a survey of practices for crowdsourcing

Charalampos Saitis
Distributed Digital Music Archives and Libraries Lab
Schulich School of Music, McGill University

Contents

1 Prelude

Humans are critical components for performing quality control in a pattern recognition process, correcting the inevitable errors that automated systems make and ensuring these errors do not compound themselves in subsequent workflow steps (Hankinson et al., 2012; Burlet et al., 2012). Correction, however, can be very time, labour and cost intensive. In the case of optical character recognition (OCR)—the textual counterpart of OMR—some unique solutions have been developed, leveraging collective intelligence and the Web to help offset the costs of this task. The reCAPTCHA project (von Ahn et al., 2008) has produced over 5 billion human-corrected OCR words by presenting the correction task as a spam-fighting challenge to prove that the user is a human and not an automated system. The Australian Newspapers Digitisation Project (Holley, 2009) has created a distributed correction system, where more than 9,000 volunteers have now corrected more than 12.5 million lines of text, with more corrections added all the time.

Both these examples are manifestations of **crowdsourcing**, a Web-enabled, distributed problem-solving model that has emerged during the past decade (Brabham, 2013). Its technical and conceptual underpinnings of crowdsourcing lie at the intersection of collective intelligence, outsourcing and the Internet (Brabham, 2013; Saxton et al., 2013). Crowdsourcing can be described as a **method** where a large number of people, known as contributors, are enlisted to help solve a problem defined by the system owners (corporations, libraries, researchers, etc.), which would normally require intensive (and often tedious), costly labour. Apart from helping reaching out to potentially thousands of users around the globe, the Internet offers a high degree of automation and unique possibilities for user management (e.g., wiki, discussion groups) (Doan et al., 2011). Ever since the term was coined in 2006 (Howe, 2006), many names (citizen science, crowd wisdom, human computation, etc.), definitions (Estellés-Arolas and de Guevara, 2012) and typologies (Geiger et al., 2011) of crowdsourcing have appeared in an attempt to provide this new concept with a theoretical framework.
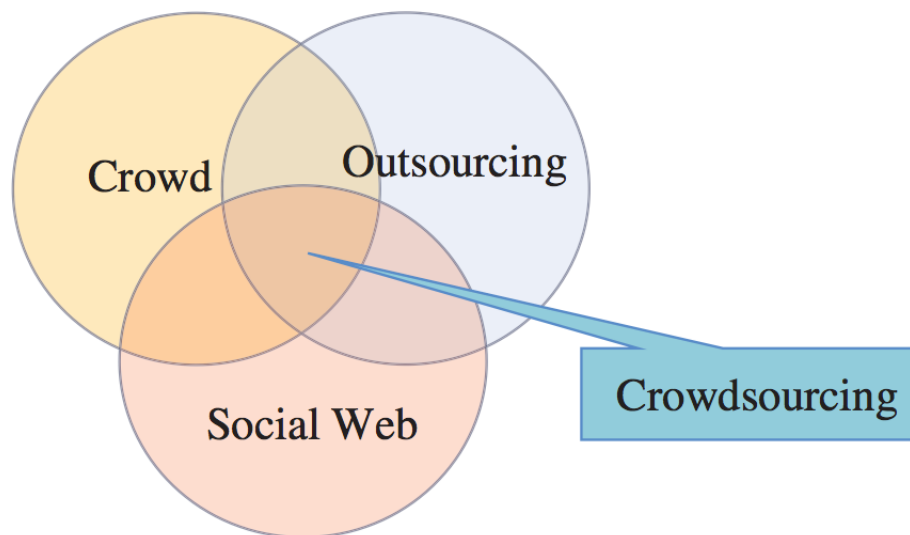

Figure 1.1: Conceptual underpinnings of crowdsourcing. From Saxton et al. (2013).

2 Crowdsourcing in practice: Some examples

Project	Description
InnoCentive	Crowdsourcing solution to problems defined by business, research, and social organizations through a cloud-based innovation management platform.
IdeaStorm by Dell	Dell is inviting people to identify ideas about computers and computer services that are most important and most relevant to the public.
Nabuur	From Wikipedia's entry: The Nabuur Foundation is a non-profit organization based in Giethoorn, Netherlands. It aims to help in the development of impoverished communities by enabling people from other countries to help the community members through Nabuur's website.
MicroPlace	A social business now owned by eBay. It facilitated peer-to-peer microfinance loans, but not on a charitable basis like its competitor Kiva.
CrowdSpirit	CrowdSpirit was an innovation community based on the idea of crowdsourcing the whole product development process, from ideation to distribution. A good example of what makes crowdsourcing *not* work (Chanal and Caron-Fasan, 2010).
Marketocracy	An online "farm system" for investors. Individual aspiring investors are invited to prove themselves/test their theories by managing a model portfolio in real time. The best managers are then allowed to apply their skills to client accounts.
Mechanical Turk	A crowdsourcing platform/marketplace by Amazon. Launched in 2005. Businesses (known as Requesters) can post so-called Human Intelligence Tasks (HITs), setting a monetary compensation. Individuals (known as Workers) can then choose among existing HITs and complete them. HITs can be placed either through the Amazon Mechanical Turk API or directly on https://requester.mturk.com/.
reCAPTCHA	The reCAPTCHA project (von Ahn et al., 2008) produces human-corrected OCR words/human-labelled images by presenting the correction task as a spam-fighting challenge (CAPTCHA) to prove that the user is a human and not an automated system. It was acquired by Google in 2009. Currently used by Google Books, Google Maps and the New York Times archives among others.
iStockphoto	A crowdsourced online marketplace for royalty-free stock images.
Spotting the Weedy Invasives	Volunteer identifying and mapping of invasive plants along hiking trails in the New York/New Jersey Highlands. Operated by Rutgers University and the New York/New Jersey Trail Conference and sponsored by the United States Department of Agriculture.
Invasive Plant Atlas of New England	Another project that invites volunteers to identify and map invasive plants (in New England).
ReClam the Bay	Community volunteers are engaged in restoring shellfish to Barnegat Bay, New Jersey.
Project Gutenberg	A book digitization and archival project where volunteers create and/or proofread digital versions of books (some formats: plain text, HTML, PDF, mobi). Founded in 1971 by Michael S. Hart, Project Gutenberg is the oldest digital library.
IBM's 2006 Innovation Jam	In July 2006, IBM hold an online open-call brainstorming session for innovative business ideas. The event reportedly attracted more than 150,000 participants from 104 countries and 67 different companies over the course of two 72-hour sessions.[1]
Cisco Systems' I-Prize	An external innovation competition aimed to find the next major business opportunity for Cisco Systems. Run in 2007.

