Collective Intelligence in Humanitarian Voluntary Geographic Information

The Case of the HOT Tasking Manager

DAGOBERTO JOSÉ HERRERA-MURILLO, Universidad de Zaragoza, Spain
HÉCTOR OCHOA-ORTIZ and UMAIR AHMED, Università degli Studi di Camerino, Italy
FRANCISCO JAVIER LÓPEZ-PELLICER, Universidad de Zaragoza, Spain
BARBARA RE and ANDREA POLINI, Università degli Studi di Camerino, Italy
JAVIER NOGUERAS-ISO, Universidad de Zaragoza, Spain

Voluntary Geographic Information initiatives are transforming the disaster response landscape. Our research provides insights into how the concept of collective intelligence is accomplished in humanitarian mapping initiatives. The main source originates from the data obtained in 746 mapping projects organised by the Humanitarian OpenStreetMap Team between December 2021 and November 2023, where 38,893 contributors completed 312,289 mapping tasks. These data include detailed attributes of the contributors and the states the tasks go through. The methodology adopts a quantitative approach, including descriptive and inferential statistics, and standard process mining techniques. Our results indicate that, in general terms, in humanitarian mapping, a group of contributors from outside the area of interest perform straightforward mapping tasks with limited collaboration among them. The "wisdom" of advanced contributors is the cornerstone that sustains the system. The discussion section elaborates on (1) how these findings suggest that humanitarian mapping projects effectively meet their short-term mapping objectives but fall short if more sustainable mapping objectives are sought and (2) possible strategies for better harnessing the collective intelligence of these efforts.

CCS Concepts: • **Human-centered computing** → **Collaborative and social computing systems and tools**; **Collaborative interaction**; **Interactive systems and tools**; • **Applied computing** → Cartography; • **Information systems** → Geographic information systems.

Additional Key Words and Phrases: Collective intelligence, Crowdsourcing, Human Computer Interaction, User Interface, Volunteered Geographic Information, OpenStreetMap

ACM Reference Format:
Dagoberto José Herrera-Murillo, Héctor Ochoa-Ortiz, Umair Ahmed, Francisco Javier López-Pellicer, Barbara Re, Andrea Polini, and Javier Nogueras-Iso. 2025. Collective Intelligence in Humanitarian Voluntary Geographic Information: The Case of the HOT Tasking Manager. 1, 1 (March 2025), 42 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Intelligence is not an exclusive property of individuals. It also arises in groups of individuals such as families, nations, companies or other human, non-human and hybrid conglomerates. Numerous proposals have been developed in recent

Authors' addresses: Dagoberto José Herrera-Murillo, dherrera@unizar.es, Universidad de Zaragoza, Zaragoza, Spain; Héctor Ochoa-Ortiz, hector.ochoaortiz@unicam.it; Umair Ahmed, umair.ahmed@unicam.it, Università degli Studi di Camerino, Camerino, MC, Italy; Francisco Javier López-Pellicer, fjlopez@unizar.es, Universidad de Zaragoza, Zaragoza, Spain; Barbara Re, barbara.re@unicam.it; Andrea Polini, andrea.polini@unicam.it, Università degli Studi di Camerino, Camerino, MC, Italy; Javier Nogueras-Iso, jnog@unizar.es, Universidad de Zaragoza, Zaragoza, Spain.

decades to define the concept of collective intelligence [32]. In general, these definitions converge on the idea of an emerging capacity within groups to solve problems, make decisions or achieve results more effectively than individuals working alone. The concept of collective intelligence finds applications in various fields, such as biology, sociology, political science and economics [32]. However, its influence is particularly noticeable in computer science, especially in the fields of human-computer interaction and computer-assisted cooperative work [12, 32].

The popularisation of the Internet, and more recently artificial intelligence, has taken the idea of collective intelligence to a new level where interconnected groups of people and computers collectively do intelligent things in multiple domains. This is the case of the production model known as crowdsourcing, in which the work traditionally performed by a designated agent is outsourced to a large, undefined group of people, usually in the form of an open call [24]. Examples of crowdsourcing include Wikipedia and open-source software development, where the collective contributions of numerous individuals result in robust, high-quality products [5]. This research focuses on crowdsourced geographic information, specifically the phenomenon of Volunteered Geographical Information (VGI). VGI describes the efforts of individuals and communities to address digital geographic information gaps, often arising due to the absence of comparable commercial platforms [14].

OpenStreetMap (OSM), often referred to as "the Wikipedia of Maps," exemplifies the potential of VGI by providing a free and open geographic platform maintained by a global community of volunteer contributors [16]. This platform brings together approximately 10.5 million registered members who have collectively contributed to creating 9.1 billion elements in the database [41]. The OSM community has naturally attracted researchers interested in crowdsourcing and collective intelligence [25, 47, 55].

In recent years, humanitarian mapping missions have become one of the major focus areas of VGI initiatives, including the OSM community [21]. This type of mapping is transforming the disaster response landscape in situations where conventional geographical data sources are inaccessible or outdated [14, 39, 43]. Humanitarian mapping differs substantially from other forms of voluntary mapping in aspects such as its purpose, geographic areas of impact, mapping interfaces, editing and validation dynamics, and resulting footprint [21]. For instance, while overall mapping in OSM tends to be heavily concentrated in regions with a very high Human Development Index, humanitarian mapping shows a distinct pattern, primarily targeting regions with medium and low human development. Furthermore, the study of the effects that micro-tasking introduces into the dynamics of OSM peer production is a promising field for research [60]. Despite the growing role of humanitarian mapping in disaster response and its particularities, the mechanisms underlying its success remain poorly understood. Questions persist about how these projects harness collective intelligence to achieve their goals, and what lessons they offer for broader applications in peer-production systems.

Herrera-Murillo et al. [23] proposed an initial analysis of the humanitarian mapping process. This paper expands on that work by analysing humanitarian mapping projects as a collective intelligence system. To achieve this, we drew on two key approaches. First, we used the collective intelligence framework proposed by Malone et al. [32] to formulate relevant research questions that shed light on the dynamics of the humanitarian mapping system. Second, we leveraged the extensive dataset provided by the Humanitarian OpenStreetMap Team (HOT)[1] —the leading humanitarian VGI initiative operating on the OSM platform— to gather evidence to answer these research questions. More specifically, these data relate to the technological tool used for the coordination of humanitarian mapping, called the HOT Tasking Manager (HOT-TM).[2]

[1] https://www.hotosm.org/
[2] https://tasks.hotosm.org/

Malone et al. [32] argue that new information technologies have fostered novel forms of collective intelligence, leading to the emergence of a distinct field of study. To address this evolution, they propose a definition of collective intelligence that encompasses these new realities while remaining consistent with previous interpretations of the concept. In their revised definition, collective intelligence is characterised by three key elements: "(1) groups of individuals (2) acting collectively (3) in ways that seem intelligent". Each of these three components is described below along with the research questions derived from them.

From the perspective of new collective intelligence systems, a **group of individuals** can consist of both human and computational agents [31, 33]. While groups consisting of a single person and a computer are at the periphery of collective intelligence, the core lies in the combination of multiple individuals and computer systems working together to collectively address problems.

Decades of research on human groups reveal that individual characteristics significantly shape group dynamics, influencing both team performance and outcomes [3]. Effectively managing group composition, the configuration of the attributes of its members, involves identifying and prioritising key attributes. When a particular attribute becomes salient, it can be a determining factor in the structure and interactions of the group [37].

According to Jiao et al. [26], the success of crowdsourcing is closely related to individual attributes. In this field, the experience of the group members is often the most highly regarded attribute. Evidence consistently shows that members with knowledge and experience in a specific domain produce higher quality results than non-experts on a wide range of tasks [19]. This evidence reinforces the significant emphasis placed on estimating, identifying, and managing the expertise of contributors to enhance task allocation in crowdsourcing settings [4, 10, 53].

As far as the non-human component of the groups is concerned, Malone [31] recognises that computer agents can participate in a group in various ways, depending on their relationship with human agents. The most common role is that of a tool in which computers serve to enhance human capabilities. Like other tools, a computer requires direct instructions from a human to perform tasks. A step further is the role of assistant. Unlike a tool, an assistant has more autonomy and can take the initiative to help humans achieve their goals. Further up the hierarchy, computers can act as peers, demonstrating an autonomy comparable to that of human group members. Finally, computers can take on managerial roles. In this capacity, they perform tasks such as determining the sequence of tasks needed, predicting which contributor is best suited for each task, automatically assigning tasks to appropriate contributors, and evaluating their work.

Given the importance of group characterisation in peer production environments, we formulate the following research question:

RQ1: What characterises the group of individuals participating in HOT-TM mapping projects?

By **acting collectively**, collective intelligence implies that the behaviour of the group is characterised by the relationships between the activities of its members. This does not mean that all members share identical objectives or cooperate at all times but emphasises the presence of interdependencies between their activities. According to Suran et al. [49], the processes of collective action can be analysed and categorised based on two key dimensions: the type of activities (creating or deciding) and the type of interactions (independent or dependent). In creating activities, contributors produce something new (e.g. ideas or artifacts), while in deciding activities, contributors evaluate and select alternatives. To fulfil their missions, organisations usually need both creation and decision capabilities. Creation capabilities require decision capabilities to select the best outcomes, and decision capabilities need creation capabilities to provide evaluation options. Since both activities can be carried out by individual actors as well as by groups, they can also be considered as dependent or independent interactions. The intersection of these two dimensions, types of activities and interactions, gives rise to four combinations:

1. Collection (Create + Independent): In these activities, individuals contribute independently, each offering unique inputs to the system based on their work.

2. Collaboration (Create + Dependent): These activities involve multiple individuals working to generate interrelated or interdependent solutions.

3. Individual Decision (Decide + Independent): Decisions are made by individuals acting independently, with outcomes that may vary from person to person. Sometimes these decisions may be influenced by information shared by others.

4. Group Decision (Decide + Dependent): Decisions are made collectively by a group, resulting in a consensus that affects the group as a whole. Important variants include voting, consensus, averaging, and prediction markets.

Understanding and describing these processes, along with identifying the dominant dimensions of collective action, are useful for effectively managing crowdsourcing efforts and designing more efficient workflows. McDonald [34] outlines an itinerary for monitoring a range of key aspects in crowdsourcing initiatives. This itinerary begins with examining the type, time and effort involved in micro-tasks and extends to analysing the interaction —both positive and negative— between the actions of the involved actors. Classic concepts within the field of computer-assisted cooperative work such as articulation work, user roles and division of labour also contribute to enriching the debate on collective action [44, 48].

This context leads us to the following research question:

RQ2: How is collective action characterised in HOT-TM mapping projects?

By using the term **seem intelligent**, the definition acknowledges that what is considered intelligent can vary depending on the perspective of the observer [32]. Characterising something as intelligent can be challenging due to the multifaceted nature of intelligence. We can detect intelligence in a system by identifying cognitive processes such as reasoning, consciousness, planning, abstraction, and learning. Alternatively, we can observe typical outcomes of intelligence, such as adaptive behaviour, problem-solving, and artifact creation, which relate to goals and interactions with the environment [32]. According to Riedl et al. [42], collective intelligence has the potential to predict group performance across a wide range of tasks.

Collective intelligence is not by default a universal property of collaborative groups [58]. In this regard, it is important to highlight what differentiates intelligent action from mere collective action. While intelligence cannot always be directly or linearly related to success in task performance, the concept of collective intelligence suggests that groups, under certain conditions, can outperform individuals working alone. This idea is closely aligned with the notion of the “wisdom of crowds” [50], an emergent property in which groups of individuals may be smarter than the smartest individuals within those groups. However, collective action alone does not guarantee the emergence of superior group performance, a highly desirable outcome in any crowdsourcing initiative. On the contrary, an ineffective organisational model for coordinating collective efforts can lead to the opposite effect, the “madness of mobs” [30]. This distinction underpins the following research question:

RQ3: What evidence of intelligent action can be identified in HOT-TM mapping projects?

In addressing these questions, our research aims to contribute to several key areas. First, it allows us to better understand the dynamics of participation in humanitarian mapping. As it will be discussed in section 2, while this topic experienced important developments in its early days, subsequent research has not kept pace with the major transformations in the field. This gap has resulted in a lag behind studies of other peer production environments, especially in terms of how contributors collaborate and interact effectively. Furthermore, this study helps to identify specific opportunities to improve the mapping process by focusing on ideas that strengthen, rather than disrupt, productive and sustainable collaborative dynamics.

The remainder of this paper is organised as follows. Section 2 reviews the existing HCI research in the field of VGI, offering insights into humanitarian mapping efforts through the lens of collective intelligence. Section 3 introduces basic notions about HOT and HOT-TM. Subsequently, Section 4 explains the process followed to answer the research questions on how the concept of collective intelligence manifests itself in HOT-TM projects. Section 5 presents the development of the case study following the methodology proposed for that purpose. Section 6 focuses on reflecting on how the current collective intelligence arrangement contributes to or hinders the humanitarian mapping process. Finally, Section 7 offers concluding remarks, opportunities for improvement and potential future directions for research.

2 RELATED WORK

Over the years, the OSM community and its humanitarian applications have served as a rich basis for research on cooperative work practices. In this section, we aim to synthesise previous findings to inform our reflection and discussion on the three components of the working definition of collective intelligence adopted in this paper. Some of these studies share the particularity of using the same data source as our study, the HOT-TM API.

2.1 Crowdsourced Work in OSM

We begin by exploring studies that provide insights into the characteristics of the groups of individuals involved in OSM. As reported by Anderson et al. [1] most efforts to study the social organisation of the OSM community have relied predominantly on qualitative research methods, such as participant observation and interviews. These studies have offered insights into the demographics and motivations of contributors. For instance, Choe et al. [7] highlight the critical role of group composition in shaping group dynamics in OSM, especially conflicts. Their study emphasises the impact of differences that often arise during interactions between various sub-groups of mappers. These sub-groups are often distinguished by factors such as gender, geographical location, relationship between mappers and the areas they map, level of experience and professional affiliation.

Several studies employing more quantitative approaches have explored the role of factors such as mapper experience in shaping the dynamics of OSM. According to Yang et al. [59], a small proportion of contributors with extensive experience in geo-data editing, proficiency in professional software, and a high level of enthusiasm and concentration are responsible for the majority of contributions. This same research suggests that higher experience consistently produces high quality geographic data. Begin et al. [2] identified that most OSM contributors are part of an inactive majority who do very little editing, in contrast to a small group of prolific contributors who are highly active. According to this study, it typically takes several years, with 4.5 years being the norm, for a user to become an advanced OSM contributor. Urrea and Yoo [54] specifically examine the effect of experience on HOT projects. The results of analysing 5,162 HOT projects show that the project completion rate improves, albeit at a decreasing rate, with the experience of contributing volunteers. Furthermore, the effect of experience on the project completion rate is influenced by the urgency of the project. In terms of retention, the results indicate that volunteers feel incentivised to return to an online volunteering platform more quickly when they are close to reaching a new rank based on experience. Dittus et al. [9] conducted an analysis of HOT-TM contributor retention using behavioural data from 1,570 first-time contributors in 99 projects. Their findings indicate that most first-time HOT contributors tend to work at a fairly steady pace, while contributors with previous OSM experience tend to work faster and remain active for longer. In addition, they found that more complex task requirements can demotivate first-time contributors, regardless of their previous OSM experience.

Another key attribute which has received a great deal of attention is the geographical location of the contributors. This factor raises important questions, such as the importance of local knowledge in mapping and the influence of mappers in certain regions on outcomes in other areas. Even the initial conceptualisation of VGI reflects the singular attention attached to the local factor. According to Goodchild [14], "the most important value of VGI may lie in what it can tell about local activities in various geographic locations that go unnoticed by the world's media, and about life at a local level". This intuition appeals to a greater knowledge of the terrain and context on the part of nearby contributors. Locally-produced VGI tends to be associated with higher quality, richness, diversity and utility [11, 27, 45]. From the earliest academic studies on HOT, the role of local mappers has been emphasised. Soden and Palen [46] highlight the need to move from the initial showcase of crowdsourcing efforts following the 2010 Haiti earthquake —the first major humanitarian response by the OSM community— to the development of sustainable, locally-owned community mapping ecosystems in at-risk regions around the world. However, studies indicate that the goal of capturing local knowledge of the terrain often falls short in VGI initiatives, leading to a significant amount, or even the majority, of VGI content being non-local [20, 28, 51].

It is also noteworthy that there are quantitative and qualitative works characterising the OSM group of contributors by developing archetypes that reflect broad patterns of behaviour [6, 61]. Such is the case of Zhang et al. [61] who developed a quantitative approach to identify the emerging editor roles in OSM based on temporal behaviour, change set type, feature diversity and geographical diversity. They identified two clusters particularly associated with humanitarian mappers through a clustering approach: humanitarian enrichers and humanitarian creators. Humanitarian creators show a notable inclination towards creating new mapping features in OSM, demonstrating a strong interest in humanitarian mapping efforts but showing a low propensity to revise existing mapping features to enrich or improve them. In contrast, humanitarian enrichers actively engage in humanitarian mapping efforts, focusing on enriching existing map features by adding detailed attribute information rather than creating new features. In general, humanitarian mappers have lower retention rates than the average among active mappers, with enrichers having higher retention than creators.

