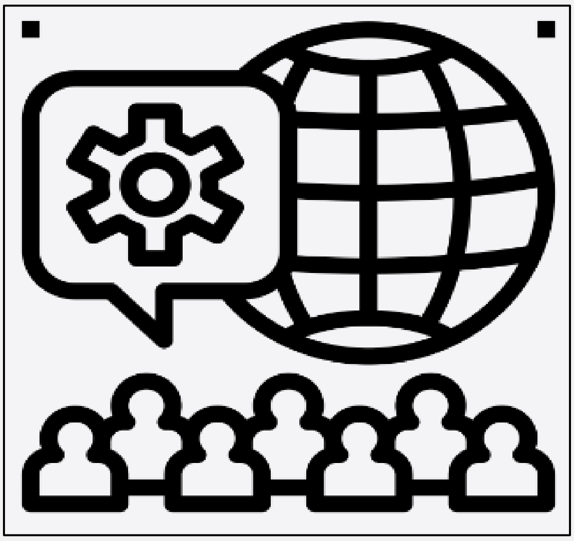
Feasibility analysis of using crowdsourcing to monitor dual quality of food in the EU single market

A review of practice and literature

Solano-Hermosilla, G.

2024



Joint Research Centre

EU Science Hub
https://joint-research-centre.ec.europa.eu/jrc

JRC133821

PDF ISBN 978-92-68-04581-7 doi:10.2760/3309 KJ-04-23-649-EN-N

Luxembourg: Publications Office of the European Union, 2024

How to cite this report: European Commission, Joint Research Centre, Solano Hermosilla, G., *Feasibility analysis of using crowdsourcing to monitor dual quality of food in the EU single market*, Publications Office of the European Union, Luxembourg, 2024, https://data.europa.eu/doi/10.2760/3309, JRC133821.

Contents

Abstract

In the context of the policy debate around business practices related to the marketing of branded food products as being identical (i.e. in their brand and appearance on the packaging) across EU Member States when, in fact, they differ significantly in composition or characteristics (i.e. in their ingredients), the European Parliament has called for a system for monitoring this issue, which is often referred to as 'dual quality'. Despite the various studies carried out by the European Commission, there currently exists no monitoring system that can be used to evaluate the presence of such practices across the EU single market. This is mainly because monitoring this practice requires readily accessible and up-to-date information on branded food products sold in supermarkets, which are generally subject to constant reformulation; in addition, new products are continually introduced to the market and older ones removed. This study aims to address this gap by assessing the feasibility of crowdsourcing (gathering citizen contributions) to collect branded food product information (i.e. photos of the front and back of pack, including information on the nutritional composition and ingredients of branded food products) to monitor dual quality cases. The analysis builds on existing practice and literature, using a data life cycle framework to identify processes and key factors for each crowdsourcing component (task, crowdsourcer, crowd, system/platform) and subcomponent (crowd management, quality assurance, incentive mechanism and technology) at each stage. The study provides insights into the challenges related to crowd participation, accuracy, representativeness and data usage. It also examines the advantages and disadvantages of crowdsourcing to monitor dual quality in the EU and makes several key recommendations for stakeholders. The study concludes that, for crowdsourcing to be a viable tool for monitoring dual quality, the crowdsourcer's value proposition must integrate the benefits and importance (e.g. access to data, knowledge, social contribution) for all participants (crowdsourcer, crowd, society) beyond any economic reward. For this to succeed, crowdsourcing must deliver a high-quality aggregated outcome and compensations that meet the value proposition (the promised benefits for all), including dual quality information. In this context, behavioural tools (e.g. nudges) and gamification (e.g. points, score tables, puzzles) can help. Finally, for crowdsourcing to function correctly, the crowdsourcer must plan and manage well, identify risks and use a valuation method that encompasses all costs and benefits to determine the value captured by the crowdsourcing organisation and its viability. The results of this study provide insights that can help Member State authorities, business and consumer representatives and other stakeholders considering implementing tools that rely on crowdsourcing for the monitoring of dual quality practices.

Authors

Gloria Solano Hermosilla, Pablo de Olavide University, Seville, Spain

Executive summary

Following various activities at the European level, including the introduction of a new provision in the unfair commercial practices directive (Directive 2005/29/EC) by the European Parliament and the Council of the European Union as part of the European Commission's 2018 New Deal for Consumers initiative to address misleading marketing practices that suggest to consumers that products sold under the same brand and in the same or similar packaging are identical across the EU when this is not the case, the European Parliament launched a preparatory action(1) aimed at 'assessing alleged differences in the quality of products sold on the Single Market'. The project builds on a common methodology developed by the Joint Research Centre, as well as on what has emerged and been learned from an EU-wide testing campaign. It focuses on extending the scope of the research to include non-food products (e.g. detergents, cosmetics, toiletries and baby products, as covered by previous pilot projects), with samples from all Member States. Furthermore, it involves an assessment of the feasibility of 'creating a permanent quality monitoring centre for products sold on the Single Market, with a view to long-term action to resolve the issue of "Dual Quality" on the Single Market.'(2) It should be noted that, while the European Commission, in the context of this feasibility study, assesses and outlines how existing information and communications technology (ICT) tools could be adapted to monitor dual quality practices in the internal market and what challenges Member State authorities, consumer and business representatives and other stakeholders might face when establishing such dual quality monitoring tools, the Commission has no competence to establish such tools at the European level. It will therefore be for Member States, consumer and business representatives and other stakeholders to implement these tools, should they see the need and wish to use them for monitoring dual quality practices in the single market.

In this context, the current study explores how timely and reliable information on branded food products can be gathered to monitor the presence of dual quality across EU Member States by building on existing ICT tools. One of the current challenges is that information on the composition (i.e. the ingredients and nutritional facts) of branded food products is not always readily available or up to date, due to the rapidity of food product formulation changes and new product introductions and removals. A potential solution is exploiting crowdsourcing (gathering citizen contributions) for data collection. This study assesses the feasibility of crowdsourcing to monitor dual quality cases in the EU – that is, the collection of information on branded food product (i.e. photos of both the front and back of pack, including information on products' nutritional composition and ingredients) from consumers visiting supermarkets. The analysis builds on existing practice and literature, using a data life cycle framework to identify processes and key factors for each crowdsourcing component (task, crowdsourcer, crowd, system/platform) and subcomponent (crowd management, quality assurance, incentive mechanisms and technology) at each stage. The study finds fundamental challenges related to crowd participation, data quality and representativeness and data usage, examines the advantages and disadvantages of crowdsourcing to monitor dual quality in the EU and provides several recommendations on design and governance mechanisms. Notably, in addition to having the right resources and skills (e.g. having a smartphone, being able to take pictures and use apps), the crowd needs to be aware of the initiative (requiring sound advertising of the initiative with a marketing campaign) and be extrinsically (e.g. through monetary rewards and information access) or intrinsically (e.g. through fun, a feeling of making a social contribution, learning skills) motivated to participate. Implementing a well-planned marketing campaign and a good mix of monetary and non-monetary incentives might have cost implications. For example, costs may considerably increase if the crowd is motivated solely by monetary rewards. In a hypothetical scenario in which the incentive offered to the consumer moves from EUR 0.01 to EUR 0.5 per photo taken, the multiplying effect on cost would be 50 and, depending on the number of photos involved, the initiative could prove unviable. Behavioural tools and gamification can help increase motivation while keeping costs under control. In addition, success in attracting a crowd is associated with the choice of information technology (IT)

(1) https://www.europarl.europa.eu/doceo/document/BUDG-DT-648406_EN.pdf.

(2) https://www.europarl.europa.eu/cmsdata/187781/budg2020-doc6-txt-2-en-original.pdf

platform, whether proprietary (and perhaps developed for a specific purpose) or commercial (generalist). Furthermore, the choice of IT platform is an economic and operational decision, and is particularly important due to the path dependencies associated with sunk costs.

Moreover, one cannot expect the contributions from the crowd to be directly usable. Adequate quality assurance and aggregating mechanisms (processing) must be applied. Most importantly, the degree of processing automation is inversely associated with cost. Therefore, current image recognition techniques and machine learning solutions, together with barcodes for product identification, should be thoroughly investigated, among other potential solutions, as tools for the conversion of images into valuable quality data, contributing to the construction of a comprehensive dataset on branded food products. In addition, the value of crowdsourcing for providing information about dual quality lies in the results available for use. Therefore, producing relevant outcomes (i.e. dual quality metrics identifying cases in which the front of packs of the different versions of the same product are very similar but the transcribed ingredient lists differ significantly) is crucial, as is gaining users' trust. The study provides several recommendations on design and governance mechanisms. It concludes that, for crowdsourcing to work as a monitoring tool, the crowdsourcer's value proposition must integrate the benefits and importance (e.g. access to data, knowledge, social contribution) for all participants (crowdsourcer, crowd, society) beyond any economic reward and deliver a high-quality aggregated outcome, compensations and processes that meet the value proposition (the promised benefits for all). Finally, the crowdsourcer must plan and manage well, identify risks, implement adequate governance mechanisms and use a valuation method that encompasses all costs and benefits to determine the value captured by the crowdsourcing organisation through crowdsourcing and its viability. The results of this study offer important insights to policymakers, business managers and other stakeholders aiming to use crowdsourcing and modern ICT solutions to monitor the occurrence of dual quality products in the EU single market.

1. Introduction

1.1. Background

A policy issue that has gained attention in recent years is the marketing of branded products as being identical (i.e. they are sold under the same brand and in the same or similar packaging) across different Member States when, in fact, they differ significantly in composition or characteristics (European Commission, 2017). Although differences in composition or characteristics do not necessarily lead to differences in quality ([3]), this issue has commonly been referred to as 'dual quality'. Interventions by the European Parliament ([4]) and the Council of the European Union ([5]) have stressed the importance of tackling the issue of dual quality (DQ) at the European level. As a result, in September 2017, the European Commission issued specific guidelines on applying EU food and consumer protection law to this issue. In November 2019, the European Parliament and the Council adopted the better enforcement and modernisation directive (Directive (EU) 2019/2161) as part of the Commission's 2018 New Deal for Consumers initiative. That directive amends the unfair commercial practices directive (Directive 2005/29/EC) by introducing a new specific provision on the issue of DQ (Article 6(2)(c)). Member States had to transpose this directive into national law by 28 November 2021 and apply it from 28 May 2022. Under the new Article 6(2)(c) of Directive 2005/29/EC, the marketing of a product with a significantly different composition or characteristics as being identical (i.e. selling it under the same brand and with the same or similar packaging) to a product marketed in another Member State can amount to a misleading practice. The competent authorities of the Member States need to assess, on a case-by-case basis, whether such DQ practices are misleading, while taking into account the impact of the practice on consumers' transactional (purchase) decisions as well as possible legitimate and objective factors that may justify differences in the composition of the same product in different Member States ([6]) (European Parliament and Council, 2019). For this assessment, it is crucial to monitor and identify the possible occurrence of DQ practices in the EU market, which requires considerable efforts in data collection and processing, requiring personnel, information technology (IT) and data coding.

Tests have been conducted by several Member States (see, for example, Council of the European Union, 2017; Croatian Food Agency, 2017; European Parliament, 2017; MPSR, 2017; Néhib, 2017) and by the Joint Research Centre at the EU level (European Commission, 2019; Nes et al., 2023); these confirmed the occurrence of DQ practices to some extent across the single market. This previous work in the DQ area was based on coverage of a limited number of food products and on the researchers' manual evaluation of the presence of DQ. In the past decade, the rapid development of digital technologies (e.g. internet, mobile devices, social media) and innovative alternative data collection methods ([7]) have offered the potential to provide valuable information for policymaking hitherto not available (Dutil, 2015; Taeihagh, 2017). Furthermore, advances in food image capturing and text recognition technologies have provided alternative means of data collection (Chen et al., 2021; Martin et al., 2008, 2014). As a result, branded food databases are becoming very valuable for nutrition research, policymaking, businesses and the general population (Pravst et al., 2021). Possible sources of images of the front and back of food product packs include private databanks, websites, social media and crowdsourcing (i.e. relying on internet/mobile apps and voluntary engagement of

---

([3]) Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning unfair business-to-consumer commercial practices in the internal market (OJ L 149, 11.6.2005, p. 22).

([4]) https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52013IP0239&from=EN.

([5]) http://data.consilium.europa.eu/doc/document/ST-8754-2016-INIT/en/pdf.

([6]) Brand owners are allowed to adapt the composition of their goods for different markets when it is justified by objective factors such as requirements under national law, availability or seasonality of raw materials, or voluntary strategies to improve access to healthy and nutritious food. In such cases, companies still need to inform consumers about the different composition of goods being offered in different markets through other means, such as advertising and product websites.

([7]) The study uses the term 'alternative data collection methods' (as opposed to traditional approaches) to refer to those innovative approaches that use data sources such as sensor inputs, web traffic, mobile devices, satellites, public records, social media and websites such as news sites that were not set up for statistical purposes and thus do not follow a statistical design (Beręsewicz et al., 2018).

citizens and stakeholders in supplying pictures) (Harrington et al., 2021; Pravst et al., 2021). Web-scraping techniques can automate searches for photos of the front and back of pack of food products posted online. However, its success is partially dependent on what companies or citizens happen to upload (e.g. in terms of variety, frequency, location and timeliness) without the influence of the data collector (unsolicited information).

In contrast, crowdsourcing involves soliciting a specific task (e.g. photographing the front-of-pack design and back-of-pack information of branded food products on the market, i.e. in supermarkets) from an undefined group of volunteer citizens (the crowd) with whom the organiser (the crowdsourcer) engages within a system (an online IT platform and incentives) (Y. Wang et al., 2017; Zhao and Zhu, 2014). This can represent a valuable and manageable information source in terms of volume, time and cost efficiency, scale, spatial representativeness and timeliness for political, business or social action (Buettner, 2015; Nassar and Karray, 2019). Furthermore, the increasing demand for data has triggered the growth of crowdsourcing platforms (e.g. Mechanical Turk (MTurk) and Appen), which can be used to solicit data for various applications (Lian et al., 2021). However, while in traditional resource management approaches businesses and organisations know their resources and control task allocation, in crowdsourcing participants are usually unknown and select themselves to perform the tasks (Cabanillas, 2016). Therefore, crowdsourcing also involves challenges such as developing effective incentive mechanisms to motivate the crowd, dealing with noise and biases, quality assurance, identifying effective processing techniques and algorithms for reliable aggregation of data into relevant information, and managing trade-offs between cost and quality, privacy concerns, the ethics of data collection and use, and uncertainty about future crowd participation and data availability (Buettner, 2015; Ding and Zhou, 2018; Liu et al., 2021; Nassar and Karray, 2019; Zhao and Zhu, 2014). Therefore, crowdsourcing initiatives often struggle to turn their promising projects into sustainable platforms (Kohler and Chesbrough, 2019). Using organisation theory, some authors argue that the main challenges in crowdsourcing are motivation and coordination (Buettner, 2015). Particularly acute are the motivational problems related to retaining a crowd to that uses a mobile app; this is more of an issue than engaging them to do so in the first place, considering that users abandon most of the apps that they download within a month (Gu et al., 2022). The motivation (and behaviour) problem can be addressed through adequate incentive mechanisms, which needs to be done while controlling costs (Nassar and Karray, 2019). The coordination problem can be tackled by establishing adequate organisational structures and processes (Buettner, 2015; Gu et al., 2022; Nassar and Karray, 2019). Some authors highlight the need for processes for efficiently verifying and aggregating multiple crowd contributions into a solution (Afuah and Tucci, 2012; Nassar and Karray, 2019).

1.2. Objective

This study aims to examine, based on a review of existing theoretical and empirical studies, the feasibility of using smartphone-captured and crowdsourced branded food product images to monitor DQ. Photographing the front- and back-of-pack information of branded food products and sharing photos is now commonplace and socially acceptable; therefore, despite the challenges, crowdsourcing may offer a practical strategy for obtaining a comprehensive dataset of branded food product images to assess potential DQ cases, with minimal effort required of participants. Furthermore, image recognition techniques combined with crowdsourcing could help to increase the number of foods that can be monitored and compared across countries (Kawano and Yanai, 2014). Nevertheless, the requirements, cost, challenges and effective governance mechanisms need attention.

We review the concept of crowdsourcing, focusing mainly on findings from the academic literature and some ‘grey’ literature. In doing so, we review research published in journals, conference papers and working papers on economics, strategic management, development and information systems. Over the past decade, research on and the practice of crowdsourcing have grown considerably (Modaresnezhad et al., 2020). However, it is only more recently that top-tier journals have been publishing research on crowdsourcing (Hossain and Kauranen, 2015). We review existing approaches to tracking food standards and food safety using citizen contributions. Furthermore, in this review we focus on the advantages and disadvantages of crowdsourcing to collect images of the front- and back-of-pack information of branded food products to track DQ in the EU and the factors and mechanisms that could make it work. We use a system view to examine the research issues from the perspective of both the crowdsourcing process and the crowdsourcing components. From the process perspective, we use a framework based on the data life cycle steps: task definition, data collection, data processing, data analytics, dissemination and usage (Dahlander et al., 2019; Matheus et al., 2018; Roth and Luczak-Roesch, 2020). Each crowdsourcing step is examined in relation to the crowdsourcing components – the task, the crowd, the organiser and the system (Karachiwalla and Pinkow, 2021; Zhao and Zhu, 2014) – considering managerial, behavioural (i.e. motivational), quality, incentive mechanism and technology aspects. Many existing crowdsourcing studies focus primarily on single or specific design elements, not developing an integrated picture of crowdsourcing; the literature lacks comprehensive guidelines for practitioners who want to initiate and manage crowdsourcing (Karachiwalla and Pinkow, 2021). Problems with motivating participants are a common concern, but lack of quality and accuracy of data collected through crowdsourcing can also lead to disappointing results and untapped potential, and we address this issue too (Hosseini et al., 2019).

The study is structured as follows. Chapter 2 explains the concept of crowdsourcing and its key aspects; reviews the benefits, costs and challenges; presents the system view used to analyse data crowdsourcing based on the data life cycle; and compares crowdsourcing with traditional data collection methods. Chapter 3 describes current approaches to tracking food quality and safety through user contributions. Chapter 4 presents a possible method of applying crowdsourcing to monitoring DQ in the EU, and Chapter 5 discusses its advantages and drawbacks. Finally, Chapter 6 presents the conclusions.

2. What is crowdsourcing?

2.1. Definition and key aspects

Crowdsourcing is a virtual sourcing method for obtaining information or a solution to a specific problem by distributing an online task to a pool of people (the crowd), leveraging the crowd's wisdom (Brabham, 2013; Surowiecki, 2004). In 2006, Jeff Howe (2006) coined the term as a combination of 'crowd' (people) and 'outsourcing' (externalisation of activities), and offered the following definition, where the main difference between crowdsourcing and outsourcing is that there is not an *ex ante* contract between the organiser and the contributor (Afuah and Tucci, 2012):

> Simply defined, 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.

The idea of crowdsourcing is not new; it has existed for centuries, with many examples of applications – for example, in 1884, the crowd corrected and updated the catalogue of the Oxford English Dictionary (Bhatti et al., 2020). Furthermore, the idea of governments relying on citizens for services or data is not new (Dutil, 2015). However, with the development of new information and communications technology (ICT), crowdsourcing has received a great deal of attention for its increased potential as a cost-efficient and operationally effective method for a variety of tasks, such as generating ideas, designing products, problem-solving, creating content, providing opinions and collecting information (Blohm et al., 2013, 2018; Daniel et al., 2018). Furthermore, organisations can benefit from crowdsourcing methods to go beyond their existing resources to obtain new data, knowledge and capabilities (Buettner, 2015; Nevo and Kotlarsky, 2020). Notably, in the 'big data' era, businesses, government policies and others rely heavily on data for decision-making and innovations. As a result, several more or less generic crowdsourcing platforms – e.g. MTurk ([8]), Appen ([9]), Wazoku ([10]) (formerly InnoCentive) – and mobile applications have emerged. Given the variety of tasks that can be solved with crowdsourcing and the different types of crowdsourcing that exist, Estellés-Arolas and González-Ladrón-De-Guevara (2012) propose a detailed and comprehensive definition that encompasses them all:

> a type of participative online activity in which an individual, an institution, a non-profit organisation, 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 … always entails mutual benefit.

Accordingly, in crowdsourcing, there is a crowdsourcer or requester – the person or organisation that launches a call to outsource a task to the public (the crowd) to achieve a particular goal. This goal translates into specific tasks that the crowd is invited to undertake through an online platform, serving the exchange between the crowdsourcer and crowd member. Crowd contributions are quality assessed and aggregated into results that can be used by the crowdsourcer and disseminated to the crowd and other stakeholders for their use. Crowd members' participation will depend on their motivation and the incentive mechanisms put in place by the crowdsourcer (Simperl, 2015; Zhao and Zhu, 2014).

A key characteristic of crowdsourcing is that, without an *ex ante* contract, potential crowd contributors need to be motivated enough to select themselves to participate (Afuah and Tucci, 2012). Motivation refers to the activation of human behaviour and the way this behaviour is sustained towards reaching the desired goal (Van Eerde, 2015). In other words, motivation describes why a person does something. Different types of incentives or a mix of them are needed to motivate participation

([8]) https://www.mturk.com/.

([9]) https://appen.com/.

([10]) https://www.wazoku.com/.

extrinsically or intrinsically. For example, economic, reputational or informational rewards are associated with extrinsic motivation, whereas fun, social contribution and improving skills are related to intrinsic motivation (Kaufmann et al., 2011; Pedersen et al., 2013). The literature is not conclusive as to whether extrinsic or intrinsic incentives are more effective. However, some authors argue that intrinsic incentives may have a more significant positive effect on quality than extrinsic ones. On the other hand, a financial incentive may speed up the attraction of contributors but may not affect the quality (Nassar and Karray, 2019; Solano-Hermosilla et al., 2022).

In contrast, Sun et al. (2015) suggest that higher incentives may encourage more significant crowd effort, resulting in higher quality. However, it is unclear whether financial incentives are always needed; it may depend on the task. For example, platforms such as Amazon's MTurk use monetary incentives to encourage people to solve tasks that humans can more effectively and efficiently solve than machines. Other initiatives provide a mix of monetary and non-monetary incentives. However, recently, Blohm et al. (2018) stressed that when the crowdsourcing task is about pooling information together, the interest in the outcome, be it extrinsic (usefulness of accessing the information) or intrinsic (altruism or social contribution), could be more relevant than the payment. Moreover, researchers stress that attracting is easier than retaining participation in crowdsourcing (Geiger et al., 2011) and that, without incentives, participants may drop out or submit low-quality information (Nassar and Karray, 2019). Therefore, it is vital to align the crowd's motivations with the incentives offered (Pedersen et al., 2013).

The literature on crowdsourcing identifies several approaches that may require different management and governance mechanisms. First, depending on the complexity of the task, it distinguishes between simple and complex tasks. Simple tasks, also known as micro-tasks, can be performed in short amounts of time by individual crowd members and do not require cognitive efforts or specific expertise. Examples of micro-tasks are tagging/labelling images, transcribing audio, and classification, rating and ranking, verification and validation, collection of data collection (be it numerical, or images or videos). Micro-tasks are usually rewarded with micro-payments (Bhatti et al., 2020; Blohm et al., 2018). However, when the purpose of micro-tasks is pooling information together or contributing to a social good, the crowd may be interested in the resulting information or outcome, and the payment may be less relevant, such as in crowdsourcing for crisis management or citizen science (Blohm et al., 2018). Citizen science projects that are entirely mediated by ICT are often considered a form of crowdsourcing applied to science (Wiggins and Crowston, 2011). Complex tasks may be decomposable into simpler tasks; if not, they are macro-tasks or creative tasks (Bhatti et al., 2020; Blohm et al., 2018). Complex tasks usually require specific skills and knowledge and computational efforts, and are paid higher than micro-tasks to attract solvers. Examples are writing, proofreading, and solving and creating solutions for complex problems.

