



**HAL**  
open science

# WDPprop: Web Application to Analyse Multilingual Aspects of Wikidata Properties

John Samuel Samuel

► **To cite this version:**

John Samuel Samuel. WDPprop: Web Application to Analyse Multilingual Aspects of Wikidata Properties. OpenSym 2021: 17th International Symposium on Open Collaboration, Sep 2021, Online, Spain. pp.1-12, 10.1145/3479986.3479996 . hal-03621622

**HAL Id: hal-03621622**

**<https://hal.science/hal-03621622v1>**

Submitted on 31 Mar 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# WDProp: Web Application to Analyse Multilingual Aspects of Wikidata Properties

JOHN SAMUEL, CPE Lyon, LIRIS, UMR-CNRS 5205, Université de Lyon, France

Compared to Wikipedia, Wikidata is a single domain website with the possibility to view the information in multiple languages. Translation plays a significant role in Wikidata. Unlike Wikidata items, Wikidata properties are influenced less by translation bots and require a meaningful amount of human effort. The study of Wikidata property creation and translation is, therefore, very essential. Since the inception of Wikipedia, several research works have focused on the information flow among different language Wikipedias. The attention has now shifted to the way information on Wikidata is created and translated. The focus of this article is the Wikidata properties. WDProp is a web application created to understand and obtain an integrated view on the various multilingual aspects of Wikidata properties, from their proposition to their use on multiple domains.

CCS Concepts: • **Software and its engineering** → **Software libraries and repositories**; • **Information systems** → **Collaborative and social computing systems and tools**; **Open source software**; **Data mining**.

Additional Key Words and Phrases: Wikidata, Wikidata Properties, Collaboration, Multilingual Development

## ACM Reference Format:

John Samuel. 2021. WDProp: Web Application to Analyse Multilingual Aspects of Wikidata Properties. In *17th International Symposium on Open Collaboration (OpenSym 2021), September 15–17, 2021, Online, Spain*. ACM, New York, NY, USA, 18 pages. <https://doi.org/10.1145/3479986.3479996>

## 1 INTRODUCTION

Wikidata [17], a Wikimedia project created in 2012 is a free, linked, open, collaborative, and multilingual knowledge base. There are significant differences between Wikidata and Wikipedia. Unlike Wikipedia, which contains the majority of information in unstructured form (e.g., textual description), Wikidata is a structured knowledge base. Wikipedia uses multiple subdomains for different languages, each of which is managed by the associated language community. Wikidata, on the other hand, is a single domain website. All the information related to a single topic in multiple languages is referred to by a single URL. Logged-in users may be able to view the information in their local language by configuring the local language in their settings. Users can also make use of *uselang = lc*, in the URL to view the information in any other language *lc*, where *lc* is the language code. This is indeed a major change for Wikipedia users used to multiple websites for looking up information in different languages. Such an approach of building a central store for storing facts (or statements) related to multiple domains is challenging. This approach of collaborative multilingual and multi-domain ontology development on Wikidata has been the focus of many recent works [6, 8, 11].

Nevertheless, it is important to state that not all communication on Wikidata exists in a multilingual way. There exist specific pages where contributors need to use monolingual text for expressing their opinions, like in discussion and User:Talk pages. Whether a truly multilingual experience can be obtained on Wikidata is an open and relevant research question. This article focus on a small, but the key aspect of Wikidata: properties to comprehend the multilingual

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2021 Association for Computing Machinery.

Manuscript submitted to ACM

53 aspects around them. Wikidata properties play a major role in describing knowledge, across domains. Unlike Wikidata  
 54 items, they cannot be simply created but requires discussion and voting. Depending on the need, the properties are  
 55 proposed by the Wikidata contributors. Such propositions are discussed by the community members, where some may  
 56 point to the availability of existing properties and others agreeing to the need for the creation of the property. Every  
 57 property proposition undergoes voting and if a property has achieved a sufficient number of supporting votes, the  
 58 property may be created and made available for use.

60 A newly created property may not be available in all the languages and may have one or more translations, based  
 61 on the initial translations proposed by the property proposer. This makes it difficult for users who do not speak or  
 62 comprehend any of the available translations. The role played by bilingual or multilingual speakers is therefore very  
 63 important. It is also difficult to use the properties, immediately after their creation, if their usage is not properly described  
 64 and documented. For this purpose, Wikidata WikiProjects for different domains have been created, which curates the  
 65 properties and documents the usage for describing Wikidata items belonging to the given domain. The problem of  
 66 finding relevant properties has also garnered some attention in the research community. There have been several works  
 67 on predicting properties [3, 18] for describing domain knowledge, including a recent work on their prediction from  
 68 Wikipedia (e.g., Wiki2Prop [10]). It is important to state here that such property suggestions or predictions may require  
 69 some validation by the contributors and hence their translations cannot be overlooked.

73 Wikidata is highly evolving. It is difficult to find and track the latest number of Wikidata supported languages,  
 74 datatypes, properties, translation statistics, etc. Some contributors are interested in a subset of curated information  
 75 and the associated translation statistics. In this article, we discuss the development of WDDrop as a solution to these  
 76 problems. Section 2 gives a brief introduction of Wikidata, particularly focusing on properties. Some works on Wikidata  
 77 properties are presented in section 3. Section 4 presents the development of WDDrop and the associated results. Starting  
 78 from some coarser analysis, it shows the results related to some fine-granular analysis. Finally, section 5 concludes the  
 79 article and presents the future course of action.

## 82 2 WIKIDATA PROPERTIES

84 Wikidata item pages are used to describe different classes like museum, person, archive, website, conference, scientific  
 85 event<sup>1</sup> etc. and their instances like OpenSym 2019<sup>2</sup>, Wikidata<sup>3</sup> etc. In addition to Wikidata item pages, there are property  
 86 pages (e.g., instance of<sup>4</sup>, country (P17)<sup>5</sup>, official website (P856)<sup>6</sup>, etc.), entity schema pages (Wikimedia project<sup>7</sup>) and the  
 87 recently introduced lexicographical pages (e.g., lexeme (L315)<sup>8</sup>). Properties are used to describe items (e.g. the country  
 88 of OpenSym 2019 was described using P17). Entity schemas are used to validate the Wikidata properties and items. It is  
 89 worth mentioning here that all these pages have an identifier of the form [QPLE][0-9][0-9]\*. Unlike Wikipedia, the  
 90 contributors may not know the topic of the page from the URL, without looking at the page content. For every such  
 91 page, there is a discussion page (e.g., Wikidata (Q2013) talk page<sup>9</sup>), where contributors discuss the ways to improve the  
 92 associated page.