2.2 Collective Action in OSM

We now turn to studies that explore collective action within VGI initiatives. Mapping, by its very nature, is a creation activity, in which contributors enrich the collective map by adding elements such as nodes, pathways and relationships. These contributions are made through various map editing interfaces. Humanitarian mapping shows a distinct mapping footprint compared to general OSM activities [21]. It is characterised by being more intensive in mapping buildings than in mapping roads. In addition, edits in humanitarian mapping are more often focused on creating new elements, with less emphasis on modifying or updating elements. This divergence may reflect the unique priorities and objectives of humanitarian mapping efforts, which often aim to rapidly generate essential geographic data for disaster response and relief operations.

Mooney and Corcoran [35] sought to determine whether this creation in OSM involves genuine collaboration or whether it consists primarily of individual tasks performed in isolation with minimal interaction between contributors. Using a case study that examined the entire mapping history of London in OSM between 2005 and 2011, the authors used object co-editing as an indicator of collaboration between mappers. Their findings revealed that collaboration between contributors was limited, suggesting that much mapping activity occurs independently rather than through coordinated efforts. Building on their earlier research, Mooney and Corcoran [36] subsequently extended their study of co-editing networks to include seven major cities, focusing specifically on the activity of very frequent contributors. The results were mixed. On the one hand, these contributors were found to do a great deal of mapping work independently, presumably related to objects that did not attract sufficient interest from other mappers to collaborate in their editing or development. On the other hand, frequent contributors also collaborated with the less frequent contributors by editing or updating their work.

The study of socio-behavioural phenomena in OSM lags behind research in other peer-production environments, such as Wikipedia, where such phenomena have been extensively analysed [1, 29]. The challenges posed by the manipulation of OSM data [1] and the difficulty of tracking the influence of OSM discussion channels on the immediate mapping process [29] are pointed to as possible explanations. Kogan et al. [29] delved into the analysis of user interactions in OSM, while critically addressing the limitations of defining collaboration solely through the co-editing of objects. These authors argue that this narrow definition does not capture the nuanced and multifaceted nature of collaboration as understood in the fields of CSCW and HCI. In their investigation of the dynamics of collaboration during the 2010 Haiti earthquake, they conducted a two-phase study. In the first phase, they employed network analysis techniques to identify high-value segments within the extensive OSM dataset for a more focused qualitative examination. In the second phase, they conducted a detailed content analysis of selected mapping activity and interviewed participants. This approach allowed the researchers to uncover detailed mapping practices that reveal variations in temporal, spatial and interpersonal interactions. The study also recognised the influence of HOT-TM, introduced after the Haiti earthquake, for future research examining the nature of interactions within the OSM community.

2.3 Intelligent Action in OSM

Research has also been carried out to identify the presence of intelligent action within VGI initiatives, with the aim of distinguishing it from mere collective action. Spielman [47] reviewed the existing literature and evidence to assess whether OSM fosters the conditions necessary for the emergence of collective intelligence. The review begins by highlighting the need for map quality metrics and then identifies a recurring tension in relevant literature between two key quality perspectives: credibility and accuracy [8, 15, 17, 18]. Spielman concluded that the system has mixed conditions. On the one hand, low barriers to participation enable widespread contributions but also make it difficult to assess their credibility. This creates the risk that high quality or highly credible data will be degraded when combined with lower quality contributions, which could hinder the development of an intelligent collective outcome. To mitigate this, the OSM community has implemented data validation strategies that rely on the authority of contributors or the sequence in which contributions are made. However, reliance on user authority in crowd-based systems introduces a major drawback: it can become self-reinforcing. This dynamic can lead to overemphasising the work of certain contributors and undervaluing or marginalising others, potentially limiting the diversity and adaptability of the collective intelligence process. Yin et al. [60] used HOT-TM data to evaluate the effects of a micro-tasking intervention on contribution dynamics. According to their research, micro-tasking can effectively increase both the number of contributors and the contribution rates. However, it may also deepen the concentration patterns commonly observed in such settings, where a small group of contributors accounts for the majority of contributions and efforts. They argue that in HOT-TM, as well as in other peer-to-peer production environments such as Wikipedia [38], there is a trade-off between productivity and equity.

Finally, it is worth mentioning emergent research on the effects of artificial intelligence on VGI activities. AI has a transversal impact on the collective intelligence system, as artificial agents, in their various roles, could influence the composition of the group of contributors, modify the organisation of work and may ultimately affect the quality of the final mapping products. In order to assess the impact of AI on volunteer productivity, Tipnis et al. [52] studied the effect of introducing Rapid[3] —an OSM editor designed to add buildings and roads that were identified by AI from aerial imagery— into HOT-TM. Using a difference-in-differences event study design, they found that volunteers who were introduced to AI-powered mapping reduced their weekly contributions to the platform by an average of 8.1% compared to those who were not introduced. Moreover, this effect varied depending on the volunteer experience. While the negative impact on productivity persisted for the least experienced volunteers, those with the highest levels of experience increased their contributions after being introduced to Rapid.

A critical reflection on previous research reveals a significant opportunity to advance studies on collective intelligence in the context of humanitarian mapping, an opportunity that this study aims to seize. On the one hand, the foundational efforts of OSM and humanitarian mapping attracted a great deal of scholarly interest. However, since that initial research there have been a number of noteworthy developments, including the introduction of tools such as HOT-TM and broader societal changes, such as the ongoing impact of the 2020 pandemic on perceptions of remote and distributed work. The continuation of previous research is clearly warranted.

In addition, previous research has highlighted a lag in the study of OSM compared to other peer production environments, particularly in terms of insights into contributor interactions. There is a notable appetite for research that delves into how contributors collaborate and interact to complete mapping tasks but few studies have taken full advantage of the rich data provided by tools such as HOT-TM. Those that have explored these processes have done so only superficially, leaving significant gaps in understanding the dynamics of micro-tasks and the underlying factors that influence their success. This study addresses these gaps by incorporating a collective intelligence framework, which offers a holistic perspective on the humanitarian mapping system, by examining interactions at the micro-task level and identifying factors associated with the successful completion of micro-tasks.

Looking ahead, the buoyant interest in AI and its applications in a variety of fields underscores the importance of improving our understanding of collaborative mapping. The findings of this study can serve as a basis for thoughtful integration of AI tools into humanitarian mapping workflows, ensuring that they enhance rather than disrupt productive and sustainable collaborative dynamics.

3 BACKGROUND

3.1 Understanding the Context of Humanitarian Mapping and HOT-TM

As a predominantly quantitative study, a variety of qualitative methods were employed at the outset of the project to better understand the context of humanitarian mapping. These included in-depth interviews with field experts, participant observation and inspection of the HOT-TM user interface. Although the information collected during this phase was not analysed in depth, it served as a valuable complement to designing, interpreting, validating and discussing quantitative research. We consider it constructive to highlight this aspect of the research process, as access to rich data sources, such as those provided by the HOT-TM API, could lead to assume that analysis can be carried out without a nuanced understanding of the context. However, this omission could limit the potential depth and value of the analysis.

The semi-structured interviews were conducted in two rounds. The first round focused on open-ended questions on mapping objectives, data use, challenges, bottlenecks in the mapping process, communication and collaboration practices during mapping and validation. In this round, we interviewed separately a support specialist from the HOT Hub team for Latin America and the Caribbean (HOT-LAC) and a logistical advisor from the GIS team at Médecins Sans

[3] https://rapideditor.org/

Frontières (MSF) for just over an hour each. In the second round, we analysed the data from the tasking process in HOT-TM, generated preliminary results, and presented these findings to the same interviewees from the first round. For the presentation to the HOT-LAC team, three additional members joined the discussion: a manager, a data quality associate and a technical product owner. The research team annotated the presentation slides to gather feedback on the interpretation of the specific results shown in tables and figures. Interviewees were then asked to identify any inconsistencies in the results, highlight areas for improvement and share their assessments of the quantity, quality and possible improvements of the collaboration during the mapping and validation. In both rounds, one researcher acted as interviewer while the other researcher recorded the responses. These notes were periodically reviewed by the research team to identify supporting arguments and ideas that could enrich the discussion and interpretation of quantitative results.

Interviewees agreed that validators are a scarce resource in projects. They also highlighted the coexistence of two approaches to humanitarian mapping. The first approach prioritises speed and agility to provide basic geographic information to emergency beneficiaries as quickly as possible. The second approach focuses on more elaborate objectives, such as emergency preparedness, prevention, and forecasting, and emphasises higher-quality and more detailed information, such as building types, water bodies, and health facilities.

For the participant observation, three members of our research team attended an online Mapathon organised by HOT. One researcher, with experience in humanitarian mapping missions and OSM, took part in the task validation, which is only allowed to expert contributors. The other two researchers, who were new to the process, first received a half-hour introduction to HOT-TM before starting basic mapping operations. The activity lasted two hours in total.

3.2 Understanding the HOT Tasking Manager

In this section, we introduce basic notions to understanding HOT and the HOT Tasking Manager (HOT-TM), the main technological tool for coordinating humanitarian mapping. This description focuses particularly on identifying connections with categories of collective action presented in the introduction, which will serve as a basis for the development of later sections. Figure 1 shows a typical project page in HOT-TM which contains a map with the different tasks and their current states. As can be seen, the project area is divided into tasks following a grid. The process followed by the mapping tasks can be divided into the Mapping and Validation phases [23]. As the tasks go through these two phases, they pass through various states registered by the HOT-TM system (see Figure 2).

In the mapping phase, contributors select a task or receive a random one. A built-in editor opens in HOT-TM (although contributors can also choose to use a desktop editor). In this editor, contributors must draw missing elements as specified in the project description, such as buildings or roads, making sure not to draw outside the marked area. Note that they correspond to creation activities. Then, contributors decide individually whether the task is fully mapped, not fully mapped, in need of splitting, or its imagery is poor. Splitting a task divides it into four smaller tasks. This step can be repeated several times. Since the contributions of successive mappers are directly influenced by the work of previous mappers on the same task, this implies that mapping activities allow for some degree of asynchronous collaboration. The characteristic states of this phase registered by HOT-TM are as follows:

- **Locked for mapping:** Tasks currently being edited by mappers are locked to prevent overlapping editions. The system records the duration of this state. In the end, the mapper decides if the area is completely mapped, or if more mapping is needed.

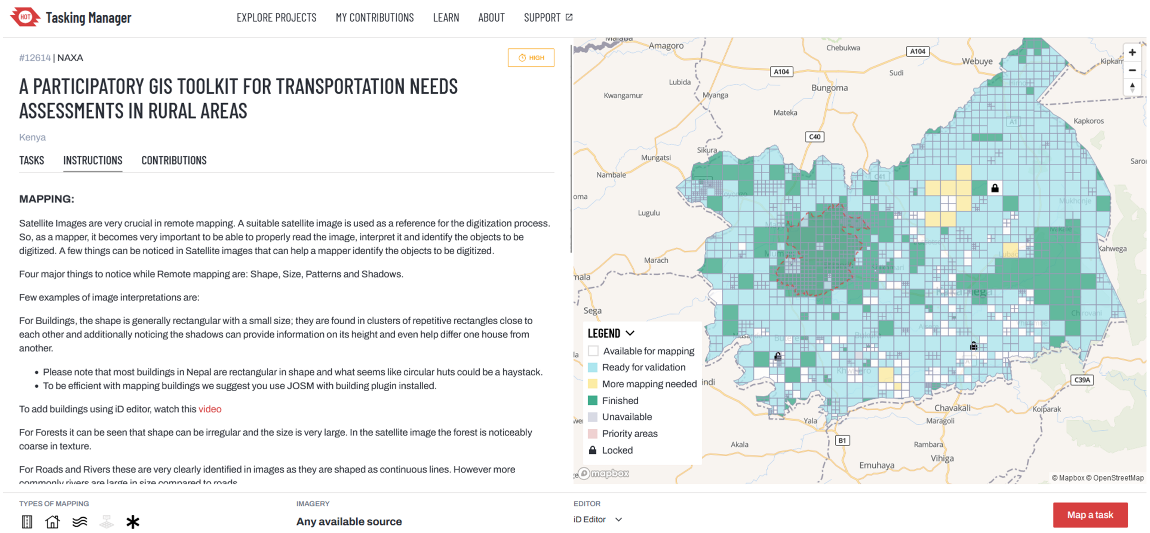

Fig. 1. Project Page in HOT-TM: the left section displays information about the project, including a list of task states, instructions, and contributor metrics. On the right, a map visually represents the task states. Contributors can select a random task to map using the button at the bottom right, or they can manually select it. Once a task is selected, contributors can choose between mapping or validate modes. Source: https://tasks.hotosm.org/projects/12614/tasks

- **Auto-unlocked for mapping:** Tasks locked for mapping are automatically unlocked after a timeout period of 2 hours. The mapper may have done some editing during the period which would still be preserved.
- **Mapped:** Once the mapping job is complete as judged by the last mapper, then the task becomes complete and is ready for validation in the next phase.
- **Split:** The mapper can split the task into smaller parts by assigning this state when the task is dense or complex. When a task is split, four new tasks are created.
- **Bad imagery:** This state is acquired when, in the opinion of a contributor, the image of the task is not of sufficient quality for proper mapping.

In the validation phase, contributors review the mapped tasks. An editor similar to the mapping phase opens offering multiple options. The contributor then decides individually whether the task is properly mapped and moves it to the validated state or, if not complete, to the invalidated state. Those tasks that are invalidated return to the mapping phase for subsequent work. The characteristic states of this phase registered by HOT-TM are as follows:

- **Locked for validation:** Mapped tasks acquire this state when a contributor selects them for validation, preventing concurrent validation. Note that the validator has the ability to self-correct the task. Therefore, the validation phase may also include creation activities. The system records the duration of this state.
- **Auto-unlocked for validation:** Tasks locked for validation are automatically unlocked after a timeout period of 2 hours. The validator may have done some work during the period it was locked but did not ultimately record its validation decision.
- **Validated:** Tasks correctly mapped according to the mapping instructions are marked as validated.
- **Invalidated:** Tasks incorrectly mapped according to the mapping instructions are marked as invalidated and return to the mapping phase.

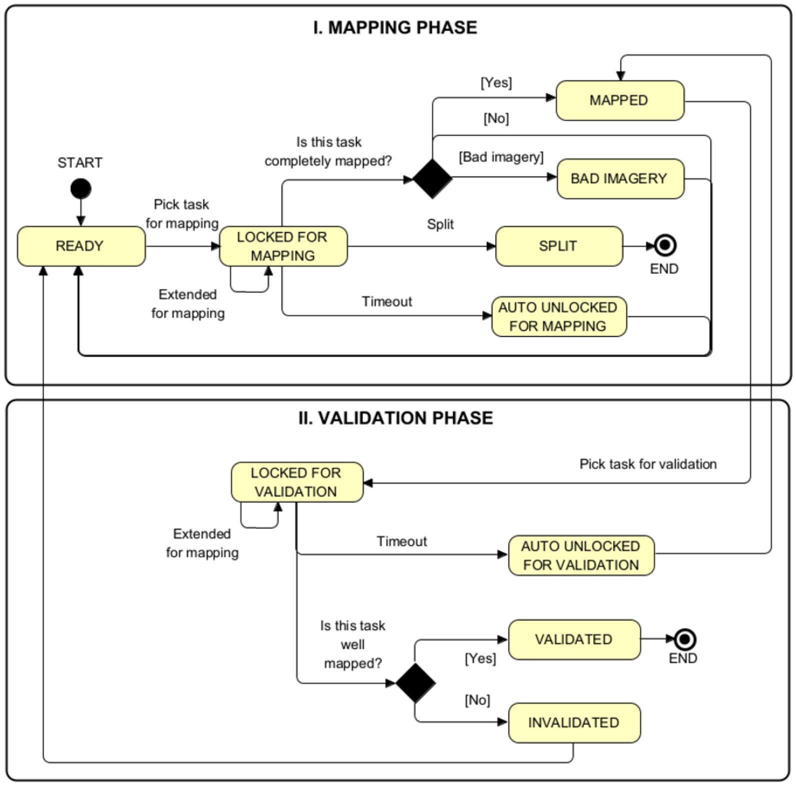

Fig. 2. Task state diagram showing the mapping workflow in HOT-TM (Herrera et al., 2024).

There is an additional state that can appear in both the mapping and validation phases called Extended for mapping. After two hours have elapsed, the contributor can extend this locked period to continue mapping or validating the task, avoiding the auto-unlock.

4 METHODOLOGY

To meet our objective, we proposed a two-stage methodology: (1) data collection, and (2) data analysis (see Figure 3). In addition, we have created a code repository[4] containing Python and R notebooks that provide a detailed description of the procedures used for data collection, preprocessing and analysis in this study.

4.1 Data Collection

This subsection describes the procedures used to collect the quantitative data used for the study. The HOT-TM API[5] provides access to numerous data on humanitarian mapping projects. The API operations are grouped using tags representing mapping-relevant concepts (e.g. *projects*). Each invoked operation returns items with associated data fields (e.g., *action*, *actionDate*, *actionBy*), which we then mapped to answer the research questions.

