Meta-moderation for crowdsourced disaster management and urban participatory applications

R.I. Ogie, M. Amirghasemi and P. Perez

SMART Infrastructure Facility, University of Wollongong, Wollongong, Australia
Email: rogie@uow.edu.au

Abstract: This paper presents the Enterprise Meta-moderation of Innovation (EMI) system, which was designed, developed, and deployed to drive innovation in MTR Corporation, Hong Kong. Although EMI was designed specifically for use in an organisation, the system can potentially be extended to serve other useful functions in society. Hence, an important aspect of this paper involves identifying and discussing key considerations that need to be addressed in order for such a system to be used in crowdsourced disaster management and other urban participatory applications. The study highlights the role of agent-based modelling in better addressing some of the socio-technical challenges identified.

Keywords: *Meta-moderation, agent-based modelling, crowdsourcing, disaster, urban participatory approach*

1. INTRODUCTION

The rapid advancement in ubiquitous technologies such as web 2.0 and smart mobile devices has provided a unique opportunity to harness user-generated contents for improving productivity, innovation and collective intelligence (O'reilly 2009). This is particularly appealing in a megacity context, where there are millions of digitally enabled residents who can potentially contribute in generating relevant contents. While many organisations and cities would like to maximise this opportunity, it can be quite challenging to sieve out the meaningful and most relevant content from the noise (Ogie et al. 2018). This is where the concept of meta-moderation proves useful. *Meta-moderation can be defined as a process whereby the moderation (review) of a user-generated content is subjected to a second layer of moderation in order to check that the initial review or ratings are fair, unbiased and consistent with expectations* (Lampe et al. 2014). In essence, meta-moderation promotes the notion that every judgement or content evaluation should be subject to further judgement or evaluation (Durante 2010). Meta-moderation is quite important because it has the potential to sieve out contributions that are irrelevant or inflammatory. In a typical distributed moderation system, the rating mechanism recognises, over time, those who are high quality contributors, assigning higher weightings to their contributions so that low quality or irrelevant contents are eventually suppressed or eliminated.

There are several benefits that can accrue from implementing a meta-moderation system. These include (i) ability to promote open conversations and obtain increased civic engagement in the delivery of public services; (ii) gain consensus and community intelligence on a phenomenon of common interest such as flood, epidemic, traffic congestion, political issue, etc.; (iii) enact norms of civility in online communities; (iv) minimise bias or partiality in online communities; (v) reduce information overload by suppressing noisy or low quality contents; (vi) gain increased diversity of ideas and synergistic explosion of creativity, in effect combining and extending ideas to achieve more creative outcomes (Klein 2011). Despite these benefits, the application of meta-moderation has remained limited to few peer-production systems such as *Slashdot* and *Reddit* (Klein 2011; Dosono and Semaan 2019).

Slashdot is a social news website that relies on a distributed meta-moderation mechanism to improve the diversity and quality of contents (Lampe et al. 2014). The Slashdot system comprises users who post stories, moderators who qualify or assign attributes to these stories (e.g., informative, insightful, normal, off topic, troll, redundant, interesting, funny) and a meta-moderation process to check that the moderations are fair and unbiased. Moderators are typically high quality contributors who have earned high reputation scores (aka karma) through active participation. Similarly, the MIT Centre for Collective Intelligence has developed *Deliberatorium*- a system for harnessing the community's collective wisdom in large-scale deliberations about complex societal problems (Klein 2011). The system relies on moderators to inspect and certify posts that have followed the formal guidelines. Certified posts can then be viewed, edited, commented, and rated by the broader community. A post that is not certified is provided with insight for alignment with guidelines. Here, the moderators are simply playing the role of ensuring that the posts follow a formal structure that is more meaningful to the community; the actual merits of the posts are judged and rated by the broader community. The outputs from Deliberatorium can be further improved through a proper meta-moderation process that adds a second layer of content rating to the previous judgement of community members.

