

Connecting Conflict Data through Wikimedia

ISA 2022

Thomas Scherer
(and others)

The Vision

A Conflict Data Set that:

- Draws upon all existing data
- Take advantage of digital corpora
- Customizable Criteria (fatalities, participants, etc)
- (relatively) Rapid Updating
- Infrastructure for issues, debates, and edits

First Steps: Connecting Existing and Wikimedia

- Conflict Data and Wikimedia Review
- Data Collection and Cleaning
- What We Got: Descriptives, ED Characteristics, ED Overlap
- Capabilities and Cautions

Conflict Datasets

Table 1: Conflict Datasets Matched to Wikimedia

- Massive efforts

- Varying Unit of Analysis

- Varying criteria*

- Fallible

Dataset	Focus	Unit Analysis	Dates	Fatality	State Actors	Other Actor(s)
COWExtra	Armed Conflict	Conflict	1816-2004	1000	1	Nonstate
COWInter	Armed Conflict	Conflict	1823-2003	1000	2	-
COWIntra4.1	Armed Conflict	Conflict-Actor	1818-2008	1000	1	Nonstate
COWintra5.1	Armed Conflict	Conflict-Actor	1818-2014	1000	1	Nonstate
COWNon	Armed Conflict	Conflict	1818-2005	1000	0	Nonstate
DividedArmies	Armed Conflict	Conflict-Actor-Phase	1800-2011	500	2	-
ICB	Crisis	Conflict; Conflict-Actor	1918-2016	NA	2	-
IMI	Intervention	Conflict-Dyad	1946-2006	NA	2	-
IWD	Armed Conflict	Conflict-Actor-Year	1823-2003	1000	2	-
NACH	Armed Conflict	Conflict	1518-1899	20 annum	1	Nat Am
PITF	Armed Conflict	Conflict-Year	1948-2018	1000	1	Rev or Ethnic
UCDP/PRIO	Armed Conflict	Conflict-Year	1946-2018	25 annum	1	Any

Wikimedia as a Source

- Quality?

- Comparable to Encyclopedia Britannica (Blumenstock 2008)
- High accuracy on political parties and elections (Brown 2011)
- Historical events akin to Dictionary of American History and American National Biography Online
 - Discrepancies on oft disputed figures: precise size of armed forces, casualties (Rector 2008)
- Most (81%) uncited claims are quickly flagged and fixed
(Adler et al. 2011; Tramullas, Garrido-Picazo, and Sánchez-Casabón 2016)

- Structure

- Wikipedia pages connected to structured wikidata pages
- Wikidata accessible by SPARQL queries
 - (or R via WikidataQueryServiceR package)
- Class structure
- Multi-lingual



- Main page
- Contents
- Current events
- Random article
- About Wikipedia
- Contact us
- Donate
- Contribute
- Help
 - Learn to edit
 - Community portal
 - Recent changes
 - Upload file
- Tools
- What links here
- Related changes
- Special pages
- Permanent link
- Page information
- Cite this page
- Wikidata item
- Print/export
 - Download as PDF
 - Printable version

World War I

From Wikipedia, the free encyclopedia

"World War One", "Great War", and "WWI" redirect here. For other uses, see [World War One \(disambiguation\)](#), [Great War \(disambiguation\)](#), and [WWI \(disambiguation\)](#).

World War I, often abbreviated as **WWI** or **WW1**, also known as the **First World War** and contemporaneously known as the **Great War** and by *other names*, was an *international conflict* that began on 28 July 1914 and ended on 11 November 1918. It involved much of [Europe](#), as well as [Russia](#), the [United States](#) and [Turkey](#), and was also fought in the [Middle East](#), [Africa](#) and parts of [Asia](#). One of the *deadliest conflicts in history*, an estimated 9 million were killed in combat, while over 5 million civilians died from occupation, bombardment, hunger or disease.^[4] The [genocides](#) perpetrated by the Ottomans and the 1918 [Spanish flu](#) pandemic spread by the movement of combatants during the war caused many millions of additional deaths worldwide.^{[5][6]}