97 <sup>1</sup><https://www.wikidata.org/wiki/Q52260246>

98 <sup>2</sup><https://www.wikidata.org/wiki/Q56259215>

99 <sup>3</sup><https://www.wikidata.org/wiki/Q2013>

100 <sup>4</sup><https://www.wikidata.org/wiki/Property:P31>

101 <sup>5</sup><https://www.wikidata.org/wiki/Property:P17>

102 <sup>6</sup><https://www.wikidata.org/wiki/Property:P856>

103 <sup>7</sup><https://www.wikidata.org/wiki/EntitySchema:E2>

104 <sup>8</sup><https://www.wikidata.org/wiki/Lexeme:L315>

<sup>9</sup><https://www.wikidata.org/wiki/Talk:Q2013>

It is difficult to remember all the identifiers of properties, items, etc., especially for usage by humans. Properties, items, and entity schemas need to be translated, i.e., these pages must have labels, descriptions, and aliases in any of the supported Wikidata languages. A property, item, and entity schema can have one label, one description, and multiple aliases in any supported language. Such translations are useful for finding the relevant items through the search interface. Thus contributors can find the relevant identifier by searching their local names.

Properties have datatypes. For example, to specify the official website of OpenSym, we may need a property with the datatype URL. Property P856 (official website) is one such example belonging to the datatype URL. There are several other datatypes like external identifiers, media, geographical coordinates, etc. Properties are not just limited to Wikidata, there are properties in other external data stores like DBpedia, VIAF, OCLC, ISBN, etc. Properties could be matched to properties of other external data stores through a special property called equivalent property<sup>10</sup>. Items can also be linked to entities in other external data stores through the properties belonging to the datatype: external identifier.

### 2.1 Property proposal and creation

Unlike Wikidata items, which require that any item created to follow the notability guidelines<sup>11</sup>, the creation of property items is a much longer process. Figure 1 shows the process of Wikidata property creation and its possible deletion.

A contributor looking to describe an item may not find a relevant property to describe an item. In this case, they may propose a new property on the property proposal page<sup>12</sup>. A selected number of topics have been created for allowing the contributors to discuss and find contributors belonging to a particular domain. Some examples include authority control, person, place, transportation, etc.

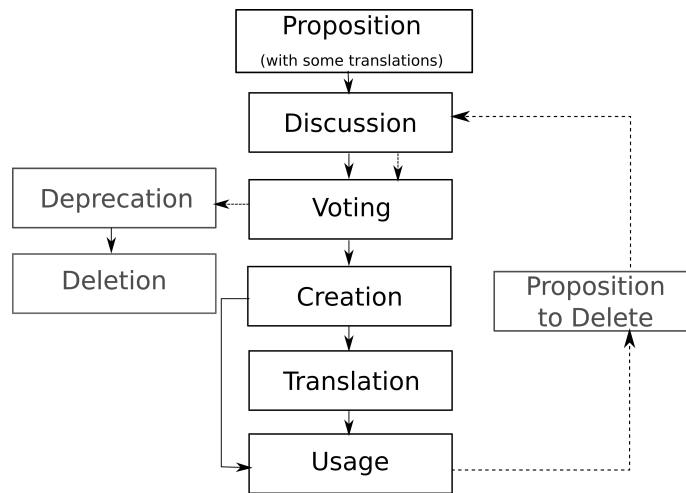


Fig. 1. Property proposition, creation and possible deletion. Dark boxes and arrows shows the steps related to property creation, translation and usage. Light gray boxes and dotted arrows shows the steps related to property deletion. Some steps are shared by property creation and deletion, like discussion and voting.

<sup>10</sup><https://www.wikidata.org/wiki/Property:P1628>

<sup>11</sup><https://www.wikidata.org/wiki/Wikidata:Notability>

<sup>12</sup>[https://www.wikidata.org/wiki/Wikidata:Property\\_proposal](https://www.wikidata.org/wiki/Wikidata:Property_proposal)

157 An example property discussion is highlighted in Figure 2. This is the property proposal discussion of P3966  
 158 (programming paradigm)<sup>13</sup>. Figure 2a shows the information proposed by the property proposer. Figure 2b shows the  
 159 voting and in Figure 2b, we see the source code of the property proposition.

160 Readers may refer to the French<sup>14</sup> or the Spanish<sup>15</sup> rendering of this property discussion page. We see that some of  
 161 the information like the French or the Spanish translations are available for some information like Support, Oppose,  
 162 Comment. These template pages support, oppose, comment have already been translated in these languages.

163 As can be seen, the property proposer (or/and some other contributors) translated the property label and description  
 164 in some languages. For this purpose, TranslateThis template has been used. At the time of property creation, these  
 165 labels and descriptions made available, as seen in Figure 1. Properties may be translated into other languages after  
 166 creation. It should be stressed that the labels, descriptions, and aliases may be modified during the lifespan of a property.  
 167 Some of these modifications may be possible vandalism and require to be detected<sup>[16]</sup>. Finally, the lifespan of a property  
 168 may be limited. Sometimes, contributors may propose that a property should be deleted<sup>16</sup>. This may be because of  
 169 the need to change the datatype or because a new and better property has been proposed with which the concerned  
 170 property can be merged.  
 171  
 172  
 173  
 174

## 175 2.2 Property curation: WikiProjects and Property Classes

176 Another major problem is the difficulty in finding relevant property items for describing an item. Take for example, the  
 177 following three properties:  
 178  
 179

- 180 • country (P17) used for specifying the country of the item.
- 181 • country of origin (P495) used for specifying the country of origin of the item, when the item is a creative work,  
 182 food, etc.
- 183 • country of citizenship (P27) for specifying the nationality of a person (item).  
 184  
 185

186 A new contributor may not be aware of all such properties, and often may end up using P17 (country) for specifying  
 187 the country of origin of a creative work. For helping the contributors, there are Wikidata WikiProjects<sup>17</sup> that curate the  
 188 properties relevant to a particular domain. For example, there are WikiProjects related to books<sup>18</sup>, museums<sup>19</sup>, etc.  
 189 Another possible way is to curate them using the property: properties for this type (P1963), that can be used to find the  
 190 properties for a given type (a class etc.)  
 191  
 192  
 193

## 194 2.3 Property translation and evolution

195 As seen in Figure 1, Wikidata properties constantly evolve. Properties are created, modified, translated and even deleted.  
 196 New languages, datatypes, and WikiProjects are also created from time to time. To improve the language coverage and  
 197 coverage of topics from across the world, the translation of properties cannot be overlooked. Property translation and  
 198 evolution is the focus of this article.  
 199  
 200  
 201

202 <sup>13</sup>[https://www.wikidata.org/wiki/Wikidata:Property\\_proposal/programming\\_paradigm](https://www.wikidata.org/wiki/Wikidata:Property_proposal/programming_paradigm)

203 <sup>14</sup>[https://www.wikidata.org/wiki/Wikidata:Property\\_proposal/programming\\_paradigm?uselang=fr](https://www.wikidata.org/wiki/Wikidata:Property_proposal/programming_paradigm?uselang=fr)

204 <sup>15</sup>[https://www.wikidata.org/wiki/Wikidata:Property\\_proposal/programming\\_paradigm?uselang=es](https://www.wikidata.org/wiki/Wikidata:Property_proposal/programming_paradigm?uselang=es)

205 <sup>16</sup>[https://www.wikidata.org/wiki/Wikidata:Properties\\_for\\_deletion](https://www.wikidata.org/wiki/Wikidata:Properties_for_deletion)

