11-17-2017 2:00 PM

# An Exploration of the Application of Crowdsourcing to Health-Related Research

Harpreet Bassi, *The University of Western Ontario*

Supervisor: Andrew M. Johnson, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Health and Rehabilitation Sciences

## Abstract

**Background:** A growing number of health research projects are employing crowdsourcing as part of their methods, leveraging it to inform everything from study design to participant recruitment to data collection and analysis. Therefore, greater understanding of how crowdsourcing is being used and how it can be applied in the research contexts warrants further exploration. **Purpose:** The purpose of this dissertation was to explore crowdsourcing as a means of research inquiry, and to locate it amidst research paradigms; understand how crowdsourcing in research is used in practice; and, create a framework, and guidelines, for researchers using crowdsourcing in their research. **Research Questions:** The following research questions were posed: a) What are the core principles and philosophies of crowdsourcing as a research paradigm? b) How and why are researchers using crowdsourcing? c) How are researchers addressing the basic characteristic of crowdsourcing in research studies? d) How should researchers address the basic characteristics of crowdsourcing in research studies? **Methodology:** To answer the first question, the ontology, epistemology, methodology and axiology of crowdsourcing as a research paradigm was explored. An observational study then analyzed 227 publically available research projects on a crowdsourcing website. Finally, a modified Delphi technique was used to determine whether there was a consensus among 18 experts regarding the use of crowdsourcing for the purposes of research. Based on these studies, a conceptual framework for crowdsourcing research studies emerged. **Findings:** The core principles and philosophies of crowdsourcing resemble those of the participatory paradigm. Crowdsourcing is being used primarily as a method for participant recruitment, data collection and analysis. The most plausible framework for the application of crowdsourcing in studies is based on the research paradigm which in turn defines the roles of the crowd. The role of the

crowd defined in generally acceptable research terms (e.g. participant, data collection, analysis, study design etc.) makes it feasible to align the role with the research paradigms to define the crowd as subjects or participants, citizen scientists, or co-researchers.

**Implications:** These findings suggest that crowdsourcing as a method should align with the research paradigm within which it is being applied. Implications for future research are discussed.

**Keywords:** crowdsourcing; research methods; co-researchers; citizen-scientists; research methodology

## Co-Authorship Statement

This doctoral dissertation consists of three complementary manuscripts that represent the collaborative work of researchers and co-authors. The primary contributor is Harpreet K. Bassi, who identified and researched the topic, designed the studies, collected, analyzed, and interpreted the data, and drafted the manuscripts. The advisory committee, consisting of Dr. Laura Misener, Dr. Christopher Lee, and Dr. Andrew M. Johnson, are co-authors on these manuscripts as they supervised the research, provided insights into the research design, the research findings, and the final manuscripts submitted for publication.

**Acknowledgements**

Thank you to the Delphi Panel members whose willingness to share their knowledge, experience and time made this dissertation possible.

To my supervisor, Dr. Andrew M. Johnson, I will forever be grateful for your academic heroism, your unwavering commitment to this dissertation, and your sense of humour. Committee members Dr. Laura Misener and Dr. Christopher Lee, I thank you for intellectually challenging me during every conversation and exchange, and for improving each iteration of this dissertation.

I must also recognize and acknowledge the steadfast encouragement and support from Dr. Laura Corbett, Dr. Kirk Nylen and Dr. Terrance Sullivan all of whom inspired and helped sustain my forward momentum.

Dr. Pierre-Gerlier Forest, Dr. Greg Marchildon, and Dr. Juan-Luis Surez your early insights were much appreciated and helped define this journey.

Finally, a special acknowledgement to my family. None of this would have been possible without your love and support. No words can begin to express my gratitude. I am truly blessed.

## Table of Contents

## List of Tables

## List of Figures

## List of Appendices

## Preface

In the spirit of self-reflection, and the practice of "outing the researcher" in qualitative research, I would be remiss if I failed to articulate the three themes that influenced my experience throughout this dissertation: (1) embracing the interplay of experiences, (2) adopting a mixed mindset, and (3) life lessons. I will endeavour to briefly address each of these in turn.

### Embracing the Interplay of Experiences

This dissertation embraces the creative, curious and innovative side of my personality – a side that is not always exposed in my day to day life. My academic background is a mix of political science, communications, and public administration. All three of these have shaped and informed my research. Influences of each can be found throughout this dissertation. My own background in political science and communications no doubt colour my discourse on crowdsourcing. Over the past 15 years, I have worked primarily in health systems administration – provincial-level organizations responsible for shaping and informing policy and care delivery through communications with and engaging providers and patients. This too is reflected throughout these pages in the way in which I consider research within the health context. The interplay of my experiences have influenced and driven this dissertation.

### A Mixed Mindset

Crowdsourcing requires adopting a mixed mindset. Aggregation is what makes the knowledge and skills of the crowd effective, but in order to engage and understand the crowd, one needs to embrace individual motivations. I liken this to mixed methods. In my professional health administration world, I oscillate between these two realities. Health care is similar insofar as the outcomes we measure include those tied to patient satisfaction and experience at a macro level; however, the stories we tell to amplify narratives of patient centeredness are of individual

people. Similarly, it is important to recognize that although clinical research is primarily grounded in quantitative methods, health policy follows headlines and occasionally reflects an isolated negative experience that captures the public's attention. The mixed mindset allows one to embrace these paradoxes and contradictions; recognizing and harnessing the value of each. This dissertation seeks to embrace the complimentary nature of both the individual and the collective, the quantitative and the qualitative, the left brain and the right. This dissertation is best read with a mixed mindset.

**Lessons Learned to Live By**

As a mature student, the experience of returning to school has been challenging and rewarding. I have had to equally learn and unlearn. This means questioning many of my underlying values and beliefs. Having encountered some challenging situations throughout this journey, I have come to the realization that I no longer am the same person who began this dissertation four years ago. I take from this experience three lessons that have had a profound impact on my personal and professional growth and development. First, undertaking a PhD is a lesson in rejection and perseverance. Whether it is not getting a scholarship (personal experience) or failing to get a manuscript published in a specific journal. Accepting rejection is insufficient – the only option is to persevere and carry on. Rejection and perseverance build character.

Second, the best ideas come from interactions with others. Throughout the course of my PhD, I have had a limited number opportunities to engage with research and academic colleagues and peers. Sharing information and knowledge through dialogue facilitates creative thinking and forces you to challenge your own thinking. Each of these interactions left me energized and

excited about revisiting my own research. If you are open to it, there is as much value in the informal learning as there is in the formal processes.

Finally, to quote Mahatma Gandhi "be the change you want to see in the world". Throughout the course of my dissertation, I have meet a handful of researchers and academics who truly embrace the changing times in which we live and rise to occasion. They do this by questioning their own assertions, demonstrating an openness and willingness to think and act differently, and challenging those around them to do the same. I believe it is these bright lights, these individuals who will help maintain the relevance of a university education. I hope to continue to be inspired by them, learn from them, and contribute in a meaningful way to future generations.

# Chapter 1

## 1.1 Overview

Crowdsourcing, a form of open collaboration, is increasingly being used in health sciences research. That said, while a growing number of health- and medicine- related studies identify crowdsourcing as part of their methodology, there is limited evidence aimed at properly understanding crowdsourcing as a research strategy. The opportunity exists, therefore, to undertake an exploratory study to understand crowdsourcing as a means of inquiry and knowledge generation, and to further examine how it can be employed as a research technique, process and paradigm. This research is situated within qualitative inquiry, and this dissertation seeks to:

1. Explore the potential of crowdsourcing as a research paradigm, and examine its core principles and philosophies as a means of inquiry;
2. Understand how crowdsourcing in research is used in practice; and,
3. Create a framework, and guidelines, for researchers using crowdsourcing in their research.

This study will begin to formulate an approach to crowdsourcing as a research paradigm and/or method. This work will support researchers and academics by examining how crowdsourcing can be applied for the purposes of research. It also aims to contribute to discussions about the co-generation and co-creation of knowledge in the publicly funded research sphere, the democratization of knowledge, and how the engagement of the crowd can potentially change health sciences research.

## 1.2 Relevance of Dissertation

In *The Structure of Scientific Revolutions*, Kuhn maintains that periodic revolutions in science alter existing paradigms and change the way knowledge is created, received and consumed (1970). Today's rapidly evolving landscape challenges existing approaches to research, redefines existing paradigms, and introduces innovative new approaches to knowledge generation. There are a number of factors that make this research relevant, including: the evolving research funding environment, the proliferation of knowledge, and the fact that many researchers already reference 'crowdsourcing' in their methods (Armstrong, Cheeney, Wu, Harskamp, & Schupp, 2012; Bevelander et al., 2014; Brown & Allison, 2014; Coley et al., 2013; King, Gehl, Grossman, & Jensen, 2013; McCoy et al., 2014; Turner-McGrievy, Helander, Kaipainen, Perez-Macias, & Korhonen, 2015).

Researchers face a growing set of criteria for grants and funding. In order to secure funding, researchers must attentively examine and rethink how research is conducted and evaluated. Research funding is increasingly tied to partnerships and collaborations, patient engagement, and knowledge translation and mobilization (Graham & Tetroe, 2009; Lavis, Lomas, Hamid, & Sewankambo, 2006; Tetroe et al., 2008). Research grants are becoming mechanisms that bring together various parties, often seeking alignment between industry and academic research for multiple purposes including spurring innovation and commercialization, matched funding opportunities, and knowledge translation and mobilization (Benner & Sandström, 2000). There are also increasing requirements of funding organizations for patient engagement in research (i.e., patient-centred research) to ensure that the needs and concerns of patients are being considered and addressed (Domecq et al., 2014; Frank et al., 2015; Légaré, Stacey, Forest, & Coutu, 2011; Shippee et al., 2015). In addition, patient-centred research has the

potential to enable greater uptake of evidence, which again fulfills the desideratum of knowledge translation and mobilization (Celi, Ippolito, Montgomery, Moses, & Stone, 2014). Crowdsourcing thus has the potential to offer researchers not only a method by which to engage with the public, and/or specific segments of the public, but also to build knowledge translation and mobilization into the method itself.

When knowledge-doubling-curve theory was first introduced, the belief was that human knowledge would double every 25 years (Fuller & Kuromiya, 1981). Today, it is estimated that the doubling time of medical knowledge went from 50 years in 1950 to 3.5 years in 2010, with some researchers proposing that the "Internet of things" would lead to the doubling of knowledge every 12 hours (Coles, Cox, Mackey, & Richardson, 2006; Densen, 2011). This exponential growth of knowledge, enabled by information and communications technologies, has influenced all aspects of human knowledge, and health related research is no exception. The ability to generate and broadly disseminate knowledge is increasingly dispersed among the masses as well as researchers and academics. Further accelerated by information technologies, the current research ecosystem includes cross-sectoral collaborations, virtual research environments, and new forms of data—all of which are shifting how research is undertaken. The sheer volume of information that now exists makes it virtually impossible to keep up with everything that is published in both peer-reviewed and grey literature. Moreover, and perhaps more importantly, the proliferation of knowledge is challenging the traditional role of researchers and their relationship to knowledge and discovery. If research is a systematic way to reveal truth and generate knowledge, then how does health related research retain both its relevance and value in today's context? One option for researchers to consider may be to actively engage the crowd's skills and knowledge in a way that compliments the researchers' own expertise.

A growing number of researchers are employing crowdsourcing as part of their methods, and are leveraging it to inform everything from study design, to participant recruitment and analysis. A systematic review of health research crowdsourcing found that crowdsourcing has been used for problem solving, data processing, surveillance/monitoring and surveying (Ranard et al., 2014). As a relatively new phenomenon in research, the deployment of crowdsourcing challenges existing philosophical assertions about the nature of knowledge, as well as practical considerations related to the quality of methods and data. This innovation, while transforming the way research is undertaken, raises fundamental questions about the relationship between researchers and participants, researchers and knowledge creation, ownership and the democratization of knowledge, knowledge mobilization, and the methodological evolution of scientific inquiry. Researchers studying and engaging in crowdsourcing will inevitably have to consider its implications, not only in the context of their research, but also more broadly.

The evolving research funding-environment, the proliferation of knowledge, and the appearance of crowdsourcing in the methods sections of research studies, presents an opportunity to explore the potential of crowdsourcing as a research method or methodology. In the absence of studies that focus solely on crowdsourcing in the context of research methods, this dissertation aims to contribute to quantitative and qualitative methods literature by exploring how crowdsourcing can be used in studies.

### 1.3 Literature Review

A specific search of 'crowdsourcing' in the health literature was undertaken by generating a list of potential published studies for inclusion by using the keyword terms "crowdsource*" and ("medical" or "health") with the following filters: English, peer-reviewed and full-text. This search resulted in 87 articles identified in PubMed, and 142 articles in

Proquest for a total of 224 articles. After the removal of editorial, opinion and comment pieces, and the removal of non-crowdsourcing studies following a review of abstracts, a total of 48 articles remained.

### 1.3.1 What is crowdsourcing?

The term "crowdsourcing" was first introduced by Jeff Howe in an article published by Wired Magazine, "The Rise of Crowdsourcing" (2006). According to Howe,

> *Crowdsourcing represents the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call. This can take the form of peer-production (when the job is performed collaboratively), but is also often undertaken by sole individuals. The crucial prerequisite is the use of the open call format and the large network of potential laborers* (Howe, 2009).

Later, Howe (2009) simplified the definition to state, "Crowdsourcing is the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call". Initially, Howe did not apply or limit the concept to focus on its features related to technology or its virtual nature. Rather, Howe emphasises the relationship between the company and the individuals undertaking the work, and the nature of the work itself. In Howe's (2009) third and abridged definition of crowdsourcing, he began to trace the roots of crowdsourcing to the more recent open source-code hacker movement. This software development was associated with the development of Gnu's Not Linux (GNU) and Linux systems, wherein thousands of individual developers contributed code and to create new products, or improve upon existing ones (Howe, 2009).

Since its introduction, numerous others have attempted to refine the definition of "crowdsourcing" and unpack its underlying components in order to draw a distinction between the term and its application. Estelles-Arolas and Gonzalez-Ladron-de-Guevara (2012) identified 40 unique definitions for the term, and coined an integrated definition that deemed 'crowdsourcing' an open call for voluntary participative, online activity, whereby the crowd contributes expertise, knowledge, skills and/or money, to perform a mutually beneficial function. While this definition included "crowdfunding" (the contribution of monetary resources) as an instance of crowdsourcing, most authors, including the author of this paper, exclude crowdfunding from the definition of crowdsourcing. Estelles-Arolas and Gonzalez-Ladron-de-Guevara (2012) further consolidated the research to identify critical components of crowdsourcing (p. 197):

(a) the crowd is clearly defined;

(b) there exists a task with a clear goal;

(c) the recompense received by the crowd is clear;

(d) the crowdsourcer is clearly identified;

(e) the compensation to be received by the crowdsourcer is clearly defined;

(f) it is an online assigned process of participative type;

(g) it uses an open call to a variable extent; and,

(h) it uses the Internet.

Noteworthy here is the introduction of the online and/or Internet provision as a characteristic of crowdsourcing, which may reflect the authors' respective fields of study in information technology, the historical open source roots in software development and the

evolution of crowdsourcing itself vis a vis technology. These critical components will be used to better understand crowdsourcing in the research context.

The lack of consensus around the definition of crowdsourcing can be attributed to a number of factors: (1) the relatively short history of the concept; (2) the evolution of the term to fit established processes ad-hoc; and (3) the application of crowdsourcing to a broad number of disciplines, each of which has its own lexicon particular to its practice. As crowdsourcing is applied in various disciplines, it is adapted to compliment discipline specific concepts.

As a point of clarification, the term "citizen science" is frequently used synonymously with crowdsourcing. Citizen science is defined as "a form of research collaboration involving members of the public in scientific research projects to address real-world problems" (Wiggins & Crowston, 2011, p. 1). Examples of projects that have been classified as either "crowdsourcing" and/or "citizen science", depending on the author, include:

- The Longitude Prize (£20,000) established in 1714 by Britain's Parliament, which is offered to anyone who could solve the problem of identifying a ship's longitudinal position within 30 miles (Ranard et al., 2014); and
- The National Audubon Society's annual bird count, which started over 100 years ago. The Society now has more than 60,000 volunteers of all ages, races, and levels of expertise, who go to more than 2,200 locations throughout the United States, Canada, Central and South America, the West Indies and the Pacific Islands, to identify and count every individual bird they see. More than 60 million birds have been documented thus far (National Audubon Society, 2014).

What is significant here is the fact that organizations have been able to mobilize and actively engage people to achieve various goals, well before the use of digital and electronic media. For the purposes of this dissertation, the terms will be used synonymously as they are frequently treated as such by researchers.

### 1.3.2 Theoretical Foundations

Much like the definition of crowdsourcing itself, the theoretical foundation for the concept varies depending on the context or field of study from which it is viewed. Indeed, there is a general sense that there is a lack of theory associated with crowdsourcing (Zhao & Zhu, 2014). As Howe (2009) puts it: "the Internet is catalyzing change so fast that theory is struggling to keep up" (p. 169); we are essentially applying theory to what is already occurring.

Despite the lack of clarity, many authors, including Howe and Brabham, cited *The Wisdom of the Crowd* (Surowiecki, 2004) to provide a degree of theoretical underpinning to the concept of crowdsourcing. Building on the works of Sir Francis Galton, Surowiecki (2004) states that "under the right circumstances, groups are remarkably intelligent, and are often smarter than the smartest people in them" (p. 64). Howe (2009) elaborates this point further, and surmises that "crowdsourcing is rooted in a fundamental egalitarian principle: every individual possesses some knowledge or talent that some other individual will find valuable" (p. 134).

Surowiecki (2004) provides examples intended to demonstrate the collective intelligence of crowds. These examples - ranging from guessing the number of jelly beans in a jar, to Iowa Electronic Markets used to predict election results - demonstrate instances where the crowd effectively addressed cognition, coordination, and cooperation problems. Further, Surowiecki (2004) identifies four characteristics of a "wise crowd": (1) diversity of opinions; (2) independence (people's opinions are their own and not influenced by others); (3) decentralization

(people can draw on specialized and local knowledge); and (4), aggregation (turns individual decisions into collective decisions).

Given the links between technology and crowdsourcing, Brabham (2008) cites the work of Levy on cyber culture and its role in connecting people, fostering creation, and communication, as the basic building blocks for crowdsourcing. According to Levy, "no one knows everything, everyone knows something, [and] all knowledge resides in humanity" (as cited in Brabham, 2008, p. 247). In order to address this fact about knowledge, we must consciously adopt the technologies and methods that harness this collective talent. For example, technologies embedded in personal devices (e.g., smart phones) "enable public and professional users to gather, analyze and share local knowledge" (Kamel Boulos et al., 2011, p. 5). Advances in computer technologies, and proliferation of connected applications and devices has, inarguably, scaffolded crowdsourcing to a degree that was previously not possible. By breaking down institutional boundaries, and eliminating geographic constraints, crowdsourcing has the potential to garner public participation in new and exciting ways. The origins of crowdsourcing are rooted in a variety of disciples, and has thus benefited from their respective strengths.

### 1.3.3 Crowdsourcing Typologies

Various types and forms of crowdsourcing, ranging from task-based functions, to idea generation and problem solving, attract different crowds of individuals based on their interests, skills and motivations. Howe (2009) identified four types of roles taken on by the crowd: (1) the professional; (2) the packager; (3) the tinkerer; and, (4) the masses. This categorization has evolved over time into various function or output-based typologies of crowdsourcing, but articulating a single typology of crowdsourcing has proven to be complicated. Table 1.1 outlines different types of crowdsourcing that have been defined by various authors.

**Table 1.1. Crowdsourcing Typologies**

| Author | Discipline/Paradigm | Typology |
|---|---|---|
| Howe (2009) | Business | 1. Prediction or information markets<br>2. Problem-solving<br>3. "Idea jam" (generate new ideas) |
| Brabham (2010) | Urban Planning | 1. Knowledge discovery and management (collecting information<br>2. Broadcast search (ideation problems with empirical solutions)<br>3. Peer-vetted creative production (creating and selecting creative ideas)<br>4. Distributed human intelligence tasking (analyzing large amounts of data) |
| Zhao & Zhu (2014) | Information Systems | 1. Design and development<br>2. Test and evaluation<br>3. Idea and consultation<br>4. Other |
| Parvanta, Roth, & Keller (2013) | Health Communications & Behavioural Studies | 1. Crowd research—gathering insights/data from the audience<br>2. Crowd labour—individuals recruited to perform specific tasks<br>3. Creative crowdsourcing—generate ideas<br>4. Crowdfunding—open call to raise money |

While there is no single accepted framework to identify all the various types of crowdsourcing, there appears to be general agreement across the literature that crowdsourcing activities range from micro, repetitive, task-based activities (such as data collection or analysis), to creative tasks (generating new ideas) and problem-solving (solutions based).

The nature of the task being crowdsourced will inherently determine the crowd it requires (and draws). As Howe (2009) noted, in order to solve the problem being put forward, the crowd

must have the appropriate qualifications. Given the significant numbers of organizations leveraging crowdsourcing and the number of people who participate, this raises the question as to who participates in crowdsourcing and why?

### 1.3.4 Examples of Crowdsourcing Across Multiple Domains

The four most frequently referenced examples of the earliest and most successful crowdsourcing activities which demonstrate the application of crowdsourcing in the private sector come from iStockphoto, Threadless, Amazon Mechanical Turk and InnoCentive. A brief summary of each of these is provided below.

iStockphoto is a royalty-free stock images/photographs/video company where approximately 50 000 plus artists/ photographers share their work (http://www.istockphoto.com/) and make a small profit every time their work is downloaded. Many of these contributors are part-time amateur photographers and hobbyists who have disrupted the professional photography market (Brabham, 2008). Another variation of crowdsourcing in business comes from Threadless, an online t-shirt company that holds regular competitions whereby designers submit their ideas for t-shirts, the crowd votes on them, and the company produces them and pays the winning designers for their ideas (Brabham, 2010). In addition, websites such as Amazon Mechanical Turk and InnoCentive provide access to a large, diverse, on-demand, scalable crowd from around the globe. Amazon Mechanical Turk is an online marketplace for work that requires human intelligence and cannot rely on artificial/computer intelligence such as identifying objects in a photo or video, performing data de-duplication, transcribing audio recordings, or researching data details (https://www.mturk.com).

InnoCentive enables organizations to post problems and challenges they are facing and has been used by companies such as Astra Zeneca, Cleveland Clinic, Eli Lilly & Company,

NASA, Procter & Gamble, The Economist, The Department of Defense, and other government agencies in the U.S. and Europe (http://www.innocentive.com).

