USING CROWDSOURCING FOR RESEARCH PROJECTS

Open Access

sciendo

USING CROWDSOURCING FOR RESEARCH PROJECTS[1]

Katarzyna Lisek, M.A.
Jagiellonian University, Poland
Institute of Sociology
kat.lisek@gmail.com
DOI: 10.14611/minib.29.09.2018.10

Summary

The phenomenon of crowdsourcing — enabling the crowd to get involved in the activities of an organization by means of new media (Estelles — Arolas, Gonzales — Lordon — de — Guevara, 2012) — is drawing the attention of research teams. It is successfully used by the biggest scientific centres both in case of life sciences and humanities. The following article is supposed to compare various strategies of applying this phenomenon to the creation and dissemination of knowledge. 40 crowdsourcing projects were subject to exploratory observation. Their description was based on the desk research analysis of Internet websites, applications and other Internet sources. On the basis of variables describing the ordered task, the character of the crowd, incentives built into the process and the method of providing answers an empirically rooted classification of the described projects was created. Its detailed description made it possible to highlight the ways research teams use crowdsourcing and to define the areas which the authors should take into consideration planning the utilization of this phenomenon in their work: connecting the kind of the offered task with the moment of the research process at which it is applied, defining the level of expertise necessary to carry out the task, building into the process appropriate incentives and choosing the appropriate method of verifying answers.

Keywords: crowdsourcing, research process, wisdom of crowd, collective intelligence

Introduction

Crowdsourcing, that is, the process in which a unit, or an organization orders an undefined group of new media users (Estelles — Arolas and Gonzales — Lordon — de — Guevara, 2012) to carry out a particular task has become a popular method of getting Internet users engaged in the life of an organization. It has been successfully used in various kinds of scientific projects. It has become a way to accelerate research when resources are scarce, a way to search for innovative solutions and popularize knowledge already at the stage of creation of this knowledge. The novelty and dynamic character of this phenomenon lead to high diversity of projects proposed by organizations and thus to controversies surrounding definitions.

A Polish reader willing to study Polish scientific literature concerning crowdsourcing will find mostly publications dealing with the application of such projects in companies (e.g. Lenart — Gansiniec, 2017). It is also projects from the area of culture management (e.g. Kopeć, 2016) and public participation (e.g. Chrisidu — Bunik and Korczak, 2014) that draw attention. Unfortunately, in Poland the application of crowdsourcing for scientific research doesn't seem to be a popular technique.

The goal of the following article is the presentation of good practices in the area of application of crowdsourcing in scientific activity. It presents a comparative analysis of 40 projects used at various stages of research activity. Based on the collected information concerning the ordered task, the characteristics of the crowd, motivation, way of working and the kind of collective thinking they were divided into three groups. The strategies of implementation of projects of every kind were analysed and summed up with recommendations for project authors resulting from the presented classification.

Research background

Definition of crowdsourcing

The term of crowdsourcing was first used by James Howe (2006) in the magazine "Wired". Describing this process he compared it to outsourcing, pointing out that tasks originally carried out by the employees of a particular company are assigned to a big, anonymous group of Internet users. The likening to the process of assigning task to entities outside the company, known earlier in the business environment, was adopted in scientific literature and the assumption that the lack of contract with the offer takers is a defining feature of crowdsourcing gained popularity (Hosseini et. al, 2015; Wazny, 2017).

In the following years crowdsourcing became popular in various branches and areas of science (Ghezzi et. al, 2018). It is used, among others, in business as a marketing tool (e.g. Gatautis, 2014), in public management (e.g. Brabham, 2009) or even in introducing constitutional changes (Castels, 2013). Strong interest in the phenomenon and its broad application in various areas of social life have led to a situation in which the boundaries of the definition of crowdsourcing have become blurred (Ghezzi et. al., 2018, Hosseini et. al., 2015). Distinguishing it from other phenomena occurring in new media (such as open innovation and citizen science) has become a difficult challenge (Wazny, 2017). Authors point out that the definitions of crowdsourcing which can be found in scientific literature, even though they are not specific for the scientific disciplines on the basis of which they were created, are often mutually contradictory, or cover a different scope of the described features of this phenomenon. (Hosseini et. al., 2015).

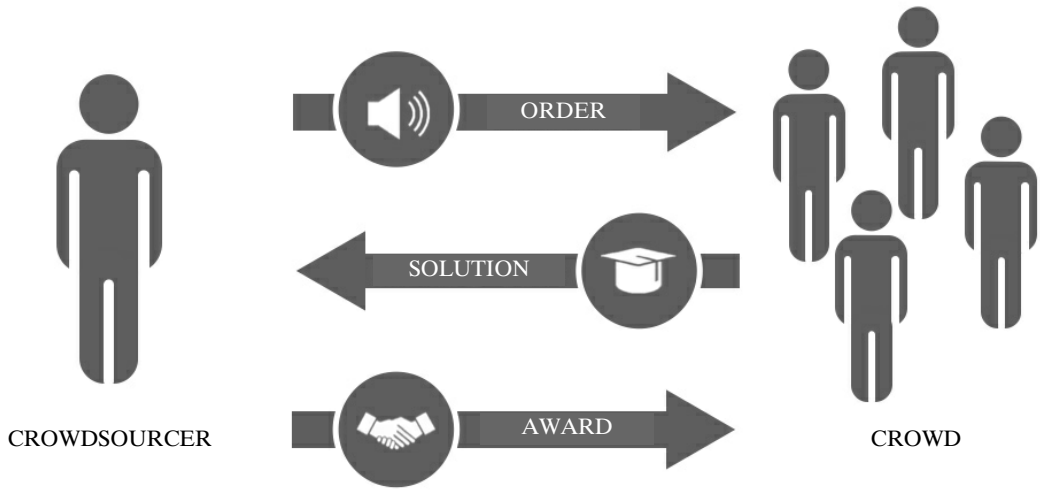
The first ones who made an effort to systematically analyse the definition of crowdsourcing in scientific literature were Estelles — Arolas and Gonzales — Lordon — de — Guevara (2012). On the basis of 40 unique definitions they drew up one, which is as follows:

"Crowdsourcing is a kind of participatory activity conducted online, in which a unit, an institution, a non-profit organization, or a company offers, by means of an open announcement, to a heterogenous group of

people with diversified knowledge carrying out a particular task. The crowd carries out a particular task with variable complexity and modularity using own work, resources, knowledge and/or experience, which always leads to mutual benefits. The user satisfies his needs associated with the economic condition, but also social recognition, the feeling of own value, or developing own skills, while the crowdsourcer receives and uses for his own benefits what the user contributed to the venture (the form of the contribution depends on the kind of the project)."

It is worth paying attention to the characteristic elements of crowdsourcing which appear in the quoted definition. The whole process starts with the crowdsourcer, or in other words the one who orders a task. The ordered work should take the form of an open invitation and should be delivered to the users by means of new technologies. On the other side of the process there is an anonymous crowd of Internet users, who accept the assigned task. Thanks to its work the crowd provides the crowdsourcer with the expected solution and the crowdsourcer rewards the people with material remuneration, social recognition, or satisfaction.

Picture 1. The model of crowdsourcing (own materials prepared on the basis of the definition by Estelles — Arolas and Gonzales — Lordon — de — Guevara, 2012)



Main actors

As the quoted definition suggests, the actors playing in the process of crowdsourcing are the crowdsourcer and the crowd. The presence of the former is an important feature distinguishing the phenomenon. The fact that the task subject to the crowdsourcing project is always ordered by someone who manages the whole process distinguishes them from open source projects, where nobody has supervisory control over the task. Crowdsourcing is a process of the "top-down" kind (Brabham, 2013).

The definitions of crowdsourcing present in scientific literature don't determine who can become the orderer. There are descriptions of projects in which the crowdsourcer was an individual, but also an institution, a company, or a non-governmental organization (Hosseini et.al., 2015). What distinguishes them is the demand for knowledge and resources, which can be offered by the crowd (Kietzmann, 2017).

Among the tasks of a crowdsourcer there is above all defining the ordered task clearly. The crowdsourcer must make sure that it will be accessible for a broad audience. It is also the crowdsourcer's duty to make sure that the crowd wants to get engaged in the task. He can do it by guaranteeing incentives in form of material goods, prestige, or fun. He is also supposed to guarantee the protection of privacy of the participants and their data (Hosseini et. al., 2015, Wazny, 2017).

On the other side of the process there is the "crowd", that is, the group of Internet users who are the recipients of the task. A defining feature and the reason for the efficiency of the crowd is diversity: It has many dimensions: spatial (associated with diverse origin of the participants), gender, or age (Hosseini et.al., 2015). Also the level of expertise of the participants in the ordered task varies. Even though what is usually emphasized is the fact that the participants are amateurs, quite often we can find specialists from a particular area in the crowd. Another frequently emphasized feature of the crowd is its big size (in contrast to the small number of employees available to the orderer on a daily basis). Talking about a "big" number of engaged users the authors of individual projects don't refer to the particular, objective number, but to the size of the crowd which is sufficient for carrying out a task efficiently. Crowds may include dozens, hundreds, or sometimes even thousands of users (Hosseini et. al., 2015).

It is also worth pointing to the anonymity of the participants of such projects. It usually works in two directions: the users don't know the side ordering a particular task and he doesn't collect detailed information about the users (Hosseini et. al., 2015). Numerous definitions of the crowd emphasize the lack of prerequisites for the users who want to become a part of the crowd and the lack of selection (Hosseini et. al., 2015). However, there are also projects where announcements were aimed at a particular group of users (Kietzmann, 2017) and preliminary selection makes it possible to control the quality of the developed product (Assis Neto and Santos, 2018).

Task

The axis of the whole process of crowdsourcing is a task ordered in form of an open announcement (Estelles — Arolas and Gonzales — Lordon — de — Guevara, 2012). In the first definitions of crowdsourcing it was emphasized that the subject of projects of that kind are tasks which were earlier carried out by the employees of a company and thanks to new media they can be assigned to people outside the organization (Howe, 2006; Hosseini et. al.., 2015). The development of this phenomenon expanded the prospects and possibilities of crowdsourcers to such an extent that they don't just use the crowd where their own resources are insufficient, but regard it as a possibility to expand their activity on a scale which up till now has been impossible (Kietzmann, 2017).

Usually in the process of crowdsourcing it is tasks which cannot be automatized, but at the same time are doable for participants at various levels of advancement that are assigned to the crowd (Hosseini et. al., 2015; Wazny, 2017). The development of new technologies has brought a change also in this area — among the most interesting solutions using crowdsourcing are those in which the potential of the crowd is combined with the application of artificial intelligence. Algorithms can guarantee high level of the solution and at the same time learn from the behaviour of the crowd (Kietzmann, 2017).