[1] http://www-03.ibm.com/ibm/history/ibm100/us/en/icons/innovationjam/

continued from previous page

Project	Description
LiveOps	Maintains a pool of crowdsourced call center agents.
99designs	Maintains a pool of crowdsourced graphic designers.
SnowTweets	The project invites people to report their local snow depth measurements, which are then mapped and used by environmental researchers at the University of Waterloo, Canada. User measurements are reported through Twitter. Launched in 2009.
Foldit	A computer game wherein players interactively reshape improperly folded protein structures in the direction they believe will lead to the highest score. The score increases as the energy of the structure decreases; the goal is to find the native state of the protein conformation (Cooper et al., 2010).
Fossil Finders	A project that brings together educators, students, and researchers from the University of Georgia and the Paleontological Research Institution in Ithaca, New York, to learn about Devonian-age fossils.
Galaxy Zoo	Astronomy project which invited people to assist in the morphological classification of more than 900,000 galaxies by eye that had been imaged by the Sloan Digital Sky Survey at the Apache Point Observatory in New Mexico, USA. Launched in 2007, it led to the creation of Zooniverse.org, a framework to support and host citizen science projects. Various versions followed, which either asked for more detailed classifications, classified images of galaxies collected by other observatories/telescopes, or compared different datasets.
Vipul's Razor	(From the source:) Through user contribution, Razor establishes a distributed and constantly updating catalogue of spam in propagation that is consulted by email clients to filter out known spam. Detection is done with statistical and randomized signatures that efficiently spot mutating spam content.
The ESP Game	The Extra Sensory Perception (ESP) Game reveals the same image to two random players and asks each to guess what the other person has written to describe it. If they agree, that word or phrase is then used to annotate the picture. The same image is presented to multiple pairs of players until a detailed label is built up. It was originally developed by Luis von Ahn in 2003 (von Ahn and Dabbish, 2004). In 2006, Google bought a license to create its own version of the game, the Google Image Labeler, which run until 2011.
helpfindjim.com	An open call to help find one Jim Gray.
International Children's Digital Library	The ICDL Foundation is a nonprofit project that was created in 2003 by researchers at the University of Maryland at College Park. Its goal is to make the best in children's literature available online free of charge. Users contribute and/or translate books. ICDL currently[2] offers 4619 ebooks in 59 languages, complete with 2 iPhone/iPad apps; users (including contributors) come from 228 countries.
TagATune	TagATune is an audio-based online game that aims to extract descriptions of sounds and music from human players. The process is relatively similar to the ESP game, i.e., two independent players have to agree on a description for a given sound (Law et al., 2007).
Soylent	From the source: Soylent is a word processor with a crowd inside: an add-in to Microsoft Word that uses crowd contributions to perform interactive document shortening, proofreading, and human-language macros. Underlying Soylent is a new programming design pattern called Find-Fix-Verify that splits tasks into a series of generation and review stages to control costs and increase quality. Paper: Bernstein et al. (2010).
MonoTrans	From the source: An iterative protocol in which monolingual human participants work together to improve imperfect machine translations.

---

[2]Website visited on December 3, 2014.

continued from previous page

Project	Description
Kasparov versus the World	An Internet-based chess game played in 1999 between Garry Kasparov and a "world team" comprised over 50,000 people from more than 75 countries who decided on moves by plurality vote (Wikipedia entry).
VizWiz	A free talking iPhone app that enables blind and visually impaired users to receive quick answers to questions about their surroundings. It combines automatic image processing, anonymous web workers (through an abstraction layer on top of the Amazon Mechanical Turk API), and members of the user's social network in order to collect fast and accurate responses.
ChaCha	Wikipedia entry: ChaCha is a human-guided search engine. It provides free, real-time answers to any question, through its website, or by using one of the company's mobile apps. From the source: It is used primarily as a marketing tool for brand advertisers to reach and engage with their target audience in real-time.
The Australian Newspapers Digitisation Project	Ongoing program that is digitising historic Australian newspapers from 1803 onwards using OCR. To improve and enhance the quality of the data, the public is correcting OCR transcriptions and/or adding tags to full-text articles. The program forms part of the larger Trove project by the National Library of Australia.
What's on the menu?	The New York Public Library is inviting people to help improve its historical restaurant menu collection by transcribing the menus, dish by dish.
Transcribe Bentham	A public collaborative transcription initiative at the University College London to support the creation of a vast digital repository of Jeremy Bentham's writings. Users are asked to transcribe text and encode their transcriptions in TEI-compliant XML (Text Encoding Initiative).
What's the score at the Bodleian?	Ongoing music score annotation project which invites people to describe newly digitized scores of the Bodleian Libraries at the University of Oxford (UK), mostly piano music from the nineteenth century. Uses the Zooniverse interface/platform.
Old Weather	Crowdsourcing transcriptions of handwritten text from historical ship logs to improve reconstructions of past weather and climate. Contributors can provide new transcriptions or correct computer-transcribed (OCR) ones. Uses the Zooniverse interface/platform.
1001 Stories Denmark	A collection/website of about 1001 cultural heritage sights in Denmark. End-users are invited to contribute by adding/recommending new places, tagging/commenting on existing stories, or uploading images and videos.
Netherlands Institute for Sound and Vision wiki	Volunteer contribution of contextual information on Dutch television programmes, broadcasters, presenters, etc.
UK_Soundmap	The British Library invited people to capture "sounds of today" through using the mobile pp Audioboo. Sound clips are uploaded together with some metadata such as geo-coordinates, which help place them on an interactive map and make them searchable.
Wir Waren So Frei	The Deutsche Kinemathek and the Bundeszentrale für politische Bildung created an archive of visual testaments to the fall of the Berlin Wall through inviting the public to contribute photographs, films, and the stories behind them.
Mass Observation Project	The original Mass Observation project was started in 1937 by a group of people who aimed to record everyday life in Britain. To this purpose, they recruited voluntary observers/writers. This work continued until the early 1950s. In 1981, a new public call was launched, continuing until today.

continued from previous page

Project	Description
Civil War Phases	The Library of Congress published a rare collection of Civil War photographs on flickr and invited people to identify soldiers and/or photographers.[3]
Waisda?	A Web-based game developed by the Netherlands Institute for Sound and Vision, KRO broadcasting, and VU University Amsterdam wherein players annotate television heritage. Similar to ESP, players gain points if they guess correctly what their opponent has written within a given amount of time.
Emotify	A Web-based (standalone or as a Facebook app) game wherein players describe induced emotions (up to three from a list of nine) upon listening to one-minute music fragments (Aljanaki et al., 2014).
Hooked	A game that collects "hook" ("the most salient, easiest-to-recall fragment in a piece of music") and "catchiness" ("the long-term musical salience") data for music songs from its players through three tasks: recognition (listen to random excerpt and recognize within a few seconds), verification (sing along recognized tune), and prediction (find catchier excerpt) (Burgoyne et al., 2013).
Click!	"Click! A Crowd-Curated Exhibition" was launched by the Brooklyn Museum in 2008. The museum first invited photographers to submit electronically images that responded to the theme "The Changing Faces of Brooklyn." The general public was then invited to assess the submissions online through a custom built evaluation tool.
Kaggle	Crowd-sourced data prediction. Companies, organizations, and researchers post their data and data analysts from all over the worlds are invited to compete for finding the best model(s).
Indaba Music	From Gomes et al. (2012): It allows to create, edit and mix music collaboratively. Crowds of people from anywhere in the world can collectively produce music without having to be professionals in the area.
Queremos!	Translated from the source: Queremos! (We want!) is an innovative platform [in Brazil] that connects fans, artists and producers to promote shows. The initiative comes from the crowd itself (fans), based on which artist(s) they "want!"
PickTheBand.com	From Gomes et al. (2012): It allows fans to participate in decisions made by their favorite artists. From the source: "The world's first fan run label."
Voices.com	An online marketplace for voice overs: Maintains a pool of crowdsourced "voices" and voice-over jobs.
Askville	A user-driven research site founded by Amazon and run between 2006 to 2013. Unlike other crowdsourced Q&A initiatives, Askville was designed with game characteristics (with a point system based on asking/answering questions). It also provided discussion boards.
webook.com	An online platform for writers, readers and literary agents geared towards discovering new writers and helping them get published.
Lending Club	From Wikipedia: Lending Club operates an online lending platform that enables borrowers to obtain a loan, and investors to purchase notes backed by payments made on loans.
Angie's list	Crowd-sourced reviews of local businesses.
Lulu	From Wikipedia: An online print-on-demand, self-publishing and distribution platform.
inkling markets	From the source: Inkling offers crowdsourced forecasting applications and solutions to help organizations improve their decision making through more accurate forecasts and greater internal transparency.