[4] https://github.com/IAAA-Lab/Collective-Intelligence-in-Humanitarian-Voluntary-Geographic-Information-ODECO-HOT-TM
[5] https://tasks.hotosm.org/api-docs

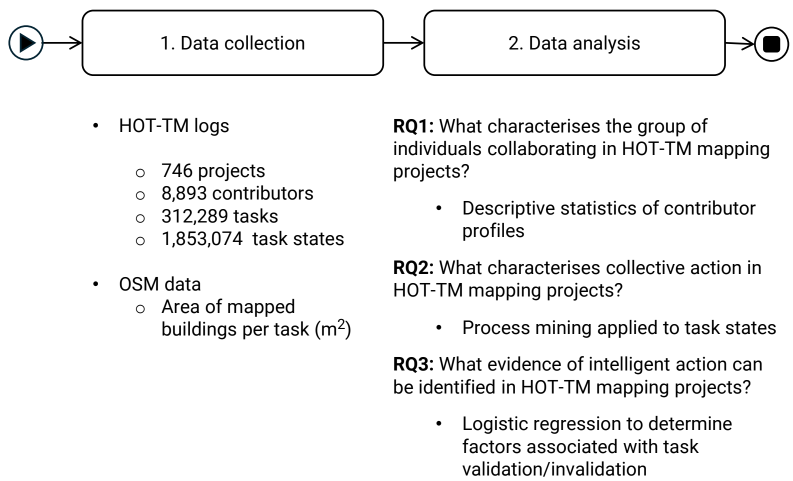

Fig. 3. Methodology at a glance.

The *users* tag was revised to address the question of group composition. This tag includes an operation (*queries*) for retrieving attributes describing each OSM user, which was queried with those participating in each humanitarian project. Among its fields, *mappingLevel* and *projectsMapped* provide information on their mapping experience. Similarly, the *country* field provides information on the location of contributors. As explained in the related work section, these attributes contribute substantially to understanding group composition. In addition, secondary fields such as *photo*, *slack*, *facebook*, *linkedin*, and *twitter* were collected.

For the collective action question, we sought data describing the tasks executed in a project and the individuals responsible for those tasks, with the aim of analysing types of activities and interactions. The *projects* tag was relevant to this purpose, particularly its *activities* operation, which retrieves the history of states through which tasks in a project pass. The key fields are *action*, which specifies the type of task state (see Section 3); *actionDate*, which records the date of the state change; and *actionBy*, which identifies the person responsible for the state change. From the same tag, we also used the *statistics* operation to retrieve detailed info for all projects. The extracted fields included *priority*, *difficulty*, *perc_validated* (the percentage of total tasks validated), *created* (the date the project was created), *total_mappers*, *organisation* and *country*. Difficulty and priority are attributes that potentially influence all tasks in the same project and will, therefore, serve as recurring variables in subsequent analyses.

Regarding the question of intelligent action, previous research underlines the importance of having map quality indicators to study the spatial collective intelligence of a VGI system [47]. Although the available API operations lack direct indicators of map quality, the *activities* operation of the *projects* tag provides basic task performance evidence through the task validation phase. In particular, the label field indicates the result of the task validation as *VALIDATED* or *INVALIDATED*. These results serve as indicators to assess whether the mapping process has been executed correctly.

The resulting dataset comprises 746 HOT-TM projects. We selected projects created between December 1st, 2021, and the collection date —December 1st, 2023— and, from those, all the complete and archived projects. In these projects, 38,893 contributors completed 312,289 mapping tasks. In addition, we collected data on the 1,853,074 states that the mapping tasks went through in the projects.

For this exercise, we mainly relied on the data provided by the HOT-TM API to maintain manageability. However, we considered it necessary to include a basic control indicator associated with the workload of the tasks. To this end, we complemented the analysis with OSM data to calculate the observed building area for each mapped task. Task grids were used as the main input for the Bunting Labs API[6] to retrieve buildings for each mapped task. Before calculating their areas, both the task grid and building datasets were reprojected from geographical coordinates to the Universal Transverse Mercator (UTM) projection.

The data collected from the APIs were stored in files for further processing. The primary goal of data pre-processing was to prepare an event log, enabling the application of the relevant quantitative analysis techniques. For more details on the data collection process and pre-processing, the code repository includes a Python script that explains step-by-step how the HOT-TM and Bunting Labs APIs were accessed.

4.2 Data Analysis

This subsection outlines the analytical strategy used to answer the research questions. Before elaborating on the strategy, it is important to note that the analysis began with the development of a profile of the projects under study. This profiling was done to provide context on key project variables, such as difficulty and priority, which can significantly influence the execution of mapping tasks. These variables will be mentioned frequently in the subsequent analysis and presentation of results.

4.2.1 RQ1: What characterises the group of individuals collaborating in HOT-TM mapping projects?

We developed a profile of the composition of the group based on two key attributes: the experience and location of the mappers, using the available data. For this initial profile, we consider as a group unit all 38,893 contributors who participated in any of the 746 mapping projects included in the dataset. Regarding expertise, HOT-TM uses a mapping level system to classify contributors into three categories: beginners (less than 250 OSM changesets), intermediate (250-499 changesets) and advanced (500 or more changesets).[7] Unlike other labels, this classification is available to all users of the system. We analysed the frequency distribution of mapping levels and their association with the number of HOT-TM projects in which contributors participated. In addition, we use complementary fields (name, city, country, photo, Slack, Facebook, LinkedIn, Twitter) to examine the completeness of individual profiles according to their mapping level. Similarly, to profile contributors location, we used data on the country from which they reported and the countries in which humanitarian projects were implemented. This allowed us to describe the distribution of contributors according to the origin and destination of contributions. In addition, since HOT organises its activities through regional hubs serving specific geographic areas, we also grouped contributors using these broader geographic units. It is important to note that location-based frequency calculations were only performed for contributors who reported their country, which represents 29% of the total. Finally, we calculated the group composition of these two attributes according to the type of project.

4.2.2 RQ2: What characterises collective action in HOT-TM mapping projects?

We conducted a quantitative characterisation of collective action using the activity and interaction framework proposed by Suran et al. [49] and the crowdsourcing activity tracking approach outlined by McDonald [34]. In this analysis, it is

[6] https://buntinglabs.com/solutions/openstreetmap-extracts
[7] https://github.com/hotosm/tasking-manager/blob/v4.7.4/backend/config.py

important to distinguish between states that directly reflect creation activities (e.g., "Locked for mapping") and those that indicate decision-making activities (e.g., "Mapped", "Validated", "Invalidated").

Our findings are presented in three parts. The first part characterises the mapping process by analysing the control flow and duration of task states in a typical mapping task. To improve readability, details on the frequency and duration of mapping states are provided in the appendix. The second part explores work articulation and contributor roles throughout the mapping process. We analysed how task states are distributed based on mapping levels and the geographic location of contributors, distinguishing between mapping and validation phases. Additionally, we examined how these patterns vary with project difficulty and priority. The final part focuses on interaction analysis, recognising that while mapping tasks can be completed individually or collaboratively, key decisions are always made individually. We began by quantifying the number of mappers involved in task execution across different scenarios. Next, we assessed the frequency of collaboration during the mapping phase of a standard task. Finally, we examined how collaboration relates to the mapping level of contributors by identifying common combinations of collaborators and applying the "handover of work" concept [56]. This concept posits that the more frequent an individual "x" performs an activity that is causally followed by an activity performed by "y", the stronger the relationship between "x" and "y" is.

To guide the analysis of this question, we took advantage of process mining techniques [56]. They use event logs to gain insights into processes from multiple perspectives. In this case, the logs consist of state changes in mapping tasks provided by the HOT-TM API. Specifically, we selected *bupaR*,[8] an open-source R package for business process data analysis. This choice was motivated by its extensive range of analytical capabilities and well-documented functionality.

4.2.3 RQ3: What evidence of intelligent action can be identified in HOT-TM mapping projects?

To address this question, we adopted the approach that tracks the footprint of intelligent activity on system outputs [32], in line with the recommendation to use a quality indicator to study spatial collective intelligence [47]. As noted in the data collection section, the main indicator available for this purpose in HOT-TM is the field that indicates the validation result of each task as either "Validated" or "Invalidated". This result serves as evidence of whether the underlying mapping process has been executed correctly. Based on this reasoning, we performed a logistic regression analysis. This technique produces coefficients that describe the relationship between quantitative independent variables and a binary dependent variable - VALIDATED (0) / INVALIDATED (1) states.

The selection of independent variables included indicators that capture the collective intelligence factors discussed in the previous sections. To account for the effect of mapper experience, we included the mapping level of the mapper who marked the task as mapped, as this mapper is assumed to have the most influence on decision-making about the task. To account for the effect of mapper location, we included the location of the mapper who marked the task as mapped. The result of this factor should be interpreted with caution given the high level of profiles with incomplete localisation and the bias of more experienced users to complete their profiles. To capture the effect of mapper interactions during task execution, we included an indicator specifying whether the task was completed by a single mapper or whether it required the contribution of multiple mappers. In addition to these variables, we included control factors such as the difficulty and priority of the project, as well as the area of buildings mapped per task, reflecting the workload and complexity associated with each task. We also included the relative validation time to account for the potential impact of corrective work performed by validators. Difficulty, priority, involvement of multiple mappers and the mapping level and location of whoever declared the task as finally mapped were coded as dummy variables. Relative validation time

[8] https://bupar.net/

was calculated as the proportion of the total processing time devoted to validation, while the area of mapped buildings was logarithmically transformed to account for scale effects.

In order to have an appropriate dataset for the experiment, we trimmed the event log to take into account only the states prior to the first validation. Tasks with states of SPLIT, AUTO-UNLOCKED FOR MAPPING and AUTO-UNLOCKED FOR VALIDATION were also excluded to remove the effect of duplicate states and because of the impossibility of determining the effective time that contributors spent on processing the auto-unlocked tasks.

More details on the analysis procedure can be directly explored on the project repository, which contains Python and R notebooks describing the step-by-step process.

5 RESULTS

This section begins by describing the main characteristics of mapping projects and then develops the findings for the three research questions.

5.1 Profiling of Humanitarian Mapping Projects

The study includes 746 mapping projects, for which the system records key variables such as difficulty, priority, number of tasks and number of collaborators, along with additional descriptors such as location and coordinating organisation. These variables are characterised according to their relevance in shaping the context of task execution.

The left panel of Figure 4 shows the distribution of the projects according to their difficulty, priority, number of tasks and contributors who participated in them. Approximately two thirds of the projects belong to the easy difficulty category, one third to moderate and only a few projects fall into the challenging category. Just over half of the projects have a low priority, almost a third have a moderate priority and the remaining fifth are divided between high priority and urgent projects. The number of tasks and contributors per project exhibits a noticeable dispersion and a positive skew due to some large observations.

The right panel of the same figure elaborates on the relationship between the above variables using a series of scatterplots segregated by level of difficulty that display the number of tasks per project on the horizontal axis and the number of contributors on the vertical axis. Priority levels are encoded using different colours. Between the two main categories of difficulty, there is a difference in the proportion of low priority projects, which tends to be higher for easy projects than for moderate difficulty projects (60% vs. 31%). The majority of urgent projects are of moderate difficulty (69%). There is a positive correlation between the number of tasks and the number of contributors to the projects regardless of the level of difficulty. One fifth of the projects of moderate difficulty were completed by a pair of contributors, presumably one in the role of mapper and the other as validator. No strong patterns are apparent with respect to the association between the priority level of projects and their number of tasks and contributors. Although in moderate projects, more urgent cases seem to recruit more contributors than less urgent projects given an equivalent number of mapping tasks.

Other notable features of the projects are their location and the coordinating organisation. HOT-TM projects are mainly organised around four regional hubs serving specific geographical areas of particular interest (Asia-Pacific —AP—, East and Southern Africa —ESA—, West and North Africa —WNA—, Latin America and the Caribbean —LAC—). ESA is the hub with the highest number of projects (41.4%), followed by LAC (26.0%), WNA (17.3%), and AP (11.0%). Projects outside the hubs are infrequent (4.3%). The membership of certain countries in a given hub is a factor that will be taken into account later in the analysis of the geographical origin and destination of contributions. Finally, each

Projects		
Difficulty	nº	%
Easy	501	67.2
Moderate	239	32.0
Challenging	6	0.8

Projects		
Priority	nº	%
Low	377	50.5
Medium	232	31.1
High	92	12.3
Urgent	45	6.0

	# Tasks	# Contributors
Avg.	418.6	91.2
Std.	487.3	227.5
Min.	8.0	1.0
25%	107.3	14.0
50%	286.0	38.0
75%	564.0	86.8
Max	4,483.0	3,608.0

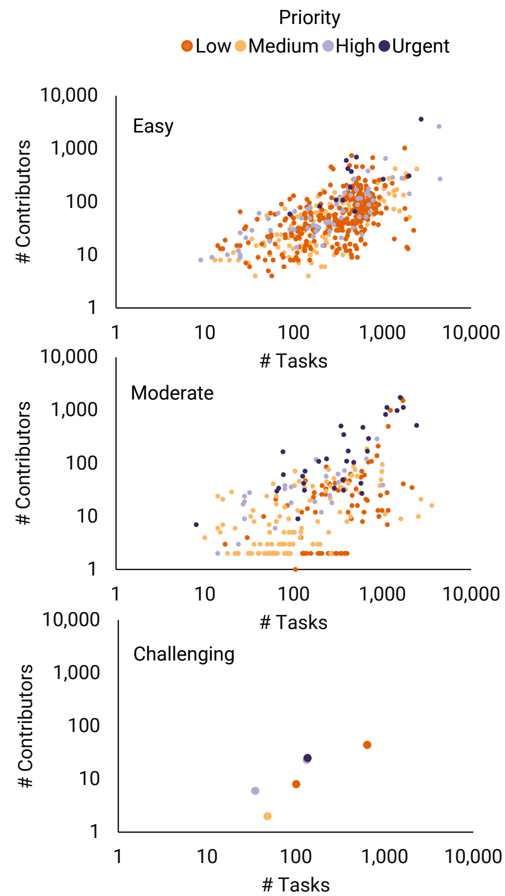

Fig. 4. Analysed projects at a glance

project is associated with one of 56 organisations. The organisations with the highest number of projects are HOT (24.9%) and Médecins Sans Frontières —MSF— (17.2%). The other organisations report marginal frequencies.

5.2 Data Analysis

5.2.1 *RQ1: What characterises the group of individuals collaborating in HOT-TM mapping projects?*

With the aim of profiling the members of the group according to their level of experience, Table 1 shows the breakdown of the total number of unique contributors who participated in the humanitarian mapping projects analysed according to their mapping level. It can be seen that the vast majority (approximately nine out of ten) of contributors are beginner OSM mappers. Only a small proportion of contributors are advanced and intermediate OSM mappers. The increase in the mapping level is also followed by a higher frequency of participation in HOT-TM humanitarian mapping projects. Most novice users have participated in a single humanitarian mapping project, while intermediate and advanced contributors are repeat participants.

Mapping Level	Contributors		Nº HOT projects		
	nº	%	Mean	Median	Std.
Beginner	35,387	91.0	1.9	1.0	3.4
Intermediate	980	2.5	12.2	8.0	13.3
Advanced	2,526	6.5	35.7	12.0	79.4
TOTAL	38,893	100	4.4	1.0	22.3

Table 1. Overview of total unique contributors by mapping level

Differences in mapping levels are also evident in the completeness of profile fields, as shown in Figure 5. Beginners tend to have substantially more incomplete profiles than intermediate and advanced contributors, who have similar completeness rates for most attributes. The exception is the profile picture field, which advanced contributors complete more frequently than intermediate contributors.

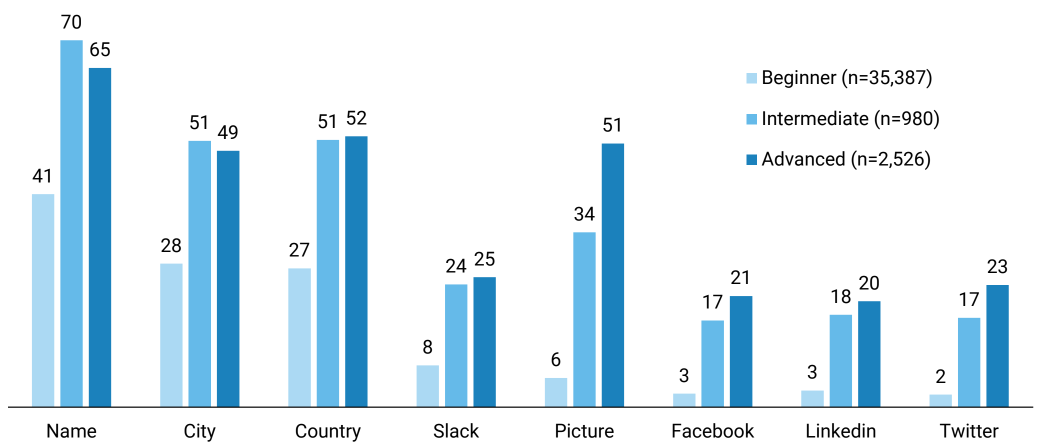

Fig. 5. Completeness of the contributor profile by mapping level - % of total contributors-

Another important attribute to characterise the members of the group is their location. In this respect, the reported countries give a clue about the geographical origin and destination of contributions in HOT-TM, as shown in Table 2. The table presents the weighted proportion of contributors according to their declared location, categorised by the hubs in which projects are located. The upper part of the table breaks down contributions by hub of origin, while the lower part of the table highlights contributions from the 10 most active countries overall. It must be noted that the calculations exclude contributors who did not report their location, which, as mentioned above, constitute the majority. Approximately two-thirds of the contributions with identifiable origins come from outside the HOT hubs, a trend that holds for all reference regions. Generally, the second largest group of contributors within a hub is those located in the same hub. Looking at the results by country, the United States, the United Kingdom and the Netherlands together account for about one-third of the total contributions, both overall and within each destination hub. This reflects a mapping dynamic in which contributors are still mainly located in the Global North, while most projects are in the Global South. When examining the distribution of users between the different hubs according to their mapping level, some differences can be observed, although they are not drastically pronounced. The proportion of experts ranges from 15.5% in the LAC Hub to 29.8% in the WNA Hub.