In each of these instances, crowdsourcing disrupted each sector by evolving the business model through access to knowledge and expertise that exists outside of each institution. In addition, examples of crowdsourcing from academia and government also demonstrate how the crowd can support public initiatives and programs. For example, following the 2011 tsunami in Japan, it was critical for officials to monitor the spread of radiation resulting from the severely damaged Fukushima-Daiichi nuclear power plant. A team of researchers designed the Japan Nuclear Crowd Map (JNCM) to monitor and map real-time radiation data (Kamel Boulos et al., 2011). Within two weeks of the disaster, individual citizens had deployed 577 Geiger counters across the country to help the monitor and track the spread of the nuclear cloud (University of Southampton, 2013). The map combines sensor information with and crowdsourced radiation data and has provide more than 27 million readings since the day of the Fukushima disaster (University of Southampton, 2013). A key incentive for people to take part in crowdsourcing projects is to help them understand these large quantities of data by feeding back the results to the data contributors (University of Southampton, 2013). In this example, the geographically dispersed crowd was able to distribute Geiger counters across the country in a much more efficient and timely manner than would have been the case if the researchers had to deploy the counters themselves. The crowd was also able to provide a large amount of real-time information for monitoring consistently over a prolonged period of time. Other examples, briefly highlighted below, include the Cities at Night project and United States Patent and Trademark Office.

The International Space Station has captured approximately two million images of Earth, and while the images are clear, the specifics in the images are not always easy to determine without analysis and categorization therefore rendering them useless for scientific purposes (NASA, 2015). The Complutense University of Madrid (UCM) launched a project called Cities at Night to catalog the images so they could be used to create an open atlas (NASA, 2015). Given the large number of images and the volume of work that would be required, UCM researchers decided to engage the crowd. In an open call, people were invited to undertake three tasks of varying degrees of complexity including sorting images, and identifying locations and cities (NASA, 2015). To date, approximately 20,000 images have been categorized by hundreds of volunteers. To ensure accuracy, each image is being categorized multiple times by different individuals (NASA, 2015). In addition to creating the atlas, the project will also help determine the optimum number of individuals required to assess each image (NASA, 2015). In this example, the crowd is able to bring their collective interest in space, and their individual geographic knowledge, to help contribute to the analysis and identification of the images as part of the larger project. This project may not have otherwise been feasible due to its magnitude, and the number of people and time required to complete it.

The United States Patent and Trademark Office ("USPTO") leveraged the crowd to address challenges during its patent crisis (Noveck, 2006). The combination of increasing demand (a backlog of approximately one million applications), limited expertise and time of patent examiners, and an exposition of information on the Internet, rendered the existing approach ineffective and inefficient (Noveck, 2006). In order to address this gap in human resources and expertise, USPTO launched "Ask Patents" through Stack Exchange, a site that engages the public interested in improving and participating in the US patent system

(https://patents.stackexchange.com/). This site allows people to share information on "prior art," which is typically an onerous and time consuming task for patent reviewers. Users can help by either identifying dangerous patent applications or by finding prior art related to dangerous patent applications. The system then blocks applications by alerting the patent examiner when good prior art is found for an application that he or she is reviewing (Noveck, 2006). In this example, the crowd contributes information based on their knowledge, essentially increasing the capacity of the patent's office by supporting the work of patent reviewers, and improving efficiency of the review process and ensuring greater rigour in identifying prior art.

### 1.3.5 Understanding the Crowd

Any further understanding of the nature of crowdsourcing calls for elaboration on the characteristics of the crowd (e.g. size, age, education, and profession) and their motivations for participation. Due to the limited information available, the analysis of this point is necessarily preliminary. However, given its importance to understanding "crowdsourcing" as a methodology, it is worth elaborating on as much as the modest amount of information available will permit.

Howe (2009) makes two assertions about the crowd related to its size and qualifications. First, the crowd includes the billions of people who have access to the Internet. Second, "the crowd must have some level of qualification" to solve the problem being put forth. Each of these will be addressed in turn.

Based on their review of the literature, Estelles-Arolas and Gonzalez-Ladron-de-Guevara (2012) identify a characteristic of "the crowd" as being a large and undetermined number of people. This appears to align with Howe's (2009) suggestion that everyone who has access to the Internet could be a part of the crowd. However, access to (and use of) the Internet is limited, a

fact that calls into question the actual diversity of the crowd. Approximately 3 billion people currently use the Internet, which accounts for roughly 40% of the world's population. In 2014, nearly 75% of Internet users lived in 20 countries. Moreover, English remains the dominant language used on the Internet (Internet Live Stats, 2015).

Canadians are the second heaviest users of the Internet, following the US, spending an average of 45.6 hours online per month (Canadian Internet Registration Authority, 2014). In 2012, 83% of Canadian households had access to the Internet, with 85% of households located in metropolitan areas having home Internet access, compared with 75% of households outside these areas (Statistics Canada, 2013). The urban/rural gap is even more pronounced in the Canadian North. A 2010 report showed that while 83.5% of households in the Northwest Territories (NWT) and 100% of communities in the Yukon had Internet access, only 27% of communities in Nunavut had access (Statistics Canada, 2013). In the NWT, community-level Internet access ranged from 17% in the tiny hamlet of Wrigley, to 89.9% in Yellowknife (Canadian Internet Registration Authority, 2014). Further, approximately 95% of Canadians in the highest income quartile are connected to the Internet, while only 62% in the lowest income quartile have Internet access (Statistics Canada, 2013). Other international studies on the digital divide and Internet use suggest that the typical web user is white, middle- or upper-class, English-speaking, has higher levels of education, and is equipped with a high-speed connection (Brabham, 2008).

When it comes to online content creation, we know that a relatively small portion of participants account for vast majority of content when it comes to crowdsourcing (Zhao & Zhu, 2014). This builds on the 1:10:89 rule whereby 1/100 people will create something, 10/100 will vote on it and 89/100 will simply consume it (Howe, 2009, p. 227). Therefore, while crowdsourcing has the potential to draw a large number of participants, a single crowd may not

be considered broadly representative, as only a small segment of a given crowd actively participates, and an even smaller segment creates content.

Age also appears to play a role in content creation, and certain authors identify the most productive individuals in the crowd as likely to be young, most likely under the age of 30, and probably under the age of 25 (Lenhart, Fallows, & Horrigan, 2004; Lenhart & Madden, 2005). Given that the under-30 age group is the most active in the so-called Web 2.0 environment of massive content creation, they also seem to be a reasonable cohort of active members within the crowd (Fox & Madden, 2006; Raine, 2005). Although these generalizations may hold true, it is difficult to draw conclusions related to the age of crowdsourcing contributors, based on the limited demographic information available.

The small number of studies on specific crowdsourcing initiatives provides insight into the educational and professional backgrounds of participants. For example, problem solvers on InnoCentive are well-educated, with a majority (65.8%) self-reporting to hold a PhD (Lakhani, Lohse, Panetta, & Jeppesen, 2007). Meanwhile, a study of participants on iStock by Brabham (2008) found that the most common occupations identified by users were being self–employed (30.2%) and having a professional or technical background (28.2%), while professional photographers and designers only comprised 3.9% of users. Participants on iStock were well educated, with 77.6% of respondents having completed at least a U.S. associate's degree (or an equivalent two–year, post–high school degree) and 43.5% holding Bachelor's degrees. Finally, high–speed home Internet connections were extremely commonplace (97.4%), and 98.3% of participants considered themselves skillful at using the Web (Brabham, 2008). The crowd for the Next Stop Design, an urban planning initiative, ranged from architects to engineers to architect teachers, along with many who had previously studied architecture (Brabham, 2013). These three

examples would supports Howe's (2009) assertion that the crowd needs to have some level of qualifications to adequately participate and address the issue at hand.

Another characteristic of a crowd is the actual role they perform in relationship to the crowdsourcing host. Howe (2009) suggests that crowdsourcing blurs the lines between consumer and producer. Interestingly, the crowd at Threadless does exactly this by essentially acting as the producers, decision-makers, and consumers of the product (Brabham, 2010; Howe, 2009). While there are other examples such as the Doritos' Crash the Super Bowl campaign (https://crashthesuperbowl.doritos.com/) and Dell Ideastorm (http://www.ideastorm.com/) where the role of the consumer has evolved into advertiser or designer, these initiatives have not be formally researched and documented.

Unfortunately, it is difficult to compare the characteristics and composition of different crowds for various forms of crowdsourcing due to the limited information available. It is important to note that as demographics change and Internet usage around the world shifts, the make up of the crowd will also likely evolve. In addition, the rapid evolution of technology will likely play a role in what future crowdsourcing activities will aim to achieve and who they will attract. Needless to say, regardless of the crowds' composition, their motivations are of equal interest.

### 1.3.6 Motivations of the Crowd

In parallel with the information available on crowdsourcing contributors, research into the motivation of participants remains limited as well (Smith, Manesh, & Alshaikh, 2013). Where there is research available, there are conflicting findings when it comes determining what motivates participation, which once again appears to be context specific (Brabham, 2010; Smith et al., 2013; Zhao & Zhu, 2014). Different authors have developed different frameworks for

identifying motivators based on various characteristics of crowdsourcing. At the most rudimentary level, Vukovic and Bartolini (2010) differentiate the crowd into internal and external groups based on their relationship with the organization hosting the crowdsourcing. An internal crowd represents employees of the organization who are motivated by career and professional advancement as well as recognition, while an external crowd is comprised of members outside of the organization who is motivated by other rewards (Vukovic & Bartolini, 2010). In this instance, the motivation is based on the nature of the relationship been the individual participant and the host organization.

Drawing from his own work and existing studies, Brabham (2010) identifies the following motivations for individuals in crowds: the desire to

- earn money;
- develop one's creative skills;
- network with other creative professionals;
- build a portfolio for future employment;
- challenge oneself to solve a tough problem;
- socialize and make friends;
- pass the time when bored;
- contribute to a large project of common interest;
- share with others; and
- have fun.

This list of motivators has been categorized and collapsed by some authors. For example, Parvanta et al. (2013) summarize motivations for participating in crowdsourcing as the "Four Fs"—fun, fulfillment, fame and fortune. The notions of socialization, contributing to a larger

project, and fulfillment may feed into the characterization of participant motivation, rooted in psychology. Seltzer and Mahmoudi (2013) suggest that participation can be motivated by the intrinsic needs of an individual (e.g. fulfilling cognitive, emotional and social desires) and/or extrinsic factors (e.g. public recognition). The sense of community created by the crowd likely has greater meaning for some individuals than it does others and the drivers of participation in crowdsourcing activities requires further exploration, particularly in the context of research studies.

### 1.3.7 Crowdsourcing in Health-Related Research

There have been several research projects that have used crowdsourcing as part of their methodology, including protocol design, participant recruitment, and data analysis. The following examples illustrate the numerous ways in which health researchers are engaging the crowd.

From a protocol design perspective, engaging participants (usually patients), may encourage participation in the research study. On the other hand, researchers want to ensure that the protocol maintains sufficient integrity, and that the project is very specific in terms of what is being asked of participants. Examples of engaging the crowd in the design of a research protocol include seeking input into the development of a cancer clinical trial (Leiter et al., 2014), to patient-led research related, to the effects of lithium use in (Amyotrophic Lateral Sclerosis) ALS patients (Swan, 2012a). Websites such as PatientsLikeMe.com (PLM), with open patient registries, are increasingly driving patient-directed studies and self-experimentation—whereby the researcher acts as an advisor or engages in the clinical trial after preliminary results from patient-directed studies show promise (Swan, 2012a). User-driven research can accelerate and improve the innovation adoption process of a solution or new knowledge (Celi et al., 2014).

While this approach aligns with qualitative methodologies such as critical theory or participatory action more readily, it is interesting to contemplate how user-driven research effects quantitative research principles and processes which tend to be more structured.

Crowdsourcing is also being used to recruit study participants as subjects for research studies or as participants who contribute to the overall project. For example, participants in many studies either brought knowledge and experience that the researchers required, and/or helped the researchers with a specific task such as identifying predictors of obesity or evaluating literature (Bevelander et al., 2014; Brown & Allison, 2014). In instances where participants were contributing knowledge, researchers frequently conducted assessments to ensure they were qualified to partake in the task at hand. Furthermore, the crowd can also provide access to real-time, geographically specific data, which is particularly important in the case of infectious disease surveillance (Chunara, Smolinski, & Brownstein, 2013; Kamel Boulos et al., 2011).

Although crowdsourcing provides access to large pools of potential participants, issues with population representativeness and self-selection nevertheless need to be addressed. From a data perspective, self-reported data can be fraught with issues of accuracy and validity. Various mechanisms have been identified to address these issues, including: bringing reported data together with diagnostic or other clinical measures (Chunara et al., 2013); in-house calculations and physician verification (Swan, 2012b); and, reputation metrics for evaluating user-generated content (McCoy et al., 2014). There are studies that suggest that the quality of self-reported or crowdsourced data is as good researcher collected data if not better and is also comparable to other types of sampling (Behrend, Sharek, Meade, & Wiebe, 2011; Swan, 2012a).

Other examples of crowdsourcing in research relate to data analysis and problem solving, such as one study on the disorder ALS. Prize4Life and the Neurological Clinical Research

Institute (NCRI) at Massachusetts Hospital hosted a competition that invited the crowd to develop algorithms for the prediction of disease progression of 1,822 ALS patients from standardized, anonymized phase 2/3 clinical trials (Küffner et al., 2015). More than 1000 participants were involved in the challenge, resulting in 37 potential solutions and ultimately two teams securing first prize (Küffner et al., 2015). The two best algorithms outperformed a method designed by the challenge organizers as well as predictions by ALS clinicians and estimates, suggesting that using both winning algorithms in future trial designs could reduce the required number of patients by at least 20% (Küffner et al., 2015). This crowdsourcing competition enabled researcher to harness the collective intelligence of a team of researchers and an individual who was external to the team, to potentially improve the lives of people living with ALS.

Crowdsourcing has been deemed successful by all the authors whose works are noted; however, they all acknowledge its limitations from a research methodical perspective such as the lack of sample representativeness and self-selection bias. The advantages and benefits for researchers include access to large volumes of data and information, access to resources and low cost, and novel science (Ranard et al., 2014). As a result, from a research prospective the crowd has much to offer, and researchers are beginning shift their paradigms from engaging individuals qua subjects to individuals qua active participants.

## 1.4 Overview of Dissertation

This thesis is organized into five related chapters. In addition to setting the context for the dissertation, Chapter 1 also consists of a literature review of crowdsourcing, and lays the foundation for the remainder of the study. Chapters 2, 3 and 4 explore crowdsourcing as a research paradigm and form of research inquiry, describe how researchers are using

crowdsourcing in practice, and finally develop a framework for how crowdsourcing can be used in research. Chapter 2 explores the core principles and philosophies of crowdsourcing as a research paradigm. This paper examines crowdsourcing as a form of inquiry by considering questions of ontology, epistemology, methodology and axiology. Chapter 3 reviews a series of research projects on the Crowdcrafting website (http://crowdcrafting.org) with the aim of understanding how crowdsourcing is being used in practice. This two-stage process first seeks to understand the type of tasks the crowd is undertaking, and then maps the projects against the Estellés-Arolas and González-Ladrón-de-Guevara (2012). The ultimate goal of this chapter is to describe the extent to which the identified criteria can be applied within a research context. Chapter 4 presents findings from a Delphi panel of experts – researchers who have used, or are knowledgeable about, crowdsourcing – and builds a framework/guidelines for researchers to consider when deploying crowdsourcing in their research. Chapter 5 concludes the dissertation with an overview of the work described in Chapters 2 through 4, and also presents some final thoughts on crowdsourcing as a research paradigm and method.

## 1.5 References

Armstrong, A. W., Cheeney, S., Wu, J., Harskamp, C. T., & Schupp, C. W. (2012). Harnessing the power of crowds: Crowdsourcing as a novel research method for evaluation of acne treatments. *American Journal of Clinical Dermatology*, *13*(6), 405–416. http://doi.org/10.2165/11634040-000000000-00000

Behrend, T. S., Sharek, D. J., Meade, A. W., & Wiebe, E. N. (2011). The viability of crowdsourcing for survey research. *Behavior Research Methods*, *43*(3), 800–813. http://doi.org/10.3758/s13428-011-0081-0

Benner, M., & Sandström, U. (2000). Institutionalizing the triple helix: research funding and

norms in the academic system. *Research Policy*, *29*(2), 291–301. http://doi.org/10.1016/S0048-7333(99)00067-0

Bevelander, K. E., Kaipainen, K., Swain, R., Dohle, S., Bongard, J. C., Hines, P. D. H., & Wansink, B. (2014). Crowdsourcing novel childhood predictors of adult obesity. *PLoS ONE*, *9*(2), e87756. http://doi.org/10.1371/journal.pone.0087756

Brabham, D. C. (2008). Crowdsourcing as a Model for Problem Solving: An Introduction and Cases. *Convergence: The International Journal of Research into New Media Technologies*, *14*(1), 75–90. http://doi.org/10.1177/1354856507084420

Brabham, D. C. (2010). Moving the Crowd at Threadless: Motivations for participation in a crowdsourcing application. *Information, Communication & Society*, *13*(8), 1122–1145. http://doi.org/10.1080/13691181003624090

Brabham, D. C. (2013). Crowdsourcing: A model for leveraging online communities. In A. Delwiche & J. J. Henderson (Eds.), *The Participatory Cultures Handbook* (pp. 120–129). New York, NY: Routledge.

Brown, A. W., & Allison, D. B. (2014). Using crowdsourcing to evaluate published scientific literature: Methods and example. *PLoS ONE*, *9*(7), e100647. http://doi.org/10.1371/journal.pone.0100647

Canadian Internet Registration Authority. (2014). The 2014 CIRA Factbook. Retrieved January 28, 2015, from http://www.cira.ca/factbook/2014/the-canadian-internet.html

Celi, L. A., Ippolito, A., Montgomery, R. A., Moses, C., & Stone, D. J. (2014). Crowdsourcing knowledge discovery and innovations in medicine. *Journal of Medical Internet Research*. http://doi.org/10.2196/jmir.3761

Chunara, R., Smolinski, M. S., & Brownstein, J. S. (2013). Why we need crowdsourced data in

infectious disease surveillance. *Current Infectious Disease Reports*, *15*(4), 316–319. http://doi.org/10.1007/s11908-013-0341-5

Coles, P., Cox, T., Mackey, C., & Richardson, S. (2006). The toxic terabyte. *Technology*, (July), 1–12.

Coley, H. L., Sadasivam, R. S., Williams, J. H., Volkman, J. E., Schoenberger, Y. M., Kohler, C. L., … Houston, T. K. (2013). Crowdsourced peer- versus expert-written smoking-cessation messages. *American Journal of Preventive Medicine*, *45*(5), 543–550. http://doi.org/10.1016/j.amepre.2013.07.004

Densen, P. (2011). Challenges and Opportunities Facing Medical Education. *Transactions of the American Clinical and Climatological Association*, *122*(319), 48–58. http://doi.org/10.4172/2155-9619.S6-009

Domecq, J. P., Prutsky, G., Elraiyah, T., Wang, Z., Nabhan, M., Shippee, N., … Murad, M. H. (2014). Patient engagement in research: a systematic review. *BMC Health Services Research*, *14*(1), 89. http://doi.org/10.1186/1472-6963-14-89

Estellés-Arolas, E., & González-Ladrón-de-Guevara, F. (2012). Towards an integrated crowdsourcing definition. *Journal of Information Science*, *38*(2), 189–200. http://doi.org/10.1177/0165551512437638

Fox, S., & Madden, M. (2006). Riding the Waves of "Web 2.0." Retrieved from http://www.pewinternet.org/2006/10/05/riding-the-waves-of-web-2-0/

Frank, L., Forsythe, L., Ellis, L., Schrandt, S., Sheridan, S., Gerson, J., … Daugherty, S. (2015). Conceptual and practical foundations of patient engagement in research at the patient-centered outcomes research institute. *Quality of Life Research*, *24*(5), 1033–1041. http://doi.org/10.1007/s11136-014-0893-3

Fuller, R. B., & Kuromiya, K. (1981). *Critical Path* (1st ed.). New York, NY: St. Martin's Press.

Graham, I. D., & Tetroe, J. M. (2009). Getting evidence into policy and practice: perspective of a health research funder. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, *18*(1), 46–50.

Howe, J. (2006). The rise of crowdsourcing. *Wired Magazine*, (14), 1–5.

Howe, J. (2009). *Crowdsourcing: Why the power of the crowd is driving the future of business*. New York: Random House.

Internet Live Stats. (2015). Internet Users. Retrieved January 28, 2015, from http://www.internetlivestats.com/internet-users/

Kamel Boulos, M. N., Resch, B., Crowley, D. N., Breslin, J. G., Sohn, G., Burtner, R., … Chuang, K.-Y. S. (2011). Crowdsourcing, citizen sensing and sensor web technologies for public and environmental health surveillance and crisis management: trends, OGC standards and application examples. *International Journal of Health Geographics*, *10*(1), 67. http://doi.org/10.1186/1476-072X-10-67

King, A. J., Gehl, R. W., Grossman, D., & Jensen, J. D. (2013). Skin self-examinations and visual identification of atypical nevi: Comparing individual and crowdsourcing approaches. *Cancer Epidemiology*, *37*(6), 979–984. http://doi.org/10.1016/j.canep.2013.09.004

Küffner, R., Zach, N., Norel, R., Hawe, J., Schoenfeld, D., Wang, L., … Leitner, M. L. (2015). Crowdsourced analysis of clinical trial data to predict amyotrophic lateral sclerosis progression. *Nature Biotechnology*, *33*(1), 51–7. http://doi.org/10.1038/nbt.3051

Kuhn, T. S. (1970). *The Structure of Scientific Revolutions* (2nd ed.). Chicago, IL: The University of Chicago Press.

Lakhani, K. R., Lohse, P. A., Panetta, J. A., & Jeppesen, L. B. (2007). *The Value of Openness in*

*Scientific Problem Solving*. Division of Research, Harvard Business School. http://doi.org/10.1007/s00268-006-0616-1

Lavis, J. N., Lomas, J., Hamid, M., & Sewankambo, N. K. (2006). Assessing country-level efforts to link research to action. *Bulletin of the World Health Organization*, *84*(8), 620–8. http://doi.org/10.2471/BLT.06.030312

Légaré, F., Stacey, D., Forest, P.-G., & Coutu, M.-F. (2011). Moving SDM forward in Canada: milestones, public involvement, and barriers that remain. *Zeitschrift Für Evidenz, Fortbildung Und Qualität Im Gesundheitswesen*, *105*(4), 245–253. http://doi.org/10.1016/j.zefq.2011.04.011

Leiter, A., Sablinski, T., Diefenbach, M., Foster, M., Greenberg, A., Holland, J., … Galsky, M. D. (2014). Use of crowdsourcing for cancer clinical trial development. *Journal of the National Cancer Institute*, *106*(10), 8–10. http://doi.org/10.1093/jnci/dju258

Lenhart, A., Fallows, D., & Horrigan, J. B. (2004). Content Creation Online. Retrieved from http://www.pewinternet.org/2004/02/29/content-creation-online/

Lenhart, A., & Madden, M. (2005). Teen Content Creators and Consumers. Retrieved from http://www.pewinternet.org/2005/11/02/teen-content-creators-and-consumers/

McCoy, A. B., Wright, A., Rogith, D., Fathiamini, S., Ottenbacher, A. J., & Sittig, D. F. (2014). Development of a clinician reputation metric to identify appropriate problem-medication pairs in a crowdsourced knowledge base. *Journal of Biomedical Informatics*, *48*, 66–72. http://doi.org/10.1016/j.jbi.2013.11.010

NASA, U. S. G. (2015). Space station sharper images of earth at night crowdsourced for science.