[3] https://www.flickr.com/photos/library_of_congress/sets/72157625520211184/

Project	Description
eYeka	From the source: eYeka is a crowdsourcing platform that connects creative individuals with brands to boost their return on marketing investment.
OpenDataParis	An open data project by the Mayor of Paris to enable citizen participation in urban issues such as accessibility and bicycle parking.
Staying Alive	Formerly known as Arrêt Cardiaque, Staying Alive is a mobile app mapping defibrillators worldwide and allowing users to report new ones.
SeeClickFix	A Web tool for community activism. Citizens can publicly report neighborhood issues to appropriate services such as local government.
Elance	From Nakatsu et al. (2014): An online employment platform. Using Elance, a requestor can solicit work (e.g., design a website, write a program, create a marketing plan) from qualified professionals throughout the world who bid to be hired.
Mark the Spot by AT&T	A mobile app that enables AT&T customers to submit wireless coverage problems in real time, together with location information.
CureTogether	From Nakatsu et al. (2014): A website where patients with medical conditions can report on the performance of treatments they received. By gaining access to the millions of ratings reported to the website, patients can learn from the experiences of other patients with the same medical condition.
40 Fires Foundation	An open source hardware design project by Riversimple. The company has published the design schematics of their hydrogen-powered car under a Creative Commons Attribution-Noncommercial license, inviting the public to take part in the design process.
Mob4Hire	Crowd-sourced mobile app testing services and market research. It enables collaboration between developers and professional users who evaluate app usability under a fixed-fee structure.

3 Definitions

Source	Definition
Howe (2006, 2008)	Crowdsourcing is the act of taking a task traditionally performed by a designated agent (such as an employee or a contractor) and outsourcing it by making an open call to an undefined but large group of people. Crowdsourcing allows the power of the crowd to accomplish tasks that were once the province of just a specialized few. Or to put it another way, crowdsourcing is to take the principles which have worked for open source software projects and apply them right across the entire spectrum of the business world.[1]
Brabham (2013)	Crowdsourcing is an online, distributed problem-solving and production model ... It is a model capable of aggregating talent, leveraging ingenuity while reducing the costs and time formerly needed to solve problems.[2]
Geerts (2009)	Crowdsourcing is the online outsourcing of a task to (a group of) private individuals in the form of an open call.
Erickson (2011)	Tapping the perceptual, cognitive or enactive abilities of many people to achieve a well-defined result such as solving a problem, classifying a data set, or producing a decision.
Doan et al. (2011)	A system is a crowdsourcing system if it enlists a crowd of humans to help solve a problem defined by the system owners, and if in doing so, it addresses the following four fundamental challenges: How to recruit and retain users? What contributions can users make? How to combine user contributions to solve the target problem? How to evaluate users and their contributions?
Geiger et al. (2012)	Crowdsourcing information systems are a special case of information systems that produce informational products and/or services for internal or external customers by harnessing the potential of crowds.[3]

---

[1] On his crowdsourcing website/blog (http://www.crowdsourcing.com/) Howe rearranges this definition in two "versions:"

- **The White Paper Version:** 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.

- **The Soundbyte Version:** The application of Open Source principles to fields outside of software.

[2] Brabham draws a line between (what he defines as) crowdsourcing and such massive online participatory phenomena as Wikipedia and Linux. For Brabham, crowdsourcing combines traditional top-down production with bottom-up user production: "the locus of control is between organization and online community." Wikipedia and Linux lack the former: contributors are not told what articles/code to create.

[3] Here the authors built on the definition of information systems (IS) as work systems proposed by Alter (2008): "an IS is a system in which human participants and/or machines perform work (processes and activities) using information, technology, and other resources to produce informational products and/or services for internal or external customers."

Source	Definition
Estellés-Arolas and de Guevara (2012)	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.[4]
Lebraty and Lobre-Lebraty (2013)	Crowdsourcing is the externalization by an organization, via an application using the Internet protocol, of an activity to a large number of individuals whose identities are most often anonymous.
Saxton et al. (2013)	Crowdsourcing is a sourcing model in which organizations use predominantly advanced internet technologies to harness the efforts of a virtual crowd to perform specific organizational tasks.
Pedersen et al. (2013)	A collaboration model enabled by people-centric web technologies to solve individual, organizational, and societal problems using a dynamically formed crowd of interested people who respond to an open call for participation.
Nakatsu et al. (2014)	We define crowdsourcing as a four-step process in which: (1) a requestor (either an individual or organization) identifies a specific task to be performed or problem to be solved; (2) the requestor broadcasts the task or problem online; (3) the crowd performs the task or solves the problem; (4) depending on the nature of the task, the requestor either – (4a) sifts through the solutions and selects the best solution (selective crowdsourcing), or – (4b) aggregates/synthesizes the crowd's submissions in a meaningful way (integrative crowdsourcing).
Merriam-Webster.com	Crowdsourcing is the practice of obtaining needed services, ideas, or content by soliciting contributions from a large group of people and especially from the online community rather than from traditional employees or suppliers.[5]

---

[4]Estellés-Arolas and de Guevara derived this definition through textual analysis of 40 definition-like descriptions of crowdsourcing extracted from 203 publications in ACM, IEEE, ScienceDirect, SAGE, and Emerald. The authors identified 8 characteristics which were then integrated into a more overarching definition: 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; h) it uses the internet. An attempt to verify this integrated definition through 11 Internet-based projects such as Wikipedia and Threadless is also reported (see Table 4, page 197).

[5]The Wikipedia entry for "crowdsourcing" cites this definition.

4 Describing crowdsourcing: Typologies and taxonomies

Source	Description	Examples
Surowiecki (2004)	**Types of problems that can affect *collective intelligence*[1]**	
	• *Cognition problems* (problems that have a definitive solution) • *Coordination problems* (requires the crowd to coordinate their behaviour with one another) • *Cooperation problems* (involve getting people to collaborate and work jointly together)	
Howe (2008)	***What the crowd-***	
	-*knows*: Use and application of collective intelligence Subtypes: prediction markets, crowdcasting, brainstorming Further subtype (Geerts, 2009): crowdcasting/storming hybrid	InnoCentive (crowdcasting), IdeaStorm (brainstorming), Nabuur (hybrid)
	-*creates*: Production of mass creative works Subtypes (Geerts, 2009): product (all contributions are combined), market (contributions are used in a p2p form)	Wikipedia (product), Threadless (market)
	-*thinks*: Filtering and organizing of vast information stores	Amazon, Threadless (consumer reviews/ratings)
	-*funds*: Use of the crowd's collective pocketbook Subtypes (Geerts, 2009): product (all contributions are combined), market (contributions are used in a p2p form)	Kickstarter (product), MicroPlace (market)
Vukovic (2009)	**Two primary dimensions**	
	Mode (competition / marketplace)	Threadless (competition), CrowdSpirit (marketplace), Marketocracy (competition), Mechanical Turk (marketplace)
	Function (design and innovation / development and testing / marketing and sales / support)	Threadless (design and innovation), CrowdSpirit (development and testing), Marketocracy (marketing and sales), Mechanical Turk (support)
Corney et al. (2009)	**Three dimensions**	
	Nature of the task (creation / evaluation / organisation)	InnoCentive (creation), market surveys (evaluation), reCAPTCHA (organisation)

---

[1] Nakatsu et al. (2014): "Predates the crowdsourcing literature by a few years, but provides a useful framework for studying problems solved by a crowd."