In 1914, the [Great Powers](#) were divided into two opposing alliances: the [Triple Entente](#), consisting of [France](#), [Russia](#), and [Britain](#), and the [Triple Alliance](#), made up of [Germany](#), [Austria-Hungary](#), and [Italy](#). Tensions in the [Balkans](#) came to a head on 28 June 1914 following the [assassination of Archduke Franz Ferdinand](#), the Austro-Hungarian heir, by [Gavrilo Princip](#), a [Bosnian Serb](#). Austria-Hungary blamed [Serbia](#) and the interlocking alliances involved the Powers in a series of diplomatic exchanges known as the [July Crisis](#). On 28 July



Clockwise from the top: The road to Bapaume in the aftermath of the Battle of the Somme, 1916 • British Mark V tank in the mud, 1917 • HMS Irresistible hitting a mine in the Dardanelles, 1915 • German biplane fighters near Douai, France, 1918

WIKIDATA

[Main page](#)
[Community portal](#)
[Project chat](#)
[Create a new item](#)
[Recent changes](#)
[Random item](#)
[Query Service](#)
[Nearby](#)
[Help](#)
[Donate](#)

Statements

instance of	<div><div><div><div><div></div><div>world war</div></div></div><div><div><div></div><div>series ordinal</div></div><div><div>1</div></div></div><div><div><div></div><div>0 references</div></div></div></div></div>
historical period	<div><div><div><div><div></div><div>historical period</div></div></div><div><div><div></div><div>0 references</div></div></div></div></div>
part of	<div><div><div><div><div></div><div>late modern period</div></div></div><div><div><div></div><div>0 references</div></div></div></div></div>

English | Not logged in | Talk | Contributions | Create account | Log in

Item | Discussion | Read | View history | Search Wikidata

World War I (Q361)

1914–1918 global war, centered in Europe, between the Allied and Central Powers

World War One | WW1 | World War 1 | World War | the First World War | 1st World War | the 1st World War | WW I | The Great War | WW 1 | World War, 1914-1918 | Great War | WWI | The War to End All Wars | First World War

[In more languages](#)

Configure


Language	Label	Description	Also known as
English	World War I	1914–1918 global war	World War One

Wikidata Class Browser : war (Q198)

Class item ID or Label/Alias


Name/description language code


[view root classes](#)

Name: [war \(Q198\)](#) 


Description: [organised and prolonged violent conflict between states](#)

Parent classes: [armed conflict \(Q350604\)](#), [occurrence \(Q1190554\)](#), [social issue \(Q1920219\)](#)



Subclass is instance of: 1 [\(SPARQL query\)](#) 

Direct subclasses: 34 [\(SPARQL query\)](#) 









Indirect subclasses: 30

Direct instances: 1 179 [\(SPARQL query\)](#) 

Indirect instances: 524

Lists of: 221 [\(SPARQL query\)](#)  List type qualifiers: [SPARQL query](#) 

Direct subclasses

Name	Wikidata link	Direct subclasses	Indirect subclasses	Direct instances	Indirect instances
civil war	Q8465 	6	0	226	18
religious war	Q1827102 	2	0	34	5
war of succession	Q1348385 	2	0	32	0
war of independence	Q21994376 	1	0	16	25
military intervention	Q5919191 	1	2	15	101
proxy war	Q864113 	2	0	6	0
world war	Q103495 	0	0	6	0
range war	Q7292715 	0	0	4	0

Coding Conflicts

- Connect each row to a wikipedia page QCode
- Focus on actors, locations, time-spans
- Record confidence on match 1-3
- Allow 2 matches
- 2 Coders* per dataset independently
- Encouraged to use non-English pages (with translation assistance)

QCoding Case (COW Nonstate)