		Hub where project is located					
Contributor location		TOTAL	ESA	LAC	WNA	AP	OTHER
Hub	OTHER	63.6	63.5	54.5	69.6	56.5	74.8
	ESA	10.6	20.1	2.9	7.0	5.7	5.8
	LAC	2.2	1.6	5.0	1.0	1.6	1.8
	WNA	6.5	5.8	3.4	16.9	4.8	5.3
	AP	17.	8.9	34.2	5.5	31.4	12.2
Top 10 countries	United States	15.8	16	18.3	11.2	15.4	16.4
	United Kingdom	11.7	12.1	12.0	11.4	11.3	11.2
	Netherlands	5.2	4.9	1.9	11.0	5.3	5.3
	Philippines	4.7	2.2	16.1	0.7	2.0	2.5
	India	4.5	1.8	11.2	1.3	7.6	2.8
	Nepal	4.5	2.3	4.9	1.7	13.5	3.4
	Germany	3.4	3.6	2.5	4.1	3.3	3.8
	France	3.2	3.0	2.0	6.2	2.4	3.6
	Kenya	2.9	5.7	0.9	0.9	1.1	2.1
	Nigeria	2.6	2.6	1.2	7.3	2.1	1.6

Table 2. Origin and destination of HOT contributions - weighted proportion of contributors from project locations .

To complete the overview of the composition of the groups, Figure 6 shows the distribution of contributors according to their mapping level and location for different kinds of projects. The values shown correspond to the weighted average percentage of contributors for each project category. The calculation of the proportion of users by location (same country as the project, another country in the same hub as the project and countries outside the hub) excludes contributors who do not report their country. A breakdown of the association between the type of project and the composition of the mapping level of the participants shows that as the difficulty assigned increases, the proportion of advanced users increases substantially to the detriment of beginners. Priority does not seem to be a determining factor in the mapping level mix of a project. Projects with a lower number of tasks are associated with a higher proportion of advanced contributors. The breakdown of proportions by contributor location suggests that the bulk of activity is carried out by contributors outside the country and the hub where the projects are located regardless of the category of project in question. However, the proportion of mappers located in the same country tends to be higher for lower priority projects with fewer tasks.

5.2.2 *RQ2: What characterises collective action in HOT-TM mapping projects?*

a) The Mapping Process: How Does the Work Happen?

Figure 7 illustrates the typical lifecycle of a mapping task, highlighting the various states it progresses through. This directly-follows graph, which represents the frequency and duration of mapping task states, was derived from the 85% most common traces. In terms of frequency, the nodes indicate the absolute number of state instance executions, while the edges represent the absolute number of times the source and target states were executed sequentially. Both the

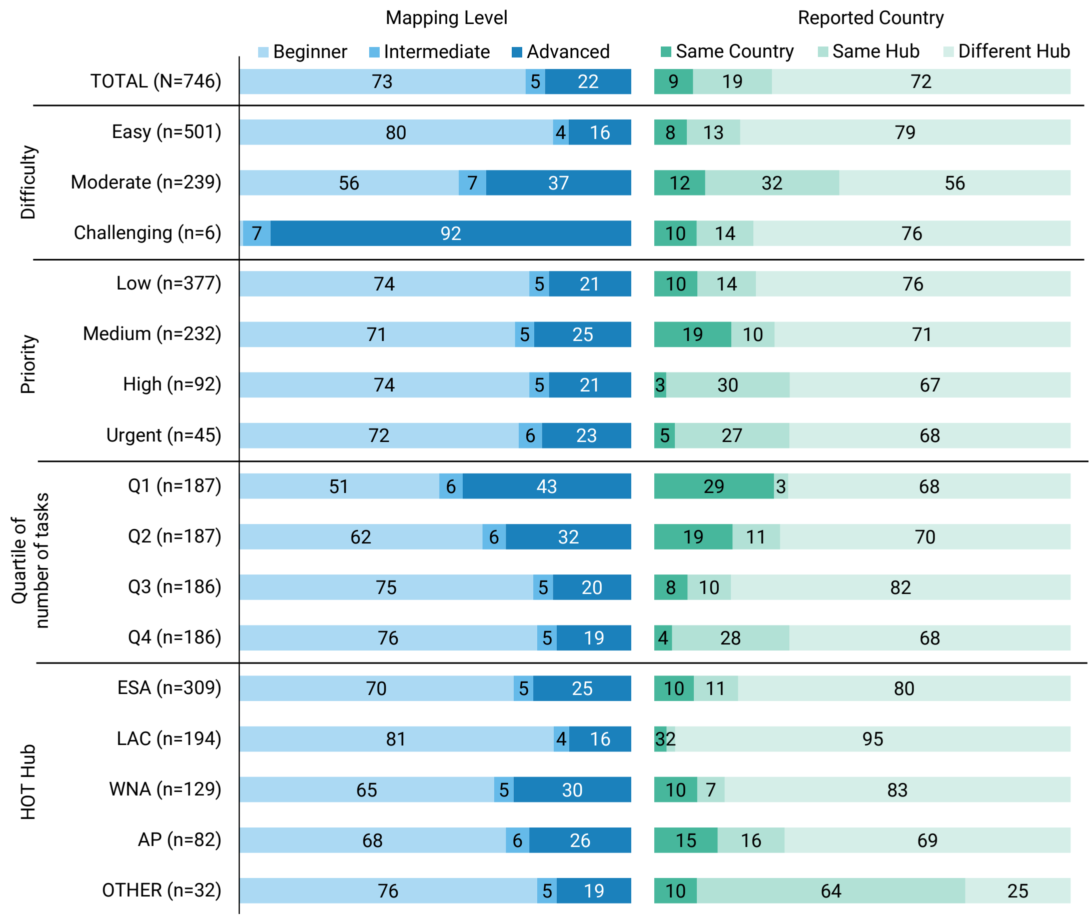

Fig. 6. Participation structure in projects - weighted proportion of contributors per project-

thickness of edges and vertices are proportional to the frequency of occurrences, with higher frequencies resulting in thicker edges and vertices. The median duration of states or transitions is shown just below each frequency count, either in minutes, hours or days as appropriate. Those transitions whose median duration was longer than one day are highlighted in orange.

The flow reveals a primary pathway for mapping tasks that follows a straightforward progression. Typically, contributors lock a task for mapping and complete one or two mapping cycles before it is marked as Mapped. The task then moves to the Locked for validation state and is usually declared Validated on the first attempt. However, alternative, less frequent pathways also exist, as indicated by states such as Split, Auto-unlocked for mapping, and Invalidated. It is important to note that the frequency of these deviations from the standard mapping process varies considerably depending on the project type, increasing for projects with higher difficulty and priority (see Table 7 in Appendix A). This variation underscores how project characteristics shape the complexity of task execution and decision-making processes.

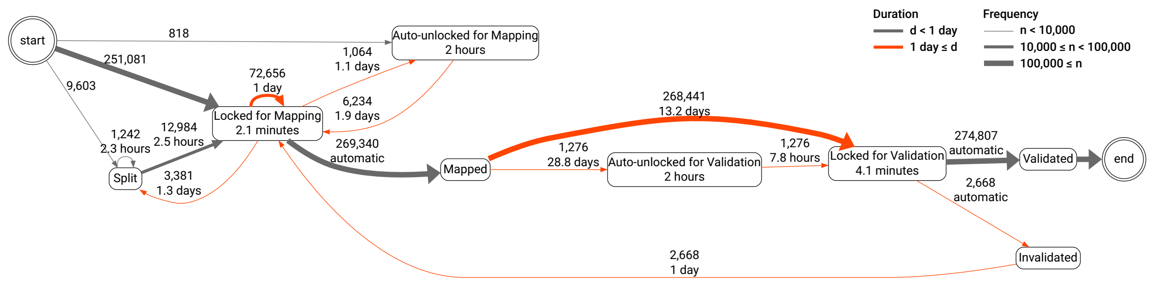

Fig. 7. Frequency and duration map of task states and transitions (85% of most frequent traces)

Shifting focus to the temporal dimension of the flow, we find that the median duration of a mapping cycle is just 2.1 minutes. However, if a task is not declared Mapped at the end of a cycle, it faces a median wait time of one day before re-entering another mapping cycle. Once a task is marked as Mapped, it remains in a queue for a median of 13.2 days before being Locked for validation, a process that itself takes a median of 4.1 minutes. Similarly, if a task is Invalidated, it typically waits another day before re-entering the mapping phase. Notably, the actual processing times for mapping and validation are brief—just a few minutes—whereas the transitions between these stages introduce significant delays, often measured in days. This contrast underscores how waiting time constitutes one of the main "costs" of decision-making. The bottleneck becomes particularly evident when a mapper hesitates to mark a task as Mapped or when a validator decides to Invalidate a task. Another key observation, which will be explored further, is that the median time a task remains in the Locked for validation state is twice as long as in the Locked for mapping state. Once again, the duration of states and transitions varies significantly by project type (see Table 8 in Appendix A). For instance, tasks in low-priority projects can wait nearly a month for validation after mapping, whereas for other projects this delay is typically less than a week.

b) Roles and Responsibilities: Who Does What in Mapping?

It is now time to address the issue of role configuration, with a particular focus on identifying who assumes the burden of creation and decision-making within the system. To this end, Table 3 shows the breakdown of the states according to the mapping level of the contributors responsible for their execution, depending on whether they were beginners or advanced contributors. Intermediate contributors, reporting a generally marginal presence, were omitted for the sake of readability. The first row of results takes the weighted average number of contributors per project for each mapping level as a benchmark. The remaining rows show the proportion of the total instances of each state that was performed by one or another contributor profile according to the type of project.

The first key finding concerns the creation activities during the mapping phase. According to the proportion of Locked for mapping states, advanced mappers take on a significantly higher mapping workload compared to their proportional representation within the group of mappers across all project types. However, this distribution depends on the difficulty of the project. In terms of overall volume of activity, beginners carry the main mapping load on easy projects. This trend is reversed in projects of higher difficulty, where the mapping load is concentrated on advanced mappers. Advanced mappers also play a dominant role in the decisions made during the mapping phase, as suggested by a substantially higher than expected frequency of Mapped and Split states in all project categories. In contrast, the proportion of Auto-unlocked for mapping states is mainly associated with beginners, which could indicate a

		TOTAL N=312,289		Easy n=222,729		Moderate n=88,468		Challenging n=1,092	
		Beginner	Advanced	Beginner	Advanced	Beginner	Advanced	Beginner	Advanced
Proportion of contributors mapping phase		76	20	82	13	58	34	1	91
MAPPING PHASE	Locked for mapping	51	42	61	32	29	65	0	99
	Mapped	44	50	55	39	17	77	0	99
	Auto unlocked for mapping	74	20	79	16	65	29	0	100
	Split	25	71	35	60	15	82	0	100
	Bad imagery	52	44	82	12	14	85	0	100
Proportion of contributors validation phase		4	91	7	89	1	94	0	97
VALIDATION PHASE	Locked for validation	1	98	1	98	0	99	0	100
	Auto unlocked for validation	2	96	2	95	0	96	0	100
	Validated	0	98	0	98	0	99	0	100
	Invalidated	2	96	3	96	0	96	0	100
		Low n=166,911		Medium n=80,768		High n=38,119		Urgent n=26,491	
		Beginner	Advanced	Beginner	Advanced	Beginner	Advanced	Beginner	Advanced
Proportion of contributors mapping phase		77	18	74	22	76	19	73	21
MAPPING PHASE	Locked for mapping	53	41	44	49	47	45	57	35
	Mapped	47	47	41	53	37	56	43	50
	Auto unlocked for mapping	71	23	68	24	74	21	77	17
	Split	27	69	27	69	18	79	25	71
	Bad imagery	75	22	34	64	76	11	41	48
Proportion of contributors validation phase		6	90	7	89	2	95	0	92
VALIDATION PHASE	Locked for validation	1	99	0	98	0	96	0	98
	Auto unlocked for validation	2	96	2	97	1	98	1	91
	Validated	0	99	0	98	0	96	0	98
	Invalidated	6	89	1	97	0	98	0	99

Table 3. State execution based on the contributor mapping level - % of states-

higher likelihood of interface-related misunderstandings, lack of confidence in making decisions or deviations from recommended mapping practices. In the validation phase, all states are executed almost exclusively by advanced contributors. This is due to the participation criteria set by project managers at the beginning of the project, which usually exclude beginners.

Incomplete data from the location field limits a detailed analysis of the distribution of states by mapper location. However, to provide an overview of the intensity of mapper contribution as a function of their location with respect to the project, Figure 4 presents the distribution of total MAPPED states by project priority and project size (measured by number of tasks). These two variables showed the greatest variation in the concentration of national mappers, as highlighted in Figure 6. As a benchmark for assessing the volume of activity, the first row presents the total weighted proportion of users in each location category across all projects. It is important to note that this calculation excludes MAPPED states executed by users with unknown location, and interpretation should take into account a possible bias towards expert users, who tend to have more complete profiles. The results indicate that national mappers tend to contribute proportionally more than their representation within the overall group of contributors. In addition, their relative activity appears to be concentrated on lower priority projects with a lower volume of tasks. Projects with these characteristics are more suitable for preventive, preparatory and follow-up mapping activities.

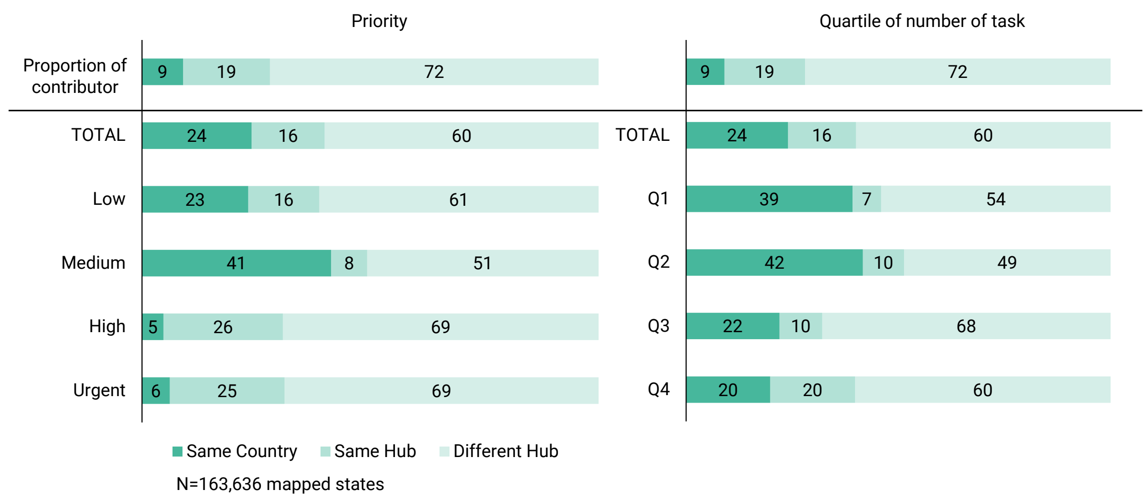

Table 4. Execution of mapped states based on contributor location - % of mapped states where the location of the mapper is known-

c) Collaboration in Mapping: How Do Contributors Interact?

After analysing the nature of the work and the configuration of roles based on mapping states, the next step is to examine the interactions between the mappers in completing the tasks. To do this, Figure 8 elaborates on the number of contributors involved in the completion of a task. We take as a basis the tasks that do not report frictions such as splits or invalidations, which are the majority. Then we show the case of tasks with splits, invalidations or both states. For each task, a distinction is made as to whether such participation occurred for the total number of states or for the states corresponding to the mapping and validation phases. Descriptive statistics such as average number, standard deviation, and quartiles are displayed for each scenario.

If we look at the averages, the number of contributors needed to carry out the mapping operations of a task multiplies almost 4-fold when Split, and 5-fold when Invalidated states occur. In the case of the validation phase, the presence of these states also reports an increase, albeit more discrete, in the number of validators. This suggests that, in addition to the waiting time costs discussed above, decisions to perform a split or an invalidation must also be assessed in terms of the number of mappers needed to process the task.

