Working for the Invisible Machines or Pumping Information into an Empty Void? An Exploration of Wikidata Contributors' Motivations

CHARLES CHUANKAI ZHANG* and MO HOUTTI*, University of Minnesota, USA
C. ESTELLE SMITH, University of Colorado Boulder, USA
RUOYAN KONG, University of Minnesota, USA
LOREN TERVEEN, University of Minnesota, USA

Structured data peer production (SDPP) platforms like Wikidata play an important role in knowledge production. Compared to traditional peer production platforms like Wikipedia, Wikidata data is more structured and intended to be used by machines, not (directly) by people; end-user interactions with Wikidata often happen through intermediary "invisible machines." Given this distinction, we wanted to understand Wikidata contributor motivations and how they are affected by usage invisibility caused by the machine intermediaries. Through an inductive thematic analysis of 15 interviews, we find that: (i) Wikidata editors take on two archetypes—Architects who define the ontological infrastructure of Wikidata, and Masons who build the database through data entry and editing; (ii) the structured nature of Wikidata reveals novel editor motivations, such as an innate drive for organizational work; (iii) most Wikidata editors have little understanding of how their contributions are used, which may demotivate some. We synthesize these insights to help guide the future design of SDPP platforms in supporting the engagement of different types of editors.

CCS Concepts: • **Human-centered computing → Empirical studies in collaborative and social computing**; **Wikis**;

Additional Key Words and Phrases: Wikidata; interview study; social computing; peer production; structured data

ACM Reference Format:
Charles Chuankai Zhang, Mo Houtti, C. Estelle Smith, Ruoyan Kong, and Loren Terveen. 2022. Working for the Invisible Machines or Pumping Information into an Empty Void? An Exploration of Wikidata Contributors' Motivations . *Proc. ACM Hum.-Comput. Interact.* 6, CSCW1, Article 135 (April 2022), 21 pages. https://doi.org/10.1145/3512982

1 INTRODUCTION

Founded in 2012, Wikidata [38] is a free and open structured data peer production (SDPP) platform that provides information about humans, objects, locations, concepts, etc. and the relationships between them. The information stored on Wikidata enables it to serve as a centralized knowledge base for other Wikimedia projects—e.g., Wikipedia and Wikivoyage—which can link to structured

*Both authors contributed equally to this research.

Authors' addresses: Charles Chuankai Zhang; Mo Houtti, University of Minnesota, 200 Union St SE, Minneapolis, Minnesota, 55455, USA; C. Estelle Smith, University of Colorado Boulder, 80305, USA; Ruoyan Kong, University of Minnesota, Minneapolis, MN, 55455, USA; Loren Terveen, University of Minnesota, Minneapolis, MN, 55455, USA.

2573-0142/2022/4-ART135 $15.00
https://doi.org/10.1145/3512982

information on Wikidata instead of re-encoding the same information locally.[1] This is Wikidata's primary use-case across the web as well; unlike Wikipedia, where articles are mainly consumed by human readers, Wikidata is mainly used as a data source for various software tools.

Wikidata's database of information is constructed entirely by volunteer editors and the bots they create. Creating an accurate and functional database requires editors to take on several different types of tasks. Data modeling and data entry are two of the major task types on Wikidata. Modeling tasks involve making sure that the database represents items in ways that match their real-world ontology—for example, deciding that country of citizenship should be a property for human items but not for other types of items. Making these kinds of modeling decisions often requires multiple iterations that involve discussion among Wikidata community members. The contributions of editors who engage in modeling tasks serve as guidelines for their fellow editors. Data entry tasks then consist of following these guidelines to link or import data from other Wikimedia projects or external databases. For example, data entry might consist of adding the correct country of citizenship to one or more human items.

A foundational question in the study of peer production systems in general is *what motivates contributors to volunteer their time and effort?* Prior work has demonstrated both intrinsic [28, 47] and extrinsic [30] motivations for contributing. However, the nature of working on SDPP platforms and the outcomes of that work differ from conventional peer production systems. For instance, work such as drafting, writing or editing Wikipedia articles is constrained by meeting Wikipedia's standards for formatting, quality, notability, neutrality, citations, etc., and the results are viewed primarily by humans. The semantic and stylistic structure of individual articles can evolve organically over time without impacting the rest of Wikipedia. Because of variation in article structure, as well as differences across Wikipedia's many language editions, there is no straightforward way to accurately and automatically extract semantic relationships between articles (e.g., by scraping info from the sidebar or article itself).

On the other hand, working on Wikidata involves tasks like creating items, setting statements, terms, and sitelinks, and protecting or reverting items or properties [25]. This work is constrained by the rules of the database that generate semantic meaning across all items, and the outcomes of this work are used primarily by machines. Bots are used to help populate fields in Wikidata items. However, human editors need to ensure that the information is accurate and consistent with the intended purpose of the field; otherwise machine errors could propagate across the entire Wikimedia ecosystem and beyond. Given these fundamental distinctions in the nature of work between conventional peer production platforms and SDPP platforms, there is a gap in our understanding of what motivates people to work on SDPP platforms. Thus, we formulated the following research questions to guide our work:

RQ1: *What personal characteristics lead Wikidata editors to take on certain types of tasks over others?*

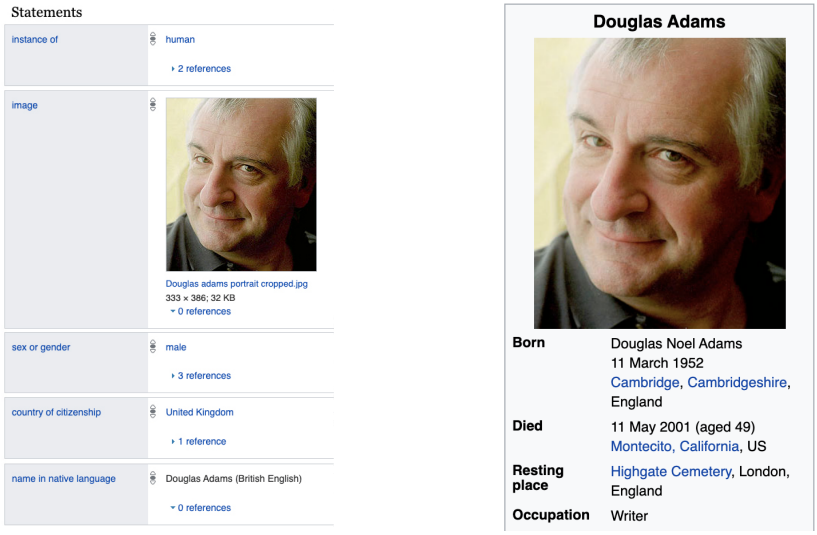
RQ2: *What editor motivations are fulfilled by the structured nature of Wikidata?*

RQ3: *How are editors motivated (or demotivated) by the perceived usage of their contributions?*

We carried out an interview study to explore these questions and performed an inductive thematic analysis [8] of interview transcripts. Our findings are as follows:

- Based on our observation that user characteristics along several dimensions go hand-in-hand—e.g., editors who prefer low cognitive burden are often the same users who prefer solitary work—we propose two contributor archetypes—*Architects* and *Masons*—as heuristics

[1] The process of linking information on Wikidata to its sister projects is ongoing, so it is unclear how much of the information on sister projects that hypothetically *could* be linked from Wikidata actually *is*.



(a) Screenshot of the Wikidata item for Douglas Adams

(b) Screenshot of the Wikipedia infobox for Douglas Adams

Fig. 1. Example of a Wikidata item (left) and its corresponding Wikipedia infobox (right). Note that these screenshots are partial representations. The item for Douglas Adams has over 40 properties, including: date of birth, place of birth, date of death, place of death, place of burial, image of grave, occupation, father, mother, sibling, spouse, child, etc.

for understanding patterns of editing activity, preferences, personal characteristics, and motivations.
- We identify novel editor motivations derived from the structured nature of Wikidata, such as an innate drive for categorization and organization.
- We find that most Wikidata contributors have little understanding of how their contributions are actually *used* and, therefore, the *value* of their contributions.