Depending on the skills and capabilities needed for the crowdsourcing task, the task can be launched as (1) an open call to the general public, (2) an open call to a limited crowd with the required skills or (3) a mixed call in which an open call to the general public is made, but the crowdsourcer controls for specific skills (Estellés-Arolas and González-Ladrón-De-Guevara, 2012). Open call is a term used to describe how to broadcast an initiative to the crowd. However, it is worth noting that 'open' does not imply that the broadcast to the crowd cannot be addressed to a particular audience (Cullina et al., 2015). Moreover, depending on the type of solution, the literature distinguishes between integrative tasks (the contributions are complementary, and the value relies on their integration) and selective tasks (the contributions are competitive, and only one delivers the optimal solution). Researchers also refer to iterative tasks when the contribution depends on previous or affects subsequent contributions (Bhatti et al., 2020; Blohm et al., 2018; Karachiwalla and Pinkow, 2021).

Moreover, the crowdsourcing process includes verification and quality assurance of the process and results. Quality control approaches can be established for crowd participant selection, task design and collected data (Nassar and Karray, 2019). The collected data must undergo a preprocessing (e.g. transforming and filtering) and aggregation phase to obtain a quality solution. In crowdsourcing, aggregation refers to finding a solution (mining the hidden ground truth) from the crowd's answers (Barbier et al., 2012; Blohm et al., 2018; Geiger et al., 2011; Nassar and Karray, 2019). Preprocessing is the phase during which crowd contributions are transformed into suitable formats for analysis and cleaned so that various data-mining and machine learning algorithms can be applied to them (Barbier et al., 2012).

More importantly, in crowdsourcing, the type of task (i.e. integrative or selective) determines the aggregation method for establishing the solution (Blohm et al., 2018; Kamoun et al., 2015). Most crowdsourcing initiatives rely on redundancy – that is, assigning the same task to multiple crowd members to verify, quality control and aggregate their contributions (Gadiraju et al., 2019; Lian et al., 2021), thus leveraging the crowd's wisdom (Surowiecki, 2004). A typical approach to detecting and discarding low-quality observations while relying on redundancy is comparing each contribution to other contributions asked for in the same task and applying majority consensus (or majority voting or decision) or answer agreement, bearing in mind that the more contributions there are the more expensive the process is, leading to cost–quality trade-offs (Hirth et al., 2013; Nassar and Karray, 2019). Following this method, observations that do not agree with the majority in certain aspects are discarded. Another approach to detecting low quality is comparing contributions with a control group or a gold standard, such as authoritative data (Gadiraju et al., 2019; Nassar and Karray, 2019). However, the drawback here is often the cost of a control group or the lack of a gold standard to compare with. Importantly, Hirth et al.'s (2013) cost analysis revealed that the majority decision approach is more suitable for low-paid, routine, simple tasks, whereas the control group approach performs better for complex, high-priced tasks. Furthermore, the crowdsourcing system can include an additional crowdsourcing task where crowd members can verify and validate peers' contributions (Blohm et al., 2018) or the hiring of external experts to check the quality independently (Chen et al., 2021). For instance, Harrington et al. (2021) used two crowd participants to classify food pictures, and their agreement was assessed before aggregating the data. Another important aspect is whether the verification and quality control process is manual or automated. Researchers suggest the importance of automation to ensure the effectiveness and cost efficiency of the process. Accordingly, they suggest that majority voting processes can be easily implemented and automated, for example with algorithms that compare the different contributions and discard those that deviate from the majority on specific parameters (Arbia et al., 2018, 2023; Hirth et al., 2013; Nassar and Karray, 2019).

Concerning aggregation, for integrative tasks, a common aggregation mechanism is averaging (e.g. averaging prices for the same product variety, geographical location and time), whereas for selective tasks, majority voting (the solution mostly voted for is the one chosen) is a commonly used method. Moreover, again, the aggregation method's degree of automation is crucial to crowdsourcing's success. The literature proposes algorithms that can use rules and strategies, such as the assessment of specific parameters, to choose the most appropriate solution or, in the case of equality in the parameters, to choose randomly (Nassar and Karray, 2019). These algorithms can also take into account the reliability of the participants. Moreover, the literature suggests two alternatives of aggregation from which to choose: non-iterative (using the rules to find a single solution) and iterative (implying several rounds of probability estimation) (Nassar and Karray, 2019). Again, it will depend on the task whether iterative or non-iterative aggregation is most appropriate. After data are cleaned, Barbier et al. (2012) stress that machine learning or data-mining techniques can be applied to extract relevant information. Accordingly, they refer to three types of machine learning techniques: classification (e.g. support vector machine, Bayesian approaches and regression methods), clustering (e.g. density-based and spectral-based clustering techniques) and semi-supervised learning (e.g. expectation maximisation).

Finally, regarding crowdsourcing uses, there are many examples of applications giving people and decision-makers better information and insight into events that impact communities and society (Barbier et al., 2012; Bhatti et al., 2020). Wikipedia ([11]) is an example of a widely used result of crowdsourcing. Crowdsourced data with the addition of a spatial reference, often referred to in the literature as 'volunteered geographic information', is another example (Goodchild, 2007; Goodchild and Li, 2012; Senaratne et al., 2017). In policy intervention, several studies highlight the importance

([11]) https://wikipedia.org/.

of the real-time component of crowdsourcing for disaster management (Poblet et al., 2018) or monitoring food price development (Adewopo et al., 2021; Zeug et al., 2017). For example, Ushahidi ([12]) and Sahana ([13]) are well-known crisis management platforms (Barbier et al., 2012). The use of crowdsourcing for climate and atmospheric sciences is also well documented (Muller et al., 2015). Other studies stress the potential of using crowdsourcing in combination with official statistics (Sternberg and Lantz, 2018). Finally, and relevant to the topic of this study, are several initiatives that use crowdsourcing, inspired by the remote food photography method (Martin et al., 2008), to collect food pictures to assess food intake, nutrient content or ingredient composition (Harrington et al., 2021; Pravst et al., 2021). Importantly, when people, including decision-makers and policymakers, use crowdsourcing results, this can provide meaningful feedback to the crowdsourcing process.

2.2. Benefits, costs and challenges, and how to manage them

Crowdsourcing is an innovative approach that can be used for tasks where, for example, a vast amount of data needs to be collected and thus can benefit from a large group of people completing the task. Benefits may be related to cost savings (through saving on human resources, IT, energy and time), speed (real-time data are collected), flexibility, scalability, diversity, and citizen participation and interaction (Liu et al., 2021; Pravst et al., 2021). Importantly, these benefits can be achieved only if the task is suitable for crowdsourcing. For example, Nassar and Karray (2019) suggest that crowdsourcing is appropriate for tasks that need human intelligence rather than machine intelligence or for tasks that a crowd can perform with higher time and cost efficiency than employed experts. In the same vain, Afuah and Tucci (2012) specified several conditions for crowdsourcing to work effectively: (1) that the task is simple, modular and easily transferable, bearing in mind that any task above simple is risky for crowdsourcing (Liu et al., 2021); (2) that a vast number of contributions are needed, which involve distant search (as opposed to local search), since the knowledge or information required to solve the problem falls outside the focal agent's knowledge neighbourhood; (3) that there is a potential crowd motivated to contribute; (4) that contributions can be efficiently verified and aggregated into a solution; and (5) that there are IT tools suitable for the task (Afuah and Tucci, 2012; Blohm et al., 2018; Karachiwalla and Pinkow, 2021). We will use these criteria later to assess the suitability of the task that is the focus of this study. However, even if these criteria are met, crowdsourcing also entails costs and challenges that require appropriate management mechanisms (Blohm et al., 2018).

For example, crowdsourcing for data collection entails set-up costs (i.e. for IT, marketing, personnel) and running costs (i.e. for incentives, marketing, personnel, IT, materials), like any other data collection system. In crowdsourcing, it is argued, the runtime data collection costs may be cheaper than hiring experts, but data collection may also be challenging given the cost–quality trade-off (the need to minimise costs while maximising quality) (Ding and Zhou, 2018; Hirth et al., 2013; Kawano and Yanai, 2014; Nassar and Karray, 2019). For example, incentives may have to be increased to get enough valid contributions to provide information on the issue that the crowdsourcing was started to address, implying higher costs. Moreover, higher data quality in crowdsourcing is linked to redundancy and therefore to a higher number of contributions; this enables more efficient quality control processes, but also entails higher costs. Moreover, if quality control and data aggregation are too time-consuming and resource-intensive, this can outweigh the benefits of crowdsourcing (Simperl, 2015). Therefore, in crowdsourcing it is crucial to obtain enough valid contributions to infer the answer at a reasonable cost; an unlimited budget would make it possible to hire experts or large pools of workers, so that quality could be established a priori, so in crowdsourcing costs need to be controlled (Singh et al., 2021). The literature identifies several areas to consider in managing the crowdsourcing process and its challenges, which are described below. They are crowd management, incentive mechanism, crowdsourcing task, quality assurance and aggregation, technology and information use (Blohm et al., 2018; Karachiwalla and Pinkow, 2021; Nassar and Karray, 2019; Pedersen et al., 2013). In

([12]) https://www.ushahidi.com/.

([13]) https://sahanafoundation.org/.

addition, some authors also refer to ethical challenges regarding the use of crowdsourcing, which we also consider (De Stefano, 2016; Standing and Standing, 2018).

2.2.1. Crowd management

In crowdsourcing, in the absence of a contract, the organiser must motivate enough potential crowd contributors to select themselves to participate and provide quality inputs (Afuah and Tucci, 2012). Therefore, a key challenge in crowdsourcing is attracting and managing the crowd to ensure and sustained participation. A real risk is that the crowd will not sustain participation. This is exemplified very well by Gu et al.'s (2022) study, showing that 95 % of downloaded mobile apps are no longer used within 1 month. Attracting users and sustaining participation requires understanding crowd motivations and aligning incentives. Similarly, managers must understand employees' motivations and set appropriate incentives (salaries and other benefits). Therefore, several authors use a human resources (HR) management perspective to study crowdsourcing (Buettner, 2015). For example, researchers and employers have used surveys for many years to address the challenge of employee motivation (Wiley, 1997). In a similar way, to assess motivation to participate and stay engaged in crowdsourcing, a small registration and end-of-contribution survey could be helpful (Harrington et al., 2021). However, it is also worth noting conflicting ideas on motivation survey results. For example, while HR professionals often believe that employees over-report the importance of pay, research finds the opposite (Rynes et al., 2004).

Of course, the crowd must first be aware of the initiative to be motivated to take part. Crowdsourcing starts by effectively launching and broadcasting the task to the potential crowd, reaching the public and communicating the value of and rewards for the task to raise awareness and motivate the public (Dahlander et al., 2019; Pravst et al., 2021). It must first be decided whether to conduct an open call to all possible participants or invite a specific group. Here, the competencies and skills that the task requires should guide the answer (Dahlander et al., 2019). Moreover, the challenge of promoting crowdsourcing initiatives is finding the right channels to target the crowd appropriately, such as websites, social media, blogs, media (press, TV, radio) and events, while keeping costs down and taking into account the local context. A quick registration and end-of-contribution survey can also be helpful for this (Harrington et al., 2021), for example to find out through which means most participants learned about the initiative, to design future campaigns (Solano-Hermosilla et al., 2020).

Furthermore, another decision is whether to use intermediaries for dissemination and what kind of intermediaries (Dahlander et al., 2019). Diffusion can be carried out by hiring marketing specialists in project dissemination or by networking and partnering. A downside if an intermediary is responsible for the broadcasting is that the participants may be less motivated to contribute to the crowdsourcing initiative because of the feeling of affiliation (Dahlander et al., 2019). Nevertheless, networking and partnering with other organisations may be helpful, as well as word of mouth, to communicate about the project and, at the same time, cost-saving.

In crowdsourcing, it is essential to communicate the value to the crowd to increase motivation, explicitly making a value proposition that includes the benefits for the organiser, the crowd and society (Aluchna, 2018; Fedorenko et al., 2017; Fung, 2015). Crowd motivations can be extrinsic, related for example to money, reputation or information, or intrinsic, related to entertainment, fun, curiosity or altruism (Karachiwalla and Pinkow, 2021; Kaufmann et al., 2011; Malone, 1981; Nassar and Karray, 2019; Pedersen et al., 2013). Moreover, Maslow (1958) proposed an order of human needs that he linked to the motivation of individuals, starting with basic needs as the most important, followed by psychological needs and, finally, self-esteem needs.

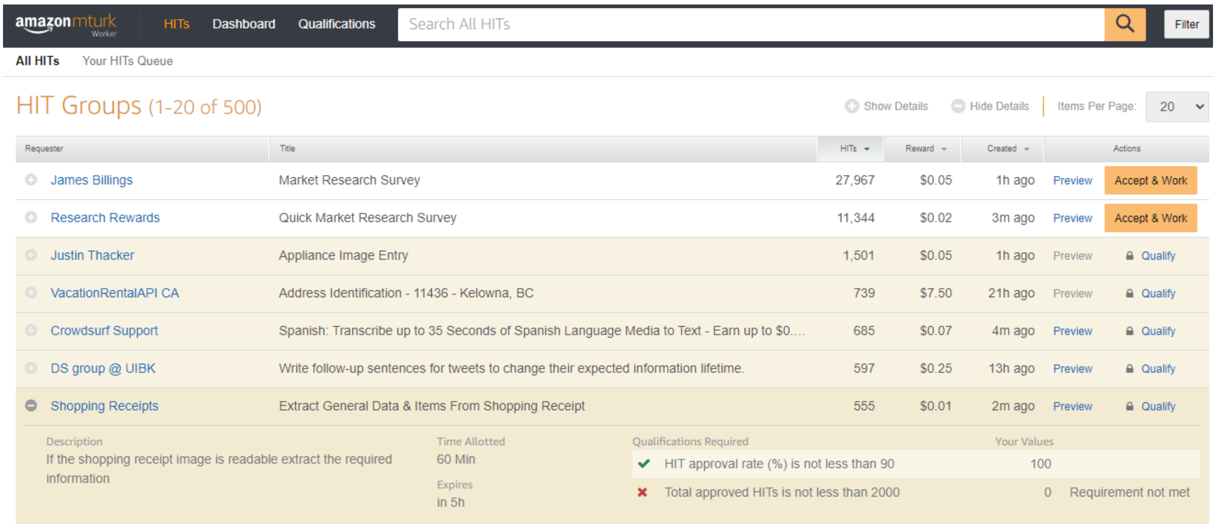
2.2.2. Incentive mechanism

It is fundamental to understand what the potential crowd motivations are and use the right incentive mixes to address them. In addition, some studies point to crowd segmentation to ensure adequate targeting of participants (Fedorenko et al., 2017). Particularly relevant is the finding that intrinsic incentives may help to save costs. Behavioural science provides an additional tool to help sustain crowd contributions by activating behavioural factors that strengthen the engagement of individuals with the crowdsourcing platform while controlling the cost. These tools, for example in the form of informational nudges (e.g. sharing social norms, indicating how a group of peer contributors is performing), can be included in the design of crowdsourcing mechanisms to maximise the number of contributions (Solano-Hermosilla et al., 2022). Game elements can also help to increase or sustain participation (Gu et al., 2022; Tinati et al., 2017). Examples of game elements are points awarded for the accurate completion of tasks, leader board displays, a real-time ticker of users online, puzzle pieces, competitions and a real-time chat (Tinati et al., 2017). In order to improve the design of crowdsourcing initiatives, it is essential to research which motivational tools work best. In the literature, we found several studies using randomised controlled trials to assess motivation tools in crowdsourcing initiatives (Gu et al., 2022; Solano-Hermosilla et al., 2022). Moreover, Tinati et al. (2017) surveyed participants in a citizen science project that included game elements. They concluded that participants are motivated mainly by the contribution to science, followed by learning and personal interests, then by gaming and entertainment, and only a few by the feeling of being part of a community. Particularly important is the finding that in simple co-creating tasks, such as building information, which the crowd can access and understand the co-created value of, the monetary payment may become less critical (Blohm et al., 2018; Fedorenko et al., 2017; Nassar and Karray, 2019; Tinati et al., 2017). It is therefore crucial to understand how jointly created value (e.g. results, indicators, publications, websites, web dashboards) can motivate the crowd to make an appropriate value proposition.

Nevertheless, incentives may be an important cost factor in crowdsourcing. Simple, routine tasks are usually compensated with micro-payments (Bhatti et al., 2020; Blohm et al., 2018). To get an idea of the level of micro-payments, we list below some examples of crowdsourcing tasks and the incentive paid, as published in crowdsourcing platforms and research articles. Specifically, the crowdsourcing platform MTurk, often used by researchers for obtaining data (Zhou et al., 2018), offers a list of 'human intelligence tasks' (HITs) that crowd members (MTurk workers) can complete for minimal payments, as seen in Figure 1 (Felstiner, 2011). For instance, they can earn from EUR 0.01 for 'Extract General Data and Items from Shopping Receipt' to EUR 0.02 for a 'Quick Market Research Survey' and EUR 7.50 for an address identification task. If a HIT takes 1 minute and pays out EUR 0.02, it is equivalent to a EUR 1.20 per hour pay rate; if it takes 10 minutes and pays out EUR 7.50, then EUR 45 per hour. For example, the Food Price Crowdsourcing Africa ([14]) platform in Nigeria paid EUR 4 for a task that, on average, took about 7 minutes to complete (contingent on being one of the first 30 contributions and up to 2 contributions per week). In a recent study, to document eating behaviours from food pictures, project participants took and uploaded pictures to MTurk, where MTurk workers were to classify the images into food categories. Each HIT, a picture within a batch of 100, was estimated to take 15 seconds to complete, providing a EUR 0.05 reward, equivalent to a EUR 12 per hour rate (Harrington et al., 2021). Another recent study looked to assess nutritional information from crowdsourced restaurant food pictures entered into a deep-learning model. However, it does not report on the existence of a reward (Chen et al., 2021); presumably participants were motivated by their own interest in the project. Furthermore, the study of Kawano and Yanai (2014) assessing the automatic expansion of a food image dataset paid EUR 0.03 for one HIT consisting of excluding food pictures that were not relevant and EUR 0.05 for one HIT consisting of drawing bounding boxes on the selected food images.

([14]) https://sites.google.com/prod/view/foodprice/home/crowdsourcing.

Figure 1. Task overview on the Amazon MTurk platform



Source: Screenshot from MTurk platform (www.mturk.com/)

It is important that the crowdsourcer understands the effort involved in the task to align it with the incentives (Kittur et al., 2011). The effort can be measured and monitored by the time participants need to conduct the task (Cullina et al., 2015; Ford et al., 2015). Simple tasks requiring minimal effort take only seconds to a few minutes to complete (Simperl, 2015). In the example above, contributions to the Food Price Crowdsourcing Africa platform (submitting a minimum of four food product prices) took a time ranging from 4 minutes (the fastest) to 10–12 minutes (those who took more time) (Solano-Hermosilla et al., 2020). All this suggests a significant trade-off between the number of contributions and the incentive cost.

Moreover, to keep the crowd motivated, the crowdsourcer must have in place channels for communication with and feedback to the crowd at different stages of the process. Researchers distinguish three types of communication in crowdsourcing initiatives: unidirectional (suggestion boxes), bidirectional (email) and multidirectional (forums, wikis) (Karachiwalla and Pinkow, 2021; Schäfer et al., 2017). According to researchers, unidirectional and multidirectional communication are needed more before data collection, whereas bidirectional communication is needed more during and after data collection. Most importantly, through feedback and communication, contributors see that their contributions are important to the crowdsourcing organisation, and this can increase their motivation (Blohm et al., 2018). The crowdsourcer should remember to plan time and resources for a support service to communicate with and answer any questions from crowd members.

2.2.3. Crowdsourcing task

Other factors that influence motivation are task complexity (simple tasks are more appealing than complex ones) (Karachiwalla and Pinkow, 2021; Pedersen et al., 2013) and the IT platform (attractive, easy to use and working as expected without substantial effort from the participant) (Davis, 1989; Schenk et al., 2019). The literature highlights that, other than rewards, one of the primary mechanisms for attracting participation and thus for crowdsourcing to work successfully is reducing the complexity of the crowdsourcing task (Cullina et al., 2015; Karachiwalla and Pinkow, 2021). Specifically, the literature uses expectancy theory to relate the expected effort (linked to the complexity level) of the crowd and the expected probability of solving the task (connected to crowd capabilities) with the expected reward (Sun et al., 2015). If participants believe they can succeed in solving the task and thus can obtain the prize, their motivation to participate will be stronger. Therefore, researchers in the field recommend breaking down complex tasks into simpler ones that can be run independently to attract and keep participants engaged (Kittur et al., 2011). Interestingly, researchers also find that expert citizens (with knowledge of the specific crowdsourced topic) are more motivated to participate in both complex and simple tasks than non-experts (Seidel et al., 2013). This is reminiscent of the idea above, that citizens interested in the topic, or its outcome, are more motivated to participate in a crowdsourcing initiative, with payment being less critical (Blohm et al., 2018; Tinati et al., 2017).

2.2.4. Quality assurance and aggregation

Another key challenge in crowdsourcing is ensuring the quality of contributions and the use of quality assurance and aggregation methods to produce a valuable output from the raw crowdsourced contributions. Of course, one cannot expect the crowdsourced contributions to be directly usable, and adequate preprocessing, quality assurance and aggregation techniques must be in place (Barbier et al., 2012; Blohm et al., 2018; Daniel et al., 2018; Karachiwalla and Pinkow, 2021; Zhao and Zhu, 2014). Data are transformed and cleaned through preprocessing by removing irrelevant or incorrect contributions and made ready for the aggregation phase. Aggregating the data collected from the crowd is essential in inferring the correct answer to the task. Critical challenges of data collection in crowdsourcing are accuracy and duplication (Pravst et al., 2021) and data representativeness (Arbia et al., 2018, 2023). Human error and spammers are some issues that can affect data quality and severely impact aggregation accuracy (Singh et al., 2021). Furthermore, crowdsourcing typically follows a 'convenience sampling' approach (a non-probabilistic sampling approach) that does not ensure that the data accurately represent the population. Therefore, information users have to be very careful when analysing and extrapolating information from convenience samples. In contrast, probability sampling strategies are sampling methods that use some form of random selection that ensures that different members of the target population have an equal probability of being chosen. However, convenience samples provide valuable information that might otherwise be too costly to obtain (Jager et al., 2017).

Accuracy has been positively associated with higher rewards through the need for multiple observations (i.e. redundancy) for quality assessment (Sun et al., 2015). In practice, to promote the quality of contributions, researchers argue that incentives should be paid only to those contributions that pass the quality assurance process (Goncalves et al., 2015). They further argue that paying for invalid tasks has negative impacts, as it may encourage crowd members to continue cheating or make mistakes and may imply reputation loss (Hirth et al., 2013). At the same time, participants who continuously upload incorrect or low-quality data could be banned from the crowdsourcing platform to prevent noise in the data and waste of processing time (Gadiraju et al., 2019).

Furthermore, processing crowdsourcing data into quality information can be costly, depending on the methods and resources required for aggregation, again highlighting the trade-off between quality and cost (Ding and Zhou, 2018; Kawano and Yanai, 2014). On the one hand, minimising the costs of crowdsourcing implies ensuring high quality of contributions so that the total cost remains low; on the other hand, ensuring quality may also imply high costs for quality controls or paid rewards (Zhao and Zhu, 2014). It should be noted that quality control can be done *ex ante* – through crowd pre-selection and task design (setting quality thresholds), in runtime during the contribution and *ex post* by the crowdsourcer, the crowd (through an additional task) or a third party (Zhao and Zhu, 2014). However, more importantly, as manual data curation and quality control are costly and time-consuming, for crowdsourcing to be effective it is necessary to automate this process as much as possible, reducing the impact of the cost–quality trade-off. For example, Hirth's 2013 study found that, for simple, routine tasks, implementing automatic majority voting (contribution comparison) algorithms to rule out erroneous contributions and selecting the right solution is more effective and cost-efficient than third-party controls. Therefore, for each crowdsourcing task, the most suitable method of processing and extracting quality information has to be established, while also considering which of the available techniques (e.g. machine learning and data-mining algorithms) to will enable as much automation of this process as possible.

