An analysis of discussions in collaborative knowledge engineering through the lens of Wikidata

Elisavet Koutsiana[a], Gabriel Maia Rocha Amaral[a], Neal Reeves[a], Albert Meroño-Peñuela[a], Elena Simperl[a]

[a]*King's College London, Informatics Department, Bush House, Strand Campus, 30, Aldwych, London, WC2B 4BG, , UK*

Abstract

We study *discussions* in Wikidata, the world's largest open-source collaborative knowledge graph (KG). This is important because it helps KG community managers understand how discussions are used and inform the design of collaborative practices and support tools. We use descriptive statistics, thematic analysis, and statistical tests to investigate how much discussions in Wikidata are used, what they are used for, and how they support knowledge engineering (KE) activities. The study covers three core sources of discussion, the talk pages that accompany Wikidata items and properties, and a general-purpose communication page. Our findings show low use of discussion capabilities and a power-law distribution similar to other KE projects such as Schema.org. When discussions are used, they are mostly about KE activities, including activities that span across the entire KE lifecycle from conceptualisation and implementation to maintenance and taxonomy building. We hope that the findings will help Wikidata devise improved practices and capabilities to encourage the use of discussions as a tool to collaborate, improve editor engagement, and engineer better KGs.

Keywords: collaborative knowledge engineering, knowledge graph, discussion analysis, Wikidata

1. Introduction

Knowledge engineering (KE) is about processes, methods, languages and tools to create, maintain, and use knowledge bases [51]. *Knowledge graphs* (KGs) are the latest exemplar of knowledge base; they organise data from various domains and sources into graphs for widespread use in downstream (AI) applications [13].

To reach the scale that modern applications demand, KGs today routinely reach millions, if not billions of entities. Engineering them involves a mix of manual and automatic means, and, despite considerable advances, remains challenging for a range of sociotechnical reasons [47]. Some of the most popular KGs are built through online peer-production, that is, with the help of volunteers who self-organise to co-create a knowledge artefact [6]. In the case of Wikidata [60], the world's largest open-source KG, this involves around 24 thousand active editors.[1] The size of the Wikidata community and its use in anything from web search to virtual assistants makes it a primary case study to investigate and support practices of large decentralised KE projects [48].

Prior studies have looked at various sociotechnical concerns in Wikidata [11, 36, 39, 40, 43] by interviewing editors and analysing activity logs; Wikidata discussions are missing from this line of research. This is the starting point for this paper - as shown by a rich body of literature in online peer-production, discussions play a critical role in the success of Wikipedia [58], but also other types of peer-production systems such as question-answering sites (e.g. Quora [38], Stack Overflow [35]) and open-source software development (e.g. GitHub [18], Apache server [32]). In the case of Wikipedia, which shares a lot of common ground with Wikidata, the subject of our paper, Viegas et al. [58] reported that discussions about articles serve multiple essential functions, including strategic planning of editing activities and the enforcement of Wikipedia policies and conduct guidelines.

We aim to provide similar insights for Wikidata by studying discussions across several channels - the *talk pages* that accompany every *item* and *property*, the main building blocks of the KG, as well as a general-purpose communication page, the project chat. Through a mixed-methods approach with descriptive statistics, thematic analysis, and statistical tests, we get a sense of how much discussions are used ($RQ1$), what they are used for ($RQ2$), and how they support KE activities ($RQ3$). $RQ1$ is motivated by the fact that, despite the existence of the aforementioned studies, it remains unclear how much discussion pages are being used in Wikidata, and how such usage compares in similar peer-production systems. Moreover, the fact that Wikidata is a collaborative, structured KG may be triggering very specific types of conversations relating to KE –on e.g. how to organise knowledge, what is the class type for a certain item and its hierarchy, etc.– that are generally harder to find in other peer-production platforms. With $RQ2$ we aim at, therefore, understanding the content of these discussions and discerning which of them relate to KE, and which relate to other, general peer-production subjects that can also be found elsewhere –on e.g. authenticity of sources, coordination and organisation of work, etc. Finally, with $RQ3$ we zoom into the former group of conversations, those about KE activities,

[1] `https://www.wikidata.org/wiki/Wikidata:Statistics`

and map their correspondence to processes commonly found in well-established KE methodologies.

Our contributions are two-fold: (i) an account of the use and content of discussions across Wikidata; and (ii) a coding scheme to apply thematic analysis to Wikidata discussions with a focus on KE. Our study shows that:

- Editors do not use talk pages a lot: only 0.02% of items have them.
- Discussions follow a power-law distribution, with an overwhelming majority having one editor posting an issue without response (50% for items, 8% for properties, and 16% for project chat) and only a small portion present longer discussions (3% for items, 9% for properties, and 31% for project chat).
- Discussions in Wikidata are rarely conflictual.
- The main topic of discussion revolves around KE activities, including KE processes and actions, and taxonomy building.
- Editors use discussions to analyse KE issues with main KE activities in classical and collaborative KE, conceptualisation, implementation, maintenance, role specialisation, version control, project flexibility, and tool support.

Our study is the first attempt to understand KE practices in Wikidata through the lens of discussions, a key tool in any online peer-production system. We discuss the implications of the findings for Wikidata practices and collaboration support tools. Similar to other systems in knowledge and software engineering (e.g. Schema.org [22], GitHub [18]), we believe more has to be done to encourage editors to use discussions and improve the instructions available, especially for newcomers. The large share of talk pages consisting of a single, unattended post could impact participation, again, especially for newcomers [18, 50], and the quality of the KG [7]. Our analysis has shown that editors discuss various KE concerns; this could potentially lead to inconsistencies, especially as we found little evidence of planning via discussions, which is quite common in discussions in Wikipedia [58]. Overall, we believe that our study confirms how important discussions are as a source of insight into the doings of a large KE community [15], as we find that a large portion of item and property talk pages are empty or contain unanswered posts; and that KE is a prominent discussion topic, albeit with no systematic guidance. We use these to suggest design improvements and topics for follow-up studies in KG quality and editor engagement.

2. Background: Collaborative knowledge engineering and Wikidata

KE is the branch of artificial intelligence concerned with engineering and constructing knowledge-based systems. Collaborative knowledge engineering has a special place in KE because of several reasons: for instance, as knowledge bases are costly to build, researchers have investigated ways to distribute and decentralise efforts; in addition, in ontology engineering, ontologies are meant to capture shared, agreed understandings of domains of interest, which are achieved through collaboration across stakeholders [48]. There is a rich body of literature in collaborative KE, ranging from methodologies [25, 26, 33, 52, 61] to semi-automatic methods [1, 59] and technologies [1, 44, 53].

Wikidata is an open-source KG built through online peer production [60]. It is part of the Wikimedia Foundation[2] and anyone can contribute to it via a browser interface.The underlying system is a wiki with a collection of web *pages* connected via hyperlinks. Content-wise, the graph consists of two main types of pages, *items* and *properties*, and there are labels, descriptions, and aliases in natural language to allow people to understand what these represent. Items represent entities, e.g., `Marie Curie` or `London`, and classes of entities, e.g., `human` or `capital`, and properties represent the relations between entities. Items and properties form *statements*, which may be accompanied by *qualifiers* to specify further the statement, e.g., the statement `London - population - 8,908,081` can be specified with the qualifier: `point in time - 2018`.

Wikidata is built by a mix of people and *bots*, software that executes simple, repetitive edits. Editors can have different editing rights depending on their experience and contributions. In general, any editor can work on items and properties, but some have higher access rights, specific *roles*,[3]. They can be *Administrators*, trusted editors responsible for tasks like deleting and blocking, *Bureaucrats*, editors with technical experience able to change other editors' rights, and *Property creators*, editors with the right to create a new property. The higher access rights are granted to ensure high-quality results. We can find similarities in editors with and without right with classical KE roles. In classical KE the team developing the ontology consists of: knowledge engineers, ontology engineers, and domain experts [48]. Usually, we can distinguish between ontology editors and ontology contributors, with the first controlling the ontology and the latter proposing changes [48]. However, roles in collaborative KE depend on the model. Particularly, in Wikidata we can consider the editors with higher rights as ontology editors, as they can give permission for actions such as creating a bot, blocking an editor, and protecting entities. In addition, we can consider editors without specific rights as ontology contributors.