We draw a set of implications from our findings, directed towards improving contributor engagement. Notably, we suggest ways in which our identified roles might provide novel utility to SDPP platform designers, and we propose solutions to mitigate the existence of **"invisible machines"**—the many unknown, distinct software entities that obscure Wikidata's usage by downloading and querying its data off-platform.

2 BACKGROUND INFORMATION

This section introduces Wikidata in detail by providing information about its data modeling structure, types of editor tasks, and how Wikidata is used.

2.1 Data Structure and Task

The primary building blocks of Wikidata are items, which contain statements. Statements are how information about an item is encoded in Wikidata. Each statement consists of a property that is paired with at least one value. A property represents a characteristic of an item, such as the coordinate location of a city or a person's sibling. A value is then applied to a property to encode information about that characteristic. The item for Douglas Adams [41], for example, contains a

statement with the property *country of citizenship* that is paired with the value *United Kingdom*, along with references for verifying the sources of the information (Fig. 1). Items on Wikidata also include labels and descriptions, which make them easily searchable by users using different languages.

Editors can get involved in Wikidata through various tasks such as data modeling, data entry, patrolling for vandalism, and fixing item properties [43]. Data entry tasks can often be quick and easy—e.g., adding a professor's official website link to his Wikidata page by inputting it as a value for the property *official website P*586. Data modeling tasks are typically more complex, as they require making ontological decisions, that is, how to represent entities consistently and rigorously. For example, each property has *constraints*, which are rules for how it should be used. Data modeling might involve making decisions about the appropriate constraints for a property. Since data modeling tasks such as the engineering of properties affect not only the properties themselves, but also items correlated with particular properties, decisions are often made through iterative discussions and refinement from multiple Wikidata editors. Discussion pages of Wikidata properties often include a considerable amount of discussion, as editors exchange perspectives, justify modifications, and request changes. Therefore, while *some* data modeling tasks such as merging duplicates or connecting similar properties could be automated using bots, the major bulk of data modeling tasks, especially those ontology engineering tasks that require judgement and reasoning, must be accomplished by human editors.

2.2 Contribution Reuse

Wikidata functions in part as a central repository of structured data for its Wikimedia sister projects. For example, some Wikipedia pages use Wikidata-powered infoboxes to fetch data directly from the connected Wikidata item instead of storing the information on Wikipedia itself. The "In other projects" and "In other languages" links on Wikivoyage are also partially populated from Wikidata.

Wikidata usage, however, is not limited to Wikimedia projects. Wikidata's intro page [42] states that *"Imposing a high degree of structured organization allows for easy reuse of data by Wikimedia projects and third parties, and enables computers to process and 'understand' it."* In other words, Wikidata is intended as an open database for structured information that can easily be pulled into tools across the internet. These tools function as "machine intermediaries" that consume Wikidata information on Wikidata and repackage it for end-users. One such tool, Scholia, is *"a service that creates visual scholarly profiles for topics, people, organizations, species, chemicals, etc. using bibliographic and other information in Wikidata"* [33]. Scholia illustrates one way in which Wikidata's data might be distributed through a machine intermediary. The scope of information intended to be documented by Wikidata is, however, quite comprehensive. Wikidata's licensing is also highly non-restrictive; it freely allows commercial use of its data, so it is used by some for-profit businesses to power online tools. For example, Google use Wikidata as a source for improving their Knowledge Graph.

In the remainder of the paper, we situate our study with respect to related work, describe our methods, discuss our results in detail, and conclude by reflecting on design implications and directions for future work.

3 RELATED WORK

We introduce related work in the subsections below. Each subsection situates and motivates the research question stated at the beginning. We first discuss literature on roles in online communities, then research in peer production editor motivations, and finally engagement maintenance.

3.1 Roles in Online Communities

RQ1: *What personal characteristics lead Wikidata editors to take on certain types of tasks over others?*

Previous studies have identified different types of roles in various contexts. In the Open Source Software space, Nakakoji et al. [26] studied four different OSS projects and came up with eight roles that members take on, such as Project Leader, Core Member and Active Developer. These roles reflect differences in users' influence rather than a strict hierarchical structure [46]. Jin et al. [21] developed different member archetypes such as founder/officer, meeting and facilities coordinator, and public relations officer based on studies on Linux user groups from a user-oriented perspective. Through an interview study with 17 OSS contributors, Trinkenreich et al. [37] identified the existence of both community-centric roles and project-centric roles. They suggested that people build career pathways through two kinds of roles and could move from one to another. They also highlighted the importance of making the roles visible and recognized, as some of the roles may not directly contribute to the software itself. In the field of citizen science, where there is an asymmetric structure between the volunteers who make contributions and the scientists who benefit from them [29], there has been research studying the formation of different user groups and the governance structures within different projects [12]. For example, Bowser et al. [7] identified two user groups—nature participants and gamer participants—in a gamified citizen science mobile application, which differ from each other in motivations as well as how they perceive the value of their contributions.

Within the Wikipedia space, Bryant et al. [9] and Zhu et al. [48] observed the behavior and editing patterns that distinguish peripheral users from central users. Specifically, novice editors edit *what they know* while expert editors aim at *building Wikipedia* and perceive the improvement of Wikipedia quality as motivation [9]. Welser et al. [40] examined qualitative data of user comments and identified four social roles within Wikipedia: substantive experts, technical editors, vandal fighters, and social networkers. Arazy et al. [3] examined functional roles in Wikipedia and presented role transition dynamics from periphery users to core contributors.

For Wikidata specifically, Piscopo and Simperl [31] and Müller-Birn et al. [25] defined user roles based on quantitative analyses of editing activity. Müller-Birn et al. [25] used k-means to cluster the participation patterns of both human and bot Wikidata editors based on their revision histories. They found that the majority of Wikidata contributors take on one of the following specialized roles: reference editor, item creator, item editor, item expert, property editor, and property engineer. The first 5 roles can be occupied by both humans and bots, as they involve following the existing Wikidata data model to enter new data. The property engineer role, however, is the only one occupied solely by humans, as it requires defining and communicating the structure that other humans and bots follow to enter new data, primarily through: (1) the creation of new properties and (2) communication on talk pages.

Our work builds on the same line of inquiry as Müller-Birn et al. [25]. Our own findings substantiate theirs using alternative methods—which is valuable in and of itself—but also provide novel and unique insights that could not be gleaned from the quantitative inquiry they conducted (or any other purely quantitative inquiries, for that matter). The primary way in which we build on their prior work is by tying their editing-pattern-based roles to editors' personal characteristics—none of which are observable in the data traces available for quantitative analysis. Our work fills this gap in understanding by applying an in-depth qualitative approach that seeks to elucidate the various ways in which the occupants of these roles might differ in their motivations and personal characteristics. In our results section, we describe how these quantitatively identified roles relate to the two main Architect and Mason archetypes that emerged from our analysis. Further, we provide

novel insights into how these archetypes differ along three dimensions—cognitive engagement, desire for collaboration, and domain expertise—and reveal how these contrasting characteristics lead to differences in motivations. The similarities and distinctions between the two archetypes identified in this paper shed light on important design implications for developing tools and systems that could assist different stakeholder groups in the community.

3.2 Motivation in Online Collaboration Platforms

RQ2: *What editor motivations are fulfilled by the structured nature of Wikidata?*

One of the core characteristics of peer production is its ability to harness different kinds of motivations [5]. According to self-determination theory [17], contributors are motivated by intrinsic and extrinsic factors that lead members to contribute time, effort and expertise to the community. In OSS communities, it has been found that the enjoyment-based intrinsic motivation is one of the most important motivators [24]. Learning new skills [24, 46], feelings of competence [23], and reciprocity [6] are other factors that motivate OSS contributors. In citizen science projects where volunteers take on tasks like data collection and curation, Rotman et al. [32] found that participants are motivated by personal interests as well as external factors such as recognition and attribution. Nov et al. [29] conducted a survey study and identified collective and intrinsic motivations as the key motivation factors.

In the context of Wikipedia, through an online survey, Yang and Lai [45] identified that Wikipedia contributors are motivated by having their internal standards met through positive feedback. Using the taxonomy of volunteering activity (6 motivational categories), Nov [28] described a survey aimed at categorizing the motivations of Wikipedians, the results of which showed that Wikipedia editors are driven primarily by fun and ideology.