- Central American Confederation War:
 - Conservative Confederation and Liberals , 1826-1829
- Search Engine → [Federal Republic of Central America](#)
 - 1938 -1940 war between Cons and Libs - “Dissolution of the Central American Confederation”
 - “Honduras and Nicaragua joined the rebellion and Arce was deposed in **1829**. The victors led by the Honduran [Francisco Morazán](#) took power and Morazán was proclaimed president in 1830.”
- Francisco Morazán
 - Section on the civil war unhelpful
 - Go to Spanish page and get a nice link:

Guerra civil centroamericana [editar]

Artículo principal: Guerra civil centroamericana

Después de su victoria en 'La Trinidad', Morazán surgió como a ser reconocido por sus habilidades militares en toda Am

IA
pedia

First Central American Civil War

From Wikipedia, the free encyclopedia

The **First Central American Civil War** was a civil political and military conflict within the [Federal Republic of Central America](#) which lasted from 1826 until 1829. The civil war was fought between Liberal and Conservative lines with [Francisco Morazán](#) leading

First Central American Civil War

Part of the [Central American Civil Wars](#)



The [Battle of La Trinidad](#) on a Honduran 5 lempira bill.

Date 10 October 1826 – 25 June 1829

Coding Results

Table 2: Summary of Results from Matching to Wikimedia

dataset	n	No Matches	Average Match Count	Average Match Confidence	Average Max Confidence
COWExtra	198	5	1.4	2.6	2.7
COWInter	337	0	2.1	2.9	3.0
COWIntra4.1	442	1	2.1	2.5	2.7
COWIntra5.1	420	14	1.0	2.4	2.4
COWNon	62	1	2.0	2.3	2.5
DividedArmies	825	17	1.3	2.5	2.8
ICB	1,065	29	1.8	2.5	2.7
IMI	1,114	89	1.2	2.2	2.3
IWD	628	1	1.0	2.9	3.0
NACH	148	6	2.0	2.5	2.7
PITF	1,598	0	1.2	2.5	2.5
UCDP/PRIO	2,384	0	2.4	2.7	2.9

Identify ‘Best Match’

- 1. Multiple QCodes (5,066)
- 2. Highest confidence (1,133)
- 3. Most votes (942)
- 4. Most connected (284)
- 5. Most ‘conflictiness’ (124)
- 6. Reported first (22)

‘Conflictiness’	
1.	War
2.	Civil War
3.	Rebellion
4.	Revolution
5.	Insurgency
....	
12.	Military Unit
....	
21.	Occurrence

dataset	identifier	dataset_unique_identifier	QCode	Confidence
COWExtrastate	300	COWExtrastate_300_200_-8	Q891806	3
DividedArmies	1	DividedArmies_1_16_220_5/2/1808	Q152499	3
ICB	393	ICB_393_IRQ_1990_11_29	Q37643	3
IWD	WorldWarI	IWD_WW1Ger-France_220_1914	Q361	3
NACH	685	NACH_685	Q8049168	2
PITF	133	PITF_133_2000	Q1960733	3
UCDPPRIO	249	UCDPPRIO_249_1959	Q8740	3

(n = 8,077)

Instance_Of	All	UCDP/PRIO	COWInter	COWIntra	COWIntraV5	COWNon	COWExtra	IMI	ICB	DividedArmies	IWD	PTTF	NACH
war	15.8%	18.7%	63.0%	12.7%	13.3%	24.5%	32.8%	13.9%	16.1%	46.2%	61.0%	19.4%	40.4%
conflict	14.6%	18.7%	7.0%	15.4%	15.6%	16.3%	17.5%	9.8%	7.8%	17.9%	10.0%	14.5%	27.3%
civil war	5.4%	14.3%	3.0%	22.8%	23.5%	10.2%	1.5%	6.4%	5.2%	8.4%	2.0%	26.7%	1.0%
battle	4.0%	0.6%	4.0%	1.5%	1.4%	6.1%	3.6%	0.9%	7.3%	4.4%	2.0%	-	6.1%
rebellion	3.6%	2.0%	1.0%	9.3%	9.5%	8.2%	5.1%	1.4%	0.7%	2.4%	1.0%	2.4%	4.0%
military operation	3.4%	1.5%	-	0.4%	1.0%	-	0.7%	5.9%	4.7%	-	-	-	4.0%
coup d'état	2.6%	6.7%	-	1.5%	1.0%	-	-	2.3%	1.7%	-	-	1.8%	-
human	2.5%	0.3%	1.0%	3.5%	4.1%	10.2%	4.4%	1.4%	0.5%	0.8%	1.0%	1.2%	1.0%
revolution	1.6%	2.3%	-	6.6%	5.4%	2.0%	2.2%	0.9%	1.2%	0.8%	-	4.8%	-
organization	1.4%	0.3%	-	0.4%	0.7%	-	-	3.6%	0.5%	-	-	-	-