2.2.5. Technology

A further challenge in crowdsourcing is the choosing the technology. From Afuah and Tucci (2012), we know that, for crowdsourcing to work, there has to be available and readily accessible technology for collecting, verifying and aggregating contributions into a solution for the task at hand. Therefore the first significant challenge is choosing the best IT platform for the crowdsourcing task and the technology to process the contributions into a solution. Specifically, the crowdsourcing platform is a mobile app or a website providing an interface between the crowdsourcer and the crowd member. The IT platform provides services to the crowdsourcer or requester to submit tasks and to crowd members to perform tasks. The IT platform and its capabilities may differ based on the crowdsourcer's goal and the characteristics of the task. According to the task–technology fit theory, the platform must be easy to use and appropriate for the specific crowdsourced task. If this is the case, it is more likely to positively impact the crowd individuals' performance (Davis, 1989). It is particularly challenging when there is a mismatch between the mechanisms that are implemented to conduct the task and the support and service that the crowdsourcing platform provides; these are often rudimentary (Simperl, 2015). For example, the platform technology may not allow for the setting of specific quality control mechanisms during data entry processes. Furthermore, a recent study on mobile apps suggests that apps need to be easy to use, engaging and entertaining (i.e. including game elements) and to perform as expected to keep users engaged (Gu et al., 2022).

Moreover, a key decision in crowdsourcing, with implications for cost and functionalities, is whether to use a customisable marketplace platform or develop a proprietary one (Modaresnezhad et al., 2020; Schenk et al., 2019). The literature analyses this choice based on transaction costs, network externalities and available skills and competencies. It suggests that a marketplace platform reduces transaction costs, and that it is adequate when the crowdsourcer does not have a well-known brand name and a very large crowd is required, and the crowdsourcer does not have the competencies internally and does not aim to create a specific community (Schenk et al., 2019). On the other hand, the main disadvantage of a market platform is that it may not have all the necessary functionalities for the specific task. Furthermore, the choice of IT platform is an economic and operational decision and is particularly important due to the path dependencies associated with sunk costs.

Another important aspect related to the IT platform is trust. On one side, the platform must behave according to individuals' expectations regarding rewards payment, functionalities and interaction to build trust (M.-M. Wang and Wang, 2019). On the other side, for exchanging information from and to the end-users, smart devices communicate through an open channel, such as the internet, which is not sufficiently secure (Bodkhe and Tanwar, 2021). Therefore, crowdsourcing platforms must address security concerns regarding possible attackers (given the large-scale and potentially lucrative datasets) by implementing secure data exchange (Bodkhe and Tanwar, 2021). In addition, the platform must comply with strict legal regulations on data privacy, such as the General Data Protection regulation ([15]) (Lian et al., 2021). In summary, the platform must create trust by ensuring that it behaves as expected by the crowd participants, addresses security concerns and complies with data privacy regulations.

2.2.6. Information use

Another critical challenge is associated with the use of crowdsourced information. Crowdsourcing has value only if its results are used, achieving which is a challenge that entails building trust in the solution and the tool and implementing the proper dissemination formats and channels; to succeed in this, it is crucial to involve the information users from the design stage onwards, as is true of any other data-sourcing system (Matheus et al., 2018). It is also essential to enable users of the information to give feedback, which can help to improve the crowdsourcing model for data collection.

2.2.7. Ethical considerations

The adoption of crowdsourcing is sometimes ethically controversial (De Stefano, 2016; Standing and Standing, 2018). Some authors consider that crowdsourcing presents challenges relating to labour law, social protection and the employment relationship. In crowdsourcing, labour costs are minimised; however, the crowd members can be likened to independent contractors. As a result, there is some

([15]) https://ec.europa.eu/info/law/law-topic/data-protection_en.

concern about whether the rewards of crowdsourcing bears costs such as social security contributions and correspond to the statutory minimum wage (De Stefano, 2016). Therefore, one criticism of crowdsourcing is that it exploits cheap or unpaid labour (Standing and Standing, 2018). Moreover, there is no requirement for a relationship between the organiser and crowd members outside the sourcing platform.

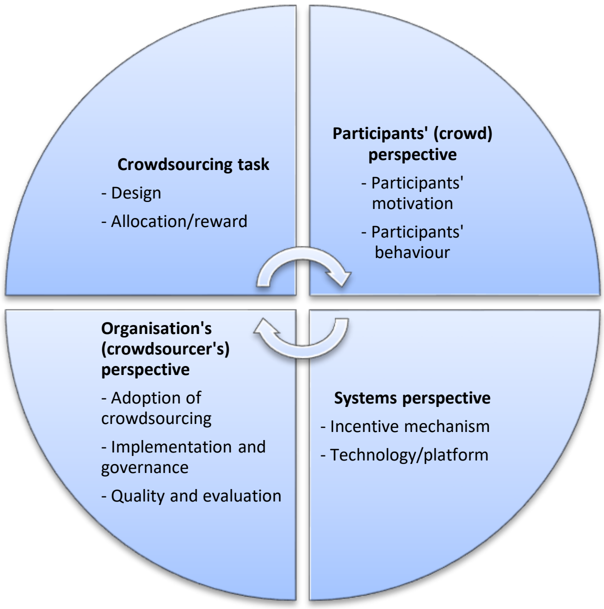
In contrast, other authors consider crowdsourcing as part of the collaborative economy (Aluchna, 2018) and suggest that two questions must be asked: ‘Does the initiative deliver primarily economic value or another type of value?’ and ‘Who is the value mainly created for?’ These authors focus on the fact that crowdsourcing can create shared value for the crowdsourcer, the participant crowd and the general public, with the results responding to different stakeholders’ expectations.

Finally, following the remarks of Blohm et al. (2018) and Kamoun et al. (2015), crowdsourcing, like any other method of allocating a task to achieve a goal, needs effective management and risk assessment mechanisms and adequate resources and capabilities to achieve the goal effectively and efficiently. Successful crowdsourcing entails developing a sound business case in terms of problems to be solved and objectives to be achieved, conducting a stakeholder analysis, analysing existing alternatives and recommendations, and establishing the main assumptions and constraints. Other activities include setting the budget and the schedule and estimating resources. In addition, potential problems – for example lack of participation, poor data quality and difficulties in aggregating contributions and using information – that may arise during the project should be identified and managed (Kamoun et al., 2015). Poor planning and management of a crowdsourcing initiative can lead to failure to achieve the desired result. Accordingly, researchers suggest deciding on success metrics (e.g. valid contributions number, product and regional coverage), estimating and monitoring costs of resources (e.g. marketing, rewards, personnel, IT) and managing risks to contribute to crowdsourcing success (Ford et al., 2015; Karachiwalla and Pinkow, 2021).

2.3. System view

The literature stresses the importance of looking at the crowdsourcing system as a whole and not just the individual components (Karachiwalla and Pinkow, 2021). Accordingly, Zhao and Zhu (2014)propose an analytical framework based on three interacting components: the crowd, the crowdsourcer and the crowdsourcing system. In addition, some authors refer to a fourth component, the crowdsourcing task (Bhatti et al., 2020; Blohm et al., 2018; Hossain and Kauranen, 2015; Karachiwalla and Pinkow, 2021), as depicted in Figure 2. First, the task is what is to be outsourced, with instructions; here, task design and allocation, and usually rewards, are important. Second, the crowdsourcer or requester is the individual or organisation that launches the task to obtain a solution. Third, the crowd members are the participants who use their resources and skills to perform the task in exchange for an expected benefit (extrinsic and intrinsic motivations). Fourth, the system, including the incentive mechanisms and the IT platform, enables the exchange between the crowdsourcer launching the task and the crowd members performing it (Bhatti et al., 2020).

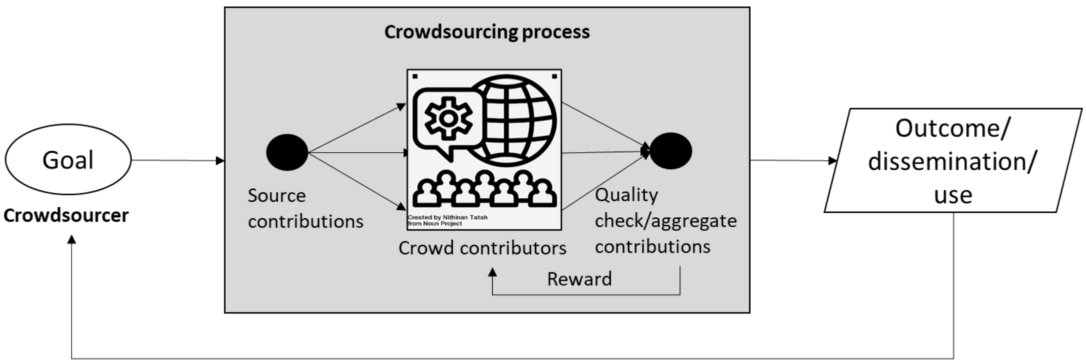
Figure 2. The main interacting crowdsourcing components



Source: Author, based on Zhao and Zhu (2014) and Karachiwalla and Pinkow (2021).

Other authors analyse crowdsourcing as a method for data collection from the perspective of the data life cycle. All data, whether they come from sensors, official statistics or crowdsourcing, go through a set of time-ordered stages (processes) in the data life cycle (i.e. goal/task definition, data collection, processing, analytics, storage and dissemination and usage) (Dahlander et al., 2019; Matheus et al., 2018; Roth and Luczak-Roesch, 2020). Data life cycle management comprises the processes, policies and procedures of managing data within an organisation through its life (from creation to use and retirement) (Figure 3). Therefore, it is helpful to think about data management in crowdsourcing within this framework. It includes all actions and sequential processes related to the data, making it possible to analyse factors and elements at each stage that may require different resources and capabilities and involve different risks (Kamoun et al., 2015).

Figure 3. Overview of the crowdsourcing process and actors



Source: Prototypical crowdsourcing approach based on Geiger et al. (2011).

Whether analysed from the perspective of the key components (task, crowd, crowdsourcer, system/platform) or the processes (initiation, task performance and data aggregation, use and dissemination), the literature refers to several key elements to consider in crowdsourcing (Blohm et al., 2018; Nassar and Karray, 2019): crowd management, quality assurance, incentive mechanism and technology. Moreover, in the crowdsourcing process, it is necessary to define the task and specify its value, the necessary skills, and remuneration and other incentives, while considering the effort involved in the task and the potential crowd's motivations. Then, publicise it via an open call with a campaign to attract the public. Once participants have been attracted, to maintain participation it is essential to continue encouraging, communicating with and giving feedback to the crowd and, if necessary, initiating new awareness campaigns or adjusting incentives. Furthermore, in the crowdsourcing process, the multiple contributions have to go through a quality assurance process that verifies and aggregates them into a solution that can be disseminated and from which indicators and information can be extracted.

2.4. Crowdsourcing versus traditional methods for data collection

An organisation can consider different options to collect data, such as (1) using professional enumerators, usually mediated by or outsourced to a contractor and following some technical specifications related to the sampling approach, coverage or number of observations, among other things; (2) using internal resources; or (3) using crowdsourcing (mediated by a contractor or not) (Afuah and Tucci, 2012). In addition, new technologies nowadays offer alternative data collection methods, such as web scraping, sensors or social media mining, which will be applicable or not depending on the data to be collected.

Here, using the system view (processes and components), we compare collecting data using crowdsourcing and conventional methods (using professional enumerators and defined samples). Recall that using crowdsourcing for data collection is suited mainly to simple, modular and transmittable tasks (Afuah and Tucci, 2012; Blohm et al., 2018; Karachiwalla and Pinkow, 2021) and when many observations and distant knowledge are required (Afuah and Tucci, 2012). Recall also that simple tasks are associated with low job effort. Therefore, the crowd may participate for minimal, or even without, payment if their interest is in the result (extrinsic or intrinsic motivation) (Blohm et al., 2018). In any case, in crowdsourcing, the focal organisation (crowdsourcer) broadcasts a task to the crowd in an open call, as it often does when searching for a contractor (e.g. hiring professional data collectors) (Afuah and Tucci, 2012). However, there are differences in several aspects; these are summarised in Table 1 and explained below.

While in traditional resource management, resources are known and task allocation is usually controlled, in crowdsourcing, participants are unknown and select the tasks they want to do (Cabanillas, 2016). Therefore, even if a task is solicited in both cases, a crucial difference between crowdsourcing and contracting or outsourcing the data collection to professionals is that, in crowdsourcing, the organisation does not select the crowd (it is self-selected), whereas in contracting or outsourcing the data collection, the organisation would evaluate and select a contractor, with several implications. First, when participants are self-selected, the data collected follow a convenience sample, a non-probabilistic sampling approach that, as mentioned above, fails to ensure that the data obtained represent the population (Arbia et al., 2018, 2023). Accordingly, given the biases inherent in a convenience sample, researchers and analysts should be cautious when generalising results. Moreover, the crowd is often large, anonymous and characterised by diverse and uncertain knowledge, unlike professional data collectors, who are usually trained and have specific knowledge and skills. However, in simple and modular tasks requiring general knowledge, crowd members with different types and levels of expertise can work effectively on different task modules in parallel. Moreover, if particularly distant knowledge is required, crowd diversity (e.g. geographical) may be positive (Afuah and Tucci, 2012). On the other hand, if specific knowledge is required, the organisation must target the crowd well, or contracting may be more effective.

More importantly, in crowdsourcing, there is no contractual relationship between the focal organisation (crowdsourcer) and the crowd, implying that participation relies exclusively on crowd motivation and incentives. In contrast, the contract regulates the relationship in traditional data collection. Crowd management and communication play an essential role in crowdsourcing but may be more complex than communicating with a contractor or a more limited number of hired professional data collectors bound by contract (Afuah and Tucci, 2012).

In terms of IT technology, there may be no significant differences; both crowdsourcing and traditional methods can use modern IT solutions. Pravst et al. (2021), for instance, compare the effectiveness of researcher-driven data collection using a mobile app with that of previous data collection when the mobile app was not used. Moreover, in crowdsourcing, the quality of contributions is more uncertain and difficult to establish a priori (known *ex post*) even if particular crowd, task design and data entry controls can be introduced. Some argue that quality may be lower than that of professional data collectors due to entry errors of non-trained participants, or potential cheaters or spammers (Hirth et al., 2013). The implications for users' trust in the data and the usability of the crowdsourced solution are important, as data may remain unused as a result (Han et al., 2019; Hosseini et al., 2019). In contrast, several empirical studies find that crowdsourcing achieves an accuracy level comparable to that achieved by experts and saves time (Khare et al., 2016; Mortensen and Hughes, 2018). Differences in accuracy findings can be attributed to differences in the type of task (e.g. degree of simplicity), the process of filtering and aggregating to obtain the results, and the management of the crowdsourcing process itself (e.g. rewards and quality controls). Given the uncertainty and possible low quality of the multiple raw data, in terms of accuracy and coverage, and the variety of multiple contributions, data processing (preprocessing and aggregation) could be more complex in crowdsourcing than in traditional methods of data collection, but depends on the level of automation achieved for the specific task. In simple crowdsourcing tasks, automatic majority voting mechanisms to control quality are easy to implement, saving costs compared with manual data curation methods (Hirth et al., 2013). Buettner (2015) describes quality assurance in crowdsourcing as a coordination problem and an interdisciplinary challenge containing human–computer or human–agent interaction issues.

Regarding cost, crowdsourcing should function with relatively low payments offered for simple tasks, making the cost of running crowdsourcing relatively modest compared with hiring professional data collectors, although upfront costs and processing costs should not be underestimated (Khare et al., 2016). Finally, recent studies suggest that crowdsourcing should not replace but complement expert-driven studies and traditional data collection with speed and timeliness and volume (Khare et al., 2016; Pravst et al., 2021).

An important conclusion to note is that the advantages or disadvantages of one method over the other will depend on the task, mainly the simplicity and modularity of the task and whether it requires distant knowledge and specific or general knowledge and skills. In addition, it will depend on the availability of an effective aggregation method and the use of effective management mechanisms, mainly to motivate the crowd and control the whole process.

Table 1. Comparison of crowdsourcing and traditional data collection

	Crowdsourcing data collection	**Traditional data collection**
Data collection		
Task	Solicited	Solicited
Participants	Self-selected crowd	Enumerators (professional)
Sample	Convenient sample 😐	Defined sample ☺
Participant number	Typically large ☺	Contract-based ☺
Knowledge	Diverse 😐	Specific 😐
Relationship	Volunteer (crowdsourcing platform)	Contract-based
Motivation	Extrinsic/intrinsic incentives 😐	Contract ☺
Management/ communication	More complex ☹	Less complex (contract) ☺
Tech	IT (mobile app, web)	IT, paper
Quality	Lower/uncertain ☹ (mainly *ex post*)	Higher (contract; trained professionals) ☺ (*ex ante*)
Data processing	Automatic or human (aggregation)	Automatic or human
Cost	Lower (data collection) ☺ High upfront costs ☹	Higher (data collection) ☹

Source: Author

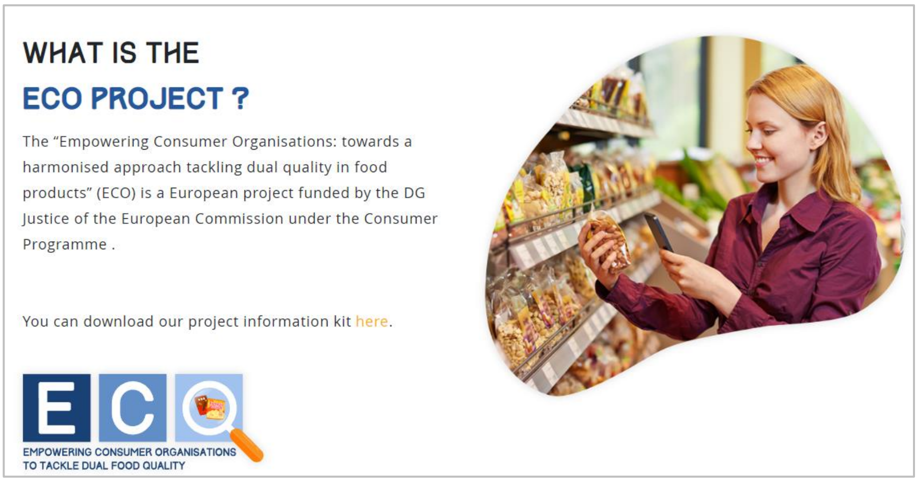
3. Current approaches to tracking food quality and food safety with citizen input

This section reviews the experience of several online platforms and initiatives that monitor food standards and safety, and how crowdsourcing initiatives are already used or could be implemented to feed into databases to provide information about nutritional aspects and support policies, or contribute to helping consumers make more informed choices when buying food products.

3.1. ECO project interactive platform

'Empowering consumer organisations: Towards a harmonised approach tackling "dual quality" in food products' (ECO [16]) is a European project funded by the European Commission Directorate-General for Justice and Consumers under the consumer programme (Figure 4).

Figure 4. The ECO project



Source: *https://www.fightdualfood.eu/*.

The ECO project is intended to strengthen the capacity of consumer organisations to test food products and identify potential DQ cases, disseminate test results and report potential unfair practices through a set of guides (including definitions and principles) and a manual to harmonise testing and an interactive web platform to upload the results. The guidance and manual for harmonised testing indicate that a testing campaign may include only products marketed under the same brand and packaging in at least three Member States. Moreover, sufficient product samples must be collected at retailers for all the anticipated testing activities (i.e. sensory analysis and laboratory testing). In addition to these general guidelines, the following steps should be performed.

- **Verify to what degree the front of packs of the compared products might be considered identical.** Expert panels can be hired for this purpose. Only food products with practically identical front-of-pack designs and marketing strategies go on to the next step, as only those potentially breach the unfair commercial practices directive.
- **Compare the nutritional values and the list of ingredients declared on the label (i.e. back-of-pack information).** Only food products with significant differences in the ingredients and nutritional values and practically identical front of packs go on to the next step.

(16) https://www.fightdualfood.eu/.

- **Use sensory analysis** to better understand the nature of any differences. Only food products that present significant differences after the sensory testing, have significant differences in ingredients and nutritional values and have practically identical front of packs go on to the next step.
- **Conduct laboratory tests** to establish whether the products are of different quality.

Following the first two steps, the web platform classifies the quality of seemingly identical products as 'good' (no DQ) if ingredients and nutritional values are the same and the front-of-pack design is the same or similar. On the other hand, it classifies the quality as 'medium' (needs further information) if the list of ingredients and/or nutritional values vary among countries included in the testing and the front of packs are similar or identical with regard to their design.

It should be noted that the interactive web platform enables consumer organisations and any citizen to report potential DQ cases by uploading images of the front and back of pack of food products and describing the case. This, then, can be analysed as a type of crowdsourcing implemented to collect the information in question. Below, the data life cycle elements and enabling factors, among other aspects, are outlined (Matheus et al., 2018).

Organiser (institutional set-up). The platform is a three-partner collaboration between the Association for the Defence and Orientation of Consumers, an Italian association of consumers recognised by the Ministry of Economic Development and present in all 20 Italian regions; Safe Food Advocacy Europe, a European non-governmental organisation (NGO) whose objectives are to ensure that consumers' health and concerns remain at the core of the EU's food legislation; and InfoCons, a Romanian consumer organisation that protects the right of consumers and raises their awareness.

Data collection

- Who – consumer associations/citizens across the EU
- what – uploading food pictures and explanations in text form
- why (motivation) – obtain feedback (extrinsic); social contribution (intrinsic)
- how (tool for interaction) – web page.

Data processing. Manual (packaging examination, sensory and laboratory tests).

Data analytics. Both manual and automatic. The outcome is a database reporting and sharing misleading practices related to branded products, using information uploaded by consumer organisations and citizens.

Dissemination. Web platform

Usage. Policymakers, consumer associations and companies.

Service management. Managed by ADOC, Safe Food Advocacy Europe and InfoCons.

Cost drivers. Laboratory testing (tends to be costly); manual data processing.

Funding. Public (European Commission).

Business model. Public service.

Advantages. Harmonised guides and handbook for collecting data on DQ; an interactive web platform reachable by any organisation or citizen.

Challenges. Citizen awareness about the platform; manual data processing and analytics; budget needed and funding after the end of the project is unclear. No information has been published on sensory analysis and laboratory tests, nor on how policymakers and other stakeholders are using the data.

3.2. Lebensmittelklarheit initiative by German consumer centres

Lebensmittelklarheit (food clarity) is a web portal operated by German consumer centres and funded by the German Federal Ministry of Justice and Consumer Protection. The purpose of the portal is to provide consumers with clear and transparent information about the food and beverage products they buy. The website aims to empower consumers by giving them the knowledge and tools to make informed purchasing decisions. This is achieved by publishing product tests and articles and background information on various food-related topics, such as food labelling, nutrition, additives and sustainability. The portal also aims to promote fair and transparent practices in the food industry by highlighting misleading or deceptive practices and calling for more transparency in food labelling and marketing. Ultimately, the goal of the portal is to help consumers make healthier and more sustainable choices, while also holding food producers and sellers accountable for their products and practices (Figure 5).

Figure 5. The Lebensmittelklarheit initiative

Source: *https://www.lebensmittelklarheit.de/.*

The Lebensmittelklarheit web portal offers a web space where consumers can file complaints about food products that they believe are misleading, unclear or false in their labelling or advertising ([17]). The portal team reviews the complaints and, if they find them to be valid, contacts the food company in question to request corrective action. If the food company fails to take corrective action, the complaint may be escalated to the relevant authorities, such as the German Federal Office of Consumer Protection and Food Safety, to take legal action against the company. The portal aims to promote greater transparency and accountability in the food industry and protect consumers' rights. Companies are encouraged to respond to complaints promptly and take corrective action to avoid escalation of the complaint.

Organiser (institutional set-up). Operated by German consumer centres ('Verbraucherzentrale').

Data collection

- who – consumers in Germany
- what – uploading food pictures, complaints and explanations in text form
- why (motivation) – obtain feedback (extrinsic); social contribution (intrinsic)

([17]) Other services offered by the portal include a search option (enables consumers to search for specific food products and check their ingredients, nutritional value and possible health effects); ratings (consumers can rate and comment on different food products based on their personal experiences, which can help other consumers make informed decisions); news (the portal provides the latest news and updates related to food safety, food labelling and consumer rights); and campaigns (the portal runs various campaigns to promote transparency and consumer awareness).