206 <sup>17</sup><https://www.wikidata.org/wiki/Wikidata:WikiProject>

207 <sup>18</sup>[https://www.wikidata.org/wiki/Wikidata:WikiProject\\_Books](https://www.wikidata.org/wiki/Wikidata:WikiProject_Books)

208 <sup>19</sup>[https://www.wikidata.org/wiki/Wikidata:WikiProject\\_Museums](https://www.wikidata.org/wiki/Wikidata:WikiProject_Museums)

209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260

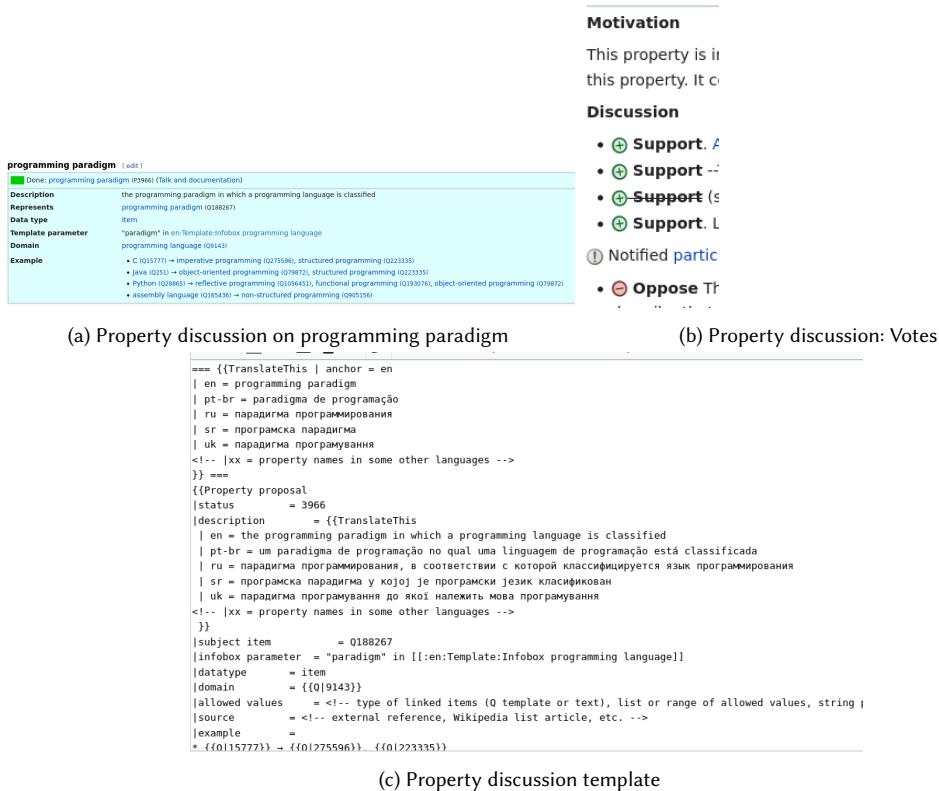


Fig. 2. Property proposition and discussion of P3966 (programming paradigm)

### 3 STATE OF THE ART

Wikidata provides a SPARQL endpoint and Mediawiki API for accessing the structured data as well as the commit history. A number of tools have been created during the past few years<sup>20</sup>, which help contributors to describe, analyse and visualize Wikidata items. Tools like PropBrowser<sup>21</sup>, SQID property browser<sup>22</sup> focus on Wikidata properties. Some commonly used tools on Wikidata by contributors for generating lists are Listeria<sup>23</sup> and InteGraality<sup>24</sup>. This has been used by certain contributors to maintain a list of datatypes<sup>25</sup>. However, these tools generate the lists in a periodic manner (like weekly updates etc.)

Collaborative [11] and multilingual aspects [6] of Wikidata has been the focus of many recent works. Many of these works make use of periodic Wikidata datadumps, focusing on the global coarser analysis. But the contributors focus on some properties or a subset of curated properties for their works. Certain visualizations like the visualization of deletion discussions considered by Notabilia [15] or information flow across different language Wikipedias [1, 4] give interesting insights. While analysing and visualizing the translations of Wikidata, one may observe the multilingual

<sup>20</sup><https://hay.toolforge.org/directory/>  
<sup>21</sup><http://tools.wmflabs.org/hay/propbrowse>  
<sup>22</sup><https://sqid.toolforge.org/#/browse?type=properties>  
<sup>23</sup><https://www.wikidata.org/wiki/Wikidata:Listeria>  
<sup>24</sup><https://integraality.toolforge.org/>  
<sup>25</sup>[https://www.wikidata.org/wiki/Wikidata:WikiProject\\_Properties/Reports/Datatypes](https://www.wikidata.org/wiki/Wikidata:WikiProject_Properties/Reports/Datatypes)

nature of the contributors and also the role played by the bots[5]. Some recent work have focuses on the stability of property labels[16]. It must be noted that the property labels are meant for human consumption and undesired or possible vandalism on the properties may produce unexpected semantics of information. Hence it is important to obtain real-time and fine-granular analysis of Wikidata properties.

Translation of properties may aid to extracting structured information from Wikipedia articles to populate infoboxes [9] and Wikidata<sup>26 27</sup>. Finally, it is equally important to point the role played by Wikidata WikiProjects [7, 8] in curation of properties. The focus of WDProp proposed in this article is on Wikidata properties and the multilingual aspects around the properties. The goal is to navigate and explore properties by languages, datatypes, WikiProjects, etc.

#### 4 WDPROP: FROM COARSER ANALYSIS TO FINE-TUNED ANALYSIS

One approach for analysis of Wikidata items and properties is to download Wikidata datadump<sup>28</sup> with the commit history for the study. Another approach is to make use of SPARQL endpoints and Wikidata Mediawiki API, especially when the focus is on a small subset of information. *Multilingual Wikidata Property Translation Flow Dataset* [2] was created using the latter approach. Datadumps, complete or selective are quite useful for coarser analysis and may provide some useful insights [6, 11].

Table 1 shows the details of *Multilingual Wikidata Property Translation Flow Dataset*. This dataset contains the translation flow of labels, descriptions, and aliases of Wikidata properties collected on July 7, 2019. The dataset consists of four columns: timestamp of the translation, the identifier of the property translated, the language of translation, and the type of translation (label, description, or alias). The 6347 properties considered in the study showed that a mean of 21.44 property labels in different languages was available. The minimum and the maximum number of property labels are 1 and 154 respectively. The distribution of the count of properties and the number of languages is shown in Figure 3.

Measure	Value
count	6347
mean	21.444935
std	20.897675
min	1.000000
max	154.000000

Table 1. Details of *Multilingual Wikidata Property Translation Flow Dataset* [2]. The dataset contains the details of every type of translation (labels, description and aliases) on 6347 properties.

Further analysis of these property labels was made on the count of combined occurrences of languages on different properties and the results of the top seven combinations are shown in Table 2. For example, considering a combination of four languages, English (en), Dutch (nl), Arabic (ar) and Ukrainian (uk) property labels are available on 6257 properties.