Figure 1: Overlap of Conflict Datasets

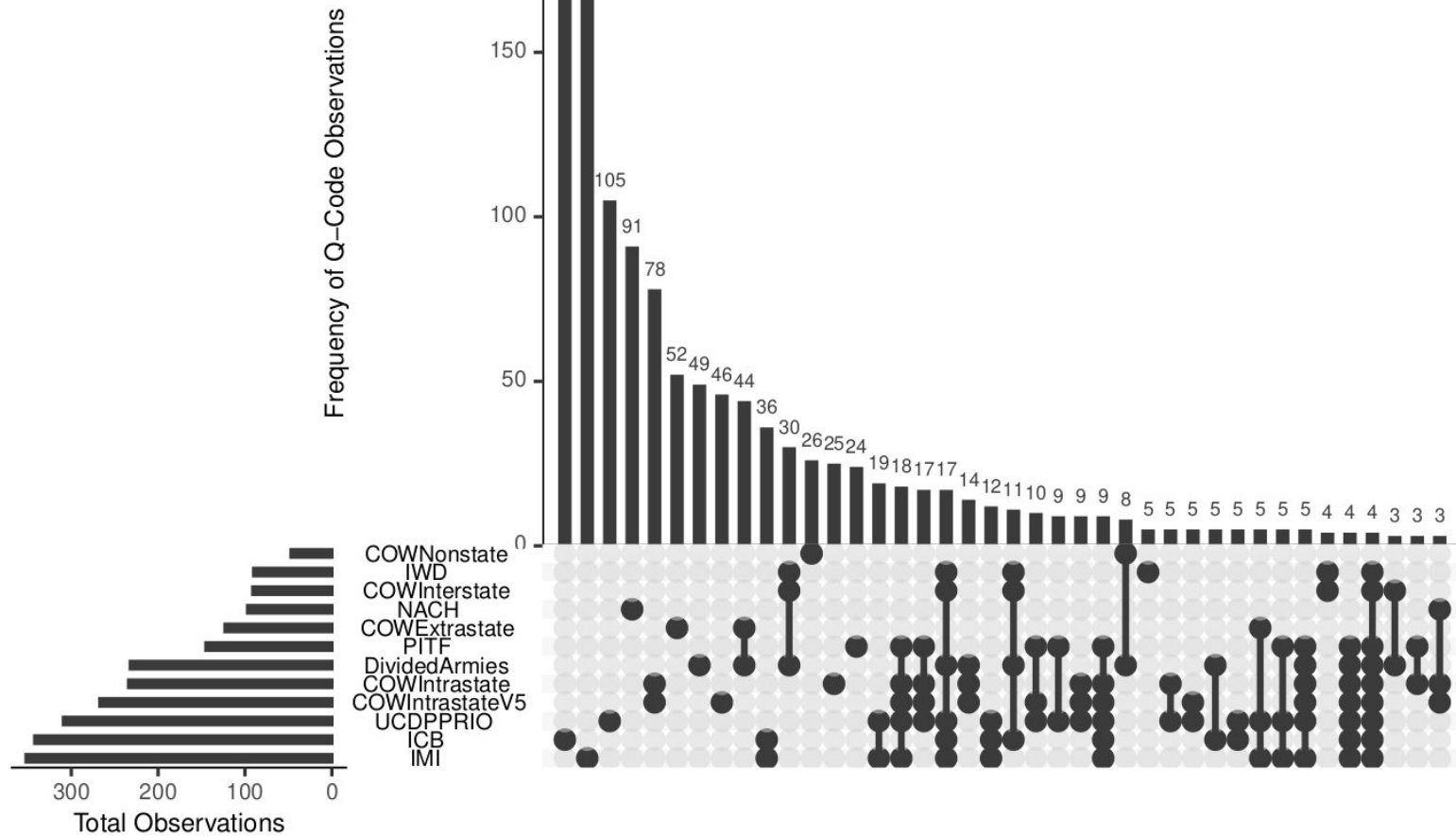


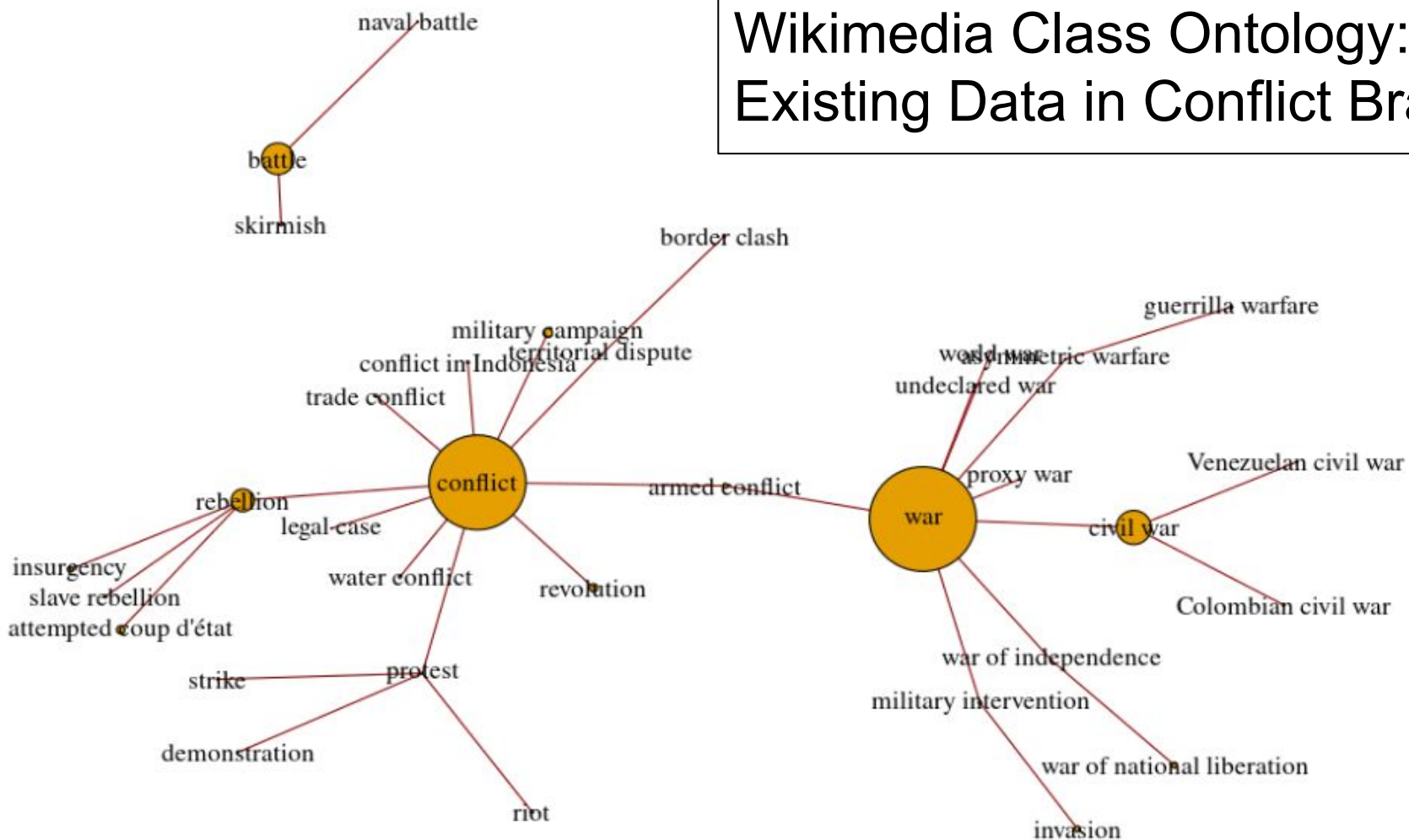
Table 5: Percent of each dataset's entries (row) with QCodes found in the other datasets (column)

