An Assessment of Privacy Preservation in Crowdsourcing Approaches: Towards GDPR Compliance

Vasiliki Diamantopoulou
School of Computing, Engineering
and Mathematics
University of Brighton, UK
v.diamantopoulou@brighton.ac.uk

Aggeliki Androutsopoulou, Stefanos Gritzalis, Yannis Charalabidis
Department of Information and
Communication Systems Engineering
University of the Aegean, Greece
{ag.andr,sgritz,yannisx}@aegean.gr

***Abstract*—The increasing use of Social Media has transformed them into valuable tools, able to provide answers and decision support in public policy formulation. This has resulted in the emergence of new e-participation paradigms, such as crowdsourcing approaches, aiming to drive more constructive interactions between governments and citizens or experts, in order to exploit their knowledge, opinions, and ideas when tackling complex societal problems. However, the continuous exposure of the average users, without or with limited awareness of the dangers of the disclosure of sensitive data, remains a threat to the preservation of their information privacy. The upcoming EU General Data Protection Regulation (GDPR) about the protection of personal data is especially well timed, and forces for revision of the processes followed related to the manipulation of personal data within public participation methods. Towards this direction, a thorough examination of three advanced methods of crowdsourcing in public policy-making processes is conducted in the current paper, analysing the data collection and processing methods they encompass. Then, an assessment of their compliance with fundamental privacy requirements is presented. The research contributes to the identification of challenges that crowdsourcing, and in general, e-participation approaches impose with regard to privacy protection. Further research directions include the implementation of techniques that can satisfy the identified requirements.**

***Index Terms*—E-Participation, Crowdsourcing Methods, Personal Identifiable Information, Privacy Requirements, Privacy Enhancing Technologies**

I. Introduction

During the last two decades, governments have put effort in incorporating Information and Communication Technologies (ICT) to provide a new means of communication, allowing citizens' participation and engagement in shaping public policy. This "practice of consulting and involving members of the public in the agenda-setting, decision-making and policy forming activities of organisations or institutions responsible for policy development" is known as public participation [1]. At the same time, Web 2.0 technologies and architectures have brought great potential in collecting and processing a large amount of external information to convey useful insights in the policy formulation process [2]. Social Media provide a valuable source of information concerning the different issues perceived by different problem stakeholder groups, as well as the different solutions they propose and arguments in favour and against them, and in general their different concerns [3].

The above have attracted attention of many researchers conducting research under the field of e-participation, which aim to drive more constructive interactions between governments, in order to tackle complex societal problems. For example, there is a growing adoption of crowdsourcing and open innovation initiatives in the public sector aiming to harness the "wisdom of the crowd" [4], [5], [6], [7]. Both concepts emerged in the private sector and rely on the utilisation of external knowledge and ideas, either to solve a specific problem or to develop innovations [8], [9], [10]. This makes a perfect alignment with the notion of e-participation, where individuals are recognised as carriers of a wealth of (tacit) knowledge and experience that can be explicited to better understand social needs, identify expectations, and assess the effectiveness of policies. Applications of the above areas have introduced new opportunities to exploit towards the design of better public policies ideas and knowledge, on the one hand, of citizens-general public (citizen-sourcing) [11], [6], [12], and on the other hand, of experts (expert-sourcing) [13], incorporating them into the work of the governments. As open innovation consists a form of crowdsourcing, the focus of current research is on the utilisation of social media by governmental agencies for the collection of external knowledge through crowdsourcing. Web 2.0 capabilities have played a significant role for the development of crowdsourcing, as they allow the efficient participation and interaction of numerous and geographically dispersed individuals, and also the analysis of their contributions [14], [15].

Despite that these applications in the public sector have already been explored [16], [17], [18], few studies have examined the privacy requirements that have to be taken into account in order for users to preserve their privacy. Moreover, even though the advantages and the opportunities that crowdsourcing offer to decision makers, the collection of knowledge, opinions and ideas also create risks to privacy of the contributors of this content.