The above dataset can also be used to study the property translation flow for a given property as shown in Figure 4. It took a period between 2013 and 2019 for the translation of the labels (violet dots) of the property P17 (country) to be available in multiple languages. The first descriptions (light blue dots) were subsequently translated and the first aliases (red dots) were added after a long time. Check the distances between the violet, blue and red dots for some languages.

<sup>26</sup><https://pltools.toolforge.org/harvesttemplates/>

<sup>27</sup><https://petscan.wmflabs.org/>

<sup>28</sup>[https://www.wikidata.org/wiki/Wikidata:Database\\_download](https://www.wikidata.org/wiki/Wikidata:Database_download)

313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364

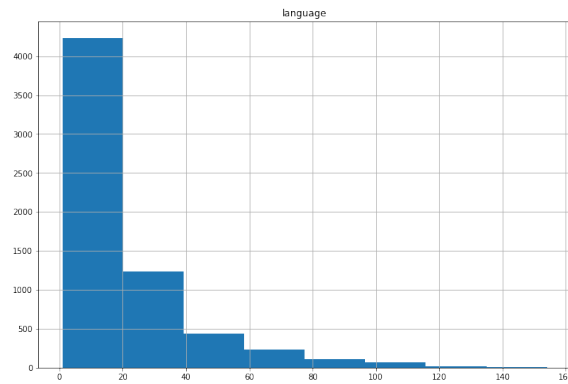


Fig. 3. Distribution of labels in different languages. The X-axis corresponds to the number of languages and the Y-axis corresponds to the number of properties. Very less number of properties have label translations in more than 100 languages.

Count	Language Combinations	Count	Language Combinations	Count	Language Combinations
6347	en	6259	ar, uk	6257	en, nl, ar, uk
6338	ar	5923	fr, ar	5919	en, nl, fr, ar
6306	nl	5905	fr, uk	5903	en, nl, fr, uk
6263	uk	4819	ar, ca	5901	en, ar, fr, uk
5928	fr	4795	ca, uk	5899	nl, fr, uk, ar
4823	ca	4530	fr, ca	4818	en, nl, ar, ca
4019	de	4015	ar, de	4794	en, nl, ca, uk

Table 2. Total number of properties with the label translations in the language combinations (Top 7)

By focusing on modifications (i.e. changes made after the first translations), a modified form of this visualization can be used to detect possible vandalism.

However, the above analysis is limited to a snapshot of data at a given time. Considering the rapidly evolving nature of Wikidata, there is a need to have some real-time view of Wikidata properties. WDProp was developed to understand and obtain an integrated view on the different aspects of properties, their creation, and translation. The focus here is on obtaining real-time statistics, on a much fine-granular basis, i.e., instead of focusing on the global analysis of properties (as seen above), the users can focus on a single property or a combination of curated properties. WDProp provides a visual interface to such information.

Property translations can be analyzed in different ways. There are two major translations: property labels and property descriptions. Properties may also have some aliases in certain languages. Whether all the aliases in one language need to be available in other languages is an open question. For a given language, the language community may wish to ensure the translation of all the property labels and descriptions. For some generic properties like country (P17), the goal of the Wikidata community members is to ensure the translation in all the supported languages.

WDProp separates property translations in the following manner:

- translated and untranslated property labels
- translated and untranslated property descriptions
- translated and untranslated property aliases



365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416

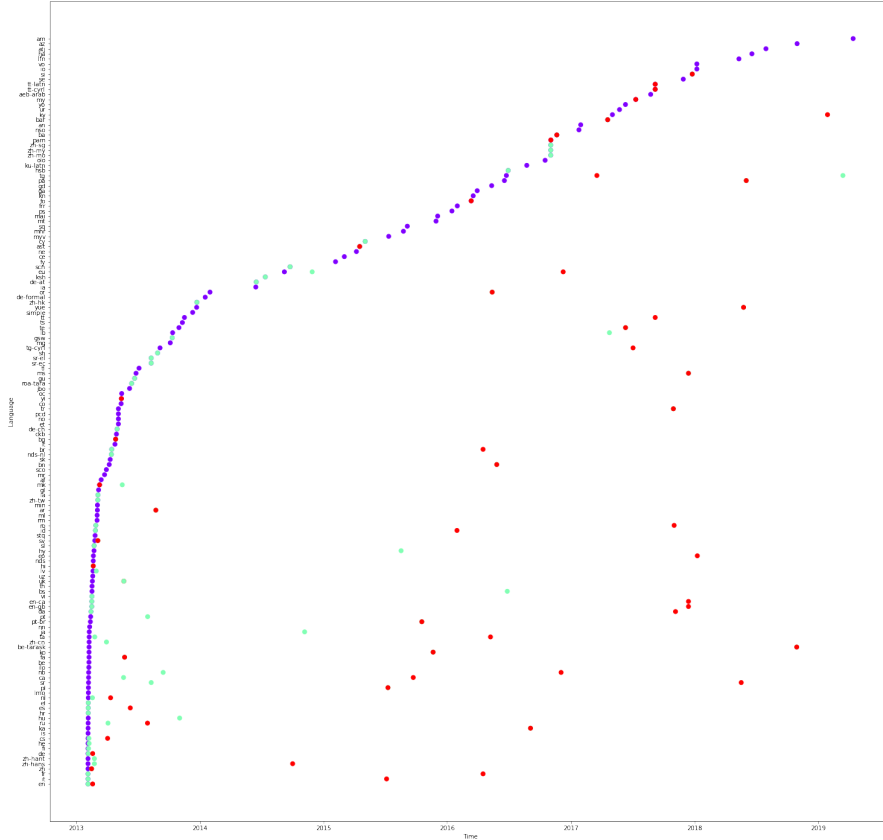


Fig. 4. Timeline of translation of labels (violet dots), descriptions (blue dots) and aliases (red dots) of property P17 (country). The Y-axis shows the languages and the X-axis shows the period between the year 2013 and 2019.

A Wikidata contributor may find one or more of the above pieces of information relevant, some of whom may wish to obtain the latest translation statistics, and some others looking for properties not yet translated.