Having realised that tasks that suffer disruptions such as splits or invalidations by nature require a combined effort between several contributors, the following results focus on the analysis of interactions in ordinary tasks, which comprise the vast majority of cases. We then concentrate on states in the mapping phase corresponding to tasks that have not undergone splits or invalidations to analyse interactions. As explained previously, the HOT-TM micro-tasking scheme allows several mapping operations of the same or different mappers on a task before declaring it as mapped, prior to its validation. These multiple operations, if they occur, are not done simultaneously, as the task is locked for mapping. Each subsequent edit builds on the work done by the previous mappers, as the map is done on the shared OSM database.

Figure 9 shows the average percentage of tasks per project whose mapping states were executed by more than one mapper, broken down by difficulty and priority. It is important to note that, proportionally, tasks are predominantly carried out by individual mappers rather than through collaborative efforts in all project categories. However, there are

N=312,289		Avg.	Std.	25%	50%	75%
No split/No invalidated n=263,165 (84.3%)	Total	2.4	0.9	2	2	2
	Mapping	1.4	0.9	1	1	1
	Validation	1.0	0.2	1	1	1
Split n=42,223 (13.5%)	Total	7.2	6.2	3	5	9
	Mapping	5.9	5.8	2	4	8
	Validation	1.5	1.0	1	1	2
Invalidated n=19,750 (6.3%)	Total	10.0	7.4	5	7	13
	Mapping	8.1	7.0	3	6	10
	Validation	2.3	1.1	2	2	3
Split & Invalidated n=12,849 (4.1%)	Total	12.5	7.9	7	11	16
	Mapping	10.5	7.3	5	9	14
	Validation	2.6	1.2	2	2	3

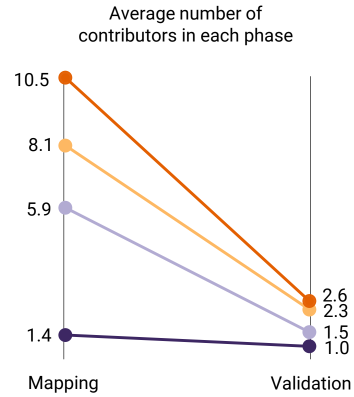

Fig. 8. Number of contributors per task, depending on the occurrence of Split or Invalidated states (average number, standard deviation, and quartiles)

variations. Higher priority levels mean an increase in the number of tasks with multiple mappers. In terms of project difficulty levels, easy projects report the highest proportion of collaborative tasks, followed by projects of moderate difficulty and challenging projects.

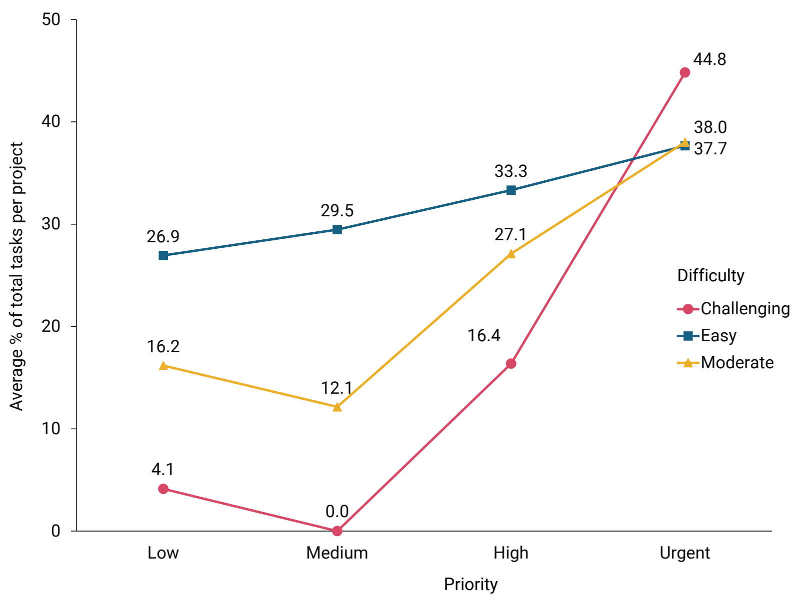

Fig. 9. Tasks mapped by more than one contributor by project difficulty and priority

In the next step, we focus on understanding the most common combinations of mapping levels that are observed in collaborations in the mapping phase. Figure 10 shows the top 5 popular combinations of mapping levels involved in Locked for mapping and Auto-unlocked for mapping states before the first validation of a task. Next to each combination, the percentage of the total collaborative tasks it represents is shown. It can be observed that the 5 most common combinations alone account for approximately two thirds of the total collaborative tasks for almost all project types. Most of these popular combinations are binary and to a much lesser extent trinary. The combination involving two beginners is the most common, except for projects of moderate and challenging difficulty and high priority. An equivalent analysis based on the location of mappers was not carried out due to noise caused by incomplete user profiles.

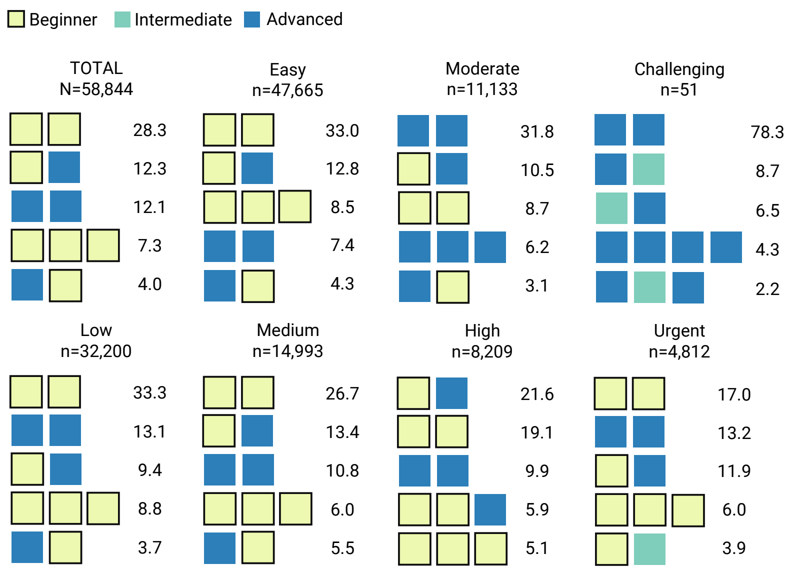

Fig. 10. Mapping level profile of mapping states for tasks mapped by more than one contributor -% of tasks-

Figure 11 elaborates on the dynamics of collaborative tasks using the “handover of work” concept [56]. In this figure, the nodes identify the mapping levels and the weights of the arcs between them are based on the proportion in which a handover occurs from one mapping level to another, 100% of the handovers leaving each mapping level to the others. The shaded area of the nodes shows the proportion of the total number of mapping states executed by each mapping level that belong to collaborative tasks.

As can be seen, most of the mapping states of novice contributors occur in collaborative tasks and that proportion decreases considerably for intermediate and advanced contributors. For the latter, task mapping is mostly solitary. Regardless of the mapping level, the proportion of collaborative mapping increases progressively with the urgency of the projects. In terms of difficulty, the proportion of collaborative tasks increases for beginners when they move from easy to intermediate projects, while for advanced users the proportion of collaborative tasks decreases with increasing difficulty. The arcs suggest that the handover from beginner and advanced users is to contributors of the same mapping level. For intermediate contributors, the results are mixed.

The frequency of handovers between groups may simply reflect the overall volume of mapping activities carried out by each group within the respective project categories, rather than indicating a greater or lesser tendency for interaction between groups. To account for this, the Table 5 compares the observed handover frequencies with the expected values based on the proportional distribution of mapping states between groups. These expected values are estimated using

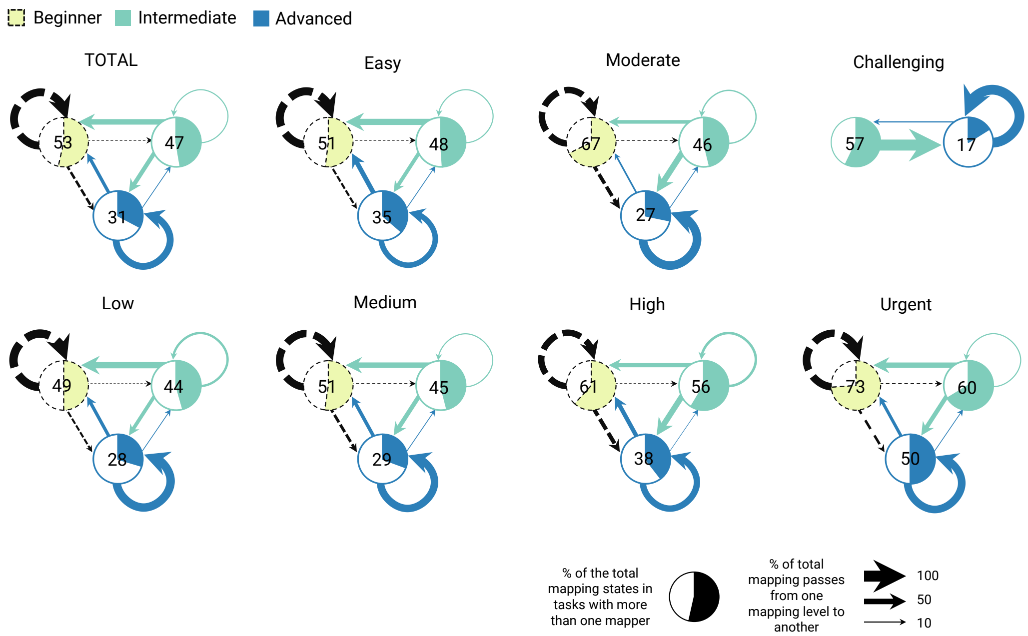

Fig. 11. Handover of mapping tasks

the combined frequencies of the Locked for mapping and Auto-unlocked for mapping states of each group, as presented in Table 3. It can be observed that handovers between novice users are substantially higher than expected to the detriment of handovers from an advanced user to a novice user which is consistently lower than expected.

The overall picture not only reflects a scenario where mapping is primarily driven by individual contributions (collection) but also highlights a limited level of meaningful collaboration. By limited we do not only refer to the lower frequency of collaborative task execution but to the nature of these interactions. Collaboration seems to be the result of novice mappers leaving tasks incomplete or being unsure whether their work is sufficiently accurate, thus handing over responsibility for completion and decision to another mapper. This handover often comes at the cost of delaying the progression of the task by one or more days. This is in stark contrast to the behaviour of advanced users who try to complete tasks individually, helping to dispatch tasks quickly and without friction. In this context, collaboration - defined as the involvement of several mappers to complete a task - seems to indicate inefficiency rather than representing a desirable behaviour.

5.2.3 RQ3: What evidence of intelligent action can be identified in HOT-TM mapping projects?

So far, performance considerations have focused on aspects such as the complexity of the mapping process in terms of sequence, duration and interaction, as reflected in task states. However, these indicators alone do not directly reveal whether mapping activities have been carried out properly. Therefore, the following section aims to identify factors that contribute to the success of mapping outcomes, providing potential evidence of intelligent collective action.

We performed a logistic regression to determine the effect of the task characteristics on the first validation result, that is, to determine if there are factors that make a task more likely to be validated or invalidated. As mentioned in the

Handover	TOTAL N=97,847 Obs.	Exp.	Diff.	Easy n=77,508 Obs.	Exp.	Diff.	Moderate n=20,288 Obs.	Exp.	Diff.	Challenging n=51 Obs.	Exp.	Diff.
Beginner - Advanced	16	21	-6	16	19	-3	14	20	-6			
Beginner - Beginner	49	28	21	55	39	16	25	11	14			
Beginner - Intermediate	4	4	1	5	4	0	3	2	1			
Intermediate - Advanced	3	3	0	2	2	0	4	4	0	8	1	6
Intermediate - Beginner	4	4	0	4	4	-1	3	2	1			
Intermediate - Intermediate	1	0	0	1	0	0	1	0	1			
Advanced - Advanced	15	16	-1	9	9	-1	39	38	1	82	97	-15
Advanced - Beginner	7	21	-14	7	19	-12	7	20	-13			
Advanced - Intermediate	2	3	-1	1	2	-1	4	4	0	10	1	8

Handover	Low n=49,058 Obs.	Exp.	Diff.	Medium n=24,470 Obs.	Exp.	Diff.	High n=13,868 Obs.	Exp.	Diff.	Urgent n=10,451 Obs.	Exp.	Diff.
Beginner - Advanced	13	22	-9	17	22	-4	26	21	4	15	20	-5
Beginner - Beginner	54	29	25	45	21	25	41	25	16	45	37	7
Beginner - Intermediate	4	3	1	4	3	1	4	4	0	5	4	1
Intermediate - Advanced	2	2	0	3	3	0	4	3	1	3	2	1
Intermediate - Beginner	3	3	0	4	3	1	3	4	-1	4	4	-1
Intermediate - Intermediate	1	0	0	1	0	0	1	1	0	1	1	0
Advanced - Advanced	15	16	-1	15	23	-7	11	18	-7	18	10	7
Advanced - Beginner	7	22	-15	8	22	-13	8	21	-13	8	20	-12
Advanced - Intermediate	2	2	-1	2	3	-1	2	3	-2	2	2	0

Table 5. Observed and expected frequency of handovers -% of handovers-

methodology, the dataset includes states prior to the first validation. Tasks with Split, Auto-unlocked for mapping and Auto-unlocked for validation states were discarded to remove noise. This results in a total of 281,653 tasks.

Table 6 lists each of the independent variables along with their regression coefficient, standard error, z statistic, and p-values. Larger regression coefficients indicate higher probabilities of invalidation. However, since these numbers are often not very intuitive, they are accompanied by odds ratios -OR-. The odds ratios represent an exponential transformation of the regression coefficient, implying a multiplication factor of the dependent variable for each one-unit increase in an independent variable. For each categorical variable where hot encoding was applied, the proportion of the total number of tasks belonging to that category is shown in parentheses, along with the average invalidation rate of the group, including the reference categories. For the numerical variables, Relative Validation Time and Building Area Mapped, the invalidation rates for tasks below and above the median for each indicator are displayed.

Starting with the effect of the mapping level of the mapper who marked the task as mapped, used as an indicator of experience, tasks are less likely to be invalidated when declared as mapped by more advanced mappers. Continuing with the effect of mapper location, the highest performance is observed among mappers from different hubs, followed by national mappers. However, this result should be interpreted with caution, as the unknown category comprises an unidentified mix of origins, and the location data is skewed toward expert mappers due to their more complete profiles. As for the effect of the involvement of several mappers, used as an indicator of interactions, the probability of task invalidation increases. Among the control factors, higher difficulty levels, higher priority levels and larger building area mapped are associated with higher invalidation rates. Finally, longer relative validation time tends to correspond to lower probabilities of task invalidation.

	N=281,653	% Inval.	β	OR	S.E	z	p value	
	(Intercept)	-	-2.30	0.10	0.03	-83.21	<2e-16	***
Experience of mappers	Mapped [Beginner] yes (vs. no) (44.7%)	5.9	-	-	-	-	-	-
	Mapped [Intermediate] yes (vs. no) (5.9%)	4.9	-0.59	0.55	0.04	-14.52	<2e-16	***
	Mapped [Advanced] yes (vs. no) (49.3%)	3.1	-1.15	0.32	0.02	-46.46	<2e-16	***
Location of mappers	Mapped [Unknown] yes (vs. no) (41.9%)	5.5	-	-	-	-	-	-
	Mapped [Same Country] yes (vs. no) (14.5%)	4.4	0.23	1.26	0.03	7.27	3.7e-13	***
	Mapped [Same Hub] yes (vs. no) (7.8%)	6.8	0.35	1.42	0.03	10.40	<2e-16	***
	Mapped [Different Hub] yes (vs. no) (35.8%)	2.9	-0.36	0.70	0.02	-14.96	<2e-16	***
Collaborative mapping	Multiple Mappers no (vs. yes) (75.3%)	3.0	-	-	-	-	-	-
	Multiple Mappers yes (vs. no) (24.7%)	8.8	0.44	1.55	0.02	21.72	<2e-16	***
Difficulty	Difficulty [Easy] yes (vs. no) (72.0%)	3.7	-	-	-	-	-	-
	Difficulty [Moderate] yes (vs. no) (27.6%)	6.4	0.66	1.93	0.02	27.47	<2e-16	***
	Difficulty [Challenging] yes (vs. no) (0.3%)	12.9	1.30	3.67	0.10	12.38	<2e-16	***
Priority	Priority [Low] yes (vs. no) (54.5%)	3.1	-	-	-	-	-	-
	Priority [Medium] yes (vs. no) (26.9%)	4.0	0.21	1.23	0.02	8.37	<2e-16	***
	Priority [High] yes (vs. no) (11.9%)	5.6	0.46	1.58	0.03	15.48	<2e-16	***
	Priority [Urgent] yes (vs. no) (6.7%)	15.5	1.16	3.19	0.03	40.48	<2e-16	***
	Building Area Mapped log([Square meters of buildings]+1)	<M 3.5, >M 5.4	0.06	1.06	0.00	23.72	<2e-16	***
	Relative Validation Time % Validation time/ (Validation time + Mapping time)	<M 7.4, >M 1.5	-0.03	0.97	0.00	-74.18	<2e-16	***

*, **, *** significant at p ≤ 0.10, 0.05 and 0.01 respectively.
M: median.