- how (tool for interaction) – web page.

Data processing. Manual (complaint examination and response by companies).

Data analytics. Manual.

Dissemination. Web platform.

Usage. Consumer associations, consumers and companies.

Service management. German consumer centres.

Cost drivers. Manual examination.

Funding. Public (German Federal Ministry of Food and Agriculture).

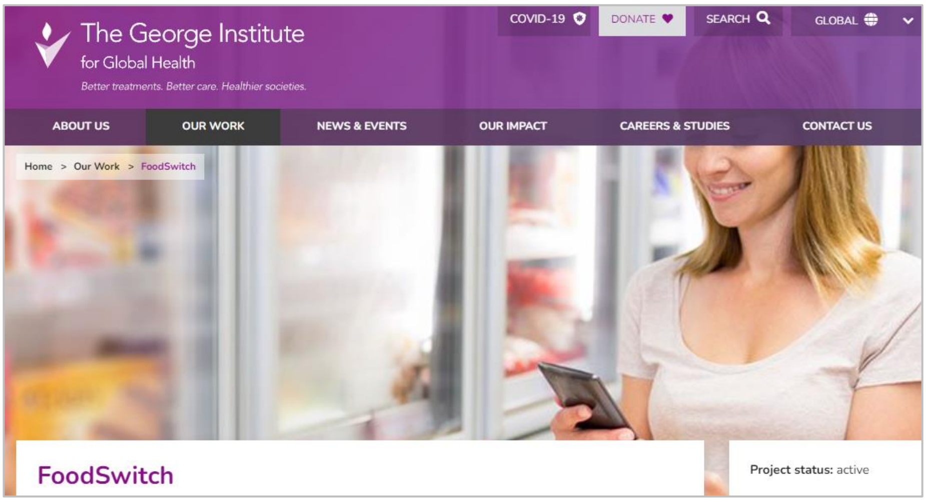
Advantages. Easy-to-use platform for consumer interaction on food quality aspects.

Challenges. There is a need to raise citizen awareness of the platform; data processing and analytics are done manually; the budget/funding needed is uncertain; it is unclear what the implications / next steps are if the issue is not resolved through the platform.

3.3. FoodSwitch app

The FoodSwitch app is an initiative launched in 2012 by the George Institute for Global Health to bring transparency to the world's food supply with a vision of an optimised food system for human health and the health of our planet. It collects data from packaged food labels and supports a food composition database, aiming to help consumers make healthier choices. The app has more than 700 individual food categories (Figure 6).

Figure 6. The FoodSwitch app



Source: *https://www.georgeinstitute.org/projects/foodswitch*.

The FoodSwitch app helps consumers to make better food choices by providing easy-to-understand health and nutrition information on a scanned product, suggestions for healthier alternatives to 'switch' to ([18]) and educational information about healthy eating and the importance of a balanced diet. A crowdsourcing function was built into the app, enabling users to submit missing items. It resulted in the submission of approximately 1 million photos of new food items by users, which

([18]) For instance, if a consumer scans a white bread product, the app recommends healthier white bread products.

translated into more than 100 000 new products being added to the database (Dunford and Neal, 2017).

Organiser (institutional set-up). George Institute for Global Health.

Data collection

- who – user contributions (new items) by consumers and food companies
- what – uploading food pictures (users)
- why (motivation) – consumers obtain feedback and food companies ensure that enhancements to their products are rapidly captured within the system (extrinsic); social contribution by supporting a tool for healthier food choices (intrinsic)
- how (tool for interaction) – mobile app.

Data processing. A data management centre, overseen by the George Institute, checks and adds the information to the respective country's database. The data management teams undertake annual surveys in each country to ensure up-to-date and accurate data.

Data analytics. Nutrient profiling method underpinning the health star rating (HSR).

Dissemination. Mobile app (free access).

Usage. Consumers, government and public health researchers.

Service management. Data management centre.

Cost drivers. Initially, a labour-intensive exercise with high costs.

Funding. Private – Public.

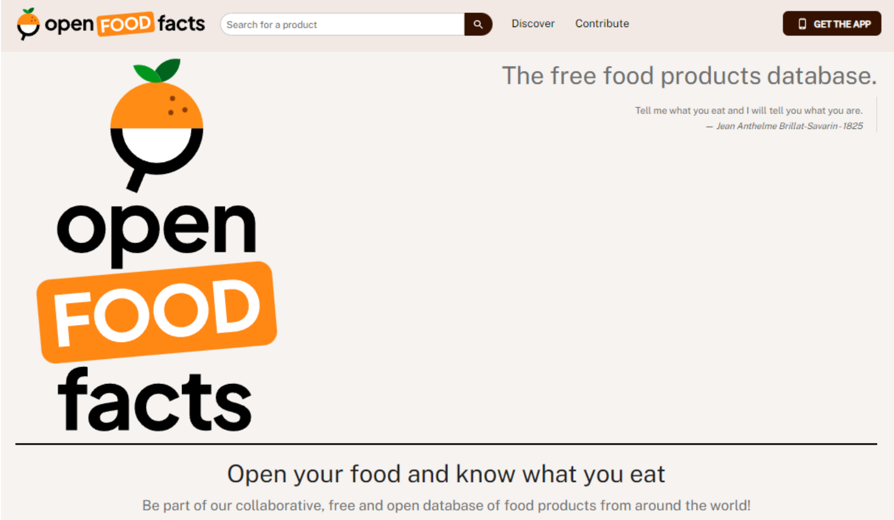
Advantages. Easy-to-use platform; valuable information fed back to users (motivation not monetary).

Challenges. Need to streamline data management systems, data entry automation and food classification to further reduce costs and enhance data processing.

3.4. Open Food Facts – World

Open Food Facts is a database of crowdsourced food products launched in 2012 and is managed by a non-profit organisation based in France. The goal of Open Food Facts is to provide transparent and reliable information on food products to consumers, researchers and public health officials. The platform enables users to search for products by brand name, category and nutritional information. It collects information on ingredients, allergens, nutritional value and other details on product labels. The crowd consists of thousands of volunteers from around the world who contribute to the database by scanning barcodes and entering information about products. Currently, it has 2 674 645 products. The information is then verified by a team of moderators to ensure accuracy and completeness (Figure 7).

Figure 7. Open Food Facts database



Source: https://world.openfoodfacts.org/.

Consumers can use Open Food Facts to make informed food choices, and, as it is open data, anyone can reuse the information for any purpose (e.g. for research and for public officials to study trends and patterns in the food industry). Over 15 000 contributors have added over 1 000 000 products from 150 countries using a smartphone app to scan barcodes and upload pictures of products and their labels.

Organiser (institutional set-up). Non-profit organisation.

Data collection

- who – user contributions by consumers and food companies
- what – uploading pictures of products and their labels
- why (motivation) – consumers obtain feedback and food companies ensure that enhancements to their products are rapidly captured within the system (extrinsic); social contribution by supporting a tool for healthier food choices (intrinsic)
- how (tool for interaction) – mobile app or camera.

Data processing. Consists of several steps: data cleaning (the collected data are cleaned to ensure consistency and accuracy), data standardisation (the collected data are standardised to ensure uniformity and consistency, such as the product names, ingredients and nutritional information), data enrichment (the collected data are enriched with additional information, such as allergens, packaging details and images) and data verification (the enriched data are verified by a team of volunteers who check for accuracy and completeness). The crowd can contribute to reviewing/completing the information.

Data analytics. Provides information on products' Nutri-Score and Nova nutritional classifications.

Dissemination. Mobile app (free access).

Usage. Consumers, government and researchers.

Service management. Non-profit organisation with eight employees.

Cost drivers. Initially, a labour-intensive exercise with high costs.

Funding. Private contributions.

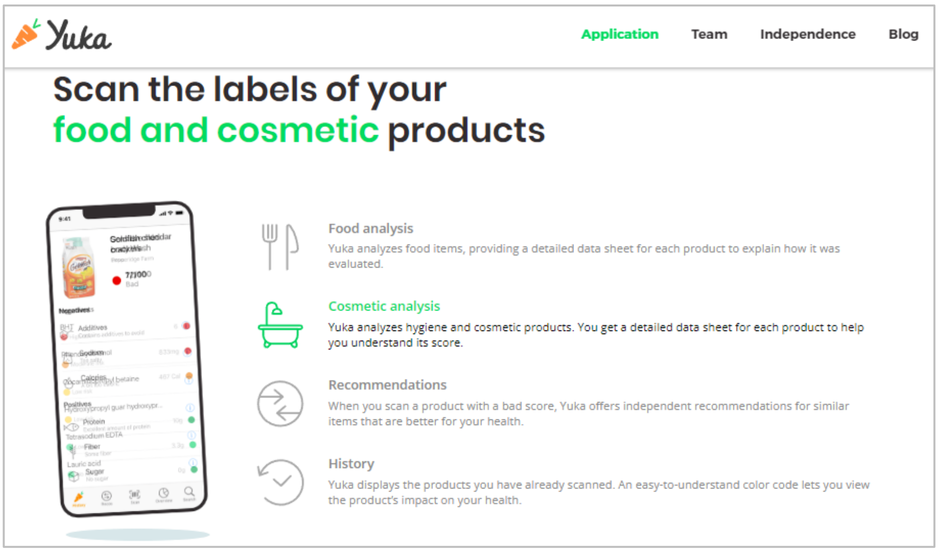
Advantages. Easy-to-use platform; valuable information fed back to users (motivation not monetary).

Challenges. Country differences; quality processing; representativeness.

3.5. Yuka mobile app

Yuka is a private start-up created in 2016 to support food and cosmetics purchasing decisions through a mobile app that relies on a comprehensive database of 2.5 million food products (around 1 200 new products are added every day) and 1.5 million cosmetic products. It automatically analyses each scanned product's ingredients based on its scoring methodology to help consumers understand how healthy a product is (Figure 8).

Figure 8. The Yuka application



Source: https://yuka.io/en/app/.

Yuka mobile app is designed to help consumers make informed choices by scanning the barcode of a product and receiving a score based on its nutritional value and composition. The app analyses the product based on three criteria: nutritional quality, presence of additives and use of organic farming. According to 2021 data, the service provided by the app is fully financed by premium subscriptions (EUR 575 566, down from EUR 777 711 in 2020), book sales (EUR 536 734, up from EUR 485 286 in 2020), calendar sales (EUR 160 699, down from EUR 271 826 in 2020) and other revenue streams (e.g. partnerships with various companies, data analytics on users' behaviour) (EUR 8 142). However, losses amounting to EUR 395 853 were reported in 2021 (in contrast, in 2020, it obtained a net margin of 1.1 %), where the largest cost items were external purchases and services, amounting to EUR 913 043 (e.g. verification and transcriptions of ingredient lists, IT services), followed by salaries, amounting to EUR 409 021. Overall, Yuka's revenue model is based on a combination of subscription services, affiliate marketing, partnerships and data monetization.

Organiser (institutional set-up). Private company.

Data collection

- who – user contributions and brands
- what – uploading food pictures (users) / sharing pictures (brands)

- why (motivation) – extrinsic (obtain product assessment) and intrinsic motivation (social contribution).
- how (tool for interaction) – mobile app.

Data processing. Automatic control/processing system via image and text recognition for ingredient list transcription; manual verification/processing by Yuka employees, external services and other users. Yuka does not process or sell user data. All user data remain strictly confidential.

Data analytics. Automatic.

Dissemination. Mobile app (free and premium access).

Usage. Consumers.

Service management. Two full-time employees are dedicated to managing the database, verifying the contributions and correcting them if necessary.

Cost drivers. External services (data processing, IT services); personnel.

Funding. Private (business income).

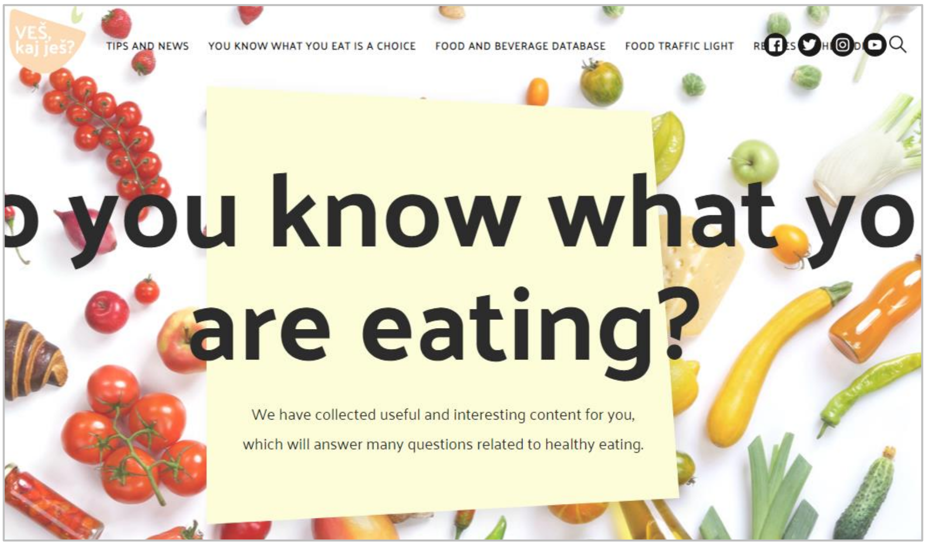
Advantages. Easy-to-use platform; valuable information fed back to users (extrinsic motivation).

Challenges. High processing costs; scoring methodology; engaging users.

3.6. 'Veš, kaj ješ?' (#VKJ; 'Do you know what you are eating?')

The app was launched in mid 2019 by the Nutrition Institute, the Jožef Stefan Institute and the Slovenian national consumer organisation, and is supported by the Slovenian Ministry of Health. The app enables users (consumers) to scan the barcode of selected food products and receive feedback information on the product's nutritional composition (Figure 9). It also interprets the nutritional information based on the nutrient profile using the food traffic light labelling system, to support healthier food choices. In addition, the application has a crowdsourcing function, which is activated when a user scans a food barcode that is not yet included in the database or when there is a difference between the nutritional composition of the scanned food and the information available in the mobile application (Pravst et al., 2021).

Figure 9. 'Veš, kaj ješ'?' website



Source: https://veskajjes.si/.

Since its release, the application's users have contributed more than 11 000 unique items, which has translated into more than 9 000 processed items – that is, individual products on which information is available in its database. It currently has around 24 000 active users.

Organiser (institutional set-up). Nutrition Institute, Jožef Stefan Institute and the Slovenian national consumer organisation, supported by the Slovenian Ministry of Health.

Data collection

- who – user contributions by consumers
- what – uploading pictures of products and their labels
- why (motivation) – extrinsic (obtain feedback) and intrinsic motivation (social contribution)
- how (tool for interaction) – mobile app.

Data processing. Data are processed in a dedicated web application (bazil.si) developed by the Jožef Stefan Institute, enabling researchers to view user data. The application transcribes the data of interest from the labels that can be seen on the images.

Data analytics. Interpretation of the nutritional profile using the food traffic light labelling system.

Dissemination. Mobile app (free access).

Usage. Consumers, government and researchers.

Service management. Nutrition Institute; Jožef Stefan Institute.

Cost drivers. Labour costs; IT.

Funding. Public.

Advantages. Reduced costs; speed; flexibility; scalability; diversity; and participation of citizens.

Challenges. Accuracy and duplication.

3.7. LEDA database by the Netherlands Food Information Resource

In 2016, the Dutch minister of health, welfare and sport asked the Netherlands Nutrition Centre to develop an app to help consumers make healthy food choices. Following this request, the Netherlands Nutrition Centre, under the umbrella of the Netherlands Food Information Resource (NethFIR), hosted the Dutch-branded food database LEDA (short for Levensmiddelendatabank, or 'Food Database') (Figure 10). The LEDA database also includes the Dutch food composition database, dietary supplement database and portion sizes database, which are hosted by the Institute for Public Health and the Environment (Westenbrink et al., 2021).

Figure 10. LEDA, the branded food database in the Netherlands

Source: *https://www.rivm.nl/publicaties/leda-branded-food-database-in-netherlands-data-challenges-and-opportunities*

The food industry provides data on a voluntary basis via intermediate organisations. These data are automatically uploaded overnight using application programming interfaces. The LEDA database includes product name, brand, manufacturer, data provider, barcode (global trade item number (GTIN) / European article number (EAN)), food group classification (the GS1 global product classification or other), serving size, net weight, ingredients, nutrient values (per 100 g or 100 mL), instructions for use and product images. Information identifyin foods (product name, brand, barcode) is checked during the validation process within LEDA. However, due to licence agreements with data providers, NethFIR can use the LEDA dataset only for its consumer information and nutritional research, and cannot share the data with third parties or make it publicly available online.

Organiser (institutional set-up). Netherlands Nutrition Centre.

Data collection

- who – food companies
- what – product name, brand, manufacturer, data provider, barcode (GTIN/EAN), food group classification (GS1 or other), serving size, net weight, ingredients, nutrient values (per 100 g or 100 mL), instructions for use and product images
- why (motivation) – licence agreement with NethFIR (some require payment)
- how (tool for interaction) – electronic data exchange facilities.

Data processing. Quality control procedures are assumed to start at the food producers when generating label information. The automated validation process at the Netherlands Nutrition Centre focuses on accuracy at the individual food level. Manual inspection by a nutritionist for new branded products.

Data analytics. Automatic and manual.

Dissemination. Offline and online health nutrition tools.

Usage. Consumers, policymakers and researchers.

Service management. Netherlands Nutrition Centre.

Cost drivers. Manual steps in data processing and data analytics.

Funding. Public.

Advantages. High product coverage (75 % of the market). High quality of data assumed at the data provider level.

Challenges. Data processing (some manual work required). Missing foods (alternative efforts, for example crowdsourcing or web scraping, are needed). Inconsistencies and inaccuracies need improvement, preferably during data entry by food producers. Using documentation standards and validation rules and enrolling in quality assurance programmes run by data providers such as GS1 or Brandbank would help to improve data quality and lead to more efficient data interchange procedures and data use.

3.8. Australian Branded Food Database

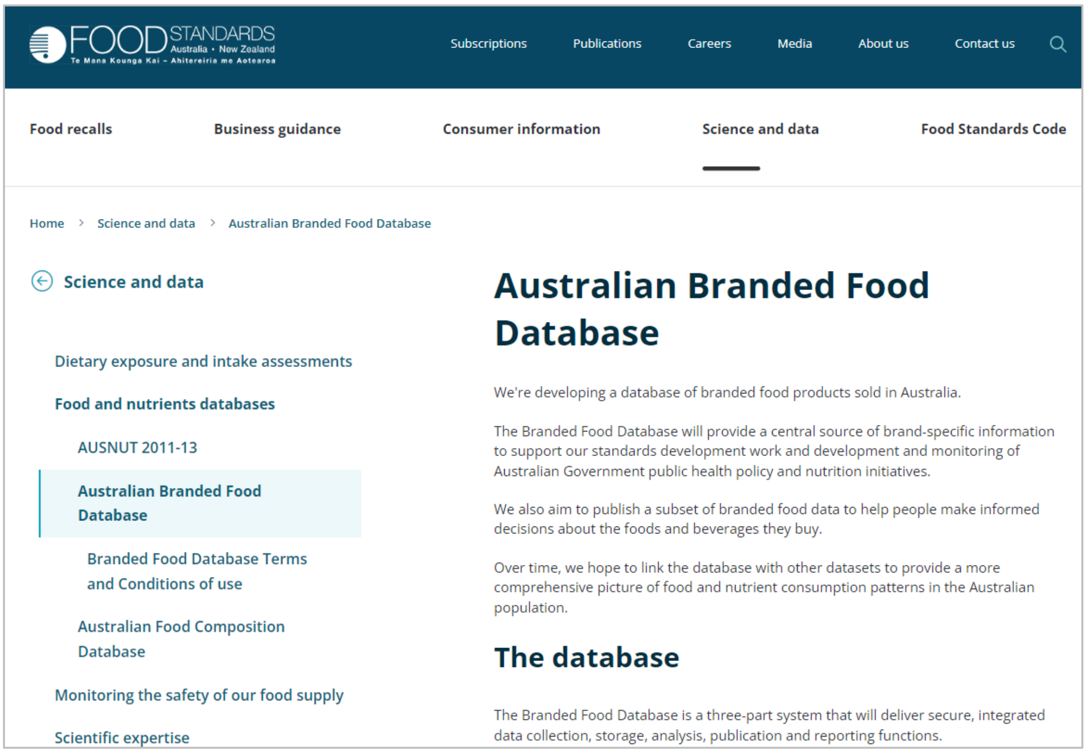
The Australian Branded Food Database ([19]) is a publicly available database that contains detailed information on the nutrient composition and other characteristics of foods commonly available in Australia. It is maintained by Food Standards Australia New Zealand (FSANZ) – an independent statutory agency that is part of the Australian government's health portfolio – and contains information on over 50 000 branded and generic foods, including packaged foods and beverages. The database provides information on a wide range of nutrients, including energy, protein, fat, carbohydrates, vitamins and minerals. It also provides information on serving sizes, ingredient lists and allergens, as well as other product characteristics such as brand name and packaging information. The database can be accessed online through the FSANZ website, and users can search the database by food type, brand name or nutrient content (Figure 11).

FSANZ also aims to publish a subset of branded food data to help people make informed decisions about the foods and beverages they buy. It is expected that, over time, the database will link with other datasets to provide a more comprehensive picture of the Australian population's food and nutrient consumption patterns.

The partnership with GS1 Australia (barcode provider) enables the agency to work directly with food manufacturers and retailers to collect branded food data. This collaboration with the industry is intended to capture a wide range of foods and product information, with the aim of covering 85 % of all packaged food and beverage products sold by Australian retailers by June 2023.

([19]) The database is under development.

Figure 11. Information page on the Australian Branded Food Database



Source: https://www.foodstandards.gov.au/science-data/monitoringnutrients/Branded-food-database.

Organiser (institutional set-up). FSANZ, an independent statutory agency that is part of the Australian government's health portfolio.

Data collection

- who – food companies
- what – GTIN, manufacturer, brand and food name, nutrition information panel, listed ingredients, pack and serve size, and HSR, if displayed
- why (motivation) – extrinsic (support standards development) and intrinsic (support public health policy and nutrition initiatives) motivations
- how (tool for interaction) – GS1 Australia-registered members can provide data to FSANZ via the national product catalogue, while other manufacturers and retailers (brand owners) can submit data via a free FSANZ online portal.

Data processing. All data provided are checked against predefined rules as part of the data collection and exchange process with GS1 Australia. Data cannot be incorporated into the branded food database until these rules have been followed. Targeted in-store audits will also be undertaken to ensure the accuracy and currency of the data collected.

Data analytics. Data link with other datasets to provide a more comprehensive picture of food and nutrient consumption patterns analytics: HSR

Dissemination. Website with the permission of data providers.

Usage. Consumers and researchers, and to inform public health initiatives (not for compliance or enforcement purposes).

Service management. FSANZ.

Cost drivers. Manual steps in data processing and data analytics.

Funding. Public.

Advantages. High product coverage (it aims to cover 85 % of the market).

Challenges. Need to link with other datasets to provide a more comprehensive picture of food and nutrient consumption patterns.

3.9. Food monitoring using the Composition and Labelling Information System

The Nutrition Institute of Slovenia uses the Composition and Labelling Information System (CLAS) ([20]) as a tool to support nutrition research and monitoring of the food supply in Slovenia. The institute developed a mobile application that enables researchers to collect data on the nutrient content and composition of branded food products sold in Slovenia. The app is linked to the CLAS and enables researchers to access detailed information on the nutrient content and ingredient composition of thousands of branded food products sold in Slovenia. The app enables researchers to scan the barcode of a food product and automatically retrieve information on the product's nutrient content and ingredient composition from the CLAS database. This information can be used to support nutrition research and monitoring of the food supply, as well as to inform public health policy and nutrition education efforts (Figure 12) (Pravst et al., 2021).

Figure 12. Web page about the CLAS

Source: https://www.nutris.org/en/composition-and-labelling-information-system.