In literature we can find many classifications of work assigned to the crowd, according to their various features. Taking into consideration the

character of the task we can distinguish selective and integrating projects (Schenk and Guittard, 2011). The former involve looking for one correct answer to the assigned task in the crowd. The solution is picked from the answers delivered by the crowd through objective verification, or on the basis of a vote (Brabham, 2013). Integrating tasks involve the division of the essence of the project into numerous micro-tasks. Answers delivered by the Internet users are summed up, or averaged and only this constitutes the final solution to the assigned problem (Brabham, 2013).

Another perspective for regarding the assigned task is the perspective of the way it is carried out. Authors distinguish projects which involve cooperation and competition (Ghezzi et. al., 2018). In case of the first kind, the participants of the project have to cooperate to carry out the task. In the second case the participants compete with each other. Literature on the subject also mentions the third way of carrying out tasks: collaborative (Assis Neto and Santos, 2018). The participants of the project don't cooperate with each other directly, but it is only the combination of their efforts that brings the solution.

Contribution of the crowd

The task assigned by the crowdsourcer has to be enthusiastically welcomed by the crowd. To make it possible it is necessary to secure incentives of an appropriate kind. In literature on the subject we can find a division into external and internal motivation built into the structure of crowdsourcing projects (Hosseini et. al., 2015; Ghezzi et. al., 2018). In the first group there are material incentives such as small remuneration, or other rewards, the fact that the participants build their reputation, or good image, and finally, the fact that participants develop their professional careers. In the second group we can find such motivators as satisfaction, self-satisfaction, the opportunity to develop and learn, the need to share knowledge and creativity, as well as belonging to a community.

When the crowd encouraged by the crowdsourcer shares with him its solutions, it is necessary to verify the product delivered by the crowd. It

takes place on three levels: pre-task, in-task and post-task level (Assis Neto and Santos, 2018). In the first group there are strategies raising the average competences of the crowd within the scope of the task, such as selection of the participants, training and simplifying the task. In the second task we can find strategies which involve the elimination of answers not matching the standards, e.g. not matching the majority of answers of the crowd. In the last group there are analytical strategies carried out on the taken number of observations.

The efficiency of the crowd in delivering correct answers to the assigned tasks is rooted in the phenomena of joint, group thinking. One of the phenomena used by crowdsourcing is the wisdom of the crowd described by James Surowiecki (2010). In his work he analyses numerous examples in which average answers of big groups of amateurs gave a more precise answer to a given problem than specialists' expertise. This surprising result can be explained with Condorcet's theorem — if most of the participants have a higher than random chance to give a correct answer to a given question, the likelihood of getting the correct answer through voting will be higher than through individual opinions (Bottom et. al., 2002). Obviously, the necessary condition is avoiding the group effect and the influence of individuals on their answers (Sustein, 2009).

A strategy used to obtain the correct answer to a given question is not just averaging, but also adding up the contribution of the crowd. This happens thanks to the phenomenon of micro-experts — people who may not be exceptional experts in a particular area, but have very narrow, specialist knowledge on a given, clearly outlined subject. The confirmation of this knowledge is not always associated with formal certification (Nielsen, 2011). Authors of scientific literature point out that there are areas of interest of scientists, entrepreneurs, or their institutions, in which the knowledge is dispersed among numerous actors and crowdsourcing may constitute a way to systematically gather this knowledge (Kietzmann, 2017).

The most advanced method of cooperation and collective thinking is the phenomenon of collective intelligence. Levy (1997) defines it as: "(…) a form of commonly dispersed intelligence, continuously strengthened, coordinated in real time and leading to effective mobilization of skills. (…) Its basis and at the same time goal is mutual recognition and

enrichment of units." Crowdsourcing in this meaning uses not just individual answers of the crowd, but is becoming a platform for cooperation and mutual growth. Units become wiser thanks to being connected in a network and thanks to the fact that they can cooperate within the network (Johnson, 2010; Wooley et. al., 2010).

The goal of the article

The definitional features of particular elements of crowdsourcing presented above show huge diversification of this phenomenon. When you decide to apply it in a scientific project, you need to take a series of decisions concerning the form of the ordered task, the method of verification of the solution, characteristics of the crowd, or the kind of provided incentives. The goal of this article is comparing crowdsourcing projects created within various scientific fields and presenting efficient strategies of implementing these projects.

Four basic dimensions of the model of Collective Intelligence Genome (Malone et. al., 2009) were used as a theoretical structure for the comparison of particular cases. The authors of the model claim that the most efficient method of investigating the phenomenon of collective intelligence is searching for an answer to four key questions:

(1) What work is conducted?
(2) Who takes action?
(3) Why do the participants take part in the task? What is their strongest motivation: money, glory, or passion?
(4) How is the work done? Who decides that the task has been done well?

Method

In order to answer the formulated research questions the history of 40 successful crowdsourcing projects used in scientific processes was scrutinized. Each project had to satisfy the following three criteria:

(1) the presence of elements found in the definition of crowdsourcing by Estelles — Arolasa and Gonzales — Lordon — de — Guevara (2012): crowdsourcer, open announcement, crowd, solution and reward;
(2) implementation in course of projects conducted by scientific units, or whose effects have been used to create, or disseminate knowledge;
(3) achieving the defined goals of the project, or the level of progress of implementation suggesting that the project is successful.