National Audubon Society. (2014). History of the Christmas bird count. Retrieved January 31, 2017, from http://www.audubon.org/history-christmas-bird-count

Noveck, B. S. (2006). “Peer to Patent”: Collective intelligence, open review and patent reform. *Harvard Journal of Law & Technology*, *20*(1).

Parvanta, C., Roth, Y., & Keller, H. (2013). Crowdsourcing 101: A Few Basics to Make You the Leader of the Pack. *Health Promotion Practice*, *14*(2), 163–167. http://doi.org/10.1177/1524839912470654

Raine, L. (2005). The State of Blogging. Retrieved from http://www.pewinternet.org/2005/01/02/the-state-of-blogging/

Ranard, B. L., Ha, Y. P., Meisel, Z. F., Asch, D. A., Hill, S. S., Becker, L. B., … Merchant, R. M. (2014, January). Crowdsourcing: Harnessing the masses to advance health and medicine, a systematic review. *Journal of General Internal Medicine*. http://doi.org/10.1007/s11606-013-2536-8

Seltzer, E., & Mahmoudi, D. (2013). Citizen Participation, Open Innovation, and Crowdsourcing: Challenges and Opportunities for Planning. *Journal of Planning Literature*, *28*(1), 3–18. http://doi.org/10.1177/0885412212469112

Shippee, N. D., Domecq Garces, J. P., Prutsky Lopez, G. J., Wang, Z., Elraiyah, T. A., Nabhan, M., … Murad, M. H. (2015). Patient and service user engagement in research: a systematic review and synthesized framework. *Health Expectations*, *18*(5), 1151–1166. http://doi.org/10.1111/hex.12090

Smith, D., Manesh, G. M. M., & Alshaikh, A. (2013). How Can Entrepreneurs Motivate Crowdsourcing Participants? *Technology Innovation Management Review*, *3*(2), 23–30.

Statistics Canada. (2013). Canadian Internet Use Survey (2012). Retrieved January 28, 2015, from http://www.statcan.gc.ca/daily-quotidien/131126/dq131126d-eng.htm

Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few, and how*

*collective wisdom shapes business, economies, societies, and nations*. New York: Anchor Books.

Swan, M. (2012a). Crowdsourced Health Research Studies: An Important Emerging Complement to Clinical Trials in the Public Health Research Ecosystem. *Journal of Medical Internet Research*, *14*(2), e46. http://doi.org/10.2196/jmir.1988

Swan, M. (2012b). Health 2050: The Realization of Personalized Medicine through Crowdsourcing, the Quantified Self, and the Participatory Biocitizen. *Journal of Personalized Medicine*, *2*(3), 93–118. http://doi.org/10.3390/jpm2030093

Tetroe, J. M., Graham, I. D., Foy, R., Robinson, N., Eccles, M. P., Wensing, M., … Grimshaw, J. M. (2008). Health research funding agencies' support and promotion of knowledge translation: An international study. *Milbank Quarterly*, *86*(1), 125–155. http://doi.org/10.1111/j.1468-0009.2007.00515.x

Turner-McGrievy, G. M., Helander, E. E., Kaipainen, K., Perez-Macias, J. M., & Korhonen, I. (2015). The use of crowdsourcing for dietary self - monitoring: Crowdsourced ratings of food pictures are comparable to ratings by trained observers. *Journal of the American Medical Informatics Association*, *22(e1)*, e112-9. http://doi.org/10.1136/amiajnl-2014-002636

University of Southampton. (2013). ECS researchers at Southampton University develop new tool to provide radiation monitoring in Japan. Retrieved January 31, 2017, from http://www.ecs.soton.ac.uk/news/4237

Vukovic, M., & Bartolini, C. (2010). Towards a Research Agenda for Enterprise Crowdsourcing. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 6415 LNCS, pp. 425–434).

http://doi.org/10.1007/978-3-642-16558-0_36

Wiggins, A., & Crowston, K. (2011). From conservation to crowdsourcing: A typology of citizen science. In *Proceedings of the Annual Hawaii International Conference on System Sciences* (pp. 1–10). http://doi.org/10.1109/HICSS.2011.207

Zhao, Y., & Zhu, Q. (2014). Evaluation on crowdsourcing research: Current status and future direction. *Information Systems Frontiers*, *16*(3), 417–434. http://doi.org/10.1007/s10796-012-9350-4

# Chapter 2

## Crowdsourcing: A Potential Research Paradigm

### 2.1 Overview

Crowdsourcing is now finding its way into the methods section of research studies as a means of developing protocols, recruiting participants, collecting data, and harnessing analytical capabilities. A systematic review of health research found that crowdsourcing has been used for problem solving, data processing, surveillance/monitoring and surveying (Ranard et al., 2014). Examples of crowdsourcing in health related research range from seeking input into the development of a cancer clinical trial protocol (Leiter et al., 2014), to patient-led research related to the effects of lithium use in ALS patients (Swan, 2012). Crowdsourcing is also being used to recruit individuals to serve as both study participants and contributors to research studies. These participants either brought knowledge and experience that the researchers required, or helped them with a specific task, such as identifying predictors of obesity or evaluating literature (Bevelander et al., 2014; Brown & Allison, 2014). The crowd has also provided access to real-time, geographically-specific data which is particularly important in the case of infectious disease surveillance (Chunara, Smolinski, & Brownstein, 2013; Kamel Boulos et al., 2011).

While the limited body of available evidence would appear to suggest that researchers are beginning to deploy crowdsourcing, its application in research has yet to be considered from a research paradigm and methods perspective. This scarcity of information presents an opportunity to examine crowdsourcing as a research paradigm. This paper aims to explore the concept of crowdsourcing as a form of inquiry and method by considering its philosophies and principles. The premise of this discourse assumes that crowdsourcing warrants consideration as a research paradigm and method given that researchers are employing it as such and, therefore, has

implications for the quality and integrity of studies. Further, this paper assumes that quality research hinges upon alignment and cohesion between research methodology and methods. Methodology, here, is defined from a qualitative viewpoint as an overall approach that aligns the research paradigm (or philosophical stance) and methods (process). To anchor this conceptual exploration of crowdsourcing, the research paradigm approach established by Lincoln and Guba (2011) serves as a framework for the methodological discourse.

## 2.2 Crowdsourcing as a Research Paradigm

The basic philosophical foundations of a paradigm rest upon its ontology, epistemology, and methodological position. Ontology is about truth and the nature of reality, addressing questions such as *what is the form and nature of reality?* Epistemology is about the relationship between the inquirer and knower, and addresses questions around the relationship between these two. Then there is the methodology, which is about process and procedure, and aims to address how we uncover what is known (Lincoln & Guba, 2011; Ponterotto, 2005). Heron and Reason (1997) added axiology as a fourth philosophical dimension, important for paradigmatic discussions. Axiology seeks to understand the value of knowledge or, as Ponterotto (2005) suggests, the role of researcher values in the research process. These fundamental concepts create the basis for any paradigm thus each of these will be considered in turn for crowdsourcing based on the existing definitions and theories presented in the previous section. Examples from health research studies have been presented as tangible applications to illustrate the paradigmatic arguments.

### 2.2.1 Contemplating Questions of Ontology and Epistemology

When considering ontological questions regarding the nature of reality, the deconstruction of crowdsourcing reveals many possible perceptions. First and foremost, there is

the belief of the crowdsourcer (in a research context, this would be the researcher) with regards to the nature of reality. This spectrum of perspectives ranges from the objective reality typically espoused by positivists (primarily quantitative researchers), to a more subjective reality endorsed by primarily qualitative researchers. This likely informs the research context in which crowdsourcing is applied and how the crowd is engaged. The perspective of the individuals who construct the crowd will also impact on ontology. Each of these individuals has a view of reality that is shaped by their own beliefs, interactions, and experiences, and this diversity and independence of opinions, experiences and knowledge significantly improves the effectiveness of crowdsourcing (Surowiecki, 2004). By the nature of its engagement, however, the ontology of crowdsourcing is derived from the convergence of diverse beliefs on the nature of reality, through a shared, lived experience. Thus, the crowdsourcing interaction facilities the production of a co-created reality based on the parameters established by researchers and contributions of the crowd. The participatory nature of crowdsourcing embraces each individuals' subjective experience in the objective physical world.

The definitions presented by Braham (2010) and Estellés-Arolas and González-Ladrón-de-Guevara (2012) highlight the participatory and collective nature of the crowdsourcing interaction. The participatory nature of online crowdsourcing results in the creation of a collective subjective reality through the contributions of each individual. This collective subjective reality is experienced differently by each individual, based on their interactions and experiences in the crowdsourcing environment. If one considers the online community to be a natural extension of one's objective world, a subjective reality is created through ones' participation within it. The online community is a shared space in which individuals come together with their own experiences and knowledge. The online community requires the

participation of individuals to exist, and thus it is a participative reality. In this participative reality, individuals collectively co-create reality, within this virtual objective shared space, based on their subjective knowledge.

Thus, we can see that crowdsourcing benefits from a diverse set of opinions, ideas and experiences, that begin with the recognition that individuals amass knowledge through their interactions and experience with (and in) the world. In crowdsourcing, this is the notion that "every individual possesses knowledge or talent" (Howe, 2009) and "all knowledge resides in humanity" (Brabham, 2008). This acknowledgement of the dispersion of knowledge among the masses democratizes and decentralizes intellectual capital, moving it from the few to the many. Knowledge in crowdsourcing is decentralized and people can leverage specialized and local knowledge, while still remaining rooted in their own unique experience (Surowiecki, 2004). Through its collective and participatory nature, crowdsourcing attributes and demonstrates value of the knowledge held among the masses – and in a research context, crowdsourcing enables researchers and participants to co-create knowledge through their respective contributions in the interaction. As both the researcher and individual members of the crowd bring experience and knowledge, the act of crowdsourcing facilitates an open collaboration for problem-solving to co-produce new knowledge.

As the crowdsourcer and the crowd (or the researcher and participants) bring knowledge to this relationship, the prescribed dichotomy of roles evolves. Howe (2009) notes that crowdsourcing blurs the lines between consumer and producer. Each individual has knowledge that others may find valuable regardless of role, profession, or expertise. By its very design, through the co-creation of knowledge, crowdsourcing shifts the relationship between the researcher (the crowdsourcer) and participants (the crowd). Given the participatory nature of

crowdsourcing in its harnessing of collective intelligence, research participants (members of the crowd) have an active role in the co-production of knowledge. This potential evolution of the roles of researcher and participant (e.g., researcher as participant, participant as researcher), with each individual playing various roles based on their expertise and experience, would be contingent upon the nature of problem-solving required by the research and the composition of the crowd. It is possible that in the crowdsourcing context, the researcher and participants act as both knowledge producers and knowledge consumers.

### 2.2.2 Methodology & Axiology: Exploring Implementation and Value

The use of crowdsourcing for creative ideas, solutions to problems, or fulfillment of tasks, arises from a particular challenge facing an individual or an organization that cannot be resolved by the "internal" team or with existing resources. To do so, Estellés-Arolas and González-Ladrón-de-Guevara (2012) have established core characteristics of crowdsourcing which can potential serve as a procedural checklist for implementation. While the issue, task, or problem at hand is of interest to both parties (the crowdsourcer and the crowd), the demand for the solution is typically generated organically. This is not necessarily dissimilar to the research process, whereby a problem is identified and researchers seek to address the issue through various types of inquiry and engagement. The online and Internet characteristics of crowdsourcing do require further contemplation. Very early examples of crowdsourcing such as The Longitude Prize (Ranard et al., 2014) and The National Audubon Society's annual bird count (National Audubon Society, 2014), existed well before the advent of the Internet, and so the open call in this instance was through other channels that offered a similar opportunity for participant self-selection based on interest. From a research methods perspective, when applying

the core components of crowdsourcing, researchers need to consider the impact of an open call, and the use of online technologies in terms of sample, representativeness, bias and self-selection.

While crowdsourcing originated within a business context, its application in research necessitates contributions to the greater good, given the very nature of research and its purpose to reveal truth, and create knowledge. It is this focus on the greater good that defines the axiology of crowdsourcing. Further, a mutually beneficial altruism in research-based crowdsourcing that motivates both the researchers/hosts and participants/community. One benefit of crowdsourced research is the potential to enable greater uptake of evidence to support knowledge translation and mobilization (Celi, Ippolito, Montgomery, Moses, & Stone, 2014). Knowledge translation and mobilization potential is likely of interest to all parties involved. However, the motivations of each individual within the crowd will likely differ and could range from monetary to participation in community.

Table 2.1 summarizes the proposed crowdsourcing research paradigm. In short, crowdsourcing facilitates the co-production and co-creation of knowledge based on the premise that everyone holds some knowledge. The participatory and collaborative nature of crowdsourcing has the potential to shift traditional roles of the researcher and participant towards a more egalitarian relationship whereby both act as knowledge producers and knowledge consumers. A mutually beneficial exchange underpins crowdsourcing whereby both the crowdsourcer (researcher) and the crowd (participant) consider it worthwhile to engage in the interaction. In a research context, this purpose could be an altruistic mutually benefit to society as a whole. Finally, in its application, crowdsourcing is open and online to enable participation by any who wish to contribute.

**Table 2.1. Summarizing the Crowdsourcing Research Paradigm**

| | Philosophy/Principles |
|---|---|
| **Ontology** | Subjective-objective reality of individuals; co-created shared reality online. |
| **Epistemology** | Co-creation of knowledge; Experiential, practical and local knowledge. |
| **Methodology** | Collaborative; use of language grounded in shared experiential context. |
| **Axiology** | Shared knowledge within a community, contributing to the greater good or a collective cause meaningful to the community. |

## 2.3 The Crowdsourcing Paradigm in Health-Related Research

One of the most compelling examples of crowdsourcing in research is related to data analysis and problem solving for ALS. Prize4Life and the Neurological Clinical Research Institute (NCRI) at Massachusetts Hospital hosted a competition that invited the crowd to develop algorithms for the prediction of disease progression among 1,822 ALS patients from standardized, anonymized phase 2/3 clinical trials (Küffner et al., 2015). More than 1,000 participants were involved in the challenge, resulting in 37 potential solutions with two teams ultimately securing first prize. The two best algorithms outperformed a method designed by the challenge organizers as well as predictions by ALS clinicians and estimates, suggesting that using both winning algorithms in future trial designs could reduce the required number of patients by at least 20% (Küffner et al., 2015). Thus, this crowdsourcing competition allowed researchers to harness the collective intelligence of a team of external researchers to potentially improve the lives of people living with ALS. This exemplifies the participative and collaborative nature of crowdsourcing. Further, this example illustrates how leveraging the experience and expertise that exists within the crowd can be used to problem solve, and co-create and co-produce knowledge for the greater good.

Another ALS-related example of crowdsourcing is patient-led research related to the effects of lithium use in ALS patients (Swan, 2012). Online communities such as PatientsLikeMe (PLM), are increasingly driving patient-directed studies and self-experimentation whereby the researcher acts as an advisor or engages in the clinical trial after preliminary results from patient-directed studies show promise (Swan, 2012). In the case of lithium use for ALS, patients self-organized and mirrored a European trail with researcher-observers. When the results appeared promising, the patients engaged researchers to conduct a formal clinical trial. This example speaks to both the experience and knowledge that exists in the crowd as well as the collaborative relationship between researchers and participants where the role of the researcher was undertaken by patients (who comprise the crowd in this instance). Further, this pre-trial and its outcome was of shared value to both researchers and patients as it provided a collective and individual benefit to each party.

Another example that demonstrates a mutual value exchange from crowdsourcing for research comes from the 2011 tsunami in Japan. Following the tsunami, people were worried about radiation levels, and it was critical for officials to monitor the spread of radiation resulting from the severely damaged Fukushima-Daiichi nuclear power plant. A team of researchers designed the Japan Nuclear Crowd Map (JNCM) to monitor and map real-time radiation data (Kamel Boulos et al., 2011). Within two weeks of the disaster, individual citizens had deployed 577 Geiger counters across the country to help the monitor and track the spread of the nuclear cloud. The map combines sensor information with crowdsourced radiation data readings, resulting in more than 27 million readings since the day of the Fukushima disaster (University of Southampton, 2013). A key incentive for people to take part in crowdsourcing projects is access to the results and outcomes. In this example, the geographically dispersed crowd was able to

distribute Geiger counters across the country in a much more efficient and timely manner than would have been possible if the researchers had to deploy the counters themselves. The crowd was also able to provide a large amount of real-time information for monitoring consistently over a prolonged period of time. Thus, the crowd identifying a need also has much to offer from a solution/research perspective such as access to large volumes of data and information, access to resources and low cost, and researchers are beginning to move from recruiting individuals as subjects to engaging them as active collaborators.

These three examples highlight the core paradigmatic features of crowdsourcing: subjective-objective realities, co-creation of knowledge, the metamorphosis of the researcher-participant relationship and the shared value and mutual benefit derived from crowdsourcing health-related research studies. In each of these instances, the crowd was able to support researchers with their knowledge, interests and experience. The crowd in each example was purposely diverse, including individuals with knowledge in math and science undertaking an intellectual challenge, patients with lived experience and a vested interest in the outcome, as well as regular citizens who were located in a specific geographic area.

## 2.4 Future Directions

On the surface, the principles of crowdsourcing resemble those of the participatory paradigm. Participatory research is "a way of creating knowledge that involves learning from investigation and applying what is learned to collective problems through social action" (Park, 1992). Moving from its purpose and application, the participatory paradigm emphasizes "the person as an embodied experiencing subject among other subjects, its assertion of the living creative cosmos we co-inhabit, and emphasis on the integration of action with knowing, is more satisfying" (Heron & Reason, 1997). According to Frisby et al. (2005), the term 'participatory'

refers to the creation of partnerships among people, including researchers, to work together, to learn about a problem, and to develop solutions based on the expertise and lived experiences of each participant. The role of the community, in cooperation with participant-researchers, enables integrated knowledge translation, and results in change, based on the research undertaken. Unlike traditional research, where knowledge is produced for understanding, and much like action research, knowledge produced in participatory research is intended for use in enacting or enabling change. These striking similarities between principle of crowdsourcing and the participatory paradigm warrants further exploration.

Another opportunity for future research is to explore crowdsourcing as a research method. There are numerous ways in which the crowd is being engaged for health-related research. From data collection (Adams, 2013; Chunara et al., 2013; Kamel Boulos et al., 2011; Neighbour, Oppenheimer, Mukhi, Friesen, & McLeod, 2010) to data analysis (King, Gehl, Grossman, & Jensen, 2013; Turner-McGrievy, Helander, Kaipainen, Perez-Macias, & Korhonen, 2015) to content creation (Coley et al., 2013), researchers are undertaking crowdsourcing for a wide range of purposes in broad array of health-related studies. As a standalone method, crowdsourcing could be applied across all research quantitative and qualitative paradigms. In such a case, researchers would need to consider paradigm specific issues related to data quality such as reliability, validity, and saturation.

## 2.5 Conclusions

The surge of open collaboration, facilitated by information technologies such as the Internet, provides unprecedented opportunities for the research community. A growing number of research projects are employing crowdsourcing as part of their methods, leveraging it to inform everything from study design, to participant recruitment, to analysis. Less often, as

evidenced by a lack of published research, researchers discuss crowdsourcing as a methodology, and address methodological questions. By looking at questions of ontology, epistemology, methodology, and axiology, this paper attempts to highlight the principles and philosophies of crowdsourcing. Crowdsourcing is rooted in subjective-objective reality that allows for the co-creation of knowledge through collaborative inquiry, ultimately undertaken for a greater good. Further, given the fundamental premises of crowdsourcing, and the manner in which crowdsourcing in being employed in health-related research, a possible home for crowdsourcing could exist within the participatory paradigm. This paper thus sets the foundation for further investigation to better understand how crowdsourcing can be used in research studies.

## 2.6 References

Adams, S. A. (2013). Maintaining the collision of accounts: crowdsourcing sites in health care as brokers in the co-production of pharmaceutical knowledge. *Information, Communication and Society*, *17*(6), 657–669. http://doi.org/10.1080/1369118X.2013.808362

Bevelander, K. E., Kaipainen, K., Swain, R., Dohle, S., Bongard, J. C., Hines, P. D. H., & Wansink, B. (2014). Crowdsourcing novel childhood predictors of adult obesity. *PLoS ONE*, *9*(2), e87756. http://doi.org/10.1371/journal.pone.0087756

Brabham, D. C. (2008). Crowdsourcing as a Model for Problem Solving: An Introduction and Cases. *Convergence: The International Journal of Research into New Media Technologies*, *14*(1), 75–90. http://doi.org/10.1177/1354856507084420

Brabham, D. C. (2010). Moving the Crowd at Threadless: Motivations for participation in a crowdsourcing application. *Information, Communication & Society*, *13*(8), 1122–1145. http://doi.org/10.1080/13691181003624090

Brown, A. W., & Allison, D. B. (2014). Using crowdsourcing to evaluate published scientific

literature: Methods and example. *PLoS ONE*, *9*(7), e100647. http://doi.org/10.1371/journal.pone.0100647

Celi, L. A., Ippolito, A., Montgomery, R. A., Moses, C., & Stone, D. J. (2014). Crowdsourcing knowledge discovery and innovations in medicine. *Journal of Medical Internet Research*. http://doi.org/10.2196/jmir.3761

Chunara, R., Smolinski, M. S., & Brownstein, J. S. (2013). Why we need crowdsourced data in infectious disease surveillance. *Current Infectious Disease Reports*, *15*(4), 316–319. http://doi.org/10.1007/s11908-013-0341-5

Coley, H. L., Sadasivam, R. S., Williams, J. H., Volkman, J. E., Schoenberger, Y. M., Kohler, C. L., … Houston, T. K. (2013). Crowdsourced peer- versus expert-written smoking-cessation messages. *American Journal of Preventive Medicine*, *45*(5), 543–550. http://doi.org/10.1016/j.amepre.2013.07.004

Estellés-Arolas, E., & González-Ladrón-de-Guevara, F. (2012). Towards an integrated crowdsourcing definition. *Journal of Information Science*, *38*(2), 189–200. http://doi.org/10.1177/0165551512437638

Frisby, W., Reid, C. J., & Millar, S. (2005). Putting "participatory" into participatory forms of action research. *Journal of Sport Management*, *19*, 367–386.

Heron, J., & Reason, P. (1997). A participatory inquiry paradigm. *Qualitative Inquiry*, *3*(3), 274–294. http://doi.org/10.1017/CBO9781107415324.004

Howe, J. (2009). *Crowdsourcing: Why the power of the crowd is driving the future of business*. New York: Random House.