Once transmitted to the online CLAS tool, individual product information is checked for quality using both automatic controls and a manual check by the researcher. Data extraction is completed in an online CLAS tool, with the support of optical character recognition (OCR) technology, and supported by manual work and cross-checking. In the 2020 study, more than 28 000 branded food products were sampled compared to the 6 348 products sampled in 2011 (traditional data collection without an app). Furthermore, in 2020, only 1 526 products matched with those sampled in 2011 (Pravst et al., 2021)

Organiser (institutional set-up). Nutrition Institute and the Jožef Stefan Institute.

Data collection

- who – researchers
- what – uploading pictures of products and their labels

([20]) CLAS is a database managed by the Food and Agriculture Organization of the United Nations that contains information on the nutrient content and ingredient composition of foods from around the world.

- why (motivation) – research (work)
- how (tool for interaction) – mobile app.

Data processing. Data are processed through the online CLAS tool, where they are checked for quality by the researcher using both automatic and manual controls for each product. Data extraction is completed using an online CLAS tool, with the support of OCR technology, and supported by manual work and cross-checking.

Data analytics. Food composition.

Dissemination. Study.

Usage. Consumers, government and researchers.

Service management. Nutrition Institute.

Cost drivers. Labour costs; IT.

Funding. Public.

Advantages. Reduced costs; scalability.

Challenges. Cross-sectional studies; product coverage.

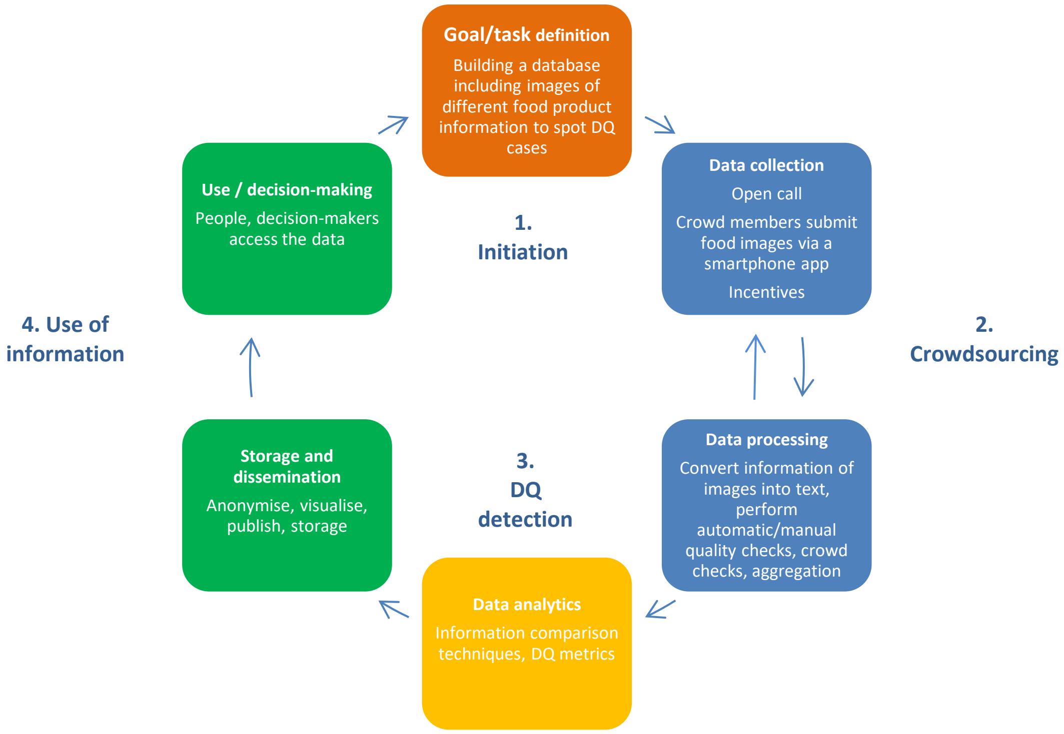
4. A possible method of applying crowdsourcing to monitoring dual quality practices

In this section, based on the review of literature and existing initiatives tracking food standards and food safety, we propose a framework to guide the application of crowdsourcing to tracking DQ. We propose a method of applying the framework, by identifying the primary success factors and the decisions that need to be taken in each phase and providing recommendations on design and governance mechanisms.

4.1. Framework

In order to assess the crowdsourcing model to monitor DQ cases, we use a systems approach (Bhatti et al., 2020). We propose a framework that integrates the crowdsourcing components (crowd, crowdsourcer, platform) and the crowdsourcing process throughout the data life cycle. Analysing crowdsourcing as a data collection method from a data life cycle management perspective enables us to assess all the processes, success factors, benefits, costs and risks at each stage throughout the cycle, from the creation of the initiative to the use of the results (Dahlander et al., 2019; Roth and Luczak-Roesch, 2020). We start by proposing a basic representation of the data life cycle to provide information about DQ in six steps, summarised as the following four phases (Figure 13): (1) initiation (goal and task definition), (2) crowdsourcing (data collection and processing and transformation, quality checking and aggregation), (3) detection of DQ and (4) use of information.

Figure 13. Data life cycle in crowdsourcing applied to the collection of branded food product images (front and back of pack, including nutritional composition and ingredients) to collect information on DQ issues



Source: Author, loosely based on the workflow proposed by Matheus et al. (2018).

The process starts with the **initiation phase**, common across different sourcing methods, which involves setting the goal and the task to be crowdsourced. In this phase, the crowdsourcing organisation (crowdsourcer*)* needs to (1) state the rationale for crowdsourcing (i.e. access to data not readily available, reduced costs, improved quality) and ensure that it aligns with the goal (i.e. providing information about DQ cases in a cost-efficient manner), (2) decide on the specific crowdsourcing task to allocate to the crowd (i.e. collect branded food product pictures of the front and back of pack, including ingredient list and nutritional composition, using a mobile app in order to create and regularly update a database of branded food product images and information) and (3) decide on how to manage the crowdsourcing – that is, design and implement processes and methods, choose the platform and technology, create an incentive mechanism and assess risks (Kamoun et al., 2015).

In this phase, the crowdsourcer must also design the workflow or process to integrate the crowdsourcing task to achieve the objective. The crowdsourcing organisation must plan the broadcasting campaign and incentive mechanism to recruit and sustain a crowd, and the quality and aggregation process to transform the crowdsourced data into information, and choose and develop the IT platform and tools. The quality and aggregating process includes converting the photos into machine-readable text and verifying and selecting product images to be part of the database that will report on DQ. Reporting on DQ requires comparing pictures of versions of the same product marketed in different EU Member States in terms of the similarity of the front-of-pack design and ingredient list to identify possible cases. Recall that DQ occurs where products are marketed as seemingly identical under the same brand and with the same or similar packaging design across different EU Member States when, in fact, they differ significantly in their composition or characteristics. Importantly, this process and how to disseminate the resulting information should also be designed in the initiation phase.

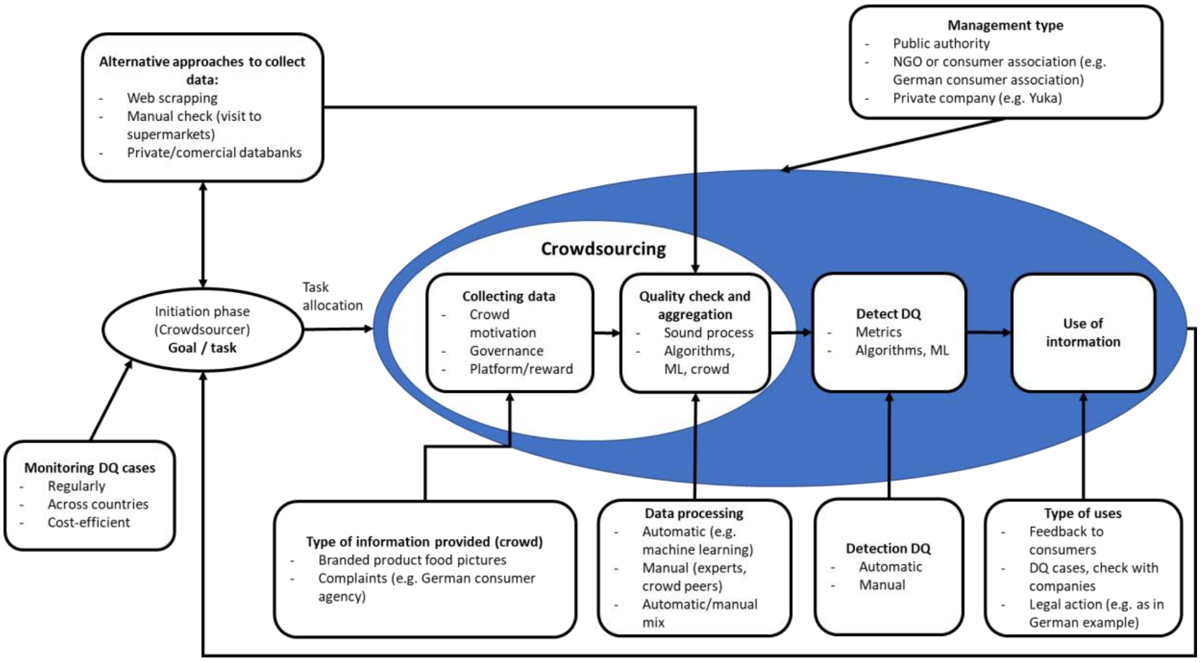
Next, in the **crowdsourcing phase,** data collection and processing is implemented and managed by (1) advertising the call for contributions to engage a crowd – that is, call for participants to contribute branded food product photos via a mobile app; (2) monitoring the response in terms of quantity and quality and aggregating information; and (3) incentivising contributions and motivating the crowd.

In the **DQ detection phase,** estimations and calculations, as designed in the initiation phase, which ideally should be fully automated, serve to assess the presence of DQ. In addition, inspection by a person with relevant experience may also be necessary.

Finally, in the **use of information phase,** data are published and disseminated for use, thus giving value to the crowdsourced information and feedback to the process. The data life cycle perspective adds this essential phase, the information dissemination and usage, which closes the crowdsourcing loop and usually receives less attention, even though value can only be created from data when data are used (Janssen et al., 2014). Accordingly, any crowdsourcing initiative should include a dissemination and utilisation strategy/plan.

In addition to the process, it is essential to consider all the actors involved in the cycle. These are usually the crowdsourcer, the crowd and the stakeholders who are the potential users of the information, such as Member State authorities, consumers and companies, and their associations or other relevant advocacy bodies. The organisation responsible for monitoring DQ cases – the crowdsourcer – could be a public authority, an NGO, a consumer association or a private company. Finally, the crowd is the public to whom the task is outsourced. Figure 14 presents the framework guiding the application of the technique, showing the key elements and processes of crowdsourcing within the data life cycle.

Figure 14. A framework for crowdsourcing applied to the collection of food images to provide information on DQ issues



Source: Author

4.2. Success factors for crowdsourcing for tracking dual quality at different stages of the data life cycle

Any organisation aiming to adopt crowdsourcing to achieve a particular goal effectively and efficiently must carefully consider all stages and characteristics of the crowdsourcing process to be implemented (Geiger et al., 2011). Therefore, in this section, using the system view proposed in our framework, we look at the key design elements and tools at different stages of the data life cycle that can ensure that crowdsourcing successfully identifies DQ cases.

4.2.1. Initiation phase: defining the goal/task

The goal of crowdsourcing in this study is to monitor cases of DQ in branded food products in the EU by regularly and cost-efficiently collecting and analysing photos of the front and back of pack of branded food products currently available on the market. Accordingly, the envisaged crowdsourcing task consists of photographing branded food products (front and back of pack, including the nutritional facts and ingredients information) available on the market by crowd volunteers using a mobile app. It achieves its goal if it obtains sufficient valid observations (i.e. pictures convertible to machine-readable text) that can be aggregated to reliably and cost-efficiently report on the issue that motivated the data collection (Blohm et al., 2018; Schenk and Guittard, 2011) – in this case, spotting DQ cases. As a result, a database of images of branded food products and their lists of ingredients can be created, where each product appears as many times as different front-of-pack and ingredient list versions are photographed, making it possible to identify DQ cases.

It is crucial to assess if the reason for choosing to crowdsource (e.g. cost-effectiveness, real-time or information needs) matches the task and the objective (Zhao and Zhu, 2014). The assumption is that, given the high cost of traditional methods that employ professionals or experts for regular data collection, crowdsourcing can be an effective and cost-efficient way to obtain observations (images) of many different branded food products across EU Member States. Furthermore, the information extracted from these pictures can also help monitor DQ cases and inform policymakers, companies and consumers. This study aims to analyse the feasibility of achieving this objective using crowdsourcing, which involves comparing it with alternative data-sourcing methods. Its aim is to provide an overview of both the advantages and challenges of using crowdsourcing for Member States' competent authorities and consumer and trade representatives or other relevant advocacy bodies that may consider it useful to set up and operate ICT solutions involving crowdsourcing to monitor the occurrence of DQ cases across the EU single market.

To achieve this, inspired by the initiatives and literature reviewed, we propose the following workflow of nine activities to meet the overall goal of providing information about DQ cases, including the definition of the crowdsourcing task.

1. **Defining, broadcasting and launching the task** through a digital tool (e.g. mobile app, web platform).
2. **Taking and uploading of branded food product pictures** (front- and back-of-pack information) by the crowd participants to a digital tool (e.g. mobile app, web platform) at their convenience (e.g. while shopping at the supermarket or when at home).
3. **Automatic conversion of image information** (i.e. ingredients, nutritional facts) into machine-readable text data.
4. **Reviewing picture and text transcript quality**, automatically or manually, by experts or non-expert participants (e.g. by the crowd through peer checks).
5. **Identifying the same or seemingly identical branded food products across countries** to enable cross-country comparisons. The literature suggests that combining crowdsourcing (for training data) and image recognition techniques / machine learning may be helpful for this (Kawano and Yanai, 2014; Yasmin et al., 2022).
6. **Aggregating data and adding food pictures to the database** (including adding food pictures to the final branded food product picture database);
7. **Creating DQ metrics** to assess the degree of similarity of product front of packs and the degree of difference of lists of ingredients across countries for the same product.
8. **Disseminating data** in the required formats.
9. **Use of data** by policymakers, businesses or the general public.

Crowdsourcing success and success factors for detecting dual quality cases

The analysis of success factors requires an understanding of what constitutes success in using crowdsourcing to detect DQ cases in branded food products. Generally, the performance or success of crowdsourcing, like those of any problem-solving task, are measured by the solution's quality, cost, speed and/or the mere fact of achieving it (Afuah and Tucci, 2012). Accordingly, success metrics are needed to evaluate the performance and effectiveness of crowdsourcing overall and at each stage (Blohm et al., 2018; Cullina et al., 2015; P. S. Ferreira et al., 2012; Karachiwalla and Pinkow, 2021; Solano-Hermosilla et al., 2020). Success in the DQ task could be considered to be achieved when sufficient valid observations (in terms of quality, quantity and accuracy) on many different branded food products (achieving representativeness) are obtained cost-efficiently, enabling comparisons of front-of-pack designs and ingredient lists of the same products in different countries. In this way, it can be determined whether there are cases of DQ – that is, product versions that appear to be the same owing to their presentation but have a substantially different list of ingredients. Examples of success metrics might include the number of products covered (valid images and total images) in each country, the number of products covered in several countries simultaneously, the average incentive cost per image and the average total cost per image.

However, a significant trade-off between quality and cost-efficiency emerges in crowdsourcing. Specifically, in crowdsourcing with volunteer workers (not hired and therefore not bound by a contract), the data quality (including quantity, accuracy and spatial coverage) is uncertain, as participation and accuracy cannot be assured a priori. For example, we cannot establish with a priori certainty whether all countries will be adequately covered, or which products or whether the same products will be covered in different countries. Therefore, finding, applying and adjusting ways to motivate people to send pictures of the appropriate quality is essential, and may have substantial cost implications. Furthermore, processing the pictures into a quality output may be costly, depending on the methods and resources needed. Therefore, reducing this trade-off is essential in applying crowdsourcing for DQ assessment. As mentioned earlier in the study, if the budget were unlimited, experts or a large pool of workers could be hired.

Afuah and Tucci (2012), as discussed in Chapter 2, suggest the following five success factors that make crowdsourcing an appropriate method to meet the desired objective. Below, we analyse these factors with reference to monitoring DQ in branded food products (Afuah and Tucci, 2012; Blohm et al., 2018; Karachiwalla and Pinkow, 2021; Liu et al., 2021).

1. The task is simple, modular and easily transferable.
2. A vast amount of contributions and distant search (as opposed to local search, i.e. knowledge or information required to solve the problem falls outside the focal agent's knowledge neighbourhood) are required.
3. There exists a potential crowd to contribute.
4. Contributions can be efficiently verified and aggregated into a desired solution.
5. There are available IT tools that are suitable for completing the task.

The first requirement (the task is simple, modular and easily transferable) refers to workflow activity 2 (taking/uploading branded food product pictures). It can be easily fulfilled, as the crowdsourcing task of photographing branded food products (front and back of pack, including ingredient list and nutritional facts) with a mobile app is simple (in principle, it requires seconds to a few minutes). Moreover, it can be made modular (different contributions can occur simultaneously and independently) and is easily transmittable to the public. The second requirement is also met. A vast number of contributions and distant search are required; the crowdsourcing must cover many different branded food products in all EU Member States. For example, from the LEDA database described in Section 3.7, we know that many food product pictures, namely 100 000, are needed to cover about 75 % of the Dutch market.

The third requirement is that there must be a potential crowd to contribute to the task. Indeed, nowadays, taking photos with smartphones and sharing them is commonplace, and most people in the EU have smartphones and mobile internet. According to Eurostat ([21]) data for 2020, 81 % of people in the EU use a smartphone. Moreover, people are becoming increasingly aware of what they eat and the quality. Knowledge of the topic usually increases the motivation to volunteer in crowdsourcing initiatives, contributing to cost efficiencies (Blohm et al., 2018; Tinati et al., 2017). Therefore, the third condition of the existence of a potential crowd is fulfilled. A further factor will be finding the right mechanisms to publicise the initiative and motivate participation while keeping costs down.

The fourth requirement is that contributions can be efficiently processed, verified and aggregated into a solution, and it refers to workflow activities 3 to 6 (automatic converting of image information, reviewing picture and text transcript quality, identifying the same or seemingly identical branded food products across countries, and aggregating data and adding food pictures to the database). Human action may be necessary at some point in the quality and aggregation process. However, if the required number of contributions (i.e. images) is vast, it may be costly and time-consuming for experts or peer crowd participants to transcribe texts, verify image and text quality, identify products and select product images for the final database. Indeed, it becomes clear from the initiatives discussed in Chapter 3 that these processes are among the most challenging and costly. Therefore, adequate, and possibly automated, processes are needed for crowdsourcing to work, which requires careful

([21]) https://digital-strategy.ec.europa.eu/en/policies/desi.

evaluation of current technologies and methodologies. Specifically, activity 3 (automatic converting of image information) requires the automatic conversion of image text into machine-readable text, for which technology is available. For instance, OCR is a widespread technology used to recognise text inside images, such as scanned documents and photos. Activity 4 (reviewing picture and text transcript quality) requires the verification of image quality (e.g. resolution, sharpness, content relevance). Recent studies suggest that machine learning and deep-learning techniques offer real-time opportunities for automatic image quality assessment (Dong and Tian, 2015; Hou et al., 2014). In DQ monitoring, a valid picture implies that the image is relevant (indeed, a branded food product) and of good quality (Dong and Tian, 2015) – good resolution, sharp, accurate colour, and adequate size and content – and it is therefore possible to convert the image to text accurately.

Furthermore, when reviewing the automatically transcribed texts (workflow activity 4), a validation task may be given to peer crowd participants (validators), who must confirm whether the input (image text) and output (text) match. Typically, crowdsourcing relies on redundancy, implying that the same task is given to several crowd members. The validation can occur by majority voting or a majority decision mechanism (most peer crowd participants agree that the image text and text match). This voting system is common in crowdsourcing. However, it requires motivating several crowd members to participate, which may be challenging and costly (it may require higher incentives than those given for submitting a picture), and majority voting does not consider the participants' diversity of expertise levels, which will turn into a problem if most are spammers (Nassar and Karray, 2019; Singh et al., 2021). To increase efficiency, the voting system could be automated by using redundancy. Redundancy means that the same task of photographing a specific branded food product in a given EU Member State is carried out by several contributors. This would enable a comparison of several transcriptions of images of the same product and validation when agreement between transcriptions is found (the majority are the same). Again, the drawback is that if most of the transcriptions are incorrect, the correct one will go unnoticed. Moreover, for comparison to take place, it needs to be possible to identify the same (or a seemingly identical) product sold in different countries (workflow activity 5). Accordingly, research shoudl be conducted on whether barcodes, GS1 and GTIN/EAN codes can help. The initiatives reviewed use these coding systems for product identification, but they are country specific. In addition, several software packages compare images' similarities, providing a percentage of similarity; these could be analysed in terms of their suitability for identifying different versions of the same product across countries. Finally, for aggregating data or finding the solution (workflow activity 6), when several pictures of the same product with precisely the same front of pack and the same list of ingredients are validated, the system could automatically select which picture to keep in the database, for example the one with the highest image quality according to specific criteria. Of course, to avoid unnecessary data collection and processing costs associated with redundancy, it seems necessary to determine a stopping criterion, namely collect only the number of photos of the same product necessary for the quality assurance process (Singh et al., 2021).

We can also apply the fourth requirement (contributions can be efficiently aggregated into a desired solution) to workflow activity 7 (dual food quality metrics). This means that achieving the objective of detecting DQ cases requires techniques that automatically assess the similarity of front-of-pack designs and lists of ingredients for versions of the same product. Recall that DQ occurs when two seemingly identical products (i.e. with similar front-of-pack designs) have substantially different ingredient lists. To detect cases of DQ, it is necessary to have sufficient valid observations for comparisons to be carried out. For this purpose, a sufficient number of observations means that photos of the front and back of pack of the same product have been captured in different countries and that the data collection covers many different products. Covering many different products ensures that the data collection covers a sizeable part of the market, and that the results are representative and meaningful. On the other hand, it may require specific requests to the crowd, possibly making the data collection more expensive, as the task of finding a specific product is no longer as simple (it possibly requires more time and effort). In conclusion, it appears that the fourth condition could be fulfilled by automating the quality assurance, aggregation and DQ detection processes through appropriate programmes and algorithms, but doubts are raised as to whether voluntary contributions can achieve the necessary product coverage in several countries to assess DQ.

Finally, the fifth requirement is that there are IT tools suitable for the task. We note that most crowdsourcing initiatives use devices and technologies that are readily available and potentially low-cost. For example, in the DQ task, crowdsourcing would rely on devices owned by crowd individuals. On the other hand, the platform (in the form of a mobile app or website) would rely on existing technology. However, workflow activities 3 to 7 related to the quality assurance process, data aggregation and calculation of DQ metrics require careful analysis to assess whether the available technologies and tools are suitable for them.

In conclusion, crowdsourcing may be a suitable method to monitor DQ in the EU cost-efficiently. However, it depends on whether it can engage the potential crowd to collect a sufficient number of valid observations, keeping costs low, and whether automatic approaches can be applied to transform, validate and select images, generate information and calculate DQ metrics.

Next, based on the review of existing initiatives and the literature, the following is an analysis of the design elements and mechanisms necessary for crowdsourcing to work at each data life cycle stage, starting with the stage we are at now, the initiation phase.

In the **initiation phase,** after setting the goal and rationale of the particular crowdsourcing approach (monitoring DQ in branded food products in the EU cost-efficiently), and envisaging the crowdsourcing task and workflow, it is necessary to establish how to manage the crowdsourcing in terms of the crowdsourcing task, the crowd, quality assurance and data aggregation, the incentive mechanism and the technology/platform. Therefore, we review these aspects, looking at key design elements and mechanisms.

Crowdsourcing task

Researchers in the field have proposed several design elements and mechanisms to consider when designing the task to be crowdsourced, which we apply to the objective of monitoring DQ by collecting branded food product pictures from the crowd.