The choice of the analysed projects was based, first of all, on the criteria of availability. It is hard to identify projects of this kind. There is no reliable selection range for drawing a representative sample. The method of searching for projects resembled a snowball — sometimes one project drew attention to another one, conducted by the same unit, or a similar project under construction. Another criterion was based on the diversification of observation: projects which were practically identical in terms of content, but covering different territories were skipped.

The description of particular cases was based above all on information contained in their Internet websites and applications offered by their creators. Moreover, articles describing the process of carrying out the projects and their effects were taken into consideration. Conclusions were supplemented with observations drawn from scientific works.

Using four dimensions highlighted by Malone's (2009) team and the collected literature on the subject 11 dichotomous variables describing each of the analysed projects were created. The analysis was supplemented with qualitative observations. Quantitative variables were used to create a classification and qualitative observations were used to describe it and thoroughly understand it.

Results

The first group of variables concerned the essence of the task. It was investigated, whether the task required from the users the possession of expert scientific knowledge prior to taking action.

Such knowledge wasn't necessary in case when all information necessary to carry out the task was conveyed to the participants in form of a training, or instructions before the start of work. Expert knowledge acquired before the start of a project was required in merely 18% of the analysed projects. In case of less than a half of all projects the user was supposed to understand the essence of the whole project before carrying out the task. In case of most projects the user could focus exclusively on his micro-task and the knowledge about how this work contributes to the whole project wasn't indispensable. In only 20% of the analysed projects the user could be creative in terms of the form of delivered solutions, in other cases he had to carry out the work precisely in the way defined by the crowdsourcer.

Another group of variables concerned the level of the participants' engagement. In 70% of the analyzed projects the user didn't have to share details about himself with the crowdsourcer (to participate in the project the user had to register by only providing his e-mail). In case of almost a half of all projects the user was asked to deliver data about himself, or his environment. Over a half of projects required from users engagement also outside the platform itself.

Looking at the motivators built into projects, it was possible to distinguish projects containing the elements of game and competition such as receiving points, advancing to higher levels, or comparing yourself to other users. Almost 1/3 of projects displayed such characteristics. Only in case of 13% of the analysed projects the participants could obtain financial rewards. In case of 40% of projects the participants learn while carrying out the task.

The last group of variables referred to the way in which a task should be carried out. It was investigated to what extent the uniqueness of the solution delivered by the crowdsourcer is important for the whole project — this happened in case of a half of all projects. In case of 80% of analysed projects the crowdsourcer could obtain the answer to the given question only after summing up, or averaging the answer of the crowd.

Table 1. The percentage of occurrence of particular elements of crowdsourcing in the analysed projects

		Total	Players	Collectors	Innovators
	N	40	16	16	8
What?	the task requires scientific knowledge	18%	0%	0%	88%
	the user has to understand the essence of the project	45%	6%	56%	100%
	creativity is required from the user	20%	0%	0%	100%
Who?	the user is anonymous	70%	94%	44%	75%
	the user shares data about himself, or his environment	48%	0%	100%	38%
	the user has to act outside the platform	53%	0%	81%	100%
Why?	the task contains elements of competition	28%	38%	6%	50%
	users can win money	13%	0%	6%	50%
	users learn in the process of solving the task	40%	6%	44%	100%
How?	uniqueness of the solution is important	50%	0%	75%	100%
	the type of the project is integrative	80%	100%	100%	0%

On the basis of 11 described variables a hierarchical analyses of groupings was carried out. It was supposed to divide the observations into coherent, but differing segments and thus, create an empirically rooted classification of analysed projects. The method was used for its ability to define the most heterogenous groupings of observations and at the same time to maximize the differences between them.

Considering the fact that all variables included in the model were coded in a binary way, it wasn't necessary to standardize them. A problematic issue in the process of building the model was very high level of correlation between some variables. High, or perfect correlation between the variables is usually associated with the fact that they represent the same theoretical construct. Including such variables in the process of clustering means that the construct will be represented twice as well as the remaining ones.

Among the analysed variables there was very high correlation between those concerning the necessity to hold scientific knowledge, the requirement of creativity in carrying out a task and the way how an answer is obtained (integrating vs. selective). Their definitions are not close to each other, they refer to different elements of a phenomenon. Thus, their collinearity may be resulting from different reasons. First of all, it may be the result of a comparably small number of observations and their lack of representative character with regard to the whole phenomenon. The second thing is that perfect correlation may result from the level of coding of the variables — dichotomous coding simplifies the description of the phenomenon, while ordinal level may reveal its greater diversification.

The simplest way to cope with collinearity is removing the variables which are closely tied to each other. In the prepared model it is important to make sure that all four dimensions of Malone's model are represented in a symmetrical way. Removing correlating variables would require introducing weights for the remaining ones. Suppose we remove from the model the variables concerning the necessity to possess scientific knowledge and creativity. The "what?" dimension will be represented by only one variable saying that the participant has to understand the essence of the project. In this case this variable should be given the weight 3, so that this dimension is represented in the model equally well as the remaining ones. Assigning this weight in the statistical sense is tantamount to introducing two perfectly correlated variables (Aldenderfer and Blashfield, 1984). Thus, the application of this method doesn't make sense.