Motivation changes at different stages of the collaborative process, and different stakeholders may have distinct goals and values that they seek in the contribution process [32]. Understanding the unique motivation factors of different types of users and different platforms can lead to design changes that improve platform effectiveness. In this paper, we find that editors have some of the same motivations described in previous research on peer production platforms. However, we also identify novel ways in which the structured nature of Wikidata creates additional motivations in the SDPP setting.

3.3 Engagement and Feedback

RQ3: *How are editors motivated (or demotivated) by the perceived usage of their contributions?*

One of the major challenges of peer production systems lies in maintaining the engagement of editors, especially newcomers who have not yet fully integrated into the community and are more likely to stop contributing [44]. Feedback plays a significant role in attempting to overcome this challenge. For example, Chen et al. [11], Zhu et al. [50], and Asadi et al. [4] showed that feedback from peers and other social aspects are key factors in keeping editors engaged in a platform. Gorbatai [18] revealed that ignoring the relationship between producers and consumers may result in negative consequences in knowledge production. It also suggested that intermediaries between these two parties can play an important role as well. In virtual communities, reviews, comments, questions and other kinds of feedback lead to improvement of contribution quality [36]. In the OSS development process, feedback is also vital for expert producers to refine their contributions [34]. Misalignment between supply and demand would significantly affect the quality of knowledge contribution in peer production platforms like Wikipedia [37]. Thus a connection between contributors and their audience through feedback and reveal of usage will direct the community to tackle demand for content in an efficient way [37]. Building on this, we describe editors' perceptions about how their

ID	Gender	Age	Time on Wikidata	Archetype
P1	Male	25-34	Since the beginning	Architect
P2	Male	18-24	Under 3 months	Mason
P3	Female	55+	Over 3 years	Architect
P4	Male	55+	Over 3 years	Architect
P5	Male	25-34	Over 3 years	Architect
P6	Male	55+	Over 3 years	Architect
P7	Male	18-24	1 to 3 years	Mason
P8	Agender	25-34	Over 3 years	Architect
P9	Female	55+	1 to 3 years	Architect
P10	Male	35-44	Under 3 months	Mason
P11	Male	25-34	1 to 3 years	Architect
P12	Male	25-34	Over 3 years	Architect
P13	Male	35-44	1 to 3 years	Mason
P14	Male	45-54	Over 3 years	Architect
P15	Female	55+	Over 3 years	Architect

Table 1. Participant demographics and archetype classifications by two researchers.

contributions are used and how Wikidata fails to provide adequate feedback to editors about the value of their contributions. We attribute this to the existence of invisible machines through which Wikidata is consumed by end-users.

4 METHODS

After obtaining approval from our Institutional Review Board, we posted our study plans on the Wikimedia research meta-wiki to elicit community feedback and ensure that our research would not harm the community in any way. Several Wikidata community members responded in support of our research efforts and methods. While our original research goals were broader than the research questions presented here, these members helped us to refine and narrow our scope to focus more specifically on the motivations and values of Wikidata contributors. Thus, we clarified the background and motivation of our study on Wikidata, and adjusted our interview questions accordingly. Because of this step, we are confident that our work provides contributions that will be useful to the Wikidata community as well as of general research interest.

4.1 Participant Recruitment

We posted our research interview invitation on the Wikidata project chat page. We simultaneously went through Wikidata edit statistics on Wikiscan, and identified users who were actively editing at that time and sent out individual invites through their personal talk pages. Editors who were interested were asked to fill out a brief survey about their contact and demographic information. We recruited 15 participants, with whom three of the authors conducted semi-structured interviews via Zoom or Google Meet. After about 10 interviews, we began to observe data saturation—i.e., subsequent interviews were not revealing new concepts that we had not already learned about from prior participants. We stopped recruiting after completing 15 interviews, which is consistent with the average sample size of qualitative work in the ACM CHI conference [10]. Demographic information for the participants can be found in Table 1.

4.2 Data Collection

We completed and transcribed the 15 interviews with an average length of 50:24 minutes (range from 28:51 to 80:37 minutes). The interviews had three parts. First, we asked general questions about contributors' work processes, for example, what types of content they edit on Wikidata. Second, we asked questions designed to uncover contributor motivations, for example, what drew them to Wikidata. Third, we asked questions to explore how contributors assessed the value of their contributions and who they thought the audience for their contributions was. A representative list of interview questions is included in the appendix (See APPENDIX:Interview Questions). Since interviews were semi-structured, additional follow-up questions were also asked.

4.3 Analysis

We followed an inductive thematic analysis approach [8] to analyze our interview data. Inductive analysis allows collective or shared meanings to emerge in a way that is commonly mentioned or perceived by the interviewees. In order to achieve this, we first transcribed and anonymized all the conducted interviews. Then four of the authors coded two interviews together to reach a consensus of how we wanted to summarize and describe the data. The rest of the interviews were coded individually. Next, after getting familiar with the codes, four of the authors clustered codes that shared unifying features together on a virtual code board. After the codes formed different clusters, we added high level topics and themes to each cluster that summarized the content.

We observed two high-level clusters of responses, giving rise to what we came to call the Architect and Mason archetypes. Codes and themes were selected in a way that could best delineate the difference of characteristics between two archetypes. Two of the researchers independently completed informal classifications of each participant into one of the two archetypes based on readings of their interview transcripts. The researchers then resolved disagreements by discussing them in detail, with a particular focus on aspects of the transcripts that led to differing decisions about the participants' positions along each of the three pertinent dimensions—domain expertise, desire for collaboration, and cognitive engagement. The two researchers were able to come to a final agreement on all classifications at the end of this process.

There were a few interesting topics and quotes that emerged from our discussions with Wikidata editors, such as the perception of Wikipedia and rules in Wikidata that we did not select as they were not tightly connected to our major themes.

4.4 Limitations

Our methods have two main limitations. First, we recruited the majority of our participants by finding highly active Wikidata editors through Wikiscan. As with other qualitative work in the Wikimedia ecosystem (e.g., Smith et al. [35]), it is notoriously challenging to recruit newcomers who have limited experience with and commitment to the project. While we were able to recruit two participants from the Wikidata project chat page who were newer to Wikidata, our findings primarily describe experienced Wikidata editors—a crucial stakeholder group whose ongoing dedication to the project is vital to its success. A distinct but important question for future work is to understand what draws newcomers to Wikidata in the first place, and what can be done to engage and retain their participation.

Second, as qualitative research, this work contributes an in-depth understanding of the motivations of Wikidata editors, and describes patterns in editor characteristics that might inform the ways in which system designers fulfill these motivations. However, our small sample size cannot be considered representative of all Wikidata editors, so future quantitative work should verify whether our results hold at scale. As one example, editors' personal characteristics and motivations

could be determined in future work by using a widely deployed survey. It would then be possible to examine the relationships between survey-derived personal characteristics and motivations (i.e., our archetypes) and editors' editing patterns (i.e., Müller-Birn et al.'s [25] roles) at scale to determine whether the relationships we identify between personal characteristics and editing patterns apply to all (or some subset) of the Wikidata editor population.

5 RESULTS

In this section, we introduce a taxonomy that delineates two editor archetypes—Architects and Masons (Fig. 2). Since we will describe editor characteristics in detail as they pertain to these archetypes, we begin with a brief description of participants to ground subsequent results.

Most of our participants were highly active and experienced Wikidata editors but were nevertheless quite diverse in terms of their backgrounds and approaches to editing Wikidata. Some editors are focused and purposeful in their editing approach. For example, P12 is an archaeologist who is interested in the ways Wikidata could be used by academics to visualize research. P9 is a plant enthusiast who aims to make finding plant names and locations in her home country easier for others. P3 *"spent about 40 years building the library that you see behind me here, of books on the topics I'm interested in"* and adds information about these topics—e.g., clothing and textiles—to Wikidata as a means of sharing information with the world. Other editors are more casual and aimless in their approaches to Wikidata. P10, for example, is a biologist who began by adding various insect species to Wikidata but, through a series of rabbit holes, now finds himself adding publication info for *"this one scientist from the 1800s."* P13 will often pick up his phone and just *"load up Wikidata"* when he comes across some interesting bit of information in a podcast.