In the following section, we present a meta-moderation system that was designed for use in an enterprise. Section 3 is devoted to discuss key considerations for using such a system in crowdsourced disaster management and other urban participatory applications. Finally, we highlight the different ways in which agent-based modelling (ABM) can contribute in ensuring that the system is able to deliver maximum benefits and meet new demands and expectations in a community-facing context as opposed to an organisational setting.

2. THE ENTERPRISE META-MODERATION OF INNOVATION

2.1. Overview of learning organisation application: MTR Corporation, Hong Kong

The Enterprise Meta-moderation of Innovation (EMI) is an application that provides meta-moderation functionalities for employees of an organisation to propose and review innovative ideas. EMI was originally developed by researchers at the SMART Infrastructure Facility, University of Wollongong, Australia in order to help the MTR Corporation, Hong Kong to pursue its goal of becoming a learning organisation. *A learning organisation is one that has purposefully developed its ability to continually learn, acquire requisite knowledge for transformation and rapidly innovate in order to survive and thrive in a rapidly changing business environment* (Thomas and Allen 2006). EMI has been serving the purpose of initiating and collectively nurturing innovative ideas in MTR since its deployment in 2018. Ideas generated by employees are initially assessed and ranked by *moderators*; the meta-moderation functionality then improves on this process through

an additional layer of moderation. In this extra level of moderation, a different set of assessors, known as *meta-moderators*, assess the previously performed moderations and provide ranking scores for the fairness (impartiality) and accuracy of the moderators.

2.2. How EMI works

EMI is implemented as a web application that requires user authorisation and authentication prior to performing any action. The system design is such that the authentication process can be seamlessly integrated with the organisation's Active Directory, enabling all users to gain access by providing their existing organisational credentials. All employees in the Active Directory can login to the system and are assumed to have an *Idea Submitter* role by default. For effective use of the application, each employee or staff member is assigned one of two different roles, namely (i) *idea submitter* and (ii) *administrators*. These roles are to be defined in the Active Directory system (using *groups*). Based on the role, each user will have a different set of permissions in the system. In effect, regular users (i.e. idea submitters) are allowed to (i) submit new ideas, (ii) moderate the ideas assigned to them, and (iii) meta-moderate other users' moderations. Administrators acquire all these permissions, in addition to (iv) user management and (v) tracking and reviewing all ideas. Once authorised and authenticated, each user is presented with appropriate navigation options based on their role. For instance, an Idea Submitter has access to Idea Submission, Moderation and Meta-Moderation pages. Each idea, once submitted, goes through multiple steps from submission, moderation, through to meta-moderation and finalisation. The idea submitter can track the status of his/her submitted idea at any stage in the process (see Figure 1). In order that this process proceeds in a timely manner, a remind/reassign routine has been devised and implemented. Figure 2 shows and explains the various stages involved in the process of initiating an idea right through to the finalisation of the idea for management action.