#### 4.1 Development

WDProp was developed using basic web technologies like HTML, Javascript, and CSS. It makes use of the SPARQL queries and Wikidata MediaWiki API to obtain the latest information from the Wikidata servers. It, therefore, does not require any installation since it contains a collection of HTML, Javascript, and CSS files. This simple approach ensures





521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572

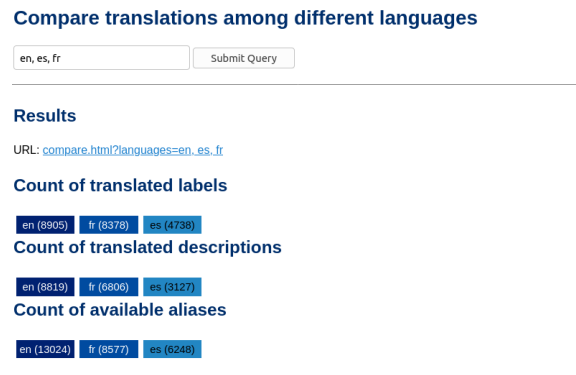
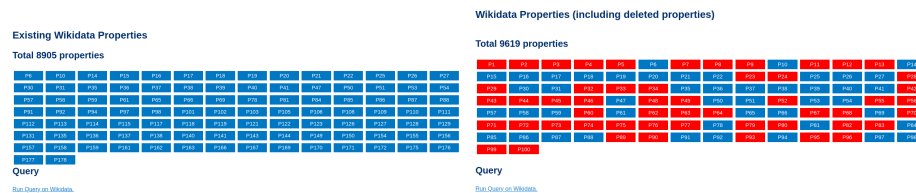


Fig. 7. Comparison of property translation statistics on WDProp: English(en), French (fr) and Spanish (es)

As discussed above, new properties are regularly proposed and voted by the community members. The properties that have been created have an associated identifier. In Figure 8, 100 properties are shown including the option to view the deleted properties<sup>34</sup>. It is interesting to observe here that many of the initial properties have been deleted.



(a) Properties (b) Properties including deleted properties

Fig. 8. List of properties on WDProp: limited to view of 100 properties

Some properties have equivalent properties in other knowledge bases. Take, for example, Property (P17) is also available on DBpedia<sup>35</sup>. Wikidata tracks such equivalent properties, which can be considered as a quality metric for property relevance. Additionally, like Wikidata items, Wikidata properties may also have references to the different statements. However, not all properties have references. In Figure 9, the usage of equivalent properties and references<sup>36</sup> are shown.

It is also possible to obtain the translation statistics for a given language. In Figure 10, current information of property translations in the Afar (aa) language<sup>37</sup> can be seen.

In Figure 11, current information of property translations in English (en) language<sup>38</sup> can be seen. When compared to Figure 10, we observe the differences in property translation in the two languages. Similar observations can be made between the English language and the languages with few speakers (or few contributors on Wikidata).

<sup>34</sup><https://wdprop.toolforge.org/properties.html>

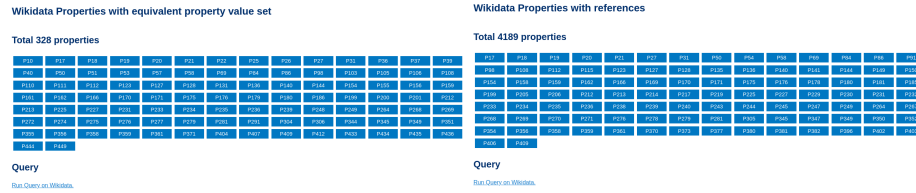
<sup>35</sup><http://dbpedia.org/ontology/country>

<sup>36</sup><https://wdprop.toolforge.org/provenance.html>

<sup>37</sup><https://wdprop.toolforge.org/language.html?language=aa>

<sup>38</sup><https://wdprop.toolforge.org/language.html?language=en>

573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624



(a) Properties with equivalent properties (b) Properties with references

Fig. 9. Equivalent properties and usage of references on Wikidata properties



(a) Property information in Afar (aa) language (b) Labels



(c) Descriptions (d) Aliases

Fig. 10. Property information in the Afar (aa) language

Figure 12 shows the current translation statistics of the key templates used for property discussions<sup>39</sup>. The numbers are lower than the total number of supported languages on Wikidata, for all four templates.

Wikidata supports several datatypes: Math, URLs, external identifiers, etc. New datatypes like wikibase:WikibaseLexeme and wikibase:WikibaseSense were recently created for representing lexicographical data. Figure 13 shows the list of available datatypes<sup>40</sup> and the properties belonging to datatype: wikibase:Math<sup>41</sup>.

Some contributors have created property classes to help to describe Wikidata items belonging to different domains, especially for new contributors who often find it difficult to find the relevant properties to describe a given entity. Figure 14 shows a snapshot of the available property classes<sup>42</sup> and highlights lighthouse (Q39715).

Figure 15 shows the details of two such curated Wikidata items: lighthouse (Q39715)<sup>43</sup> and Wikidata property related to lighthouses (Q28739677)<sup>44</sup> and shows a screenshot of some of the properties proposed by them.

<sup>39</sup><https://wdprop.toolforge.org/templates/translated.html>  
<sup>40</sup><https://wdprop.toolforge.org/datatypes.html>  
<sup>41</sup><https://wdprop.toolforge.org/datatype.html?datatype=wikibase:Math>  
<sup>42</sup><https://wdprop.toolforge.org/classes.html>  
<sup>43</sup><https://wdprop.toolforge.org/class.html?class=Q39715>  
<sup>44</sup><https://wdprop.toolforge.org/class.html?class=Q28739677>