Despite these differing backgrounds, personal characteristics, and editing approaches, we show that both Architects and Masons are intrinsically motivated by working with structured data, but that Masons are additionally motivated by the way structured data facilitates scaffolding of tasks. We conclude by explaining how Wikidata's machine intermediaries have a hand in demotivating contributors, by making it difficult for them to understand the value their contributions create.

We state the relevant RQ at the beginning of each subsection, and provide a bolded summary of how our findings answer it at the end of the subsection.

5.1 Two Contributor Types: Architects & Masons

RQ1: *What personal characteristics lead Wikidata editors to take on certain types of tasks over others?*

Architects and Masons differ along three dimensions: cognitive engagement, desire for collaboration, and domain expertise. We find that occupants of the first 5 roles in Müller-Birn et al. [25] are well described by the Mason archetype, as they primarily focus on creating and editing references and items, and editing properties. We find that Architects, who enjoy collaborative and mentally engaging work that cannot be performed by bots, take on many of the same tasks described by the property engineer role in Müller-Birn et al. [25].

This taxonomy is not meant as a hard classification tool—editors do not always fall neatly into one group or the other—but the archetypes provide a useful framework for understanding some of the ways in which editors' personal characteristics relate to their editing patterns. These personal characteristics are also useful for understanding differences in motivations, which we explore later under RQ2.

We identified 4 of our participants as Masons and the other 11 as Architects, so both archetypes represent observations we made across multiple participants. Two disagreements on editor classification arose. One of the disagreements—for P12—was caused by differing interpretations of

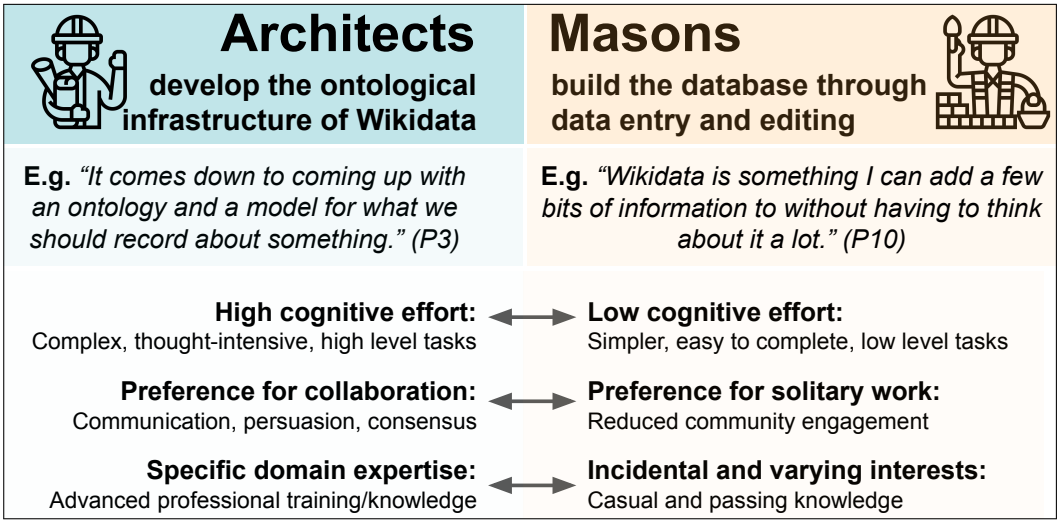


Fig. 2. Summary of Wikidata Archetypes.[2] Architects (left) resemble property engineers from Müller-Birn et al.'s *k*-means-based taxonomy [25], whereas Masons (right) resemble reference editors, item creators, item editors, item experts, and property editors.

ambiguities in the participant's responses to interview questions. After in-depth discussion of these ambiguities, the researchers agreed that P12 was more likely to be an Architect. The other disagreement arose from differences in how conservative the two researchers were in classifying the participants; one researcher leaned towards classifying P7 as a Mason, while the other researcher determined that P7 could not be confidently classified based on the information available. We treat P7 as a Mason in the rest of this paper, though we recognize that his archetype is the most ambiguous of all our participants. More importantly though, our informal classifications demonstrated to us that most of our participants could be reliably classified into one archetype or the other based on their interview transcripts.

Architects focus on developing the ontological infrastructure or "blueprints" of Wikidata. They primarily engage in modeling and proposing properties, so they help decide how information on Wikidata gets represented—e,g., what pieces of information are documented for a certain type of item. Masons, on the other hand, use Architects' blueprints to "lay the bricks," so to speak. They primarily perform data entry work, which directly and immediately creates new information on Wikidata—e.g., adding or editing items, references, and properties, or linking to external databases. Our data suggest that Architects and Masons differ on three main characteristics related to their level of desired cognitive engagement, their preference for collaboration or solitary work, and their degree of domain expertise, as we will now describe.

5.1.1 Cognitive Engagement. Wikidata relies on cognitively engaged editors to make important decisions about how information gets represented on the platform. P5, for example, spends a lot of time ensuring Wikidata's ontology can *"make sense across more cultures"* by proposing the elimination of properties—such as aunt and uncle for biographical items—that might unnecessarily introduce intercultural ambiguity. Architects are drawn to this kind of editing work because of the

[2]Icons sourced from https://www.flaticon.com/free-icon/engineer_2942433 and https://www.flaticon.com/free-icon/worker_2942457

mental stimulation it provides. P3, for example, is a recent retiree who used to *"get a real kick out of explaining things"* in her career, and now uses Wikidata as a way to *"keep [her] brain engaged."*

Masons, on the other hand, are *"happy to just plug information in rather than get involved at other levels"* (P10). As a father with limited time and energy, one of the main advantages of Wikidata for P10 is that it *"is not quite as taxing on the creative side of the brain*. Similarly, P13 often keeps a Wikidata tab open while at work because *"it'll be a nice repetitive task to do"* if he gets bored. While this usually relegates Masons to data entry tasks, some do work on proposing properties that *"are somewhat easier and less complicated"* (P7) as this still satisfies their requirement for low cognitive burden.

5.1.2 Desire for Collaboration. Architects enjoy the collaborative aspects of Wikidata. P3, for example, explained that retirement has caused her to miss working with others *"together to solve a problem"* and that Wikidata provides her with a space for doing so. After developing editing guidelines for a type of item, she often communicates them publicly so other editors have a chance to either use her suggestions or disagree with her. Communication, persuasion, and consensus-seeking are commonly involved in the process of developing Wikidata's information-modeling infrastructure, further placing modeling tasks in the Architects' wheelhouse.

Masons might be pushed in the other direction—towards data entry tasks—because they are averse to the cognitive and emotional burden of having to interact with others on the platform. P13, for example, was drawn to Wikidata from Wikipedia partly because he found interactions with Wikipedians *"just a little too intense sometimes."* Similarly, P10 is *"happy to just tinker in my own little corner"* rather than get involved in the Wikidata community.

5.1.3 Domain Expertise. Architects tend to stick to areas in which they have deep knowledge. P3, for example, works on building out areas in which she has *"expertise and maybe not too many other Wikidatans have expertise."* This aligns with Architects' focus on modeling tasks, as deciding what information is important to capture about a type of item often requires deep knowledge of the domain. P12, for example, primarily works on the area in which he has an advanced degree because he is *"not sure I have the expertise to say this is how we model the information"* in other areas.

This is a strong contrast from P13 who described his editing pattern as *"slightly aimless"* and *"kind of all over."* Like many Masons, P13 adds information to Wikidata based on things he comes across in his daily life—e.g., biographical info for guests on a podcast he listens to. We noticed this pattern in several of our participants; in the days following their interviews, at least 3 of them edited items that had come up in their conversations with us.