Independently of the specific focus of a KE project, an essential aspect is collaboration; even more so when contributors are self-organised volunteers who coordinate and interact remotely. Discussions are a primary tool for this, as identified in e.g. collaborative ontology engineering [48]. Wikidata has two ways to support them: (i) talk pages, and (ii) communication pages. As mentioned earlier, every content page can have an accompany talk page that anyone can use to ask questions, flag

[2] `https://foundation.wikimedia.org/wiki/Home`

[3] `https://www.wikidata.org/wiki/Wikidata:User_access_levels`

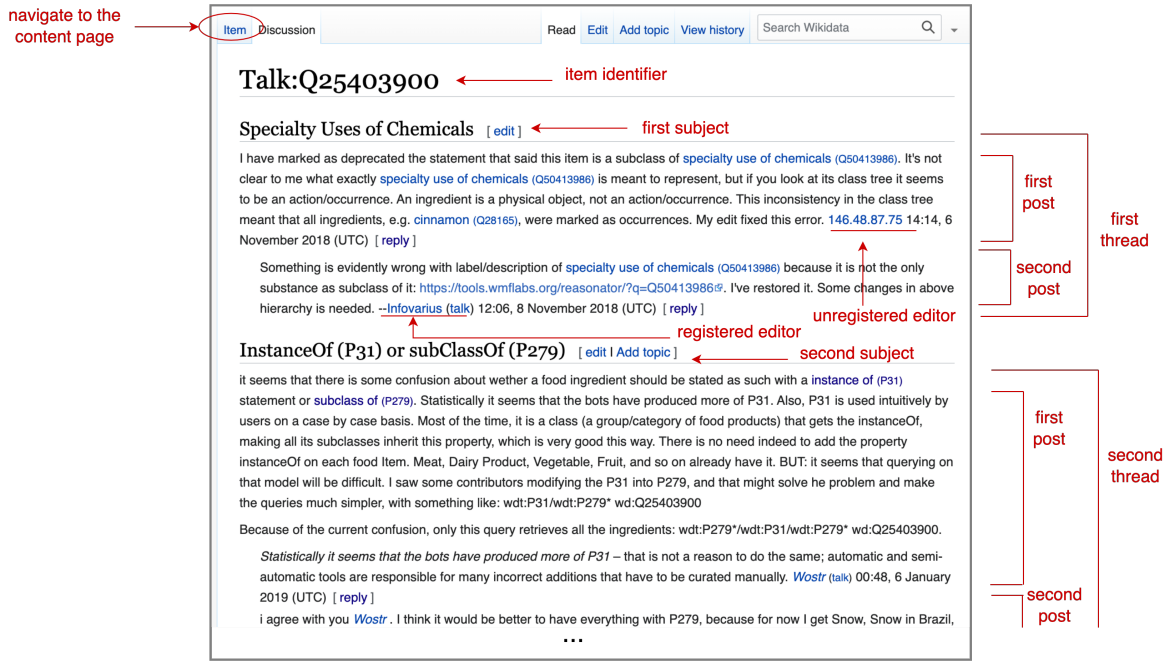

Figure 1: An example of a Wikidata talk page including two threads. We use annotations to highlight their main features.

mistakes, report actions, etc. In addition to talk pages, Wikidata also has many other *communication pages*, like project chat, used to discuss project-wide issues in a single space, which contains all discussion.[4] Table 1 presents a list of the Wikidata communication pages and the different issues they address. Both talk communication pages follow the same structure. Figure 1 shows an example of a talk page. The first line states the item identifier. The page includes *posts* organised in *threads* under specific *subjects*. Each page may include one or more discussions, which we refer to as *threads*. A thread is a set of *posts* under a *subject* title. A subject title is essential to separate the different threads, as well as editors signatures to separate the different posts.

Table 1: A list with the communication pages in Wikidata and their descriptions.

Communication page	**Description**
General discussion	
Project chat	General discussion about the project
Requests for comments	Discussion requests for specific topics
Report a technical problem	Discussion about platform complication
Requests	
Request a query	Requests for Wikidata SPARQL queries
Interwiki conflicts	Report problems with content on other wikis
Bot requests	Requests for tasks to be done by a bot
Wikidata: Property proposal	Propose the creation of a property
Administrators' noticeboard	Reporting vandalism, requesting page protections, etc.
Translators' noticeboard	Report a translation problem, ask to mark a page for translation
Bureaucrats' noticeboard	Requesting for flood flag, etc.
Requests for deletions	Deletion requests of items and pages
Properties for deletion	Deletion requests of properties
Wikidata: Requests for permissions	Permissions requests for trusted users in the community

[4] https://www.wikidata.org/wiki/Wikidata:Project_chat

3. Related work

3.1. Wikidata studies

Wikidata is a major knowledge graph providing semantically structured data. From the early stages, the success of the project motivated Google to offer the data of the Freebase knowledge graph to the Wikidata community [37]. Today, Wikidata supports many Wikimedia projects like Wikipedia, as well as many artificial intelligence tasks like question answering, recommendation systems, and information retrieval [13].

Previous studies of the Wikidata community have focused on editor engagement [39, 40, 43] and participation profiles [11, 36]. Using semi-structured interviews, Piscopo et al. [39] found that "Wikidatians" acquire a higher sense of responsibility and a deeper community engagement as they get more involved in the project, following loosely the reader-to-leader framework observed in other online communities [41]. As their involvement progresses, editors tend to do more work on the Wikidata schema instead of curating items [40]. In the same vein, Sarasua et al. [43] analysed Wikidata activity to predict editors' lifespan using machine learning. Furthermore, Muller et al. [36] identified editor roles by clustering activity logs. They found that most editors made specialised contributions, and only a small cluster of active editors contributed more widely. They matched the clusters to roles commonly found in online peer-production systems, but also in collaborative ontology engineering projects [36]. The authors also studied participation patterns in the roles, discovering that most editors tend to keep the same role over time, and there are different participation patterns for the "semantic aware users (contributors)" and contributors who engage irregularly [11].

Kanke et al. [21] used content analysis, a method to quantify and analyse certain words, themes, and concepts within

qualitative data e.g., interviews [19], to study discussions in WikiProjects.[5] These are sub-groups in Wikidata working on specific topics like biology or sports. The authors aimed to study how editors participate and work in those projects by identifying the group's main activities, norms and rules, and tools. Nonetheless, our study analyses the main discussion areas in Wikidata and aims to find how editors use discussions as a tool for collaboration, what they discuss, and particularly, whether and how discussions support KE activities. These will help understand collaboration within Wikidata community as a whole, not only the small sub-groups that work closely on specific cases, and will identify the role of discussions in the project.

3.2. Discussions in online peer-production systems

There is a large body of work in CSCW (Computer-Supported Cooperative Work), social computing and computational social sciences that investigates online peer production systems and their underlying communities [3, 15, 18, 21, 22, 31, 34]. Prior studies have shown that discussions help with community building, agenda setting, and productivity [18, 42, 45, 58]; at the same time, analysing them can give insight into community behaviour and volunteer engagement [3, 18]. In discussions, one can observe how decisions are collectively made, spot possible dysfunctions and conflicts, and think about ways to improve guidance, practices, and collaboration tools.

Various projects have been typically used as prime examples of successful peer-production communities in which peer discussion is central, such as free and open source software, Wikipedia, Slashdot, OpenStreetMap and Schema.org [5]. However, in open source projects discussions are much more constrained and centered around project development and developer engagement [18], and address specific technical issues (bugs, solutions to problems, testing) [4] that often give raise to disagreements and toxic comments [31]. In other projects such as OpenStreetMap discussions are much more open and unstructured, but are sent through a combination of various discussion channels such as mailing lists, issue trackers, and real-life events, making their analysis much more complex [22, 46]. Wikipedia, on the other hand, allows for unconstrained discussions via a centralised communication medium through discussion pages associated to each specific article, generating a large volume of discussions [45]. Previous analyses of the content of discussions of Wikipedia found that these typically revolve around coordination, references to guidelines, infoboxes, edits, and polls [45, 58], as well as conflict resolution [57]; and are primarily used to support strategic planning of edits, project coordination, and enforcement of guidelines [58]. These analyses typically resorted to quantitative and qualitative methods, most notably manual coding of samples and content analysis [4, 45, 58].

More closely to Wikidata, discussions in collaborative KE projects have also been analysed, primarily in a Semantic Web context. For example, in the collaborative ontology-engineering tool WebProtégé [56] the study of three knowledge production projects showed that editing activities and discussion participation are positively correlated [15], signaling that discussions go hand in hand with contributions. In Schema.org, a collaboratively-built ontology for structured data on the Internet, the topic of discussions was linked to the medium where they happen (GitHub for functionality, mailing lists for clarifications) [22]. Furthermore, collaborative ontology engineering methodologies like DILIGENT [61] and HCOME [25] emphasised the crucial role of discussions in collaboration. Using argumentation frameworks [24, 55] support argumentation during every step of the ontology development. Discussion analysis showed that discussions serve as the means of consensus [55] as well as an archive for the new members of the community to understand how and why decisions were made and for the existing members to remember the rational behind them [54].

Our paper contributes to this field a comprehensive study of discussions in the Wikidata online peer-production system. This system produces a KG, and hence our analysis aims to understand the link between discussions and KE practices in a large, self-organised community of volunteers who are not necessarily KE experts.