To fill the identified research gap, we present a study which aims to cover the above limitations. This study is realised following a three-step procedure: At the first step, a thorough examination on three advanced methods of crowdsourcing in public policy making processes was conducted, namely *active crowdsourcing*, *passive crowdsourcing*, and *passive expertsourcing*. The focus of the study in this step was to analyse those characteristics that shape users' identity and have to be examined when addressing privacy concerns. During the second step, we analysed relevant literature where we identified a list of privacy requirements that have to be satisfied in order to preserve individuals' information privacy. This was realised by aggregating the results of the literature review on this area, while taking into account the GDPR specifications. Moreover, this step contains the assessment of the degree to which alternative crowdsourcing practices conform to those requirements and also reveals privacy requirements that are at risk. Finally, at the third step we indicate a set of technical and administrative precautions, focusing on established Privacy Enhancing Technologies (PETs), to protect Personal Identifiable Information (PII) against unauthorised access, disclosure, misuse and theft.

The interrelation of these steps produced an adjacency matrix, presented in Section V, which concentrates the output of this analysis, resulted by matching the findings of the first and second step. This analysis contributes to pinpoint specific PETs. It is clear that while the first two steps rely on literature study on the respective fields, the final step results on the analysis and synthesis of the findings of the previous steps. In order to support the analysis of crowdsourcing methods, we were based on the recent literature [6], [19], [20], [21], [13] that describes new ways of e-participation, after the rise of Web 2.0 technologies. Additionally, in order to support the second step regarding the privacy requirements that have to be examined, we focused on fundamental studies in this area [22], [23], [24], [25], [26]. To the best of our knowledge, there is no other study that approaches the privacy issue in crowdsourcing environment in such a holistic way.

The results of this work can be found useful by the research community, as they enumerate a set of privacy challenges identified through literature review and can be used as a basis for a deeper analysis of each crowdsourcing method. Moreover, policy makers that use such platforms to communicate with massive amount of people can use our findings, realise the importance of the adoption of a privacy culture, and through the implementation of the suggested technical solutions, they can enhance their trustworthiness, by making their systems more robust in privacy threats, preventing incidents such as identity theft, losses of databases of personal information, unauthorised accesses, fraudulent connection requests, to name a few.

The rest of the paper is structured as follows. Section II discusses the related work, while Section III presents the three methods of crowdsourcing to be examined. Section IV identifies a set of privacy requirements that have to be satisfied to ensure the secure data collection and processing in these practices. Section V presents the findings of this study concentrated in an adjacency matrix, while Section VI concludes the paper by raising issues for further research.

II. Related Work

One of the contributions of our work is that it examines the three crowdsourcing practices in a holistic way. For this reason, since to the extent of our knowledge, there is no other work analyses the privacy requirements in such environment, we focused our literature review on these areas separately.

A. Crowdsourcing

Crowdsourcing is the easiest way to gather huge data on something that can be used to monitor the well-being of high-risk communities without requiring significant investment in human resources or infrastructure. Through the use of a crowdsourcing process, it is possible to gather private data and sensitive personal data on sensitive issues. Crowdsourcing allows citizens to connect with each other, governments to connect with common mass for various actions, such as the coordination of disaster response work, the mapping of political conflicts, acquiring thus information in short time, being closer to real issues that affect day-to-day life of citizens [27]. Sometimes governments develop (i.e. LEEDIR[1] in the U.S.) or allow private companies to develop citizen-monitoring tools (Internet Eyes[2] in the UK). In the name of national security, governments also use different laws to monitor or to access data from such type of digital platforms. Private companies can gather different types of personal information of online active individuals for their business promotion. They use different methods of data collection including data mining, crowdsourcing, online surveying process, etc. Thus, this type of act by private companies might constitute gross privacy risks for citizens. Different law enforcement agencies in developed countries have the technical infrastructure to use crowdsourcing process to collect personal data. They collect personal data of citizens who contribute in a different crowdsourcing platform for a different purpose hosted by third parties [27].

Under this category we can meet *spatial crowdsourcing* [28] which refers to a transformative platform that engages individuals, groups or communities and aims at the collection, analysis, and dissemination of environmental, social or other spatio-temporal information. Individuals can use their mobile devices (*mobile crowdsourcing*) to perform tasks while they are located in the relevant places of interest. If the server is not trusted, individual locations, their activities [29], the time and the place they were when interacting with the platform [30], and finally, the properties of the location, e.g., the variety of people that visit the location [31] may be disclosed, causing serious privacy implications [32], [33], [34], [35]. The data of each individual's activity can be revealed, allowing the attacker either to inspect individuals (through physical surveillance and stalking), to steal their identities, or to breach sensitive

[1] https://www.leedir.com/

[2] https://www.interneteyes.com.br

information related to individual's data (health data, lifestyle choices, political and religious views).