Source	Description	Examples
	Nature of the crowd (any individual / most people / experts)	reCAPTCHA (any individual), iStockphoto (most people), InnoCentive (experts)
	Nature of the payment (contribution: voluntary / rewarded with a bonus or prize / rewarded at a flat rate)	iStockphoto (prize), reCAPTCHA (voluntary), InnoCentive (flat rate)
Geerts (2009)	**Attributes of crowdsourcing**[2]	

- Intervention attributes
  - Role of the company (mediator / own initiative)
  - Type of participation (voting; rating; commenting; forum; weblog; social network; wiki; upload; financial)
  - Party that takes initiative (crowd initiative / company initiative)
  - Whose contributions are used (individual / crowd / combination / p2p)
  - Financial rewards (yes / no)
  - Interaction (yes / no)
  - Input control (contributions not public; type of participation; screening; registration)
  - Output control (company / crowd)
  - Intellectual Property (company / crowd / public / not relevant)
- Context attributes
  - Phase (concept; development/design; testing; customer service; production; decision making; support)
  - Field (science; design; consumer goods; music; journalism)
- Outcome attributes
  - Size and activity of crowd
  - Composition participants
  - Number and quality of contributions
  - Company effort

	Three types of public participation in scientific research	
Bonney et al. (2009)	*Contributory projects* (members of the public primarily contribute data)	Spotting the Weedy Invasives
	Collaborative projects (members of the public contribute data but also may help to refine project design, analyze data, or disseminate findings)	Invasive Plant Atlas of New England

[2]These theoretical attributes are grouped into three categories representing design propositions according to the so-called CIMO-logic: "in this class of problematic Contexts, use this Intervention type to invoke these generative Mechanism(s), to deliver these Outcome(s)" (Denyer et al., 2008, p. 395). Mechanism attributes are not included as Geerts considers them to be mainly based on empirical confirmation.

continued from previous page

Source	Description	Examples
	Co-created projects (designed by scientists and members of the public working together; at least some of the public participants are actively involved in most or all steps of the scientific process)	ReClam the Bay
Zwass (2010)	**Taxonomic framework of factors in *co-creation***	
	• Performers (the world, prequalified individuals, community members, skilled contributors) • Motivation (altruistic desire to contribute, passion, inner need to reciprocate, enjoyment / state of flow / playfulness, self-expression, identity construction, forming personal relationships, community norms, competitive spirit, learning, satisfying one's affiliation needs, self-esteem and self-efficacy, thymotic strivings, acquiring social capital and peer recognition, career advancement, own use, nonmonetary rewards, signaling, financial rewards)[3] • Governance (individual autonomy, collective norms, software code, facilitators, adhocracy, bureaucracy, market mechanism, hybrid forms) • Task characteristics (structural complexity, intellectual demands, effort intensity, time frame) • Product aggregation (searchable corpus, hyperlinking, progressive refinement, statistical ratings and rankings, competitions and voting, information markets, bottom-up taxonomy, moderators / auditors / facilitators) • Economic beneficiary (the world, the community, the sponsoring firm, the aggregator, the contributors)	
Rouse (2010)	**Three dimensions**[4]	
	Distribution of benefits (community / individualistic)	Project Gutenberg (community), Threadless (individualistic), Cisco Systems' 2007 I-Prize (individualistic)
	Supplier capabilities/Nature of the task (simple / moderate / sophisticated)	Project Gutenberg (simple), Threadless (moderate), Cisco Systems' 2007 I-Prize (sophisticated)
	Forms of motivation (*intrinsic*: social status, altruism, personal achievement and learning; *extrinsic*: self-marketing, instrumental, token compensation, market compensation)	

---

[3]Some of these terms are further explained in Table 1, page 29 of the source.

[4]A taxonomy based on these dimensions is proposed in Fig. 1 of the source.

continued from previous page

Source	Description	Examples
Malone et al. (2010)	**Describing the goals and design of a *collective intelligence system* through four main questions**	
	• *What is being done?* [Goal] (create / decide) • *Who is doing it?* [Staffing] (crowd / hierarchy or management) • *Why are they doing it?* [Incentives] (money / love / glory) • *How are they doing it?* [Structure/Process] 　– How – create (collection / collaboration) 　– How – decide (group decision / individual decisions)	
	Four characteristics	
Geiger et al. (2011)	Preselection of contributors (qualification-based / context-based / both / none)	iStockphoto (qualification-based), InnoCentive (context-based), LiveOps (both), reCAPTCHA (none)
	Accessibility of peer contributions (modify / assess / view / none)	Wikipedia (modify), iStockphoto (assess), 99designs (view), reCAPTCHA (none)
	Aggregation of contributions (integrative / selective)	reCAPTCHA (integrative), InnoCentive (selective), Wikipedia (integrative), iStockphoto (integrative)
	Remuneration for contributions (fixed / success-based / none)	LiveOps (fixed), InnoCentive (success-based), iStockphoto (success-based), reCAPTCHA (none)
	Five types of *citizen science*	
Wiggins and Crowston (2011)	*Action* (encourage participant intervention in local concerns; employ participatory action research approaches mostly based on "bottom-up" organizing)	ReClam the Bay
	Conservation (volunteer engagement focuses on data collection activities; include explicit educational goals or content)	Invasive Plant Atlas of New England
	Investigation (focused on scientific research goals requiring data collection from the physical environment; education not always an explicit goal; range in score from regional to international)	SnowTweets
	Virtual (similar to Investigation but all project activities are ICT-mediated with no physical elements whatsoever)	Foldit

continued from previous page

Source	Description	Examples
	Education (primary objectives are education and outreach)	Fossil Finders
	Classification along nine dimensions	
Doan et al. (2011)	Nature of collaboration (explicit / implicit)	Wikipedia (explicit), reCAPTCHA (implicit)
	Target problems (evaluating / sharing / building artifacts / networking / other)	Threadless (evaluating), Galaxy Zoo (sharing), Wikipedia (building), LinkedIn (networking), reCAPTCHA (other)
	How to recruit and retain users[5] (require users[6] / pay users / ask for volunteers / make users pay for service / piggyback on user traces)	Mechanical Turk (pay), Wikipedia (ask), reCAPTCHA (make pay for service), Google (piggyback)
	What can users do[5] (explicit: review; vote; tag; share knowledge; build software / implicit: play games with a purpose; solve CAPTCHAs; keyword search; browse Web sites) → Need to consider: cognitive demand, impact of a contribution, machine contributions (e.g., correcting OCR/OMR transcriptions), user interface design	Amazon (review), Threadless (vote), del.icio.us (tag), Galaxy Zoo (share knowledge), Linux (build software), ESP (gamification), reCAPTCHA, Google
	How to combine user input[5] (loosely ⤳ tightly)	tagging (loosely), Wikipedia (tightly), Amazon (reviews are not combined but ratings are)
	How to evaluate user input[5] (manually: monitoring by system owners or distributed among trusted users / automatic: machine tests)	
	Degree of manual effort (little ⤳ substantial) → Need to consider: how much the system is automated, how to divide the manual effort between the users and the system owners	combine ratings (little), combine code (substantial)
	Role of human users (slaves / perspective providers / content providers / component providers)	reCAPTCHA (slaves), Flickr (content), Wikipedia (slaves, perspective, content)
	Architecture (standalone / piggyback)	Wikipedia (standalone), reCAPTCHA (piggyback)

---

[5]These dimensions correspond to the fundamental challenges that a crowdsourcing system should address according to the authors' definition (see Sec. 3 and Sec. 5).