dataset	COWExtra	COWInter	COWIntra	COWIntraV5	COWNon	DividedArmies	ICB	IMI	IWD	NACH	PITF	UCDPPRIO
COWExtra	100%	3%	7%	6%	7%	46%	8%	16%	3%	0%	11%	23%
COWInter	2%	100%	14%	13%	3%	92%	59%	39%	90%	1%	13%	41%
COWIntra	3%	10%	100%	88%	7%	29%	32%	36%	8%	1%	46%	50%
COWIntraV5	4%	5%	84%	100%	5%	23%	21%	28%	4%	2%	40%	46%
COWNon	13%	0%	13%	11%	100%	30%	7%	6%	0%	2%	4%	4%
DividedArmies	18%	46%	19%	21%	9%	100%	37%	18%	43%	1%	11%	20%
ICB	3%	27%	18%	18%	3%	34%	100%	42%	26%	0%	17%	35%
IMI	5%	29%	27%	26%	4%	37%	56%	100%	28%	0%	29%	63%
IWD	1%	97%	16%	13%	1%	95%	69%	34%	100%	1%	14%	36%
NACH	0%	1%	4%	4%	8%	2%	0%	0%	1%	100%	0%	0%
PITF	8%	13%	76%	79%	5%	22%	34%	57%	10%	0%	100%	89%
UCDPPRIO	15%	10%	56%	62%	3%	17%	27%	46%	9%	0%	63%	100%

Validation

- Comparison where there should be overlap
 - All UCDP/PRIO with >1000 fatalities (culumlative_intensity = 1) should be in COW
 - All COW post 1946 should be in UCDP/PRIO
 - IWD and COW Interstate have same criteria (allegedly)
 - PITF should be in UCDP/PRIO
 - COWInterstate should be in Divided Armies
- Improvements
 - Consider using 'part of' property
 - Toyota War (86-87) is 'part of' 'Chadian Libyan conflict' (78-87)
 - BUT! Ethiopian-Somali War 'part of' Cold War
 - disallow 'category' pages (i.e. Category:Wars involving Panama)
 - Use most up to date (COWIntrastate v5)
- To Do: Compare properties across