There are only a few papers that deal with privacy problems in crowdsourcing. The authors in [36] brought together researchers from the crowdsourcing field and the human computation field, and among others, they raised issues related to privacy requirements in such environments, such as the preservation of anonymity. In [37] the authors focused on a privacy problem related with task instances in crowdsourcing. Next, the authors in [38] focus on privacy issues related with workers in crowdsourcing environments and they propose a crowdsourcing quality control method in order to estimate reliable provided results from low-quality ones. Our study provides a holistic approach of privacy preservation in crowdsourcing environments, by analysing three crowdsourcing methods and identifying, through the PII that are provided by the users, the privacy requirements that are compromised, providing also appropriate implementation techniques.

B. Citizen-Sourcing

The success stories of "open innovation" and "crowdsourcing" in the private sector [9], [10] have motivated government agencies to move in this direction as well, and this gave rise to the development of "citizen-sourcing" practices in government, and also of considerable research activity in this area [39], [19], [21], [40], [3]. Governments are increasingly using ICT, mainly social media, in order to collect information, knowledge, ideas and opinions from the citizens about important social problems, as well as existing or under formulation public policies for addressing them. According to a recent review of the research literature in this area [41], most of this research focuses on the analysis of the specific activities government agencies conduct for this purpose, the impact of the external context on them, as well as the demographics and the behaviour of the participating citizens; on the contrary, limited research has been conducted for the development of advanced and more effective ICT platforms for citizen-sourcing.

Most of these ICT platforms related research focuses on "active citizen-sourcing", which is based on the use of government agencies' social media accounts in order to pose a specific social problem or public policy (existing or under development), and solicit relevant information, knowledge, ideas and opinions from the general public [3], [42]. Recently, a novel approach to government citizen-sourcing has started being developed, which is based on "passive citizen-sourcing" [21], [6]. In this approach government agencies have a more passive role: they exploit policy-related content that has already been generated by citizens freely, without any active stimulation or direction by government, in various external social media (such as political fora and blogs, Facebook, Twitter accounts, etc.) not belonging to government agencies. This "passive citizen-sourcing" approach offers significant advantages over the "active citizen-sourcing" one: (i) it enables government agencies to access, retrieve and exploit much larger quantities of more diverse policy relevant content from a wide variety of social media sources of different political orientations; and (ii) this content already exists, so government agencies do not have to find ways to attract large numbers of citizens to participate in citizen-sourcing and generate new content. For the above reasons, we build on the concept of "passive citizen-sourcing" as a foundation for the design of the citizens' related components of the proposed e-participation platform.

III. Crowdsourcing Methods

During the last 25 years, governments are using ICT as a mechanism to involve citizens in policy development [43]. The extension of ICT into citizens' involvement is also called e-participation, e-democracy, e-governance, or e-government [44]. In general, the catalog of the domains that use crowdsourcing is very long [45], [46]. Hereby, we present three methods for enhancing public participation, relying on different forms of crowdsourcing. The first implements the concept of active crowdsourcing, in which government has an active role, posing a particular social problem or public policy direction in a governmental website or social media account, and soliciting relevant information, knowledge, opinions and ideas from citizens, who provide content in there. The second one relies on passive crowdsourcing, in which government has a more passive role, monitoring and collecting content on a specific topic or public policy (existing or under development) that has been freely generated by citizens without any stimulation in external various sources not owned by government. Finally, the third method is based on the automated retrieval of information about experts on various policy related topics (expert-sourcing), as well as relevant online texts and postings already published by such experts in multiple social media and websites.

For reasons of completeness, the three following subsections provide a detailed description of each method, analysing the involvement of the various users involved, as well the data collection and processing methods that are being used in each of them.