4. Data

4.1. Choice of discussion channels

As noted earlier, to get a sense of how people talk in Wikidata, we analysed the talk pages of items and properties, and the project chat communication channel. We abbreviate talk pages as `itemTP` for items and `propertyTP` for properties, and project chat as `PC`. In the following, we refer to the three sources of discussions as *discussion channels*. The choice of discussion channels mirrors prior Wikipedia study-designs, which focused mostly on article talk pages as the main place where discussions happen [17, 45, 58]. We added the project chat to acknowledge Wikidata practices, where, unlike in Wikipedia, volunteers are advised to use the project chat to discuss cross-cutting concerns and ask questions. We did not include in our current analysis any communication pages dealing with specific requests (see Table 1). While some of these discussion channels are concerned with KE activities (e.g. for requests for deletions, for SPARQL queries, for properties), most of them have to do with the operation of the underlying wiki system and with community management across specific roles.

We used publicly available data from the Wikimedia Foundation.[6] The data corpus contains content from the beginning of Wikidata in October 2012 to June 2021.

4.2. Exploratory data analysis

Our data corpus of the three discussion channels consists of 31,677 talk pages for items and properties, including a variety of number of threads, and one page including 14,404 threads for `PC`. We use the term *documents* when we refer to talk pages

[5] `https://www.wikidata.org/wiki/Wikidata:WikiProjects`

[6] `https://dumps.wikimedia.org/wikidatawiki/`

Figure 2: Example of Wikidata talk pages without threads including (i.e., *G*0).

and project chat threads as data corpus. Preliminary results on the talk pages revealed significant inconsistency in pages' content, including either a highly unbalanced number of threads, or do not include threads at all. This led us to split our data corpus into groups based on the number of threads and their length (i.e., the number of posts per thread). Table 2 presents statistics for the identified groups. We found a group of talk pages which do not contain actual discussions (i.e., *G*0), but rather meta-data, e.g. constraints or categories where an item or property could be used (see Figure 2). Furthermore, for the remaining data corpus which includes discussions, we identified documents including a single thread with one post not receiving any responses (i.e., *G*1), up to 5 posts forming a short conversation (i.e., *G*2), or more than 5 forming a long conversation (i.e., *G*3). In addition, we found documents with more than one thread including short (i.e., *G*4) or long (i.e., *G*5) conversations. The `PC` is split only into groups *G*1, *G*2, and *G*3 as we analyse threads not pages with threads. This preliminary analysis helped us to understand the peculiarity of the data and how to analyse them. We used the groups to verify we have a representative sample for the thematic analysis.

For descriptive statistics we used our complete corpus, while for thematic analysis we used a sample set (see Table 3). To calculate the size of the sample sets for each discussion channel, we applied Cochran's formula [20] (confidence level 90%, a margin of error 10%), leading to approximately 70 documents for each discussion channel. We built a stratified random sample [2] based on the identified groups *G*1 to *G*5. This means we sampled an equal number of documents from each group, summing to approximately 70 documents for each discussion channel. The size of our samples, 210 documents, is greater than similar qualitative studies for Wikipedia. Hara et al. [17] sampled 30 talk pages from three Wikipedias, summing to 120 talk pages to conduct a cross-cultural analysis, while Viegas et al [58] analysed 25 talk pages to understand coordination and organisation.

Our data and code are publicly available through GitHub.[7]

5. Methodology

5.1. Overview

We followed a mixed-methods approach for analysis and used quantitative means to investigate the use of discussions across the different discussion channels and qualitative means to explore their content. Table 3 summarises the methods and data we used to answer our research questions. For *RQ*1 we applied descriptive statistics, and for *RQ*2 and *RQ*3 thematic analysis and statistical tests.

For descriptive statistical analysis, we counted the number of posts, subjects, and words of `itemTP` and `propertyTP`

[7] `https://github.com/ElisavetK/Collaborative_knowledge_engineering_Wikidata`

Table 2: Group separation for the three discussion channels based on the thread length and number of subjects.

Group ID	itemTP %	proeprtyTP %	PC %
G0 - zero talk pages	44	72	-
G1 - single-subject post with no answers	50	8	16
G2 - single-subject short conversation	2	6	53
G3 - single-subject long conversation	1	1	31
G4 - multiple-subjects short conversations	2	6	-
G5 - multiple-subjects long conversations	1	7	-

Table 3: Summary of research methods and data

Research question	Method	Data			
		Data corpus	itemTP	propertyTP	PC threads
RQ1	Descriptive statistics	complete set	23,271	8,406	14,405
RQ2 & RQ3	Thematic analysis	sample set	70	70	69

and the number of posts and words of PC threads. We explain our methodology in detail in the following section for thematic analysis.

5.2. Thematic analysis

Thematic analysis [30] is a standard qualitative method for text analysis in which codes are used to annotate text as summary markers for further analysis and comparison. We defined the codes starting from a coding scheme used in related studies of Wikipedia discussions [58, 17].

We built a coding scheme of 42 codes for Wikidata (for the complete list of codes, see supplementary material Table 1). As part of *RQ*2 on what are discussions used for, we wanted to understand which of these discussions are used for general peer-production system topics that could be found elsewhere – e.g. organising work, authenticity of sources, etc.–, and which are used for KE issues in particular. Consequently, we designed our coding scheme using Wikipedia's discussion page thematic analysis as a basis [17], and we incorporated codes representing traditional KE activities [48]. With this scheme, we aim to investigate whether and how often editors discuss KE (e.g., describe, add, delete, merge, deprecate items, taxonomy building), and how often discussions are about content, rules and policies, connection with other Wikimedia projects, and conflicts.

Table 4 describes the themes we used in our analysis. Our coding scheme includes six themes, or coherent groups of codes: (i) KE activity; (ii) fact accuracy; (iii) regulation or instruction; (iv) connection with a Wikimedia project; (v) conflict or disagreement; and (vi) housekeeping. For each theme, we define specific codes to tag discussion text about questions, informative text (e.g., explanations and answers to questions), and actions (e.g., suggestions and requests) about these six themes.

Two of the authors of this paper conducted the thematic analysis independently, using an iterative coding process. In all documents (talk pages for items and properties or threads for project chat) the authors read the threads and assigned the codes to the post level. Each post can have one or more codes based on the size and content. We conducted this process in two stages. First, the annotators used an initial coding scheme for a testing dataset of 60 documents (20 documents for each discussion channel). The initial coding scheme included a much higher number of codes (110) attempting to identify more details about personal opinion, social expression, disagreements, hierarchy (e.g., identify discussion specific properties like `instance of`, `subclass of` and `type of`), and qualifiers. However, it was very difficult for the annotators to agree due to the large number of codes and the ambiguity of the posts (free text tends to present different meanings for readers). Next, the annotators used the adjusted coding scheme for the set of 210 documents.

We used the non-parametric Kruskal-Wallis (K-W) statistical test [27] to calculate the statistical difference between the coding results of the three discussion channels. To identify the pairs with a statistical difference when the K-W p-value is less than 0.05, we used the post-hoc Dunn's test (using the Bonferroni correction for the p-values to control the family-wise error rate) [12].

6. Results

6.1. RQ1: How much are discussions used?

We found a low number of talk pages compared to the number of items and properties in Wikidata. At the time of the analysis, Wikidata has 95, 370, 190 items and 8, 939 properties, but only 23, 271 (0.02%) `itemTP` and 8, 406 (94%) `propertyTP`. While the percentage for properties may seem high, among these pages 72% do not contain any posts at all, leading to 28% of properties for which there is at least one post. For items, the number of talk pages `itemTP` is tiny to start with, and only 69% have at least one post.

Counting the number of subjects in talk pages, we identified a high inconsistency in subject use when starting a thread

Table 4: Themes used for the thematic analysis and their description.

Theme	Description
KE activity	Any issue related to the management of the KG. The discussions are about changes related to KE processes/actions (i.e., the characteristics of items and properties, or specific actions like merge, add, delete, and deprecate), and the taxonomy building (i.e., discussions about the hierarchy related to the properties `instance of`, `subclass of` and `type of`.
Fact accuracy	Conversations about the content of item or property and the correctness of its descriptions.
Regulation or instruction	The instructions of Wikidata platform and regulations of how the community have agreed to manage editing activities.
Connection with Wikimedia project	Discussions analysing how to connect items with other Wikimedia projects like Wikipedia and Wikisource.
Conflict or disagreement	Simple disagreements or more serious threats taking place between editors.
Housekeeping	General topics like how often users refer to discussions in other channels of Wikidata or for discussions that are not in English language. For PC this theme includes issues we cannot find in the other two discussion categories, like the creation of a template, the use of a bot or an upcoming event, and the weekly summaries (i.e. announcements about news, events, tools, etc.).

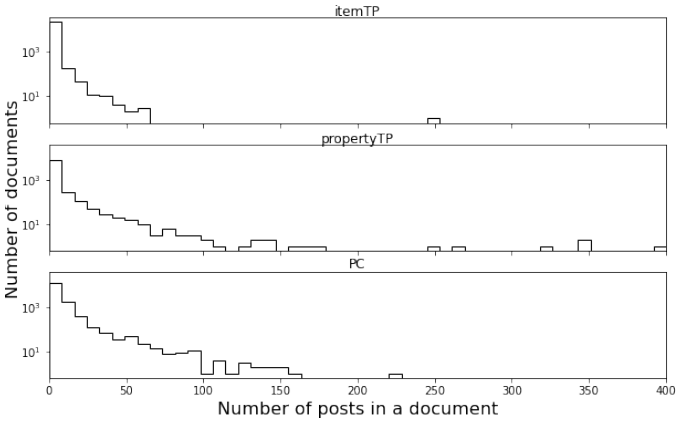

Figure 3: Histograms presenting the number of posts and the number of documents for the three discussion channels `itemTP`, `propertyTP`, and `PC`.