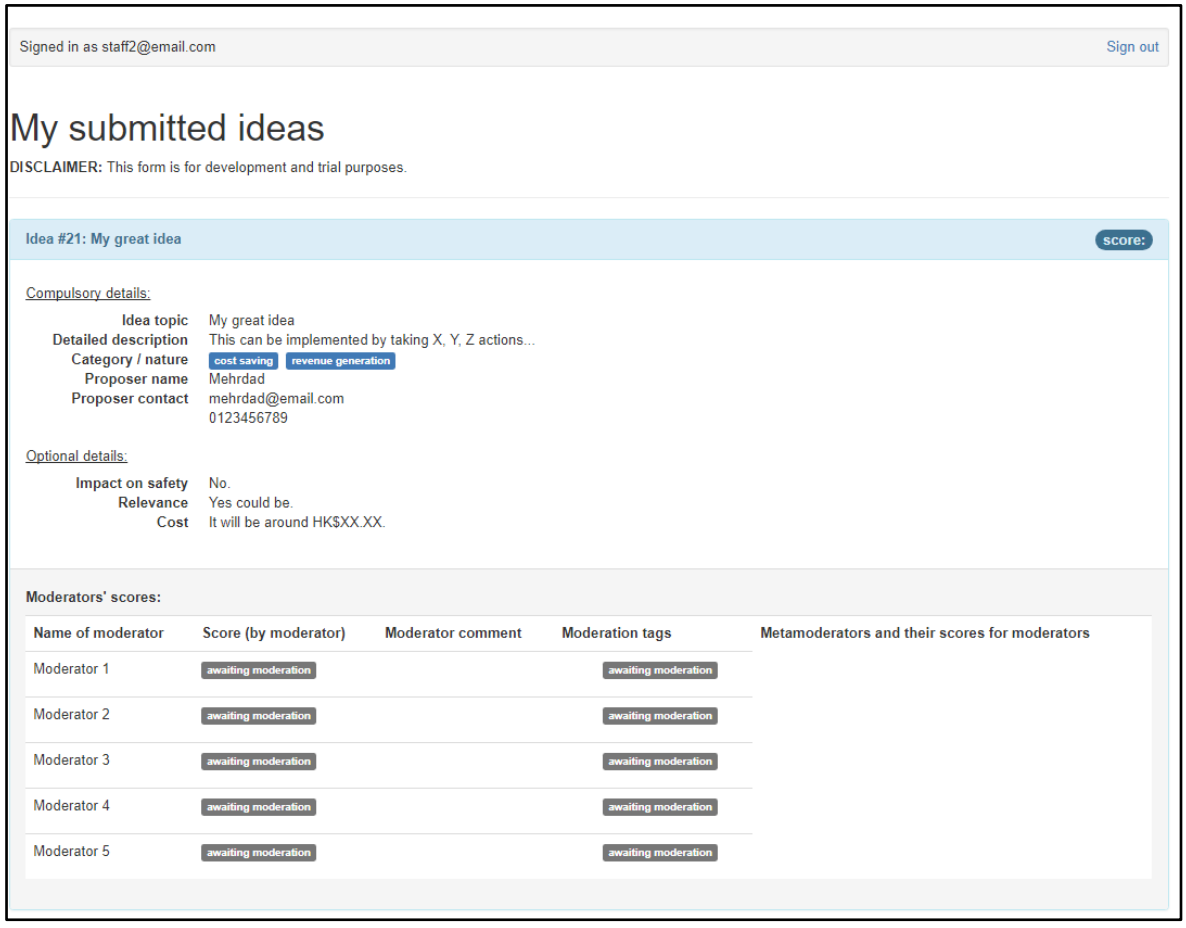

Figure 1. An idea submitter can track his/her submitted idea.

