

How to cite this paper:

Hazleen Aris. (2017). Current state of crowdsourcing taxonomy research: A systematic review in Zulikha, J. & N. H. Zakaria (Eds.), Proceedings of the 6th International Conference of Computing & Informatics (pp 392-399). Sintok: School of Computing.

## CURRENT STATE OF CROWDSOURCING TAXONOMY RESEARCH: A SYSTEMATIC REVIEW

**Hazleen Aris**

*Universiti Tenaga Nasional, Malaysia, hazleen@uniten.edu.my*

**ABSTRACT.** In this study, a systematic review was performed to identify the current state of crowdsourcing classification or taxonomy research to date. A total of 23 studies was found, which were categorised into general classification and specific classification, where specific classification was further divided into classification of processes, tasks and crowd. From these studies, a total of 21 attributes used in classifying crowdsourcing initiatives were found, which were categorised into seven themes as a result of constant comparison analysis. The seven themes are crowdsourcer, crowd, task, process, platform, content and reward. Expert evaluation involving five independent researchers in the area was used to validate the themes and the categorisation of the 21 attributes into the seven themes. Evaluation results showed that the independent researchers unanimously agreed on the seven themes and on the assignments made, after slight improvement on the latter.

**Keywords:** crowdsourcing classification, systematic study, crowdsourcing taxonomy

### INTRODUCTION

Crowdsourcing is gaining more and more popularity day by day, which is manifested by the ever increasing number of crowdsourcing applications and research papers produced lately. The latter range from discussing the fundamentals of crowdsourcing to the optimisation and improvement of specific techniques and mechanisms used in crowdsourcing initiatives. A noticeable number of research papers was also seen discussing on the classification or taxonomy of crowdsourcing initiatives with the aims to better understand the components that (should) constitute a crowdsourcing initiative (Estellés-Arolas, Navarro-Giner, & González-Ladrón-de-Guevara, 2015), to present the current state of crowdsourcing initiatives and platforms (Mahmud & Aris, 2015) and to understand how crowdsourcing initiatives have evolved (Aris & Md Din, 2016). The presence of many studies on crowdsourcing initiative taxonomy provides ample resources for a study on its current state to be performed. Therefore, in this study, existing work on crowdsourcing initiative taxonomy are systematically reviewed. Such study is able to present an overview on the current state of the studies on crowdsourcing taxonomy, indicating, amongst others, areas that are mostly addressed by researchers in creating the taxonomies and areas that need to be given attention. In fulfilling the objective, the following research questions are answered.

- RQ1. How many studies are there that discuss on crowdsourcing initiative taxonomies?
- RQ2. What are the types of crowdsourcing initiative taxonomy created?
- RQ3. What are the elements that constitute each type of the taxonomies?

The remaining of this paper is organised as follows. The method used in this study is first explained followed by the presentation of the search results. Analysis of the results and discussion on the findings are then elaborated. The last section concludes the paper.

## METHOD

Search terms were formulated prior to searching. Our preliminary literature review in the area found that apart from classification, the following words that are also used to mean crowdsourcing taxonomy; categories, categorisation, typology, survey and types. Thus, we created the following search terms for our systematic literature search.

- Search term #1: crowdsourcing AND taxonomy
- Search term #2: crowdsourcing AND categorisation
- Search term #3: crowdsourcing AND classification
- Search term #4: crowdsourcing AND categories
- Search term #5: crowdsourcing AND typology
- Search term #6: crowdsourcing AND survey
- Search term #7: crowdsourcing AND types

Title search was performed on five identified data sources; Google Scholar, IEEEExplore, ACM, Science Direct and Scopus. The data sources included Google Scholar and Scopus that should give sufficient coverage. After the duplicate removal process, unique results obtained underwent three phases of selection process; selection based on title, based on abstract and based on full-text reading. The following are the inclusion criteria used in the selection process.

- The classification or taxonomy was created (proposed) in the study. Thus, we excluded studies that merely instantiated or applied others' classification.
- The classification was based on the elements that constitute a crowdsourcing initiative. For example, we excluded a study that created taxonomy of research areas in crowdsourcing.

There was only one exclusion criterion used in this study. A study would be excluded if the scope of the classification was specific to a particular application of crowdsourcing such as citizen science or geographic information only. This was due to the limited applicability of such studies.