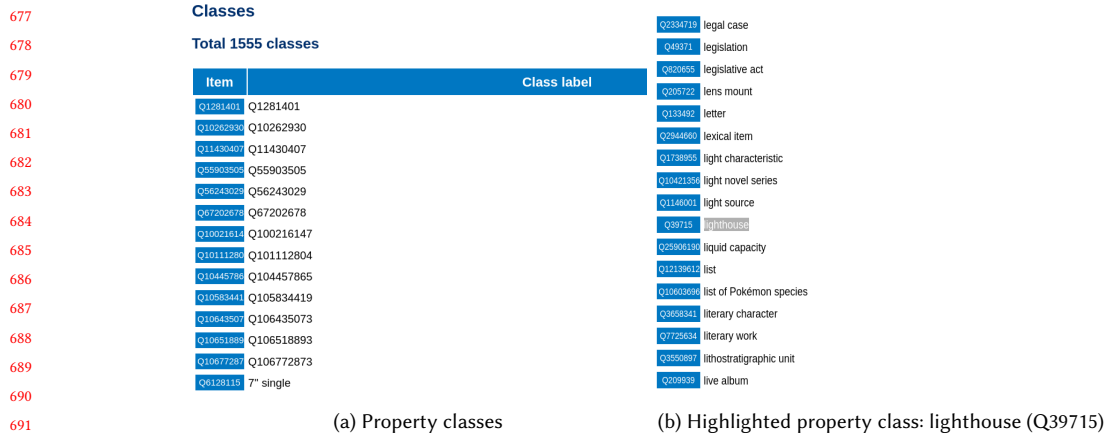


Fig. 14. Property classes: curation of properties for a particular class

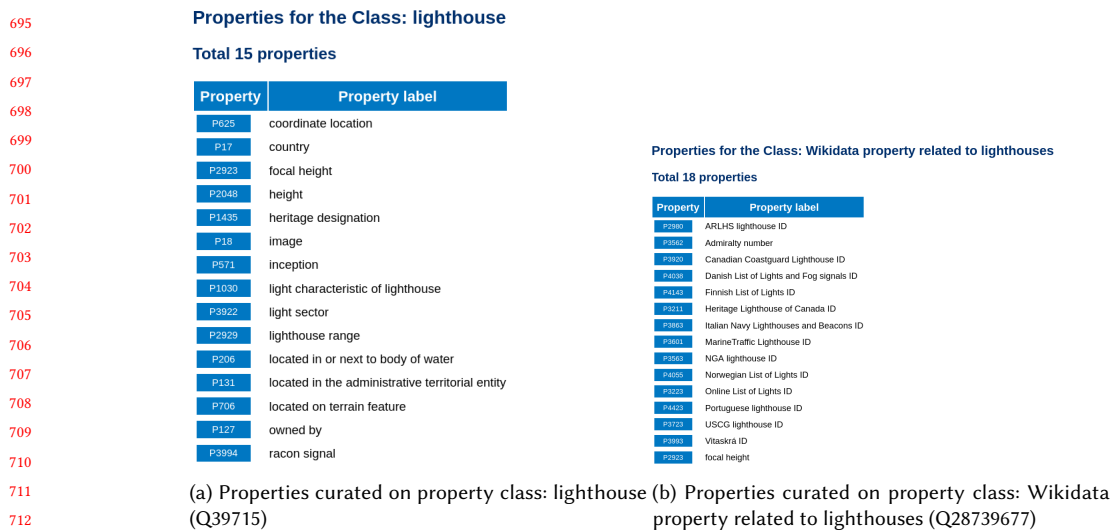


Fig. 15. Properties curated on two property classes

The curators may be interested to know the translation statistics of all the properties considered in a given class. Figure 16 shows the statistics of property translations of the two property classes.

The translation statistics of the properties curated and used on the WikiProjects can also be seen. Take, for example, Figure 17 shows the translation statistics of labels of the 37 properties used on WikiProject Museums<sup>45</sup>.

In Figure 18, the translation path of a Wikidata property, P856 (official website)<sup>46</sup> is shown. This is an extension to the work previously done in [13], where a tabular column was used to highlight the translation of property labels, descriptions, and aliases. On clicking any property page (e.g., P17<sup>47</sup>), it is possible to obtain the current translation

<sup>45</sup><https://wdprop.toolforge.org/wikiproject.html?project=Wikidata:WikiProject%20Museums>

<sup>46</sup><https://wdprop.toolforge.org/pathviz.html?property=P856>

<sup>47</sup><https://wdprop.toolforge.org/property.html?property=P17>

729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780

Count of translated labels

ca (14)	ar (15)	sv (14)	zh-hans (15)	uk (14)	sr (15)	pt (14)	nl (14)	nb (14)	mk (14)	fr (14)	es (14)	en (14)	de (14)	
				en-gb (14)	da (14)	zh (14)	tr (14)	ru (14)	it (14)	fi (14)	zh-tw (13)	lv (13)	ko (13)	
eo (13)	zh-hant (13)	be-tarask (12)	zh-hk (12)	te (12)	pl (12)	ja (12)	eu (12)	ba (11)	nn (11)	fa (11)	br (10)	bg (10)	be (10)	
				ast (10)	af (10)	he (10)	sl (10)	pt-br (10)	vi (10)	tr (10)	sq (10)	sk (10)	sh (10)	scn (10)
ru (10)	nds (10)	fr (10)	ka (10)	tr (10)	hy (10)	hu (10)	gu (10)	gsw (10)	gl (10)	et (10)	el (10)	zh-cn (10)	cy (10)	
cs (10)	bn (9)	sq (9)	th (9)	ta (9)	sco (9)	oc (9)	ms (9)	mr (9)	hsb (9)	hr (9)	ga (9)	vec (9)	sr-ec (9)	
zho (8)	ba (8)	ur (8)	rup (8)	kk (8)	ia (8)	kn (8)	dog (8)	it-cyrl (8)	nl (8)	hr (8)	sr-el (8)	zh-mo (8)	sr (8)	
zh-my (8)	zh-sg (8)	ast (7)	zue (7)	en-ca (7)	ast (7)	ilo (7)	y (7)	ku-lam (7)	ja (7)	is (7)	ha (7)	de-at (7)	pt (7)	
ne (6)	mn (6)	hr (6)	it-cyrl (6)	ce (6)	aq (6)	nis-arab (6)	ia (6)	yo (6)	kab (6)	bar (6)	fo (6)	si (6)	ti (6)	
						szl (5)	olo (5)	mg (5)	ky (5)	lij (5)	my (5)	nds-nl (4)		
gd (4)	am (4)	fi (4)	rm (4)	ksh (4)	jr (4)	eo (4)	szl (4)	an (3)	arz (3)	myv (3)	hyw (3)	ary (3)	jae (3)	
uz (3)	pcd (3)	io (3)	it-sam (3)	mt (3)	fy (3)	ts (3)	lmo (3)	aeb-arab (3)	sms (3)	pam (3)	xh (2)	pfl (2)	de-ch (2)	
lfn (2)	mzn (2)	lez (2)	vo (2)	icy (2)	sgs (2)	sa (1)	vmf (1)	nso (1)	kk (1)	bho (1)	ik (1)	aeb-lan (1)	kl (1)	
													ht (1)	
gn (1)	bxr (1)	ban (1)	as (1)	ang (1)	sma (1)	wa (1)	sw (1)	su (1)	sah (1)	pap (1)	sat (1)	stq (1)	mnj (1)	
mhr (1)	ltg (1)	roa-tara (1)	ku (1)											

(a) Properties curated on property class: lighthouse

Count of translated labels

uk (18)	nl (18)	fr (18)	en (18)	ar (18)	tr (11)	ca (11)	es (10)	da (9)	ru (9)	de (9)	sv (9)	en-gb (8)	zh-hans (8)
mk (7)	zh-hant (7)	fi (7)	nb (7)	te (7)	pt (6)	zh-hk (6)	pl (6)	lv (6)	zh (5)	eo (4)	sr (4)	it (4)	zh-tw (4)
en-ca (3)	de-ch (3)	en-it (3)	simple (3)	nds (3)	ko (3)	bar (3)	nn (2)	be-tarask (2)	fa (2)	eu (2)	br (1)	ba (1)	af (1)
									sr-ec (1)	ja (1)	sr-el (1)	be (1)	

Query

[Run Query on Wikidata.](#)

Count of translated descriptions

uk (18)	en (18)	fr (14)	tr (11)	nl (9)	en-gb (8)	fi (7)	de (7)	es (7)	te (7)	nb (6)	da (4)	en-ca (3)	simple (3)
pl (3)	ca (3)	it (3)	ko (3)	zh (2)	pt (2)	zh-hans (2)	ru (2)	sv (2)	lv (2)	nn (1)	br (1)	zh-hant (1)	zh-tw (1)
						zh-hk (1)	ja (1)	eu (1)					

Query

[Run Query on Wikidata.](#)

(b) Properties curated on property class: Wikidata property related to lighthouses

Fig. 16. Statistics of property translations curated on two property classes

statistics in different languages, links to the provenance information (number of references on statements describing the property, number of equivalent properties, usage of references in the statements using the property), translation path, and visualization of the translation path.