[6]If one has the authority to do so. For example, students may be asked to execute a task as part of their course requirements.

continued from previous page

Source	Description	Examples
Schenk and Guittard (2011)	**Two basic dimensions**	
	Nature of the process (integrative / selective)	Wikipedia (integrative), Threadless (selective)
	Task complexity (simple / complex / creative)	Wikipedia (complex), Threadless (creative)
Yuen et al. (2011)	**Categorization based on four characteristics**	
	• Application (voting system / information sharing system / game / creative system) • Algorithm • Dataset • Performance (user participation; quality management; cheating detection) → user participation factors: demographics, financial incentives, intrinsic incentives, worker behavior	Vipul's Razor (voting), Wikipedia (information sharing), ESP (game), Threadless (creative)
	Four types of crowdsourcing based on a space-time four quadrant model[7]	
Erickson (2011)	*Audience-centric*—same time/place	digital audience backchannels
	Event-centric—same time/different places	innovation jams
	Geocentric—different times/same place	FourSquare
	Global—different times/places	Wikipedia, ESP
Bratvold (2011)	**A taxonomy based on the relationship of the worker to the work**[8]	
	• Microtasks (employ workers to perform individual [micro] tasks without regard to the entire project) • Macrotasks (workers can see more, if not all, of the project and get involved with the portion they are knowledgeable in) • Crowdfunding (classifies companies that employ crowds to provide work in the form of monetary donations) • Contests (asking a crowd for work and only providing compensation to the chosen entries)	

[7]Adapted from the CSCW (Computer Supported Cooperative Work) spatial and temporal quadrants (cf. Nichols and Twidale, 1997).

[8]Bratvold claims that "when we consider grouping crowdsourcing into the relationship of the worker to the work (rather than outputs), it becomes much easier to segment into a handful of divisions that survive very well through time."

continued from previous page

Source	Description	Examples
	Classification along six dimensions	
Quinn and Bederson ([2011](#))	Motivation (pay; altruism; enjoyment; reputation; implicit work)	Mechanical Turk (pay), helpfindjim.com (altruism), ESP (enjoyment), ICDL (reputation), reCAPTCHA (implicit work)
	Quality control (output agreement; input agreement; economic models;[9] defensive task design;[10] redundancy; statistical filtering;[11] multilevel review; automatic check; reputation system)	ESP (output agreement), TagATune (input agreement), reCAPTCHA (redundancy), Soylent (multilevel review), Foldit (automatic check), Mechanical Turk (reputation system)
	Aggregation (collection; wisdom of crowds; search; iterative improvement; genetic algorithm;[12] none)	reCAPTCHA (collection), Kasparov versus the World (wisdom of crowds), helpfindjim.com (search), MonoTrans (iterative improvement), VizWiz (none)
	Human skill (visual recognition; language understanding; basic human communication)	ESP (visual recognition), Soylent (language understanding), ChaCha (basic human communication)
	Process order (1. computer→worker→requester; 2. worker→requester→computer; 3. computer→worker→requester→computer; 4. requester→worker)	reCAPTCHA (1), ESP (2), Old Weather (3), VizWiz (4)
	Task-request cardinality (one-to-one; many-to-many; many-to-one; few-to-one)	ChaCha (1-1), ESP (m-m), helpfindjim.com (m-1), VizWiz (f-1)

---

[9]Adopt compensation schemes that elicit more good-quality contributions for less reward (e.g., "quota" systems, Mason and Watts, 2009).

[10]Design tasks so that cheating is obvious or impossible.

[11]Design workflows that filter out irrelevant/malicious contributions in some way (e.g., Ipeirotis et al., 2010).

[12]Kosorukoff (2001)

continued from previous page

Source	Description	Examples
	1) Six types of crowdsourcing in cultural heritage (matched to stages of the Digital Content Life Cycle[13]) [short definition]	
Oomen and Aroyo (2011); Oomen et al. (2014)	*Correction and transcription tasks* (Describing) [Inviting users to correct and/or transcribe outputs of digitisation processes.]	The Australian Newspapers Digitisation Project, What's on the menu?, Transcribe Bentham, What's the score at the Bodleian?, Old Weather
	Contextualization (Discovering/Using&Reusing) [Adding contextual knowledge to objects, e.g., by telling stories or writing articles/wiki pages with contextual data.]	1001 Stories Denmark, Netherlands Institute for Sound and Vision wiki
	Complementing collection (Creating) [Active pursuit of additional objects to be included in a (Web)exhibit or collection.]	UK_Soundmap, Wir Waren So Frei, Mass Observation Project
	Classification (Describing) [Gathering descriptive metadata related to objects in a collection. Social tagging is a well-known example.]	Civil War Faces, Waisda?, Emotify, Hooked, ESP
	Co-curation (Using&Reusing) [Using inspiration/expertise of non-professional curators to create (Web)exhibits.]	Click!
	Crowdfunding [Collective cooperation of people who pool their money and other resources together to support efforts initiated by others.]	Kickstarter
	2) Two clusters of motivational factors	
	Connectedness and membership	Old Weather, UK_Soundmap
	Sharing and generosity	Waisda?, Wir Waren So Frei

[13]The authors use a version of the Digital Content Life Cycle model developed by the National Library of New Zealand (http://www.digitalnz.org/make-it-digital/getting-started-with-digitisation) to better understand the potential impact of crowdsourcing in the cultural heritage domain (GLAMs: Galleries, Libraries, Archives and Museums).

continued from previous page

Source	Description	Examples
	Four types of crowdsourcing information systems based on differentiation between contributions (homogeneous / heterogeneous) and the value derived from contributions (emergent / non-emergent)[14]	
Geiger et al. (2012)	*Crowd rating*—homogeneous/emergent Participants: perspective providers, specific knowledge required for some tasks; Information: capture information of low structural complexity, provide rating question; Technology: support for automated aggregation, one contribution per participant	Amazon, eBay
	Crowd creation—heterogeneous/emergent Participants: providers of self-generated content, particular skillset or domain knowledge required; Information: capture information of potentially high structural complexity, provide general guidelines; Technology: support for (manual) integration	iStockphoto, Wikipedia
	Crowd processing—homogeneous/non-emergent Participants: workers, mostly only basic human capabilities required; Information: capture information of low structural complexity, provide specific information to be processed; Technology: support for batch processing	reCAPTCHA, VizWiz
	Crowd solving—heterogeneous/non-emergent Participants: providers of self-generated content, particular skillset or domain knowledge required; Information: capture information of potentially high structural complexity, provide (specific) evaluation criteria; Technology: support for comparative selection	Foldit, Kaggle
	Six types of crowdsourcing specifically related to music	
Gomes et al. (2012)	Music co-creation	Indaba Music
	Decision support	Queremos!, PickTheBand.com
	Crowdsourced music collection and management	Last.fm
	Marketplace	Voices.com
	Release	Last.fm, SoundCloud
	Crowdfunding	Kickstarter, Queremos!