The final set of results included for further analysis after the selection process was the *evidence* in this study. Each evidence was scrutinised and reviewed in detail to identify the types of crowdsourcing classification and the elements that constitute the classification. Cross-case analysis method was used to compile a list of unique elements. This was followed by constant comparison analysis involving axial coding and selective coding to categorise the attributes of each element into its corresponding category and theme. The categorisation result was sent for review by five independent researchers who are actively involved in crowdsourcing research for validation.

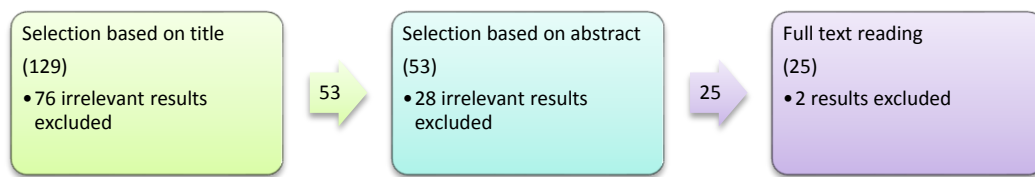
## RESULTS

Title search performed on the five selected data sources mentioned earlier returned a total of 174 initial results as shown in Table 1. From this total, 45 duplicates was removed prior to the selection process. Figure 1 shows the three phases of the selection process and the number of results included and excluded at each phase. As can be seen from the figure, a total of

23 evidence was included for further analysis at the end of the selection process. These evidence,

**Table 1. Raw Results Returned from Searched Data Sources.**

Data source	Search term #1	Search term #2	Search term #3	Search term #4	Search term #5	Search term #6	Search term #7	Search term #8
Google Scholar	13	1	15	2	5	21	8	34
IEEEExplore	2	0	5	1	2	7	2	0
ACM	5	0	9	0	0	5	3	2
Science Direct	1	0	0	0	0	3	1	2
Scopus	13	1	15	2	5	21	6	19



**Figure 1. Summary of the Selection Process.**

**Table 2. Summary of Evidence.**

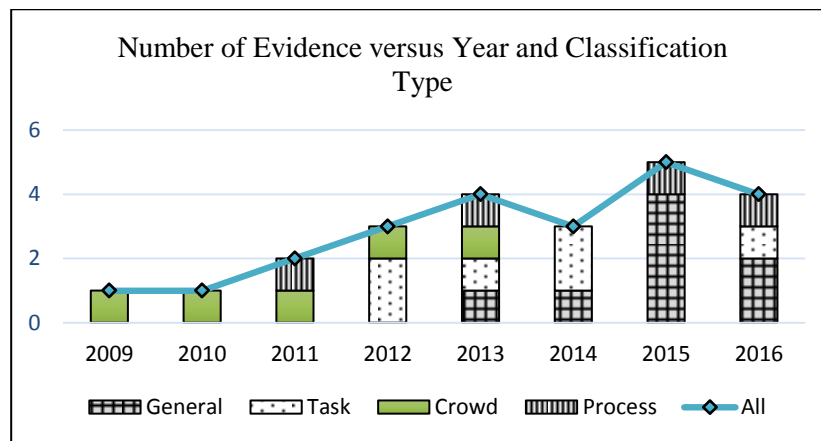
ID	Source	Type	Citation <sup>a</sup>
E1	(Rouse, 2010)	Conference	95
E2	(Nakatsu, Grossman, & Iacovou, 2014)	Journal	18
E3	(AlShehry & Ferguson, 2015)	Conference	0
E4	(Gadiraju, Kawase, & Dietze, 2014)	Conference	17
E5	(Leicht et al., 2016)	Conference	0
E6	(Aris & Md Din, 2016)	Conference	0
E7	(Hosseini, Shahri, Phalp, Taylor, & Ali, 2015)	Journal	5
E8	(Brussee, Rovers, van Vliet, Swart, & Hekman, 2013)	Conference	0
E9	(Geerts, 2009)	Master thesis	33
E10	(Geiger, Seedorf, & Schader, 2011)	Conference	209
E11	(Cullina, Conboy, & Morgan, 2015)	Conference	1
E12	(Estellés-Arolas & González-Ladrón-de-Guevara, 2012)	Journal	1
E13	(Rosselet, 2013)	Conference	1
E14	(Geiger, Rosemann, Fielt, & Schader, 2012)	Conference	52
E15	(Martineau, 2012)	Master thesis	7
E16	(Ali-Hassan & Allam, 2016)	Journal	0
E17	(Schultheiss, Blieske, Solf, & Staudtner, 2013)	Conference	4
E18	(Kazai, Kamps, & Milic-Frayling, 2011)	Conference	55
E19	(Colombo, Buzanza, Klanner, & Roiser, 2013)	Journal	9
E20	(Chittilappilly, Chen, & Amer-Yahia, 2016)	Journal	0
E21	(Luz, Silva, & Novais, 2015)	Journal	7