- **Set a simple task.** 'Simple' implies that crowd participants can upload or submit pictures in seconds to a few minutes (not overcomplicated). The literature highlights that, other than rewards, one of the primary mechanisms for attracting participation – and thus ensuring that crowdsourcing works successfully – is reducing the complexity of the task (Cullina et al., 2015; Karachiwalla and Pinkow, 2021). Simple tasks attract more participants (Karachiwalla and Pinkow, 2021; Y. Wang et al., 2017). It has been noted that the level of complexity affects the effort and skills required to conduct the task and thereby impacts crowd motivation to engage (Karachiwalla and Pinkow, 2021). Therefore, the crowdsourcer should bear in mind that verification tasks and photographing specific branded food products instead of any branded food product would entail more significant effort on the part of the crowd.
- **Set a clear and well-defined task.** 'Clear' and 'well-defined' implies that it includes all relevant attributes (parameters to explain a task), such as descriptions – for example the type of food product to photograph (i.e. branded packaged foods), which parts of the package to photograph (front and back of pack, including ingredient list and nutritional facts) and deadlines, and not forgetting the rewards.
- **Break down complex tasks into simpler ones.** Consider modularity – that is, whether tasks can be conducted simultaneously and independently by different crowd participants.
- **Assess the level of complexity** (task granularity) to understand the required effort (e.g. time to complete the task, number of steps) and align incentives.
- **Decide on task deadlines** according to the targeted frequency of updates.

Crowd management

- **Understand crowd motivation.** For example, run a pilot at registration asking participants to choose from a list of motivations (e.g. monetary reward, fun, interest in the topic, social contribution) to fine-tune the incentive mechanism. Motivations can be intrinsic (personal enthusiasm, fun, social contribution or altruism) or extrinsic (monetary reward, reputation, information). Both types of motivations may prompt participation, and organisations pursuing crowdsourcing need to understand the crowd's motivation, as any organisation needs to know what motivates their employees to do a good job (Buettner, 2015; Nassar and Karray, 2019). Accordingly, they can plan for different incentive types, knowing that insufficient incentives may result in a drop in or low-quality participation. Employee motivation has always been a core problem for leaders and managers. More critical in crowdsourcing is understanding what kind of feedback or output of the collected information could motivate their participation. If the jointly produced information motivates them, this may positively affect cost efficiency and quality, as monetary rewards become less critical, and intrinsic and extrinsic motivation favours quality. Therefore, asking them when registering what kind of information (e.g. dual food quality cases, food information) they would like to receive in return could be relevant.

- **Understand crowd segmentation.** Understanding different segments of the potential crowd may be relevant to create more effective and differentiated incentive structures (Fedorenko et al., 2017). For example, younger participants may be more motivated by entertainment or financial incentives, while older participants may be more interested in information about food products and DQ.

- **Decide on the target crowd for collecting branded food product pictures to monitor DQ.** This should be based on the knowledge and skills required for the task. Accordingly, the target crowd can be the general public, as the task requires only having a smartphone and being able to take and upload photographs in a mobile app, following online instructions, which is nowadays commonplace.

- **Decide on the size of the crowd.** Crowd size will be related to the number of contributions required to ensure the quality of the solution – that is, to provide information on DQ cases. While the crowd is self-selecting, it is vital to set a goal for level of participation and contributions in order to establish the right strategy to achieve the solution and in order to assess the cost.

- **Decide on the most appropriate channels to publicise the crowdsourcing task.** These include word of mouth, social media, networking, dedicated websites, blogs, videos/digi stories, media such as press, TV or radio, and events, depending on the task and local context characteristics (e.g. use of social media and internet) (Dahlander et al., 2019; Pravst et al., 2021). Again, asking at registration how the potential participant heard about the DQ initiative can help fine-tune current and future advertising campaigns.

- **Decide whether to use intermediaries to broadcast the crowdsourcing initiative.** For example, some individuals and organisations, such as consumer organisations, can help promote the DQ crowdsourcing initiative and influence local decision-making when it comes to participating.

- **Establish minimal crowd pre-selection criteria.** This is to prevent poor-quality contributions (Kamoun et al., 2015) and can be done, for example, through automatic assessment of the quality of a picture taken by a candidate.

- **Plan to reject the participation of volunteers who consistently deliver poor quality.** This is to avoid data noise and wasting processing time and resources.

- **Create communication and feedback channels between the crowdsourcer and the crowd.** This is to motivate and keep the crowd engaged (Blohm et al., 2018; Karachiwalla and Pinkow, 2021). Think also of creating or turning the crowd into a community (i.e. give it a social identity) by encouraging conversation and interaction (e.g. through online chats) between crowd members. Crowdsourcing managers can leverage social identity to reinforce participant motivation to contribute (Fedorenko et al., 2017).
- **Create a support service for users (crowd participants and information users).** The service should provide help with any questions related to the use of the application, including technical and reward issues and issues related to the information provided.
- **Plan for a pilot phase.** A pilot phase offers a chance to fine-tune the task and processes (Kamoun et al., 2015).

Quality assurance and aggregation

Quality control is crucial during task design and runtime since errors can happen even when high-quality crowd participants are selected (Nassar and Karray, 2019). Moreover, even with high-quality tasks and runtime controls, low-quality contributions may occur as a result of unnoticed errors (unintentional errors or errors related to spammers). Therefore, appropriate verification and validation processes, as well as aggregation processes, must be established to provide the crowdsourcing solution.

Task-related recommendations

- **Plan for mechanisms to warn of possible quality issues during data entry.** For example, include certain limits or conditions (e.g. acceptable image quality) in the task to avoid collecting data or pictures that do not meet the quality threshold, implying a waste of resources.
- **Set a stopping criterion.** That is, when a specific product is available in the database of a given country, do not allow more pictures of that product to be contributed to avoid waste of collection and processing time and resources. Accordingly, raise the stopping criterion depending on the frequency determined for updating a product in the database.
- **Limit the number of accepted product pictures per person and per day/week/month.** This can help avoid fraudulent behaviour, such as repeated uploading of the same data or picture, and people making mistakes (Solano-Hermosilla et al., 2020, 2022).
- **Enhance product coverage.** For example, by suggesting, possibly automatically, to the crowd members specific products for which photos need to be uploaded (e.g. products for which there are not enough observations to ensure validation through the quality process and for which there are photos in other countries).

Quality process-related recommendations

- **Determine the number of crowd contributions needed (associated with the crowd size needed)** to monitor DQ successfully while avoiding processing more data and rewarding more contributors than necessary (Dahlander et al., 2019; Pedersen et al., 2013). This will entail deciding on the number of branded food products to be covered in each country and the number of pictures needed per product and country, bearing in mind that providing information on DQ cases requires the comparison of different versions of the same product across several countries. It will also entail deciding on the appropriate frequency of updates. For this purpose, the validation and aggregation method must be decided on. As discussed above, it should be taken into account that crowdsourcing often relies on redundancy and majority voting mechanisms (in this case, asking for pictures of the same product in the same country from different contributors and comparing them) to assess the quality and select the solution. To give an idea of the number of products that might be required, recall the LEDA database from Section 3.7; 100 000 branded food products are needed to cover 75 % of the Dutch market. Therefore, the market coverage goal can guide the decision about the number of contributions needed per country and in total, and enable costs to be estimated. Moreover, information about how many times on average in a year products change formulation can guide a decision on how often to collect information on a given product.

- **Decide on the processing, validation and aggregation methods**, their degree of automation, and the criteria for selecting the solution (i.e. the image and related information to be passed to the database). Given that crowdsourcing often relies on redundancy to assess quality and that human verification and aggregation is time-consuming and costly, it is recommended that the quality assurance and aggregation process be automated as much as possible, as the time and cost associated with manual work can potentially outweigh the expected benefits of crowdsourcing (Simperl, 2015). To this end, it would be necessary to assess current image recognition techniques used to convert image text into machine-readable text – such as OCR, machine learning and deep learning – for automatic image quality assessment, and algorithms for assessing the similarity of images and ingredient lists.

Quality performance recommendations

- **Plan for assessing whether the results – for example in terms of product coverage – are biased owing to characteristics of the crowd**, in order to take action to correct this. Examples of relevant characteristics are age, gender, education and geographical location.
- **Choose indicators to monitor the functioning of crowdsourcing.** For example, the number of registered participants, number of contributions, number of valid contributions, number of products covered, number of countries, number of products for which DQ metrics could be calculated (only possible when the same or seemingly identical product has been photographed in several countries) and number of user entries in the final database.
- **Choose indicators to monitor individual crowd member performance.** Examples of indicators are the number of pictures submitted and the number or percentage of accurate pictures submitted (Buettner, 2015).
- **Determine contributors' quality thresholds (e.g. number or percentage of valid pictures submitted).** This assists with the decision of whether to aggregate their contributions in the solution or block their participation if contributions are consistently below the thresholds.

Incentive mechanism

- **Provide fair incentives that align the incentive level with the required task effort.** Incentives should be linked to the level of task complexity (Kittur et al., 2011; Pedersen et al., 2013). For example, the crowdsourcing task of photographing branded food products using a mobile app is simple and expected to require little effort (seconds to a few minutes during regular food purchases); therefore, it could work with a minimal payment per picture, as seen in the examples of existing crowdsourcing platforms in Chapter 2 (ranging between EUR 0.01 and EUR 0.05 per picture). However, less simple tasks, such as photographing specifically requested branded food products and validating data, may require more substantial rewards according to the effort. Therefore, keep in mind that crowd members participate to earn a reward (expected financial or non-financial benefit) that is in line with the effort, skills and capabilities required (Bhatti et al., 2020; Sun et al., 2015; Zhao andand Zhu, 2014).

- **Bear in mind that micro-tasks for collecting data, such as uploading food pictures, mostly require a financial incentive**, even if low, **but non-financial incentives can contribute to increasing participation.** Consider including behavioural tools, such as nudges (Pedersen et al., 2013; Solano-Hermosilla et al., 2022) and game elements (Tinati et al., 2017). For example, nudges in the form of information about 'social norms' provided to the participants (e.g. average number of pictures provided by the other participants) and game elements, such as points, scoreboards or puzzle games, may help to increase participation.
- **Design participation in such a way that its outcomes are meaningful to the crowd participants.** This is essential (Fedorenko et al., 2017; Fung, 2015).
- Bear in mind that being aware of the existence of the crowdsourcing initiative is also essential for people to participate. This requires appropriate broadcasting campaigns that communicate the value of crowdsourcing for the crowdsourcer, the crowd and society (Dahlander et al., 2019; Fedorenko et al., 2017; Fung, 2015).
- **Decide on different incentive mixes for different segments of the crowd.** For example, differentiate between young and older crowd participants according to their different motivations.
- **Design effective feedback mechanisms.** They should intrinsically or extrinsically motivate crowd members and help increase trust, user self-competition and usefulness (Pedersen et al., 2013).
- **Plan for additional output (value) to be given to the crowd participants.** For example, information derived from the crowdsourcing data collection (Fedorenko et al., 2017). This can increase participants' extrinsic motivation (if they consider the information valuable for themselves) and their intrinsic motivation (if they consider the information as a common good and their contribution as a social contribution or altruism), while the monetary reward could become less critical (Blohm et al., 2018; Kamoun et al., 2015). Note that none of the initiatives reviewed in Chapter 3 mention paying volunteer participants for their contributions; rather, they are expected to contribute because they find the application useful.
- **Decide if potential contributors will be authorised to review or update peers' contributions and see the aggregated solution** as part of the value proposition for the crowd. Decide also how the information will be disseminated (e.g. website, static or interactive web dashboard).
- **Plan for learning from results, for example through assessment of incentive impacts, and accordingly adjust incentives.** As with crowdsourcing, it is easier to attract than retain crowd participants (Geiger et al., 2011; Solano-Hermosilla et al., 2022).

Technology

- **Choose or develop a user-friendly, robust (e.g. functioning 24/7) IT solution (e.g. device and platform) suitable for the task.** This will motivate participants (Davis, 1989).
- **Decide between using a marketplace IT platform or developing a proprietary platform.** The platform should be suitable for the particular crowdsourcing task (Modaresnezhad et al., 2020). According to the literature, a customisable marketplace IT crowdsourcing platform could be cheaper than developing a new one and may give easier access to a larger crowd. However, how customisable the platform is should also be examined, for example whether it has the functionalities required for the crowdsourcing

task of taking/uploading front- and back-of-pack pictures of branded food products to provide information on DQ cases, possibly during purchases at the supermarket (i.e. suits the task), and whether a logo can be included to create an identity for the initiative. The platform should also be well-evaluated in terms of how the raw output of the crowdsourcing task would enter into the quality assurance process for preprocessing of the images and image information, and selection of the image and image information to be included in the database for DQ assessment. Finally, how the output from the aggregation process and the DQ assessment should be disseminated, such as on a website or a web dashboard, should also be examined.

- **Assess available technologies to automatically process, validate and aggregate branded food product images.** Such technologies include OCR, GS1 barcoding (GTIN/EAN), machine learning and deep-learning approaches, and software for comparing image similarity and ingredient lists. For example, the initiatives reviewed identify barcode scanning technology as the optimal mechanism for product identification and interaction between the mobile app and the database (Dunford and Neal, 2017; Westenbrink et al., 2021). However, most are country specific, so it is not clear how it could work across countries; furthermore, decisions on the use of barcoding and the type used lie with the manufacturer.
- **Create trust by ensuring that the platform behaves as expected by the crowd participants** (including all functionalities), is robust (functions 24/7), addresses security concerns and complies with data privacy regulations. In addition, as mentioned above, a user support service can enhance the trust of the crowd and information users.

4.2.2. Crowdsourcing phase: implementing and managing data collection and processing

Next, in the crowdsourcing phase, data collection and processing is implemented and managed as designed during the initiation phase. It starts with launching the task by advertising the call for contributions and engaging a crowd to submit branded food product pictures via a mobile app. Then the crowdsourcer must monitor the response in terms of quantity and quality, and incentivise the response. The data processing mainly involves filtering out low-quality pictures and text transcriptions and aggregating the solution (selecting the picture to be included in the database). As crowdsourcing cannot be expected to work perfectly from the beginning, it is imperative to plan for flexibility and adjustment (Simperl, 2015).

Several essential aspects and mechanisms of the data collection and processing phase of crowdsourcing to monitor DQ cases are summarised below.

Crowdsourcing task

- **Run a pilot with limited participants, testing all the functionalities and processes** (e.g. registration, picture taking/uploading and quality controls during runtime.
- **Monitor task effort by measuring the time to complete the registration and the task** to help align incentives. Attracting people to register may need extra effort in terms of incentives.
- **Monitor recurrent errors and modify the task or tool if needed.**

Crowd management

- **Monitor mobile app participant registrations** and reinforce awareness campaigns if necessary and where necessary.

- **Monitor the success of implemented advertising channels** by means of a question in the registration form about how crowd participants heard about the initiative, as decided in the initiation phase (e.g. word of mouth, social media, networking, dedicated websites, blogs, videos/digi stories, media such as press, TV or radio, events).
- Monitor the number of **contributions and their quality (individual and overall)** and reinforce/adjust incentives and awareness campaigns if needed.
- **Block spammers** to reduce the waste of time and resources.
- **Provide feedback on the individual performance and the overall performance** to the contributors.
- **Assess the motivation to participate and stay engaged;** for example, a small registration and end-of-contribution survey could help (Harrington et al., 2021).
- **Remunerate valid contributions** based on the quality assessment / data-processing phase results.

Quality assurance and aggregation

- **Review limits and thresholds** (e.g. image quality, the stopping criterion, number of pictures needed to run the quality process).
- **Develop and implement appropriate protocols and quality assurance methods** (manual, automatic or a mix of both) for the crowdsourcing task (e.g. image and text quality filtering, product identification, picture selection, calculation of the front-of-pack similarity, list of ingredients similarity and the DQ indicator) as decided in the initiation phase. Accordingly, **ensure adequate capabilities** (e.g. image text recognition techniques, machine and deep learning, redundancy and majority voting, and image similarity detection).
- **Calculate performance metrics to monitor the performance of the data collection and the quality assurance process** (including accuracy and time and resources needed), and develop appropriate feedback mechanisms for the data collection (Cullina et al., 2015; Ford et al., 2015; Serret et al., 2019).
- **Consider the use of the most up-to-date image processing and quality assessment techniques.**
- **For data mining, develop automatic solutions as much as possible** that make use of recent emerging areas of artificial intelligence and machine learning (Chen et al., 2021; Yasmin et al., 2022).

Incentive mechanism

- **Consider adjusting incentives over time** to encourage and retain participation (Geiger et al., 2011; Solano-Hermosilla et al., 2022).
- **Test and assess different incentive types and levels** – monetary and non-monetary (information fed back to users, nudges, games) – to find out which work the best, as planned in the initiation phase. For example, use experiments that require designing, establishing ‘treatment’ and control groups to distinguish those people subject to a particular incentive and those not, and measuring. Incentive types and levels can be assessed in terms of the number of contributions and quality.

Technology

- **Make sure the platform is attractive and functional in terms of the task, and robust (functions 24/7).**
- **Be sure to implement secure data exchange and dissemination techniques** for smartphone-based applications (Bodkhe and Tanwar, 2021).
- **Maintain an effective support service.**
- **Consider recent technological developments in data-mining techniques and algorithms**, and digital payment platforms, which can make it possible to streamline and automate the data quality and payment processes in crowdsourcing (Harrington et al., 2021; Solano-Hermosilla et al., 2020; Zhou et al., 2018).
- **Ensure that privacy is preserved during data aggregation** in the processing phase in crowdsourcing; anonymisation techniques are usually proposed for this (Bayardo and Agrawal, 2005).

4.2.3. Dual quality detection phase: detecting possible dual quality practices

The third phase in crowdsourcing is developing the indicators from the processed (aggregated) data and analysis that can provide information on the issue in question, in this case detecting potential DQ cases. It is a crucial stage for gaining insights from the data. This phase is less studied from a crowdsourcing perspective, since it is common for any system of data collection to shed light on a specific topic. However, it is imperative to ensure that crowdsourcing offers value.

Usually, the data produced from the quality assurance and aggregation phase (or data-processing stage) are stored in back-end database systems and used by the crowdsourcer or other stakeholders to obtain insights into the topic of interest. Statistical functions are the main tools used for this (Lian et al., 2021). For example, most of the initiatives reviewed in Chapter 3 use a rating system linked to food composition, assessing nutritional aspects and ingredients. Crowdsourcing the front- and back-of-pack images of branded food products to detect potential DQ branded food products in the EU offers several possibilities for reporting information, in addition to reporting on DQ cases..

To this end, indicators of DQ (comparing the degree of similarity of front-of-pack designs and ingredient lists usually contained on the back of pack of different versions of the same product across EU countries and indicating when two versions have similar front-of-pack designs but differ (significantly) in composition) must be assessed (possibly with the help of automation). This stage may involve setting thresholds for picture and ingredient list similarity. For example, available software that can be used to assess image similarity generally provides a similarity percentage (Hammond et al., 2020), usually suggesting that pictures are similar if the percentage is 90–100 % and that they are different if it is less than 70 %. However, these percentages will need careful assessment in this context, particularly in the case of front-of-pack images of branded food products. Regarding differences in composition (e.g. in nutrition declarations and ingredients), researchers could decide whether to use a Boolean (true or false) indicator or a percentage indicator. The approach applied in the Joint Research Centre common methodology could be used to identify differences in composition (European Commission, 2018; JRC, 2019; Nes et al., 2023).

The results could also be used to provide information about products and their composition across EU countries by making the pictures and information available in a user-friendly and navigable way. As mentioned before, feeding back such useful information to the crowd could provide the most substantial incentive for participation.

4.2.4. Use of information phase: publishing and disseminating data

The fourth phase concerns the dissemination of information and its use. The dissemination stage in crowdsourcing involves (1) the storage of the crowdsourced information and (2) dissemination of the information – that is, getting the information and findings from the crowdsourced data collection to the crowdsourcer and the various stakeholders who can use it to maximise the benefit of the crowdsourcing initiative without delay. At this stage, data security issues must be considered, including confidentiality and privacy, and secure content distribution to prevent cyberattacks (Bodkhe and Tanwar 2021; L. N. Ferreira and 2016; Matheus, Janssen, and Maheshwari 2018).

This phase is less studied from a crowdsourcing perspective, since it is common to any data collection system. Importantly, when disseminating information, it is essential to decide on the frequency of publication and to accompany the information with appropriate metadata and documentation on methodology and quality. In addition, it is important to choose the proper format; otherwise, information may remain unused (Solano-Hermosilla et al., 2022). Information can be disseminated via publications and dedicated websites that provide static or interactive visualisations (i.e. interactive online dashboards), among other options (Eurostat, 2020). Matheus et al. (2018) suggest that interactive dashboards enable flexible visualisation of consolidated datasets for a particular purpose, supporting policymaking and stakeholders' decision-making, and communicating and interacting with the public. The initiatives reviewed in Chapter 3 use interactive web pages or dashboards that allow selection by product and category, for which different indicators are displayed, depending on the initiative. It could be a suitable option for DQ monitoring. For each selected product, it seems necessary to show the various front-of-pack versions captured by the crowd in different countries, accompanied by their ingredient lists and DQ indicators (as decided in the DQ detection phase) that signal differences in ingredients for product versions with the same or very similar front of pack.

Finally, data usage refers to using crowdsourced data/information to inform policy and decision-making processes and actions. It is crucial, as the value of crowdsourcing lies in the data being used (Janssen et al., 2014). To achieve this, organisations and individuals must develop adequate capabilities to access and use information. However, despite the potential benefits of crowdsourcing, there has been less progress concerning its policymaking and decision-making support. Importantly, the use of data requires the involvement of the potential users from the beginning, which, in the particular case of DQ monitoring, may be policymakers, consumers and consumer organisations, and possibly food companies.

Various indicators could be selected to measure the use of the data. For example, tracking the number of visits to a website or web dashboard can give an idea of the extent of data use. In addition, tracking related policy or business actions can give an idea of the impact.

5. Advantages and disadvantages of using crowdsourcing to monitor dual quality

As crowdsourcing is used in an increasing number of areas, it is growing as a tool for outsourcing to the people the mass photographing of branded food products to build databases to assess nutritional and food composition aspects, and also support policymaking and consumer healthy choices (Harrington et al., 2021; Martin et al., 2008, 2014; Pravst et al., 2021). Examples of this are several of the initiatives reviewed in Chapter 3, such as the FoodSwitch app, the Open Food Facts – World database, and the Yuka and 'Veš, kaj ješ?' (#VKJ; You know what you eat) apps. We consider them to be crowdsourcing initiatives because of their use to collect data (i.e. pictures) from users, although they may be complemented by other forms of data collection. In addition, the ECO project interactive platform and the Lebensmittelklarheit initiative can also be considered a type of crowdsourcing, where citizens can raise issues or complaints about certain food aspects electronically and help track the complaint history. In addition, Chen et al. (2021), Harrington et al. (2021), Kawano and Yanai 2014) and Noronha et al. (2011) provide several examples of crowdsourcing for building food image datasets, although their purpose is not to monitor but to conduct a concrete study.

In this section, based on the literature and initiatives reviewed, we analyse the advantages and disadvantages of a typical crowdsourcing approach to monitoring DQ cases in branded food products in the EU. Accordingly, the task of photographing branded food products (front and back of pack, including the list of ingredients and nutritional facts) to detect DQ cases is given to the general public to perform online for an expected financial or non-financial benefit (e.g. fun, altruism, social contribution, information). Although the crowdsourcing initiatives reviewed in Chapter 3 of this study do not provide financial incentives, the literature does point out that these types of simple crowdsourcing tasks or micro-tasks are usually remunerated with a micro-payment. For example, a study by Harrington et al. (2021) to capture real-time eating behaviours through food images does so. Note also that the LEDA approach reviewed in Section 3.7 differs from the typical crowdsourcing approach in that specific agreements to provide data are signed with food companies, covering 75 % of the branded food product market, assuring a priori commitment, representativeness of the data and data quality. Food companies are therefore expected to be willing to ensure the quality of their product images to promote or not damage their brand image commercially. However, in crowdsourcing, quality (in the number and distribution of contributions and accuracy) is typically uncertain a priori due to a lack of knowledge about the crowd, its engagement, and its diversity of knowledge, skills and motivations. Usually, it can only be established *ex post* and must be well motivated.