Table 6. Logistic Regression for validation state of a task —Validated vs. Invalidated—

This result aligns with the findings from previous sections, suggesting that the outcome of the system is not casual but is directly shaped by the group composition and the dynamics of collective action. Advanced mappers are not only more active, report less friction in the process and lead mapping and validation decisions but their contributions tend to better mapping results. This is evidenced not only by higher validation rates but also by longer validation times associated with fewer invalidations that are likely to reflect corrective work on their part. These results also highlight the challenges posed by the contribution dynamics of beginners. It is not just that their validation success rates are lower. In easy and low-priority projects, where their contributions on mapping are concentrated, the relatively longer time invested by validators - presumably for corrective work - seems to be necessary to compensate for the performance of beginners. Finally, evidence of the shortcomings of collaborative mapping, defined as the involvement of several mappers to complete a task, is reinforced. Collaborative mapping not only slows down the process but also is associated with further invalidation.

6 DISCUSSION

Our discussion focuses on synthesising and combining the findings on the collective intelligence system in HOT-TM projects with arguments from related work to present actionable insights and contributions to the field. It also reflects on possible ways to address the limitations identified in the current system set-up.

6.1 Towards a More Sustainable Group Composition of Humanitarian Mappers: Developing Novice Mappers and Enhancing Local Participation

The results of our study provide an overview of the group composition of HOT-TM mappers based on the key attributes of experience and location. Broadly speaking, this characterisation is in line with the profile of VGI participants derived from previous research, although previous studies tend to address this issue in a more fragmented way. By incorporating insights into the processes and outcomes of the collective intelligence system, our study sheds light on how group composition both influences and is influenced by these two key components of the system.

Starting with the attribute of expertise, which appears to have the greatest impact on the current scheme, HOT-TM reproduces the pattern of a group characterised by a small proportion of advanced contributors outperforming a majority of beginners both in the relative quantity and quality of their contributions and in the stability of their participation over time [2, 40, 54, 59]. Our findings on how decision-making functions are articulated in the mapping, together with validation rates on projects of varying levels of difficulty and priority, suggest that the system is designed to capitalise on the 'wisdom' of advanced contributors. This is achieved through deliberate decisions - such as restricting the validation and mapping of more complex projects to experienced mappers - or through presumably more spontaneous behaviour, such as corrective actions by validators to address the contributions of beginners. Meanwhile, beginners are largely confined to easier, lower priority projects and have limited involvement in decision-making processes. Our study reveals that this design has at least one major cost: the slowness of the validated mapping results. This is evident from our analysis of the duration and limited availability of the advanced users. Given that some of these data are used for active disaster response efforts, this delay represents a substantial inconvenience rather than an insignificant trade-off.

Our research also complements the work of Yin et al. [60] by providing more detailed findings. While they use macro-level project statistics to suggest that the HOT-TM micro-task model reflects the dependence on user authority and self-reinforcement characteristic of OSM communities [47], our study offers a more operational perspective on these dynamics. By following this approach, the system does not encourage long-term retention or support the skills development of beginners. The marginal presence of intermediate mappers, both in terms of their proportion within the group and their overall contribution to the mapping process and its outcomes, that we detected is further evidence of this lack of investment in the development of a cadre of more experienced mappers.

In light of the above, fostering the development of entry-level mappers through a strategy that supports their retention and progression to higher levels of performance presents a promising opportunity to build a more sustainable community. This strategy could take the form of a structured 'career path' that integrates short and long-term actions.

The tendency of newcomers to disappear after contributing to a single project, combined with signs of uncertainty and hesitation as they map, suggests that an unsatisfactory user experience could contribute to their lack of return. In this respect, an onboarding process designed to create a more rewarding first mapping experience could play a key role. Regular mapathons - both physical and virtual - that facilitate interaction between mappers and foster a sense of community provide an ideal framework for welcoming newcomers. In addition, the user interface should guide users through a navigation flow that encourages exploration of relevant sections and ensures a satisfactory mapping experience. Simple actions, such as asking users to complete their profile information, could foster a greater sense of identity as contributors. However, the most significant impact lies in empowering these users to gain confidence in completing tasks and making mapping decisions. HOT-TM already provides comprehensive materials on how to edit maps effectively. These resources could be reinforced with contextual hints to help novice mappers discern when a task is ready to be marked as complete, when it requires splitting, or when it should be reported due to poor imagery.

In the medium term, gamification elements could be explored to encourage novice users to return and gradually move on to more complex mapping tasks, as suggested by the work of Watkinson et al. [57]. This is in line with the findings of Urrea and Yoo [54] who highlight the positive effect of reaching new levels of experience, especially for novice users, and suggest that further segmentation of ranks based on experience could improve retention. Detailed data on mapper profiles - including demographics, preferences and performance - can help implement a reporting system that not only assigns online volunteers to the projects where they are most productive but also matches their career progression. In terms of long-term retention strategies, it is important to recognise the central role of altruism as a motivator for participation in OSM [6], and in humanitarian mapping in particular [62]. It is therefore convenient to actively communicate to mappers the tangible impact of the projects to which they contribute and, wherever possible, to highlight their individual contributions to relief and disaster prevention efforts.

Continuing with the locality factor, it is important to note that our work can be seen as a checkpoint in assessing progress towards the goal of developing sustainable, locally owned community mapping ecosystems in at-risk regions around the world, as proposed by Soden and Palen [46] at the beginning of academic research on HOT. Our findings indicate that the majority of contributions come from mappers outside the local hub, located in the Global North. In second place, are contributions from national mappers. It further reinforces the contradiction - already noted in other mapping contexts [20, 28, 51] - that locality, the founding principle of the VGI concept [14], seems to be relegated to the background.

As already mentioned, the main initiative to decentralise the organisation and incorporate local expertise in project management has been the creation of Open Mapping Hubs.[9] However, these initiatives are relatively recent, as the current four hubs were created between 2021 and 2023. Presumably, these hubs are still in the development phase, so their impact may not yet be fully reflected in the mapping results. In this context, the results presented here can serve as a reference for future assessments of the impact of the hubs as they become more mature. However, there are promising signs. An example of this is the fact that mappers from the hub in the area of interest are the second largest group, after mappers not affiliated to any hub, in almost all regions. This highlights the presence of a critical mass of local mappers that could be tapped to further develop HOT hubs. Additionally, the contribution of national mappers far exceeds their proportion of the group, and is particularly pronounced in the lower priority projects and projects with fewer tasks. This pattern suggests a possible specialisation in preventive, preparatory and follow-up mapping activities.

[9] https://www.hotosm.org/hubs/

In addition to actions at the organisational governance level, technical solutions must also be part of the approach. The current design of HOT-TM is optimised for desktop use but is not well suited to field operations. However, new tools for localised mapping are being developed, such as the Field Mapping Tasking Manager (FMTM),[10] a tool for organising ground mobile survey mapping, which is currently in beta. These recent innovations point in the right direction but it is still too early to assess their full impact, which warrants continuous monitoring and evaluation.

6.2 Towards a More Meaningful Collective Action in Humanitarian Mapping: Facilitating Effective Collaboration Among Mappers

In examining the dimensions of collective action, HOT-TM seems to heavily prioritise collection over collaboration at the level of the mapping task. Indeed, if we adopt a loose definition of collaboration - i.e. the mere contribution of several mappers to complete a task [35, 36] - it tends to be more of an anomaly than a beneficial factor in achieving mapping goals. On the contrary, it tends to lead to rework, slows down the progress and reveals systemic problems in the user experience for beginners. In addition, collaborative instances tend to show worse validation performance. Thus, the expected improvement effect associated with the 'wisdom of crowds' [50] concept does not seem to emerge, at least at the task level. If the main objective is to cover large territories quickly, a contributor should ideally complete a task independently within a few minutes, without the need for subsequent intervention by other mapper.

The analysis shows that the interactions between novices and experts are mainly characterised by corrections made by the latter as validators of the work of the former, a pattern that is also observed in the general dynamics of OSM [36]. The correction process in OSM usually occurs discretely [29], which means that novice mappers receive little or no direct information about the specific reasons for a correction. In HOT-TM, contributors are notified when a task they have mapped is validated or invalidated, provided that they have enabled notifications. However, the notification alone does not guarantee that users receive meaningful feedback on their mapping performance, especially if the validators take on the correction on their own. This lack of guidance may hinder their development as mappers and reinforce existing group dynamics.

These results support the concerns raised by Kogan et al. [29] regarding the limitations of relying solely on co-editing to study interactions between mappers. In this sense, the lack of visibility in discussion channels in OSM [29] is also a blocking factor present in HOT-TM. Although HOT-TM includes a comments section for each task, its visibility is somewhat limited, and it may not always offer clear guidance to subsequent mappers. As a result, understanding why a previous contributor left a task incomplete or identifying aspects that still require attention can be challenging. Enhancing the system with a more actionable task history could support more effective collaboration by helping mappers complete tasks more efficiently and enrich the mapped elements.

The challenge, therefore, lies in fostering interactions between novice and advanced mappers that facilitate knowledge transfer. While this type of mentoring interaction often occurs organically during face-to-face mapping events [29], there is potential to further develop mechanisms that provide constructive and accessible feedback in response to validation outcomes. These mechanisms should avoid intimidating styles that could discourage future contributions and instead create a supportive environment that encourages continued engagement and growth.

It is also worth considering whether the intrinsic nature of current mapping tasks requires and encourages collaboration during their execution. Where tasks are limited to adding missing elements, such as buildings or roads, it is understandable that the involvement of additional mappers would not significantly enrich the outcome. In this context,

[10] https://fmtm.hotosm.org/

diversification of task types could offer valuable improvements. For example, enrichment tasks could complement creation tasks, in line with the specialisation patterns identified by Zhang et al. [61]. This approach could support a layered model in which co-editing adds significant value to map content, helps to close the detail gap that distinguishes humanitarian mapping from other types of OSM mapping [21], and further emphasises the importance of local contributions. Note that national mappers now appear to be significantly active but this is not necessarily associated with greater success in validation.

6.3 Towards More Intelligent Collective Action in Humanitarian Mapping: Fostering Greater Equity Without Compromising Productivity

As noted since the introduction of the concept of collective intelligence, defining intelligence can be challenging, as it often depends on the objectives of the observer [32]. For example, if we focus solely on the goal of productivity (covering territory and validating tasks), this perspective fits well with the urgent needs of first response efforts during an ongoing emergency. Although no formal benchmark exists, an overall task invalidation rate of 6.3% and 15.5% for urgent projects suggest that the current system is relatively efficient in producing an initial mapping output from mapping operations that typically take only a few minutes. However, the need for corrective work by validators and the waiting period of several days for validation highlight potential areas for improvement in the quality of this initial output.

If we shift the focus from productivity to equity-related objectives, the limitations of the current approach become more apparent. Equity-oriented objectives could include fostering the growth of the mapping community, promoting the integration of local communities, and, most importantly, avoiding unsustainable mapping outcomes—such as the creation of low-quality data that are not maintained and quickly lose relevance [21, 60]. Note that these other objectives are better suited to the needs of prevention, preparedness and long-term monitoring of emergencies.

Our results and observations align with and extend the findings of Yin et al. [60], who highlight that in HOT-TM, as in other micro-tasking tools used in peer production, there is an inherent trade-off between productivity and equity. This may raise questions about how to address the tension between productivity and equity objectives. In this sense, it would be beneficial to allow project managers to define the most appropriate mapping objectives - whether focused on productivity or equity - before the project is launched, depending on the specific needs and timing of the emergency. These objectives could then be linked to tasks aimed at creating or enriching the map, as appropriate. In addition, targeted recruitment efforts to assign online volunteers to projects where their contributions will have the greatest impact could further support this approach, as suggested by Urrea and Yoo [54].

The review of objectives must be accompanied by a corresponding review of system metrics, especially given the strong influence these metrics often have on system incentives. As Spielman [47] suggests, if there is any kind of spatial collective intelligence, it should be reflected in map quality measures focused on the credibility and accuracy of the output. In HOT-TM, the most direct indicator of the success of a task is achieving validation. This is a credibility metric, since it comes from the status of the validators. However, this metric aligns with the dynamics of authority and self-reinforcement, as discussed above. According to Spielman [47], ideally, credibility and accuracy complement each other. This opens up an opportunity to introduce new metrics that are less dependent on the reputation of validators and better aligned with evolving mapping objectives.

When discussing the concept of intelligence, it is pertinent to consider opportunities to better harness the 'wisdom of crowds' [50]. We have identified a validation bottleneck linked to the relative scarcity of validators, as this task is mainly limited to the most experienced users. In addition, the decision-making process is always individual rather than collective. A conventional approach to address this problem is to increase the visibility of projects requiring validation to potential validators. However, a strategy more in line with the principle of collective intelligence would be to broaden the pool of collaborators authorised to validate, potentially including less experienced mappers. While novice mappers alone may lack the expertise to accurately evaluate a task, the collective input of a large number of novice users can yield reliable results. Voting or averaging mechanisms that have proven effective in other contexts could be explored in this case [50]. To this end, validation tasks could be further fragmented, making them more accessible to novice users. MapSwipe,[11] another micro-task platform within the OSM humanitarian ecosystem, has already implemented simple, fragmented validation. However, this functionality is still limited to pilot projects and should be extended to more HOT-TM projects to maximise its potential.

To conclude our reflection, we turn to the implications that our findings may have for the presumably increasing incorporation of artificial intelligence (AI) tools in humanitarian mapping tasks. By introducing the concept of collective intelligence, it is clear that computer agents are an integral part of this equation [31, 33]. Studies such as those by Tipnis et al. [52] on the introduction of Rapid show that while AI tools can improve productivity, they can also exacerbate differences in participation between novice and experienced users, an outcome that does not necessarily align with fostering a more sustainable community.

In this context, our findings highlight several areas where AI can make a sustainable contribution. On the one hand, further productivity gains can be achieved by anticipating and mitigating events that slow down mapping workflows, especially in high-priority or more complex projects where such interruptions are more frequent. One such approach could be to predict split-prone areas and perform a preliminary split during project setup. In addition, novice users could be guided to complete tasks more efficiently by avoiding repeated self-unlocking and marking tasks as completed in a single cycle. In some cases, an extra minute of guidance for a mapper could speed up the completion of a task by a whole day. However, the greatest potential lies in the strategic use of AI to support equity-based goals. Initiatives such as gamified progression through task difficulty levels, personalised notifications, layered orchestration of task assignments with enrichment goals, and collective validation through aggregated contributions would be overwhelming for human managers to coordinate on their own. In the framework of computational participation in collective intelligence systems [31], the use of AI to implement these strategies would redefine the role of computational agents from mere tools to active assistants and managers.

7 CONCLUSIONS

This case study illustrates the value of analysing collaborative production environments through the lens of collective intelligence. Adopting this perspective fosters a more systemic understanding, revealing that the components of the framework are not independent but interdependent, continually influencing each other. For instance, the composition of a given group can influence initial decisions about task design and performance evaluation in crowdsourcing activities, which in turn can create incentives that reinforce the original composition of the group.

The collective intelligence framework has not only proven effective in coherently organising the numerous quantitative results generated in this study but has also served as a unifying lens to synthesise evidence from previous notable research efforts in the field of VGI. The dispersed nature of these studies often makes it difficult for systemic perspectives to emerge. However, the main advantage of this framework lies in its ability to support the formulation of structural recommendations for improving system intelligence. As Malone et al. [31, 33] argue, understanding the

[11] https://mapswipe.org/

factors that influence collective intelligence allows system managers to intervene more effectively. By adjusting the fundamental components of a system, managers can significantly improve their collective intelligence, an advantage not usually achieved with individual intelligence, which tends to be much less adaptive.

This case serves as an invitation to future researchers, designers and managers of crowdsourcing platforms to consider the advantages offered by this framework. The availability of such diagnostics is a valuable asset in this era of rapid and widespread adoption of artificial intelligence, a phenomenon that undoubtedly also affects computer-supported collaborative work initiatives. As discussed above, careless adoption of these technologies can undermine rather than foster productive collaboration. With proper diagnosis it is possible to better guide interventions. In this sense, the basal literature on collective intelligence is rich in conceptual ideas [31, 33, 49]. However, this abundance of concepts is not always accompanied by detailed use cases that can guide practical application. In this sense, our study, while not aiming to establish a formal methodology, can provide a source of application ideas that can be adapted and replicated in other contexts. The key prerequisite is access to an equally rich source of data.