A. Active Crowdsourcing

The first crowdsourcing method supports public participation adopting an active crowdsourcing approach. It is based on a centralised automated publishing of policy-related content (e.g., short or long text, images and videos on a public policy under formulation or modification) on multiple social media (e.g., Facebook, Twitter, YouTube, Blogspot) simultaneously through a single integrated interface. The purpose of this publishing is to stimulate citizens' discussions around this content. The citizens are able to access this content, view it and interact with it through the capabilities offered by each of these social media. Then, data on citizens' interaction with them (e.g., views, comments, ratings, votes, etc.) are monitored and collected using the application programming interfaces (API) of the targeted social media. Part of this citizens-generated content is numeric (e.g., numbers of views, likes, retweets, comments, etc., or ratings), so it can be used for the calculation of various analytics following Social Media

Monitoring practices. Furthermore, a large part of this content is in textual form, so opinion mining methods are also applied. Therefore, the interaction data collected undergoes various types of advanced processing (e.g., access analytics, opinion mining, simulation modelling) in order to extract synthetic conclusions from them and provide substantial support to government policy makers. The results of this analysis are visualised to finally present to policy makers three types of citizens' feedback:

1) Social Media Metrics (Views, Likes, Tweets, Posts, Comments, Shares, Retweets), which are used to calculate the level of reach and citizens' engagement.
2) Opinion Mining and Sentiment Analysis Results (Positive/Negative statements, which provides a classification of an opinionated text (e.g., a comment) as expressing a positive, negative or neutral opinion and extracts the main issues commented
3) Simulation results based on Decision Support Model (Forecasted Awareness, Interest, Acceptance of citizens with respect to the policy under discussion).

A comprehensive description of the method is provided in [20].

B. Passive Crowdsourcing

A method of passive crowdsourcing consisting of services for sophisticated collection and analysis on textual content published by citizens in external social media, has been developed in the context of the NOMAD project[3]. This is highly valuable, as it enables decision makers to take into account and benefit from "fresh" relevant content contributed by citizens in numerous social media, incorporating useful ideas, knowledge as well as perceptions of the general public. This method consists of services that: (i) create and maintain policy models (incorporating the main elements of public policies), (ii) use such policy models mine relevant citizen-generated data from a variety of online text sources (e.g., political blogs, social media, web-sites), (iii) perform linguistic analysis of them to transform free text into a set of structured data, (iv) discover and extract main issues discussed as well as arguments from free text, (v) perform sentiment analysis to classify text segments according to their "tone" (positive, neutral, negative), (vi) cluster arguments, based on calculated similarities, and present automatically-generated summaries, and (vii) visualise a structured view of citizens' opinions on a policy related topic (visualised through word-clouds and other kinds of charts), providing insights on what about, how much and when citizens are discussing concerning this topic.

This method is supported by a set of tools for searching and analysing public policy related content that has been generated by citizens in numerous "external social media (i.e. not belonging to government, such as various political blogs, fora, Facebook and Twitter accounts, etc.); furthermore it provides advanced tools for analysing this content in order to identify specific issues, ideas, concerns and other information hidden within the text of citizens' posting on the web, enabling in this way a "passive citizen-sourcing". What differentiates this approach from typical Social Media Monitoring tools topic [21] is that analysis is tailored against specific policy makers' goals, by properly visualising arguments, opinions and sentiments regarding a policy domain, and creating a semantically rich, accurate stream of data that can be leveraged in any workflow.

C. Passive Expert-Sourcing

Passive expert-sourcing is a social media based crowdsourcing method supporting policy making, whose inception originates from the need of policy makers to utilise knowledge and perspectives of experts as well, when addressing critical societal problems. This expert-sourcing method exploits policy-related content that has already been published by experts in numerous social media and web sites (without any direct stimulation or direction by government, so it performs "passive expert-sourcing"), adopting a selective approach. It filters this content, in order to extract the highest quality parts of it that have been authored by the most knowledgeable experts, based on reputation management and text/opinion mining techniques. In particular, textual content of documents, articles and social media posts is processed using opinion mining and sentiment classification methods, in order to identify subjective information, extract opinions and assess their sentiment (positive, negative or neutral). Furthermore, these documents undergo sophisticated processing using text/opinion mining and sentiment classification techniques, in order to assess the polarity of their orientation (positive, negative or neutral) and assess the relevance of them with relation to a topic. Regarding the experts, it is necessary to apply digital reputation techniques for assessing their reputation/credibility and provide a ranking of them per topic of interest. The results are then visualised in order for policy stakeholders to gain a comprehensive view on policy related content and get involved in a constructive public policy dialogue.