Kamel Boulos, M. N., Resch, B., Crowley, D. N., Breslin, J. G., Sohn, G., Burtner, R., … Chuang, K.-Y. S. (2011). Crowdsourcing, citizen sensing and sensor web technologies for

public and environmental health surveillance and crisis management: trends, OGC standards and application examples. *International Journal of Health Geographics*, *10*(1), 67. http://doi.org/10.1186/1476-072X-10-67

King, A. J., Gehl, R. W., Grossman, D., & Jensen, J. D. (2013). Skin self-examinations and visual identification of atypical nevi: Comparing individual and crowdsourcing approaches. *Cancer Epidemiology*, *37*(6), 979–984. http://doi.org/10.1016/j.canep.2013.09.004

Küffner, R., Zach, N., Norel, R., Hawe, J., Schoenfeld, D., Wang, L., … Leitner, M. L. (2015). Crowdsourced analysis of clinical trial data to predict amyotrophic lateral sclerosis progression. *Nature Biotechnology*, *33*(1), 51–7. http://doi.org/10.1038/nbt.3051

Leiter, A., Sablinski, T., Diefenbach, M., Foster, M., Greenberg, A., Holland, J., … Galsky, M. D. (2014). Use of crowdsourcing for cancer clinical trial development. *Journal of the National Cancer Institute*, *106*(10), 8–10. http://doi.org/10.1093/jnci/dju258

Lincoln, Y. S., & Guba, E. G. (2011). Paradigmatic controversies, contradictions, and emerging confluences, revisited. In N. . Denzin & Y. S. Lincoln (Eds.), *The Sage Handbook of Qualitative Research* (2nd ed., pp. 97–128). Thousand Oaks: Sage Publications, Inc.

National Audubon Society. (2014). History of the Christmas bird count. Retrieved January 31, 2017, from http://www.audubon.org/history-christmas-bird-count

Neighbour, R., Oppenheimer, L., Mukhi, S. N., Friesen, M. R., & McLeod, R. D. (2010). Agent based modeling of "crowdinforming" as a means of load balancing at emergency departments. *Online Journal of Public Health Informatics*, *2*(3), 1–13. http://doi.org/10.5210/ojphi.v2i3.3225

Park, P. (1992). The discovery of participatory research as a new scientific paradigm: Personal and intellectual accounts. *The American Sociologist*, *23*(4), 29–42.

http://doi.org/10.1007/BF02691929

Ponterotto, J. G. (2005). Qualitative research in counseling psychology: A primer on research paradigms and philosophy of science. *Journal of Counseling Psychology*, *52*(2), 126–136. http://doi.org/10.1037/0022-0167.52.2.126

Ranard, B. L., Ha, Y. P., Meisel, Z. F., Asch, D. A., Hill, S. S., Becker, L. B., … Merchant, R. M. (2014, January). Crowdsourcing: Harnessing the masses to advance health and medicine, a systematic review. *Journal of General Internal Medicine*. http://doi.org/10.1007/s11606-013-2536-8

Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few, and how collective wisdom shapes business, economies, societies, and nations*. New York: Anchor Books.

Swan, M. (2012). Health 2050: The Realization of Personalized Medicine through Crowdsourcing, the Quantified Self, and the Participatory Biocitizen. *Journal of Personalized Medicine*, *2*(3), 93–118. http://doi.org/10.3390/jpm2030093

Turner-McGrievy, G. M., Helander, E. E., Kaipainen, K., Perez-Macias, J. M., & Korhonen, I. (2015). The use of crowdsourcing for dietary self - monitoring: Crowdsourced ratings of food pictures are comparable to ratings by trained observers. *Journal of the American Medical Informatics Association*, *22(e1)*, e112-9. http://doi.org/10.1136/amiajnl-2014-002636

University of Southampton. (2013). ECS researchers at Southampton University develop new tool to provide radiation monitoring in Japan. Retrieved January 31, 2017, from http://www.ecs.soton.ac.uk/news/4237

# Chapter 3

## Exploring the Characteristics of Crowdsourcing: An Online Observational Study*

### 3.1 Overview

As virtual research environments and e-research grow in popularity, the way in which research is undertaken is evolving to embrace these technology-enabled approaches (Fraser & Fraser, 2005; Voss & Procter, 2009). This paper focuses on one such approach, known as crowdsourcing, in which participants are recruited over the Internet to complete a task with few or no restrictions on participation. In terms of research applications, early adopters of crowdsourcing have sought participants for a variety of tasks. For example, a systematic review examining the use of crowdsourcing in health and medical research found it has been used for a variety of tasks, ranging from problem solving, data processing, surveillance/monitoring to surveying (Ranard et al., 2014). These applications appear as proof of concept to determine feasibility, and to verify the practical potential of crowdsourcing in research for everything from obtaining of feedback on health promotion materials, to the tracking of flu symptoms, to the identification of malaria infected red blood cells.

The term 'crowdsourcing' first emerged in a business context, in reference to a public, open call to outsource an activity or work typically undertaken by employees of an organization (Howe, 2006). The concept of an open call to engage the public in activity was not, however, new – its origins have been traced to various events and authors, including the establishment of the Longitude Prize in 1714 by Britain's Parliament, which offered a monetary reward to anyone who could solve the problem of identifying a ship's longitudinal position within 30 miles

---

* A version of this chapter has been submitted for publication as follows: Bassi, H., Lee, C.J., Misener, L., & Johnson, A.M. (under review). Exploring the characteristics of crowdsourcing: An online observational study. *Submitted to the Journal of Information Sciences*.

(Ranard et al., 2014) and the works of Sir Francis Galton (Surowiecki, 2004). Further, others liken it to the open source movement in the information technology sector (Zhao & Zhu, 2014). The varied roots of crowdsourcing are accompanied by a broad spectrum of definitions of the term. Estellés-Arolas and González-Ladrón-de-Guevara (2012) identified 40 different definitions for the term crowdsourcing and articulated the following definition:

> *Crowdsourcing is a type of participative online activity in which an individual, an institution, a non-profit organization, or company proposes to a group of individuals of varying knowledge, heterogeneity, and number, via a flexible open call, the voluntary undertaking of a task. The undertaking of the task, of variable complexity and modularity, and in which the crowd should participate bringing their work, money, knowledge and/or experience, always entails mutual benefit. The user will receive the satisfaction of a given type of need, be it economic, social recognition, self-esteem, or the development of individual skills, while the crowdsourcer will obtain and utilize to their advantage what the user has brought to the venture, whose form will depend on the type of activity undertaken. (p.197)*

This definition allows for the broadest possible application of the term 'crowdsourcing.' It covers features of the process, participant characteristics, the nature of the task or problem to be addressed, as well as the nature of the interaction. Estellés-Arolas and González-Ladrón-de-Guevara (2012) further dissects the definition of crowdsourcing into eight discrete characteristics:

(a) there is a clearly defined crowd;

(b) there exists a task with a clear goal;

(c) the recompense received by the crowd is clear;

(d) the crowdsourcer is clearly identified;

(e) the compensation to be received by the crowdsourcer is clearly defined;

(f) it is an online assigned process of participative type;

(g) it uses an open call of variable extent; and

(h) it uses the Internet (p. 197)

These characteristics begin to put the qualities of crowdsourcing into methodological terms. Estellés-Arolas and González-Ladrón-de-Guevara (2012) verified these characteristics of crowdsourcing by demonstrating that sites such as InnoCentive and Amazon Mechanical Turk meet all of these criteria, while others such as Wikipedia do not. While this verification holds true with the specific platform (i.e., the websites), the question remains whether and how these characteristics of crowdsourcing can be applied to individual research projects. These characteristics serve as the framework for the content analysis within this paper, to examine how research projects address the crowdsourcing criteria proposed by Estellés-Arolas and González-Ladrón-de-Guevara (2012). This framework was chosen primarily for two reasons: first, the comprehensive and integrative process by which the authors compiled definitions and characteristics from the literature and; second, it is the only framework within the literature that has attempted to establish criteria for crowdsourcing.

The aim of the present study is to provide an understanding of how crowdsourcing is being used in research by undertaking a content analysis of studies posted in an online site purposed to crowdsourcing research. Although there are a number of such sites, the present study provides an analysis only of the Crowdcrafting website (http://crowdcrafting.org). The specific issues to be addressed are:

1. How researchers are using crowdsourcing in practice;

2. How studies address the basic characteristics of crowdsourcing as defined in the literature; and,
3. The extent to which particular characteristics of crowdsourcing relate to the completion of a project.

This paper concludes with recommendations for researchers undertaking the design and implementation of projects employing crowdsourcing.

## 3.2 Methods

This observational study analyzes publically available online data on the Crowdcrafting website (http://crowdcrafting.org). An observational paradigm was deemed appropriate as it meant that the researchers would not actively engage with the hosts or users of the site in compiling data to answer the question, how are researchers using crowdsourcing? Observation allowed the researchers to examine the application of crowdsourcing in research in practice and directly see how researchers engaged the crowd in addition to reading textual accounts from published articles. A content analysis was then undertaken to map projects against characteristics of crowdsourcing noted earlier.

Crowdcrafting partnered with CERN (the European Organization for Nuclear Research), United Nations (UNITAR), the University of Geneva, and Open Knowledge International in developing their platform. This service was chosen because of its self-proclaimed focus on scientific research:

> *Crowdcrafting is a web-based service that invites volunteers to contribute to scientific projects developed by citizens, professionals or institutions that need help to solve problems, analyze data or complete challenging tasks that can't be done by machines alone, but require human intelligence. The platform is 100%*

> *open source – that is its software is developed and distributed freely – and 100% open-science, making scientific research accessible to everyone.*
>
> (https://crowdcrafting.org/about)

In addition, the fact that the platform is free and open source, aligns with the basic principles of crowdsourcing. The information reviewed (including all project content) was all available outside of the contributor login area.

A total of 427 projects, categorized by self-identified subject area by project host (researcher), were reviewed and logged between November 20 and November 27, 2016. Table 1 presents the breakdown of the projects, by subject, at various phases of review. Of the 427 projects reviewed, the following were excluded:

- 44 in languages other than English
- 101 test or demonstration projects denoted as such on the project page (i.e. non research)
- 10 projects that required a login to participate
- 18 project pages with technical errors (e.g., "page not available" or "page would not load"); and
- 23 duplicate projects.

Of the remaining 231 projects, 53 were completed, meaning all the tasks set out for participants were finished. An additional 4 completed projects were removed at this stage of observation, as the project pages were still posted but data was no longer accessible. The remaining 227 projects broke down into the subjects identified in column three of Table 3.1. A total of 21.5% (49) of the projects reviewed were completed meaning that the tasks assigned to participants had been finished.

**Table 3.1. Summary of Projects Reviewed**

| Self-identified Subject Area | Number of Posted Projects | Number of Projects Review for Study | Completed Projects Reviewed |
|---|---|---|---|
| Social | 168 | 69 | 17 |
| Art | 40 | 37 | 1 |
| Humanities | 20 | 9 | 4 |
| Biology | 6 | 4 | 1 |
| Economics | 13 | 2 | 0 |
| Science | 180 | 106 | 26 |
| | **427** | **227** | **49** |

Each project posted on this site was assessed against the characteristics of crowdsourcing. Characteristics were coded as "met" (if it was addressed within the reviewed documentation), or "not met" (if the criteria were not addressed within the documentation). All coded data is presented in Appendix A, with "met" coded as "1", and "not met" coded as "0". A second reviewer verified the coding by randomly reviewing a selection of the coded projects in each subject category, and all discrepancies were discussed until consensus was reached.

### 3.2.1 Assumptions

The authors do not necessarily perceive the Estellés-Arolas and González-Ladrón-de-Guevara (2012) framework to be the gold standard for crowdsourcing research – rather, the relative novelty of crowdsourcing, and resulting limited evidence in the area, has led to a dearth of frameworks from which to choose. It further needs to be noted that Estellés-Arolas and González-Ladrón-de-Guevara (2012) acknowledge the disproportionate influence of computer science based evidence informing their work along with the emphasis on specific types of crowdsourcing activities with no mention of crowdsourcing for research purposes per se. Therefore, the interface design and functionality of the Crowdcrafting website causes it to meet

all of the characteristics of crowdsourcing as defined by Estellés-Arolas and González-Ladrón-de-Guevara (2012). Given the capabilities of the site, and its stated purpose (as articulated on the About Us page) of promoting research and engaging citizen scientists, 100% of the listed projects also address all the characteristics of crowdsourcing suggested by Estellés-Arolas and González-Ladrón-de-Guevara (2012). We contend, however, that using a website that meets all necessary characteristics of crowdsourcing, is insufficient to meet reasonable ethical and methodological standards of rigor for research, and additional consideration must be given to the application of the characteristics of crowdsourcing to the research study itself.

In addition, it must be recognized that while other crowdsourcing websites are being used for research, none of the other sites are as explicit in their goal of supporting crowdsourced research. Sites such as Amazon Mechanical Turk and Crowdflower have been identified in the literature as platforms for crowdsourced research studies, but these sites were not created with the primary intention of supporting crowdsourced research.

Finally, the authors of this study are also viewing crowdsourced research through a lens that has been formed through extensive experience with more mature and established research paradigms, methodologies and methods – and therefore, an inherent bias exists in reviewing these projects. It was challenging to review the research projects without considering ethical implications, notions of research quality, and methodological rigor.

## 3.3 Findings

### 3.3.1 How are Researchers Using Crowdsourcing?

Of the projects reviewed, 203 clearly identified the type of task assigned to participants. A wide array of different tasks was seen within the projects reviewed, including:

- Image identification or classification-related tasks include tagging photos, identifying objects, validating images, etc. These tasks ranged from identification of insects to naming the sport in the picture to more complex tasks such as developing descriptions of images to classifying images for melanoma. Approximately 43% of the project related tasks fell into this category.
- Information- and knowledge-gathering related tasks included administering surveys and knowledge tests and gathering opinions on a range of topics such as 3D printers to what an MBA means. In addition, data was also gathered through gaming such as connecting dots to understand pattern recognition or playing a "Graph Isomorphism Game" to problem-solve. Approximately 26% of the project related tasks fell into this category.
- Text transcription, translation, and analysis related tasks ranged from transcribing hand written notes such as Winston Churchill's diaries to translating text from English to Spanish, Italian and other languages to examining sentence analysis to classifying tweets related to specific topics such as a natural disaster. Approximately 20% of the project related tasks fell into this category.
- Sound-related tasks included analyzing sounds and sound pattern recognition. Some examples of such tasks included classifying sounds clips for certain types of music and identifying urban sounds from NYC. Approximately 5% of the project related tasks fell into this category.

- Geolocation-based data collection and analysis related tasks including identifying cities based on information on a map to geotagging rural villages to geo-mapping locations of parks to transcribing locations of oil spills to tracking the floods in France. Approximately 4% of the project related tasks fell into this category.
- Counting-related tasks focused primarily on counting objects and assigning values. Examples include counting the number of building in an image and counting the number of particles in an image. Approximately 2% of the project related tasks fell into this category.

The range of tasks assigned across the projects appeared to require varying degrees of skills and a host of different participant capabilities.

### 3.3.2 How do Studies Address the Characteristics of Crowdsourcing?

A content analysis was undertaken to test the characteristics identified by Estellés-Arolas and González-Ladrón-de-Guevara (2012) for the purposes of using crowdsourcing for research.

**(a) there is a clearly defined crowd**

This characteristic is well-supported by the Crowdcrafting website itself. When considering individual projects on the site, however, only two projects (0.88%) specifically articulated what segment of website visitors might wish to participate in the research study by providing additional information about interests or skills that would be best suited to support the project. One of the two projects that segmented the crowd based on skills, interest or experience included the following narrative:

*Are you intrigued by Winston Churchill as a public figure? Interested in World War II history? Fascinated by the day-to-day of political leadership? Good at deciphering hard-to-read handwriting?*

Another project required individuals to translate text from English to Spanish, but did not explicitly articulate that the participants should have some level of linguistics capability, instead relying on the task to imply this as a requirement of participants.

**(b) there exists a task with a clear goal**

Most of the projects (81.5%) identified a goal that was associated with the task, such as: identify the image, tag image, translate text, transcribe narratives, count objects, classify tweets, analyze maps and analyze sounds. Those that did not meet this criteria did not explain to the crowd what task needed to be undertaken to complete the assignment. Only 23.7% of projects articulated an overarching goal to which the task was contributing. Some examples include:

- *Help us test TagIT, a crowdsourcing system to create image tags which will be used to create image descriptions to improve access to online teaching materials for blind and partially sighted students.*
- *Transparency has won big victories in its 15 years or so of life as a movement, with contract transparency in particular rapidly gaining momentum. We, at OpenOil.net, are firm believers that governments publishing their oil contracts is a clear step towards better governance of the oil industry (check out repository.openoil.net for most of the world's published oil contracts).*
- *Cookbooks as any other written texts can be and are used by historians as a primary source material. Since cooking books are written by and for people,*

*this project aims to investigate the image of society that emerges through cookbooks. The main focus is on the way cooking is portrayed and the role of women in the society.*

**(c) the recompense received by the crowd is clear**

As noted on the website About Us page (http://crowdcrafting.org/about) there is no recompense, or micro-payment for contributing to the research. Nonetheless, of the projects reviewed, none explicitly reiterated this lack of compensation. It appears that for most projects, participants can contribute to the research without creating an account. For some projects, however, participants are required to login, and are thus able to track their contributions (i.e., how many tasks they have completed) in comparison to others. The login requirements to contribute are minimal, requiring only a user name and password.

Approximately 10.57% of projects described the crowds' contribution in a non-monetary fashion. Some examples of such descriptions include:

- *Thanks to you, we will be able to detect meaningful relations in raw text documents. Your contribution is really important and has a huge impact… Go ahead and be part of a multilingual world!*
- *Even though this is simple information, it will go a long way to adding this missing information to the OpenStreetMap and so (in our case) help to generate routing instructions that can be tailored to those people with reduced mobility.*
- *Help support job employment by posting about job listings.*
- *It is therefore crucial for us to measure the distribution of binding angles of a particular molecule on a given surface. This will allow us to compare our*

> *results with theoretical predictions to better understand their properties. But getting accurate data on this means measuring thousands of images, which is tremendously time-consuming for our small team. That's where you come in … Based on the average of estimates by several volunteers like you, we hope to extract new information about the subtle ways that molecules interact with the surfaces they are stuck to, and how this affects their magnetic properties. Our data will be made openly available after we have analyzed it, and we will gladly acknowledge the volunteers in any publications that result from your efforts. We also hope you will enjoy this chance to explore a beautiful phenomenon from the nanoworld!*

**(d) the crowdsourcer is clearly identified**

Of the projects reviewed, only 29.96% (75) clearly identified the crowdsourcer. Any information found in the project pages that identified the organization or individual behind the project was used to determine whether this criteria was fulfilled. Of these projects, the crowdsourcer was most frequently identified as an organization and/or an individual, typically a public institution (university, research organization, etc.) or not-for-profit organization or a software/app development company. Only 4% (10) of projects identified an individual, an organization, and contact information. In a few instances, only an email address or twitter handle was provided. Further, in most instances, this information was not easily accessible and it was difficult to determine who was undertaking the project.

**(e) the compensation to be received by the crowdsourcer is clearly defined**

None of the projects reviewed outlined the nature of compensation – monetary or otherwise – to be received by the project hosts/researchers. When looking at compensation as the benefit to the researcher, some project descriptions (5.73%) were more explicit in the benefit to the project host or society:

- *AEgIS scientists need to fine-tune their understanding of annihilation by mapping the particle tracks and counting the number of thin and fat tracks for many particle bursts…Humans are way better at interpreting the particle tracks than machines so the AEgIS team needs your big brains and keen eyes to map the particles' path through the emulsion. All you have to do is join the dots! AEgIS scientists also want to be able to classify each track as fat or thin. Please get in touch if you would like to help to write the software to carry out this classification.*
- *The game you are playing solves instances of the Graph Isomorphism Problem (for short, GIP)... Here we ask: how do human beings perform when solving GIP? Do human beings find GIP easier on certain graphs? Can we define a "human parameter of cognitive computational complexity"? And how does this relate to known mathematical parameters to quantify computational complexity? We collect data from our game, with the purpose of shedding some light on these questions and hopefully to learn something new about computational complexity in general.*

The non-monetary benefits to the researcher in these instances is related to the research study which in many cases contributes to the greater good of society.

**(f) it is an online assigned process of participative type; (g) it uses an open call of variable extent; and (h) it uses the internet**

100% of projects would meet these three criteria as a result of the very fact that the site is openly accessible on the Internet.

## 3.4 Summary

Based on the premise that the website itself addresses all the characteristics of crowdsourcing, all of the projects (100%) reviewed would be considered crowdsourced research projects. However, assuming that the design and functionality of the site addressed all the characteristics with the exception of who the crowdsourcer is, only 27.75% (63) of projects could be considered crowdsourcing. Furthermore, only 1.76% (4) of the projects addressed all characteristics, and clearly articulated:

a. what segment of the population of citizen scientists is appropriate for each project;
b. what is the larger goal to which the assigned task is contributing; and,
c. who (specifically – name, affiliation and contact information) is conducting the research.

Pearson's r was used to determine which if any of the characteristics contributed to the completion of a project (where completion refers to the tasks assigned to the crowd), and these correlations are presented in Table 3.2. Only one statistically significant positive correlation was identified. There was a positive correlation ($r = 0.48$) between the variable associated with clear delineation of the recompense received by the crowd.

**Table 3.2. Summary of Projects at http://crowdcrafting.org that Satisfy Crowdsourcing Characteristics Described by Estellés-Arolas and González-Ladrón-de-Guevara (2012)**

| Characteristic | Projects that Fulfill Characteristics (%) | Project Completion (r) |
|---|---|---|
| Clearly defined crowd | 0.88 | 0.07 |
| Task with a clear goal | 81.50 | -0.08 |
| Clear depiction of compensation to crowd | 10.57 | 0.48 |
| Identification of crowdsourcer | 29.96 | 0.28 |
| Clear depiction of compensation to crowdsourcer | 5.73 | 0.15 |
| Online | 100 | -0.13 |
| Open call | 100 | -0.13 |
| Internet | 100 | -0.13 |

## 3.5 Discussion

### 3.5.1 Crowdsourcing in Research: Methodology or Method?

There is strong philosophical and methodological alignment between crowdsourcing and the qualitative paradigm of participatory action research (Bassi, Misener, & Johnson, under review). While the relationship between participatory action research and crowdsourcing is far more complex than what can be explained here, the participative nature of crowdsourcing, the recognition of different degrees of knowledge, alongside the mutual benefit exchange presented in the Estellés-Arolas and González-Ladrón-de-Guevara (2012) definition provide some immediate and tangible anchors for this discussion:

> *Crowdsourcing is a type of **participative** online **activity** in which an individual, an institution, a non-profit organization, or company proposes to a group of individuals of **varying knowledge**, heterogeneity, and number, via a flexible open call, the voluntary undertaking of a task. The undertaking of the task, of variable complexity and modularity, and in which the crowd should participate bringing*

*their work, money, knowledge and/or experience, always* ***entails mutual benefit****…*

*(p.197)*

Most research projects reviewed in the literature and during this study apply crowdsourcing as a part of their methods and not as an entire methodological approach.