Another way to cope with the challenge is using the analysis of main components for the reduction of dimensions and using the orthogonal

variables formed this way to build a hierarchical model. An analysis of groupings was carried out twice — first time when all variables in the normal form were included in the analysis and the second time when observations from the fourth dimension are weighed (considering the fact that there are only two of them, compared to three variables in the remaining dimensions). In both cases as a result of an analysis including over 80% of the variance of the original variables three components were distinguished. Next, they were used to build the hierarchical model.

As the algorithm of hierarchical analysis of groupings carries out analyses in a two-stage process — grouping similar observations and maximizing the diversification between them — two decisions were made. The first one concerned the method of clustering. The similarity between observations belonging to a given cluster was defined by means of a square of Euclidian distance. The second decision concerned the measurement of distance, which made it possible to clearly separate observations. For the purpose of determining the distance between clusters, Ward's method was used. It minimizes the sum of squares of the deviations of observations from the centre of gravity of the group they belong to. For the purpose of determining the number of separated clusters dendrogram chart analysis and the interpretability criterion were used.

As a result four models were formed: (1) using 11 variables in normal form, (2) using weights for the fourth dimension, (3) using main components determined by 11 variables in normal form, (4) using main components determined by 11 variables with the application of weights for the fourth dimension. Models 1, 3 and 4 gave the same results with regard to the attribution of observations to clusters. On their basis 3 groups of projects similar to each other were distinguished: Players, Collectors and Innovators. In case of model 2 attribution was obtained only after applying a solution separating four clusters. The Players and Innovators groups remained unchanged, the two remaining clusters determined by this model are 4 projects from the Collectors group, which don't require prior understanding of the essence of the project and for which the uniqueness of a solution is not that important, as well as the remaining 12 projects from the group. It was decided that in further analyses the model with three groupings will be used.

Classification

Players

Task

In case of the first group of projects carrying out any of the analysed tasks didn't require prior scientific knowledge from the participants. In these projects the task assigned to the crowd was usually divided into many small parts, which are not highly complicated and easy to do. The person, or organization ordering the task granted the users access to earlier collected materials and asked the users to carry out basic analyses, or categorizations covering, for example, finding appropriate shapes, or colours on photographs, attributing photographs to particular categories, or transcribing a part of a text. To properly carry out the task the users don't have to understand that they are mapping the surface of Mars like in the Planet Four project, assessing the density of a chromosome in the project titled Clumpy, or carrying out the digitalization of historical materials in War Diary. Tasks in the group are rather schematic, it is possible to learn how to carry them out by watching a short tutorial. Creative approach to carrying out the task is undesirable.

Crowd

The key to success of projects of this kind is mobilizing as many participants as possible, even if eventually only a small group of them engage in solving the tasks systematically (Rallapalli et. al., 2015). Assuming that almost every amateur is able to carry out the task, the one assigning the task doesn't need to collect detailed data about the participants. As a result, the entry threshold is usually very low and participation remains almost anonymous. In most cases you only need to provide your e-mail to register yourself in the service. Participants of projects weren't asked to carry out work outside the crowdsourcing platform — the whole task was carried out online.

Motivation

In order to attract volunteers many crowdsourcers prepare tasks in form of a computer game. This solution lets them take advantage of numerous benefits of the virtual world. In many projects from this group we can find entertaining elements — applications resemble logical puzzles, or jigsaw puzzles, they contain interesting graphics and engrossing narration. 38% of them contain elements of competition. Using a smartphone the users can analyse x-rays of cancerous cells by navigating a spaceship (Play for Cure), or can eliminate mistakes in brain mapping by fighting against a mythical monster (EyeWire). Thanks to such form of tasks it is possible to use the potential of players, that is, their feeling of encouraging optimism, faith in epic victory, joyful productivity, or the desire for mission and significance (McConigal, 2011).

Verification of answers

When the orderer invites Internet users to carry out simple tasks such as marking all oval shapes on a rather blurry photograph, which in reality is about looking for craters on the surface of a moon, he expects to get many similar, if not identical answers. This micro-task in the CosmoQuest project is regarded as solved, when at least six people provide answers to it. Such an approach turns out to be as effective as the work of a group of specialists (Robbins et al., 2014). In case of projects of this kind the value of one answer is comparably low, due to the fact that it is easy to replace it. It is only the average that can be treated as the final result. A way to control quality in projects of this kind may also be a combination of the work of a crowd with the operation of algorithms, like in the Foldit project (Khatib et al., 2011).

Kind of joint thinking

Looking at projects of this kind, it is hard not to notice the connection to the earlier presented Condorcet's thesis (Bottom et al., 2002). An important and complicated task is divided into small, schematic parts, so that even the user with the lowest competences is able to carry it out. The likelihood that a participant of the project will provide the correct answer is high, but it can be reduced by raising the

level of difficulty of a task (Hutt et al., 2013). Due to the risk of raising the number of mistakes, the level of difficulty of a task is often raised only for the most talented and engaged players. The structure of crowdsourcing platforms makes it possible for the participants to work independently, which means that the group doesn't influence the decisions made by particular Internet users. Thanks to this, according to Concordet's theory, we can expect that the most common answer will be the correct answer.

Collectors

Task

The second type of projects is supposed to collect information about small aspects of the life of Internet users, or their surroundings — users had to share data about themselves in all analysed projects of this kind. They are most often used at the stage of collecting data. Instead of investing in professional bird watching stations, or pollution measurement stations, research teams ask Internet users to share observations about the habits of birds living in their gardens (Hummingbirds at Home), or about the quality of water in their neighbourhood (SaveCast). To properly carry out a task the participants are required to prepare in advance, which includes learning about the scope of desired information and the way in which it should be reported. The user may be asked to fill out a report from observations, send photographs, or to just mark a point on a map. Project participants don't have to understand the whole research process, however, detailed knowledge about the scope and goal of their task is necessary.