This method aims to support the efficient and effective retrieval by various actors of the democratic processes (e.g., representatives of stakeholder groups, journalists, government employees, active citizens, etc.) of diverse expert information, knowledge and ideas on a specific topic/policy, which is included in postings and texts authored by experts and published in various web-sites and social media. Furthermore, the proposed method of passive expert-sourcing aims to increase the density of interactions among the actors participating in public policy networks, which is highly important for their stability, the development of shared values and beliefs, and finally the effectiveness and the outcomes of such networks, by supporting the exchange of expertise and knowledge between network participants. A comprehensive description of this method is provided in [13].

The description of the above methods makes prominent that they all encompass sophisticated techniques for acquiring and processing user generated content, in order to extract the most significant and highest quality parts of it that can provide

[3] http://nomad-project.eu

TABLE I
IDENTIFICATION OF PERSONAL IDENTIFIABLE INFORMATION

Method	PII
Active Crowdsourcing	Views Social Media Metrics Comments
Passive Crowdsourcing	Social Media Posts Blog Posts Articles' Comments
Passive Expert-Sourcing	Documents Ratings Comments

meaningful insights for the policy formulation process. For instance, they employ data mining technologies to discover patterns from the provided input of Social Media users. In the first two analysis is conducted on aggregated level and not at in individual level, without compromising the identity of an individual user. However, since the purpose of data collection and processing is to get and convey the overall picture of citizens' or experts' opinions or knowledge with regard to policy related topics, data may contain sensitive information as well. On the other hand, in the third method results are collected and presented on the basis of individuals recognised as experts and their published statements. All these pose a series of challenges with regard to users' privacy protection. The main research question that arises is to what extent the above crowdsourcing methods process and reveal personal or sensitive data and wether they meet the security requirements imposed by data protections regulations and privacy guidelines. In the following table we summarise the data processed in each of the three methods as a first step of our their assessment against a set of privacy requirements that are listed in the next Section.

IV. PRESERVATION OF INFORMATION PRIVACY

With the upcoming of the General Data Protection Regulation (GDPR) [47], the scope of data protection is expanded and as a consequence of this, the actions related to the collection, processing, storing, and transmission of EU citizens' personal data should be revised. In parallel with the expansion of these activities, the definition of personal data has also been expanded. The definition states that *"personal data" means any information relating to an identified or identifiable natural person ("data subject"); an identifiable natural person is one who can be identified*. The trivial information combined with other information should not be neglected, especially for some users who prone to expose personal information [48].

The GDPR puts a lot of emphasis on the rights of individuals when they provide their personal data for using an online service of an organisation (or "data controller"). Individuals (or "data subjects", according to the official terminology of the GDPR) must unambiguously give their consent for their data to be processes, which must be informed and voluntary. These rights are summarised in the following points:

- **Right of Access** Data subjects have the right to obtain from the organisations confirmation as whether or not personal data concerning them are processed, and, where that is the case, they have the right to request and get access to that personal data.
- **Right to Rectification** Data subjects have the right to obtain from the organisations the rectification of inaccurate personal data. Also, they have the right to provide additional personal data to complete any incomplete personal data.
- **Right to Erasure** Data subjects have the right to obtain from the organisations the erasure of their personal data.
- **Right to Restriction of Processing** Data subjects have the right to obtain restriction of processing of their personal data, applicable for a certain period and/or for certain situations.
- **Right to Data Portability** Data subjects have the right to receive from organisations in a structured format their personal data. Also, they have the right to (let) transmit such personal data to another data controller.
- **Right to be Not Subject to Automated Individual Decision-Making** Data subjects have the right not to be subject to a decision based solely on automated processing.
- **Right to Filing Complaints** Data subjects have the right to file complaints with the applicable data protection authority on the data controller's processing of their personal data.
- **Right to Compensation of Damages** In case a data controller breaches applicable legislation on processing of a data subject's personal data, the data subject has the right to claim damages from data controller.