---

[14]Homogeneous vs. heterogeneous: (valid) contributions are valued equally vs. differently according to their individual qualities; emergent vs. non-emergent: an individual contribution delivers value as part of the entirety of contributions vs. a fixed value, which is independent of other contributions

continued from previous page

Source	Description	Examples
Thuan et al. (2013)	**1) Two dimensions**—revisiting the classification by Schenk and Guittard (2011)	
	Participation mode (individual / competitive) → referring respectively to integrative vs. selective process	Wikipedia (individual), Askville (competitive), InnoCentive (competitive)
	Task complexity (simple / skilled) → here "creative" tasks are considered as skilled	Askville (simple), Wikipedia (skilled), InnoCentive (skilled)
	2) A theoretical multilayer framework for decision making	
	1. Task—(a) Internet vs. physical; (b) interactive vs. independent; (c) sensitive information vs. non-sensitive; (d) partitioned vs. non-partitioned 2. People—(a) Employee for task: few vs. large; (b) The crowd for task: available vs. must build 3. Management—(a) Budget: small vs. large; (b) Crowdsourcing expert and experience: available vs. not available; (c) Acceptance level of low quality result risk: high vs. low 4. Environment—Platform: internal (build) vs. external (available)	
Saxton et al. (2013)	**Nine crowdsourcing models based on four primary characteristics: the type of service or product being outsourced, the role of community users, the level of collaboration, and the type of managerial control (e.g., compensation, trust building, evaluation) systems employed**	
	Intermediary	InnoCentive, 99Designs, Mechanical Turk
	Citizen media production	webook.com
	Collaborative software development	Linux
	Digital goods sales	iStockphoto
	Product design	Threadless
	Peer-to-peer social financing	Lending Club
	Consumer report	Angie's list
	Knowledge base building	Wikipedia
	Collaborative science project	ESP, reCAPTCHA

continued from previous page

Source	Description	Examples
	Ten forms of crowdsourcing	
	Crowdjobbing (multiple varying routine tasks or a single and very specific task)	Mechanical Turk
	Crowdwisdom (seeking answers to questions)	Threadless, cafepress, Lulu
Lebraty and Lobre-Lebraty (2013)	Crowdfunding (financing)	Kickstarter
	Crowdsourcing and forecasting (expectation)	inkling markets
	Crowdsourcing and innovation (response to mainly technical and sometimes conceptual problems)	InnoCentive
	Crowdsourcing and authenticity (the creation of a bond of proximity with its environment)	eYeka
	Crowdauditing (data analysis aimed at finding problems or opportunities)	OpenDataParis
	Crowdcontrol (the security of people and goods)	Kaspersky
	Crowdcuration (classification of data, information, and knowledge)	Wikipedia
	Crowdcare (the health and protection of individuals)	Staying Alive
Pedersen et al. (2013)	**Conceptual model of crowdsourcing** • Problem • Process • Governance • Technology • People: problem owner, individual, crowd • Output	
	Four types of problems best suited for crowdsourcing	
Brabham (2013); Brabham et al. (2014)	*Knowledge discovery and management* Problem: information management Approach: locate and assemble information	SeeClickFix
	Distributed human intelligence tasking Problem: information management Approach: analyze existing information	Mechanical Turk
	Broadcast search Problem: ideation Approach: empirically true solution	InnoCentive

continued from previous page

Source	Description	Examples
	Peer-vetted creative production Problem: ideation Approach: a matter of taste or market support	Threadless
	A taxonomy of crowdsourcing based on three task characteristics: structure (well-structured / unstructured); inter-dependence (individuals / virtual communities); and commitment (low / high)	
Nakatsu et al. (2014)	*Contractual hiring* [well-structured × independent] Low commitment: (i) human intelligence tasks; (ii) crowdsharing marketplaces High commitment: online employment platforms	Mechanical Turk (low-i), AirBnB (low-ii), Elance (high)
	Distributed problem-solving (Additive/Pooled Coordination) [well-structured × inter-dependent] Low commitment: (i) geo-located data collection; (ii) distributed knowledge gathering; (iii) crowdfunding	Mark the Spot (low-i), CureTogether (low-ii), Kickstarter (low-iii)
	New idea generation—solo [unstructured × independent] Low commitment: consumer-driven innovation High commitment: (i) online problem-solving platforms; (ii) contests	IdeaStorm (low), InnoCentive (high-i), Cisco Systems' 2007 I-Prize (high-ii)
	Collaboration (Reciprocal Coordination) [unstructured × inter-dependent] Low commitment: real-time idea jams High commitment: (i) open source software development; (ii) open source design of hardware; (iii) open content projects	IBM's 2006 Innovation Jam (low), Linux (high-i), 40 Fires Foundation (high-ii), Wikipedia (high-iii)
Zhao and Zhu (2014)	**1) Fundamental dimensions in crowdsourcing** (adapted from Malone et al. (2010))	
	• Who—Provider (general crowd / specific group) • What—Ownership (public / private goods) • Why—Motivation & incentive (intrinsic / extrinsic) • How—Mode (collection / competition / collaboration)	
	2) Typology based on context and function	
	Context (business / non-business)	Threadless (business), Marketocracy (business), CrowdSpirit (business), Mob4Hire (non-business), Kaggle (non-business)
	Function (design & development / idea & consultant / test & evaluation / other)	Threadless (d&d), Marketocracy (i&c), CrowdSpirit (t&e), Mob4Hire (t&e), Kaggle (other)

Source	Description	Examples
	3) IS research foci	
	• Participant: motivation, behavior • Organization: crowdsourcing adoption; implementation and governance; quality and evaluation • System: incentive mechanism, technology issues	

5 A framework for crowdsourcing OMR correction tasks

Unlike small-scale public engagement activities such as tagging or rating a song, crowdsourcing requires a greater level of commitment from the contributor in terms of effort, time and intellectual input, as well as continuous assessment and reward by the system owners Holley (2010).

How to engage contributors?

According to Doan et al. (2011), one can either *require users* (e.g., students may be required to help correct transcriptions as part of the class curriculum), *make users "pay" for service* (i.e., *implicit* contribution; e.g., reCAPTCHA (von Ahn et al., 2008)), *piggyback on the user traces* (e.g., search query logs may be used for spelling corrections (Ahmad and Kondrak, 2005) or keyword generation (Fuxman et al., 2008)), or simply *ask for volunteers.* For the latter a strategy to motivate individuals to contribute is required. To this end, methods and topics from the uses and gratifications theory can provide a useful framework for understanding intrinsic (e.g., fun, challenge, community) and extrinsic motivating factors (e.g., class assignment, practical reward) (Brabham, 2013).

What type of tasks can users perform?

Many contributors may be hesitant at the prospect of re-editing an entire score, therefore it is necessary to "chunk" up error correction of an entire OMR corpus into much smaller units of work such as a single page, line or bar. In addition, different tasks may require different levels of musicological knowledge (e.g., notation systems) and thus be assigned to appropriate subgroups of contributors. For example, musicologists specializing in early music notation may perform more complex correction tasks (e.g., replacement of incorrect neumes) than volunteers with little or no relevant background (e.g., position and pitch of single notes). A two-stage strategy for crowd management is therefore needed, both prior to the assignment of the first task (using test tasks and/or questionnaires) and following the evaluation of previously assigned tasks (which may gradually increase the abilities of a contributor (Shen et al., 2003)).

How to evaluate users and their contributions?

Crowdsourcing, as with any open call to an undefined group of people, is prone to incompetent and sometimes malicious use. Continuous evaluation of users and their contributions is essential in monitoring individuals' abilities and thus assuring effective crowd management. A manual approach to this challenge includes the further distribution of contributions by ordinary users to an independent team of trusted users, whereby "bad" work is flagged on message boards (Doan et al., 2011). For example, the same task can be assigned to three contributors. If one result differs substantially from the other two, then that contributor gets flagged. The less flags a contributor has, the more trustworthy he is. Such reliability testing can also be done on the fly; a contributor can be given a correction task for which the system already knows the solution, then flagged as reliable or unreliable based on their response (Doan et al., 2011).