4.3 Discussion

Wikidata is highly evolving and new properties are being regularly proposed and translated. The WDProp interface has also undergone several changes during the last couple of years to incorporate these changes. Take, for example, WDProp initially displayed all the properties on the properties page [14], less than 5000 properties at that time. But this number has now increased to around 9,000 properties. So now the interface shows only the first 100 properties obtained from the SPARQL query. However, the user now has the SPARQL query link to obtain the complete list. SPARQL query links and links to Wikidata Mediawiki API calls have now been integrated into all the pages on WDProp. This enables the users to verify, use and even modify these API calls to suit their needs.

Some commonly used properties also face vandalism. The commit history of some initial properties has therefore become long, which means obtaining the complete translation path as shown in Figure 18 takes a significant amount of



781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832

Count of translated labels

it (37)	fr (37)	en (37)	de (37)	ca (37)	ar (37)	uk (37)	tr (37)	sr (37)	nl (37)	mk (36)	zh-hant (36)	ru (36)	pl (36)
hu (35)	es (35)	en-gb (35)	zh-hans (34)	pt (34)	eo (33)	nb (33)	fi (32)	cs (32)	sv (32)	he (31)	da (31)	be-tarask (31)	zh-tw (31)
fa (30)	ms (29)	be (29)	id (29)	gl (28)	el (28)	bn (28)	vi (28)	eu (27)	pt-br (27)	cy (26)	it (26)	lv (25)	ro (25)
nn (23)	sco (22)	zh-hk (22)	vec (21)	ur (20)	et (19)	ba (19)	ast (19)	tg (19)	zh-cn (19)	sl (18)	sk (18)	ka (18)	th (18)
sq (18)	en-ca (17)	lg (17)	fa (17)	oc (17)	lb (17)	hr (17)	os (17)	br (16)	az (16)	zh-sg (16)	zh-mo (16)	bs (16)	it-cyrl (15)
sr-ec (15)	ta (15)	nob (15)	zh-my (15)	ml (15)	gu (15)	sr-el (14)	scrn (14)	mr (14)	it (14)	mn (14)	se-se (13)	ckb (13)	el (13)
or (13)	hi (13)	gaw (12)	yue (11)	ne (11)	la (11)	hy (11)	fr (11)	sh (10)	de-ch (10)	ti (10)	ku-lam (10)	kn (10)	ms-arab (9)
yi (9)	tg-cyrl (9)	ps (9)	min (9)	to (9)	diq (8)	arz (8)	neo (8)	pa (8)	mg (8)	ky (8)	is (8)	ilo (8)	ia (8)
hy (8)	se (7)	li (7)	szl (7)	my (7)	lv (7)	bar (7)	en (6)	am (6)	si (6)	nds-nl (6)	ha (6)	ca (6)	mt (6)
ajp (5)	rm (5)	olo (5)	yo (5)	gd (5)	ie (5)	sme (5)	smn (5)	kk (5)	kw (5)	se (5)	nap (5)	li (4)	sd (4)
pam (4)	pcd (4)	ksh (4)	vo (4)	mas (4)	szc (3)	ary (3)	uz (3)	it-lam (3)	jbo (3)	hyw (3)	ts (3)	ht (3)	mi (3)
lmo (3)	simple (3)	lcy (2)	sgs (2)	na (2)	kab (2)	wa (2)	sw (2)	stq (2)	lfn (2)	mri (2)	sat (2)	mzn (2)	vmf (1)
pfl (1)	bho (1)	tj-lam (1)	su (1)	nso (1)	myv (1)	ain (1)	cds (1)	bxr (1)	ban (1)	mhr (1)	lg (1)	lez (1)	arb-arab (1)
roa-tara (1)	sah (1)	pap (1)	nan (1)	kv (1)	ig (1)	hak (1)	dag (1)	csb (1)	xh (1)	os (1)	fur (1)	ie (1)	kg (1)
li (1)	nrm (1)	pms (1)	sc (1)	vls (1)	wo (1)	zu (1)	frp (1)						

Fig. 17. Statistics of property translations curated by Wikidata WikiProject Museums

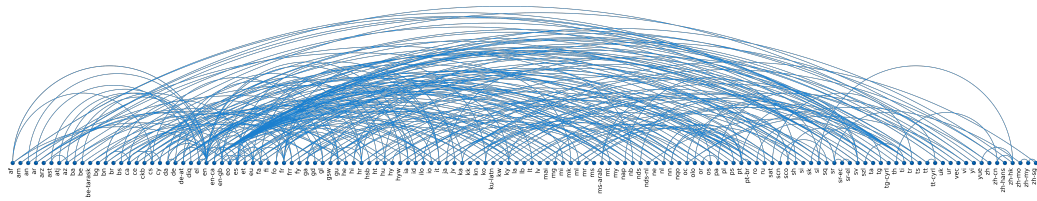


Fig. 18. Translation of Wikidata property labels: official website (P856)

time. Also, the number of property classes and WikiProjects is increasing, which means the current display of these lists may require some modifications in the future, especially the possibility to search and filter the desired information.

### 5 CONCLUSION AND FUTURE WORKS

With the growing usage of mobile devices and internet penetration across the world, web developers and content producers are looking for ways to ensure the availability of information in the local languages of the readers. Considering the number of languages, it is not easy for a few developers to produce multilingual information. Wikipedia and Wikidata have long shown how such multilingual information can be produced collaboratively and in an open manner. However, considering the disparity in the number of contributors in different languages, we find that not all languages are equally represented on these websites. This article focused on Wikidata properties and their translations and discussed the design and development of WDProp for this study. It also demonstrated how the web application can be used to perform granular analysis on Wikidata properties and their curation.

The future course of action includes options to download the data as CSV or JSON files and a multilingual user interface. Another possible direction is to explore how the property translation flow can be used to suggest the untranslated properties to the contributors of a given language. Some aspects of this work can also be integrated into Wikibase installations using SPARQL endpoints and Mediawiki API, but this requires additional experiments.

## ACKNOWLEDGMENTS

The author(s) would like to thank the Wikidata and Wikimedia community members for their feedback during Wikimedia Hackathon 2018, WikidataCon 2019, and ContribuLing 2021 for the feedback, thanks to which the web application has been improved and has helped to provide a written version of this article.