RQ1 results summary. **We answer RQ1 by describing two Wikidata editor archetypes—Architects and Masons. Architects resemble the property engineer role from [25], in that they mostly work on developing Wikidata's ontological infrastructure by creating new properties. They prefer work that is cognitively engaging and collaborative, and they tend to specialize in areas in which they are experts. The Mason archetype encompasses the remaining 5 roles described in [25]. Members of this archetype work on adding new data within the defined constraints imposed by Architects' work on properties. They prefer work that has a low cognitive burden and is solitary, and they typically do not stick to a particular domain.**

5.2 Contributor Motivations

RQ2: *What editor motivations are fulfilled by the structured nature of Wikidata?*

In 5.1, we discussed how personal characteristics relate to editing patterns by using the archetypes of Architects and Masons. We now apply this framework to explore how members of these archetypes share some motivations for contributing, but differ in other aspects of their motivations.

Many of our participants had also contributed to other peer production platforms, so there was significant overlap between their motivations and those described in previous literature on peer production platforms. These included: entertainment [15], *"some people do Dungeons and Dragons, but I like to do Wikidata"* (P15); altruism [30], *"it seems kind of selfish to have all those books [...] and not share them with anybody"* (P3); and the desire to spread knowledge in a particular domain [16], *"general knowledge about the things I'm interested in will continue to grow and spread, and I'm hoping that's one of the results of the work I do here"* (P3). The desire to share domain-specific knowledge is more characteristic of Architects, as they are often experts who are highly interested in a particular area.

We also discovered novel contributor motivations related to Wikidata's status as a SDPP platform. Architects and Masons share an intrinsic preference for structured data, but Masons additionally expressed appreciation for how structured data creates more possibilities for a wide variety of scaffolded tasks.

5.2.1 Structure as Motivator. Several participants—Architects and Masons alike—expressed an innate preference for working with structured data. In some cases, dealing with defined properties and semantic relationships—as opposed to, e.g., natural language in the case of Wikipedia—is attractive because it fits well with their perceived cognitive strengths. For example, Wikidata *"feels a lot more like how I think than Wikipedia"* (P8) or *"it's just how my brain works"* (P10). In other cases, participants described an innate drive for categorization and organization, e.g., *"liking order"* (P10) or being *"very interested in classification and categorization of different entities."*

Participants cited structure as a key differentiator in their experiences of editing Wikidata and Wikipedia. P8, for example, appreciates that *"there's less issues of article length and details of the pros and any of the other niggling areas in which you can get bogged down in editing Wikipedia."* P8 also described how editing Wikidata could be more accommodating of various editing styles:

> *I don't have anything against quality standards and style guides and things like that, but I think it's certainly a lot easier in a linked knowledge network like Wikidata than other projects might be, because if somebody edits in a particularly strange and quirky way, as long as they are contributing real structure, you know that that represents what exists for that topic in the world.*

P7 believes Wikidata's structure helps the community be *"a lot more lax than Wikipedia"* specifically because its structured nature means *"there's a lot less potential to do hard-to-revert damage"* to information on the platform. This makes Wikidata a particularly good fit for Masons like him, who dislike friction with other members of the community.

5.2.2 Scaffolded Tasks Seem More Accomplishable to Masons. In addition to their intrinsic preference for working with structured data, Masons in particular expressed strong appreciation for the way structured data facilitates the scaffolding of tasks. For P15, editing Wikipedia felt like *"staring at a blank screen"*, in contrast with editing Wikidata. *"Trying to figure out that this is what to add, and how to phrase it, and then pushing that button to say publish—it's more anxiety-producing for me than working in Wikidata"* (P15). On Wikidata, Architects handle the more ambiguous work that involves making decisions about how to represent and structure data, which Masons can then simply follow. For example, P3 spent *"a lot of time trying to figure out what the right metadata should be"* for representing tapestries. Once this work is done, however, it is quite straightforward for another editor to come along and enter information for a bunch of tapestries within the guidelines

and boundaries set by Architects. The scaffolding provided by Architects reduces ambiguity, thereby allowing Masons to contribute in ways that are meaningful while also seeming accomplishable despite their limited time and energy. P13 explained the central role this plays in his motivation to be a Wikidata contributor:

> *I have anxiety and depression, and I hate medications, and I'm high functioning. But at the same time there are books that I want to read, and I'm just anxious and I beat myself up for not being motivated to start a video game and finish it, or start a book and then finish it. But with tasks on Wikidata, they're all very accomplishable. I can do one while I'm watching TV with my girlfriend. It's a task that makes me feel like I'm doing something good.*

Other peer production platforms such as Wikipedia do also have structured tasks that can be completed—adding proper punctuation to articles, for example. The structured nature of Wikidata's data, however, allows for a wider variety of scaffolded tasks that align with editors' specific interests or goals. P13, for example, spends a lot of time adding Wikidata items for females and minorities with the hope of closing the gender gap on Wikipedia. He explained that adding these items to Wikidata causes them to show up on Google's Knowledge Graph, which some Wikipedia contributors use to find people for whom they should write articles. If the information gets *"picked up by other people,"* in this way, it might eventually be compiled into Wikipedia articles, and he can perhaps *"chip away"* at the problem through tasks that are well-defined. Due to the structured nature of Wikidata, P13 is able to make others aware of the need for new Wikipedia articles without having to deal with the ambiguities involved in actually writing those articles himself.

RQ2 results summary. **Architects and Masons share an intrinsic motivation for working with structured data because (1) they perceive themselves as being more cognitively suited to working with structured data as opposed to, e.g., natural language, and (2) they like order and have an innate preference for work that involves organization and categorization. Masons are additionally motivated by structured data's extrinsic quality of making many tasks less ambiguous and therefore more easily accomplishable.**

5.3 Perceived Impact of Contributions

RQ3: *How are editors motivated (or demotivated) by the perceived usage of their contributions?*

The desire to have impact through their contributions was common to many of our participants. P10 wouldn't be on Wikidata if he didn't think there was *"any usefulness component to it."* P1 wants to work on things that are *"profound and useful in the long term for mankind."* P8 is motivated to increase Wikidata's coverage of certain topics so it can be *"more useful"* to others. Our participants had many speculative guesses and hopes about how their Wikidata contributions might be used. These included: research, *"the Wikimedia community uses this, like researchers"* (P7); translation, *"suddenly the basic information is there regardless of your language"* (P6); historical preservation, *"I've been conceptualizing it as a way to record history"* (P10); and library cataloging, *"I wouldn't spend as much time on authority control if I didn't think it was going to be used by actual libraries"* (P7).

And yet, despite their ability to express ways in which their contributions might hypothetically be useful, most of our participants—11 of the 15—did not have a concrete understanding of how their contributions are actually used, or by whom. *"Frankly, I just don't know"* was P12's answer, when asked about his audience. *"It could be no one at all."* Similar answers were quite common—*"I don't know"* (P14), *"I don't really know"* (P2), *"I really don't know who looks at them"* (P9), *"I have no idea who they are"* (P3). This represents a stark contrast from what we might expect the norm

to be on platforms like Wikipedia, for which use cases and audiences are likely not difficult for contributors to enumerate. This raises a serious question for Wikidata editors—in contributing to the platform, are they basically just *"pumping the information into an empty void"* (P10)?

5.3.1 Different Attitudes Toward Understanding the Usefulness of Contributions. For some editors, not knowing how their contributions are used is not a major concern. P3 said knowing her audience would make no difference to her as she is *"mostly doing this for myself."* P2 described it as *"beautiful"* that he doesn't yet understand the ways in which his work will be used. P1 compared his Wikidata contributions to some scientific endeavors in that *"we just explore things"* and hope people will find the work useful sometime in the distant future.

In some cases, understanding how their contributions are used could actually be a demotivating factor. As P5 states, *"it can make people hesitate to contribute, when they know that the license is going to permit downstream for-profit reuse without giving them anything other than the credit."* P13 expressed this exact sentiment in recalling his *"struggle with the idea of creating value for a multi-trillion-dollar company"* through his previous contributions to Google Map Maker.

Many of our participants, however, require that their Wikidata work be useful, despite not having precise knowledge of how that usefulness might manifest. If he is just *"pumping the information into an empty void,"* P10 would not contribute to Wikidata at all, as he *"may as well just be filling out an Excel spreadsheet"* on his computer. P13 similarly gets *"a lot of satisfaction"* from knowing that his contributions might be used by others. Interestingly, we noticed no consistent difference between Architects and Masons with regards to the importance they place on the usefulness of their contributions.