With more and more personal, sensitive and confidential information stored, shared and managed at digital level [49], it is expected from both the individuals and the organisations that appropriate measures should be implemented to ensure privacy of such information. Moreover, it is also critical both for individuals and for organisations to realise the high importance of each piece of information they reveal during their interaction with online services, and also the dangers that their exposure can hide. However, privacy preservation is not a straightforward process, as privacy is a multifaceted concept with various parameters that need to be taken into account, at technical and social level, also various impact and different ways of achievement, which is affected, amongst other things, by the environments in which it is required to be achieved.

In this direction, the GDPR describes that one of the data controllers' obligations is to implement appropriate technical and organisational measures in order to apply data protection by design and by default [47] in the electronic services they offer to individuals. The analysis conducted in the following subsection sets the ground towards adhering to the aforementioned obligation.

A. Privacy Requirements

Information privacy refers to an individual's indefeasible right to control the ways in which personal information is obtained, processed, distributed, shared, and used by any other entity [50], [51]. The vulnerability of information privacy has increased due to the intrusion of social media platforms [52] and the intensive development of new e-participation methods on top of these. To a large extent, the raw material for most of interactions of individuals, with others, with well-established communities and with governmental authorities, include personal data of individuals. Alongside the benefits for the governmental decision making processes, which have been described in Section III, these developments are accompanied with privacy risks that can have negative impact on users' participation [53]. In view of the above, the GDPR is especially well timed.

In this study we take as basis the fundamental privacy requirements, as they have been defined and identified by the consensus of the literature of the area [22], [23], [24], [25]. The first two are mainly security concepts but they are included due to their important role in the implementation of privacy protection. The definitions of these requirements are briefly presented below:

- **Authentication** refers to the provision of assurance that a claimed characteristic of an entity is correct. The satisfaction of this requirement offers verification of a user's identity and ensures the origin integrity (the source of the data).
- **Authorisation** ensures that user's private data should only be accessed by authorised users. It allows an authenticated client to use a particular service, it deters violations of the integrity of either the systems or users resources, and deters violations of privacy.
- **Anonymity** is a characteristic of information that does not permit a personally identifiable information principal to be identified directly or indirectly. During anonymization, identity information is either erased or substituted. Anonymity supports users in accessing services without disclosing their identity. Users are more freely expressed, since freedom from user profiling is achieved (behaviour of users or other privacy-infringing practices). Moreover, we achieve freedom from location tracking. Finally, we have minimal user involvement, since they do not have to modify their normal activities for anonymity services)
- **Pseudonymity** is the utilisation of an alias instead of personally identifiable information. It supports users in accessing services without disclosing their real identity. However, the user is still accountable for its actions. Moreover, pseudonymity permits the accumulation of reputational capital, and fills the gap between accountability and anonymity. Using pseudonymity, a user may have a number of pseudonyms, thus they can hide their identity. Finally, pseudonymity prevents unforeseen ramifications of the use of online services.
- **Unlinkability** is the use of a resource or a service by a user without a third party being able to link the user with the service. It protects users' privacy when using a resource or service by not allowing malicious third parties to monitor which services are used by the user, and the intentional severing of the relationships (links) between two or more data events and their sources, ensuring that a user may make multiple uses of resources or services without others being able to link the uses together. Moreover, unlinkability requires that users and/or subjects are unable to determine whether the same user caused certain specific operations in the system. Finally, its satisfaction allows the minimisation of risks to the misuse of the privacy-relevant data and to prohibit or restrict profiling.
- **Undetectability** is the inability for a third party to distinguish who is the user (among a set of potential users) using a service. It protects users' privacy when using a resource or a service by not allowing malicious third parties to detect which services are used by the user. The attacker cannot sufficiently detect whether a particular Item of Interest exists or not, e.g., steganography, and also the attacker cannot sufficiently distinguish whether it exists or not.
- **Unobservability** is the inability of a third party to observe if a user (among a set of potential users) is using a service. It ensures that a user may use a resource or service without others, especially third parties, being able to observe that the resource or service is being used, Moreover, it requires that users and/or subjects cannot determine whether an operation is being performed.