ID	Source	Type	Citation <sup>a</sup>
E22	(Feng, Li, & Feng, 2015)	Conference	1*
E23	(Yin, Liu, Wang, Yang, & Lu, 2014)	Book chapter	5

<sup>a</sup>Taken from Google Scholar as at 6<sup>th</sup> November 2016

labelled from E1 to E23, are listed in Table 2, together with their respective types of publication and number of citations to date.

Preliminary findings showed that the 23 evidence that worked on classification of crowdsourcing initiatives can be broadly categorised into two; general classification and specific classification. General classification attempted to classify or create taxonomy for crowdsourcing initiatives by looking at the components or elements that constitute a crowdsourcing initiative while specific classification focused on one of the components. With regard to the latter, classifications based on three components of crowdsourcing initiative were found; classification of the *tasks*, classification of the *crowd* and classification of the *process*. Distribution of the evidence according to the types of classification and years published is shown in Figure 2.



**Figure 2. Distribution of Evidence According to Year and Classification Type.**

The majority of work on the classification of crowdsourcing initiatives is focusing on the general classification. As shown in Figure 2, from the 23 evidence, eight belong to general classification category. Components that constitute the general classification of crowdsourcing initiatives in these studies include goal, participant (crowdsourcer and crowd), process, task, reward, content, platform and nature of collaboration. Figure 2 also shows that general classification of crowdsourcing initiatives started relatively late in 2013. The earliest work found on crowdsourcing initiative classification was in 2009, which worked on the classification crowd. Crowd are the participants in a crowdsourcing initiative, which may include active crowd and passive crowd (Aris & Md Din, 2016). In fact, Figure 2 also shows that work on classification of crowd was dominating the earlier years with no new work on this found after 2013. Classification of participating crowd in existing studies was seen made based on the crowd motivation to participate, the extent to which crowd is engaged with the crowdsourcing initiatives and their attitudes in completing the tasks, which leads to the quality of the submitted solutions. Another form of specific classification of crowdsourcing initiatives is based tasks. Tasks refer to the activities performed by the crowd in a crowdsourcing initiative. From the 15 evidence on specific classification, six of them worked on classification of crowdsourcing tasks as shown in Figure 2. With regard to crowdsourcing tasks, classifications were seen made based on typical tasks being crowdsourced and types of activities performed in crowdsourcing, such as innovation, fund raising or voting. Finally, crowdsourc-

ing classification was also seen made based on the processes involved. Processes are the steps to be performed in a lifecycle of a crowdsourcing initiative, such as distribution of tasks, solution of tasks and evaluation of submitted tasks. By far, classification of crowdsourcing processes is the most uniformly discussed with differences found only in term of the combination of steps that constitute a complete crowdsourcing initiative lifecycle.

## ANALYSIS AND DISCUSSION

As can be seen from Figure 2, work on crowdsourcing classification began as early as 2009, approximately three years after the term crowdsourcing was coined (Howe, 2006). Overall, the number of studies can be seen increasing. It began with the work on the classification of crowd in crowdsourcing initiatives, followed by the work on the classification of crowdsourcing process, tasks and general classification. While no new work found on the classification of crowd since the 2013, the other types of classification continue to be researched until lately, with more emphasis can be seen given to general classification of crowdsourcing initiatives. With regard to the numbers, increasing pattern can be seen since 2009 that has now come to a steady state. This may indicate potential saturation on the study of crowdsourcing initiative classification where this particular branch of study is no longer the main focus of researchers. This is not difficult to understand because classification or taxonomy studies are usually performed at the beginning of a nascent research area when researchers were trying to grasp the fundamental concepts. With regard to the types of crowdsourcing initiative classification, it can be seen that finding from similar study done by Geiger et al. (2011) that related their compiled attributes to task, process and stakeholders (crowdsourcer and crowd) is still relevant because existing work on crowdsourcing classification revolve around these three. In another similar study by Soresina (2015), four classifications were made; based on the type of labour performed, the motivation to participate, how applications function and the problems that crowdsourcing initiative is trying to solve. However, the latter classification was not supported by scientific method.