Wikimedia Class Ontology: Existing Data in Conflict Branch



Wikimedia Class Tree: Entire Conflict Branch

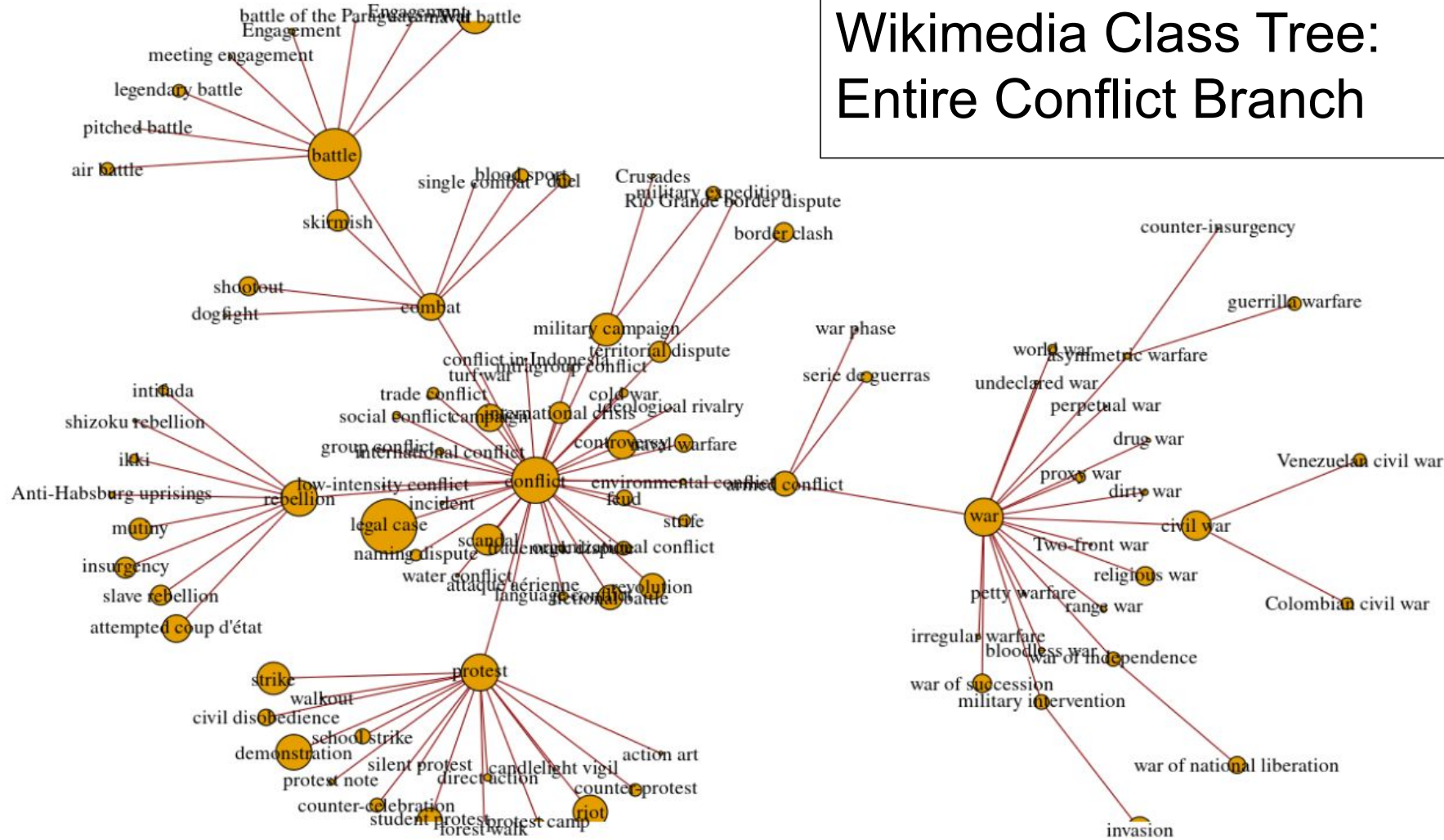


Table 6: Description of Specified Wikimedia Classes

entity type	In Existing Conflict Data	In Wikimedia	Percent with start date	Percent with end date	Percent with participants
conflict	244	1,378	1	1	0.05660377
war	259	532	1	1	0.83082707
civil war	85	134	1	1	0.43283582
battle	87	2,311	1	1	0.36088273

Conclusion

- Cautions

- Missingness, Miscodings, Unexpected Structure (legal cases)
- Challenges in levels of analysis (actors, time spans, phases, events)
Toyota War:Chadian Libyan Conflict, Ethiopian-Somali War:Cold War

- Capabilities

- We have a 'Rosetta Stone' for Conflict Datasets
 - Ease connecting new data to existing conflict datasets
- Potential source for augmenting conflict data
- Potential for larger project (updating, multi-lingual, potential for machine learning)
- Wikipedia as a forum to address differences?
 - Don't want to bake in our results (cooking the wikibooks)
 - Do want to improve for education and data extraction
 - Wikimedia Researcher Policies and Guidelines?