How to reward and retain users?

Provided that the contributions of a user are reliable, a strategy to reward and retain them should be encouraged. Such a strategy should be part of the overall crowd engagement approach (e.g., extrinsic motivation). For example, contributors may be offered exclusive beta versions of certain (possibly relevant) software or free access to digital libraries. Ideas may also be drawn from the emerging area of *gamification* (Deterding et al., 2011): contributors can earn badges that they can then redeem as points in online games; high-volume data collection and verification can be turned into a game where participants are rewarded points for their work (e.g., the ESP game (von Ahn and Dabbish, 2004); MajorMinor (Mandel and Ellis, 2008)).

Based on these considerations, Fig. 5.1 presents a framework for crowdsourcing OMR transcription corrections, adapted from the more general one proposed by Kittur et al. (2013, Fig. 2). We have identified two major types of challenges in implementing such a system (light red colour): workflow-specific (task decomposition, task assignment, contribution combination) and crowd-specific (engagement, management, evaluation, reward).
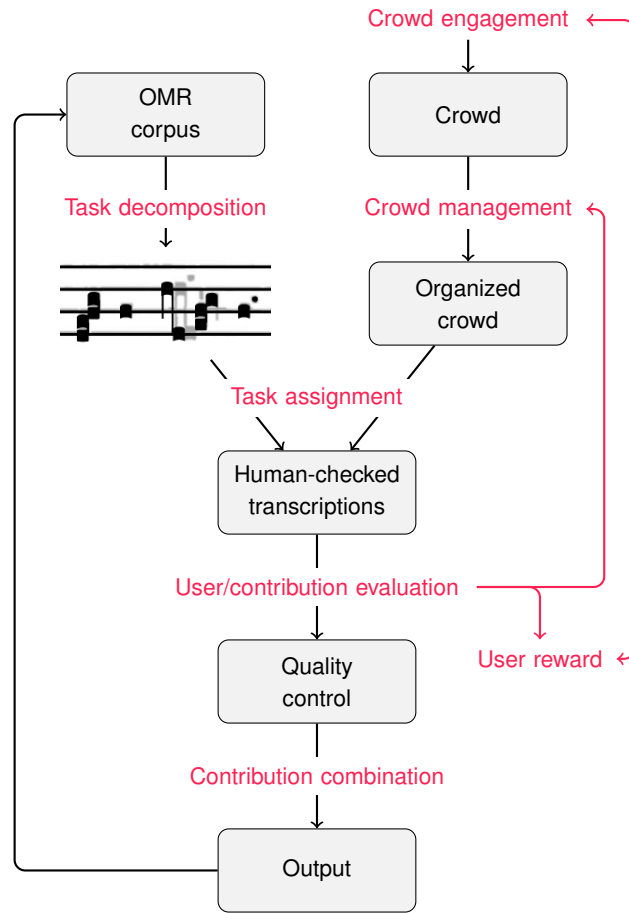
Figure 5.1: A framework for correcting large-scale OMR data with crowdsourcing. Adapted from Kittur et al. (Kittur et al., 2013, Fig. 2). Workflow- and user-specific challenges are marked in light blue.

Bibliography

Ahmad, F. and Kondrak, G. (2005). "Learning a spelling error model from search query logs", in *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language*, 955–962 (Vancouver, Canada).

Aljanaki, A., Bountouridis, D., Burgoyne, J. A., Balen, J. V., Wiering, F., Honing, H., and Veltkamp, R. (2014). "Designing games with a purpose for data collection in music research. Emotify and hooked: Two case studies", in *Revised Selected Papers from the Second International Conference in Games and Learning Alliance (GALA 2013)*, 29–40 (Paris, France).

Alter, S. (2008). "Defining information systems as work systems: implications for the IS field", European Journal of Information Systems 17, 448–469.

Bernstein, M. S., Little, G., Miller, R. C., Hartmann, B., Ackerman, M. S., Karger, D. R., Crowell, D., and Panovich, K. (2010). "Soylent: A word processor with a crowd inside", in *Proceedings of the 23rd annual ACM symposium on User Interface Softare and Technology (UIST '10)*, 313–322 (New York).

Bonney, R., Ballard, H., Jordan, R., McCallie, E., Phillips, T., Shirk, J., and Wilderman, C. C. (2009). "Public participation in scientific research: Defining the field and assessing its potential for informal science education", Technical Report, Center for Advancement of Informal Science Education (CAISE), Washington, DC.

Brabham, D. C. (2013). *Crowdsourcing* (MIT Press, Cambridge, MA).

Brabham, D. C., Ribisl, K. M., Kirchner, T. R., and Bernhardt, J. M. (2014). "Crowdsourcing applications for public health", American Journal of Preventive Medicine 46, 179–187.

Bratvold, D. (2011). "Defining crowdsourcing's taxonomy—a necessary evil", URL http://dailycrowdsource.com/content/crowdsourcing/660-crowd-leader-david-bratvold-defining-crowdsourcings-taxonomy-a-necessary-evil.

Burgoyne, J. A., Bountouridis, D., Balen, J. V., and Honing, H. (2013). "Hooked: A game for discovering what makes music catchy", in *Proceedings of the 14th International Society for Music Information Retrieval Conference*, 245–250 (Curitiba, Brazil).

Burlet, G., Porter, A., Hankinson, A., and Fujinaga, I. (2012). "Neon.js: Neume editor online", in *Proceedings of the 13th International Society for Music Information Retrieval Conference*, 121–126 (Porto, Portugal).

Chanal, V. and Caron-Fasan, M.-L. (2010). "The difficulties involved in developing models open to innovation communities: the case of a crowdsourcing platform", M@n@gement 13, 318–341.

Cooper, S., Khatib, F., Treuille, A., Barbero, J., Lee, J., Beenen, M., Leaver-Fay, A., Baker, D., and Popovic, Z. (2010). "Predicting protein structures with a multiplayer online game", Nature 466, 756–760.

Corney, J. R., Torres-Sanchez, C., Jagadeesan, A. P., and Regli, W. C. (2009). "Outsourcing labour to the cloud", International Journal of Innovation and Sustainable Development 4, 294–313.

Denyer, D., Tranfield, D., and van Aken, J. E. (2008). "Developing design propositions through research synthesis", Organization Studies 29, 393–413.

Deterding, S., Sicart, M., Nacke, L., O'Hara, K., and Dixon, D. (2011). "Gamification: Using game-design elements in non-gaming contexts", in *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, 2425–2428 (Vancouver, Canada).

Doan, A., Ramakrishnan, R., and Halevy, A. Y. (2011). "Crowdsourcing systems on the World-Wide Web", Communications of the ACM 54, 86–96.

Erickson, T. (2011). "Some thoughts on a framework for crowdsourcing", in *CHI 2011 Workshop on Crowdsourcing and Human Computation* (Vancouver, BC).

Estellés-Arolas, E. and de Guevara, F. G.-L. (2012). "Towards an integrated crowdsourcing definition", Journal of Information Science 38, 189–200.

Fuxman, A., Tsaparas, P., Achan, K., and Agrawal, R. (2008). "Using the wisdom of the crowds for keyword generation", in *Proceedings of the 17th international conference on World Wide Web*, 61–70 (Beijing, China).