## REFERENCES

- [1] Erik Borra, David Laniado, Esther Weltevrede, Michele Mauri, Giovanni Magni, Tommaso Venturini, Paolo Ciuccarelli, Richard Rogers, and Andreas Kaltenbrunner. 2015. A Platform for Visually Exploring the Development of Wikipedia Articles. In *Proceedings of the Ninth International Conference on Web and Social Media, ICWSM 2015, University of Oxford, Oxford, UK, May 26-29, 2015*. AAAI Press, 711–712. <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/view/10565>
- [2] Thibaut Chamard and John Samuel. 2019. *Multilingual Wikidata Property Translation Flow Dataset*. <https://doi.org/10.5281/zenodo.3271358>
- [3] Lars Christoph Gleim, Rafael Schimassek, Dominik Hüser, Maximilian Peters, Christoph Krämer, Michael Cochez, and Stefan Decker. 2020. SchemaTree: Maximum-Likelihood Property Recommendation for Wikidata. In *The Semantic Web - 17th International Conference, ESWC 2020, Heraklion, Crete, Greece, May 31-June 4, 2020, Proceedings (Lecture Notes in Computer Science, Vol. 12123)*, Andreas Harth, Sabrina Kirrane, Axel-Cyrille Ngonga Ngomo, Heiko Paulheim, Anisa Rula, Anna Lisa Gentile, Peter Haase, and Michael Cochez (Eds.). Springer, 179–195. [https://doi.org/10.1007/978-3-030-49461-2\\_11](https://doi.org/10.1007/978-3-030-49461-2_11)
- [4] Simon Gottschalk and Elena Demidova. 2017. MultiWiki: Interlingual Text Passage Alignment in Wikipedia. *TWEB* 11, 1 (2017), 6:1–6:30. <http://doi.acm.org/10.1145/3004296>
- [5] Lucie-Aimée Kaffee, Kemele M. Endris, and Elena Simperl. 2019. When humans and machines collaborate: cross-lingual label editing in wikidata. In *Proceedings of the 15th International Symposium on Open Collaboration, OpenSym 2019, Skövde, Sweden, August 20-22, 2019*, Björn Lundell, Jonas Gamalielsson, Lorraine Morgan, and Gregorio Robles (Eds.). ACM, 16:1–16:9. <https://doi.org/10.1145/3306446.3340826>
- [6] Lucie-Aimée Kaffee, Alessandro Piscopo, Pavlos Vougiouklis, Elena Simperl, Leslie Carr, and Lydia Pintscher. 2017. A Glimpse into Babel: An Analysis of Multilinguality in Wikidata. In *Proceedings of the 13th International Symposium on Open Collaboration (OpenSym '17)*. Association for Computing Machinery, Galway, Ireland, 1–5. <https://doi.org/10.1145/3125433.3125465>
- [7] Timothy Kanke. 2018. Preliminary Exploration of Knowledge Curation Activities in Wikidata WikiProjects. In *Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries*. ACM, Fort Worth Texas USA, 349–350. <https://doi.org/10.1145/3197026.3203878>
- [8] Timothy Kanke. 2020. Knowledge curation work in Wikidata WikiProject discussions. *Library Hi Tech* 39, 1 (Jan. 2020), 64–79. <https://doi.org/10.1108/LHT-04-2019-0087>
- [9] Dustin Lange, Christoph Böhm, and Felix Naumann. 2010. Extracting structured information from Wikipedia articles to populate infoboxes. In *Proceedings of the 19th ACM international conference on Information and knowledge management - CIKM '10*. ACM Press, Toronto, ON, Canada, 1661. <https://doi.org/10.1145/1871437.1871698>
- [10] Michael Luggen, Julien Audiffren, Djellel Eddine Difallah, and Philippe Cudré-Mauroux. 2021. Wiki2Prop: A Multimodal Approach for Predicting Wikidata Properties from Wikipedia. In *WWW '21: The Web Conference 2021, Virtual Event / Ljubljana, Slovenia, April 19-23, 2021*, Jure Leskovec, Marko Grobelnik, Marc Najork, Jie Tang, and Leila Zia (Eds.). ACM / IW3C2, 2357–2366. <https://doi.org/10.1145/3442381.3450082>
- [11] Claudia Müller-Birn, Benjamin Karran, Janette Lehmann, and Markus Luczak-Rösch. 2015. Peer-production system or collaborative ontology engineering effort: what is Wikidata?. In *Proceedings of the 11th International Symposium on Open Collaboration (OpenSym '15)*. Association for Computing Machinery, San Francisco, California, 1–10. <https://doi.org/10.1145/2788993.2789836>
- [12] John Samuel. 2017. Collaborative Approach to Developing a Multilingual Ontology: A Case Study of Wikidata. In *Metadata and Semantic Research (Communications in Computer and Information Science)*, Emmanouel Garoufallo, Sirje Virkus, Rania Siatra, and Damiana Koutsomiha (Eds.). Springer International Publishing, Cham, 167–172. [https://doi.org/10.1007/978-3-319-70863-8\\_16](https://doi.org/10.1007/978-3-319-70863-8_16)
- [13] John Samuel. 2018. Analyzing and Visualizing Translation Patterns of Wikidata Properties. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction (Lecture Notes in Computer Science)*, Patrice Bellot, Chiraz Trabelsi, Josiane Mothe, Fionn Murtagh, Jian Yun Nie, Laure Soulier, Eric SanJuan, Linda Cappellato, and Nicola Ferro (Eds.). Springer International Publishing, Cham, 128–134. [https://doi.org/10.1007/978-3-319-98932-7\\_12](https://doi.org/10.1007/978-3-319-98932-7_12)
- [14] John Samuel. 2018. Towards Understanding and Improving Multilingual Collaborative Ontology Development in Wikidata. In *Wikidata Workshop 2018*. [http://wikiworkshop.org/2018/papers/wikiworkshop2018\\_paper\\_12.pdf](http://wikiworkshop.org/2018/papers/wikiworkshop2018_paper_12.pdf)
- [15] Moritz Stefaner, Dario Taraborelli, and Giovanni L Ciampaglia. 2011. Notabilia—Visualizing Deletion Discussions on Wikipedia. <http://well-formed-data.net/archives/570/notabilia-visualizing-deletion-discussions-on-wikipedia>
- [16] Thomas Pellissier Tanon and Lucie-Aimée Kaffee. 2018. Property Label Stability in Wikidata: Evolution and Convergence of Schemas in Collaborative Knowledge Bases. In *Companion of the The Web Conference 2018 on The Web Conference 2018, WWW 2018, Lyon , France, April 23-27, 2018*. ACM, 1801–1803. <http://doi.acm.org/10.1145/3184558.3191643>
- [17] Denny Vrandečić. 2012. Wikidata: a new platform for collaborative data collection. In *Proceedings of the 21st international conference companion on World Wide Web - WWW '12 Companion*. ACM Press, Lyon, France, 1063. <https://doi.org/10.1145/2187980.2188242>

- 885 [18] Eva Zangerle, Wolfgang Gassler, Martin Pichl, Stefan Steinhauser, and Günther Specht. 2016. An Empirical Evaluation of Property Recommender  
886 Systems for Wikidata and Collaborative Knowledge Bases. In *Proceedings of the 12th International Symposium on Open Collaboration (OpenSym '16)*.  
887 Association for Computing Machinery, Berlin, Germany, 1–8. <https://doi.org/10.1145/2957792.2957804>

888  
889  
890  
891  
892  
893  
894  
895  
896  
897  
898  
899  
900  
901  
902  
903  
904  
905  
906  
907  
908  
909  
910  
911  
912  
913  
914  
915  
916  
917  
918  
919  
920  
921  
922  
923  
924  
925  
926  
927  
928  
929  
930  
931  
932  
933  
934  
935  
936