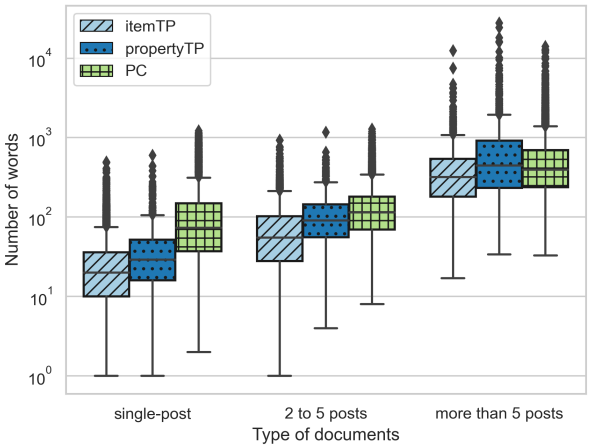

Figure 5: Box plots with the number of words for the three types of documents *single-post,* 2 *to* 5 *posts*, and *more than* 5 *posts*. The different colors represent the discussions channels `itemTP`, `propertyTP`, and `PC`

for the `itemTP` (29%). This is very different from what we find in `propertyTP` (1%). The difference between the two corpora may be due to the inconsistent instructions for creating a new talk page to start a discussion and editing a previously created talk page. Existing talk pages ask separately with a text box for a subject title, while a new one only asks for a post. Properties show a low number of pages without subject due to the description table in their talk pages. This may need an adjustment to the instructions to present consistency.

We showed that the three discussion channels present differences in the post distribution. Figure 3 presents histograms with the number of posts and the number of documents. The exploratory data analysis in Section 4.2 showed an unbalanced distribution of posts for the three discussion channels. Based on the identified groups but without considering the different subjects, we classified our data corpus into three classes. The *single-post* class includes documents from group $G1$, 2 *to* 5 *posts* class from $G2$ and $G4$ with short conversations, and *more than* 5 *posts* class from $G3$ and $G5$ with long conversations. Figure 4 shows the unbalanced size of these classes for the discussion channels. For `itemTP`, a high number of documents, 50%, are single-post, and a lower number of documents contain more than five exchanged posts, 2%. This is probably due to the graph's size, which makes it impossible to follow the discussions in millions of items. This gap may need further

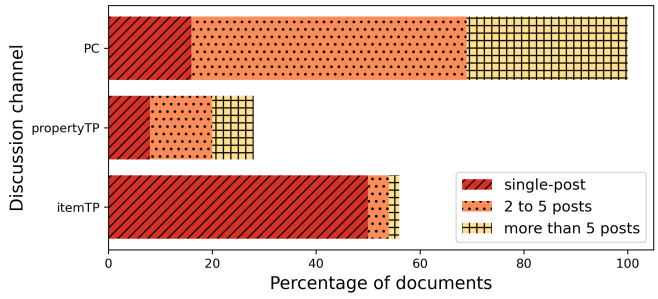

Figure 4: Stacked bars presenting the percentage of documents for the identified classes.

investigation to find a way to track the discussions without response, answer the possible questions, and resolve issues. We found similar distributions despite the lower number of single-post documents in the other two discussion channels. It is worth mentioning that we found 16% (2,296 documents) single-post in `PC` even though the project encourages editors to use it.

Figure 5 presents box plots with the number of words for the classes of documents. We found that `itemTP` show the shortest length in discussions between the three classes and `PC` the longest except for the class *more than* 5 *posts* where `propertyTP` presents the longest length in discussions. This indicates that editors tend to describe in shorter posts in talk pages than in the communication page. Particularly, in the single-post class, where we have only one post, in `itemTP` the posts are closer to 10 words, while in `PC` closer to 100 words. Possible explanations can be either that in talk pages editors have smaller issues which do not need long explanations, or that in communication pages, they are more descriptive to be perceivable in the community.

6.2. *RQ2: What are conversations used for?*

Table 5 presents the thematic analysis codes separated based on the themes in the coding scheme, with the percentage of the codes (mean of the two annotators) assigned to every discussion channel. For every discussion channel, we noted different characteristics and patterns of themes.

`itemTP`. In items, editors discuss all the themes, with *KE activity* issues presenting the greatest (46%) and *regulation or instruction* the least (4%) frequency. *Fact accuracy* was the second most discussed topic (19%).

`propertyTP`. In properties, we observed substantial differences between the discussed themes. The discussions are mainly focused on the *KE activity* (73%). By contrast, the rest themes present very low frequency.

`PC`. In project chat, the dominant theme is *KE activity* (56%).

Table 5: The percentage of codes used for every discussion category in the different themes. The sixth column presents the results of the Kruskal-Wallis(K-W) test between the three discussion categories for the seven themes.

Theme	Code	itemTP %	propertyTP %	PC %	K-W p-value
KE activity					
KE process/action	Question	8	16	11	0.9
	Explanation	11	31	27	
	Suggest (curation, merge, add, delete, deprecate)	7	16	9	
	Request (curation, merge, add, delete, deprecate)	6	9	2	
Taxonomy building	Question	3	0.1	2	0.05
	Sharing information	6	1	4	
	Suggest	2	0	1	
	Request	3	0.3	0	
Total		46	73	56	
Fact accuracy	Question	3	1	0.3	0.3
	Sharing information	15	8	6	
	Suggest	0.4	0.1	0	
	Request	0.6	0	0	
Total		19	9	6	
Regulation or instruction	Question	0.3	1	2	0.6
	Sharing information	3	3	7	
	Suggest	0.1	0.3	1	
	Request	0	0	0	
Total		4	4	10	
Connections with Wikimedia project	Question	2	0.3	2	0.03
	Sharing information	5	1	3	
	Suggest	6	0	0.3	
	Request	4	0.1	0.3	
Total		17	1	6	

However, between the remaining themes we found similar frequency.

The most frequently discussed theme in these three discussion channels is *KE activity*, suggesting that the main aim of the community is the KG maintenance and growth. We compared the significant statistical differences between the three channels for the themes and determined that *taxonomy building* and *connection with Wikimedia project* show significant differences. The Dunn's test adjusted showed that `itemTP` and `propertyTP` are the channels with significant differences. This means that hierarchical conversations about properties, like `subclass of`, and `instance of`, or the connection with Wikipedia for example, are discussed mainly in the `itemTP`. To extend our analysis to the whole data corpus we investigated the frequency of hierarchical terms in `itemTP` and found that 3% of the posts include the term `subclass of` and 2% `instance of` while in `propertyTP` we found 1% and 2% respectively. This can be one more indication that discussion about taxonomy and particularly classes take place mostly in `itemTP`.

We could find no significant statistical difference between the three discussion channels for the *conflict or disagreement* theme. The results show disagreement in all discussion channels with a very low level of intensified conflict (7-9% disagreement and up to 1% intensified conflict). For `itemTP`, this could be because conversations are typically short, while for `propertyTP`, the number of documents with talks may be very low to identify controversies. In addition, for `PC` the community may be able to control any conflicts before they escalate.

6.2.1. Inter-annotator agreement

We used Cohen's Kappa measure of agreement [9], with a value of 0.64 for 210 documents, during coding to assess inter-annotator agreement. This value is interpreted as substantial [28], or fair to good [16] agreement. The themes *KE activity* and *fact accuracy* presented high inter-annotator agreement scores (see supplementary material Table 2). In contrast, the

Table 6: Examples of posts from the three discussion channels.