B. Privacy Enhancing Technologies through Privacy Process Patterns

In the recent Workshop about Privacy Engineering Research and the GDPR [54], what has been argued is that a lot of work has already been conducted related to Privacy Enhancing Technologies (PETs) which is adequate enough and can be the basis and can be effectively used during the implementation phase of any software development project. The challenge now is not to develop different PETs, but to work efficiently with the already and well-established existing ones, and also for the privacy engineers to find a way to facilitate the Privacy by Design and Privacy by Default requirements, by focusing on the provision of efficient privacy process patterns [55], [56].

In the context of software development, a pattern is considered as a reusable package that offers solutions to specific problems. It incorporates expert knowledge and represents a recurring structure, activity, behaviour or design [57], [58]. Privacy process patterns have been developed for the facilitation of modelling privacy issues in order to specify the way that the respective privacy issues will be realised through a specific number of steps, including activities and flows connecting them. In this work, we focus on design patterns since they have already proposed by various domains as a format that allows the capturing and sharing design knowledge [59], [60]. Privacy process patterns, since they encapsulate expert knowledge of PETs implementation at the operational level, aim to fill the gap between the design stage, i.e. from the identification of the privacy requirements, to the implementation stage, i.e. to use the appropriate PETs.

PETs are solutions aiming at helping organisations and individuals to protect their privacy [23]. According to [26], they can be categorised to the following main categories: i) administrative tools, ii) information tools, iii) anonymizer products, iv) services and architectures, v) pseudonymizer tools, vi) track and evident erasers, and vii) encryption tools. Each category contains specific technical implementation techniques which can be used as a basis for the designer along with the stakeholders and/or the organisation's developer team to decide and propose the most appropriate ones that will satisfy the identified privacy requirements. The definition of selection criteria for the most appropriate and adequate PET is out of the scope of this paper. By applying the relevant privacy pattern on the respective privacy requirement, the identification of the appropriate PET leading to the successful satisfaction is achieved.

V. Creating an adjacency matrix

Each of the previously presented crowdsourcing methods has been assessed against the list of seven privacy requirements described in Section IV. In Table II the results of this assessment are illustrated. There, we indicate with "✓" the requirement that is satisfied by each method, with "X" the requirement that is possibly infringed, while with "–" we indicate a requirement that may be at risk at certain circumstances. Before explaining the information privacy examination conducted, it is necessary to mention the particularities of each method that affect users' data privacy. First of all, while the first method relies on requests of users to provide content, the latter two do not require from people to create new content, instead they conduct selective "passive" crowd-sourcing. This constitutes feasible for the authors of the content in the "active crowdsourcing" to be aware of the processing taking place and offer them the rights imposed by the GDPR (Section IV). On the contrary, in the passive approaches any information that individuals disclose publicly in Social Media (without any restrictions on access rights to specific groups of people) can be subject to processing without users being informed, since data controller is the responsible body carrying out the crowdsourcing initiative.

The first two requirements are inherited by the privacy specifications of the Social media platforms and Web 2.0 sources, where users contribute with content only after they are registered and authenticated. These platforms embed security mechanisms that control access only by authorised users, therefore both authentication and authorisation are safeguarded in all methods. For this reason, the three approaches collect solely data that are open to the public. With respect to the reservation of the rest requirements in the two crowdsourcing approaches, a distinction among the concept of citizen-sourcing and expert-sourcing has to be made. The two first citizen-sourcing methods process only aggregated data resulting to automatically generated summaries. Although the results do not compromise the identity of authors, it is possible that textual content (e.g., comments) may include sensitive information, concerning the name, demographics or the beliefs of the citizens authoring this content. Through this information a third party can infer the identity of the author of this content. Moreover, the extraction of a textual segment can help to track the original source (e.g., a comment) and thus allow to a third party to link the user with the particular resource, distinguish the Social Media user, and observe that the specific user is using the relevant Social Media capability. All the above pose risks at the anonymity, unlinkability, undetectability and unobservability of individuals interacting through Social Media services within the active and passive crowdsourcing method. Finally, pseudonymity is satisfied as it can be retained as far as the Social Media platforms allow.