### 3.5.2 How is Crowdsourcing Being Used in Research?

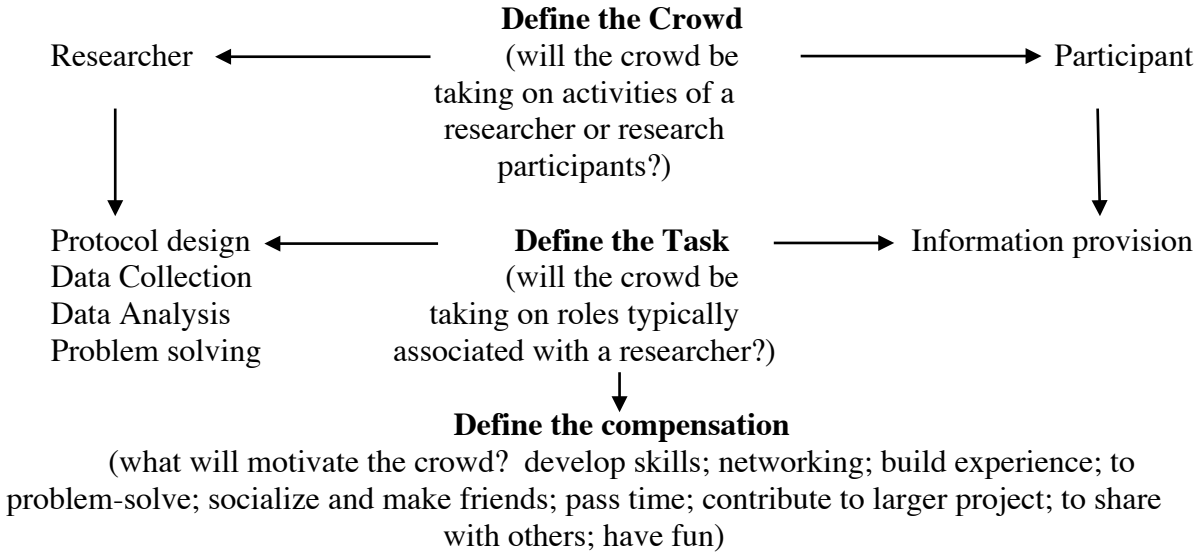
While there is no accepted framework that is appropriate for all types of crowdsourcing, there are a few categories that have been described in the literature, including micro, repetitive, task-based activities (such as data collection or analysis), creative tasks (generating new ideas), and problem-solving (solutions based) tasks (Brabham, 2013; Howe, 2009; Parvanta, Roth, & Keller, 2013; Zhao & Zhu, 2014). Further, a systematic review examining the use of crowdsourcing in health and medical research found it has been used for a variety of tasks, ranging from problem solving, data processing, surveillance/monitoring and surveying (Ranard et al., 2014). The results of the present research found that most of the activities undertaken were micro, repetitive, task-based activities related to data processing or analysis such as image and sound identification, and text translation. To a lesser degree, information gathering tasks were administered. It would appear that crowdsourcing served as a tool for participant recruitment, data collection and analysis. Largely absent from the projects reviewed were creative or complex problem-solving activities. To a certain extent, the nature of the activities assigned to participants could be attributed to the design and functionality of the site, and this is discussed in further detail below.

Another way to interpret these results is to recast the way in which the crowd is being used, against the typical role of the researcher. In crowdsourcing, the line between researcher and the crowd (the "citizen scientists"), begins to blur the lines of traditional researcher-participant

roles and responsibility. This aligns with Howe's reference for crowdsourcing as a public, open call to outsource an activity or work typically undertaken by employees of an organization. Where the crowd is being used for purposes such as data collection, analysis and problem-solving, participants are acting more as researchers (or co-researchers) rather than research participants. Alternatively, where crowdsourcing is used for participant recruitment, or information and knowledge extraction, participants act as research subjects or participants. This then allows researchers to reconsider how the characteristic of crowdsourcing apply in their context.

The definition of crowd and task should be considered together in the research context. This process is outlined in Figure 3.1.

**Figure 3.1. Defining the Crowd and the Task in Research-oriented Crowdsourcing**



As illustrated in Figure 3.1, consideration should be given to the purpose for which the crowd is being engaged, be it recruitment, data collection or data analysis. While crowdsourcing requires an open call, clearly defining the skills, interests and experience of participants best suited for the study (where appropriate) may improve the researchers' ability to engage the crowd, thereby improving crowd responsiveness (and increasing completion rate). Specification and segmentation of the crowd may not be appropriate where crowdsourcing is being used for broader participant recruitment. Research project and method-specific adaptations should be considered in the application of this criteria. Despite the self-selection and voluntary nature of crowdsourcing, further research is needed to identify how research ethics standards and guidelines apply in this context.

Reflecting upon the crowd as either researchers or research participants, and the associated tasks being assigned, may provide insights into motivations for participation. Brabham (2010) identified the following motivations for individuals in crowds: the desire to earn money; to develop one's creative skills; to network with other creative professionals; to build a portfolio for future employment; to challenge oneself to solve a tough problem; to socialize and make friends; to pass the time when bored; to contribute to a large project of common interest; to share with others; and to have fun. Estellés-Arolas and González-Ladrón-de-Guevara (2012) articulate the motives of crowdsourcing participants as "the satisfaction of a given type of need, be it economic, social recognition, self-esteem, or the development of individual skills" (p.197). Understanding why the crowd would be willing to participate can enhance the success of a crowdsourced research study.

Based on the results of the present research, the clear delineation of recompense received by the crowd appears to be positively correlated with the completion of the project. The About

Us page of the site articulates the audience they are targeting (citizen scientists) and the purpose (make research and science accessible to everyone). Further, the description of site sets the stage for volunteers to help scientists without any recompense for either party – and so it may not be necessary to compensate the crowd. The recompense must, however, be clearly articulated, which means that the crowd must understand how task completion contributes to the overall research study. While individual motivations within the crowd may vary, from altruistic to fulfilling human needs such as belonging and recognition, researchers can harness these motivations to engage the crowd effectively.

Even if the compensation received is nonmonetary, explaining how participants' contributions impacts research or a greater good or describing who is participating to create a sense of community, is likely to improve the responsiveness of the crowd.

In most of the individual projects reviewed, the crowdsourcer was not clearly identified, nor was the compensation to be received by the crowdsourcer. By the very nature of the Internet, content posted on sites is not always associated with the owner of the platform or the host/owner of the page. It only makes sense, therefore, for researchers to explicitly identify themselves when engaging the crowd. Where the study design or methodology of the research precludes this explicit identification, then the same assumptions hold true for crowdsourcing. As with all research, the researchers should be forthcoming about any personal gains and benefits they will receive as a result of the crowdsourcing and acknowledge if they stand to make any financial gains as a result of the efforts of the crowd.

Three of the characteristics defined by Estellés-Arolas and González-Ladrón-de-Guevara (2012) are primarily a function of the information technology that is used in delivering content (and retrieving participant responses), namely the notion that crowdsourcing is an online

participative process, that uses an open call, over the Internet. By the very existence of a research project on the Internet these characteristics are met. The open call of a variable is addressed by the open access to the site and projects. However, project hosts can require participants to login to contribute to their project should they chose.

### 3.5.3 The Characteristics of Crowdsourcing and Technology

When analyzing online content, it is important to consider the interface/design of site, project/host content, and the user-generated content (Neuendorf, 2002), and indeed, in the present study, the crowdsourcing characteristics described by Estellés-Arolas and González-Ladrón-de-Guevara (2012) are primarily addressed by the design and expressed purpose of the site. Therefore, the design of the site for crowdsourcing is essential to ensuring that the characteristics of crowdsourcing are met.

It could be argued that most crowdsourcing characteristics are inextricably bound up in the platform that hosts the project, rather than within the individual research projects themselves. The design and functionality of a site like Crowdcrafting facilitates small, independent tasks to be assigned to the crowd. The About Us page of the site articulates the audience they are targeting (citizen scientists) and the purpose of the site (make research and science accessible to everyone). Further, the description of site sets the stage for volunteers to help scientists without any recompense for either party and also provides background on the features of the online platform including the associated software company. For all intents and purposes, the website meets and addresses all eight characteristics of crowdsourcing. But does this mean that researchers need only to ensure that the website hosting their projects meet the characteristics of crowdsourcing? Does this free them from the obligation of ensuring that their projects meet acceptable standards for crowdsourced research?

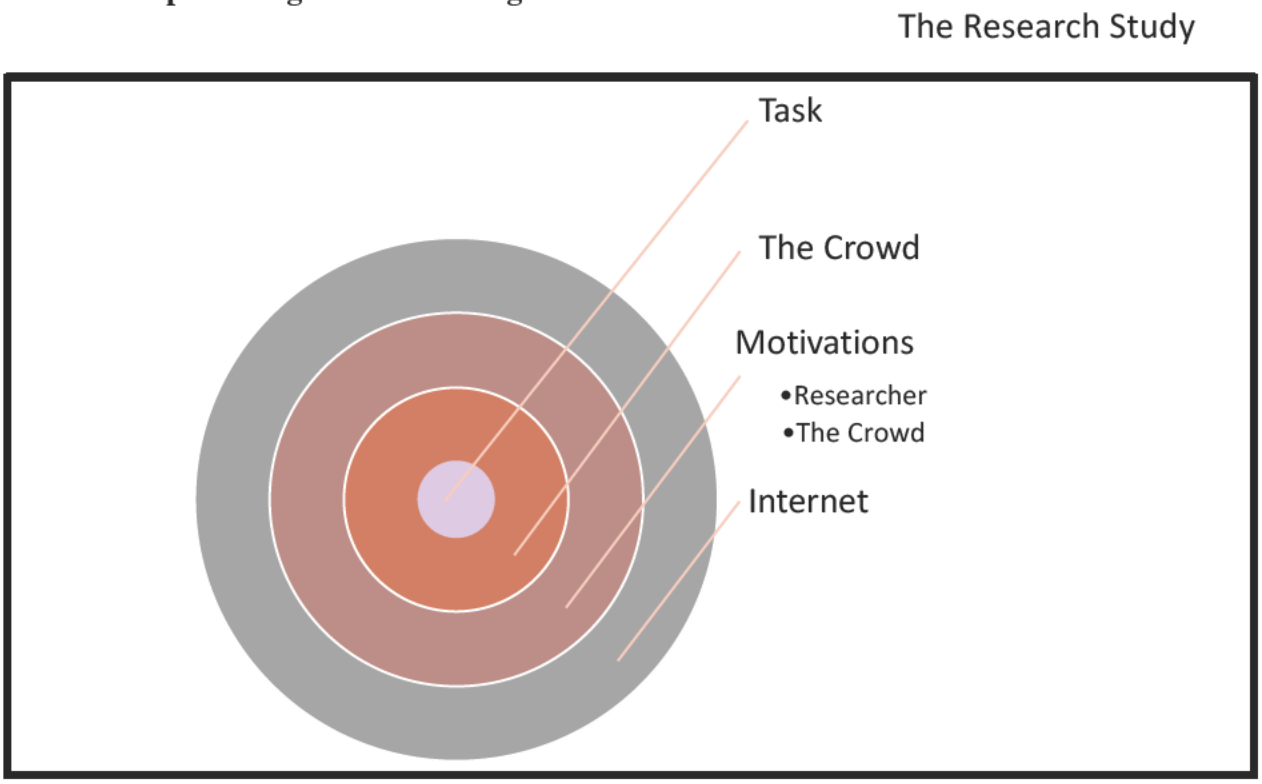
The challenge for researchers is not to merely ensure that the characteristics exist in the interface/platform, but rather to incorporate these criteria into their research study design, and to utilize them to adhere to acceptable standards for research, and (more practically) to increase participant adherence, and (ultimately) project completion rates. Relying on the generic description on the "About Us" page of a crowdsourcing website being used to conduct the research is inadequate as it does not clearly separate the project from the platform. Based on the observations made for this study, the participant is frequently left to determine who is leading the research, the project goals and researcher recompense – and this may explain the low overall (21.5%) project completion rate. Furthermore, while the interface design and functionality can clearly create small, discrete tasks for participants to undertake, an explanation of what is expected should be clearly articulated by the researchers. Where possible, an explicit description of the overarching goal of the research, not simply the task, provides important context for participants and could inform their decision to contribute. This also presents researchers with the opportunity to motivate the crowd. Further, this speaks to the overall "transparency" dimension of the research and possibly contributes to the willingness of participants to complete research. In addition, the provision of an overall goal or aim of the research allows participants to understand how the small discrete tasks they are undertaking contributes to the larger projects. The use of a platform that facilitate crowdsourcing should not negate the roles and responsibilities of the researcher in designing and implementing the protocol.

### 3.5.4 Conceptualizing the Characteristics of Crowdsourcing for Research

Figure 3.2 shows how the various components of crowdsourcing come together. The research study itself (represented as the box in which the concentric circles are contained) provides the context within which the constituent components of crowdsourcing interact with

each other. This context determines how the characteristics are applied, including how the crowd could potentially contribute – and understanding the purpose of why the crowd is being engaged in the research is central to engaging them in research. The outer ring of the circle diagram represents the use of the Internet in the presentation of tasks to participants. As the entire engagement occurs virtually via the Internet and online, it becomes the *de facto* space for all the other characteristics to converge and convene. In the absence of this space to facilitate the engagement and participation, modern day crowdsourcing would not occur. Within this rests the motivations of the researcher or oneself in wanting to engage the crowd, and determining how that fits methodologically. At the same time effort must be made to understand the motivations of the crowd – why they may want to participate – and understanding the motivations of the researcher – what benefits they receive from this engagement. This helps inform the characteristics of the crowd, specifically what particular skills or expertise they may bring that benefits the research. Once the crowd as been defined, the researcher then must assign the individual tasks to each participant.

**Figure 3.2. Conceptualizing Crowdsourcing for Research Studies**




### 3.5.5 Crowdsourcing and Research Ethics

As observed, none of the projects examined demonstrated traditional ethical standards for participation. This warrants further consideration as these studies are excluded from ethical guidelines of research. Ideally, research studies using crowdsourcing would include the same level of accountability and transparency as research studies that use more mature and well documented methodologies and methods. To maintain legitimacy and credibility of research studies that leverage crowdsourcing, researchers should endeavour to provide robust descriptions of the benefits to the crowd, who is conducting the study and, what compensation and benefit the researchers are receiving if any.

## 3.6 Concluding Thoughts

Ultimately, researchers should use the characteristics of crowdsourcing described in the present study in two ways. First, and most obviously, as criteria for ensuring the online platform of their choosing for their research actually enables crowdsourcing. Second, as an anchor for how they will deploy crowdsourcing in relation to their respective research project.

As novel approaches to research emerge, researchers are presented with exciting new opportunities to expand the boundaries of paradigms, methodologies and methods. The characteristics of crowdsourcing provide a useful framework to guide researchers undertaking crowdsourcing within their studies. While all the characteristics can be adequately addressed by the crowdsourcing interface/platform, it is important to translate and interpret these criteria in the context of research. From a website perspective, consideration should be given to the overall quality of the content posted by researchers to ensure a level of quality that offers credibility and legitimacy for both the crowdsourcing site and the research project.

Clearly defining and openly articulating the research purpose, roles of the crowd (researchers versus participants), alongside full disclosure of the researchers involved, will help ensure the integrity of the research. The crowd acting as co-researchers by taking on roles such as analysis and content creation is an important shift in the way research is evolving. The findings from this paper provide an opportunity for additional research. While the site specifically targets the crowd for research purposes, the literature reveals that sites such as Amazon Mechanical Turk and Crowdflower are also being used for crowdsourcing by researchers. Future research could involve reviewing projects from various crowdsourcing websites to further examine the role of the hosting site on research projects which employ crowdsourcing.

Finally, researchers have ethical and methodological obligations when interacting with the public that need further consideration in the context of crowdsourcing and its application within research studies. The relative newness of crowdsourcing and the challenges that come with its application do not excuse researchers of their professional requirements to respect public participants, despite the shifting definition of the term in crowdsourced research studies, and ensure the integrity of their research. While the application of crowdsourcing in research continues to be tested, and the body of literature develops, researchers have an exciting opportunity to rethink, redesign and reinvent how research is conducted.

## 3.7 References

Bassi, H., Misener, L., & Johnson, A. M. (under review). Crowdsourcing: A Potential Research Paradigm. *Qualitative Research*.

Brabham, D. C. (2010). Moving the Crowd at Threadless: Motivations for participation in a crowdsourcing application. *Information, Communication & Society*, *13*(8), 1122–1145. http://doi.org/10.1080/13691181003624090

Brabham, D. C. (2013). Crowdsourcing: A model for leveraging online communities. In A. Delwiche & J. J. Henderson (Eds.), *The Participatory Cultures Handbook* (pp. 120–129). New York, NY: Routledge.

Estellés-Arolas, E., & González-Ladrón-de-Guevara, F. (2012). Towards an integrated crowdsourcing definition. *Journal of Information Science*, *38*(2), 189–200. http://doi.org/10.1177/0165551512437638

Fraser, M., & Fraser, M. (2005). Virtual research environments: overview and activity. *Ariadne*, *3*(44), 1–9. Retrieved from http://eprints.ouls.ox.ac.uk/archive/00001072/

Howe, J. (2006). The rise of crowdsourcing. *Wired Magazine*, (14), 1–5.

Howe, J. (2009). *Crowdsourcing: Why the power of the crowd is driving the future of business*. New York: Random House.

Neuendorf, K. A. (2002). *The content analysis guidebook*.

Parvanta, C., Roth, Y., & Keller, H. (2013). Crowdsourcing 101: A Few Basics to Make You the Leader of the Pack. *Health Promotion Practice*, *14*(2), 163–167. http://doi.org/10.1177/1524839912470654

Ranard, B. L., Ha, Y. P., Meisel, Z. F., Asch, D. A., Hill, S. S., Becker, L. B., … Merchant, R. M. (2014, January). Crowdsourcing: Harnessing the masses to advance health and medicine, a systematic review. *Journal of General Internal Medicine*. http://doi.org/10.1007/s11606-013-2536-8

Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few, and how collective wisdom shapes business, economies, societies, and nations*. New York: Anchor Books.

Voss, A., & Procter, R. (2009). Virtual research environments in scholarly work and communications. *Library Hi Tech*, *27*(2), 174–190. http://doi.org/10.1108/07378830910968146

Zhao, Y., & Zhu, Q. (2014). Evaluation on crowdsourcing research: Current status and future direction. *Information Systems Frontiers*, *16*(3), 417–434. http://doi.org/10.1007/s10796-012-9350-4

# Chapter 4

## Crowdsourcing for Research: Perspectives from a Delphi Panel

### 4.1 Overview

Public and patient engagement, alongside activities such as knowledge translation and mobilization, are becoming standard requirements of health sciences and services research funding (Domecq et al., 2014; Frank et al., 2015; Tetroe et al., 2008). While some existing methodologies, such as participatory research, embrace non-researcher involvement in research, new methods are also emerging to encourage public involvement in research. Crowdsourcing, "an online, distributed problem solving and production model" (Brabham, 2010, p. 5), is one method that researchers are using to engage the public. Crowdsourcing is a nascent method, but it appears to follow in the traditions of other more established qualitative techniques, and shares many characteristics with participatory action research (see Chapter 2). The central shared characteristic between crowdsourcing and the participatory paradigm is the premise of subjective-objective reality. This informs the co-creation of knowledge through collaborative inquiry, ultimately undertaken for a greater good in the research context. The term citizen science is frequently used synonymously with crowdsourcing, and aims to address the same notion of engaging the public in research. Citizen science is defined as "a form of research collaboration involving members of the public in scientific research projects to address real-world problems" (Wiggins & Crowston, 2011, p. 1). Researchers are increasingly using websites such as Amazon Mechanical Turk and Crowdcrafting to engage the crowd for the purposes of recruitment, data collection or data analysis for their studies (Bassi, Misener, Lee, & Johnson, under review). This study sought to explore crowdsourcing as a research methodology by understanding how it is being used, why and for what purpose, and focused on addressing the

following questions: How and why are researchers currently using crowdsourcing? In doing so, we sought to develop a conceptual framework for crowdsourcing research studies.

## 4.2 Methods

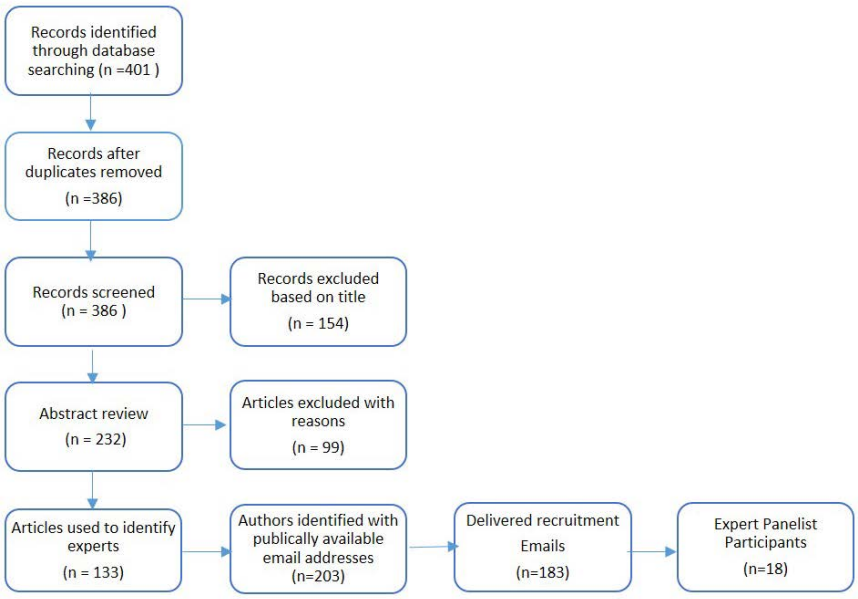
The Delphi technique, developed in the 1950s by the RAND Corporation, is a method used to achieve consensus among experts (Okoli & Pawlowski, 2004). It has been recognized that Delphi technique can also be used to determine the extent to which experts agree or disagree, and to understand the array of positions on a topic (Mullen, 2003). According to Linstone and Turoff (1975), the Delphi process facilitates group communication to enable collective problem solving. The present study employed a modified Delphi technique to determine whether there was a consensus among experts regarding the use of crowdsourcing for the purposes of research. The Delphi technique is frequently used where little evidence exists, and where the knowledge base is limited. Both of these criteria apply, within the general domain of "crowdsourcing in research." The exploratory nature of this study makes the Delphi technique appropriate as it allows for insights and knowledge to be gained, which may scaffold the induction of a general model or theory (Steinert, 2009). In addition, a panel study (as opposed to the responses of any individual expert) may provide the most relevant "answers" to our research questions, given the limited numbers of experts in this area.

