ilitch Tchaïkovski	compositeur	
Italian	Petr Il'ic Čajkovskij	compositore russo del romanticismo	Pyotr Ilyich Tchaikovsky Petr Čajkovskij
German	Pjotr Iljitsch Tschaikowski	russischer Komponist	Peter Tschaikowski Pjotr Tschaikowski Peter Tschaikowsky

All entered languages

Statements

instance of human edit

↪ 2 references

+ add value

image edit

The Wikidata item (specifically named Q7315) which represents the composer of music for the Nutcracker ballets, Mr. Tchaikovsky.

The first section we see here, contains its name in 4 languages (FR, EN, IT, DE), languages proposed by Wikidata according to the user's editions.



Interwiki links

Wikipédia (142 entrées) modifier

af

Piotr Tsjajkofsکی

als

Pjotr Iljitsch Tschajkowski

an

Pjotr Illich Chaikowski

ar

بيتر ايليتش تشايكوفسكي

arz

تشاريكوفسكي

ast

Piotr Illich Chaikowski

ay

Piotr Tchaikowski

az

Pjotr Çaykovski

ban

Pjotr Ilyich Tchaikovsky

bar

Pjotr Iljitsch Tschajkowski

bat_smg

Piotros Čaikovskis

ba

Чайковский Пётр Ильич

bcl

Pëtr Čajkovskij

be_x_old

Пётар Чаікоўскі

be

Пётр Ільіч Чайкоўскі

bg

Пьотър Чайковски

bh

प्योत्र इल्लिच चाइकोवस्की

bi

Pjotr Tchaikovsky

bn

পিয়ত্র ইলিচ চাইকোভস্কি

br

Pjotr Ilyitch Tchaikovsky

bs

Petar Iljić Čajkovski

bxr

Пётр Чайковский

ca

Pjotr Illich Txaikowski

cbk_zam

Pjotr Ilyich Tchaikovsky

cdo

Pjotr Ilyich Tchaikovsky

ceb

Pjotr Ilyich Tchaikovsky

ce

Чайковский, Пётр Ильич

ch

Pjotr Ilyich Tchaikovsky

ckb

پۆتەر ئیللیچ چایکوفسکی

Wikilivres (0 entrées) modifier

Wikinews (1 entrée) modifier

ru

Категория:Пётр Чайковский

Wikiquote (16 entrées) modifier

bg

Пьотър Чайковски

bs

Pjotr Iljić Čajkovski

cs

Petr Iljič Šajkovskij

el

Πιότρ Τσαϊκόφσκι

en

Pjotr Ilyich Tchaikovsky

eo

Pjotr Iljĉo Ĉajkovskij

es

Piotr Illich Chaikovski

hr

Petar Iljić Čajkovski

hy

Պյոտր Իլյիչ Շայկովսկի

it

Pëtr Iljĉ Šajkovskij

ka

პეტრე ილიჩ ჩაიკოვსკი

lt

Piotras Čaikovskis

pt

Pjotr Ilyich Tchaikovsky

ru

Пётр Ильич Чайковский

th

ปิเตอร์ อิลลิช ไชคอฟสกี

uk

Чайковский Петро Ілліч

Wikisource (3 entrées) modifier

cs

Author:Petr Iljić Čajkovskij

en

Author:Pjotr Ilyich Tchaikovsky

ru

Пётр Ильич Чайковский

At the end of the page (or at the top right if your window is large enough to allow it), you will find **the interwiki links**, those that link Wikipedia pages in different languages.

Characters indicate the language of the Wiki, such as **FR** for French and **EN** for English.

In addition to Wikipedia, you'll find the Wikibooks project, Wikinews, Wikiquote, Wikisources...

Statements or triples

Continuing with our example, looking at the statements on the Wikidata page of "The Nutcracker" (Q193705), we find the statement "composer" (P86), with "Piotr Ilyich Tchaikovsky" (Q7315) as a value.

<https://www.wikidata.org/wiki/Q7315>

<https://www.wikidata.org/wiki/Q193705#P86>

Implicite item
The Nutcracker (Q193705)

location of first performance	<div><div><div></div><div>Mariinsky Theatre</div></div><div><div></div><div>> 1 reference</div></div></div> <div><div></div><div>edit</div></div> <div><div></div><div>+ add value</div></div>
composer	<div><div><div></div><div>Piotr Ilyich Tchaikovsky</div></div><div><div></div><div>> 0 references</div></div></div> <div><div></div><div>edit</div></div> <div><div></div><div>+ add reference</div></div> <div><div></div><div>+ add value</div></div>
librettist	<div><div><div></div><div>Marius Petipa</div></div><div><div></div><div>> 0 references</div></div></div> <div><div></div><div>edit</div></div> <div><div></div><div>+ add reference</div></div> <div><div></div><div>+ add value</div></div>
choreographer	<div><div><div></div><div>Lev Ivanov</div></div><div><div></div><div>> 1 reference</div></div></div> <div><div></div><div>edit</div></div> <div><div></div><div>+ add reference</div></div> <div><div></div><div>+ add value</div></div>
	<div><div><div></div><div>Marius Petipa</div></div><div><div></div><div>> 0 references</div></div></div> <div><div></div><div>edit</div></div> <div><div></div><div>+ add reference</div></div> <div><div></div><div>+ add value</div></div>

Statements or triples (continuation)

Other basic statements shared by many humans are family ties, mother, father and siblings/sority.