ID	KE activity	Discussion content
a	Conceptualisation	This item used to include all female organisms, but now it excludes humans. This doesn't make much sense, because item "female" is a subclass. Either it shouldn't be a subclass, or this item should include humans.
b	Conceptualisation	I propose to convert all relationships to gender-neutral and age-neutral relationships, because age and gender are already stored in the person items, so storing it twice is unnecessarily, redundant and error-prone
c	Implementation	It seems that there is some confusion about whether a food ingredient should be stated as such with a property "instance of" statement or "subclass of". Statistically it seems that the bots have produced more of "instance of". Also, "instance of" is used intuitively by users on a case by case basis. Most of the time, it is a class (a group/category of food products) that gets the "instance of", making all its sub-classes inherit this property, which is very good this way. There is no need indeed to add the property "instance of" on each food Item. Meat, Dairy Product, Vegetable, Fruit, and so on already have it. BUT: it seems that querying on that model will be difficult. I saw some contributors modifying the "instance of" into "subclass of", and that might solve the problem and make the queries much simpler, with something like: *state a query*
d	Implementation	Please do not use URLs like "`//www.freebase.com`". Use "`https://www.freebase.com`" instead.
e	Maintenance	I notice some items like "film award" and "C.L. de Carvalho-Heineken Prize for Cognitive Science" use property "field of work" and have a constraint violation because "field of work" is only applicable to people and organizations. Is there a correct way to express that relationship?
f	Maintenance	I'm planning to write a RFC(request for a comment) for the adoption of Wikidata:Living_people (page with policies regarding how to edit in Wikidata living people). Does anybody here think that the policy needs changes before it goes into the RFC?
g	Key roles	Please merge items "Olkelda" and "Olkelda". One has the Dutch Wikipedia's article about a place in Iceland, together with a Commons-hosted picture of the place, while the other has the English and Icelandic Wikipedias' articles about the same place. Aside from the shared name, the English and Dutch articles even use the same image.
h	Key roles	Dear administrators! Please consider to change the Formatter URL from "`http://dispatch.opac.d-nb.de/DB=1.2/CMD?ACT=SRCHA&IKT=8529&TRM=$1`" to "`https://w3id.org/isil/$1`" with a value "(?:DE—ZDB—US—IT—JP)-.*" for the "format as a regular expression" property. Since I'm not allowed to change this.
i	Version control	I found more occurances of this problem related to food, this user, reverse property "subclass of", and this source: item "CAC/GL 36-1989: Class names and the international numbering system for food additives, 2017 English revision". The user apparently read this document a bit superficially, and entered a bunch of wrong "subclass of" relations. You can check all my edits and you will understand (i have much less edits than him, since i started only yesterday to be interested in Wikidata).
j	Project flexibility	Currently this item has item "Bundesgesetzblatt", property "subclass of", item "periodical". Given this is an actual "periodical" and not a "subclass of" "periodical", this relation is wrong according to Help:Basic_membership_properties AFAICT
k	Project flexibility	If you add a property "official name", it would be helpful if you included a reference as well.
l	Project flexibility	There are two main ways of giving these identifiers. -'institution/database/html/<ID>', e.g. 'raa/bbr/html/21310000000827' or -'institution/database/<ID>' e.g. 'raa/bbr/21310000000827' The first is a redirect to the human readable webbpage (`https://www.bebyggelseregistret.raa.se/bbr2/anlaggning/visaHistorik.raa?page=historik&visaHistorik=true&anlaggningId=21310000000827` for the first example). The second gets you to the official entity in the database (rdf). I would definitely recommend using the latter for wikidata where the linked data interface makes more sense.
m	Tool support	If you want to continue the discussion, I think the best place is on the "Project chat" or the property "software version identifier" in order to keep track easily and to collect more opinion.
n	Tool support	There is now also some discussion on Wikidata:Project_chat#Are_there_rules_for_the_main_label.
o	Tool support	As per Dog, I suggest we should have this item as property "instance of", property "common name of", item "felis catus", see `https://www.wikidata.org/w/index.php?title=Wikidata_talk:WikiProject_Taxonomy/Archive/2017/03#Cats`.
p	Tool support	Until we have a definitive solution, I'll resort to using the wmflabs tool "wikidata-externalid-url", as suggested in that thread.

lowest scores were observed for the themes *Regulation or instruction* and *Conflict or disagreement*. The low scores were due to ambiguous posts (subjective meaning of natural language text). One example of ambiguation for the *regulation or instruction* theme is *"As I understand it, a number of property values assigned to an item are assumed to apply also to its daughter items, unless another explicit statement replaces that value."* where one annotator assigned this as "sharing information about regulation or instruction", but the other considered it as "sharing information about taxonomy building".

6.3. RQ3: How do discussions support KE activities?

Having established in Section 6.2 that a majority of discussions are about KE activities around the Wikidata KG, we analyse here in depth what exactly these KE activities are about and how they relate to well-known KE workflows. To do this, we identify core activities in classic [10, 23] and collaborative [26, 48] KE methodologies, and we map specific examples we found in the thematic analysis to these activities (Table 6). We first look in classical ontology engineering activities where iterative processes develop the ontology [23] and then we search for activities in collaborative KE [48]. As a result of this, we classify discussion examples within the classic KE activities of **conceptualisation**, **implementation** and **maintenance**; and to the collaborative KE activities of **role specialisation**, **version control**, **project flexibility**, and **tool support**. We describe these activities below in relation to the examples in Table 6.

Conceptualisation in classical ontology engineering refers to the "conceptualisation of the model, integration and extension of existing solutions" [49]. In Wikidata, conceptualisation can be found when editors discuss domain terms for the creation of classes and their hierarchy. Editors often wonder how Wikidata could better express the relations on a specialised domain, like "female organisms" in the examples (a) Table 6. Furthermore, another example is the discussion about whether a new classification can be applied to simplify entities' relations, like the suggestion of gender-neutral and age-neutral relations in example (b).

Implementation in Wikidata is less straightforward than in classical ontology engineering. While in classic KE implementation relates to the translation of the outcomes of the conceptualisation phase into a formal knowledge representation language (e.g. RDF(S) [63] or OWL [62]), in Wikidata implementation is done directly by contributors by editing the KG using the interface or gaming tools. Editors can create new instances or define classes by creating new items. They can link

the instances and define the hierarchy by adding new relations, e.g., properties, to items. The theme *KE activity* show that editors use the discussions to ask for help about their contributions, explain how to carry out, and suggest or report the addition, deletion, deprecation and merge of entities. In the example (c) Table 6 editors discuss whether a certain category of items should be linked with the property `instance of` or `subclass of` and how this influence the outcome of a SPARQL query. Furthermore, in the case of linking properties, we found discussions about their domain and range. An example about the range of the property "formatter URL" is (d) in Table 6.

Maintenance in classical ontology engineering is the " adaptation of the ontology according to new requirements" [49]. In Wikidata, maintenance can be thought of as the most common ontology development activity in the analysed discussions. We can see it in corrections related to property refinement, and adjusting rules. In example (e) Table 6 editors discuss about the property constrains, i.e., rules that specify how to use properties. The Wikidata model does not prevent the use of any entity in any way, e.g. indicate that the type for `horse` is `insect`; however, constraints can describe the proper use and reasoning for properties. Often, the increasing use of a particular property by various groups of items may trigger the violation of such constraints, which is typically followed by discussion and review. Another form of maintenance is changing the rules and regulations in Wikidata. An example is (f) Table 6, where editors start a discussion to change the policy about how to edit items about living people in Wikidata.

Subsequently, we map the coding and examples of the thematic analysis with four collaborative KE features: role specialisation, version control, project flexibility, and tool support. As mentioned in Section 2 in Wikidata we can consider editors with higher rights as ontology editors, and editors without specific rights as ontology contributors. In discussions, we observe a clear division between these **key roles** in cases where an editor suggests and talks about an action, like merging items or deleting properties, and another proceeds with this action. An example is the merging of items in (g) Table 6 or the replacement of a URL in (h). At a higher aggregation level, our thematic analysis also shows this role distinction: codes around suggesting changes in KE activities (9% in `itemTP`, 14% in `propertyTP`, 10% in `PC`) are more prevalent than codes on reporting that such changes have actually been committed to the graph (9% in `itemTP`, 10% in `propertyTP`, 2% in `PC`).

Version control in collaborative KE projects is more complex than in classical scenarios due to the size of the community, the asynchronous edits and the continues development of the KG. Wikidata, in order to preserve versioning, stores the history of all edits, including reverts and deletions. In discussions, we find references to edits, edit history and revision. Editors use the history of edits in practice to identify and revert erroneous contributions. An example is (i) Table 6. Furthermore, they also use the edits in discussions to point to specific revisions. In this way they can be specific about what was the edit, what was the revision, and argue about where they thought the error was.

Another distinctive KE activity in collaborative projects is the **flexibility** of the organisational structure of the project with regards to their policies and rules, the adaptability of conceptualisation and design workflows, and the required level of rigour in quality control. In Wikidata, we find the classical meaning of conceptualisation in ontology engineering when editors draw the hierarchy, but collaborative conceptualisation is guided by a set of policies and rules in order to keep consistency, quality and usability. In our thematic analysis we observe frequent cases where editors pose questions, share information, suggest or report regarding the project's policies and rules (*regulation or instruction* theme, 4% of the codes in `itemTP` and `propertyTP`, and 10% in `PC`). We can find discussions where editors refer to Wikidata regulations in order to correct edits that do not comply with the desired level of quality. An example is (j) 6. Furthermore, in order to maintain the quality, editors use the discussions to analyse the references used in content pages. We find examples of suggesting to add a reference like (k) Table 6, as well as long conversations about the use of the correct source as reference in order to be valid and consistent like (l).