Crowd

To make sure that such a project will end up with success, it is necessary to gather a big group of participants. In almost a half of analysed projects they can remain anonymous. However, information about the users and the environment in which they live is crucial for the task. Participation in the project is more challenging than in case of the previous group of projects. Even in case of one-off contribution carrying

out the task requires more time and effort. In 81% of projects the research process involves tasks which have to be carried out offline, such as searching for interesting objects, making observations, or taking photographs. Thus, the participants have to employ higher skills than in case of the first group of projects. Nevertheless, scientific expertise is still not required from them. Lack of access to important information due to territorial boundaries and different life experiences may constitute an entry barrier for many potential users. It is hard to imagine that someone living in Poland could participate in the eddMaps project mapping the locations of invasive plants in Florida. However, the entry threshold is not that high in all projects.

Motivations

It seems that the very desire to carry out an epic mission and get involved in an important cause, which comes up in case of the projects of the first kind is also present among the participants of this group of projects. This time the participants are offered not just entertainment and good fun, but also the possibility of learning something new and fascinating. When an Internet user becomes a participant of a project, he is not just trained to carry out simple, repetitive tasks, but he is also invited to provide an exceptional, unique contribution to the greater challenge. On the websites of projects it is often emphasized how important the individual contribution of the participants is. Often on the website of a project users can identify their point of data by, for example, clicking on a map with observations. The feeling of authorship and community among the participants is strengthened by project authors and keeping in touch on message boards and at offline events.

Verification of answers

In case of collecting data from a dispersed crowd of users it is hard to expect that every reported observation will be diligent and correct. Every user is anonymous and isn't monitored by a researcher, that's why it is up to his will whether the information he reports is correct. A way to check the correctness of the collected data is testing their internal coherence and removing the data, or paying special attention to the data that don't match the general pattern (e.g. Sullivan et al. 2009). Again,

the crowdsourcer is interested in collecting as many observations as possible. This time he also wants the observations to differ from each other. The value of an individual piece of information is not crucial, but much higher than in the first case.

Kind of joint thinking

Searching for people who have unique information about a section of reality interesting for the researcher and combining their efforts to draw conclusions about a big and important problem are the defining elements of micro experts (Nielsen, 2011). Hobbyists interested in watching frogs don't have to be professional herpetologists to have almost expert knowledge about the habits of these amphibians in their closest surroundings. Their ability and knowledge, but also energy and engagement devoted to their passion may be used for a greater goal.

Innovators

Task

The situation of a person, or organization assigning tasks in the third type of projects is substantially different than in case of the previous two. Usually, at the starting point of a project data are collected and basic analyses of the data are carried out. Yet, attempts to find answers to the main research question are still unsuccessful. Some of these projects started with one post on a blog, where a prominent scientist shared his unsuccessful efforts to solve a particular problem (Polimath), others are official announcements from big concerns, or public institutions and guaranteed rewards (MIT Clean Energy Prize). The problem awaiting a solution in these projects is usually very hard. It requires from users expert knowledge and skills, as well as creativity and innovativeness. The goal of these projects is finding solutions which by now haven't been used, or known in a given discipline. An organization, in order to be able to assign such complicated tasks needs to share already worked out solutions and collected data. Projects involve not just collecting new scientific knowledge, but also looking for ways to implement it in applicable solutions.

Crowd

The success of such projects doesn't unequivocally depend on the number of people engaged in them. Online teams which take up the assigned tasks are in this case several times smaller than in the projects described above. An important issue is broad dissemination of the announcement concerning the task, so that the authors of the project can get people with various specializations, experiences and origins engaged in the implementation of the project. In the search for the best-working algorithm for the measurement of nano-distortions in photographs of space caused by the influence of dark matter, the authors of the most important solutions were experts from the areas of glaciology, computer vision, or verification of signatures (Kaggle). In order to participate in a project the participants need to have broad knowledge, specialist skills and the ability to transfer them to various contexts. Their expert knowledge doesn't have to be confirmed with a university diploma, but has to be confirmed in course of work on a task.

Motivations

A half of the analysed projects from the group are built like contests in which the best answer wins a monetary prize. Even though the highest prizes reach up to dozens of thousands of Dollars, patenting a solution posted there would probably be a more profitable option for the users. Apart from financial motivation, in this phenomenon we can find all three motivators which Alon (2010) identified as most significant for scientists in their activities: the possibility of becoming better, autonomy in actions and being a part of a team with a strong mission. The possibility of learning in course of carrying out the task was present in all analysed projects from the group. In many of them the form of participation was very loose — it was enough, for example, to write a comment under a blog post (Polymath Project). The participants were also required to cooperate and inspire each other.

Verification of answers

Taking into consideration the fact that projects of this kind are supposed to find one, brilliant answer to a complex and complicated

question, the verification of its correctness is not a problematic issue. In most cases the author of a question has the right (given by the regulations of a project) to decide which of the proposed solutions satisfies the expectations best. In some projects it happens that the final solution is formed in the process of long discussion among the participants and is the result of a consensus (e.g. Cranshaw and Kittur, 2011).