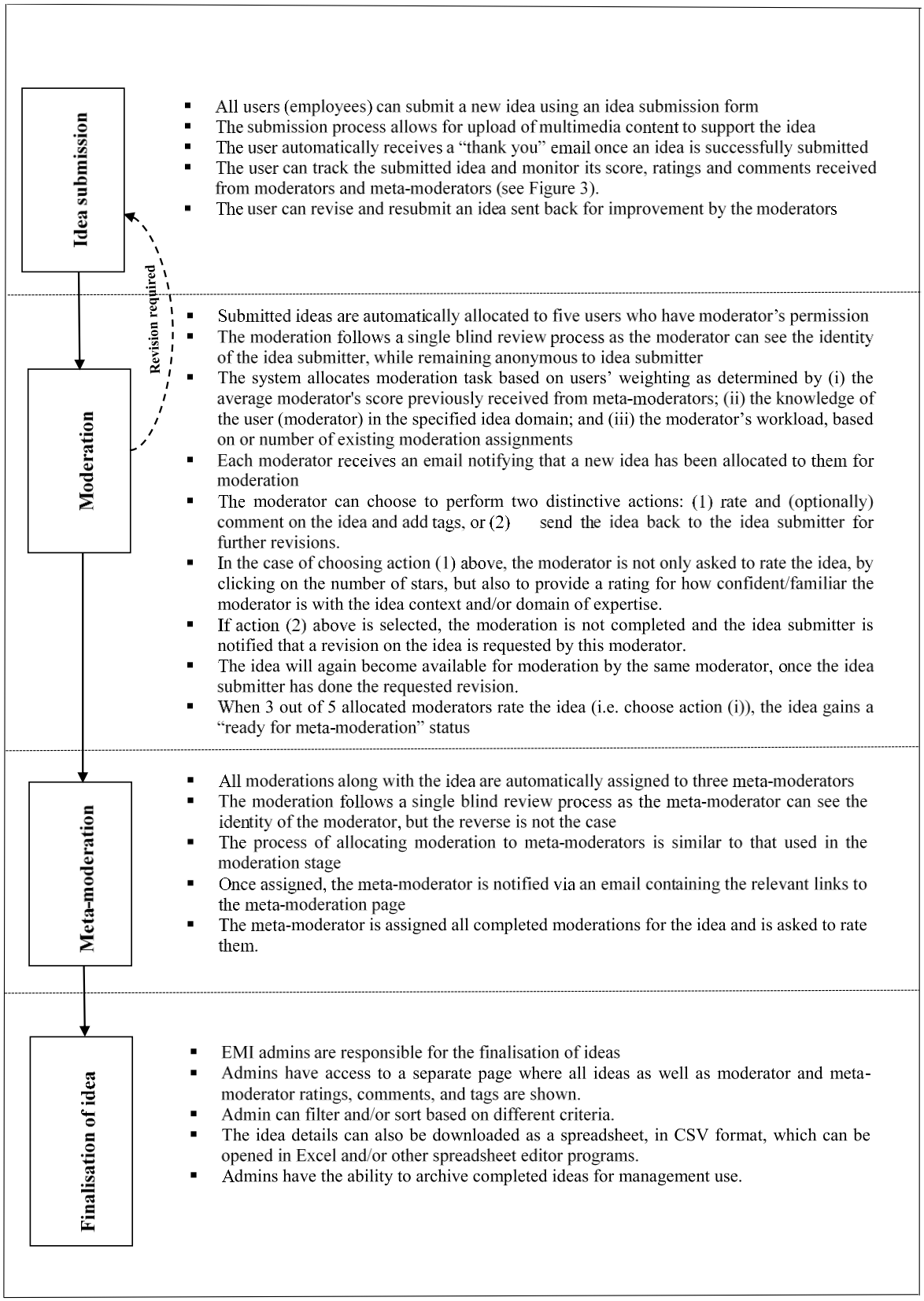

Figure 2. EMI workflow for idea generation and review.

The admins also have the permission to track each idea across its lifecycle. Figure 3 shows how an individual idea is presented on the tracking page. In cases where a moderator has not yet provided review for an idea, a small red button appears in front of that moderator's name, allowing the administrators to manually (i) send email reminders or (ii) re-assign this moderation to a new randomly selected user. Admin users can analyse and review moderators' performance and statistics, including the average score received from meta-moderators, the total number of allocated ideas, moderated ideas, and ideas pending moderation.