Discussions about the ontology, the domain, and the engineering process are common in collaborative KE, raising the need for **tools** in order to facilitate and support them. Our analysis in Wikidata shows that editors often redirect the discussion to other channels (a talk page, for example in a WikiProject, or a communication page that fits the issue). We find various examples of such cross-channel discussion references, e.g. suggesting more suitable places to discuss the topic (example (m) in Table 6), pointing at related discussions elsewhere (example (n) in Table 6), or using other discussions as established know-how to solve issues (example (o) in Table 6). This is quite widespread among Wikidata discussions, as in our thematic analysis the code "connection with a discussion in another channel" shows in 5% of `itemTP`, 7% of `propertyTP`, and 8% of `PC` discussions annotating user comments that contain references to other discussions within Wikidata. The use of tools goes, however, beyond simply supporting discussions. In some cases, we find references to other Wikimedia tools providing analytics, bot support, and web services, as shown in (p) in Table 6. Moreover, editors tend to focus their attention on checking the quality of their work, by e.g. using queries to test if their changes are valid, as shown in example (c) Table 6.

7. Discussion

Our study analyses Wikidata discussions to find how much editors use the discussions, what they discuss, and how this supports KE activities for the maintenance and growth of the KG. Answering *RQ*1, we show that editors do not often use `itemTP` and `propertyTP` to discuss their issues and the existing talk pages follow a power law distribution with a substantial number of posts without response. For *RQ*2, we find that several themes are discussed, predominantly KE process/action, and others like taxonomy building and connection with other Wikimedia projects tend to be centralised in certain Wikidata areas. In addition, we did not find indications of intensified conflicts. Addressing *RQ*3, we identify KE activities in classical and collaborative KE like conceptualisation, implementation, mainte-

nance, key roles, version control, conceptualisation in collaborative manner, and tools.

Regarding *RQ*1, the results show that in Wikidata, editors do not often use the item and property talk pages to resolve issues. This contrasts with the critical importance of talk pages in other peer-production systems: in Wikipedia, 27% of the articles have discussions with a total number of 11,041,246 posts [29], while in Wikidata we find 0.02% of items having a talk page with a total number of 35,274 posts. Wikipedia talk pages added even faster than articles themselves (11x vs. 9x) and used as a prime medium for coordination and decision-making [58], increasing an article's number of edits and editors [45]. This is consistent with the abundant documentation and support that Wikipedia provides around talk pages: abundant help,[8] introductions,[9] and guidelines[10] are directly linked from high-traffic pages, including the front page. Conversely, the front page of Wikidata only links visitors to one internal community page, the project chat page,[11] then a live chat application[12] and a Wikimedia group chat[13]. The rich support regarding the use of talk pages in Wikipedia may come from previous analyses showing that editors often do not use the talk page rules [58]. Authors found that in 33% of their analysed corpus editors did not use a signature at the end of the post making it challenging to separate the different posts in a thread. Our study showed similar inconsistencies in the Wikidata talk page rules with 29% of the item talk pages did not start with a subject title. A clear indication of instructions related to the use of talk pages and communication channels could benefit Wikidata and improve communication. Furthermore, previous research on Schema.org [22] and GitHub [18] discussions report the crucial role of discussion instructions in advancing projects' development. Lack of guidance tends to centralise discussions in one place, duplicate work in multiple places, and leave large shares of posts on core topics unattended.

To have a first understanding of why Wikidata editors do not often use the items and property talk pages we looked at the use of the communication pages (see Table 1). We randomly checked a small sample of discussions (10 discussions) for each page to understand the main purpose of use. For the *general communication pages* we have extensively analysed the *project chat* with more than 14K discussions about items, properties, policies, process etc., while the *request for comments* with less than 200 discussions includes general topics like suggestion for protection for all the property pages. The page *report a technical problem* (previously named *contact the development team*) includes a high number of discussions reporting failures with the interface or suggestions for changes.

The *requests communication pages* we can separate them in two categories: contacting editors with higher rights; asking for a service. The first category includes most of them, *administrators' noticeboard, bureaucrats' noticeboard, translators' noticeboard, interwiki conflict, Wikidata: requests for permission, Wikidata:property proposal, properties for deletion*, and *requests for deletions*, giving the option to editors to ask specific group of admins to proceed with actions based on their rights. Using these communication pages editors have the chance to request actions like the inspection of a suspicious editor who maybe cause vandalism or request permission for the creation of a property or a bot. The page *Wikidata: requests for permission* have high traffic with editors asking to gain high level rights in Wikidata like to become administrators, translators etc., as well as to remove rights and permission to create a bot. In addition, the pages *Wikidata: property proposal*, and *requests for deletion* pages include more than 10K discussions requesting to add a new property or to delete or merge items in the KG. For *Wikidata: property proposal* the discussions follow a certain pattern as the editors need to agreed while fill in a template with the specifications of the new property. The discussion has the following structure: a table with the template (e.g., a description, the type of the property, several examples for each use) which will later be included on the top of its talk page; the motivation behind its creation; a voting process. The argumentation in the voting process concerns the template with the specifications of the property. In the *Wikidata: requests for deletion* there is not extent argumentation in the discussions with the majority to include two posts, one with the request and one with the action taken (e.g., deleted, merged etc.).

In the second category asking for a service, we can find the pages *bot requests* and *request a query*. In this case editors do not ask for permission to create for example a bot but ask for the community to create a bot with specific operation, as well as for a query code with particular results.

Our short investigation in the communication pages indicates that editors use these means of communications, some more than others, to address specific requests, questions or concerns. The low number in `itemTP` may is due to the *project chat* page that covers the general inquires, while the preliminary discussion about property specifications in the *Wikidata: property proposal* maybe justify the low number of discussions in `propertyTP`. However, the option to discuss in talk pages may duplicates work and leaves topics unattended. Furthermore, the purpose of certain pages can be confusing for editors as we have found discussions like a request to delete an item in the *request for comments* page, with a first reply suggesting to transfer the request on *requests for deletions* pages and a second suggesting that these requests are suitable for *project chat*.

Results also show that the posts in the studied discussion channels follow a power law distribution; an overwhelming majority of documents are *single-post* (50% for items, 8% for properties, and 16% for Project chat) documents. Similar power laws have been observed in Schema.org's user participation [22]. The lack of responses in posts combined with the lack of discussions' instructions can impact new members' engagement. A study about guidelines for the induction of new members in open source projects suggests answering quickly to newcom-

[8] `https://en.wikipedia.org/wiki/Help:Talk_pages`

[9] `https://en.wikipedia.org/wiki/Help:Introduction_to_talk_pages/1`

[10] `https://en.wikipedia.org/wiki/Wikipedia:Talk_page_guidelines`

[11] `https://www.wikidata.org/wiki/Special:MyLanguage/Wikidata:Project_chat`

[12] `https://web.libera.chat/?channel=#wikidata`

[13] `https://meta.wikimedia.org/wiki/Telegram#Wikidata`

ers in order to keep their motivation and make a good impression [50]. Furthermore, the participation of core contributors in discussions is crucial for the success of the project, not only to coordinate the activities but also to mentor the newcomers [18]. It is also suggested that newcomers should have a specific page in the project for training and discussion [18, 50]. Further analysis of the participation of new members could reveal more details about the impact of discussion in their motivation and encouragement, however, a suggestion for Wikidata could be to create specific instructions for the use of discussions and may a certain space for new members to seek help and guidance.

KE processes and actions are the most frequent topics in the analysed discussion channels. In talk pages, editors discuss the management of the content page, ask for help and specifications, and suggest or report changes. Similar use has been detected in Wikipedia article talk pages [58, 45]. However, the authors in Wikipedia argue that editors use these discussions to coordinate their editing activities in advance, while in Wikidata, `itemTP`, `propertyTP`, and `PC` are mostly used for instant changes in the KG. Nevertheless, a study by Kanke et al. [21] identifies that in Wikidata Wikiproject talk pages, it is more common for editors to organise and coordinate the necessary tasks [21].

We find that the taxonomy building collaboration occurs mostly in `itemTP`, where many posts remain without response. A study in Wikidata taxonomy hierarchies by Brasileiro et al. [7] shows violations in the use of properties like `instance of` and `subclass of`. Our findings, combined with the inconsistencies in the hierarchy, may indicate that the way editors discuss taxonomy building is inefficient, leaving requests unattended and leading to inconsistencies in the KG. Forming certain areas for the editors to discuss in more specialised topics like ontology development could enhance collaboration and improve the KG's building. However, the inefficiency in discussions may not come from the topic, in this case the taxonomy, but from how editors discuss it. KE discussions typically concern many relations between multiple resources (e.g. groups of items), while in item and property talk pages the discussions are organised in page-centric way (e.g. discuss about one specific item). This may indicate the need for tools to support cross-references in discussions. Our analysis shows that editors often use references to other discussions as examples (see supplementary material Table 2 code *Connection with a discussion in another channel*). The reference to previous discussions has been suggested to be good practice for community health [14]. While these references may support argumentation, the lack of support in tracking multiple discussions on similar topics across Wikidata may cause inconsistencies and confusion.