Geerts, S. A. (2009). "Discovering crowdsourcing: Theory, classification and directions for use", Master's thesis, Dept. of Industrial Engineering and Innovation Sciences, Eindhoven University of Technology, Eindhoven, Netherlands.

Geiger, D., Rosemann, M., Fielt, E., and Schader, M. (2012). "Crowdsourcing information systems—Definition, typology, and design", in *Proceedings of the Thirty Third International Conference on Information Systems* (Orlando, Florida).

Geiger, D., Seedorf, S., Schulze, T., Nickerson, R., and Schader, M. (2011). "Managing the crowd: Towards a taxonomy of crowdsourcing processes", in *Proceedings of the Seventeenth Americas Conference on Information Systems*, paper 430 (Detroit, MI).

Gomes, C., Schneide, D., Moraes, K., and de Souza, J. (2012). "Crowdsourcing for music: Survey and taxonomy", in *Proceedings of the 2012 IEEE International Conference on Systems, Man, and Cybernetics*, 832–839 (Seoul, Korea).

Hankinson, A., Burgoyne, J. A., Vigliensoni, G., Porter, A., Thompson, J., Liu, W., Chiu, R., and Fujinaga, I. (2012). "Digital document image retrieval using optical music recognition", in *Proceedings of the 13th International Society for Music Information Retrieval Conference*, 577–582 (Porto, Portugal).

Holley, R. (2009). "Many hands make light work: Public collaborative OCR text correction in Australian historic newspapers", National Library of Australia Staff Papers .

Holley, R. (2010). "Crowdsourcing: How and why should libraries do it?", D-Lib Magazine 16.

Howe, J. (2006). "The rise of crowdsourcing", WIRED magazine, available at http://www.wired.com/wired/archive/14.06/crowds.html.

Howe, J. (2008). *Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business* (Crown Business, New York).

Ipeirotis, P. G., Provost, F., and Wang, J. (2010). "Quality management on Amazon Mechanical Turk", in *Proceedings of the ACM SIGKDD Workshop on Human Computation HCOMP'10*, 64–67 (Washington, DC).

Kittur, A., Nickerson, J. V., Bernstein, M. S., Gerber, E. M., Shaw, A., Zimmerman, J., Lease, M., and Horton, J. J. (2013). "The future of crowd work", in *Proceedings of the 16th ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW)*, 1301–1317 (San Antonio, TX).

Kosorukoff, A. (2001). "Human based genetic algorithm", in *Proceedings of the 2001 IEEE International Conference on Systems, Man, and Cybernetics*, volume 5, 3464–3469 (Tucson, AZ).

Law, E. L. M., von Ahn, L., Dannenberg, R. B., and Crawford, M. (2007). "TagATune: A game for music and sound annotation", in *Proceedings of the 8th International Society for Music Information Retrieval Conference*, 361–364 (Vienna, Austria).

Lebraty, J.-F. and Lobre-Lebraty, K. (2013). *Crowdsourcing: One Step Beyond* (Wiley & ISTE, London & Hoboken, NJ).

Malone, T. W., Laubacher, R., and Dellarocas, C. (2010). "The collective intelligence genome", MIT Sloan Management Review 51, 21–31.

Mandel, M. I. and Ellis, D. P. W. (2008). "A web-based game for collecting music metadata", Journal of New Music Research 37, 151–165.

Mason, W. and Watts, D. J. (2009). "Financial incentives and the "performance of crowds"", ACM SIGKDD Explorations Newsletter 11, 100–108.

Nakatsu, R. T., Grossman, E. B., and Iacovou, C. L. (2014). "A taxonomy of crowdsourcing based on task complexity", Journal of Information Science 1–12.

Nichols, D. M. and Twidale, M. B. (1997). "Computer supported cooperative work and libraries", VINE: The Journal of Information and Knowledge Management Systems 27, 10–15, available at http://www.comp.lancs.ac.uk/computing/research/cseg/projects/ariadne/docs/vine.html.

Oomen, J. and Aroyo, L. (2011). "Crowdsourcing in the cultural heritage domain: Opportunities and challenges", in *Proceedings of the 5th International Conference on Communities and Technologies*, 138–149 (Brisbane, Australia).

Oomen, J., Gligorov, R., and Hildebrand, M. (2014). "*Waisda?*: Making videos findable through crowdsourced annotations", in *Crowdsourcing our Cultural Heritage*, eds. M. Ridge (Ashgate, Farnham, UK/Burlington, VT).

Pedersen, J., Kocsis, D., Tripathi, A., Tarrell, A., Weerakoon, A., Tahmasbi, N., Xiong, J., Deng, W., Oh, O., and deVreede, G.-J. (2013). "Conceptual foundations of crowdsourcing: A review of IS research", in *Proceedings of the 46th Hawaii International Conference on System Sciences*, 579–588 (Maui, HI).

Quinn, A. J. and Bederson, B. B. (2011). "Human computation: A survey and taxonomy of a growing field", in *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI 2011)*, 1403–1412 (Vancouver, BC).

Rouse, A. C. (2010). "A preliminary taxonomy of crowdsourcing", in *Proceedings of the 21st Australasian Conference on Information Systems*, paper 76 (Brisbane, Australia).

Saxton, G. D., Oh, O., and Kishore, R. (2013). "Rules of crowdsourcing: Models, issues, and systems of control", Information Systems Management 30, 2–20.

Schenk, E. and Guittard, C. (2011). "Towards a characterization of crowdsourcing practices", Journal of Innovation Economics & Management 7, 93–107.

Shen, M., Tzeng, G.-H., and Liu, D.-R. (2003). "Multi-criteria task assignment in workflow management systems", in *Proceedings of the 36th Hawaii International Conference on System Sciences* (IEEE).

Surowiecki, J. (2004). *The Wisdom of Crowds* (Doubleday, New York).

Thuan, N. H., Antunes, P., and Johnstone, D. (2013). "Factors influencing the decision to crowdsource", in *Collaboration and Technology: Proceedings of the 19th International Conference, CRIWG 2013, Wellington, New Zealand, October 30–November 1, 2013*, eds. P. Antunes, M. A. Gerosa, A. Sylvester, J. Vassileva, and G.-J. deVreede, 110–125 (Springer, Berlin Heidelberg).

von Ahn, L. and Dabbish, L. (2004). "Labeling images with a computer game", in *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, 319–326 (Vienna, Austria).

von Ahn, L., Maurer, B., McMillen, C., Abraham, D., and Blum, M. (2008). "reCAPTCHA: Human-based character recognition via web security measures", Science 321, 1465–1468.

Vukovic, M. (2009). "Crowdsourcing for enterprises", in *Proceedings of the 2009 IEEE World Conference on Services - I*, 686–692 (Los Angeles, CA).

Wiggins, A. and Crowston, K. (2011). "From conservation to crowdsourcing: A typology of citizen science", in *Proceedings of the 44th Hawaii International Conference on System Sciences*, 1–10 (Kauai, HI).

Yuen, M.-C., King, I., and Leung, K.-S. (2011). "A survey of crowdsourcing systems", in *Proceedings of the 3rd IEEE International Conference on Privacy, Security, Risk and Trust (PASSAT) and 3rd IEEE International Conference on Social Computing (SocialCom)*, 766–773 (Boston, MA).

Zhao, Y. and Zhu, Q. (2014). "Evaluation of crowdsourcing research: Current status and future direction", Information Systems Frontiers 16, 417–434.

Zwass, V. (2010). "Co-creation: Toward a taxonomy and an integrated research perspective", International Journal of Electronic Commerce 15, 11–48.