Data dependency serves as a call to crowdsourcing platform managers to follow the example of HOT in collecting comprehensive data on contributors, workflows and outcomes, as well as to share these data with relevant communities, ideally in open formats. While specific data fields and metrics may vary across crowdsourcing scenarios, this case study highlights the value of quantifying the components of a collective intelligence system. The general action lines of (1) describing relevant attributes of contributor profiles, (2) analysing the mechanics of collective action through techniques such as process mining, and (3) tracking factors associated with successful or unsuccessful outcomes, can be broadly applied to a variety of contexts. Finally, the specific recommendations proposed in this study to improve the sustainability of the HOT-TM mapping community may also be valuable for other peer production environments, given that the uneven distribution of human activity in collaborative efforts seems to be the norm rather than the exception [38].

Adopting such an ambitious framework as collective intelligence offers significant advantages but it comes with inherent challenges. Efforts to address these challenges, however comprehensive, can often seem insufficient. Some limitations become more apparent and, rather than solving issues, may raise more questions. However, we see this as a constructive outcome. Below we set out several key limitations and threats to validity of our study and identify corresponding opportunities for future research.

Starting with group composition, our analysis, based on attribute profiling, was sufficient to provide valuable information. In addition, we consider that the mapping level classification system used by HOT-TM, although simple, is effective in identifying distinct user behaviours. However, this approach may be somewhat simplistic and overlook broader or more nuanced patterns of behaviour. Future research could explore characterisations based on emergent group properties [13] or adopt mapper categorisation frameworks [61].

When analysing collective actions, it is important to keep in mind that HOT-TM operates on two interconnected levels: the OSM editor and the tasking manager. Our analysis focuses on what the states of the tasking manager can reveal about the behaviour of the mapper. This focus is a defining characteristic of this type of mapping compared to other mapping approaches. However, we do not ignore the importance of object-level edit histories, which provide a more granular view of behaviours and actions. We see the exploration of these detailed operations and their relationship to tasking manager states as a valuable direction for future research. For example, it would be valuable to measure the impact of micro-task structure—particularly task boundaries—on spatial interactions, such as the alignment and connection of adjacent or continuous elements [29]. This consideration highlights two more general limitations of our study. The first is our reliance primarily on HOT-TM API data, which were chosen for reasons of manageability. The incorporation of external data, such as OSM-level edit histories, could significantly enrich our analysis and provide a more complete understanding of mapping dynamics. The second limitation relates to the dominance of quantitative analysis. Qualitative research activities played a secondary role, serving to complement the interpretation and discussion of the results. Future research could deepen the understanding of HOT-TM mappers by employing qualitative approaches, such as those applied by Kogan et al. [29] or through usability testing to further explore the mapping experience within the task manager [22]. This approach could help clarify the nature of behaviours such as hesitation of beginners, shedding light on whether such behaviours stem more from intrinsic mapper factors or from the design of the mapping interface.

In assessing the intelligence of the system through its outputs, we have used validation as a credibility metric. However, this approach may not represent the most objective measure of system quality. In this regard, the extensive OSM literature on data quality offers opportunities to incorporate more objective indicators, such as positional accuracy of geographic data, attribute accuracy, data completeness and other dimensions of quality [8, 15]. Leveraging these indicators require further exploration of the OSM database to collect additional data. Another important limitation of the study is that it focuses on the impact of collaboration on task outcomes. A broader and more insightful approach would involve examining how individual tasks and their integration contribute to the overall quality of the total project area.

Other considerations include the volatility of platforms like HOT-TM and OSM, which are constantly evolving in terms of policies and user interfaces. This means that certain behaviours we observed may change suddenly with new updates, making them no longer directly comparable to what we have documented here. Additionally, we prioritised the analysis of the mapping level of contributors because it was accessible to all users, while other demographic segmentation criteria, such as the location of the mappers, received secondary treatment due to their limited availability. Information about mapping levels on the HOT-TM API was restricted to the time of data retrieval. Consequently, a contributor identified as advanced or intermediate may have been a beginner during their participation in certain project. To reduce this impact, we selected archived projects launched within a two-year period.

ACKNOWLEDGMENTS

This paper is partially supported by the Aragon Regional Government through the project T59_23R. The work of Dagoberto José Herrera-Murillo, Héctor Ochoa-Ortiz, and Umair Ahmed is supported by the ODECO project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 955569.

REFERENCES

[1] Jennings Anderson, Robert Soden, Kenneth M. Anderson, Marina Kogan, and Leysia Palen. 2016. EPIC-OSM: A Software Framework for OpenStreetMap Data Analytics. In *2016 49th Hawaii International Conference on System Sciences (HICSS)*. 5468–5477. https://doi.org/10.1109/HICSS.2016.675

[2] Daniel Bégin, Rodolphe Devillers, and Stéphane Roche. 2018. The Life Cycle of Contributors in Collaborative Online Communities-the Case of Openstreetmap. *International Journal of Geographical Information Science* 32, 8 (2018), 1611–1630. https://doi.org/10.1080/13658816.2018.1458312

[3] Suzanne T. Bell, Shanique G. Brown, Anthony Colaneri, and Neal B. Outland. 2018. Team Composition and the ABCs of Teamwork. *American Psychologist* 73 (2018), 349–362. https://doi.org/10.1037/amp0000305

[4] Amal Ben Rjab, Mouloud Kharoune, Zoltan Miklos, and Arnaud Martin. 2016. Characterization of Experts in Crowdsourcing Platforms. In *Belief Functions: Theory and Applications*, Jiřina Vejnarová and Václav Kratochvíl (Eds.). Springer International Publishing, Cham, 97–104. https://doi.org/10.48550/arXiv.1609.09748

[5] Daren C. Brabham. 2013. *Crowdsourcing*. The MIT Press.

[6] Nama R. Budhathoki and Caroline Haythornthwaite. 2013. Motivation for Open Collaboration: Crowd and Community Models and the Case of OpenStreetMap. *American Behavioral Scientist* 57, 5 (2013), 548–575. https://doi.org/10.1177/0002764212469364

[7] Youjin Choe, Martin Tomko, and Mohsen Kalantari. 2023. Assessing Mapper Conflict in OpenStreetMap Using the Delphi Survey Method. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 546, 17 pages. https://doi.org/10.1145/3544548.3580758

[8] Nicholas Chrisman. 1991. *The error component in spatial data*. Vol. 1.

[9] Martin Dittus, Giovanni Quattrone, and Licia Capra. 2016. Analysing Volunteer Engagement in Humanitarian Mapping: Building Contributor Communities at Large Scale. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (San Francisco, California, USA) *(CSCW '16)*. Association for Computing Machinery, 108–118. https://doi.org/10.1145/2818048.2819939

[10] Jean-Christophe Dubois, Laetitia Gros, Mouloud Kharoune, Yolande Le Gall, Arnaud Martin, Zoltan Miklos, and Hosna Ouni. 2019. *Measuring the Expertise of Workers for Crowdsourcing Applications*. Springer International Publishing, Cham, 139–157. https://doi.org/10.1007/978-3-030-18129-1_7

[11] Melanie Eckle and João Porto de Albuquerque. 2015. Quality Assessment of Remote Mapping in OpenStreetMap for Disaster Management Purposes. In *International Conference on Information Systems for Crisis Response and Management*. https://api.semanticscholar.org/CorpusID:4506679

[12] David Engel, Anita Williams Woolley, Ishani Aggarwal, Christopher F. Chabris, Masamichi Takahashi, Keiichi Nemoto, Carolin Kaiser, Young Ji Kim, and Thomas W. Malone. 2015. Collective Intelligence in Computer-Mediated Collaboration Emerges in Different Contexts and Cultures. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 3769–3778. https://doi.org/10.1145/2702123.2702259

[13] Martin G. Everett and Steve P. Borgatti. 1999. The Centrality of Groups and Classes. *The Journal of Mathematical Sociology* 23, 3 (1999), 181–201. https://doi.org/10.1080/0022250X.1999.9990219

[14] Michael F. Goodchild. 2007. Citizens as Sensors: The World of Volunteered Geography. *GeoJournal* 69, 4 (2007), 211–221. https://doi.org/10.1007/s10708-007-9111-y

[15] Michael F. Goodchild and Linna Li. 2012. Assuring the Quality of Volunteered Geographic Information. *Spatial Statistics* 1 (2012), 110–120. https://doi.org/10.1016/j.spasta.2012.03.002

[16] A. Yair Grinberger, Marco Minghini, Levente Juhász, Godwin Yeboah, and Peter Mooney. 2022. OSM Science—The Academic Study of the OpenStreetMap Project, Data, Contributors, Community, and Applications. *ISPRS International Journal of Geo-Information* 11, 4 (March 2022), 230. https://doi.org/10.3390/ijgi11040230

[17] Hans W. Guesgen and Jochen Albrecht. 2000. Imprecise Reasoning in Geographic Information Systems. *Fuzzy Sets and Systems* 113, 1 (2000), 121–131. https://doi.org/10.1016/S0165-0114(99)00016-0

[18] Muki Haklay. 2010. How good is OpenStreetMap information? A comparative study of OpenStreetMap and Ordnance Survey datasets for London and the rest of England. *Environment and Planning B: Planning and Design* 37 (2010), 682–703. https://doi.org/10.1068/b35097

[19] Yuyan Han and David Dunning. 2024. Metaknowledge of Experts Versus Nonexperts: Do Experts Know Better What They Do and Do Not Know? *Journal of Behavioral Decision Making* 37, 2 (2024). https://doi.org/10.1002/bdm.2375

[20] Brent J. Hecht and Darren Gergle. 2010. On the "Localness" of User-Generated Content. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work* (Savannah, Georgia, USA) *(CSCW '10)*. Association for Computing Machinery, New York, NY, USA, 229–232. https://doi.org/10.1145/1718918.1718962

[21] Benjamin Herfort, Sven Lautenbach, João Porto de Albuquerque, Jennings Anderson, and Alexander Zipf. 2021. The Evolution of Humanitarian Mapping Within the Openstreetmap Community. *Scientific Reports* 11 (2021). https://doi.org/10.1038/s41598-021-82404-z

[22] Dagoberto José Herrera-Murillo, Javier Nogueras-Iso, Paloma Abad-Power, and Francisco J. Lopez-Pellicer. 2023. User Interaction Mining: Discovering the Gap Between the Conceptual Model of a Geospatial Search Engine and Its Corresponding User Mental Model. In *Perspectives in Business Informatics Research*, Knut Hinkelmann, Francisco J. López-Pellicer, and Andrea Polini (Eds.). Springer Nature Switzerland, Cham, 3–15. https://doi.org/10.1007/978-3-031-43126-5_1

[23] Dagoberto José Herrera-Murillo, Héctor Ochoa-Ortiz, Umair Ahmed, Francisco Lopez-Pellicer, Barbara Re, Andrea Polini, and Javier Nogueras-Iso. 2024. Process Analysis in Humanitarian Voluntary Geographic Information: the case of the HOT Tasking Manager. *AGILE: GIScience Series* 5 (2024), 1–12. https://doi.org/10.5194/agile-giss-5-5-2024

[24] Jeff Howe. 2006. The Rise of Crowdsourcing. *Wired Magazine* 14, 6 (2006).

[25] Xiao Huang, Siqin Wang, Di Yang, Tao Hu, Meixu Chen, Mengxi Zhang, Guiming Zhang, Filip Biljecki, Tianjun Lu, Lei Zou, Connor Y. H. Wu, Yoo Min Park, Xiao Li, Yunzhe Liu, Hongchao Fan, Jessica Mitchell, Zhenlong Li, and Alexander Hohl. 2024. Crowdsourcing Geospatial Data for Earth and Human Observations: A Review. *Journal of Remote Sensing* 4 (2024), 0105. https://doi.org/10.34133/remotesensing.0105

[26] Yuanyuan Jiao, Yepeng Wu, and Steven Lu. 2021. The Role of Crowdsourcing in Product Design: The Moderating Effect of User Expertise and Network Connectivity. *Technology in Society* 64 (2021). https://doi.org/10.1016/j.techsoc.2020.101496

[27] Isaac L. Johnson, Yilun Lin, Toby Jia-Jun Li, Andrew Hall, Aaron Halfaker, Johannes Schöning, and Brent Hecht. 2016. Not at Home on the Range: Peer Production and the Urban/Rural Divide. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 13–25. https://doi.org/10.1145/2858036.2858123

[28] Isaac L. Johnson, Subhasree Sengupta, Johannes Schöning, and Brent Hecht. 2016. The Geography and Importance of Localness in Geotagged Social Media. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 515–526. https://doi.org/10.1145/2858036.2858122

[29] Marina Kogan, Jennings Anderson, Leysia Palen, Kenneth M. Anderson, and Robert Soden. 2016. Finding the Way to OSM Mapping Practices: Bounding Large Crisis Datasets for Qualitative Investigation. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San

Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 2783–2795. https://doi.org/10.1145/2858036.2858371

[30] Andrew W Lo and Ruixun Zhang. 2022. The Wisdom of Crowds Versus the Madness of Mobs: An Evolutionary Model of Bias, Polarization, and Other Challenges to Collective Intelligence. *Collective Intelligence* 1, 1 (2022). https://doi.org/10.1177/26339137221104785

[31] Thomas W. Malone. 2018. *Superminds : The Surprising Power of People and Computers Thinking Together* (first edition. ed.). Little, Brown and Company, New York.

[32] Thomas W. Malone and Michael S. Bernstein. 2022. *Handbook of Collective Intelligence*. MIT press.

[33] Thomas W. Malone, Robert Laubacher, and Chrysanthos Dellarocas. 2010. The Collective Intelligence Genome. *Sloan Management Review* 51, 3 (2010), 21–31. http://www.lhstech.com/chair/Articles/malone.pdf

[34] David W. McDonald. 2011. Task Dependency and the Organization of the Crowd. In *ACM SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*.

[35] Peter Mooney and Padraig Corcoran. 2012. How social is OpenStreetMap?. In *Proceedings of the AGILE'2012 International Conference on Geographic Information Science* (Avignon, France). 282–287.

[36] Peter Mooney and Padraig Corcoran. 2014. Analysis of Interaction and Co-editing Patterns amongst OpenStreetMap Contributors. *Transactions in GIS* 18 (2014). https://doi.org/10.1111/tgis.12051

[37] Richard Moreland, John Levine, and Melissa Wingert. 2018. *Creating the Ideal Group: Composition Effects at Work*. Lawrence Erlbaum Associates, Inc., 11–35. https://doi.org/10.4324/9781315789293-2

[38] Lev Muchnik, Sen Pei, Lucas C. Parra, Saulo D. S. Reis, José S. Andrade, Shlomo Havlin, and Hernán A. Makse. 2013. Origins of Power-Law Degree Distribution in the Heterogeneity of Human Activity in Social Networks. *Scientific Reports* 3 (2013). https://doi.org/10.1038/srep01783

[39] Pascal Neis, Peter Singler, and Alexander Zipf. 2010. Collaborative mapping and Emergency Routing for Disaster Logistics - Case studies from the Haiti earthquake and the UN portal for Afrika. (Jan. 2010).

[40] Pascal Neis and Alexander Zipf. 2012. Analyzing the Contributor Activity of a Volunteered Geographic Information Project — The Case of OpenStreetMap. *ISPRS International Journal of Geo-Information* 1, 2 (2012), 146–165. https://doi.org/10.3390/ijgi1020146

[41] OpenStreetMap Wiki. 2025. Stats - OpenStreetMap Wiki. https://wiki.openstreetmap.org/wiki/Stats

[42] Christoph Riedl, Young Ji Kim, Pranav Gupta, Thomas W. Malone, and Anita Williams Woolley. 2021. Quantifying Collective Intelligence in Human Groups. *Proceedings of the National Academy of Sciences* 118, 21 (2021), e2005737118. https://doi.org/10.1073/pnas.2005737118

[43] Lucia Saganeiti, Federico Amato, Beniamino Murgante, and Gabriele Nolè. 2017. VGI and Crisis Mapping in an Emergency Situation. Comparison of Four Case Studies: Haiti, Kibera, Kathmandu, Centre Italy. *GEOmedia* 21, 3 (Aug. 2017). https://mediageo.it/ojs/index.php/GEOmedia/article/view/1445 Number: 3.

[44] Kjeld Schmidt and Liam Bannon. 1992. Taking CSCW seriously: Supporting Articulation Work. *Computer Supported Cooperative Work* 1 (03 1992), 7–40. https://doi.org/10.1007/BF00752449

[45] Shilad W. Sen, Heather Ford, David R. Musicant, Mark Graham, Os Keyes, and Brent Hecht. 2015. Barriers to the Localness of Volunteered Geographic Information. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 197–206. https://doi.org/10.1145/2702123.2702170

[46] Robert Soden and Leysia Palen. 2014. From Crowdsourced Mapping to Community Mapping: The Post-earthquake Work of OpenStreetMap Haiti. In *COOP 2014-Proceedings of the 11th International Conference on the Design of Cooperative Systems, 27-30 May 2014, Nice (France)*. Springer International Publishing, Cham, 311–326. https://doi.org/10.1007/978-3-319-06498-7_19

[47] Seth Spielman. 2014. Spatial Collective Intelligence? credibility, accuracy, and Volunteered Geographic Information. *Cartography and Geographic Information Science* 41 (03 2014), 1–10. https://doi.org/10.1080/15230406.2013.874200

[48] Oliver Stiemerling and Armin B. Cremers. 1998. The use of cooperation scenarios in the design and evaluation of a CSCW system. *IEEE Transactions on Software Engineering* 24, 12 (1998), 1171–1181. https://doi.org/10.1109/32.738345

[49] Shweta Suran, Vishwajeet Pattanaik, and Dirk Draheim. 2020. Frameworks for Collective Intelligence: A Systematic Literature Review. 53, 1 (2020), 36 pages. https://doi.org/10.1145/3368986

[50] James Surowiecki. 2004. *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*. Abacus. 295 pages.