Conflicts are less usual in Wikidata than they are in e.g., Wikipedia [8], indicating a less confrontational community. In Wikidata, editors often disagree over their contributions, however, we find a low frequency of intensified conflicts, supporting that most of the conversations in these three discussion channels develop without intense disagreement. To further investigate our findings we studied a small sample of threads from the *Wikidata: Property proposal* and *Properties for deletions* communication pages. We sample 23 threads from each channel to explore further explore controversies and conflicts. We identify that 21% of the analysed threads include a controversy, while only 7% include conflicts. This is similar to the results we have about conflicts in `itemTP, porpertyTP` and `PC`. Besides the origin or the length of the discussion, the granularity of Wikidata's assertions may positively influence the accuracy of these discussions. While the limited evidence of conflict in Wikidata discussions can be interpreted as a signal of a healthy and pragmatic community, it might also be flagging that important discussions are not taking place or that there is simply too low content and domain discussion in talk pages for intellectual disagreement to arise. However, we did not analyse all of Wikidata's communication pages (see Table 1), and hence this might be based on partial observations. Future work could investigate how the community disagrees to better understand how decisions are made in Wikidata.

Through mapping discussion examples in the thematic analysis with classic KE activities (conceptualisation, implementation, maintenance), as well as other KE activities proper of collaborative projects (roles, version control, project flexibility, tool support), we find a rich overlap with the latter, but a partial one with the former. Numerous examples and the result of the thematic analysis show evidence that Wikidata has abundant activity in processes that are characteristic of collaborative KE, such as specialised role definitions, version control (with frequent interactions between discussions and the edit history), project flexibility (especially regarding the size and the openness of the community of contributors, and the required level of rigor in quality control [48]), and tool support (especially around user interaction, querying, and automating data processes). However, finding conversational evidence of more classic KE activities was much more challenging. For example, we could not find activities for the initial design of the ontology, like domain analysis, because it was completed before the publication of Wikidata. This is similar to other collaborative ontology engineering projects like Schema.org where Kanza et al. [22] show that the discussions support clarifications, extension functionalities and modification; but rarely a formal conceptualisation stage in the traditional KE sense. Similarly, in Wikidata it is hard to find traces of the implementation of the knowledge representation, typically done by translating the outcomes of the conceptualisation into a formal ontology language like RDF(S) or OWL. The majority of the analysed discussions (see Table 6) are written with the aim of discussing semantic relations without using technical terms, which are more characteristic of ontology experts in classic KE. This might be due to the level of experience of contributors discussing in the item space. A study by Muller et al. [36] on the Wikidata community found that only 2% of editors contribute to the core of ontology. This could imply that most editors working in the item space may not have ontology expertise. The study shows that for editors who work on items, approximately 72% of editors connect Wikidata to Wikipedia or other Wikimedia projects, 21% create items, and 27% edit items by translating terms in other languages. For more complex edits requiring an understanding of domain and hierarchy, 16% adds statements to items, 6% adds values to properties, and only 2% contributes to the property

pages as property engineers. Future studies on editors' expertise in relation to their contributions could shed more light on how Wikidata is built, and how the field of KE has transformed over the years.

The limitations of this study are the exclusion of languages different from English, the focus on three out of many communication pages in Wikidata, and the exclusion of information regarding the users participating in these conversations. The former was due to English being our common language, but the analyses of comments in other languages could show different discussion patterns. Moreover, the decision to analyse discussions in items, properties, and Project chat was made based on the attempt to study the main blocks of Wikidata. Other communication pages like Wikidata: property proposal or Wikidata: Requests for permissions could include another perspective and more details regarding how discussions in these channels are used.

8. Conclusions and future work

The community interactions constitute a crucial part of maintaining and growing a KG. We investigated three Wikidata discussion channels (item and property talk pages, and project chat discussions) using descriptive statistics, thematic analysis, and statistical tests to understand their structure and content. The results of this study showed that Wikidata editors do not usually discuss in item and property talk pages. There were many posts without responses and unclear instructions regarding using talk pages. Furthermore, editors most frequently discuss about KE activities around the KG, and we can identify patterns of KE activities in classical and collaborative KE in the main building areas of Wikidata. In general, following conversations between editors and understanding their follow-up actions to the KG from their dialogues were challenging tasks. We could find no evidence of guidance providing support to such knowledge engineering-oriented discussions, which can turn very specific (e.g. ontology evaluation through SPARQL queries directly pasted on the discussion). Therefore, we recommend these issues to be addressed through e.g. an accurate monitoring of empty or single-posted talk pages; the production of guides and recommendations about where, and how, KE issues should be discussed; and the deployment of tooling for linking discussions to the edit history (which continues to be generally missing in collaborative KE [48]). These, and other measures, could help to mitigate possible effects on KG quality and editor disengagement.

In future work, we will extend this research to the other Wikidata communication pages with high traffic like *Wikidata: property proposal* and *requests for deletion*, comparing them to the ones of this study. It would be interesting to explore discussion topics in other discussion channels as well as to further investigate disagreements in the community. Finally, we did not include information regarding the editors in the analyses, e.g., *How many editors participating in conversations?, How long are the editors editing in Wikidata?, Who answers more often the questions raised in the different discussion channels?*. Future work, will examine the interaction between the editors in the communication pages, the study of editors who maintain the communication pages, and the discussed themes for the different types of editors.

9. Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement no 812997 (CLEOPATRA ITN).

References

[1] Ahmad Alobaid, Daniel Garijo, María Poveda-Villalón, Idafen Santana-Perez, Alba Fernández-Izquierdo, and Oscar Corcho. Automating ontology engineering support activities with ontoology. *Journal of Web Semantics*, 57:100472, 2019.

[2] Hirojiro Aoyama. A study of stratified random sampling. *Annals of the Institute of Statistical Mathematics*, 6(1):1–36, 1954.

[3] Jaime Arguello, Brian S Butler, Elisabeth Joyce, Robert Kraut, Kimberly S Ling, Carolyn Rosé, and Xiaoqing Wang. Talk to me: Foundations for successful individual-group interactions in online communities. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 959–968, 2006.

[4] Deeksha Arya, Wenting Wang, Jin LC Guo, and Jinghui Cheng. Analysis and detection of information types of open source software issue discussions. In *2019 IEEE/ACM 41st International Conference on Software Engineering (ICSE)*, pages 454–464. IEEE, 2019.

[5] Yochai Benkler and Helen Nissenbaum. Commons-based peer production and virtue. *Journal of political philosophy*, 14(4), 2006.

[6] Yochai Benkler, Aaron Shaw, and Benjamin Mako Hill. Peer production: A form of collective intelligence. *Handbook of collective intelligence*, 175, 2015.

[7] Freddy Brasileiro, João Paulo A Almeida, Victorio A Carvalho, and Giancarlo Guizzardi. Applying a multi-level modeling theory to assess taxonomic hierarchies in wikidata. In *Proceedings of the 25th International Conference Companion on World Wide Web*, pages 975–980, 2016.

[8] Anamika Chhabra, Rishemjit Kaur, and SRS Iyengar. Dynamics of edit war sequences in wikipedia. In *Proceedings of the 16th International Symposium on Open Collaboration*, pages 1–10, 2020.

[9] Jacob Cohen. Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit. *Psychological bulletin*, 70(4):213, 1968.

[10] Oscar Corcho, Mariano Fernández-López, and Asunción Gómez-Pérez. Methodologies, tools and languages for building ontologies. where is their meeting point? *Data & knowledge engineering*, 46(1):41–64, 2003.

[11] To Tu Cuong and Claudia Müller-Birn. Applicability of sequence analysis methods in analyzing peer-production systems: a case study in wikidata. In *International Conference on Social Informatics*, pages 142–156. Springer, 2016.

[12] Olive Jean Dunn. Multiple comparisons using rank sums. *Technometrics*, 6(3):241–252, 1964.

[13] Lisa Ehrlinger and Wolfram Wöß. Towards a Definition of Knowledge Graphs. *SEMANTiCS (Posters, Demos, SuCCESS)*, 48:1–4, 2016.

[14] M Elliott and Walt Scacchi. Communicating and mitigating conflict in open source software development projects. *Projects & Profits*, pages 25–41, 2002.

[15] Sean Falconer, Tania Tudorache, and Natalya F Noy. An analysis of collaborative patterns in large-scale ontology development projects. In *Proceedings of the sixth international conference on Knowledge capture*, pages 25–32, 2011.

[16] Joseph L Fleiss, Bruce Levin, Myunghee Cho Paik, et al. The measurement of interrater agreement. *Statistical methods for rates and proportions*, 2(212-236):22–23, 1981.

[17] Noriko Hara, Pnina Shachaf, and Khe Foon Hew. Cross-cultural analysis of the Wikipedia community. *Journal of the American Society for Information Science and Technology*, 61(10):2097–2108, 2010.

[18] Hideaki Hata, Nicole Novielli, Sebastian Baltes, Raula Gaikovina Kula, and Christoph Treude. Github discussions: An exploratory study of early adoption. *Empirical Software Engineering*, 27(1):1–32, 2022.