In fact, in the statements for The Nutcracker, you can find at least 13 different properties, and in the statements of type identifiers, at least a dozen or so...



Statements			Other info (0 on)
instance of	<div><div><div><div></div><div></div><div></div></div><div><div>ballerina</div><div>0 references</div></div></div><div><div>add reference</div><div>add value</div></div></div>	<div><div>edit</div></div>	
image	<div><div><div><div></div><div></div><div></div></div><div><div></div><div><div>broderick.jpg</div><div>2,048 x 2,048, 309 KB</div></div></div><div><div>1 reference</div></div></div><div><div>add value</div></div></div>	<div><div>edit</div></div>	
named after	<div><div><div><div></div><div></div><div></div></div><div><div>nutcracker doll</div><div>0 references</div></div></div><div><div>add reference</div><div>add value</div></div></div>	<div><div>edit</div></div>	
title	<div><div><div><div></div><div></div><div></div></div><div><div>Scenarij (Plot)</div><div>0 references</div></div></div><div><div>add reference</div><div>add value</div></div></div>	<div><div>edit</div></div>	
based on	<div><div><div><div></div><div></div><div></div></div><div><div>The Nutcracker</div><div>0 references</div></div></div><div><div>add reference</div><div>add value</div></div></div>	<div><div>edit</div></div>	
derivative work	<div><div><div><div></div><div></div><div></div></div><div><div>The Nutcracker in 3D</div><div>1 reference</div></div></div><div><div>The Nutcracker and the Four Realms</div><div>1 reference</div></div><div><div>De Nutcracker</div><div>1 reference</div></div><div><div>Q6712066</div><div>1 reference</div></div></div> <div><div>add value</div></div>	<div><div>edit</div></div>	
location of first performance	<div><div><div><div></div><div></div><div></div></div><div><div>Mariinsky Theatre</div><div>1 reference</div></div></div><div><div>add value</div></div></div>	<div><div>edit</div></div>	
composer	<div><div><div><div></div><div></div><div></div></div><div><div>Pjotr Iljyč Tchaikovsky</div><div>0 references</div></div></div><div><div>add reference</div><div>add value</div></div></div>	<div><div>edit</div></div>	
librettist	<div><div><div><div></div><div></div><div></div></div><div><div>Marius Petipa</div><div>0 references</div></div></div><div><div>add reference</div><div>add value</div></div></div>	<div><div>edit</div></div>	
choreographer	<div><div><div><div></div><div></div><div></div></div><div><div>Léon Ivanov</div><div>1 reference</div></div></div><div><div>Marius Petipa</div><div>1 reference</div></div></div> <div><div>add value</div></div>	<div><div>edit</div></div>	



External identifiers: a special kind of statement

At the bottom of the page, we can find specialized statements, the external identifiers, which are in fact "keys" to perfectly link the Wikidata item with its equivalent item in another system .

This kind of information is vital to have **linked data**. Most automated mechanisms use these values to exchange with external data systems.

<https://www.wikidata.org/wiki/Q7315#P950>

Identifiers			
VIAF ID	99258155		
	> 2 references		+ add value
ISNI	0000 0001 2144 9518		
	> 1 reference		+ add value
BNA authority ID	000021617		
	= 0 references		+ add reference
			+ add value
Vatican Library ID	ADV12465838		
	= 0 references		+ add reference
	ADV12570349		
	= 0 references		+ add reference
	ADV12570350		
	= 0 references		+ add reference
NORAF ID	90142966		
	= 0 references		+ add reference
			+ add value

External identifiers


























Here are a few identifier for "Elisapie Isaac (Q596774)".

Key external identifiers for artists include **VIAF**, **ISNI**, **MusicBrainz**, **IMDb**, as well as social media IDs.

Note: The value for most identifiers is a literal value (the ID number itself) not a hyperlink. However, Wikidata renders these values as clickable hyperlinks.

<https://www.wikidata.org/wiki/Q596774>

Identifiers

VIAF ID	 53894581 			
ISNI	 0000 0000 0249 2038 			
Bibliothèque nationale de France ID	 150759469 			
FAST ID	 1632456 			
Library of Congress authority ID	 no2006006080 			
WorldCat Identities ID	 lccn-no2006006080 			



Recap



A huge knowledge base

Key features of Wikidata:

- Collaborative
- Licensed CC0
- Multilingual

Factual statements in triple format:

- **Item + Propriety + Value = Statement**



Benefits of Wikidata and Wikimedia project

Wikimedia sister projects are interdependent.

- **Wikipedia** ⇨ articles in several languages
- **Wikimedia Commons** ⇨ images and other media files
- **Wikidata** ⇨ structured data

Data and content from Wikimedia projects are used to populate InfoBoxes on Wikipedia, as well as Google Knowledge Cards.



Anatomy of a Wikidata item

A Wikidata item is comprised of:

- Title (unique identifier)
- Header with labels, description and aliases
- Statements
- External identifiers
- Interwiki links

External identifiers are important: they connect a Wikidata item to other open databases that also hold information about the same item.

Acknowledgements

Project partners



CANADIAN ARTS
PRESENTING ASSOCIATION
ASSOCIATION CANADIENNE
DES ORGANISMES ARTISTIQUES



Funding partners



Canada Council
for the Arts

Conseil des arts
du Canada

Canada



WIKIMEDIA
FOUNDATION

A short horizontal bar with a teal segment on the left and a magenta segment on the right.

End of the workshop #1

Have fun with Wikidata!