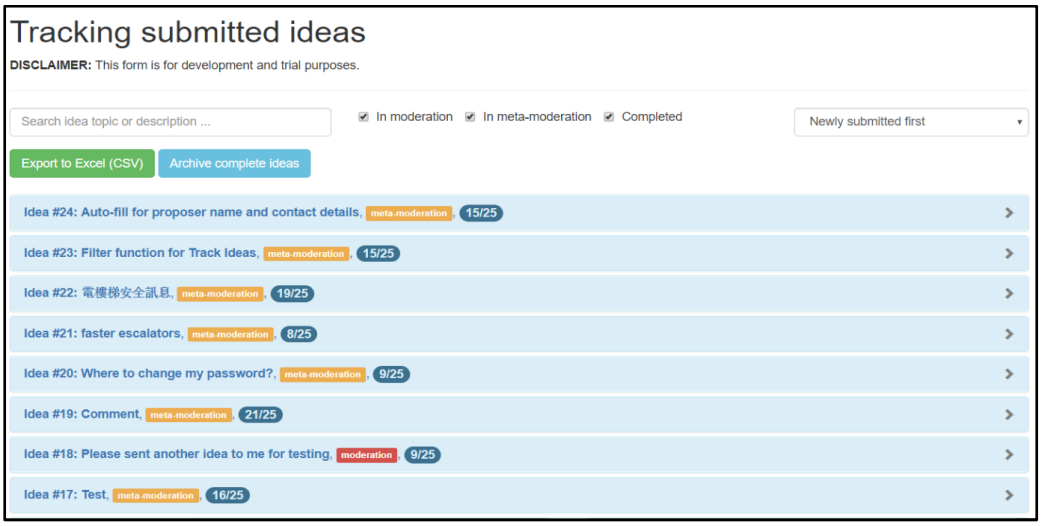

Figure 3. Tracking ideas, with all idea panels initially collapsed

3. KEY CONSIDERATIONS FOR META-MODERATION IN CROWDSOURCED DISASTER MANAGEMENT AND OTHER URBAN PARTICIPATORY APPLICATIONS

In this section, we focus on how EMI can be used in the context of crowdsourced co-management of disasters as well as for other urban participatory applications like traffic congestion, city planning, public space design, epidemic intelligence, and so on. Effective monitoring and response can be quite challenging if adequate real-time data is not readily available to gain situation awareness and improve decision making about socially disruptive events such as flooding, traffic, disease outbreak, etc. (Ogie et al. 2018). However, the rapid advancement in mobile communication technologies combined with the burgeoning concepts of "online communities", "virtual communities", crowdsourcing, social software, "open innovation", "mass collaboration", and "communities of practice" have provided a unique opportunity to harness the "wisdom of the crowd" in improving decision making about phenomena of common interest (Albors et al. 2008). The *PetaJakarta* crowdsourcing system, for example, harnesses user-generated social media content to generate real-time flood maps that enable citizens, disaster-specific agencies and the government of Jakarta, Indonesia to make informed decisions during flooding (Ogie et al. 2018). There is also the *Ushahidi* platform for crowdsourcing crisis information (Okolloh 2009), the *Waze* application for real-time traffic monitoring (Amin-Naseri et al. 2018), the *ArkiCity* project that allows citizens to engage in the transformation of their city (Amirghasemi et.al. 2019) and many more applications. The quality of outputs and the sustained coordination of these systems can potentially be improved through meta-moderation. However, the application of meta-moderation has been very limited in this type of open collaborative approach to solving problems.

An important step forward is to understand the broader issues or keys considerations for applying meta-moderation in crowdsourced disaster management and other urban participatory applications. First is to acknowledge that a meta-moderation system is not simply a technological concern; it involves people and a social dimension as well. To better appreciate the workings of these socio-technical systems, there is therefore the need to explore deeper theoretical perspectives as well as sound computational methods such as agent-based modelling (ABM). A key aspect will be to investigate the social norms regulating the online community and the co-production process (Durante 2010). Should these social norms be strongly guided by formal rules or allowed to develop through a self-organising adaptive mechanism that naturally evolves from human communication? How do different rules affect the co-production process or outcomes? These issues can be explored through an agent-based modelling (ABM) approach. ABM can simulate the actions and interactions of different autonomous agents under different participatory contexts to understand the overall effects on the co-creation process or outcomes. To this end, there is the need to explore appropriate social psychology theories

(e.g., social action theory, social identity theory, social exchange theory) in order to attain a more nuanced understanding of the complexities characterising this type of socio-technical systems (Burke 2006). Issues relating to personal values, societal values, online culture and language use will need to be considered in addition to the technological platform itself.