[19] Hsiu-Fang Hsieh and Sarah E Shannon. Three approaches to qualitative content analysis. *Qualitative health research*, 15(9):1277–1288, 2005.

[20] Glenn D Israel. Determining sample size. Technical report, University of Florida, Cooperative Extension Service, Institute of Food and Agricultural Sciences, 1992.

[21] Timothy Kanke. Knowledge curation work in wikidata wikiproject discussions. *Library Hi Tech*, 2020.

[22] Samantha Kanza, Alex Stolz, Martin Hepp, and Elena Simperl. What Does an Ontology Engineering Community Look Like? A Systematic Analysis of the schema. org Community. In *European Semantic Web Conference*, pages 335–350. Springer, 2018.

[23] Maria Keet. *An introduction to ontology engineering*, volume 1. Maria Keet Cape Town, 2018.

[24] Konstantinos Kotis. On supporting hcome-3o ontology argumentation using semantic wiki technology. In *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*, pages 193–199. Springer, 2008.

[25] Konstantinos Kotis and George A Vouros. Human-centered ontology engineering: The HCOME methodology. *Knowledge and Information Systems*, 10(1):109–131, 2006.

[26] Konstantinos I Kotis, George A Vouros, and Dimitris Spiliotopoulos. Ontology engineering methodologies for the evolution of living and reused ontologies: status, trends, findings and recommendations. *The Knowledge Engineering Review*, 35, 2020.

[27] William H Kruskal and W Allen Wallis. Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, 47(260):583–621, 1952.

[28] J Richard Landis and Gary G Koch. The measurement of observer agreement for categorical data. *biometrics*, pages 159–174, 1977.

[29] David Laniado, Andreas Kaltenbrunner, Carlos Castillo, and Mayo Fuster Morell. Emotions and dialogue in a peer-production community: the case of wikipedia. In *proceedings of the eighth annual international symposium on wikis and open collaboration*, pages 1–10, 2012.

[30] David F Marks and Lucy Yardley. *Research methods for clinical and health psychology*. Sage, 2004.

[31] Courtney Miller, Sophie Cohen, Daniel Klug, Bogdan Vasilescu, and Christian Kästner. "did you miss my comment or what?" understanding toxicity in open source discussions. In *2022 IEEE/ACM 44th International Conference on Software Engineering (ICSE)*, pages 710–722, 2022.

[32] Audris Mockus, Roy T Fielding, and James Herbsleb. A case study of open source software development: the apache server. In *Proceedings of the 22nd international conference on Software engineering*, pages 263–272, 2000.

[33] Aldo de Moor, Pieter De Leenheer, and Robert Meersman. DOGMA-MESS: A meaning evolution support system for interorganizational ontology engineering. In *International Conference on Conceptual Structures*, pages 189–202. Springer, 2006.

[34] Marie T Morris, Aaron Daluiski, and Christopher J Dy. A thematic analysis of online discussion boards for brachial plexus injury. *The Journal of hand surgery*, 41(8):813–818, 2016.

[35] Iraklis Moutidis and Hywel TP Williams. Community evolution on stack overflow. *Plos one*, 16(6):e0253010, 2021.

[36] Claudia Müller-Birn, Benjamin Karran, Janette Lehmann, and Markus Luczak-Rösch. Peer-production system or collaborative ontology engineering effort: What is wikidata? In *Proceedings of the 11th International Symposium on Open Collaboration*, pages 1–10, 2015.

[37] Thomas Pellissier Tanon, Denny Vrandečić, Sebastian Schaffert, Thomas Steiner, and Lydia Pintscher. From freebase to wikidata: The great migration. In *Proceedings of the 25th international conference on world wide web*, pages 1419–1428, 2016.

[38] Yumeng Peng and Xiang Zhou. A cross-cultural analysis of the modes and effectiveness of collaborative production of knowledge on quora. *Aslib Journal of Information Management*, 2020.

[39] Alessandro Piscopo, Christopher Phethean, and Elena Simperl. Wikidatians Are Born: Paths to Full Participation in a Collaborative Structured Knowledge Base. In *50th Hawaii International Conference on System Sciences*, pages 4354–4363. University of Hawaii, 2017.

[40] Alessandro Piscopo and Elena Simperl. Who Models the World? Collaborative Ontology Creation and User Roles in Wikidata. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW):1–18, 2018.

[41] Jennifer Preece and Ben Shneiderman. The reader-to-leader framework: Motivating technology-mediated social participation. *AIS transactions on human-computer interaction*, 1(1):13–32, 2009.

[42] Debbie Richards. Collaborative knowledge engineering: Socialising expert systems. In *2007 11th International Conference on Computer Supported Cooperative Work in Design*, pages 635–640. IEEE, 2007.

[43] Cristina Sarasua, Alessandro Checco, Gianluca Demartini, Djellel Difallah, Michael Feldman, and Lydia Pintscher. The evolution of power and standard wikidata editors: comparing editing behavior over time to predict lifespan and volume of edits. *Computer Supported Cooperative Work (CSCW)*, 28(5):843–882, 2019.

[44] Sebastian Schaffert. Ikewiki: A semantic wiki for collaborative knowledge management. In *15th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE'06)*, pages 388–396. IEEE, 2006.

[45] Jodi Schneider, Alexandre Passant, and John G Breslin. A content analysis: How Wikipedia talk pages are used. In *The Web Science Conference 2010 (WebSci '10). Raleigh, North Carolina, USA*, 2010.

[46] Moritz Schott, Asher Yair Grinberger, Sven Lautenbach, and Alexander Zipf. The Impact of Community Happenings in OpenStreetMap—Establishing a Framework for Online Community Member Activity Analyses. *ISPRS International Journal of Geo-Information*, 10(3):164, 2021.

[47] Juan Sequeda and Ora Lassila. Designing and building enterprise knowledge graphs. *Synthesis Lectures on Data, Semantics, and Knowledge*, 11(1):1–165, 2021.

[48] Elena Simperl and Markus Luczak-Rösch. Collaborative ontology engineering: a survey. *The Knowledge Engineering Review*, 29(1):101–131, 2014.

[49] Elena Simperl, Malgorzata Mochol, and Tobias Bürger. Achieving maturity: the state of practice in ontology engineering in 2009. *International Journal of Computer Science and Applications*, 7(1):45–65, 2010.

[50] Igor Steinmacher, Christoph Treude, and Marco Aurelio Gerosa. Let me in: Guidelines for the successful onboarding of newcomers to open source projects. *IEEE Software*, 36(4):41–49, 2018.

[51] Rudi Studer, V Richard Benjamins, and Dieter Fensel. Knowledge engineering: Principles and methods. *Data & knowledge engineering*, 25(1-2):161–197, 1998.

[52] Mari Carmen Suárez-Figueroa, Asunción Gómez-Pérez, and Mariano Fernández-López. The NeOn methodology for ontology engineering. In *Ontology engineering in a networked world*, pages 9–34. Springer, 2012.

[53] York Sure, Michael Erdmann, Juergen Angele, Steffen Staab, Rudi Studer, and Dirk Wenke. OntoEdit: Collaborative ontology development for the semantic web. In *International semantic web conference*, pages 221–235. Springer, 2002.

[54] Christoph Tempich, H Sofia Pinto, York Sure, Denny Vrandecic, Núria Casellas, and Pompeu Casanovas. Evaluating diligent ontology engineering in a legal case study. In *IVR 22nd World Congress–Law and Justice in a Global Society. International Association for Philosophy of Law and Social Philosophy*, 2005.

[55] Christoph Tempich, Elena Simperl, Markus Luczak, Rudi Studer, and H Sofia Pinto. Argumentation-based ontology engineering. *IEEE Intelligent Systems*, 22(6):52–59, 2007.

[56] Tania Tudorache, Natalya F Noy, Samson Tu, and Mark A Musen. Supporting collaborative ontology development in protégé. In *International Semantic Web Conference*, pages 17–32. Springer, 2008.

[57] Fernanda B Viégas, Martin Wattenberg, and Kushal Dave. Studying cooperation and conflict between authors with history flow visualizations. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 575–582, 2004.

[58] Fernanda B Viegas, Martin Wattenberg, Jesse Kriss, and Frank Van Ham. Talk before you type: Coordination in Wikipedia. In *2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07)*, pages 78–78. IEEE, 2007.

[59] Denny Vrandečić and Aldo Gangemi. Unit tests for ontologies. In *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*, pages 1012–1020. Springer, 2006.

[60] Denny Vrandečić and Markus Krötzsch. Wikidata: A Free Collaborative

Knowledge Base. *Communications of the ACM*, 57(10):78–85, 2014.
[61] Denny Vrandečić, Sofia Pinto, Christoph Tempich, and York Sure. The DILIGENT knowledge processes. *Journal of Knowledge Management*, 9(5):85–96, 2005.
[62] W3C. OWL Web Ontology Language Overview. `https://www.w3.org/TR/owl-features/`.
[63] W3C. RDF Schema 1.1. `https://www.w3.org/TR/rdf-schema/`.