[51] Jacob Thebault-Spieker, Aaron Halfaker, Loren G. Terveen, and Brent Hecht. 2018. Distance and Attraction: Gravity Models for Geographic Content Production. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173722

[52] Vinit Tipnis, Eunae Yoo, Gloria Urrea, and Fei Gao. 2024. AI-Powered Philanthropy: Effects on Volunteer Productivity. *Social Science Research Network* (2024), 1–24. https://doi.org/10.2139/ssrn.4701631

[53] Long Tran-Thanh, Sebastian Stein, Alex Rogers, and Nicholas R. Jennings. 2014. Efficient Crowdsourcing of Unknown Experts Using Bounded Multi-Armed Bandits. *Artificial Intelligence* 214 (2014), 89–111. https://doi.org/10.1016/j.artint.2014.04.005

[54] Gloria Urrea and Eunae Yoo. 2023. The Role of Volunteer Experience on Performance on Online Volunteering Platforms. *Production and Operations Management* 32, 2 (2023), 416–433. https://doi.org/10.1111/poms.13879

[55] Niels van Berkel and Henning Pohl. 2024. Collaborating with Bots and Automation on OpenStreetMap. *ACM Transactions on Computer-Human Interaction* (2024). https://doi.org/10.1145/3665326

[56] Wil Van der Aalst. 2016. *Process Mining: Data Science in Action*. Springer, NL.

[57] Kirsty Watkinson, Jonathan Huck, and Angela Harris. 2023. Using Gamification to Increase Map Data Production During Humanitarian Volunteered Geographic Information (VGI) Campaigns. *Cartography and Geographic Information Science* 50 (2023), 1–17. https://doi.org/10.1080/15230406.2022.2156389

[58] Anita Williams Woolley, Christopher F. Chabris, Alex 'Sandy' Pentland, Nada Hashmi, and Thomas W. Malone. 2010. Evidence for a Collective Intelligence Factor in the Performance of Human Groups. *Science* 330 (2010), 686 – 688. https://doi.org/10.1126/science.1193147

[59] Anran Yang, Hongchao Fan, and Ning Jing. 2016. Amateur or Professional: Assessing the Expertise of Major Contributors in OpenStreetMap Based on Contributing Behaviors. *ISPRS International Journal of Geo-Information* 5 (02 2016). https://doi.org/10.3390/ijgi5020021

[60] Yaxuan Yin, Longjie Guo, and Jacob Thebault-Spieker. 2024. Productivity or Equity? Tradeoffs in Volunteer Microtasking in Humanitarian OpenStreetMap. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1, Article 113 (apr 2024), 34 pages. https://doi.org/10.1145/3637390

[61] Bowen Zhang, Jennings Anderson, Dipto Sarkar, and Robert Soden. 2024. A Quantitative Approach to Identifying Emergent Editor Roles in Open Street Map. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 487, 14 pages. https://doi.org/10.1145/3613904.3641963

[62] Radim Štampach, Lukáš Herman, Jakub Trojan, Kateřina Tajovská, and Tomáš Řezník. 2021. Humanitarian Mapping as a Contribution to Achieving Sustainable Development Goals: Research into the Motivation of Volunteers and the Ideal Setting of Mapathons. *Sustainability* 13, 24 (2021). https://doi.org/10.3390/su132413991

A A CLOSER LOOK AT THE MAPPING PROCESS

This appendix provides a more detailed analysis of the mapping process, examining the frequency and duration of task states based on the specific project type.

Table 7 analyses the frequency of task states. Whereas the first column presents the absolute frequency of each state across the entire dataset, while the subsequent columns display the percentage of task coverage across all projects, as well as within different difficulty and priority categories. A darker tone indicates a higher coverage.

Task state	Frequency of states	TOTAL TASKS	Difficulty			Priority			
			Easy	Moderate	Challenging	Low	Medium	High	Urgent
	N=1,853,074	N=312,289	n=222,729	n=88,468	n=1,092	n=166,911	n=80,768	n=38,119	n=26,491
Locked for mapping	616,659	100	100	100	100	100	100	100	100
Locked for validation	379,266	100	100	100	100	100	100	100	100
Mapped	349,713	100	100	100	100	100	100	100	100
Validated	330,290	100	100	100	100	100	100	100	100
Split	70,612	14	10	21	54	10	9	20	43
Auto unlocked for mapping	62,632	9	8	12	15	7	5	12	33
Invalidated	29,334	6	5	9	17	4	4	10	23
Bad imagery	7,562	2	2	2	0	1	3	2	2
Auto unlocked for validation	3,863	1	1	1	0	1	1	1	2
Extended for mapping	3,143	0	0	0	1	0	0	0	0

Table 7. Task states according to frequency and case coverage

Overall, the results indicate that the most frequent state is Locked for mapping, where the creation activity takes place, typically requiring an average of two cycles per task. The Auto-unlocked for mapping state, which is also linked to map creation, appears in almost one tenth of the tasks. The frequencies of states associated with decision-making at the end of the mapping phase suggest that tasks are generally declared as Mapped only once, while the option Split is used infrequently, and the Bad imagery state is rarely selected by mappers. The average frequency of the Locked for validation and Validated states per task, combined with the low case coverage of the Invalidated states (6.3%), indicates that the validation phase is generally straightforward. Other states, such as Auto-unlocked for validation and Extended for mapping, occur only marginally, regardless of the type of project.

Table 8 expands on the duration of different states and transitions. Regarding the relative duration of mapping versus validation, the top section of the table presents the median duration (in minutes) of the Locked for mapping and Locked for validation states for the different project types.

		Difficulty			Priority			
	TOTAL	Easy	Moderate	Challenging	Low	Medium	High	Urgent
Locked for mapping (minutes)	2.1	2.0	2.9	7.5	1.6	2.5	2.9	4.8
Locked for validation (minutes)	4.1	5.2	2.8	1.8	4.0	4.3	4.9	3.8
% of total tasks where total validation time is higher than total mapping time	57.5	59.6	51.4	21.1	61.4	57.9	50.4	27.5
Locked for mapping -> Locked for mapping (days)	1.0	1.1	0.8	3.0	0.8	1.7	1.5	0.2
Mapped -> Locked for validation (days)	13.2	25.1	3.2	3.8	27.0	6.8	3.8	4.7
Invalidated -> Unlocked for mapping (days)	1.0	1.4	0.6	74.4	1.0	1.8	0.8	0.4

Table 8. Median duration of task states and transitions (85% of most frequent traces)

Our observations show that as the difficulty of the project increases, the median time required for mapping also increases, while the median time required for validation decreases. Furthermore, as the priority level increases, the median duration of the Locked for mapping state increases, while the Locked for validation state remains relatively stable. The middle section of the table illustrates the percentage of tasks in which the total duration of the Locked for validation states exceeds that of the Locked for mapping states. In just over half of the tasks - except for challenging and urgent projects - the validation time exceeds the mapping time. In the case of easier projects, it appears that validation activities go beyond mere verification, often taking on a greater burden of correcting the work done during the mapping phase. Regarding the cost of waiting, the lower section of the table shows the median waiting times between states, measured in days. The cost of waiting for another Locked for mapping by deciding not to declare a task as Mapped or Invalidated is maximised for challenging projects and minimised for urgent projects. Meanwhile the waiting time for a Mapped task to move to the Locked for validation state is more pronounced in easy and low priority project tasks and significantly reduced in other categories.

AUTHORS STATEMENT

Dagoberto José Herrera-Murillo

The last 5 author relevant publications are:

- Identifying the evolution of Open Government Data initiatives and their user engagement (Journal paper, DOI: 10.1109/ACCESS.2024.3414282)
- **Process Analysis in Humanitarian Voluntary Geographic Information: the case of the HOT Tasking Manager** (Conference paper, DOI: 10.5194/agile-giss-5-5-2024)
- User interaction mining: discovering the gap between the conceptual model of a geospatial search engine and its corresponding user mental model (conference paper, DOI: 10.1007/978-3-031-43126-5$_1$)
- Analysing User Involvement in Open Government Data Initiatives (conference paper, DOI: 10.1007/978-3-031-16802-4$_1$4)
- Towards a sustainable Open Data Ecosystem (conference paper, DOI: 10.26754/jjii3a.20227009)

D.J.H. has written papers in the Collective Intelligence, Process Analysis, and Human Interfaces domains. The paper "Process Analysis in Humanitarian Voluntary Geographic Information: the case of the HOT Tasking Manager" (marked in bold) is the most related one, as the current paper is a direct follow-up and extension of that conference paper. A detailed explanation is provided below in a separate subsection. Other papers from the author are not related to the current paper.

Héctor Ochoa-Ortiz

The last 5 author relevant publications are:

- **Process Analysis in Humanitarian Voluntary Geographic Information: the case of the HOT Tasking Manager** (conference paper, DOI: 10.5194/agile-giss-5-5-2024)
- Open data co-creation, a path for better public services (conference paper, URL: https://ceur-ws.org/Vol-3514/short66.pdf)
- Diet Map: Participatory Mapping Application for Specific Diets (conference abstract, DOI: 10.5194/ica-abs-6-4-2023)
- Perception of the filmed urban space: an analysis of the imaginaries of Madrid constructed through series and social networks (journal paper, DOI: 10.21138/bage.3294)
- Pedestrian routing of periodically changing areas using Volunteered Geographical Information (OpenStreetMap) (conference abstract, DOI: 10.5194/ica-abs-5-92-2022)

His currently submitted work is:

- The Commercial Landscape in OpenStreetMap: Why and how are they contributing to the project? (journal paper)

H.O. has written papers in the OpenStreetMap and VGI domain. The paper "Process Analysis in Humanitarian Voluntary Geographic Information: the case of the HOT Tasking Manager" (marked in bold) is the most related one, as the current paper is a direct follow-up and extension of that conference paper. A detailed explanation is provided below in a separate subsection. Other papers from the author are not related to the current paper.

Umair Ahmed

The last 5 author relevant publications are:

- BRYT: Automated Keyword Extraction for Open Datasets (Journal Paper, DOI: 10.1016/j.iswa.2024.200421)
- **Process Analysis in Humanitarian Voluntary Geographic Information: the case of the HOT Tasking Manager** (conference paper, DOI: 10.5194/agile-giss-5-5-2024)
- Towards Cross-Domain Linking of Data: A Semantic Mapping of Cultural Heritage Ontologies (conference paper, DOI: 10.1145/3657054.3657077)
- Reimagining open data ecosystems: a practical approach using AI, CI, and Knowledge Graphs (conference paper, URL: https://ceur-ws.org/Vol-3514/paper65.pdf)
- Efficient Water Quality Prediction Using Supervised Machine Learning (Journal Paper, DOI: 10.3390/w11112210

U.A. has written papers in the Artificial Intelligence and Collective Intelligence domain. The paper "Process Analysis in Humanitarian Voluntary Geographic Information: the case of the HOT Tasking Manager" (marked in bold) is the most related one, as the current paper is a direct follow-up and extension of that conference paper. A detailed explanation is provided below in a separate subsection. Other papers from the author are not related to the current paper.

Francisco Javier López-Pellicer

The last 5 author relevant publications are:

- Identifying the evolution of Open Government Data initiatives and their user engagement (Journal paper, DOI: 10.1109/ACCESS.2024.3414282)
- An empirical study of the limitations of minimum bounding boxes for defining the extent of geospatial resources: the use of DGGS and other alternatives for improving the performance of spatial searches (Journal paper, DOI: 10.1080/13658816.2024.2361274)
- Towards the Development of Interoperable Open Data Ecosystems: Harnessing the Technical, Semantic, Legal, and Organizational (TSLO) Interoperability Framework (Conference paper, DOI: 10.1145/3657054.3657160)
- **Process Analysis in Humanitarian Voluntary Geographic Information: the case of the HOT Tasking Manager** (Conference paper, DOI: 10.5194/agile-giss-5-5-2024)
- Effects of data time lag in a decision-making system using machine learning for pork price prediction (Journal paper, DOI: 10.1007/s00521-023-08730-7)

F.J.L. is one of two academic supervisors of D.J.H.. The paper "Process Analysis in Humanitarian Voluntary Geographic Information: the case of the HOT Tasking Manager" (marked in bold) is the most related one, as the current paper is a direct follow-up and extension of that conference paper. A detailed explanation is provided in a separate subsection. Other papers from the author are not related to the current paper.

Barbara Re

The last 5 author relevant publications are:

- **Process Analysis in Humanitarian Voluntary Geographic Information: the case of the HOT Tasking Manager** (conference paper, DOI: 10.5194/agile-giss-5-5-2024)
- FloBP: a model-driven approach for developing and executing IoT-enhanced business processes (journal paper, DOI: 10.1007/s10270-024-01150-8)
- Design and Development of a Digital Twin Prototype for the SAFE Project (conference paper, DOI: 10.1007/978-3-031-54712-6_7)
- A Methodology for the Analysis of Robotic Systems via Process Mining (conference paper, DOI: 10.1007/978-3-031-46587-1_7)
- A Flexible Approach to Multi-party Business Process Execution on Blockchain (journal paper, DOI: 10.1016/j.future.2023.05.006)

B.R. is the academic supervisor of H.O.. The paper "Process Analysis in Humanitarian Voluntary Geographic Information: the case of the HOT Tasking Manager" (marked in bold) is the most related one, as the current paper is a direct follow-up and extension of that conference paper. A detailed explanation is provided below in a separate subsection. Other papers from the author are not related to the current paper.

Andrea Polini

The last 5 author relevant publications are:

- BRYT: Automated Keyword Extraction for Open Datasets (journal paper, DOI: 10.1016/j.iswa.2024.200421)

- **Process Analysis in Humanitarian Voluntary Geographic Information: the case of the HOT Tasking Manager** (conference paper, DOI: 10.5194/agile-giss-5-5-2024)
- Towards Cross-Domain Linking of Data: A Semantic Mapping of Cultural Heritage Ontologies (conference paper, DOI: 10.1145/3657054.3657077)
- Towards value-creating and sustainable open data ecosystems: A comparative case study and a research agenda (journal paper, DOI: 10.29379/jedem.v13i2.644)
- Process-oriented knowledge management and learning in public administrations (journal paper, DOI: 10.1504/EG.2020.110615)

A.P. is the academic supervisor of U.A.. The paper “Process Analysis in Humanitarian Voluntary Geographic Information: the case of the HOT Tasking Manager” (marked in bold) is the most related one, as the current paper is a direct follow-up and extension of that conference paper. A detailed explanation is provided below in a separate subsection. Other papers from the author are not related to the current paper.

Javier Nogueras-Iso

The last 5 author relevant publications are:

- Identifying the evolution of Open Government Data initiatives and their user engagement (Journal paper, DOI: 10.1109/ACCESS.2024.3414282)
- An empirical study of the limitations of minimum bounding boxes for defining the extent of geospatial resources: the use of DGGS and other alternatives for improving the performance of spatial searches (Journal paper, DOI: 10.1080/13658816.2024.2361274)
- Discrete Global Grid Systems with quadrangular cells as reference frameworks for the current generation of Earth observation data cubes (Journal paper, DOI: 10.1016/j.envsoft.2023.105656)
- Approaches for the Clustering of Geographic Metadata and the Automatic Detection of Quasi-Spatial Dataset Series (Journal paper, DOI: 10.3390/ijgi11020087)
- **Process Analysis in Humanitarian Voluntary Geographic Information: the case of the HOT Tasking Manager** (Conference paper, DOI: 10.5194/agile-giss-5-5-2024)

J.N. is one of two academic supervisors of D.J.H.. The paper “Process Analysis in Humanitarian Voluntary Geographic Information: the case of the HOT Tasking Manager” (marked in bold) is the most related one, as the current paper is a direct follow-up and extension of that conference paper. A detailed explanation is provided below in a separate subsection. Other papers from the author are not related to the current paper.

Regarding the conference paper “Process Analysis in Humanitarian Voluntary Geographic Information: the case of the HOT Tasking Manager”

The paper "Process Analysis in Humanitarian Voluntary Geographic Information: the case of the HOT Tasking Manager" was submitted to the AGILE Conference and presented in June 2024. This earlier work performs a basic process mining exploration of three dimensions of humanitarian mapping tasks: control flow, time, organisation and outcome of mapping tasks.

The new work presented to TOCHI extends that work by incorporating the conceptual framework of collective intelligence to holistically analyse the humanitarian mapping system. In terms of analysis and results, this involved adding and integrating new findings on group, process and system outcomes, especially those that seek to describe the interaction between contributors at the micro-task level and the factors associated with task invalidation. The new discussion focuses on how the current organisation of collective intelligence affects the humanitarian mapping process.