### 4.2.1 Identifying the Expert Panel

According to Rowe and Wright (2001), the composition of the panel of experts should be heterogeneous to ensure that their combined experience and knowledge is representative of the full research domain. The long-standing debate of who qualifies as an expert for the purpose of a Delphi has resulted in very broad inclusion criteria such as informed individuals to more narrowly defined specialists in a field (Baker, Lovell, & Harris, 2006). The nascent nature of

crowdsourcing in research required the term "expert" to be interpreted broadly as those with experience in the application of crowdsourcing for research as well as those with knowledge on crowdsource based on study of the topic itself. Given that the purpose of this study was to identify salient characteristics of crowdsourcing within research settings, we conducted a literature review to create a list of potential participants on an expert panel of researchers and/or academics who either use crowdsourcing in their research methods, or research the topic of crowdsourcing. Figure 4.1 presents a graphical depiction of how panel members were selected.

**Figure 4.1: Participant Identification and Recruitment Process**




In October 2016, a list of published studies was assembled by using the keyword terms "crowdsourc*" and ("medical" or "health") with the filters "English" and "peer-reviewed." This search resulted in 275 articles identified in PubMed, and 126 articles in Proquest for a total of 401 articles – 15 of which were duplicates. The titles of these articles were reviewed for relevance, and 154 articles were removed that neither discussed the use of crowdsourcing nor

employed crowdsourcing as a primary research methodology. An additional 99 articles that included editorials and commentaries, articles that only referenced the term crowdsourcing in a non-substantive manner (primarily in a broader social media context), focused on crowdfunding (which is not considered to be crowdsourcing for the purposes of this study), and/or did not deploy crowdsourcing for their research, were removed post abstract review. This resulted in a total of 133 articles.

From those articles, where publicly available and when possible, the first author and corresponding authors email addresses were located. Although a total of 203 researchers were solicited to participate in this research study, 20 of those email addresses "bounced" back, suggesting that a maximum of 183 emails were delivered. Of those 183 emails delivered, 18 individuals agreed to participate in the study.

### 4.2.2 Crowdsourcing Framework

Working from the more than 40 different definitions for the term crowdsourcing, Estellés-Arolas and González-Ladrón-de-Guevara (2012) developed an integrated definition of crowdsourcing which consists of eight discrete characteristics (p. 197):

a) there is a clearly defined crowd;

b) there exists a task with a clear goal;

c) the recompense received by the crowd is clear;

d) the crowdsourcer is clearly identified;

e) the compensation to be received by the crowdsourcer is clearly defined;

f) it is an online assigned process of participative type;

g) it uses an open call of variable extent; and

h) it uses the internet

These characteristics serve as a starting point for constructing a framework for understanding crowdsourcing within a research context. In the absence of a commonly agreed-upon definition for crowdsourcing, these characteristics provide a common language to help facilitate an understanding of its application. Despite the information science undertone, the application of these characteristics within a research context was deemed appropriate, given that they were informed by a non-discipline-specific review of the literature. Further, the characteristics were identified as a result of the comprehensive and integrative process by which the authors developed them (Estellés-Arolas & González-Ladrón-de-Guevara, 2012).

### 4.2.3 First and Second Round Delphi Questions

The method involved two rounds of questionnaires and content analysis to identify whether there was consensus of expert opinion in the use of crowdsourcing for research – and if so, where. For both rounds, a mix of questions were used, including open-ended, editing, ranking, and rating questions. The questions for both rounds can be found in Tables 4.1 and 4.2. In round one, participants were asked to identify key characteristics of crowdsourcing for research, and to rate the importance of characteristics identified by Estellés-Arolas and González-Ladrón-de-Guevara (2012). Round two questions aimed to further understand why researchers were using crowdsourcing and move towards a framework for using crowdsourcing in researcher by trying to improve upon and adapt the Estellés-Arolas and González-Ladrón-de-Guevara (2012) characteristics. The threshold for consensus on positions was set at 70% for rating based questions. While there is no universally agreed upon proportion that is deemed a generally acceptable level for consensus (Powell, 2003), 70% was identified as appropriate for the purposes of this study as a signal of stability given the novelty of the subject matter. A third round was not undertaken as researchers determined that there would be no further consensus

based on the responses in the first two rounds. The two rounds of surveys took place between January 2017 and May 2017.

**Table 4.1. Delphi Survey Questions, Round One**

| |
|---|
| How have you used crowdsourcing in your research? |
| How experienced are you in the application of crowdsourcing? (Sliding scale from 0 – 100).<br>Please explain your rating. |
| How do you see crowdsourcing being applied within the research literature?<br>How is this similar to your own approach / utilization of crowdsourcing?<br>How is this different from your own approach / utilization of crowdsourcing? |
| In your opinion, what are the key characteristics of crowdsourcing research methodology? |
| Estellés-Arolas and González-Ladrón-de-Guevara (2012) suggested that crowdsourcing should consist of the following characteristics. Please indicate the importance of each characteristic, on a scale of 0 to 100.<br>_____ there is a clearly defined crowd<br>_____ there exists a task with a clear goal<br>_____ the recompense received by the crowd is clear<br>_____ the crowdsourcer is clearly identified<br>_____ the compensation to be received by the crowdsourcer is clearly defined<br>_____ it is an online assigned process of participative type<br>_____ it uses an open call to a variable extent<br>_____ it uses the Internet |

**Table 4.2. Delphi Survey Questions, Round Two**

| |
|---|
| Based on the results of round 1, we found that researchers are using crowdsourcing for the following research activities: study design; instrument design; participant recruitment; data collection; and data analysis. Are there any other research related activities that you are aware of that crowdsourcing is being used for? |
| Rank the following, in order of the applicability of crowdsourcing to these research activities with one being the lowest applicability and five being the highest:<br>Study design, Instrument design, Participant recruitment, Data collection, Data analysis |
| Comment on the potential pros and cons of using crowdsourcing for the following:<br>Study design, Instrument design, Participant recruitment, Data collection, Data analysis |

**Table 4.2. Delphi Survey Questions, Round Two (continued)**

| | |
|---|---|
| In round 1, we asked researchers to rate the importance of each of the eight characteristics of crowdsourcing as identified by Estellés-Arolas and González-Ladrón-de-Guevara (2012). The rating scale went from 0 to 100. With the exception of one characteristic ("there exists a task with a clear goal"), the ratings varied significantly and there was little to no consensus on the characteristics of crowdsourcing. | |
| | For the characteristic defined by Estellés-Arolas and González-Ladrón-de-Guevara (2012) as "there is a clearly defined crowd" the average rating was 65 (out of 100) with the range from 12 to 100 and a standard deviation of 30. Why do you think there is so much variability in the importance of this characteristic of crowdsourcing? |
| | For the characteristic defined by Estellés-Arolas and González-Ladrón-de-Guevara (2012) as "the recompense received by the crowd is clear" the average rating was 62 (out of 100) with the range from 19 to 90 and a standard deviation of 25. Why do you think there is so much variability in the importance of this characteristic of crowdsourcing? |
| | For the characteristic defined by Estellés-Arolas and González-Ladrón-de-Guevara (2012) as "the crowdsourcer is clearly identified" the average rating was 57 (out of 100) with the range from 13 to 100 and a standard deviation of 28. Why do you think there is so much variability in the importance of this characteristic of crowdsourcing? |

**Table 4.2. Delphi Survey Questions, Round Two (continued)**

| | |
|---|---|
| | For the characteristic defined by Estellés-Arolas and González-Ladrón-de-Guevara (2012) as "the compensation to be received by the crowdsourcer is clearly defined" the average rating was 55 (out of 100) with the range from 10 to 98 and a standard deviation of 25. Why do you think there is so much variability in the importance of this characteristic of crowdsourcing? |
| | For the characteristic defined by Estellés-Arolas and González-Ladrón-de-Guevara (2012) as "it is an online assigned process of participative type" the average rating was 50 (out of 100) with the range from 10 to 82 and a standard deviation of 22. Why do you think there is so much variability in the importance of this characteristic of crowdsourcing? |
| | For the characteristic defined by Estellés-Arolas and González-Ladrón-de-Guevara (2012) as "it uses an open call to a variable extent" the average rating was 58 (out of 100) with the range from 26 to 100 and a standard deviation of 25. Why do you think there is so much variability in the importance of this characteristic of crowdsourcing? |
| | For the characteristic defined by Estellés-Arolas and González-Ladrón-de-Guevara (2012) as "it uses the Internet" the average rating was 58 (out of 100) with the range from 7 to 100 and a standard deviation of 34. Why do you think there is so much variability in the importance of this characteristic of crowdsourcing? |

**Table 4.2. Delphi Survey Questions, Round Two (continued)**

| | |
|---|---|
| One possible reason for the variability that we are seeing in attitudes towards characteristics of crowdsourcing research is that the terms need to be further operationalized. We are considering additional descriptors for each of the characteristics, and will outline these modifications within this section. For each supplemental descriptor, please rate its importance to the description of crowdsourcing characteristics, on a scale of 1 to 10 (with 1 being not important and 10 being very important). | |
| | The original characteristic is "there is a clearly defined crowd". For each supplemental descriptor, please rate its importance to the description this original characteristic, on a scale of 1 to 10 (with 1 being not important and 10 being very important):<br>_____ The crowd should be defined in terms of skills and/or experience and/or knowledge required.<br>_____ The crowd should include anyone who chooses to participate<br>Please provide any comments or edits to the proposed descriptors. |
| | The original characteristic is "there exists a task with a clear goal". For each supplemental descriptor, please rate its importance to the description this original characteristic, on a scale of 1 to 10 (with 1 being not important and 10 being very important):<br>_____ The overarching purpose of the study is defined.<br>_____ The task to be completed by the participant is explicitly defined.<br>Please provide any comments or edits to the proposed descriptors. |

**Table 4.2. Delphi Survey Questions, Round Two (continued)**

| | |
|---|---|
| | The original characteristic is “the recompense received by the crowd is clear". For the supplemental descriptor, please rate its importance to the description this original characteristic, on a scale of 1 to 10 (with 1 being not important and 10 being very important):<br>_____ If participants are to be compensated, the compensation is explicitly defined<br>Please provide any comments or edits to the proposed descriptor. |
| | The original characteristic is “the crowdsourcer is clearly identified". For each supplemental descriptor, please rate its importance to the description this original characteristic, on a scale of 1 to 10 (with 1 being not important and 10 being very important):<br>_____ The crowd should know who is conducting the research.<br>_____ The crowdsourcer's contact information should be available.<br>Please provide any comments or edits to the proposed descriptors. |

**Table 4.2. Delphi Survey Questions, Round Two (continued)**

| | |
|---|---|
| | The original characteristic is "the compensation to be received by the crowdsourcer is clearly defined". For each supplemental descriptor, please rate its importance to the description this original characteristic, on a scale of 1 to 10 (with 1 being not important and 10 being very important):<br>_____ The crowdsourcer should disclose any compensation to be received as a result of the research.<br>_____ The crowdsourcer should declare any conflict of interest.<br>Please provide any comments or edits to the proposed descriptors. |
| | The original characteristics are "it is an online assigned process of participative type" and "it uses an open call to a variable extent". For each supplemental descriptor, please rate its importance to the description this original characteristic, on a scale of 1 to 10 (with 1 being not important and 10 being very important):<br>_____ Crowdsourcing is an open online participatory process.<br>Please provide any comments or edits to the proposed descriptor. |
| Other than the Internet, what other channels can be used for crowdsourcing? | |
| Do you strongly agree, agree, somewhat agree, neither agree nor disagree, somewhat disagree, disagree or strongly disagree with the following statement:<br>*The same research ethics standards apply for the use of crowdsourcing in research as with any other type of method.*<br>Please share any comments you have regarding research ethics standards when using crowdsourcing in research studies. | |

## 4.3 Results

The findings presented below represent a summary of the feedback from the panelists across both rounds of questions. The findings are organized in four sections:

- Characterizing the Experts Panelists
- The Use of Crowdsourcing for Research
- The Benefits and Challenges of Using Crowdsourcing for Research
- The Characteristics of Crowdsourcing for Research

### 4.3.1 Characterizing the Expert Panelists

In addition to establishing the level of expertise of the panelists, this characterization is important in the context of the Delphi method, owing to its reliance on the expertise of the panel. The panelists were considered to be "crowdsourcing expects", owing to their having applied this nascent technique. Of the 18 respondents who agreed to participate, 15 completed the round one survey and 12 completed the round two survey. The survey participants were a mix of researchers who had used crowdsourcing in their research (83%) and those who studied the topic of crowdsourcing (16%). Panelists had published studies that included both quantitative and qualitative methods.

When asked panelists to self-report (on a scale of 0 – 100) their level of experience with either the application, or subject matter, of crowdsourcing, the range of scores was 21 – 100, with a mean score was 66 and a standard deviation of 23. When asked to explain their ratings, the relative newness of crowdsourcing in research was frequently identified as one of the reasons alongside having employed the approach once or a very limited number of times. Some panelists qualified their expertise in crowdsourcing:

> *I employ crowdsourcing in multiple ways across many platforms, I am an expert in citizen science (a form of crowdsourcing) and regularly review papers on the topic. I am an invited speaker on crowdsourcing across many disciplines.*
>
> Panelist Q

> *In the last four years I have been actively engaged in investigating what would motivate people to participate in social responsible crowdsourcing projects.*
>
> Panelist L

Panelists further suggested that their expertise ranges from *applying it for research purposes* to *knowledge focused on a specific aspect of crowdsourcing*. In addition to identifying their areas of expertise in relation to crowdsourcing, many panelists did qualify their experience and knowledge, acknowledging areas for further growth.

> *I am regarded as an expert in using crowdsourcing as a source of convenience samples. I have also used crowds to code sentiment. However, I have very limited experience in other human computation applications and almost all of my experience is confined to Amazon Mechanical Turk.*
>
> Panelist M

> *I have been involved in ethics approvals, developing web sites, recruiting citizens, supporting them, and generating research results based on their research and presenting these at conferences. However there are many aspects of crowdsourcing that I have yet to experience.*
>
> Panelist N

> *I spend a significant amount of my professional work on crowdsourced technologies for health but certainly have room to grow in my knowledge in this*

*area.*

Panelist O

### 4.3.2 The Use of Crowdsourcing for Research

Panelists identified numerous uses of crowdsourcing in research, based both on the literature and their own experience, including: recruiting research participants; data collection; data analysis; and developing interventions. Individually, panelists used crowdsourcing for participant recruitment, data collection, and data analysis. In some instances, the purpose of crowdsourcing in their research studies was tied to the fulfillment of traditional participant or subject role such as recruitment and the provision of data.

*I have used crowdsourcing to recruit convenience samples of research participants….*

Panelist M

*My project recruited citizens via the web from across the world to contribute data….*

Panelist N

This type of role includes inviting the crowd to complete tasks such as questionnaires, providing personal information, and undertaking other online activities to generate data for research purposes. For example:

*I have used crowdsourcing to get participants to take surveys.*

Panelist G

Panel members who undertook clinical or medical quantitative research studies tended to identify these types of uses for crowdsourcing. In this case, where the primary purpose is to access participants, crowdsourcing appears to be regarded no differently than other recruitment

methods.

Researchers are also using crowdsourcing to engage the crowd in activities such as data collection and analysis – activities that have been more traditionally the role of researchers.

> *Used crowdsourcing to develop intervention tools…*
>
> Panelist E
>
> *…to annotate histopathological images.*
>
> Panelist F
>
> *As a form of data collection from human participants.*
>
> Panelist K

This type of crowd engagement required a different type of involvement based on the needs of the research project. Furthermore, panelists also recognized similar types of crowd uses identified in the literature:

> *…particularly in public health and infectious disease, there are studies that crowdsource information from the public on things like the flu…*
>
> Panelist C

In these instances, the crowd supports the research study through the provision of their knowledge, experience and skills. There is a deeper level of engagement and perhaps an underlying trust factor that the crowd has the capability to undertake such tasks. Leveraging the data collection and analytical capabilities of the crowd are, however, contingent upon the nature of the research, and range from simple tasks such as tracking and monitoring, to more complex types of problem solving.

In limited instances, researchers are building capacity through the engagement of the crowd to undertake co-researcher type activities, and providing education and training to the

crowd:

> *...citizen scientists volunteer their time towards the scientific process in an active research study. They go through extensive training (ethics, enrollment procedures, data prep and some analysis).*
>
> Panelist Q

While this type of research capacity building is common practice with qualitative research methods such as participatory action research, it was only alluded to by the panelists.

The least frequently identified uses of crowdsourcing in research were study and instrument design, with expert panel members citing concerns with lack of knowledge and expertise within the crowd. Most of the expert panelists mentioned the need for role clarity, to distinguish between the roles, and more importantly the skills and training, of the crowd versus those of the researchers. This underscores the fact that specific research expertise and skills are required for many studies, and so areas such as study and instrument design, or even data analysis in some instances, as areas that may extend beyond the capabilities of the crowd. However, this blurring of roles is common in non-research crowdsourcing activities (Howe, 2009):

> *There is a small literature that uses crowds to provide other services traditionally performed by experts (e.g. psychological therapy for subclinical issues, or screening medical images).*
>
> Panelist M

In addition, panelists distinguished between the crowd as general members of the public, and a crowd of experts:

> *Sometimes you need a special crowd, other times any crowd will do. So*

*depending on the task...*

Panelist R

*I also sometime see crowdsourced views of experts.*

Panelist C

This may suggest that the panel recognized the level of skill required in relation to the composition of the crowd.

### 4.3.3 The Benefits and Challenges of Using Crowdsourcing for Research

Panelists were asked why they used crowdsourcing, and to identify some of the benefits and challenges associated with its use. Members of the Delphi panel tended to view the crowd as a supplement to the capacity and capabilities of professional researchers – in other words, participants were seen to be an on-demand pool of resources. The benefits and challenges were categorized into five broad themes: process, people, knowledge, data and experience. Table 4.3 summarizes panelist responses within those categories.

**Table 4.3: The Benefits and Challenges of using Crowdsourcing in Research**

| | Benefits | Challenges |
|---|---|---|
| Process | • Low cost<br>• Fast | |
| People | • Access to large numbers of people<br>• Diverse population | • Self-selected<br>• Lack of representativeness |
| Knowledge | • Outsider perspective<br>• Knowledge mobilization | • "Colloquial" knowledge of subject matter |
| Data | • Large volumes that would not otherwise be possible | • Quality, validity and reliability issues |
| Experience | • Innovation spurred by the diversity of ideas<br>• The crowd benefits from their access to researchers | • Lack of research experience and understanding of research practices |

Based on feedback from the panel, the use of crowdsourcing for research is a highly effective and efficient process for overcoming barriers such as time limitation, data volumes, and costs, regardless of how the crowd is being leveraged.

### 4.3.4 The Characteristics of Crowdsourcing for Research

In an effort to identify a potential framework for crowdsourcing in research, panelists were asked to indicate the importance of each of characteristic of crowdsourcing in the research context, as initially identified by Estellés-Arolas and González-Ladrón-de-Guevara (2012), by rating it on a scale of 0 – 100, with 0 being the lowest rating and 100 being the highest. Table

4.4 summarizes the rating scores and provides the average for each characteristic.

**Table 4.4. Importance of Characteristics of Crowdsourcing in Research**

| Characteristic | Min | Max | Mean | SD |
|---|---|---|---|---|
| there is a clearly defined crowd | 12 | 100 | 65.38 | 30.32 |
| there exists a task with a clear goal | 20 | 100 | 83.62 | 20.43 |
| the recompense received by the crowd is clear | 19 | 90 | 62.31 | 25.46 |
| the crowdsourcer is clearly identified | 13 | 100 | 57.08 | 28.01 |
| the compensation to be received by the crowdsourcer is clearly defined | 10 | 98 | 55.62 | 25.09 |
| it is an online assigned process of participative type | 10 | 82 | 50.23 | 22.77 |
| it uses an open call to a variable extent | 26 | 100 | 58.77 | 25.74 |
| it uses the internet | 7 | 100 | 58.15 | 34.60 |

The only characteristic that achieved an acceptable level of consensus among panelists was “there exists a task with a clear goal” with an average rating of 83.62. When asked to explain the lack of consensus in the importance of each crowdsourcing characteristic, three common themes emerged across the responses from the panelists: (1) issues related to the definitions of terms; (2) the specificity of the task being assigned to the crowd; and (3) the nature of the study in which crowdsourcing is being applied.

On issues related to the definitions of terms and the lack of clarity around language, panelists noted:

*We all have different assumptions of what crowdsourcing is…. Not sure what definitions others are using.*

Panelist Q

*It largely depends on how you interpret this. When Estellés-Arolas and González-Ladrón-de-Guevara talk about a "clearly defined crowd," I interpret that to mean….*

Panelist P

*The issue of compensation in crowdsourcing is always murky, because some scholars interpret the word "compensation" (or in this case "recompense") to mean strictly monetary reward. Of course, many crowdsourcing efforts involve no monetary reward at all,*

Panelist P

This questioning of definitions and interpretation remained a consistent theme throughout the two rounds of the Delphi process.

Panelists also noted the disagreement in what characteristic of crowdsourcing are important for research could result from the specific task being assigned to the crowd:

*The variability of the response may depend on how people leverage the crowd in their work.*

Panelist O

*I interpret that to mean that a task is designed for and targets a particular kind of person…I don't think a clearly defined crowd is nearly as important as a clearly defined problem and solution parameters.*

Panelist P

*So depending on the task we assumed, the rating can vary.*

Panelist R

As noted by Panelists O, P and R, the characteristics of crowdsourcing are also context-specific based on the needs of the study. This, in turn, could influence how researchers are interpreting and applying each characteristic.

Finally, the variation in responses from panelists was also attributed to the nature of the study in which the crowdsourcing was being undertaken:

*It really depends on the study design and the background of the researcher….*

Panelist C

*It will depend on your research question and goals how much you need the crowd accurately defined.*

Panelist I

*Different study fields may have different ideas on this. The requirements of different studies may be varying…*

Panelist E

Given the lack of consensus around the characteristics of crowdsourcing as defined by Estellés-Arolas and González-Ladrón-de-Guevara (2012), expert panel members were provided with supplementary descriptors and statements aimed at clarifying each of the characteristics for the research context and asked to rate its importance in relation to the description of the original characteristic on a scale of 1 to 10 (1 being not important and 10 being very important). This information is summarized in Table 4.5.