In answering the third research question, elaborations of the components that constitute the classifications were compiled. Compilation of the descriptions were necessary due to the fact that although similarities were found regarding the terms used to describe components that constitute crowdsourcing initiatives across these studies, further analysis found that relying on the terms used in the evidence can be misleading due to the different interpretations of the same term. For example, nature of collaboration in E16 was referring to how contributions from the crowd are evaluated, whether integrative or selective, while the same term was used in E22 to mean the way objective of a task is achieved, whether explicit or implicit. Constant comparison analysis method was used in extracting unique descriptions, i.e. attributes. Open coding initially produced a total of 32 attributes, which were reduced to 22 during axial coding. Seven themes emerged from the 22 attributes at the end of the selective coding phase. Categorisation of each attribute into its respective theme was validated by five independent researchers. Validation results showed that majority of the researchers agreed with the seven themes that constitute crowdsourcing initiative and with the categorisation of the attributes into their respective themes with average rating of less than 2 and less than 3 respectively based on a 5-point Likert scale, where 1 being strongly agree. Nevertheless, after consulting the comments and evaluations given, the number of attributes was further reduced to 21. Table 3 shows the validated attributes and themes. However, results of the validation could not be elaborated further in this paper due to space constraint.

In summary, it can be said that the main components of a crowdsourcing initiative are still prevalent around participants, tasks and processes. However, with the addition of more evidence published between 2011 and 2016, almost triple the number of evidence in Geiger et al.

(2011), three more components were discovered, which are content, platform and reward. This indicates the evolution of studies on crowdsourcing classification or taxonomy. It is suggested that new research in this area to concentrate on discovering attributes for the emerging crowdsourcing initiative components, i.e. platform, content and reward. Examples of such research include features and functionalities of crowdsourcing platforms and incentive mechanisms to reward the crowd. Our study is also able to provide a fresher look to the current state of crowdsourcing taxonomy, which is also more comprehensive. This is proven with little overlaps between the evidence in our study and Geiger et al. (2011), and with most number of evidence published after 2011 as shown in Figure 2.

**Table 3. Attributes of Classification and their Respective Themes**

ID	Category	Theme
S1	Type of requestor, whether it is an organisation, a company or an individual.	Crowd-sourcer
S2	Crowd motivation: Classification made based on the reasons that motivate participation, such as for altruistic or communal.	Crowd
S3	Crowd commitment: Classification made based on degree of involvement with the crowdsourcing initiative, such as active or passive.	
S4	Ability of the crowd: Classification made based on whether certain skill/knowledge is required or not from the crowd in order to participate.	
S5	Attitude towards tasks: Classification made based on crowd behaviour in completing the tasks.	
S6	Crowd involvement: Classification made based on frequency of crowd interaction with the crowdsourcing initiative.	
S7	Types of task: Classification made based on types of activities performed, such as crowd creation, crowd funding and et cetera.	
S8	Task complexity: Classification made by considering the levels of complexity of the tasks, either simple, moderate or complex.	
S9	Task granularity: Classification made based on whether or not a task can be divided into smaller tasks or microtasks in order to complete it.	
S10	Task duration: Classification made based on the timeframe given to complete a task, whether open ended or limited	
S11	Task definition: Classification made based on the way tasks are being presented to the crowd, whether well-structured or unstructured.	
S12	The way final outcome is processed, i.e. through selection or aggregation	Process
S13	Dependency: Classification made based on the way tasks are being performed, whether independently or inter-dependently.	
S14	The way objective of a task is achieved, whether implicit (known to the crowd) or explicit (unknown to the crowd).	
S15	Scope of call: Classification made based on whether invitation to participate made limited (internal) or open (external).	
S16	Processes or steps involved in the crowdsourcing initiative, e.g. some require evaluation, some do not, some require aggregation and some do not.	
S17	Flow of data: Classification based on the direction of main data used, whether from requestor to crowd (e.g. asking crowd to categorise images provided) or vice versa (e.g. asking crowd to snap photos of some creatures).	Content
S18	Type of content: Classification made based on the type of content resulting from the tasks performed.	
S19	Type of platform: Classification made based on the type of platform used, whether internal or external such as intermediaries.	Platform
S20	Implementation channel: Classification made based on the means of executing	