Given that only a few users actually contribute the majority of the input (Dosono and Semaan 2019), how can these systems be adapted so that the few users are not overexposed to a problem of burnout? This will seem to suggest the need for an appropriate incentive mechanism to attract and retain more high-quality participants while discouraging free riding and other uncivilised acts (e.g., inflammatory messages) (Parameswaran and Whinston 2007). Choosing an appropriate incentive mechanism is crucial for the sustainability of any community-based co-production initiative and there is research to guide that choice (Ogie 2016). Nevertheless, ABM can help to evaluate various options by simulating the impacts of different incentive mechanisms on the overall performance and behaviour of the system. The study of contributor's motivations in online virtual communities is not new; there are well established theories (e.g., Motivation crowding theory) and constructs (e.g., altruism, recognition, anticipated reciprocity, sense of community, need for affiliation, self-efficacy, enjoyment, self-expression, learning as motivation, extrinsic motivation) that can help inform a robust ABM study. Economic theories such as the "tragedy of the commons" model (Ostrom 1991) has been at the centre of this type of research and a well-designed ABM study will not only build on this existing knowledge, but would also extend it.

Many online participatory systems start with a decentralised horizontal structure that promotes the principles of freedom and equality, but the need for control, responsibility and regulation eventually results in a vertical (hierarchical) structure that gives unequal powers and permissions to users (Durante 2010). Regardless of whether the system runs a horizontal or hierarchical structure, participants have some basic expectations that the values of inclusion, civility, and rationality will be maintained (Dosono and Semaan 2019). "*Inclusion means that anyone can participate. Civility means that everyone must treat others with respect. Rationality means that people engage in reasoned argumentation and provide facts to support their claims*" (Dosono and Semaan 2019, p. 6). One would imagine that some levels of control and regulation are needed in the system to enact these values, but to what extent before other users start feeling unequal and powerless, both in terms of the weighting of their contributions and the influence they have on the system. These are important concerns in reducing the number of users with unfulfilled expectations in any civic participatory system. Again, ABM can be used to simulate the overall behaviour of the system with multiple interacting agents with varying levels of (dis)satisfaction and expectations. The theory of disconfirmed expectancy or "failed prophecy" can help in this regard (Zehrer et al. 2011).

There are other key issues that arise with specific consideration to using the EMI system for meta-moderation in crowdsourced disaster management and other urban participatory applications. The EMI system was specifically designed for use in an organisational setting where policy and codes of conduct formally exist to guide the expected behaviours and social norms shared by the users (staff members). For the use of a system such as EMI in crucial community initiatives like crowdsourced disaster management, Parameswaran and Whinston (2007) recommend a formally designed reputation system, with stronger governance. There is also the need for stronger security measures (e.g., digital signatures and cryptographic tools) to protect the community-facing application against cyber attacks. Veglis (2014) recommends using CAPTCHA to ensure that a content or response is actually generated by a human user, not a machine or malicious code. Many users are already familiar with the CAPTCHA tests. CAPTCHA is an acronym for "completely automated public Turing test to tell computers and humans apart". Authentication is also an important concern when accessing the system. EMI requires user login. This can be extended to other broader community-based applications involving citizens. The challenge here is that unlike organisation-specific applications where each user is known and identity can be validated against the users in the Active Directory, accountability in community-based applications can be more difficult if user accounts cannot be validated against any reliable identity database. For example, disgruntled users may reinvent themselves by creating new accounts to cause havoc, share hostile or uncivil messages and undermine the integrity of the system. A robust implementation would therefore consider the possibilities of such issues arising and carefully incorporate appropriate mitigation mechanisms (e.g., federated authentication using social media login credentials like Facebook or Google).

Another issue that will need to be considered is scalability and quality trade-off. While each system may not necessarily behave the same way, there are lessons to learn from the Slashdot application that was recently criticised for inadvertently allowing for noisy and redundant contents due to rapid scale up (Parameswaran and Whinston 2007). This is a technical issue that should never be overlooked. Furthermore, the EMI system will need to extend its single-blind review process to a double-blind approach that can better ensure the system does not leave room for partiality or conflict of interest. This is particularly important in situations where the (meta-

)moderators are subject matter experts who may be known to each other and may share a common interest or agenda. It will also ensure that moderators do not become "enemies of progress" to an idea submitter who is noted to be consistently outperforming others in submitting innovative ideas.