According to the reviewed literature and online initiatives collecting branded food product images (mainly to provide information on nutritional aspects), the crowdsourcing process offers clear advantages, such as the potential to reach a large pool of people to provide branded food product pictures, with minimal effort required of participants, at low cost and in real time, and involving people interacting and possibly community creation. However, achieving the benefits comes with disadvantages, difficulties and challenges (e.g. Blohm et al., 2018; Buettner, 2015; Fedorenko et al., 2017; Liu et al., 2021; Pravst et al., 2021).

As discussed previously, the DQ task fulfils the fundamental requirements for crowdsourcing to be a suitable method and achieve its advantages (Afuah and Tucci, 2012). One requirement is that a large amount of data and distant knowledge (i.e. geographical) is needed, which is the case if a considerable part of the market for branded food products in all EU Member States is to be covered. A further requirement is that there is a crowd for it. Citizens represent an ideal crowd to accomplish the task with little effort during regular purchases, as photographing and sharing photos via an app is commonplace and simple. Task simplicity is also an essential requirement for crowdsourcing to work successfully. Another condition is that there are efficient methods to process the images into quality information. Today's text recognition technologies in images and machine learning offer significant opportunities to create efficient processes based on automatic algorithms that potentially require little manual intervention. Finally, the technology is available, such as mobile applications and websites, and is accessible to most to transmit the task (the crowdsourcer) and perform it (the crowd).

In this context, we discuss the advantages and disadvantages of crowdsourcing to monitor DQ, which are summarised in Table 2 and Table 3. It is worth mentioning that a crowdsourcing initiative is like any project, whether business or social, in that all the challenges and risks must be assessed, as well as the advantages.

Advantages

Reaches a large number of people who own a smartphone and are ready to contribute, via a mobile app, images of branded food products at any time during regular food purchases, with minimal effort required of participants. Monitoring DQ in the EU would require a large number of contributions to cover a representative sample of the same branded food products in each EU Member State to enable cross-country comparison of the same product's composition and front of pack. However, while reaching a large number of people is one of the main advantages of crowdsourcing, crowdsourcing initiatives and mobile apps often suffer from a lack of participation and engagement with people. Therefore, the main challenge in crowdsourcing is effectively activating people's motivation to engage them and sustain their participation.

From the analysis of the existing initiatives collecting branded food product images, it becomes clear that the primary motivation for participating for consumers (the crowd) is using the results. The initiatives thus offer value propositions highlighting the support to consumer decisions (concerning health, mainly) and policymaking that the results will provide. Therefore, it appears to be essential to have and transmit a value proposition that includes the benefits to the crowd and society in order to engage participants effectively. This study has revealed several mechanisms and tools to motivate voluntary participation in crowdsourcing, and has found that communication and marketing/awareness campaigns are essential. Particularly effective could be partnering with consumer organisations as intermediaries to promote the initiative. They have contact with consumers and the capacity to influence participation decisions. Monetary incentives may be a further option, as seen in the crowdsourcing tasks distributed by MTurk in the study of Harrington et al. (2021) assessing eating behaviours from people's pictures and using behavioural tools (e.g. nudges and game elements). Moreover, it should be noted that simple tasks are more appealing to the crowd. It should also be noted that the decision to use a proprietary or a marketplace platform may significantly influence costs and people's participation, and requires careful assessment. Well-known marketplace platforms may be helpful for reaching a large crowd, but it may not be possible to build into them all the functionalities required. Whether or not they succeed in creating an identity will depend on how customisable they are, for example when it comes to including logos and slogans.

Finally, given the participation potential and challenges, it seems particularly important to run a pilot to assess the crowd's motivations and the dissemination/marketing channels that work best using brief surveys of participants and experimental designs to understand which incentives work best. The latter is essential, as the literature suggests that surveys may underestimate financial incentives.

Access to diverse and dispersed (e.g. geographically) knowledge, covering different countries. This is a crucial advantage of crowdsourcing when it comes to monitoring DQ in the EU, which requires contributions from many countries. However, whether it is used as an advantage depends again on how compellingly the crowdsourcer motivates people to participate. Moreover, it may be necessary to use tailor-made motivational tools for each country, representing an added difficulty.

Collects real-time, up-to-date data. A large and diverse crowd provides, through collaboration and participation at different times, the opportunity to update the dataset frequently. However, the challenge here, again, is people's regular and sustained participation.

Creates a community around the topic. Creating an online community can be an advantage of crowdsourcing by helping to share and generate knowledge and, in turn, can be very beneficial in encouraging participation; however, online communities also face the difficulty of no active participation. Often only a few people are active, requiring motivational tools to be put into practice.

Saves on the cost of data collection, compared with hiring experts. This is one of the most frequently noted advantages of crowdsourcing. The assumption is that the crowd members receive only minimal payment or that their motivation is interest in the outcome (personal, reputational or social) rather than payment. Indeed, the initiatives reviewed do not mention payment to crowd contributors, and people seem to contribute because they are interested in the results. However, if payment is necessary, the amount and the number of contributions needed to keep the initiative economically viable must be carefully assessed. Therefore, it is essential to establish how many observations are required to assess the extent of DQ in the EU and for the crowdsourcing quality process to work successfully, given the potentially low quality of non-professional contributions. As mentioned above, it could help significantly to run a pilot assessing different levels of financial incentive (including the option of no financial incentive). At the same time, enhancing the value proposition, including the benefits to the crowd other than financial benefits, also seems to be crucial.

Saves on the cost of data processing through automatic quality assurance and aggregation processes. In crowdsourcing, contributions delivered by the crowd must be checked for quality, since people with unknown and varied skills and motivations produce them. Tools such as automated majority voting mechanisms, image text recognition software and machine learning approaches offer solid potential to automate the quality assurance and aggregation process, which is essential to harness the potential of crowdsourcing, given the expected huge number of observations. However, implementing such automated processes effectively so that they work as expected is a challenge that requires robust statistical, data and computer science capabilities and testing; a pilot data collection can be beneficial here. The online initiatives reviewed in this study illustrate the challenge, with several indicating the need to implement both automated and manual processes, which are resource- and time-intensive. For example, one of the initiatives shows how data extraction can be done through an online application with the support of OCR technology, but also suggests that manual work is necessary for cross-checking.

Table 2. Advantages of using crowdsourcing to monitor DQ, the challenges and how to address them

Advantages	Challenges	How to manage
Reaches a large number of people who own a smartphone and are ready to contribute images of branded food products at any time during regular food purchases	Effective activation of people's motivation to participate and sustain their participation	• Set a simple task • Effective marketing/awareness campaign • Value proposition, including the crowd's benefits • Extrinsic incentives (monetary, reputation, information) • Intrinsic incentives (nudges, game elements, social contribution) • Partner with local consumer organisations, supermarkets • Foster interaction among people (e.g. via chat) • Use online events and social media (e.g. LinkedIn to raise awareness) • Thoroughly assess whether to use a marketplace or a proprietary platform • Run a pilot to test the task and the whole process, and assess awareness-raising mechanisms, motivations and different types of incentives
Access to diverse and dispersed (e.g. geographically) knowledge, covering different countries		
Collects real-time, up-to-date data		
Creates a community around the topic	No active participation from crowd participants	Communicate, create newsletters, create a hashtag, use social media and game elements (e.g. competition) and find leaders
Saves on the cost of data collection, compared with hiring experts	Increasing monetary incentives could make the initiative unviable if a very large number of pictures is required	• Enhance the value proposition and include non-monetary incentives • Run a pilot to assess different incentive levels
Saves on the cost of data processing through automatic quality assurance and aggregation processes (e.g. automated majority voting mechanisms, image text recognition tools, machine learning approaches)	Effective automation of most of the data processing, including transformation, cleaning and aggregation, with appropriate methods and tools	• Need for capabilities in statistics, data and computer science • Run a pilot to test the methodological quality approaches

Source: Author

Disadvantages

Crowdsourcing is a convenience sampling method (using a non-probability sampling approach). Convenience sampling is a non-probability sampling method where the sample is taken from a group of people easy to contact or to reach (e.g. people making regular purchases), and consequently, people select themselves to be part of the sample. As a result, the data collected may not accurately represent the population. On the positive side, using convenience samples requires fewer resources and less time than collecting data from full population surveys. The drawback is that the inferences from these convenience samples are usually biased and probably do not accurately reflect the extent of the phenomenon or its policy impact on the general population. Therefore, information users must be careful when generalising results. However, convenience samples may help in monitoring a phenomenon when the cost of traditional data collection approaches is prohibitive and may be used to supplement specific studies following traditional data collection. The way to reduce bias is to enhance participation and contributions.

Trust and data usability concerns. Trust issues may lead to unused data, which could significantly reduce participation if the data are considered unreliable and not used. Therefore, it is recommended that a robust quality methodology be developed and published to increase trust. Moreover, trust concerns regarding the crowdsourcing tool can be addressed by piloting the crowdsourcing task and all processes in advance to ensure everything works as expected. Otherwise, it could generate distrust and weariness among the participants.

Communication and management of a large number of anonymous crowd members. This may be more complex and costlier than coordinating a few hired data collectors. Therefore, it is important to develop effective and efficient communication channels that enable unidirectional (e.g. suggestion boxes, SMSs), bidirectional (e.g. email) and multidirectional (e.g. chats) communication. In turn, good communication is essential to enhance crowd motivation.

With no contract, crowd members can decide at any time whether to participate or not, making it challenging to guarantee participation. With no contract, it is not easy to know a priori how many people will participate, and how many products and how many countries will be covered, as people can decide to participate or not at their own convenience. This difficulty can be faced by building a vast crowd, making it possible for people to take turns to keep the data up to date. Building a large crowd is associated with incentives and communication.

Confidentiality issues. It has been pointed out that uncertainty about confidentiality could discourage participation. It is therefore imperative to establish protocols and processes to ensure confidentiality and to make this known to participants.

Ethical considerations. Some authors argue that crowdsourcing implies exploiting cheap or unpaid labour, where participants are not hired, and and that through crowdsourcing rewards organisations avoid contributing to social security costs or statutory minimum wages

Other authors suggest that it is important to consider crowdsourcing as part of the collaborative economy, arguing that it ought to deliver not only economic value for the crowdsourcer and society but also value of some other type for the crowd and society.

Costly data collection if non-monetary benefits are not provided. As discussed above, relying primarily on monetary rewards to motivate the crowd to contribute can substantially increase costs. Costs may considerably increase if the crowd's participation is incentivised solely by monetary rewards. For example, in a hypothetical scenario in which the reward offered to a consumer increases from EUR 0.01 to EUR 0.5 per photo taken, the cost will be 50 times higher. Therefore, depending on the number of photos involved, the crowdsourcing initiative could prove economically unviable. When considering the previous example (and following the LEDA database example), where 100 000 products are needed in each EU Member State, and, for each product, a minimum of three photo deliveries updated twice a year are required (implying 600 000 photo deliveries per EU Member State and year or 16 200 000 photo deliveries in the EU per year), the total yearly costs would increase from EUR 162 000 for a reward of EUR 0.01 per picture to EUR 8 100 000 for a reward of EUR 0.50 per picture ([22]). Therefore, a key question is whether and how much monetary reward is needed. Substituting monetary rewards (at least partially) with other crowd benefits can play a critical role in reducing costs.

([22]) Note that these costs do not include other expenses related to the development of crowdsourcing mobile apps or conducting awareness campaigns.

Table 3. Disadvantages of using crowdsourcing to monitor DQ and how to manage them

Disadvantages	How to manage
Convenience samples that possibly lead to data biases	Enhance participation
Trust and data usability concerns	• Publish methodology • Pilot and test all tasks as processes to ensure they work as expected
Communication with and management of a large number of crowd members may be more complex and costlier than coordinating a few hired data collectors	Develop effective communication channels
With no contract, crowd members can decide at any time whether to participate or not, making it challenging to guarantee participation	• Build a large crowd • Incentives and communication
Confidentiality issues	Develop adequate processes and protocols
Ethical considerations	Apply collaborative economy concepts to ensure the value proposition includes crowd participants
Costly data collection if non-monetary benefits are not provided	Provide non-monetary benefits

Source: Author

It can be concluded that the advantages or disadvantages depend on the particular task and its management and implementation. For example, whether cost efficiencies can be achieved will depend on the number of contributions required to assess DQ in the EU, whether a financial incentive is needed, or whether the task and its outcome sufficiently motivate participants. Another example is the following: whether a convenience sample is a disadvantage or not depends on the task. Specifically, in the case of monitoring DQ, it is essential to collect information on many branded food products in many countries; if this is achieved, it may not matter as much if the sample is a convenience sample.

The task of monitoring DQ seems to be well-suited to take advantage of the benefits of crowdsourcing. However, it requires addressing the important challenges of motivating participation while keeping costs down and effectively developing and implementing automatic data transformation, verification and aggregation processes that enable a large amount of data to be processed. These two challenges suggest the need for marketing, people (HR) management, behaviour, communication, statistics and data and computer science capabilities. All appear to be essential capabilities for crowdsourcing, situating crowdsourcing at the intersection of various knowledge areas and not only in information systems. Moreover, a pilot to assess the motivation and process challenges and the effectiveness of the mechanisms to manage them would help assess the feasibility further. Appropriate methods and experimental designs should be used in the pilot.

Nassar and Karray (2019) provide an important clue that is helpful in assessing the feasibility of crowdsourcing for monitoring DQ. They suggest that crowdsourcing is appropriate for tasks that need human intelligence rather than machine intelligence. For example, for monitoring DQ through the collection of photos of branded food products, one might think of that an automatic process for collecting photos from the internet (i.e. web scraping) would be more efficient. The disadvantage is that web scraping is a type of unsolicited crowdsourcing that relies on what is published on the web. On the other hand, e-commerce is commonplace in all EU Member States, suggesting that it would be possible to obtain photos and lists of ingredients from websites. One might even think about combining the two types of crowdsourcing in the data collection part, and passing the photos and data through the subsequent quality process independently of the origin of the photos.

6. Conclusions

The overall purpose of this study was to assess the feasibility of crowdsourcing for collecting branded food product pictures and data from citizens using a mobile app to provide information on and monitor DQ cases in the EU, which occur when brand owners sell different versions (i.e. with a substantially different composition) of the same product in different countries with the same or similar front-of-pack design. To this end, we first analysed the literature to understand the crowdsourcing process and its key components, the benefits, costs and challenges, and how it compares with traditional data collection methods. Second, we looked at current online initiatives tracking food standards and safety through citizen contributions, some of which have crowdsourcing functionalities to collect food product pictures, in addition to other data-sourcing methods. Third, based on the literature and initiatives reviewed, we discussed the possible application of crowdsourcing to DQ monitoring in the EU, explaining the importance of defining the crowdsourcing task, describing the processes and phases involved in crowdsourcing, and specifying the main aspects and mechanisms to be taken into account in each phase to make the method work. Finally, based on the analysis carried out when reviewing the literature and existing initiatives, we discussed the advantages and disadvantages of using crowdsourcing to monitor DQ by processing photos of branded food products (front and back of packs, including ingredient lists) collected by citizens across the EU using a mobile application, and provided several recommendations.

Crowdsourcing can be defined as a virtual sourcing method for obtaining information or a solution to a specific problem achieved by distributing an online task to a pool of people (the crowd), leveraging the crowd's knowledge. Accordingly, in crowdsourcing, there is a crowdsourcer or requester – the person or organisation that launches a call to outsource a task to the public (the crowd) to achieve a particular goal. This goal translates into specific tasks that the crowd is invited to undertake using an online platform, serving the exchange between the crowdsourcer and crowd member. Crowd contributions are quality assessed and aggregated into results that can be used by the crowdsourcer and disseminated to the crowd and other stakeholders for their use. Crowd members' participation will depend on their motivation and the incentive mechanisms put in place by the crowdsourcer.

The following five success factors indicate, from a technical point of view, the suitability of crowdsourcing for monitoring DQ in branded food products:

1. the task is simple, modular and easily transferable;
2. a vast amount of contributions are required, as is distant search (i.e. good geographical coverage);
3. there exists a potential crowd to contribute;
4. contributions can be efficiently verified and aggregated into a desired solution;
5. there are available IT tools that are suitable for completing the task.

Crowdsourcing for monitoring DQ largely fulfils these requirements. It involves the crowdsourcing task of photographing branded food products (front and back of pack) with a mobile app, which is a rather simple activity (in principle, it requires seconds to a few minutes). Moreover, it can be made modular (different contributions can occur simultaneously and independently) and is easily transmittable to the public. The second requirement is also met. A vast amount of contributions and distant search are needed to monitor DQ; the crowdsourcing must cover many different branded food products in all EU Member States. Taking photos with smartphones and sharing them is commonplace, and most people in the EU have smartphones and mobile internet; hence, the third requirement is also fulfilled. The fourth requirement is among the most challenging and costly. This requirement can potentially be met by deploying adequate, and possibly automated, processes for data processing, verification and quality checking, aggregation of contributions, and DQ detection processes. However, human action may be necessary at some point in the quality checking, aggregation and DQ detection processes. Finally, the fifth requirement is that IT tools that are suitable for completing the task are available. Based on the experiences of the existing online initiatives tracking food quality and safety using citizen contributions, devices and technologies relevant to crowdsourcing for DQ monitoring seem to be readily available and potentially at low cost.

A basic representation of the crowdsourcing framework for data life cycle management to provide information on DQ comprises four phases.

- The **initiation phase** involves setting the goal and the task to be crowdsourced. In this phase, the crowdsourcing organisation (crowdsourcer) needs to (1) state the rationale for crowdsourcing and ensure that it aligns with the goal, (2) decide on the specific crowdsourcing task to be allocated to the crowd and (3) decide on how to manage the crowdsourcing – that is, design and implement processes and methods, choose the platform and technology, create an incentive mechanism and assess risks.
- The **crowdsourcing phase** involves implementation and management of the data collection and processing by (1) advertising the call for contributions to engage a crowd, (2) monitoring the response in terms of quantity and quality and aggregating information, and (3) incentivising contributions and motivating the crowd.
- The **detection of DQ phase** involves assessing the presence of DQ.
- The **use of information phase** involves the publication and dissemination of the information obtained so that it can be used.

Furthermore, it is essential to consider all the actors involved in the data life cycle management. These are primarily the crowdsourcer, the crowd and the stakeholders. The crowdsourcer is the organisation responsible for monitoring DQ cases using crowdsourcing, which could be a public authority, an NGO, a consumer association or a private company. The crowd is the members of the public who voluntarily contribute their skills, time and/or resources to complete a task (e.g. providing front- and back-of-pack photos of branded food products). Finally, the stakeholders are the potential information users, such as policymakers, consumers and companies.

Overall, the advantages and disadvantages of applying crowdsourcing to monitor DQ can be summarised as follows.

Advantages

- It involves gaining access to a large number of people who own a smartphone and are prepared to contribute, via a mobile app, images of branded food products at any time during regular food purchases, with minimal effort required of participants.
- It provides access to diverse and dispersed (e.g. geographically) knowledge, covering different countries.
- It collects real-time, up-to-date data.
- It creates an online community around the topic, which can encourage participation.
- It may save on the cost of data collection, compared with alternative approaches (e.g. hiring experts).
- It may save on the cost of data processing by applying automatic quality assurance and aggregation processes.

Disadvantages

- Crowdsourcing is a convenience sampling method (using a non-probability sampling approach), as a result of which the data collected may be non-representative.
- Trust and data usability concerns may emerge, which could significantly reduce participation.
- Communication with and management of a large number of anonymous crowd members may be costlier than using alternative approaches (e.g. coordinating a few hired data collectors).
- With no contract, crowd members can decide at any time whether to participate or not, making it challenging to guarantee participation.

- Confidentiality issues may emerge, which may discourage participation.
- Ethical considerations may emerge.
- Data collection may be costly if the crowd is offered only monetary compensation for contributions. Non-economic benefits need to be made clear.

Discussion

The reviewed literature and existing initiatives tracking food quality and safety issues using citizen contributions show that crowdsourcing can bring together a large group of participants on the same platform to contribute for a purpose that will eventually benefit them all. The systematic use of food product images, often referred to as the 'remote food photography method', to assess nutritional and compositional food aspects has recently received significant attention (Pravst et al., 2021). Moreover, smartphone-based crowdsourcing provides a potential platform for engaging a large number of people (the crowd) to solicit contributions by carrying out a simple micro-task, such as photographing and sharing pictures of branded food products during regular purchases at people's own convenience. However, for crowdsourcing to succeed in providing information about DQ across the EU, it requires contributions covering, in each country, a high number of branded food products, to ensure representativeness and enable comparison across countries of products' ingredient lists and front-of-pack designs. Recall that the LEDA database collects data on 100 000 food products, covering 75 % of the Dutch market. If it is assumed that a similar number of products needs to be covered in each EU Member State, the total number of products and observations involved seems considerable.

Crowdsourcing for data collection appears to be financially most effective in cases where the contributors are paid minimal rewards or are not financially rewarded, but instead the crowd benefit is using the results produced collaboratively. If the reward were a more sizeable financial one, it could make the crowdsourcing initiative economically unviable. On the other hand, if there were no budget limitations, experts or other workers could be hired, setting the quality in terms of quantity, accuracy and geographical distribution a priori. The initiatives reviewed refer mainly to the support they expect to provide to consumer decisions on food purchases, primarily health-related, and to policymaking, as the primary motivation to participate. Note that none refers to paid rewards. This implies that the crowdsourcing initiative must include a value proposition that states benefits for the crowd, and possibly society, other than financial remuneration. Therefore, finding, developing and communicating the benefits that the crowd can expect from the DQ monitoring tool is essential.

Moreover, unlike traditional resource management, where the resources are known and the allocation of tasks is controlled, in crowdsourcing the organiser does not select the crowd; it is self-selected, and therefore crowdsourcing results in a convenience sample that requires caution when generalising findings but may be suitable for specific tasks. Furthermore, there is no contract. These are fundamental differences between crowdsourcing and traditional forms of data collection using experts or other hired workers that imply important challenges regarding crowd motivation and coordination of the process.

Crowdsourcing difficulties and challenges mainly involve motivating sustained participation and issues with quality, quality assurance and data aggregation. For example, according to a recent study, the lack of continuity in using mobile apps is dramatic; 95 % of downloaded apps are no longer used within a month (Gu et al., 2022). This issue can be addressed by adequately understanding and managing people's motivations, aligning the value proposition of the crowdsourcing initiative and incentives (extrinsic and intrinsic), and using behavioural science tools (e.g. nudges and game elements) to enhance motivation while keeping costs down. It should be noted that simple tasks appeal more to a crowd, so it is advisable not to overcomplicate the task. Furthermore, encouraging motivation to participate in a DQ monitoring initiative would require adequate marketing/awareness channels, communication and possibly partnering with local organisations, such as consumer associations. Furthermore, all these activities, plus the development of the online platform, would require a budget, and the upfront costs may be high and should not be underestimated. The costs may be affected by the choice between a customisable online platform from the market (less expensive) and proprietary development (more expensive). However, a customisable platform may not provide enough flexibility to implement all required functionalities, and the choice must be made bearing in mind that trade-off.