Kind of joint thinking

In order to obtain an answer to a complicated question it is not enough to sum up the answer of many participants of project. Crowdsourcing platforms are becoming platforms for communication, inspiration and exchange of ideas. The basis of the process is teaching and inspiring each other. The phenomenon of crowdsourcing creates room for an outburst of collective intelligence.

Table 2. The characteristics of distinguished groups of the analysed projects

	Players	**Collectors**	**Innovators**
What?	schematic tasks	collecting data	creating a solution
Who?	random crowd	hobbyists	experts
Why?	fun	hobby	solving a problem
How?	averaging	adding up	arbitrary choice/consensus
Genome	wisdom of the crowd	micro expertise	collective intelligence

Conclusions

Conclusions arising from the analysed classification

Comparing the analysed projects points to the diversity of support that crowdsourcing can offer to scientific research and the dissemination of scientific research. It is worth pointing out here how many decisions about the shape of such a process have to be made by the research team that would like to use this phenomenon in its work. On the basis of the conducted classification it is possible to highlight a few areas, which project authors should particularly pay attention to.

First of all, the kind of the task that the orderer wants to assign to the crowd has to be associated closely with the stage of the research process at which it is supposed to be used. At the beginning of the research process, at the moment of collecting data it is the potential of Collectors that is most useful for the crowdsourcer. They can provide him with information which he wouldn't be able to obtain in a different way. They can also substantially expand his access to data. Moreover, this is possible without the necessity to raise spending. A researcher who has access to a huge amount of source materials — it is hard to automatize the preliminary analysis and categorization of these materials — can use the help of Players. In the analysed projects it was materials such as original historical texts, space photographs, or x-rays that were distributed among the project participants and coded by them. If scientists in their work want to establish cooperation with specialists from fields of science other than their own and if they want to exchange ideas with innovators unknown to them, they should look into crowdsourcing projects from the third group.

The second thing is that the decision concerning the choice of recipients of our invitation to a project should be associated with the level of complexity of a task. It is impossible to overlook the fact that particular types of analysed projects differ in terms of the difficulty of offered tasks and thus are aimed at participants with varied levels of expertise. If a crowdsourcer wants to assign a simple task characterized by low entry requirements to the crowd, he can address his invitation to a very broad audience. Projects of this kind don't require high level of engagement from the participants and their entry barriers are low. Even

one-off engagement of a random person is precious from the perspective of the received solution. In case of tasks consuming much more time and energy, with the necessity to take actions outside the platform, it is necessary to mobilize hobbyists and experts. In order to get from them a solution to the problem, the crowdsourcer wants to make sure that they want to remain in the project longer.

Third thing is that various kinds of projects apply different kinds of motivators. Projects mobilizing a random crowd carrying out simple tasks most often used elements of entertainment and competition. Projects mobilizing hobbyists (who don't necessarily have scientific knowledge, but are ready to devote more time and attention) referred to their unique contribution and the feeling of joint participation in an important cause. Experts, in case of the hardest tasks, were offered monetary rewards, but also the possibility of participating in a scientific debate and joint work on solving the most pressing challenges.

Fourth thing is that the method of verification of the provided answers is closely associated with the type of the offered task. In case of projects which involve collecting information from participants about themselves, or their closest environment, the final product of the process is summed up in form of a database (in some cases provided to the participants e.g. in form of marks on a map). Incorrect answers can be identified by the crowdsourcer through the comparison of submissions with the general observed trends. In projects which involve carrying out easy and repetitive tasks such as tagging, classification, solving simple puzzles, often the answers given most frequently by the participants, or verified by an algorithm are recognized as correct. In the last type of projects the correct answer is chosen by the crowdsourcer himself, or produced as a result of a consensus between the participants.

Limitations of the research

Using the presented classification we should remember that the presented research is exploratory in its character, so it is exposed to a few limitations. First of all, projects used for the research sample weren't chosen randomly. Crowdfunding projects (in which the crowd pays small

amounts of money supporting the implementation of a project) and projects carried out by means of platforms such as Amazon Mechanical Turk were excluded from the sample on purpose. Second thing is that due to a comparably small number of observations and their high diversification we cannot make conclusions with regard to the stability of the presented solution of the analysis of groupings. It is supposed to serve the purpose of comparing analysed projects, not for the creation of a theoretical model. Due to high dynamics of the described phenomenon and the fact that its new, surprising applications keep coming up, it is necessary to conduct further search for a comprehensive description of the phenomenon.

References

[1] Scientific paper financed with budget funds for science in 2013/2014 as a research project within the programme titled "Diamentowy Grant".

Bibliography

1. Aldenderfer, M.S., Blashfield, R.K. (1984). *Cluster Analysis*. Sage Publications, Inc.
2. Alon, U. (2010). How to Build a Motivated Research Group. *Molecular Cell*, 37, 151–152.
3. Assis Neto, F.R., Santos, C.A.S. (2018). Understanding crowdsourcing projects: A systematic review of tendencies, workflow, and quality management. *Information Processing and Management*, 54, 490–506.
4. Bottom, W., Ladha, K., Miller, G. (2002). Propagation of individual bias through group judgement: Error in the treatment of asymmetrically informative signals. *The Journal of Risk and Uncertainty*, 25, 147–163.
5. Brabham, D.C. (2009). Crowdsourcing the Public Participation Process for Planning Projects. *Planning Theory*, 8, 242–262.
6. Brabham, D.C. (2013). Crowdsourcing: a model for leveraging online communities. W: A. Delwiche (red.), J.J. Henderson (red.), *The Participatory Cultures Handbook* (120–129). New York: Routledge.
7. Castells, M. (2013). *Networks of Outrage and Hope: Social Movements in the Internet Age*. Cambridge: Polity Press, pp. 31–44.
8. Chrisidu-Budnik, A., Korczak, J. (2014). Crowdsourcing — aktywizacja społeczności lokalnej do wspólnego działania. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, 372, 60–72.
9. Cranshaw Justin, Kittur Aniket (2011) The polymath project: lessons from a successful online collaboration in mathematics. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI'11, 1865–1874.