ID	Category	Theme
	the crowdsourcing process, either physical, virtual or both	
S21	Type of reward: Classification made based on the types of reward offered, whether monetary or non-monetary.	Reward

## CONCLUSION

This paper presents the results of a study performed with the aim to understand the current state of research on the classification or taxonomy of crowdsourcing initiatives. Systematic literature review was conducted to obtain a list of relevant literature, which was subsequently analysed to obtain answers to the research questions. Findings from the analysis showed that research work on the classification of crowdsourcing initiatives are moving from focusing on the classification of specific components of crowdsourcing initiative to focusing on the classification of crowdsourcing initiatives in general, with the number of studies seen entering steady state of late. Work of crowdsourcing classification largely belong to the three main types, which are crowd, tasks and process, with the addition of three more emerging types, which are platform, content and reward, extracted from the more recent studies. Based on the findings from this study, it is proposed that future work on crowdsourcing classifications to focus more on the emerging types of crowdsourcing classifications.

## REFERENCES

- Ali-Hassan, H., & Allam, H. (2016). Comparing crowdsourcing initiatives: Toward a typology development. *Canadian Journal of Administrative Sciences / Revue Canadienne Des Sciences de l'Administration*, n/a--n/a. <https://doi.org/10.1002/cjas.1395>
- AlShehry, M. A., & Ferguson, B. W. (2015). A Taxonomy of Crowdsourcing Campaigns. In *Proceedings of the 24th International Conference on World Wide Web* (pp. 475–479). New York, NY, USA: ACM. <https://doi.org/10.1145/2740908.2741747>
- Aris, H., & Md Din, M. (2016). Crowdsourcing evolution: Towards a taxonomy of crowdsourcing initiatives. In *2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)* (pp. 1–6). <https://doi.org/10.1109/PERCOMW.2016.7457122>
- Brussee, R., Rovers, M., van Vliet, H., Swart, D., & Hekman, E. (2013). Crowdsourcing Classification, costs, benefits, and usage. In *Proceedings of the 2nd CARPE Conference*. Retrieved from <http://hu.surfsharekit.nl:8080/get/smpid:48761/DS1>
- Chittilappilly, A. I., Chen, L., & Amer-Yahia, S. (2016). A Survey of General-Purpose Crowdsourcing Techniques. *IEEE Transactions on Knowledge and Data Engineering*, 28(9), 2246–2266. <https://doi.org/10.1109/TKDE.2016.2555805>
- Colombo, G., Buganza, T., Klanner, I.-M., & Roiser, S. (2013). CROWDSOURCING INTERMEDIARIES AND PROBLEM TYPOLOGIES: AN EXPLORATIVE STUDY. *International Journal of Innovation Management*, 17(2), 1350005. <https://doi.org/10.1142/S1363919613500059>
- Cullina, E., Conboy, K., & Morgan, L. (2015). Measuring the Crowd: A Preliminary Taxonomy of Crowdsourcing Metrics. In *Proceedings of the 11th International Symposium on Open Collaboration* (p. 7:1–7:10). New York, NY, USA: ACM. <https://doi.org/10.1145/2788993.2789841>
- Estellés-Arolas, E., & González-Ladrón-de-Guevara, F. (2012). Tasks-based classification of crowdsourcing initiatives. *El Profesional de La Informacion*, 21(3), 283–291. Retrieved from <http://www.elprofesionaldelainformacion.com/contenidos/2012/mayo/09.html>
- Estellés-Arolas, E., Navarro-Giner, R., & González-Ladrón-de-Guevara, F. (2015). Crowdsourcing Fundamentals: Definition and Typology. In F. J. Garrigos-Simon, I. Gil-Pechuán, & S. Estelles-Miguel (Eds.), *Advances in Crowdsourcing* (pp. 33–48). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-319-18341-1\\_3](https://doi.org/10.1007/978-3-319-18341-1_3)
- Feng, J.-H., Li, G.-L., & Feng, J.-H. (2015). A survey on crowdsourcing. *Jisuanji Xuebao/Chinese Journal of Computers*, 38(9), 1713–1726. <https://doi.org/10.11897/SP.J.1016.2015.01713>