Moreover, the quality issue can be addressed by coordinating the use of proper processes and tools. Quality can be controlled in task design and during data entry (e.g. automatic image quality checks attending to resolution, sharpness and content relevance). However, even if the task is well designed and data entry controls are in place, mistakes can happen, and it is imperative that adequate methods and techniques (e.g. OCR, machine learning) are used to develop automatic quality assurance and aggregation processes to harness the potential of crowdsourcing. The initiatives reviewed highlight that this can be a bottleneck, as many combine both automatic and manual processes, which can be very time- and resource-intensive. It is also imperative that automatic methods of calculating DQ indicators – signalling the existence of DQ, with versions of the same product with similar front-of-pack designs and significantly different composition being marketed in different countries – are developed. Furthermore, one cannot expect crowdsourcing to work perfectly from the beginning. Therefore, developing success metrics such as the number of complete and valid observations and the extent of product and regional coverage may help in assessing the initiative's progress on monitoring DQ and indicate whether there is a need to start corrective action. All these challenges involve a range of capabilities that crowdsourcing organisations need, such as knowledge of marketing, communication, crowd management (HR), economics, statistics and data, and computer sciences, placing crowdsourcing challenges at the intersection of several knowledge fields in addition to information systems. Further research and implementation of such strategies are needed to establish a crowdsourcing model for monitoring DQ. A pilot is recommended to assess awareness channels, motivation and the effectiveness of different types and levels of incentives by including participant surveys and robust experimental testing. The pilot would also test the methodology and the quality process, considering the possible level of automation. Finally, a suggested additional research line is testing web scraping as a type of unsolicited crowdsourcing, based on the idea that a crowdsourcing model that distributes a task to people is suitable mainly for tasks that need human intelligence or contribution rather than machine intelligence. In this context, photo web scraping might be an automatic option, based on capturing what is available on the internet. This could be helpful, since e-commerce for food products is widely implemented in the EU, and pictures are available on the internet (even if, in some cases, not of the most recent version of the product). More importantly, it could be a promising method to complement solicited crowdsourcing, with photos being passed on to the quality process regardless of their origin. Finally, from a methodological perspective, it can be concluded that analysing the feasibility of a potential initiative is more challenging than assessing initiatives already in existence. However, the study provides important insights for data collectors, managers and practitioners aiming to use crowdsourcing to collect branded food images to monitor DQ in the EU.

References

Adewopo, J. B., Solano-Hermosilla, G., Colen, L., and Micale, F. (2021). Using crowd-sourced data for real-time monitoring of food prices during the COVID-19 pandemic: Insights from a pilot project in northern Nigeria. *Global Food Security*, *29*, 100523.

Afuah, A., and Tucci, C. L. (2012). Crowdsourcing as a solution to distant search. *Academy of Management Review*, *37*(3), 355–375.

Aluchna, M. (2018). Sustainable business models: The case of the collaborative economy. In *Sustainable business models* (pp. 41–62). Springer.

Arbia, G., Solano-Hermosilla, G., Micale, F., Nardelli, V., and Genovese, G. (2018). *Post-sampling crowdsourced data to allow reliable statistical inference: the case of food price indices in Nigeria*. http://meetings3.sis-statistica.org/index.php/sis2018/49th/paper/viewFile/1090/64

Arbia, G., Solano-Hermosilla, G., Nardelli, V., Micale, F., Genovese, G., Amerise, I. L., and Adewopo, J. (2023). From mobile crowdsourcing to crowd-trusted food price in Nigeria: statistical pre-processing and post-sampling. *Scientific Data*, *10*(1), 446.

Barbier, G., Zafarani, R., Gao, H., Fung, G., and Liu, H. (2012). Maximizing benefits from crowdsourced data. *Computational and Mathematical Organization Theory*, *18*(3), 257–279.

Bayardo, R. J., and Agrawal, R. (2005). Data privacy through optimal k-anonymization. *21st International Conference on Data Engineering (ICDE'05)*, 217–228.

Beręsewicz, M., Lehtonen, R., Reis, F., Di Consiglio, L., and Karlberg, M. (2018). *An overview of methods for treating selectivity in big data sources*. Statistical Working Papers. Eurostat. Doi: https://doi.org./10.2785/312232.

Bhatti, S. S., Gao, X., and Chen, G. (2020). General framework, opportunities and challenges for crowdsourcing techniques: A comprehensive survey. *Journal of Systems and Software*, *167*, 110611.

Blohm, I., Leimeister, J. M., and Krcmar, H. (2013). Crowdsourcing: how to benefit from (too) many great ideas. *MIS Quarterly Executive*, *12*(4), 199–211.

Blohm, I., Zogaj, S., Bretschneider, U., and Leimeister, J. M. (2018). How to manage crowdsourcing platforms effectively? *California Management Review*, *60*(2), 122–149.

Bodkhe, U., and Tanwar, S. (2021). Secure data dissemination techniques for IoT applications: Research challenges and opportunities. *Software: Practice and Experience*, *51*(12), 2469–2491.

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

Buettner, R. (2015). A systematic literature review of crowdsourcing research from a human resource management perspective. *2015 48th Hawaii International Conference on System Sciences*, 4609–4618.

Cabanillas, C. (2016). Exploring human resource management in crowdsourcing platforms. *Workshop on Enterprise and Organizational Modeling and Simulation*, 113–128.

Chen, X., Johnson, E., Kulkarni, A., Ding, C., Ranelli, N., Chen, Y., and Xu, R. (2021). An exploratory approach to deriving nutrition information of restaurant food from crowdsourced food images: Case of Hartford. *Nutrients*, *13*(11), 4132.

Council of the European Union. (2017). *Experience of certain EU Member States with dual quality of foodstuffs in free movement within the EU. Council of the European Union*. http://data.consilium.europa.eu/doc/document/ST-6716-2017-INIT/en/pdf

Croatian Food Agency. (2017). *Quality research of (seemingly) identical products on the markets of old and new EU member states. Osijek.* https://www.hah.hr/wp-content/uploads/2015/10/Dual-quality-final-report.pdf

Cullina, E., Conboy, K., and Morgan, L. (2015). Measuring the crowd: a preliminary taxonomy of crowdsourcing metrics. *Proceedings of the 11th International Symposium on Open Collaboration*, 1–10.

Dahlander, L., Jeppesen, L. B., and Piezunka, H. (2019). How organizations manage crowds: Define, broadcast, attract, and select. In *Managing inter-organizational collaborations: Process views*. Emerald Publishing Limited.

Daniel, F., Kucherbaev, P., Cappiello, C., Benatallah, B., and Allahbakhsh, M. (2018). Quality control in crowdsourcing: A survey of quality attributes, assessment techniques, and assurance actions. *ACM Computing Surveys (CSUR)*, *51*(1), 7.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 319–340.

De Stefano, V. (2016). Introduction: crowdsourcing, the gig-economy and the law. *Comparative Labor Law and Policy Journal*, *37*(3).

Ding, Y.-X., and Zhou, Z.-H. (2018). Crowdsourcing with unsure option. *Machine Learning*, *107*(4), 749–766.

Dong, Z., and Tian, X. (2015). Multi-level photo quality assessment with multi-view features. *Neurocomputing*, *168*, 308–319.

Dunford, E. K., and Neal, B. (2017). FoodSwitch and use of crowdsourcing to inform nutrient databases. *Journal of Food Composition and Analysis*, *64*, 13–17.

Dutil, P. (2015). Crowdsourcing as a new instrument in the government's arsenal: Explorations and considerations. *Canadian Public Administration*, *58*(3), 363–383.

Estellés-Arolas, E., and González-Ladrón-De-Guevara, F. (2012). Towards an integrated crowdsourcing definition. *Journal of Information Science*, *38*(2), 189–200.

European Commission. (2017). *Commission Notice on the application of EU food and consumer protection law to issues of Dual (Vol. 327).* European Union.

European Commission. (2018). *Framework for selecting and testing of food products to assess quality related characteristics: EU harmonised testing methodology.* https://knowledge4policy.ec.europa.eu/file/eu-harmonised-testing-methodology-framework-selecting-testing-food-products-assess-quality_en

European Commission. (2019). *Results of an EU wide comparison of quality related characteristics of food products.*

European Parliament. (2017). *Dual quality of branded food products: Addressing a possible east-west divide. Briefing*.

Eurostat. (2020). *European Statistical System handbook for quality and metadata reports*. https://doi.org/10.2785/666412

Fedorenko, I., Berthon, P., and Rabinovich, T. (2017). Crowded identity: Managing crowdsourcing initiatives to maximize value for participants through identity creation. *Business Horizons*, *60*(2), 155–165.

Felstiner, A. (2011). Working the crowd: employment and labor law in the crowdsourcing industry. *Berkeley J. Emp. and Lab. L.*, *32*, 143.

Ferreira, L. N., and Zhao, L. (2016). Time series clustering via community detection in networks. *Information Sciences*, *326*, 227–242.

Ferreira, P. S., Shamsuzzoha, A. H. M., Toscano, C., and Cunha, P. (2012). Framework for performance measurement and management in a collaborative business environment. *International Journal of Productivity and Performance Management*.

Ford, R. C., Richard, B., and Ciuchta, M. P. (2015). Crowdsourcing: A new way of employing non-employees? *Business Horizons*, *58*(4), 377–388.

Fung, A. (2015). Putting the public back into governance: The challenges of citizen participation and its future. *Public Administration Review*, *75*(4), 513–522.

Gadiraju, U., Demartini, G., Kawase, R., and Dietze, S. (2019). Crowd anatomy beyond the good and bad: Behavioral traces for crowd worker modeling and pre-selection. *Computer Supported Cooperative Work (CSCW)*, *28*(5), 815–841.

Geiger, D., Seedorf, S., Schulze, T., Nickerson, R. C., and Schader, M. (2011). Managing the crowd: towards a taxonomy of crowdsourcing processes. *AMCIS 2011 Proceedings - All Submissions. 430.*

Goncalves, J., Hosio, S., Rogstadius, J., Karapanos, E., and Kostakos, V. (2015). Motivating participation and improving quality of contribution in ubiquitous crowdsourcing. *Computer Networks*, *90*, 34–48.

Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal*, *69*(4), 211–221.

Goodchild, M. F., and Li, L. (2012). Assuring the quality of volunteered geographic information. *Spatial Statistics*, *1*, 110–120.

Gu, Z., Bapna, R., Chan, J., and Gupta, A. (2022). Measuring the impact of crowdsourcing features on mobile app user engagement and retention: A randomized field experiment. *Management Science*, *68*(2), 1297–1329.

Hammond, W. M., Stone, M. E. B., and Stone, P. A. (2020). Picture worth a thousand words: Updating repeat photography for 21st century ecologists. *Ecology and Evolution*, *10*(24), 14113–14121.

Han, S., Huang, H., Luo, Z., and Foropon, C. (2019). Harnessing the power of crowdsourcing and Internet of Things in disaster response. *Annals of Operations Research*, *283*(1), 1175–1190.

Harrington, K., Zenk, S. N., Van Horn, L., Giurini, L., Mahakala, N., and Kershaw, K. N. (2021). The Use of Food Images and Crowdsourcing to Capture Real-time Eating Behaviors: Acceptability and Usability Study. *JMIR Formative Research*, *5*(12), e27512.

Hirth, M., Hoßfeld, T., and Tran-Gia, P. (2013). Analyzing costs and accuracy of validation mechanisms for crowdsourcing platforms. *Mathematical and Computer Modelling*, *57*(11–12), 2918–2932.

Hossain, M., and Kauranen, I. (2015). Crowdsourcing: a comprehensive literature review. *Strategic Outsourcing: An International Journal*.

Hosseini, M., Phalp, K., Taylor, J., and Ali, R. (2019). On the configuration of crowdsourcing projects. In *Crowdsourcing: Concepts, Methodologies, Tools, and Applications* (pp. 33–52). IGI Global.

Hou, W., Gao, X., Tao, D., and Li, X. (2014). Blind image quality assessment via deep learning. *IEEE Transactions on Neural Networks and Learning Systems*, *26*(6), 1275–1286.

Howe, J. (2006). The rise of crowdsourcing. *Wired Magazine*, *14*(6), 1–4.

Jager, J., Putnick, D. L., and Bornstein, M. H. (2017). II. More than just convenient: The scientific merits of homogeneous convenience samples. *Monographs of the Society for Research in Child Development*, *82*(2), 13–30.

Janssen, M., Estevez, E., and Janowski, T. (2014). Interoperability in big, open, and linked data--organizational maturity, capabilities, and data portfolios. *Computer*, *47*(10), 44–49.

JRC. (2019). *Results of an EU wide comparison of quality related characteristics of food products*. Publications Office of the European Union. https://doi.org/10.2760/43870

Kamoun, F., Alhadidi, D., and Maamar, Z. (2015). Weaving risk identification into crowdsourcing lifecycle. *Procedia Computer Science*, *56*, 41–48.

Karachiwalla, R., and Pinkow, F. (2021). Understanding crowdsourcing projects: A review on the key design elements of a crowdsourcing initiative. *Creativity and Innovation Management*, *30*(3), 563–584.

Kaufmann, N., Schulze, T., and Veit, D. (2011). More than fun and money. Worker Motivation in Crowdsourcing-A Study on Mechanical Turk. *Americas Conference on Information Systems (AMCIS)*, *11*(2011), 1–11.

Kawano, Y., and Yanai, K. (2014). Automatic expansion of a food image dataset leveraging existing categories with domain adaptation. *European Conference on Computer Vision*, 3–17.

Khare, R., Good, B. M., Leaman, R., Su, A. I., and Lu, Z. (2016). Crowdsourcing in biomedicine: challenges and opportunities. *Briefings in Bioinformatics*, *17*(1), 23–32.

Kittur, A., Smus, B., Khamkar, S., and Kraut, R. E. (2011). Crowdforge: Crowdsourcing complex work. *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, 43–52.

Kohler, T., and Chesbrough, H. (2019). From collaborative community to competitive market: The quest to build a crowdsourcing platform for social innovation. *Randd Management*, *49*(3), 356–368.

Lian, R., Zhou, A., Zheng, Y., and Wang, C. (2021). Towards secure and trustworthy crowdsourcing with versatile data analytics. *International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness*, 42–53.

Liu, Z., Sakulyeva, T., Mikheev, A., and Stepanova, D. (2021). Management Problems in Global Crowdsourcing. *Journal of Global Information Management (JGIM)*, *30*(3), 1–15.

Malone, T. W. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive Science*, *5*(4), 333–369.

Martin, C. K., Han, H., Coulon, S. M., Allen, H. R., Champagne, C. M., and Anton, S. D. (2008). A novel method to remotely measure food intake of free-living individuals in real time: the remote food photography method. *British Journal of Nutrition*, *101*(3), 446–456.

Martin, C. K., Nicklas, T., Gunturk, B., Correa, J. B., Allen, H. R., and Champagne, C. (2014). Measuring food intake with digital photography. *Journal of Human Nutrition and Dietetics*, *27*, 72–81.

Maslow, A. H. (1958). *A Dynamic Theory of Human Motivation.*

Matheus, R., Janssen, M., and Maheshwari, D. (2018). Data science empowering the public: Data-driven dashboards for transparent and accountable decision-making in smart cities. *Government Information Quarterly*.

Modaresnezhad, M., Iyer, L., Palvia, P., and Taras, V. (2020). Information Technology (IT) enabled crowdsourcing: A conceptual framework. *Information Processing and Management*, *57*(2), 102135.

Mortensen, K., and Hughes, T. L. (2018). Comparing Amazon's Mechanical Turk platform to conventional data collection methods in the health and medical research literature. *Journal of General Internal Medicine*, *33*(4), 533–538.

MPSR. (2017). *Záverečná správa porovnania kvality potravín v Rakúsku a na Slovensku*. http://www.mpsr.sk/download.php?fID=12627

Muller, C. L., Chapman, L., Johnston, S., Kidd, C., Illingworth, S., Foody, G., Overeem, A., and Leigh, R. R. (2015). Crowdsourcing for climate and atmospheric sciences: current status and future potential. *International Journal of Climatology*, *35*(11), 3185–3203.

Nassar, L., and Karray, F. (2019). Overview of the crowdsourcing process. *Knowledge and Information Systems*, *60*(1), 1–24.

Néhib. (2017). *Kiadvány a Magyarországon és külföldön forgalmazott élelmiszerek 2017. Évi* *összehasonlító vizsgálatának tapasztalatairól*. https://portal.nebih.gov.hu/-/magyarorszagon-es-kulfoldon-forgalmazott-elelmiszerek-2017-evi-nyari-osszehasonlito-vizsgalata

Nes, K., Antonioli, F., Di-Marcantonio, F., and Ciaian, P. (2023). *EU-wide comparison of the characteristics and presentation of branded food products (2021)*.

Nevo, D., and Kotlarsky, J. (2020). Crowdsourcing as a strategic IS sourcing phenomenon: Critical review and insights for future research. *The Journal of Strategic Information Systems*, *29*(4), 101593.

Noronha, J., Hysen, E., Zhang, H., and Gajos, K. Z. (2011). Platemate: crowdsourcing nutritional analysis from food photographs. *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, 1–12.

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

Poblet, M., García-Cuesta, E., and Casanovas, P. (2018). Crowdsourcing roles, methods and tools for data-intensive disaster management. *Information Systems Frontiers*, *20*(6), 1363–1379.

Pravst, I., Hribar, M., Žmitek, K., Blažica, B., Seljak, B. K., and Kušar, A. (2021). Branded Foods Databases as a Tool to Support Nutrition Research and Monitoring of the Food Supply: Insights From the Slovenian Composition and Labeling Information System. *Frontiers in Nutrition*, *8*.

Roth, S., and Luczak-Roesch, M. (2020). Deconstructing the data life-cycle in digital humanitarianism. *Information, Communication and Society*, *23*(4), 555–571.

Rynes, S. L., Gerhart, B., and Minette, K. A. (2004). The importance of pay in employee motivation: Discrepancies between what people say and what they do. *Human Resource Management: Published in Cooperation with the School of Business Administration, The University of Michigan and in Alliance with the Society of Human Resources Management*, *43*(4), 381–394.

Schäfer, S., Antons, D., Lüttgens, D., Piller, F., and Salge, T. O. (2017). Talk to Your Crowd: Principles for Effective Communication in Crowdsourcing A few key principles for communicating with solvers can help contest sponsors maintain and grow their base of participants. *Research-Technology Management*, *60*(4), 33–42.

Schenk, E., Guittard, C., and Pénin, J. (2019). Open or proprietary? Choosing the right crowdsourcing platform for innovation. *Technological Forecasting and Social Change*, *144*, 303–310.

Seidel, C. E., Thapa, B. E. P., Plattfaut, R., and Niehaves, B. (2013). Selective crowdsourcing for open process innovation in the public sector: are expert citizens really willing to participate? *Proceedings of the 7th International Conference on Theory and Practice of Electronic Governance*, 64–72.

Senaratne, H., Mobasheri, A., Ali, A. L., Capineri, C., and Haklay, M. (2017). A review of volunteered geographic information quality assessment methods. *International Journal of Geographical Information Science*, *31*(1), 139–167.

Serret, H., Deguines, N., Jang, Y., Lois, G., and Julliard, R. (2019). Data quality and participant engagement in citizen science: comparing two approaches for monitoring pollinators in France and South Korea. *Citizen Science: Theory and Practice*, *4*(1), 22.

Simperl, E. (2015). How to use crowdsourcing effectively: Guidelines and examples. *Liber Quarterly: The Journal of European Research Libraries*, *25*(1).

Singh, R., Hélouët, L., and Miklos, Z. (2021). Reducing the cost of aggregation in crowdsourcing. In *Transactions on Large-Scale Data-and Knowledge-Centered Systems XLIX* (pp. 33–69). Springer.

Solano-Hermosilla, G., Adewopo, J., Peter, H., Arbia, G., Nardelli, V., Barreiro-Hurle, J., Gorrín González, C., Micale, F., and Ceccarelli, T. (2020). *A quality approach to real-time smartphone and citizen-*

driven food market price data: The case of Food Price Crowdsourcing Africa (FPCA) in Nigeria. https://doi.org/10.2760/366116

Solano-Hermosilla, G., Barreiro-Hurle, J., Adewopo, J., and Gorrín González, C. (2022). Increasing engagement in price crowdsourcing initiatives: using nudges in Nigeria. *World Development*, *152*, 105818.

Standing, S., and Standing, C. (2018). The ethical use of crowdsourcing. *Business Ethics: A European Review*, *27*(1), 72–80.

Sternberg, H., and Lantz, B. (2018). Using crowdsourced data to analyze transport crime. *International Journal of Logistics Research and Applications*, *21*(2), 133–147.

Sun, Y., Wang, N., Yin, C., and Zhang, J. X. (2015). Understanding the relationships between motivators and effort in crowdsourcing marketplaces: A nonlinear analysis. *International Journal of Information Management*, *35*(3), 267–276.

Surowiecki, J. (2004). *The wisdom of crowds: why the many are smarter than the few and how collective wisdom shapes business, economics, society and nations* (Vol. 39). Random House Large Print.

Taeihagh, A. (2017). Crowdsourcing: a new tool for policy-making? *Policy Sciences*, *50*(4), 629–647.

Tinati, R., Luczak-Roesch, M., Simperl, E., and Hall, W. (2017). An investigation of player motivations in Eyewire, a gamified citizen science project. *Computers in Human Behavior*, *73*, 527–540.

Van Eerde, W. (2015). Motivation and reward systems. *Wiley Encyclopedia of Management*, 1–4.

Wang, M.-M., and Wang, J.-J. (2019). Understanding solvers' continuance intention in crowdsourcing contest platform: An extension of expectation-confirmation model. *Journal of Theoretical and Applied Electronic Commerce Research*, *14*(3), 17–33.

Wang, Y., Jia, X., Jin, Q., and Ma, J. (2017). Mobile crowdsourcing: framework, challenges, and solutions. *Concurrency and Computation: Practice and Experience*, *29*(3), e3789.

Westenbrink, S., van der Vossen-Wijmenga, W., Toxopeus, I., Milder, I., and Ocké, M. (2021). LEDA, the branded food database in the Netherlands: Data challenges and opportunities. *Journal of Food Composition and Analysis*, *102*, 104044.

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

Wiley, C. (1997). What motivates employees according to over 40 years of motivation surveys. *International Journal of Manpower*.

Yasmin, R., Hassan, M. M., Grassel, J. T., Bhogaraju, H., Escobedo, A. R., and Fuentes, O. (2022). Improving Crowdsourcing-Based Image Classification Through Expanded Input Elicitation and Machine Learning. *Frontiers in Artificial Intelligence*, *5*, 848056.

Zeug, H., Zeug, G., Bielski, C., Solano-Hermosilla, G., and M'barek, R. (2017). *Innovative Food Price Collection in Developing Countries: Focus on Crowdsourcing in Africa*. https://doi.org/10.2788/53432

Zhao, Y., and Zhu, Q. (2014). Evaluation on crowdsourcing research: Current status and future direction. *Information Systems Frontiers*, *16*(3), 417–434.

Zhou, N., Siegel, Z. D., Zarecor, S., Lee, N., Campbell, D. A., Andorf, C. M., Nettleton, D., Lawrence-Dill, C. J., Ganapathysubramanian, B., and Kelly, J. W. (2018). Crowdsourcing image analysis for plant phenomics to generate ground truth data for machine learning. *PLoS Computational Biology*, *14*(7), e1006337.

Abbreviations

CLAS	Composition and Labelling Information System
DQ	dual quality
EAN	European article number
ECO	‘Empowering consumer organisations: Towards a harmonised approach tackling “dual quality” in food products’
FSANZ	Food Standards Australia New Zealand
GS1	global product classification
GTIN	global trade item number
HIT	human intelligence task
HR	human resources
HSR	health star rating
ICT	information and communications technology
IT	information technology
LEDA	Levensmiddelendatabank (Food Database)
MTurk	Mechanical Turk
NethFIR	Netherlands Food Information Resource
NGO	non-governmental organisation
OCR	optical character recognition

List of figures

List of tables

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct centres. You can find the address of the centre nearest you online (european-union.europa.eu/contact-eu/meet-us_en).

On the phone or in writing

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

— by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

— at the following standard number: +32 22999696,

— via the following form: european-union.europa.eu/contact-eu/write-us_en.

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website (european-union.europa.eu).

EU publications

You can view or order EU publications at op.europa.eu/en/publications. Multiple copies of free publications can be obtained by contacting Europe Direct or your local documentation centre (european-union.europa.eu/contact-eu/meet-us_en).

EU law and related documents

For access to legal information from the EU, including all EU law since 1951 in all the official language versions, go to EUR-Lex (eur-lex.europa.eu).

Open data from the EU

The portal data.europa.eu provides access to open datasets from the EU institutions, bodies and agencies. These can be downloaded and reused for free, for both commercial and non-commercial purposes. The portal also provides access to a wealth of datasets from European countries.

Science for policy

The Joint Research Centre (JRC) provides independent, evidence-based knowledge and science, supporting EU policies to positively impact society

EU Science Hub
joint-research-centre.ec.europa.eu

Publications Office
of the European Union