10. Estelles — Arolas, E., Gonzales — Lordon — de — Guevara, F. (2012). Towards an integrated crowdsourcing definition. *Journal of Information Science*, 38, 189–200.
11. Gatautis, R. (2014). Crowdsourcing Application in Marketing Activities. *Procedia — Social and Behavioral Sciences*, 110, 1243–1250.
12. Ghezzi, A., Gabelloni, D., Martini, A., Natalicchio, A. (2018). Crowdsourcing: A Review and Sugestions for Future Research. *International Journal of Management Reviews*, 20, 343–363.
13. Hosseini, M., Shahri, A., Phalp, K., Taylor, J., Ali, R. (2015). Crowdsourcing: A taxonomy and systematic mapping study. *Computer Science Review*, 17, 43–69.
14. Howe, J. (2006), The rise of Crowdsourcing. *Wired Magazine*, 14.
15. Hutt, H., Everson, R., Grant, M., Love, J., Littlejohn, G. (2013), *How clumpy is my image? Evaluating crowdsourced annotation tasks, Computational Intelligence (UKCI)*, 13th UK Workshop, s.136–143.
16. Johnson, S. (2010). *Where Good Ideas Come From: The Natural History of Innovation*. Riverhead Books.
17. Khatib, F., DiMaio, F., Foldit Contenders Group, Foldit Void Crushers Group, Cooper, S., Kazmierczyk, M., Gilski, M., Krzywda, S., Zabranska, H., Pichova, I., Thompson, J., Popović, Z., Jaskolski, M., Baker, D. (2011), Crystal structure of a monomeric retroviral protease solved by protein folding game players. *Nature Structural and Molecular Biology,* 18, 1175–1177.
18. Kietzmann, J.H. (2017). Crowdsourcing: A revised definition and introduction to new research. *Business Horizons*, 60, 151–153.
19. Kopeć, K.D. (2016). Digital participation in cultural heritage. The case of open monuments crowdsourcing platform. *Journal of Modern Science*, 1/28, 47–64.
20. Lenart-Gansiniec, R. (2017). Crowdsourcing — systematyczny przegląd literatury. *Przegląd Organizacji*, 3/2017, 25–34.
21. Levy, P. (1997). *Collective Intelligence: Mankind's Emerging World in Cyberspace?*. Plenum Trade.
22. Malone, T.W., Laubacher, R., Dellarocas, C. (2009) *Harnessing crowds: Mapping the genome of collective intelligence*. Massachusetts Institute of Technology, Working Paper No. 2009–001, 1–0.
23. McGonigal, J. (2011). *Reality is Broken: Why Games Make Us Better and How They Can Change the World*. Jonathan Cape.
24. Nielsen, M. (2011). *Reinventing discovery: the new era of networked science.*, Princeton University Press, 13–89.
25. Rallapalli, G., Fraxinus Players, Saunders, D.G., Yoshida, K., Edwards, A., Lugo, C.A., Colin, S., Clavijo, B., Corpas, M., Swarbreck, D., Clark, M., Downie, J.A., Kamoun, S., Cooper, T., MacLean, D. (2015). Lessons from Fraxinus, a crowd-sourced citizen science game in genomics. *eLife*, 4, e07460.
26. Robbins, S.J., Antonenko, I., Kirchoff, M.R., Chapman, C.R., Fassett, C.I., Herrick, R.R., Singer, K., Zanetti, M., Lehan, C., Huang, D., Gay, P.L., (2014) The variability of crater identification among expert and community crater analysts. *Icarus*, Volume 234, s. 109–131.
27. Schenk, E., Guittard, C. (2011). Towards a characterization of crowdsourcing practices. *Journal of Innovation Economics & Management*, 1, 93–107.

28. Sullivan, B.L., Wood, C.L., Iliff, M.J., Bonney, R.E., Fink, D., Kelling, S. (2009). eBird: A citizen-based bird observation network in the biological sciences. *Biological Conservation*, 142, 2282–2292.
29. Sunstein, C. (2009), *Infotopia: How Many Minds Produce Knowledge*. Oxford University Press, 11–81.
30. Surowiecki, J., (2010). *The Wisdom of Crowds*. OnePress.
31. Wazny, K. (2017). „Crowdsourcing" ten years in: A review. *Journal of Global Health*, 7, 1–13.
32. Woolley, A.W., Chabris, C.F., Pentland, A., Hashmi, N., Malone, T.W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science*, 330, 686–688.

Katarzyna Lisek, Jagiellonian University, Poland — MA in sociology, doctoral student at the Department of Sociology of Economy, Education and Methods of Social Research of the Institute of Sociology at the Jagiellonian University, expert at the Centre of Evaluation and Analyses of Public Policies of the Jagiellonian University. Scientific interests: evaluation, public policies based on evidence, crowdsourcing, methodology of social research.