- Gadiraju, U., Kawase, R., & Dietze, S. (2014). A Taxonomy of Microtasks on the Web. In *Proceedings of the 25th ACM Conference on Hypertext and Social Media* (pp. 218–223). New York, NY, USA: ACM. <https://doi.org/10.1145/2631775.2631819>
- Geerts, S. A. M. (2009). Discovering crowdsourcing: theory, classification and directions for use. Eindhoven University of Technology.
- Geiger, D., Rosemann, M., Fiel, E., & Schader, M. (2012). Crowdsourcing Information Systems - Definition, Typology, and Design. In *Proceedings of the International Conference on Information Systems, {ICIS} 2012, Orlando, Florida, USA, December 16-19, 2012*. Retrieved from <http://aisel.aisnet.org/icis2012/proceedings/ResearchInProgress/53>
- Geiger, D., Sedorf, S., & Schader, M. (2011). Managing the crowd: Towards a taxonomy of crowdsourcing processes. In *17th Americas Conference on Information Systems*. Retrieved from <http://aisel.aisnet.org/amcis2011/submissions/430>
- Hosseini, M., Shahri, A., Phalp, K., Taylor, J., & Ali, R. (2015). Crowdsourcing: A taxonomy and systematic mapping study. *Computer Science Review*, 17, 43–69. <https://doi.org/http://dx.doi.org/10.1016/j.cosrev.2015.05.001>
- Howe, J. (2006). The Rise of Crowdsourcing. *Wired Magazine*, 14(6). Retrieved from <http://www.wired.com/wired/archive/14.06/crowds.html>
- Kazai, G., Kamps, J., & Milic-Frayling, N. (2011). Worker Types and Personality Traits in Crowdsourcing Relevance Labels. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management* (pp. 1941–1944). New York, NY, USA: ACM. <https://doi.org/10.1145/2063576.2063860>
- Leicht, N., Durward, D., Haas, P., Zogaj, S., Blohm, I., & Leimeister, J. M. (2016). An Empirical Taxonomy of Crowdsourcing Intermediaries. In *Academy of Management Annual Meeting (AOM)*. Retrieved from <https://www.alexandria.unisg.ch/247976/>
- Luz, N., Silva, N., & Novais, P. (2015). A survey of task-oriented crowdsourcing. *Artificial Intelligence Review*, 44(2), 187–213. <https://doi.org/10.1007/s10462-014-9423-5>
- Mahmud, F., & Aris, H. (2015). State of mobile crowdsourcing applications: A review. In *Software Engineering and Computer Systems (ICSECS), 2015 4th International Conference on* (pp. 27–32). <https://doi.org/10.1109/ICSECS.2015.7333118>
- Martineau, E. (2012). A typology of crowdsourcing participation styles. Concordia University.
- Nakatsu, R. T., Grossman, E. B., & Iacovou, C. L. (2014). A Taxonomy of Crowdsourcing Based on Task Complexity. *J. Inf. Sci.*, 40(6), 823–834. <https://doi.org/10.1177/0165551514550140>
- Rosselet, U. (2013). Towards a goal-oriented taxonomy for crowdsourcing initiatives. In *24th Australasian Conference on Information Systems (ACIS)*.
- Rouse, A. C. (2010). A Preliminary Taxonomy of Crowdsourcing. In *21st Australasian Conference on Information Systems*. Retrieved from <http://aisel.aisnet.org/acis2010/76/>
- Schultheiss, D., Blieske, A., Solf, A., & Staudtner, S. (2013). How to encourage the crowd? a study about user typologies and motivations on crowdsourcing platforms. In *Proceedings of the 2013 IEEE/ACM 6th International Conference on Utility and Cloud Computing* (pp. 506–509).
- Soresina, C. (2015). Types of Crowdsourcing. Retrieved from <http://www.skipsolabs.com/en/blog/types-of-crowdsourcing>
- Yin, X., Liu, W., Wang, Y., Yang, C., & Lu, L. (2014). What? How? Where? A Survey of Crowdsourcing. In S. Li, Q. Jin, X. Jiang, & J. J. (Jong H. Park (Eds.), *Frontier and Future Development of Information Technology in Medicine and Education: ITME 2013* (pp. 221–232). Dordrecht: Springer Netherlands. [https://doi.org/10.1007/978-94-007-7618-0\\_22](https://doi.org/10.1007/978-94-007-7618-0_22)