4. CONCLUSION

This study has presented the Enterprise Meta-moderation Innovation (EMI). EMI is a novel meta-moderation system developed by the SMART Infrastructure Facility, University of Wollongong, Australia and deployed as a learning organisation tool to drive innovation in MTR Corporation, Hong Kong. Furthermore, we discuss key considerations for using such a system in crowdsourced disaster management and other urban participatory applications. Key aspects covered include identity and authentication management, accountability issues, incentive mechanisms, scalability, cyber security, permissions/control and the social norms that regulate the system. Importantly, we highlight several opportunities to explore agent-based modelling for improved understanding of the socio-technical issues faced.

ACKNOWLEDGMENTS

EMI was funded by the MTR Corporation, Hong Kong. The authors acknowledge contributions from the MTR team, as well as other key contributors, including Dr Matthew Berryman and A/Prof. Rodney J. Clarke.

REFERENCES

Albors, J., Ramos, J.C. and Hervas, J.L. (2008). New learning network paradigms: Communities of objectives, crowdsourcing, wikis and open source. International Journal of Information Management, 28, 194-202.

Amin-Naseri, M., Chakraborty, P., Sharma, A., Gilbert, S.B. and Hong, M. (2018). Evaluating the reliability, coverage, and added value of crowdsourced traffic incident reports from Waze. Transportation Research Record, 2672, 34-43.

Amirghasemi M., Perez P., Frisk R., Pilehchian Y., Arin E., Hendrigan C. (2019). ArkiCity – co-create your city (Version 1.0.15) [Mobile Application]. Beta version accessible at https://testflight.apple.com/join/P9DwHBjm

Burke, P.J. (2006). Contemporary social psychological theories. Stanford University Press.

Dosono, B. and Semaan, B. (2019). Moderation Practices as Emotional Labor in Sustaining Online Communities: The Case of AAPI Identity Work on Reddit. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (pp. 1-13). ACM.

Durante, M. (2010). Re-designing the Role of Law in the Information Society: Mediating between the Real and the Virtual. European journal of legal studies.

Klein, M. (2011). How to harvest collective wisdom on complex problems: An introduction to the MIT deliberatorium. Center for Collective Intelligence working paper.

Lampe, C., Zube, P., Lee, J., Park, C.H. and Johnston, E (2014). Crowdsourcing civility: A natural experiment examining the effects of distributed moderation in online forums. Government Information Quarterly, 31, 317-326.

Ogie, R.I. (2016). Adopting incentive mechanisms for large-scale participation in mobile crowdsensing: from literature review to a conceptual framework. Human-centric Computing and Information Sciences, 6, 1-31.

Ogie, R.I., Forehead, H., Clarke, R.J. and Perez, P. (2018). Participation patterns and reliability of human sensing in crowd-sourced disaster management. Information Systems Frontiers, 20, 713-728.

Okolloh, O. (2009). Ushahidi, or 'testimony': Web 2.0 tools for crowdsourcing crisis information. Participatory learning and action, 59, 65-70.

Ostrom, E. (1991). Governing the Commons: The Evolution of Institutions for Collective Action. Cambridge University Press, New York.

O'reilly, T. (2009). What is web 2.0. " O'Reilly Media, Inc.".

Parameswaran, M. and Whinston, A.B. (2007). Social computing: An overview. Communications of the Association for Information Systems, 19, 762-780.

Thomas, K. and Allen, S. (2006). The learning organisation: a meta-analysis of themes in literature. The learning organization, 13, 123-139.

Veglis, A. (2014). Moderation techniques for social media content. In International Conference on Social Computing and Social Media (pp. 137-148). Springer, Cham.

Zehrer, A., Crotts, J.C. and Magnini, V.P. (2011). The perceived usefulness of blog postings: An extension of the expectancy-disconfirmation paradigm. Tourism management, 32, 106-113.