However, as P10 explains, the machines that exist as intermediaries between Wikidata and end-users are difficult to identify—a fact that makes it less than straightforward to uncover the myriad ways in which end-users consume Wikidata. According to P10, any serious tool would download everything on Wikidata and query it locally to avoid latency. Google, for example, *"caches [Wikidata] because they have to respond in a second."* P2 echoed this in saying he would not use Wikidata's built-in query service because *"it's too much delay"* and is prone to time-outs. It is unclear how many end-user interactions with Wikidata information happen through what we have decided to name "invisible machine intermediaries", that operate in the way P2 and P10 recommend—by downloading and querying their own local copies of Wikidata. We delve into the ramifications of this in the latter part of the Discussion section.

RQ3 results summary. **Wikidata editors have many speculations about how their contributions *might* be used, but little concrete knowledge of how and by whom they are *actually* used. End-users often interact with Wikidata contributions through other software-based tools that use all or part of Wikidata as a database. These invisible machine intermediaries make it difficult for contributors to access information about how their contributions are used by others. This may pose a problem for some Wikidata editors who want to understand how their contributions are used and useful.**

6 DISCUSSION

We begin by discussing the Architect and Mason roles. We outline how they relate to and build upon previous findings about roles in online communities, and focus particularly on how our own findings can inform design of SDPP platforms in novel ways.

We then discuss usage opacity and the invisible machines—what causes the machines to arise in the context of Wikidata, and why their existence negatively impacts Wikidata in particular.

We offer multiple suggestions for immediate action that could reduce usage opacity, and thereby mitigate the damage done to Wikidata contributor motivation by the invisible machine reuse.

6.1 Architects & Masons

In this paper, we describe two Wikidata editor archetypes that differ along the dimensions of cognitive engagement, desire for collaboration, and domain expertise. Their characteristics are encapsulated by the metaphor of Masons and Architects; while Masons prefer working alone on "brick-laying" tasks with lower cognitive effort, Architects with domain expertise work closely with others in the community to develop the "blueprints" that define how data are represented on Wikidata.

Our findings fit with much of the prior literature on user characteristics in online communities. For example, Kozinets [22] identified the role of devotees in virtual consumption communities that maintain a strong interest and enthusiasm in activity, but few social attachments to the community, similar to Masons on Wikidata. Core members in OSS projects [46] guide and coordinate the development of the projects in a manner that is similar to the activities that Architects perform on Wikidata. However, there are nuances that differentiate our Wikidata archetypes from roles identified in other online communities. Perhaps most importantly, the two editor roles we identified are less likely to follow an onion ring model [13, 14, 19, 20], wherein users begin with simpler tasks and take on more central decision-making roles over time. Our results give us little reason to believe that Wikidatans progress from Masons to Architects as they gain experience; almost all our participants were highly experienced editors, and we saw no indication that the Architects started out by contributing in the ways Masons usually do.

Our results build on Müller-Birn et al.'s quantitatively derived roles based on editing patterns [25]. In conducting a qualitative analysis that relates these editing-pattern-based roles to editors' personal characteristics, we provide several novels ways in which SDPP platform designers can use knowledge about editor roles in their work. We now outline two such examples, involving algorithm design and newcomer engagement, respectively.

6.1.1 Different Algorithms for Architects & Masons. Zhu et al. [49] highlighted that designing algorithms first necessitates an understanding of the stakeholders affected by them. As editor characteristics vary between the two archetypes, algorithm design for Wikidata should acknowledge members of both archetypes as distinct stakeholder groups. Therefore, understanding how the motivations and other personal characteristics of each archetype relate to editing patterns allows us to design algorithms in ways that better fit with the preferences of their respective audiences. A task-routing system for property engineering tasks, for example, would be expected to primarily serve Architects, and would therefore recommend tasks within the editor's single domain of interest, whereas a task-routing system for all other types of work might provide recommendations based on features other than topic area—e.g., favoring items that have relatively sparse editing activity so as to accommodate Masons' aversion to interacting with other editors.

6.1.2 Classifying Editors Before They Start Editing. If we rely solely on editing patterns to classify editors into roles, we are limited to classifying only those editors who have already edited Wikidata in the past—and in sufficient quantities for us to algorithmically identify the clusters to which they belong. By tying Müller-Birn et al.'s 6 clusters to two sets of personal characteristics, we open up the possibility of predicting the eventual role an editor will occupy even before they have completed their first edit.

Upon account creation, new Wikidata editors could be presented with a quick survey about their motivations and the types of tasks they might enjoy working on—e.g., "Would you prefer working alone or working collaboratively with others?" The answers to these questions could

be used to determine where an editor falls along the dimensions of domain expertise, desire for collaboration, and cognitive engagement, and then present them with an orientation that focuses on either property engineering work (if they are an Architect) or more item-focused work (if they are a Mason), with the goal of better facilitating newcomer engagement. While such an instrument would have to be validated quantitatively, our qualitative insights provide guidance on the types of questions this kind of introductory survey should contain if its purpose is to predict eventual editing patterns and/or preferences.

6.2 Usage Opacity and the Invisible Machines

We now discuss usage opacity on Wikidata—the phenomenon we observed whereby most of our participants indicated having no knowledge of how or if their contributions were being used. Although some of our participants did not care about how their contributions would be used, there is a clear subset of editors who would benefit from having this knowledge. Furthermore, since most of our participants were experienced Wikidata contributors, our results are biased towards people for whom Wikidata currently provides sufficient motivation. Survivorship bias would therefore conceal any large contingent of would-be contributors who care so much about understanding the uses of their contributions that they decide not to contribute at all. Regardless, the most straightforward solution is to reduce usage opacity, so editors who *want* to know how their contributions are used can access that information.

We begin by re-iterating—with additional detail—what we learned from our participants about *why* Wikidata's usage in particular is so opaque, as this reveals several avenues by which we might make its usage more transparent. We then provide reasons why this opacity may be more consequential for Wikidata than for larger, more established platforms like Wikipedia, justifying the need for Wikidata to invest in reducing usage opacity. We then outline three recommendations for solutions that Wikidata system designers could immediately begin implementing. We conclude with thoughts on the potential unintended side-effects of increased transparency, especially when Wikidata contributions are used by for-profit third parties.

6.2.1 Rise of the Invisible Machines. Two of our participants gave us some insight into issues that cause invisible machine intermediaries to emerge on Wikidata. P2 mentioned that Wikidata's query service is slow and unstable, incentivizing tool developers to query their own local downloads of Wikidata's data. P10 similarly highlighted the advantages of querying local copies of Wikidata data to avoid latency causing queries to be dispersed across a bunch of distinct entities—"invisible machines"—each of which holds its own copies of data that originated from Wikidata. Because it is an SDPP platform that provides structured data for use in other software, it is unsurprising that Wikidata's data is used in this way.

However, this means the Wikimedia Foundation (WMF) does not have access to complete usage information for Wikidata, making it difficult to communicate meaningful usage statistics—such as how many times a particular item is viewed by end-users—to editors. The absence of such readily communicable metrics poses a problem for Wikidata editors who might be motivated by the usefulness of their contributions. Furthermore, given that our participants were selected from a pool of highly active Wikidata editors, the demotivating effect this has is likely to be even more pronounced in a sample that includes former contributors who might have otherwise been retained.

6.2.2 Why Usage Opacity is a Problem for Wikidata. Prior work indicates that signals between end-users and editors might be similarly weak on other peer production platforms such as Wikipedia. Warncke-Wang et al. [39], for example, find that editors focus more on topics that interest them than on topics that are most read by end-users. This would seem to indicate that peer production platform

editors seldom care about usage statistics, even when they have full access to them. This interpretation is belied by the existence of myriad pages on Wikipedia dedicated to documenting highly viewed pages (e.g., [1, 2]). While it *is* clear that usage statistics are not *sufficient* for communicating contribution usefulness to editors, they do have a clear role on Wikipedia and, we will now argue, would have an even stronger role on Wikidata for at least two reasons.