**Table 4.5. Importance of Supplementary Statements**

| | Min | Max | Mean | SD |
|---|---|---|---|---|
| **"there is a clearly defined crowd"** | | | | |
| The crowd should be defined in terms of skills and/or experience and/or knowledge required | 1.00 | 10.00 | 6.17 | 3.56 |
| The crowd should include anyone who chooses to participate | 1.00 | 10.00 | 6.00 | 3.02 |
| **"there exists a task with a clear goal"** | | | | |
| The overarching purpose of the study is defined | 1.00 | 10.00 | 7.75 | 3.00 |
| The task to be completed by the participant is explicitly defined | 1.00 | 10.00 | 7.92 | 2.90 |
| **"the recompense received by the crowd is clear"** | | | | |
| If participants are to be compensated, the compensation is explicitly defined | 6.00 | 10.00 | 8.50 | 1.38 |
| **"the crowdsourcer is clearly identified"** | | | | |
| The crowd should know who is conducting the research | 1.00 | 10.00 | 6.50 | 3.10 |
| The crowdsourcer's contact information should be available | 1.00 | 10.00 | 6.08 | 3.68 |
| **"the compensation to be received by the crowdsourcer is clearly defined"** | | | | |
| The crowdsourcer should disclose any compensation to be received as a result of the research | 2.00 | 10.00 | 7.50 | 2.72 |
| The crowdsourcer should declare any conflict of interest | 1.00 | 10.00 | 7.92 | 2.60 |
| **"it is an online assigned process of participative type" and "it uses an open call to a variable extent"** | | | | |
| Crowdsourcing is an open online participatory process | 0.00 | 10.00 | 5.00 | 3.65 |

The only characteristics where the panelists thought the supplementary statements improved and clarified the original statements were “there exists a task with a clear goal”, “the recompense received by the crowd is clear” and “the compensation to be received by the crowdsourcer is clearly defined”. When asked to provide comments and/or edits to each of the supplementary statements, the majority of the comments suggested the supplementary statements did not added anything to the characteristics: “not really sure what this still means still” to “these are 2 totally different things” to “I don’t like the word….”.

For the characteristics related to an open call, online and using the Internet, panelists noted that there were other channels that could be used to facilitate crowdsourcing in research, including: texting, audience response in a live setting, in person events, public spaces, traditional media, sensor systems, community meetings, and recruitment from public places.

> *Ornithologists have been doing crowdsourcing of bird observations since before the internet and are/were organized in birders clubs. If that's one idea of crowdsourcing you have then it's clear that it doesn't need to be online.*
>
> Panelist K

> *Also, some people may see plenty of great crowdsourcing examples that use SMS text messages …. which isn't technically the internet.*
>
> Panelist P

However, panelists did appear to support the idea of an open call:

> *Being open to a large number of relevant people.*
>
> Panelist C

*Crowdsourcing places no particular requirements on the people that comprise the crowd.*

Panelist M

In the context of research, this 'open call' or 'invitation to participate' could speak to the need for inclusivity, sample size and representativeness. Alternatively, it could be tied to the composition of the crowd and ensuring the right mix of skills, knowledge and experience.

Finally, when asked whether the same ethical standards apply when using crowdsourcing in research studies, 67% of the panelists strongly agreed or agreed, 8% were uncertain and 25% somewhat disagreed and disagreed. The panelists who disagreed noted that sometimes crowdsourcing is used because it is easier from a requirements perspective and should not be considered human subjects research.

## 4.4 Discussion

This modified Delphi study demonstrates a broad range of applications for crowdsourcing for the purposes of research alongside the various benefits and challenges in its use. While no general consensus was achieved on the characteristics of crowdsourcing for research purposes, the findings revealed gaps in knowledge, related to the application of crowdsourcing in research both from a methodological and a methods perspective. Recognizing the nouveau nature of crowdsourcing in the research context, this suggests a need to establish a framework that aims to contextualize crowdsourcing as a research method within existing forms of inquiry and research paradigms. It also suggests a strong and pressing need to evaluate the ethics of crowdsourcing.

### 4.4.1 A Conceptual Framework: Using Crowdsourcing in Research

The way in which crowdsourcing is used in research is contingent upon the task that is

assigned, and is therefore fundamentally driven by the needs of the research study. These uses of crowdsourcing can be mapped along a continuum (Figure 4.2). At one end of the continuum, crowdsourcing is used for basic research purposes such as subject or participant recruitment, while at the other end, crowdsourcing serves as a mechanism for capacity building and co-researcher type activities. As you move from left to right the level of expertise, skill and experience required of the crowd increases. Considering the research task with the level of crowd expertise, skill and experience, allows for the role of the crowd to be defined as one of subject/participant, citizen scientist, or co-researchers. Furthermore, these research tasks and roles must be considered in the context of the research methodology – quantitative or qualitative – as each has a different set of implications. The application of crowdsourcing in research should align philosophically and methodologically with the research paradigm in which it is being deployed and therefore should align with the standards of those methods.

**Figure 4.2. Continuum of Crowdsourcing in Research**

This spectrum aligns with the positivist to critical theory paradigms continuum originally created by Lincoln and Guba in 1994 (as cited in Lincoln & Guba, 2011) and later modified by Heron and Reason (1997) with the addition of the participatory paradigm. The continuum allows for fluidity between the categories where the complexity of the task dictates where it rests along the continuum. Further, the role of the researcher also evolves along the continuum, moving from sole conductor of research study to more collaborative, which may entail activities such as educating and training the crowd to facilitate their participation.

The task, therefore, will also dictate the composition and size of the crowd. Where the task is complex, for example developing algorithms for the prediction of disease progression for Amyotrophic Lateral Sclerosis (Küffner et al., 2015), the task is likely to draw experts in the field who are qualified to address the challenge and have an interest in doing so, thus, limiting the size of the crowd. On the opposite end of the spectrum, where the task is more general, such as rating food choices (Turner-McGrievy, Helander, Kaipainen, Perez-Macias, & Korhonen, 2015), the crowd is likely to be larger, with a range of skills and background. Therefore, it is important for researchers to clearly articulate the goal of the study, the task that is being assigned to the crowd, and how the task relates to the study, so participants can self-select based on what they perceive they can contribute. Furthermore, the task will also determine whether there is a need for researchers to invest in crowd capacity building when the task being assigned is more complex and requires specific skills.

When cross-referencing panelist uses of crowdsourcing for research, and its benefits to the published literature on the topic, conceptualization of the crowd as engaging in a participatory, collaborative, co-research approach is seen to a comparatively limited degree among the experts solicited to participate in this study. Concepts related to building public

capacity and training the crowd, knowledge mobilization, and two-way engagement between professional scientists and citizen scientists, appeared to be tertiary objectives. Thus, leveraging the crowd to build capacity for research in the community, and to mobilize knowledge, appear to be underutilized opportunities – particularly given research that suggests that user-driven research can accelerate and improve the innovation adoption process of a solution or new knowledge (Celi, Ippolito, Montgomery, Moses, & Stone, 2014).

### 4.4.2 Definitions of Crowdsourcing for Research

One possible way to consider crowdsourcing is in the context of the research paradigm in which the crowd will be engaged. The paradigm thus defines the roles of the crowd. If the role of the crowd can be defined in generally acceptable research terms (e.g. participant, data collection, analysis, study design, etc.) it makes it possible to develop a lexicon or terminology that aligns with the roles and paradigms from research subject or participant, to citizen scientist, to co-researcher.

One particular characteristic of crowdsourcing, its online nature and use of the Internet, warrants mentioning in the context of defining crowdsourcing for research. Despite the vast majority of definitions referencing the online and Internet aspects of crowdsourcing, panelists in the present study expanded the scope to include other mechanisms and channels, while still maintaining the open call component that enables the inclusion of anyone who wishes to participate. This expansion aligns with inclusivity and equity principles of research. Thus, for the purposes of research, the application of crowdsourcing expands beyond the online and Internet space.

### 4.4.3 Issues of Integrity and Quality

The use of crowdsourcing in research studies has the same demands for integrity and

quality as do studies that deploy other methods. When used for the purposes of recruitment, researchers should acknowledge and recognize issues related to sample representativeness, self-selection, and generalizability where these are important factors based on the research study design. As quantitative and qualitative research methodologies and approaches have differing views on participant recruitment, the way in which each researcher addresses this will be contingent upon his or her mode of inquiry. Similarly, issues related to quality of data will likely be addressed according to research methodology or approach. Various methods to ensure quality have, however, been identified, including bringing reported data together with diagnostic or other clinical measures (Chunara, Smolinski, & Brownstein, 2013); in-house calculations and physician verification (Swan, 2012); and reputation metrics for evaluating user-generated content (McCoy et al., 2014). While research suggests the quality of crowdsourced data is similar to that of non-crowdsourced data (Behrend, Sharek, Meade, & Wiebe, 2011; Swan, 2012), researchers should build mechanisms to ensure quality into their study design where appropriate.

#### 4.4.4 Adherence to Research Standards

When applied specifically to research, crowdsourced studies should adhere to the same ethical standards as other studies. The question that remains, however, is whether the assigned task positions the crowd as participants, citizen scientists or co-researchers. This is a critical question, as it informs how and which ethical standards apply. What remains particularly unclear is whether the crowd represents a group of research participants or researchers – and this opacity is exacerbated in studies where the crowd is actively involved in complex areas of the study.

One area that appears certain is the need for transparency around the benefits to both participants and researchers. The expert panel identified the need to be explicit in explaining the

compensation, monetary or otherwise, to both the crowd and the researchers. In sum, despite the reasons for using crowdsourcing, if it is being used in a research study, the appropriate ethical and professional standards should be maintained.

### 4.4.5 Lessons Learned

There are numerous definitional challenges when considering crowdsourcing in research. While Estellés-Arolas and González-Ladrón-de-Guevara (2012), provide a common definition and framework, they do so in an information technology context that, although not directly transferable to a more general research context, can be adapted to some degree. Furthermore, while created in the information science context, the crowdsourcing characteristics described by Estellés-Arolas and González-Ladrón-de-Guevara (2012) appear to be applicable to design and functionality requirements that are important for undertaking online crowdsourcing (Bassi, Misener, Lee, et al., under review). They do not provide sufficient context appropriate guidance to researchers who would like to undertake crowdsourcing activities beyond online and Internet activities. Additional research is required, therefore, on the application of non-Internet-based crowdsourcing for research.

The Delphi panel experts may have also interpreted the questions differently, based on their own experiences. In some instances, the provided responses reflected a lack of certainty, in terms of what the survey questions were asking, and how it specifically pertained to their work. There was also a range of knowledge and experience in using crowdsourcing for research among the panelists, thus, making it difficult to come to consensus. This was further amplified by the relative novelty of crowdsourcing and the limited body of literature on its use in research.

### 4.4.6 Directions for Future Research

As crowdsourcing is further developed as a method, there is a risk of losing the research context, within the novelty of possibilities presented by information technologies and new communications channels. While these new opportunities should be embraced, this should be done in a manner that maintains the integrity of research paradigms. The ease with which researchers have access to the data and capabilities beyond their institutions and communities, through the crowd, should be leveraged in a responsible manner.

Future research should supplement the information uncovered in this study with case studies and interviews of researchers using crowdsourcing. This may provide an opportunity to further explore and examine the implementation of crowdsourcing in specific settings and implementations. This additional research could also highlight contextual differences that may be dependent upon the research area in which crowdsourcing is deployed.

## 4.5 References

Baker, J., Lovell, K., & Harris, N. (2006). How expert are the experts? An exploration of the concept of "expert" within Delphi panel techniques. *Nurse Researcher*, *14*(1), 59–70. http://doi.org/10.7748/nr2006.10.14.1.59.c6010

Bassi, H., Misener, L., Lee, C., & Johnson, A. M. (n.d.). Exploring the characteristics of crowdsourcing: An online observational study. *Journal of Information Sciences*.

Behrend, T. S., Sharek, D. J., Meade, A. W., & Wiebe, E. N. (2011). The viability of crowdsourcing for survey research. *Behavior Research Methods*, *43*(3), 800–813. http://doi.org/10.3758/s13428-011-0081-0

Brabham, D. C. (2010). Moving the Crowd at Threadless: Motivations for participation in a crowdsourcing application. *Information, Communication & Society*, *13*(8), 1122–1145.

http://doi.org/10.1080/13691181003624090

Celi, L. A., Ippolito, A., Montgomery, R. A., Moses, C., & Stone, D. J. (2014). Crowdsourcing knowledge discovery and innovations in medicine. *Journal of Medical Internet Research*. http://doi.org/10.2196/jmir.3761

Chunara, R., Smolinski, M. S., & Brownstein, J. S. (2013). Why we need crowdsourced data in infectious disease surveillance. *Current Infectious Disease Reports*, *15*(4), 316–319. http://doi.org/10.1007/s11908-013-0341-5

Domecq, J. P., Prutsky, G., Elraiyah, T., Wang, Z., Nabhan, M., Shippee, N., … Murad, M. H. (2014). Patient engagement in research: a systematic review. *BMC Health Services Research*, *14*(1), 89. http://doi.org/10.1186/1472-6963-14-89

Estellés-Arolas, E., & González-Ladrón-de-Guevara, F. (2012). Towards an integrated crowdsourcing definition. *Journal of Information Science*, *38*(2), 189–200. http://doi.org/10.1177/0165551512437638

Frank, L., Forsythe, L., Ellis, L., Schrandt, S., Sheridan, S., Gerson, J., … Daugherty, S. (2015). Conceptual and practical foundations of patient engagement in research at the patient-centered outcomes research institute. *Quality of Life Research*, *24*(5), 1033–1041. http://doi.org/10.1007/s11136-014-0893-3

Heron, J., & Reason, P. (1997). A participatory inquiry paradigm. *Qualitative Inquiry*, *3*(3), 274–294. http://doi.org/10.1017/CBO9781107415324.004

Howe, J. (2009). *Crowdsourcing: Why the power of the crowd is driving the future of business*. New York: Random House.

Küffner, R., Zach, N., Norel, R., Hawe, J., Schoenfeld, D., Wang, L., … Leitner, M. L. (2015). Crowdsourced analysis of clinical trial data to predict amyotrophic lateral sclerosis

progression. *Nature Biotechnology*, *33*(1), 51–7. http://doi.org/10.1038/nbt.3051

Lincoln, Y. S., & Guba, E. G. (2011). Paradigmatic controversies, contradictions, and emerging confluences, revisited. In N. . Denzin & Y. S. Lincoln (Eds.), *The Sage Handbook of Qualitative Research* (2nd ed., pp. 97–128). Thousand Oaks: Sage Publications, Inc.

Linstone, H. A., & Turoff, M. (1975). *The Delphi method: techniques and applications*. Reading, Mass.: Addison-Wesley Publishing Company.

McCoy, A. B., Wright, A., Rogith, D., Fathiamini, S., Ottenbacher, A. J., & Sittig, D. F. (2014). Development of a clinician reputation metric to identify appropriate problem-medication pairs in a crowdsourced knowledge base. *Journal of Biomedical Informatics*, *48*, 66–72. http://doi.org/10.1016/j.jbi.2013.11.010

Mullen, P. M. (2003). Delphi: myths and reality. *Journal of Health Organization and Management*, *17*(1), 37–52. http://doi.org/10.1108/14777260310469319

Okoli, C., & Pawlowski, S. D. (2004). The Delphi method as a research tool: an example, design considerations and applications. *Information & Management*, *42*(1), 15–29. http://doi.org/10.1016/j.im.2003.11.002

Powell, C. (2003). The Delphi technique: myths and realities. *Journal of Advanced Nursing*, *41*(4), 376–82. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12581103

Rowe, G., & Wright, G. (2001). Expert opinions in forecasting: The role of the Delphi technique. In *Principles of Forecasting* (pp. 125–144). Boston: Kluwer Academic.

Steinert, M. (2009). A dissensus based online Delphi approach: An explorative research tool. *Technological Forecasting and Social Change*, *76*(3), 291–300. http://doi.org/10.1016/j.techfore.2008.10.006

Swan, M. (2012). Crowdsourced Health Research Studies: An Important Emerging Complement

to Clinical Trials in the Public Health Research Ecosystem. *Journal of Medical Internet Research*, *14*(2), e46. http://doi.org/10.2196/jmir.1988

Tetroe, J. M., Graham, I. D., Foy, R., Robinson, N., Eccles, M. P., Wensing, M., … Grimshaw, J. M. (2008). Health research funding agencies' support and promotion of knowledge translation: An international study. *Milbank Quarterly*, *86*(1), 125–155. http://doi.org/10.1111/j.1468-0009.2007.00515.x

Turner-McGrievy, G. M., Helander, E. E., Kaipainen, K., Perez-Macias, J. M., & Korhonen, I. (2015). The use of crowdsourcing for dietary self - monitoring: Crowdsourced ratings of food pictures are comparable to ratings by trained observers. *Journal of the American Medical Informatics Association*, *22(e1)*, e112-9. http://doi.org/10.1136/amiajnl-2014-002636

Wiggins, A., & Crowston, K. (2011). From conservation to crowdsourcing: A typology of citizen science. In *Proceedings of the Annual Hawaii International Conference on System Sciences* (pp. 1–10). http://doi.org/10.1109/HICSS.2011.207

# Chapter 5

## 5.1 Emerging Insights

This doctoral dissertation is comprised of three inter-related articles that aim to (i) explore the foundations of crowdsourcing as a means of inquiry, and (ii) put forward a theoretical framework to guide researchers using crowdsourcing in their research. Central to this study was understanding how crowdsourcing is currently being used in research, and how crowdsourcing corresponds to established methods of inquiry. Crowdsourcing is being used for participant recruitment, data collection, data analysis, and less frequently for study and intervention design. In most instances, crowdsourcing is being used to supplement existing research methods, and therefore it could be argued that that crowdsourcing is simply another method for participant recruitment or data collection. However, the underlying principles of crowdsourcing suggest that it is a value-laden methodology within qualitative research paradigms. This chapter discusses emerging insights from these three manuscripts and provides an overall conclusion across the papers. Given the limitations of this work, it also points to future directions that build on the potential of crowdsourcing for research purposes.

The first manuscript (Chapter 2), "Crowdsourcing: A Potential Research Paradigm," examined the concept of crowdsourcing as a form of inquiry and method by considering its core philosophies and principles. By looking at questions of ontology, epistemology, methodology and axiology, Chapter 2 explored 'crowdsourcing' as a research paradigm. The core paradigmatic features of crowdsourcing discussed included: subjective-objective realities, co-creation of knowledge, the metamorphosis of the researcher-participant relationship and the shared value and mutual benefit derived from crowdsourcing health-related research studies. The principles of crowdsourcing resemble those of the participatory paradigm. As a standalone

method, crowdsourcing could be applied across all research quantitative and qualitative paradigms.

The second manuscript (Chapter 3), "Exploring the Characteristics of Crowdsourcing: An Online Observational Study," analyzed the content of studies posted to an online site (http://crowdcrafting.org) specifically focused on crowdsourcing research. Building on the manuscript presented in Chapter 2, the aim of this chapter was to explore how researchers are using crowdsourcing in practice. In addition, Chapter 3 mapped research projects against the eight characteristics of crowdsourcing proposed by Estellés-Arolas and González-Ladrón-de-Guevara (2012), to determine whether these characteristic can serve as a framework for crowdsourced research studies. Findings from this chapter suggest that most of the activities undertaken were micro, repetitive, and task-based, and were usually related to data processing or analysis (such as image and sound identification and text translation). It would appear that crowdsourcing served as a tool for participant recruitment, data collection and analysis. In reframing these findings, these results may also suggest the crowd is taking on roles, or functions, which have typically been the responsibility of the researcher. These new roles of the researcher and the crowd (the "citizen scientists") begins to blur the lines between traditional researcher-participant and participant-researcher relationships. The characteristics of crowdsourcing put forward by Estellés-Arolas and González-Ladrón-de-Guevara (2012) provide a useful framework to guide researchers in their use of crowdsourcing within their own research, particularly in the context of information technology enabled crowdsourcing research studies. While all the characteristics should be addressed by the crowdsourcing interface/platform, it is also important to translate and interpret these criteria in the context of research. Clearly defining and openly articulating the research purpose, the roles of the crowd (researchers vs participants),

in addition to the full disclosure of the researchers involved, should help ensure the integrity of the research. All of this should occur within the ethical and methodological confines required for research studies.

The final manuscript (Chapter 4), "Crowdsourcing for Research: Perspectives from a Delphi Panel," further explores how and why researchers are using crowdsourcing by hosting a Delphi panel of crowdsourcing experts. This expert panel consists of researchers and academics who either use crowdsourcing in their research methods, or research the topic of crowdsourcing itself. The purpose of this panel was to refine the characteristics of crowdsourcing for research and to help inform the theoretical framework of the methodology. While Estellés-Arolas and González-Ladrón-de-Guevara (2012) provide a common definition and framework, they do so in an information technology context which may not be transferable to the use of crowdsourcing in research, but has the potential to be adapted to some degree. The most plausible framework for the application of crowdsourcing in research is based on the research paradigm which in turn defines the roles of the crowd. If the role of the crowd can be defined in generally acceptable research terms (e.g. participant, data collection, analysis, study design, etc.) it makes it possible to align the role with the research paradigms to define the crowds as subjects or participants, citizen scientists, or co-researchers.

As a result of the three interrelated studies found in chapters 2, 3 and 4, a theoretical framework emerges that relies on researchers to understand how crowdsourcing fits into their research paradigm and to ensure it aligns with the key constructs of the paradigm. Further, crowdsourcing should be methodological cohesive and coherent with the research paradigm. Finally, the use of crowdsourcing does not absolve researchers of their ethical responsibilities

and obligations when it comes to conducting research studies, and the same stringent level of professional standards should be employed as would be when using more traditional methods.

## 5. 2 Future Directions

Taken together, the manuscripts presented in this dissertation suggest a number of promising lines of research for the future.

### 5.2.1 Crowdsourcing, Big Data and Artificial Intelligence

The term "big data" refers to a data set that is large in size, consists of various types and formats of information, and has continuous growth (Boyd & Crawford, 2012; Gandomi & Haider, 2015; Kaisler, Armour, Espinosa, & Money, 2013; Raghupathi & Raghupathi, 2014). In the context of crowdsourcing, the concept 'big data' is interesting from at least two perspectives: (i) the crowd as data contributors and (ii) the crowd as data processors. The crowd can create large data sets, as is the frequently cited case with social media sites such as Twitter and Facebook. This can be leveraged to generate significant volumes of data that can be used for research purposes. For example, following the 2011 tsunami in Japan, it was critical for officials to monitor the spread of radiation resulting from the severely damaged Fukushima-Daiichi nuclear power plant. A team of researchers designed the Japan Nuclear Crowd Map to monitor and map real-time radiation data (Kamel Boulos et al., 2011). Within two weeks of the disaster, individual citizens had deployed 577 Geiger counters across the country to help the monitor and track the spread of the nuclear cloud (University of Southampton, 2013). The map combined sensor information with crowdsourced radiation data, and has provide more than 27 million readings since the day of the Fukushima disaster (University of Southampton, 2013). This significant data set would not have been available had it not been for the contributions of the crowd.