The first is size. In 2020, Wikipedia was the 5th most viewed website in the world, with approximately 6.1 billion monthly visitors [27]. Its community of 27 million active editors also dwarfs Wikidata's base of 23,000. We expect this difference in scale is a major driver of the difference in the impact of usage opacity between the two platforms. It is likely that most Wikipedia editors can identify people in their immediate social circles who use the online encyclopedia. Thus, even if they don't happen to pay attention to the specific use-cases or number of views their particular edits are getting, Wikipedia editors may have a general understanding that, by contributing to such a widely used project, they are highly likely to have some level of real impact on the information people consume.

There is also an important difference in how often end-users that interact with each platform do so *knowingly*. It would not be enough for Wikidata to be widely used; its editors would have to *understand* that it is widely used, which is more difficult when its data are mostly consumed through third parties that do not necessarily credit the platform. This is a contrast from Wikipedia, where most interactions with its data happen through the platform's GUI, which prominently displays Wikipedia's name and logo at the top.

These are likely not the only consequential differences between the two platforms, but they already make it clear that WMF must do more work to highlight the usefulness of Wikidata contributions in ways that may be less necessary for Wikipedia. We therefore present three recommendations for immediate solutions that Wikidata system designers could implement to reduce usage opacity.

- **Focus on the Wikimedia Context.**
  One major goal of Wikidata is to serve as a centralized knowledge base for structured data that is used by all other Wikimedia projects. Connecting the usage statistics from these other projects to Wikidata would perhaps provide substantial motivation for contributors, who might be encouraged by the uses of their contributions on other platforms. Wikidata system designers might choose to transparently display, for example, the number of times a particular item property has been shown in a Wikipedia article viewed by a user.
- **Improve Wikidata's Infrastructure.**
  As mentioned previously, many of the machines that use Wikidata do so by downloading local copies that are queried off-platform, despite the existence of a Wikidata query service based on SPARQL. One of our participants mentioned that Wikidata's built-in query service was too slow and unstable to use reliably in other tools. Wikidata system designers could encourage more use of their built-in query service by improving its performance. This would allow them direct access to a larger portion of the usage analytics, which could then be communicated to contributors.
- **Cooperate with the Machines.**
  One way to figure out who the machines are is to simply ask them. Restricting and tracking third-party usage of Wikidata's data is at odds with the community's desire for a fully open platform, but many who re-use Wikidata might already collect, and be willing to provide, end-user usage analytics. Implementing an optional registration and analytics reporting process for third party developers that use Wikidata would allow the platform to reveal the machines, while also providing a way in which those machines could give back to the

community of contributors and help motivate the very people who produce the data they use.

While full usage statistics are ideal, most third party developers are unlikely to provide them, so any implementation should also allow developers to provide smaller pieces of helpful information—e.g., a link to the tool for which data are being downloaded, which could be used to compile a list of third party tools that are known to use Wikidata. As certain tools use particular types of data, this could allow Wikidata to give contributors insight into the types of contributions that are most valuable to third parties.

6.2.3 Aversion to Volunteering for Corporations. Two of our participants mentioned the aversion many Wikidata editors feel towards the idea of volunteering their time in a way that creates profit for large businesses, e.g., Google. Revealing the machines and increasing usage transparency can therefore function as a demotivating factor for some editors, if the machines are run by for-profit corporations. One simple solution to this would be for Wikidata to only highlight reuse of Wikidata by non-profit third parties. For obvious ethical reasons, we do not recommend this approach. We instead recommend that future work explore potential solutions to this tension between community values and the ways in which Wikidata is used.

7 CONCLUSION

In this paper, we investigated the personal characteristics of Wikidata editors that lead them to take on different kinds of tasks. We conducted 15 semi-structured interviews with Wikidata editors and performed inductive thematic analysis to understand their characteristics, motivations and perception of contribution reuse. We discovered that the possession of domain expertise along with the preference of collaboration lead the editors to take on more challenging and complicated Architect tasks like data modelling. In the meantime, Masons focus on data entry tasks with low cognitive burden (RQ1). While Architects and Masons share similar intrinsic motivations, Masons are motivated by the structured nature of Wikidata (RQ2). We also discovered that Wikidata editors, both Architects and Masons, do not have a concrete knowledge of how their contributions are being used due to invisible machine intermediaries being the primary user of Wikidata (RQ3). Design implications were discussed for different kinds of editors and raised the possibility of classifying an editor before the first edit. Finally, we discussed the usage of Wikidata, why it is so opaque, and offered suggestions to reveal this 'invisible machine'. As for the future work, there are two major directions. First, our work represents an initial step of identifying motivations specific to the SDPP settings. Future work could build tools and systems based on our findings or explore other non-editor archetypes in order to unfold the community structure of SDPP. Second, our work revealed the invisible machine intermediaries that restrained Wikidata editors from knowing the usage of their contributions. We pointed out three solutions that could directly mitigate this issue. Researchers and system developers could make use of our findings to reveal the invisible machines.

ACKNOWLEDGMENTS

We would like to thank the reviewers for their thoughtful criticisms, which helped us enormously in improving on earlier drafts of this work. We would also like to thank our participants, with whom we shared many fun and deeply enlightening conversations. This work was supported by the National Science Foundation (NSF) under Award No. IIS-1816348.

REFERENCES

[1] 2021. Wikipedia:Multiyear ranking of most viewed pages. https://en.wikipedia.org/w/index.php?title=Wikipedia:Multiyear_ranking_of_most_viewed_pages&oldid=1027228532 Page Version ID: 1027228532.

[2] 2021. Wikipedia:Popular pages. https://en.wikipedia.org/w/index.php?title=Wikipedia:Popular_pages&oldid=1015389995 Page Version ID: 1015389995.
[3] Ofer Arazy, Felipe Ortega, Oded Nov, Lisa Yeo, and Adam Balila. 2015. Functional roles and career paths in Wikipedia. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. 1092–1105.
[4] Saeid Asadi, Shadi Ghafghazi, and Hamid R Jamali. 2013. Motivating and discouraging factors for Wikipedians: the case study of Persian Wikipedia. *Library Review* (2013).
[5] Yochai Benkler. 2016. Peer production and cooperation. In *Handbook on the Economics of the Internet*. Edward Elgar Publishing.
[6] Magnus Bergquist and Jan Ljungberg. 2001. The power of gifts: organizing social relationships in open source communities. *Information Systems Journal* 11, 4 (2001), 305–320.
[7] Anne Bowser, Derek Hansen, Jennifer Preece, Yurong He, Carol Boston, and Jen Hammock. 2014. Gamifying citizen science: a study of two user groups. In *Proceedings of the companion publication of the 17th ACM conference on Computer supported cooperative work & social computing*. 137–140.
[8] Virginia Braun and Victoria Clarke. 2012. Thematic analysis. (2012).
[9] Susan L Bryant, Andrea Forte, and Amy Bruckman. 2005. Becoming Wikipedian: transformation of participation in a collaborative online encyclopedia. In *Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work*. 1–10.
[10] Kelly Caine. 2016. Local standards for sample size at CHI. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 981–992.
[11] Langtao Chen, Aaron Baird, and Detmar Straub. 2019. Why do participants continue to contribute? Evaluation of usefulness voting and commenting motivational affordances within an online knowledge community. *Decision Support Systems* 118 (2019), 21–32.
[12] Cathy C Conrad and Krista G Hilchey. 2011. A review of citizen science and community-based environmental monitoring: issues and opportunities. *Environmental monitoring and assessment* 176, 1 (2011), 273–291.
[13] Kevin Crowston and James Howison. 2003. The social structure of open source software development teams. (2003).
[14] Kevin Crowston, Kangning Wei, James Howison, and Andrea Wiggins. 2008. Free/Libre open-source software development: What we know and what we do not know. *ACM Computing Surveys (CSUR)* 44, 2 (2008), 1–35.
[15] Utpal M. Dholakia, Richard P. Bagozzi, and Lisa Klein Pearo. 2004. A social influence model of consumer participation in network- and small-group-based virtual communities. *International Journal of Research in Marketing* 21, 3 (Sept. 2004), 241–263. https://doi.org/10.1016/j.ijresmar.2003.12.004
[16] Nuša Farič and Henry WW Potts. 2014. Motivations for Contributing to Health-Related Articles on Wikipedia: An Interview Study. *Journal of Medical Internet Research* 16, 12 (2014), e260. https://doi.org/10.2196/jmir.3569 Company: Journal of Medical Internet Research Distributor: Journal of Medical Internet Research Institution: Journal of Medical Internet Research Label: Journal of Medical Internet Research Publisher: JMIR Publications Inc., Toronto, Canada.
[17] Marylène Gagné and Edward L Deci. 2005. Self-determination theory and work motivation. *Journal of Organizational behavior* 26, 4 (2005), 331–362.
[18] Andreea D Gorbatai. 2014. The paradox of novice contributions to collective production: Evidence from Wikipedia. *Available at SSRN 1949327* (2014).
[19] Chris Jensen and Walt Scacchi. 2007. Role migration and advancement processes in OSSD projects: A comparative case study. In *29th International Conference on Software Engineering (ICSE'07)*. IEEE, 364–374.
[20] Corey Jergensen, Anita Sarma, and Patrick Wagstrom. 2011. The onion patch: migration in open source ecosystems. In *Proceedings of the 19th ACM SIGSOFT symposium and the 13th European conference on Foundations of software engineering*. 70–80.
[21] Leigh Jin, Daniel Robey, and Marie-Claude Boudreau. 2007. Beyond development: a research agenda for investigating open source software user communities. *Information Resources Management Journal (IRMJ)* 20, 1 (2007), 68–80.
[22] Robert V Kozinets. 1999. E-tribalized marketing?: The strategic implications of virtual communities of consumption. *European management journal* 17, 3 (1999), 252–264.
[23] Karim R Lakhani and Eric Von Hippel. 2004. How open source software works:"free" user-to-user assistance. In *Produktentwicklung mit virtuellen Communities*. Springer, 303–339.
[24] Karim R Lakhani and Robert G Wolf. 2003. Why hackers do what they do: Understanding motivation and effort in free/open source software projects. (2003).
[25] 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*. 1–10.
[26] Kumiyo Nakakoji, Yasuhiro Yamamoto, Yoshiyuki Nishinaka, Kouichi Kishida, and Yunwen Ye. 2002. Evolution patterns of open-source software systems and communities. In *Proceedings of the international workshop on Principles of software evolution*. 76–85.

[27] Dorothy Neufeld. 2021. The 50 Most Visited Websites in the World. https://www.visualcapitalist.com/the-50-most-visited-websites-in-the-world/
[28] Oded Nov. 2007. What motivates wikipedians? *Commun. ACM* 50, 11 (2007), 60–64.
[29] Oded Nov, Ofer Arazy, and David Anderson. 2011. Dusting for science: motivation and participation of digital citizen science volunteers. In *Proceedings of the 2011 iConference*. 68–74.
[30] Shaul Oreg and Oded Nov. 2008. Exploring motivations for contributing to open source initiatives: The roles of contribution context and personal values. *Computers in human behavior* 24, 5 (2008), 2055–2073.
[31] Alessandro Piscopo and Elena Simperl. 2018. Who Models the World? Collaborative Ontology Creation and User Roles in Wikidata. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–18.
[32] Dana Rotman, Jenny Preece, Jen Hammock, Kezee Procita, Derek Hansen, Cynthia Parr, Darcy Lewis, and David Jacobs. 2012. Dynamic changes in motivation in collaborative citizen-science projects. In *Proceedings of the ACM 2012 conference on computer supported cooperative work*. 217–226.
[33] Scholia. 2020. Homepage. https://scholia.toolforge.org/
[34] Pankaj Setia, Balaji Rajagopalan, Vallabh Sambamurthy, and Roger Calantone. 2012. How peripheral developers contribute to open-source software development. *Information Systems Research* 23, 1 (2012), 144–163.
[35] C Estelle Smith, Bowen Yu, Anjali Srivastava, Aaron Halfaker, Loren Terveen, and Haiyi Zhu. 2020. Keeping Community in the Loop: Understanding Wikipedia Stakeholder Values for Machine Learning-Based Systems. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.
[36] Monic Sun and Feng Zhu. 2013. Ad revenue and content commercialization: Evidence from blogs. *Management Science* 59, 10 (2013), 2314–2331.
[37] Bianca Trinkenreich, Mariam Guizani, Igor Wiese, Anita Sarma, and Igor Steinmacher. 2020. Hidden Figures: Roles and Pathways of Successful OSS Contributors. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 1–22.
[38] Denny Vrandečić and Markus Krötzsch. 2014. Wikidata: a free collaborative knowledgebase. *Commun. ACM* 57, 10 (2014), 78–85.
[39] Morten Warncke-Wang, Vivek Ranjan, Loren Terveen, and Brent Hecht. 2015. Misalignment Between Supply and Demand of Quality Content in Peer Production Communities. *Proceedings of the International AAAI Conference on Web and Social Media* 9, 1 (April 2015). https://ojs.aaai.org/index.php/ICWSM/article/view/14631 Number: 1.
[40] Howard T Welser, Dan Cosley, Gueorgi Kossinets, Austin Lin, Fedor Dokshin, Geri Gay, and Marc Smith. 2011. Finding social roles in Wikipedia. In *Proceedings of the 2011 iConference*. 122–129.
[41] Wikidata. 2020. Douglas Adams (Q42). https://www.wikidata.org/wiki/Q42
[42] Wikidata. 2020. Wikidata: Introduction. https://www.wikidata.org/wiki/Wikidata:Introduction/en-gb
[43] Wikidata. 2020. Wikidata:Contribute/Suggested and open tasks. https://www.wikidata.org/wiki/Wikidata:Contribute/Suggested_and_open_tasks
[44] Dennis M Wilkinson. 2008. Strong regularities in online peer production. In *Proceedings of the 9th ACM conference on Electronic commerce*. 302–309.
[45] Heng-Li Yang and Cheng-Yu Lai. 2010. Motivations of Wikipedia content contributors. *Computers in human behavior* 26, 6 (2010), 1377–1383.
[46] Yunwen Ye and Kouichi Kishida. 2003. Toward an understanding of the motivation of open source software developers. In *25th International Conference on Software Engineering, 2003. Proceedings.* IEEE, 419–429.
[47] Xiaoquan Zhang and Feng Zhu. 2006. Intrinsic motivation of open content contributors: The case of Wikipedia. In *Workshop on Information Systems and Economics*, Vol. 10. Evanston, IL, 4.
[48] Haiyi Zhu, Robert E Kraut, Yi-Chia Wang, and Aniket Kittur. 2011. Identifying shared leadership in Wikipedia. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 3431–3434.
[49] Haiyi Zhu, Bowen Yu, Aaron Halfaker, and Loren Terveen. 2018. Value-sensitive algorithm design: Method, case study, and lessons. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–23.
[50] Haiyi Zhu, Amy Zhang, Jiping He, Robert E Kraut, and Aniket Kittur. 2013. Effects of peer feedback on contribution: a field experiment in Wikipedia. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2253–2262.

A APPENDIX: INTERVIEW QUESTIONS

A.1 General Questions about Work Process

1. Could you tell me about one of your recent Wikidata edits? (What page did you edit? What property did you update?)
2. Why do you want to edit this page?

3. Do you have any general preference in editing certain types or topics of items?

A.2 Motivation Related Questions

1. How and when you got started editing Wikidata?
2. Did it go well? Did you face any difficulties or obstacles?
3. Does editing Wikidata or Wikipedia have any relationship to your professional career?
4. If not, what motivates you to edit Wikidata? (What else motivates you to edit Wikidata?.)

A.3 Questions about Perception of Contribution Reuse

1. When you choose a topic to edit, how do you assess or estimate its value to the Wikimedia community?

2. Do you know of specific ways that your contribution will be used or applied outside the Wikimedia community? Or in simple words, who do you think your audience is?

3. If you do not know who your audience is and how are people using your contributions, how do you decide whether your contribution is interesting to other people?

Received January 2021; revised July 2021; accepted November 2021