The crowd can, however, be employed for purposes that go beyond simply providing data – they may be drawn into an analysis of the data itself. This type of application leverages the knowledge and experience of the crowd, along with its sheer size, to create a "machine" that has substantial collective analytical powers – and this can be particularly useful in scenarios where machine analysis has not yet been fully perfected. For example, the International Space Station has captured approximately two million images of Earth, and while the images are clear, the specifics in the images are not always easy to determine without analysis and categorization, thus rendering them useless for scientific purposes (Gaskill, 2015). In 2015, the Complutense University of Madrid (UCM) launched the project "Cities at Night," to catalog these images to create an open atlas (Gaskill, 2015). Given the large number of images, and the volume of work that would be required, UCM researchers decided to engage the crowd to sort images into those of cities, stars and other objects. This process required the crowd use their knowledge of local geography to identify points in night images, and to match them to positions on map by identifying cities in images and their surrounding area. To date, approximately 20,000 images have been categorized by hundreds of volunteers. To ensure accuracy, each image is being categorized multiple times by different individuals. In addition to creating the atlas, the project should also help determine the optimum number of individuals required to accurately assess each image (Gaskill, 2015). In this example, researchers are able to harness the collective knowledge of local geographies of the crowd in a way that contributes to the analysis and identification of the images as part of the larger project. Importantly, this project may not have otherwise been feasible due to its magnitude—specifically the significant number of people and time required to complete it.

In addition to contributing directly to a specific study, crowdsourcing activities can also support advancements in machine learning and artificial intelligence. Activities such as sorting images, as noted in the Cities at Night project above, which require human intelligence can create large volumes of information to inform the development of algorithms to enhance technologies that may be able to analyse and process this type of data.

### 5.2.2 Conscious Contributions

One of the key distinguishing features of crowdsourcing, informed by this research, is what this author calls 'conscious contributors' or 'conscious contributions.' What makes the contributions of the crowd valuable is that they are willingly and knowingly participating by sharing their knowledge and expertise. In comparison, in some methods used for research—such as leveraging social media data, assessing online behaviour via click-through features, or using CAPTCHA (Completely Automated Public Turing Test to Tell Computers and Humans Apart)—the individuals are not knowingly contributing to research. Often these seemingly innocuous online activities are being used to harness human capabilities without the direct knowledge of the individuals. In the case of crowdsourcing, the crowd is consciously contributing to something they deem valuable and worthy of their time, expertise and skills.

### 5.2.3 Proprietorship of Knowledge

Finally, there are a number of ethical questions that arise when crowdsourcing is used for private benefit. In the business context, some authors have called out the use of crowdsourcing as exploitive—benefiting from the use of low-cost, or even free, labour (Kleemann, Voß, & Rieder, 2008). In the research context, there are a number of questions that should be explored including, but not limited to, the impact of the crowd's contributions on traditional academic and research

performance measures (such as authorship, citations, publications, and grants) and the role of the researcher as it relates to knowledge creation and ownership.

## 5.3 Concluding Thoughts

The volume of information on crowdsourcing for health research has grown throughout the time span covered by this dissertation research, as evidenced by the increase in articles resulting from a Google Scholar search for the terms 'crowdsourcing' and 'health research.' From 2000–13, the search found 478 results vs 2,070 results from 2000–17. Swan (2012) notes similar growth in crowdsourced health studies. Health-related research is only beginning to see the potential of crowdsourcing. Researchers leveraging crowdsourcing can harness unprecedented amounts of data to improve the health and wellbeing of the population. While crowdsourcing presents significant opportunities, it also requires researchers to consider its implications on research methods and methodology to ensure that it meets the appropriate level of quality and rigour required to maintain research standards.

## 5.4 References

Boyd, D., & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society*, *15*(5), 662–679. https://doi.org/10.1080/1369118X.2012.678878

Estellés-Arolas, E., & González-Ladrón-de-Guevara, F. (2012). Towards an integrated crowdsourcing definition. *Journal of Information Science*, *38*(2), 189–200. https://doi.org/10.1177/0165551512437638

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*(2), 137–144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007

Gaskill, M. (2015). Space Station Sharper Images of Earth at Night Crowdsourced For Science.

Retrieved July 27, 2017, from https://www.nasa.gov/mission_pages/station/research/news/crowdsourcing_night_images

Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013). Big Data: Issues and Challenges Moving Forward. In *2013 46th Hawaii International Conference on System Sciences* (pp. 995–1004). IEEE. https://doi.org/10.1109/HICSS.2013.645

Kamel Boulos, M. N., Resch, B., Crowley, D. N., Breslin, J. G., Sohn, G., Burtner, R., … Chuang, K.-Y. S. (2011). Crowdsourcing, citizen sensing and sensor web technologies for public and environmental health surveillance and crisis management: trends, OGC standards and application examples. *International Journal of Health Geographics*, *10*(1), 67. https://doi.org/10.1186/1476-072X-10-67

Kleemann, F., Voß, G. G., & Rieder, K. (2008). Un(der)paid innovators: The commercial utilization of consumer work through crowdsourcing. *Science, Technology & Innovation Studies*, *4*(1), 5–26.

Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*, *2*(1), 3. https://doi.org/10.1186/2047-2501-2-3

Swan, M. (2012). Crowdsourced Health Research Studies: An Important Emerging Complement to Clinical Trials in the Public Health Research Ecosystem. *Journal of Medical Internet Research*, *14*(2), e46. https://doi.org/10.2196/jmir.1988

University of Southampton. (2013). ECS researchers at Southampton University develop new tool to provide radiation monitoring in Japan. Retrieved January 31, 2017, from http://www.ecs.soton.ac.uk/news/4237

# Appendix A: Mapping Research Projects against Crowdsourcing Criteria

| Project name | (1) a clearly defined "crowd" | (2) a task with a clear goal | (3) a clear depiction of the compensation given to the crowd – what does the crowd get | (4) a clear identification of the crowd-sourcer | (5) a clear depiction of the compensation given to the crowd-sourcer | (6) an online, assigned, participative process | (7) an open call to participation in the research | (8) it uses the internet |
|---|---|---|---|---|---|---|---|---|
| 1234 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| 5367 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 1 |
| 1980 BYTE Magazine Comps | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| aaaaa | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| actors of around the world in eighty days | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Air Quality with Biomarkers: Lichens | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Animal Classifier | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| Antimatter Alpha | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Antimatter science project | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| Arthropod Interactions | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Athletics | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 |
| Avatar directed by Cameron¸James | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Balloon Mapping Madrid | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Bardomatic | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Bergman Ingmar | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| BikeFinder | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Biomaterials | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 |
| BLCardSorter | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| Bolidos-UCM | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 1 |

| Project name | (1) a clearly defined "crowd" | (2) a task with a clear goal | (3) a clear depiction of the compensation given to the crowd – what does the crowd get | (4) a clear identification of the crowd-sourcer | (5) a clear depiction of the compensation given to the crowd-sourcer | (6) an online, assigned, participative process | (7) an open call to participation in the research | (8) it uses the internet |
|---|---|---|---|---|---|---|---|---|
| BotOrNot2 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| Bundesanzeiger Captchas | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| Cat and Dog | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Categorize Changing Visual Culture of Medical Journals, 1865-1875 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| Cats Classification | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| CEH Wildlife | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| CERN IT Computing | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 |
| CERN IT Historical Photos | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 |
| CERN Photos | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| CernVM | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| CET Google Scholar SR v2 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| Child Labor | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| Cigarette Commericals | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| Classify factories in China | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 1 |
| Classify Water Images | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| Company Filings | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| country of around the world in eighty days | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Crime, Sex, and Violence | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| CrowdIntent2 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Damage Tagger 2 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Dark Skies ISS | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |

| Project name | (1) a clearly defined "crowd" | (2) a task with a clear goal | (3) a clear depiction of the compensation given to the crowd – what does the crowd get | (4) a clear identification of the crowd-sourcer | (5) a clear depiction of the compensation given to the crowd-sourcer | (6) an online, assigned, participative process | (7) an open call to participation in the research | (8) it uses the internet |
|---|---|---|---|---|---|---|---|---|
| DescribeIT: Supporting blind and partial sighted students | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| Designing Factoria Cultural | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Detect sidewalk information from street... | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 |
| dfg | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| director of around the world in eighty days | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| director of Fearless | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| director of Mulan | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Discover the domestic cats. | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| driftwood3 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| driftwood4 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| Emergency hacklab Kit | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| Emily | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| English Hindi Translation Improvement | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Ernesto | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| European Illegal Parking | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 |
| Facial Features Collector | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| Fearless starred with Li¸Jet | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Female Image in "Pulps" | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| Feynman's flowers | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| Flickr Person Finder Reloaded | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |

| Project name | (1) a clearly defined "crowd" | (2) a task with a clear goal | (3) a clear depiction of the compensation given to the crowd – what does the crowd get | (4) a clear identification of the crowd-sourcer | (5) a clear depiction of the compensation given to the crowd-sourcer | (6) an online, assigned, participative process | (7) an open call to participation in the research | (8) it uses the internet |
|---|---|---|---|---|---|---|---|---|
| FOMC Minutes Redundancy Evaluation | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| fourAM | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| France Floods | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Gaceta Redundancy Evaluation | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| Game of Life | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| GamePro Resemblance | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| Gender and Tech Magazines | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| geotagMars | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| Grace Darlington Project | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Graph Isomorphism Game | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| haiza_firstapp | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Health app quality | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Health website annotation | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| Hello Technology | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| Hidden in the Cover(s) | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| http://crowdcrafting.org/project/test55/ | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 |
| Hysteria and Charcot | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| Identifying sounds | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| Image Clicker | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 |
| Image GeoClicker | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| Insect Catalog | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Is this a good 3D printer ? | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |

| Project name | (1) a clearly defined "crowd" | (2) a task with a clear goal | (3) a clear depiction of the compensation given to the crowd – what does the crowd get | (4) a clear identification of the crowd-sourcer | (5) a clear depiction of the compensation given to the crowd-sourcer | (6) an online, assigned, participative process | (7) an open call to participation in the research | (8) it uses the internet |
|---|---|---|---|---|---|---|---|---|
| jobs4u | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 |
| JoelLichens | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| Jons Person Finder | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| Kodak Trade Circular Ads | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| Landfill Hunter | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 |
| links | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Living Crystals | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| LobbyFacts: Who is networking? | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Localizing Pune's Budget | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| Lost at Night | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| Magicicada | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| Mali Villages | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Man made objects identity | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Map Knitter | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| Mapping Out the Unknown | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| marikana readers notes - Afrikaans | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Math tests with multiple answers | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| May 2013 Oklahoma Tornado Damage | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| MBA | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Melanoma | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| MEP Declarations of Interests | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |

| Project name | (1) a clearly defined "crowd" | (2) a task with a clear goal | (3) a clear depiction of the compensation given to the crowd – what does the crowd get | (4) a clear identification of the crowd-sourcer | (5) a clear depiction of the compensation given to the crowd-sourcer | (6) an online, assigned, participative process | (7) an open call to participation in the research | (8) it uses the internet |
|---|---|---|---|---|---|---|---|---|
| Mining the American X-Ray | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| Mosquito alert | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| movie directed by Allen,Woody and acted by Johansson,Scarlett | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| movies acted by Johansson,Scarlett | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| movies acted by Johansson,Scarlett and Slattery,John | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| movies acted by Slattery,John | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| movies directed by Cameron,James | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| movies directed by Cameron,James2 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| movies directed by Cameron,James3 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| movies directed by Caro,Marc | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| movies directed by Jeunet,Jean-Pierre | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| movies directed by Jeunet,Jean-Pierre and Caro,Marc | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| movies directed by Kubrick,Stanley | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| movies directed by Zhang,Yimou | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| movies in which Kinski,Klaus played | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |

| Project name | (1) a clearly defined "crowd" | (2) a task with a clear goal | (3) a clear depiction of the compensation given to the crowd – what does the crowd get | (4) a clear identification of the crowd-sourcer | (5) a clear depiction of the compensation given to the crowd-sourcer | (6) an online, assigned, participative process | (7) an open call to participation in the research | (8) it uses the internet |
|---|---|---|---|---|---|---|---|---|
| movies in which Ledger,Heath has played | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| movies played by Clooney,George | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| movies played by Gere,Richard | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| movies played by Kinski,Klaus | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| movies played by Ledger,Heath | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| movies where both Hanks,Tom and Spielberg,Steven worked together | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| movies where Tarantino,Quentin appears as an actor | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| movies where Tarantino,Quentin appears as an director | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| ms_fr-640 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| Murtuza Nooranis Photo App2 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Mustafa | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 |
| Neurosurgery and Imagery | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| new project | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Nicholas Cage | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Night Cities ISS | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| nli | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |

| Project name | (1) a clearly defined "crowd" | (2) a task with a clear goal | (3) a clear depiction of the compensation given to the crowd – what does the crowd get | (4) a clear identification of the crowd-sourcer | (5) a clear depiction of the compensation given to the crowd-sourcer | (6) an online, assigned, participative process | (7) an open call to participation in the research | (8) it uses the internet |
|---|---|---|---|---|---|---|---|---|
| North? | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 |
| NYCLichen | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| Oklahoma City Tornado Damage | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| one-two-three | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| Open Science data | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Open Trials FDA Indications | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| OpenOil | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| OpenTrialsFDA drugs indications | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| p2 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Particle Motion | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Particle Motion v2 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| Particle Motion v3 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| PDF Transcription | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| Pentos2 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 1 |
| Personal BotShopper | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| Pharmaceutical Ads | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| Phase2project | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 |
| Philippines Typhoon | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| Picture balance | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| Picture Classifier | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Pinyin Card Catalogue | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Real time results | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |

| Project name | (1) a clearly defined "crowd" | (2) a task with a clear goal | (3) a clear depiction of the compensation given to the crowd – what does the crowd get | (4) a clear identification of the crowd-sourcer | (5) a clear depiction of the compensation given to the crowd-sourcer | (6) an online, assigned, participative process | (7) an open call to participation in the research | (8) it uses the internet |
|---|---|---|---|---|---|---|---|---|
| recent movies directed by Zhang,Yimou | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Resource Annotation system | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Result List One | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Result List Three | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Result List Two | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| ResXplorer (Part I) | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| ResXplorer (Part II) | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| River Ice | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| riverice3 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| riverice4 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| riverice5 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| Robbery Zone | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Rocket Counter | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Rural Geolocator | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Ryan acted by Hanks,Tom | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Say What You See | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Scarasm in Twitter | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| Science photography | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| Shell JIV Transcription | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Skin Lesion Photo Detection | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 |
| SkyTruth FrackFinder | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 |
| Society through Cookbooks | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |

| Project name | (1) a clearly defined "crowd" | (2) a task with a clear goal | (3) a clear depiction of the compensation given to the crowd – what does the crowd get | (4) a clear identification of the crowd-sourcer | (5) a clear depiction of the compensation given to the crowd-sourcer | (6) an online, assigned, participative process | (7) an open call to participation in the research | (8) it uses the internet |
|---|---|---|---|---|---|---|---|---|
| SONYC: Urban Sound Annotation | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| SONYC: Urban Sound Annotation Variant 2 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| SONYC: Urban Sound Annotation Variant 3 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| SONYC: Urban Sound Annotation Variant 4 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| SONYC: Urban Sound Annotation Variant 5 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| SONYC: Urban Sound Annotation Variant 6 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| SONYC: Urban Sound Annotation Variant 7 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| SONYC: Urban Sound Annotation Variant 8 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Sound Cloud | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 |
| Sporting | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| SportPictures | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Steampunk Investigation | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Summer Palace D-Archive | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| Tag these pics | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Tagging pictures | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| TagIT | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Template - Simple Q+A | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| Text Audio Accuracy | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| TextThresher Highlighter | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |

| Project name | (1) a clearly defined "crowd" | (2) a task with a clear goal | (3) a clear depiction of the compensation given to the crowd – what does the crowd get | (4) a clear identification of the crowd-sourcer | (5) a clear depiction of the compensation given to the crowd-sourcer | (6) an online, assigned, participative process | (7) an open call to participation in the research | (8) it uses the internet |
|---|---|---|---|---|---|---|---|---|
| The Face We Make | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Titanic acted by Jack and Rose | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Translate PyBossa to Italian | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| translateEnES | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Tweet Clicker | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Tweet GeoClicker | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| Twitter Emotion Annotator Phase 1 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| UCB Ezproxy link checker | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| Understand the meaning of words | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 |
| Urban Garbage Monitoring | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Urban Parks | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| Valid telephone number identification | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 |
| Video GeoClicker | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| Vimeo Search | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 1 |
| Vital Signs | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| Walmart Parking Lots | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| War Instruments | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| Wasps or Bees | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| WDG Relation Marker | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Whale Flukes | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| What sport is this? | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 |

| Project name | (1) a clearly defined "crowd" | (2) a task with a clear goal | (3) a clear depiction of the compensation given to the crowd – what does the crowd get | (4) a clear identification of the crowd-sourcer | (5) a clear depiction of the compensation given to the crowd-sourcer | (6) an online, assigned, participative process | (7) an open call to participation in the research | (8) it uses the internet |
|---|---|---|---|---|---|---|---|---|
| wingID | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 |
| Winston Churchill Engagement Diaries | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 |
| Women in Pulp Fiction | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| World Science Festival Twitter Analysis | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| WW1 Diary TwitterBot Tweets | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 |
| X-Ray Jounral | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| year of around the world in eighty days | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Zebra Lungs | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| zxcasd | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |

**Western University Non-Medical Research Ethics Board**

## NMREB Delegated Initial Approval Notice

**Principal Investigator**: Dr. Andrew Johnson
**Department & Institution**: Health Sciences\Health & Rehabilitation Sciences,Western University

**NMREB File Number**: 108655
**Study Title**: An examination of the key characteristics of crowdsourcing as a research methodology.

**NMREB Initial Approval Date**: December 20, 2016
**NMREB Expiry Date**: December 20, 2017

**Documents Approved and/or Received for Information**:

| Document Name | Comments | Version Date |
|---|---|---|
| Western University Protocol | Received November 15, 2016 | |
| Recruitment Items | Recruitment Email | 2016/11/15 |
| Letter of Information & Consent | Including Round 1 Survey Questions - Received December 15, 2016 | |

The Western University Non-Medical Research Ethics Board (NMREB) has reviewed and approved the above named study, as of the NMREB Initial Approval Date noted above.

NMREB approval for this study remains valid until the NMREB Expiry Date noted above, conditional to timely submission and acceptance of NMREB Continuing Ethics Review.

The Western University NMREB operates in compliance with the Tri-Council Policy Statement Ethical Conduct for Research Involving Humans (TCPS2), the Ontario Personal Health Information Protection Act (PHIPA, 2004), and the applicable laws and regulations of Ontario.

Members of the NMREB who are named as Investigators in research studies do not participate in discussions related to, nor vote on such studies when they are presented to the REB.

The NMREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000941.

Ms. Harpreet Bassi
Measurement and Methods Field
Health and Rehabilitation Sciences, Western University

---

## EDUCATION

### Degrees

1. PhD Candidate, Western University, Health and Rehabilitation Science (current)

2. Masters of Public Administration, Queen's University (2010)

3. Honours Bachelor of Arts in Political Science, University of Toronto (2000)

### Workshops

1. Moving beyond Interviews: Using Creative Methods in Qualitative Health Research, University of Toronto, June 2016

### Certificates

1. Essential Evaluation Skills, Canadian Evaluation Society (2016)

2. Healthcare Innovation and Entrepreneurship, Duke University via Coursera (2013)

3. Lean Operational Excellence for Managers, Schulich School of Business, York University (2012)

4. Alternate Dispute and Conflict Resolution Certificate, Mediation Centre of Southeastern Ontario, (2010)

5. Project Management Certificate, Ryerson University (2009)

6. Corporate Communications Post-Graduate Diploma, Seneca College of Applied Arts and Technology (2001), Dean's Honour List

---

## PUBLICATIONS & PRESENTATIONS

Bassi, H., Johnson, A. J., Misner, L. "A preliminary inquiry into the utility of crowdsourcing as a research method" Poster Presentation. Canadian Psychological Association National Convention, Toronto, June 2017.

Seminar Presenter, Considering Career Options, Cancer Care Ontario, February 2016.

Seminar Presenter, Communication for Non-Academic Careers, Applied Clinical Pharmacology, University of Toronto, October 2015

Snowdon, A., Bassi, H., Scarffe, A., Smith, A., "Reverse innovation: an opportunity for strengthening health systems". Globalization and Health 2015, 11:2 (7 February 2015).

Snowdon, Anne W., Alessi, Charles., Bassi, Harpreet., DeForge Ryan T., and Schnarr Karin., "Enhancing Patient Experience through Personalization of Health Services". Canadian Healthcare Management Forum Magazine.

Keynote Speaker: Trends in Health Innovation, TVN NCE Technology Symposium (September 20, 2014).

Invited Guest Commentary, Canadian Healthcare Management Forum Magazine, Spring 2014

Preszcator, S., Cramm, H., Schnarr, K., Bassi, H., Kunsch, A., & Snowdon, A. "Reducing the burden of mental illness through innovation in recovery: The expansion of emerging novel models for psychosocial rehabilitation through employment," Presentation to be made at Annual Meeting, North America Primary Care Research Group, New York, NY, November 2014.

Schnarr, K., Kunsch, A., Bassi, H. & Snowdon, A. "Personalization in healthcare: Lessons from industry," 25th Annual Forum, Institute for Healthcare Improvement, Orlando, FL, December 2013.

Guest Lecturer, Joint Motion Program (JuMP), CIHR Strategic Training Program in Musculoskeletal Health Research and Leadership, Western University, November 2013.

Guest Lecturer, Health Innovation and Leadership, Western University, October 2013.

---

## PROFESSIONAL EXPERIENCE

**Director, Centre for Effective Practice**
June 2015 - present

**Director, Strategy, Corporate Affairs and Communications, Ivey International Centre for Health Innovation**
May 2012- June 2015

**Director, Strategy and Project Management, Health Quality Ontario**
November 2010 – April 2012

**Acting Vice President and Director, Public Affairs Planning & Stakeholders Relations, Cancer Care Ontario**
February 2008 – November 2010

**Director, Communications, Ontario Health Quality Council, Ministry of Health and Long-Term Care**
January 2006 – February 2008

**Senior Advisor, MPP /Critic, Health and Long-Term Care**
January 2004 – December 2005

**Advisor, Public Affairs, Canadian Association of Chain Drug Stores (CACDS)**
October 2001 - December 2003

**Nutrition House Services Inc., Public Relations and Marketing Consultant**
March 2002 - March 2003

**Canadian College of Naturopathic Medicine, Public and Donor Relations Officer**
January 2001 - October 2001

---

## AWARDS, VOLUNTEER EXPERIENCE, AFFILIATIONS & MEMBERSHIPS

- ACE Award of Excellence, Stakeholder Relations, Government of Ontario, Ministry of Health and Long-Term Care
- Communications Disabilities Access Canada, Board Member (2012 – 2015)
- Federal, provincial and municipal campaigns (2003 – present, various roles including Campaign Co-chair, and Communications Chair)
- Ontario Health Study, Communications Advisory Committee Member (2008 – 2011)
- Board Secretary, Condo Corporation (2006 – 2008)