According to the [26], [55], [56] the techniques that best support and implement the requirement of anonymity are:

- Anonymizer products, services and architectures: Browsing pseudonyms, Virtual Email Addresses, Trusted third parties, Crowds, Onion routing, DC-nets, Mix-nets (Mix Zone), Hordes, GAP, Tor, Aggregation Gateway, Dynamic Location Granularity
- Track and evident erasers: Spyware detection and removal, Hard disk data eraser, User data confinement pattern, Use of dummies

Next, for the satisfaction of unlinkabililty, the techniques that can be implemented are:

- Anonymizer products, services and architectures: Trusted third parties, Surrogate keys, Onion routing, DC-nets, Mix-nets, Hordes, GAP, Tor, Aggregation Gateway
- Pseudonymizer tools: CRM personalisation, Application data management
- Track and evident erasers: Spyware detection and removal, Browser cleaning tools, Activity traces eraser, Hard disk data eraser, Use of dummies, Identity Federation Do Not Track Pattern

For the satisfaction of undetectability, the techniques that can be implemented are:

- Administrative tools: Smart cards, Permission management
- Information tools: Monitoring and audit tools
- Anonymizer products, services and architectures: Hordes, GAP, Tor
- Track and evidence erasers: Spyware detection and removal, Browser cleaning tools, Activity traces eraser, Hard disk data eraser, Identity Federation Do Not Track Pattern
- Encryption tools: Encrypting email, Encrypting transactions, Encrypting documents

Finally, the techniques that best support and implement the requirement of unobservability are:

- Administrative tools: Smart cards, Permission management
- Anonymizer products, services and architectures: Hordes, GAP, Tor
- Track and evidence erasers: Spyware detection and removal, Hard disk data eraser, Identity Federation Do Not Track Pattern

What differentiates the third method is that validation of experts' knowledge is a prerequisite to perform expert-sourcing. If we add on this, the utilisation of digital reputation management techniques, authors' identity, occupation and are revealed to final users of the method. For that reason, the expert-sourcing method is not compliant with the rest privacy requirements. This necessitates that expert-sourcing takes place in full awareness and users' agreement with the collection and processing of experts' personal data.

VI. Conclusions

In light of the recent GDPR regulation, the protection of users' privacy is an increasingly important aspect of e-participation initiatives, where different stakeholders express views and opinions in policy related topics. Nevertheless, during the design of systems supporting such practices, privacy is usually considered as an afterthought due to the lack of expertise of system designers and developers. In the previous sections, three crowdsourcing approaches have been analysed under the perspective of the information privacy of users generating content in the resources they exploit. The analysis has revealed useful insights concerning the challenges that different forms of crowdsourcing impose on personal data and privacy protection. Although the main requirements, namely authentication and authorisation, as well as pseudonymity are fulfilled in both citizen-sourcing and expert-sourcing practices, anonymity, unlinkability, undetectability, and unobservability are either infringed (in the case of expert-sourcing) or not ensured (in the case of active and passive crowdsourcing). By addressing such privacy issues, through a set of PETs suggested in the current study, substantial advantages can be achieved in the area of e-participation, as it can increase stakeholders trust and engagement and drive the wider adoption of such methods by governmental organisations.

One of the limitations of the current study is that the examination takes place on empirical data collected through generic applications of these methods rather on case studies focusing on the evaluation of information privacy aspects. At technical level, an already identified by the literature obstacle is the selection and implementation of appropriate PETs during the development of the ICT systems supporting crowdsourcing approaches. One of the proposed future steps is a further analysis on which implementation technique is appropriate in each case (i.e. active and passive crowdsourcing). This will be achieved after a thorough discussion with the relevant stakeholders of the examined systems in order to identify the goals of their strategy and their available technical sources.

References

[1] G. Rowe and L. J. Frewer, "Evaluating public-participation exercises: a research agenda," *Science, technology, & human values*, vol. 29, no. 4, pp. 512–556, 2004.

[2] J. I. Criado, R. Sandoval-Almazan, and J. R. Gil-Garcia, "Government innovation through social media," 2013.

[3] I. Mergel and K. C. Desouza, "Implementing open innovation in the public sector: The case of challenge. gov," *Public administration review*, vol. 73, no. 6, pp. 882–890, 2013.

[4] S. M. Lee, T. Hwang, and D. Choi, "Open innovation in the public sector of leading countries," *Management decision*, vol. 50, no. 1, pp. 147–162, 2012.

[5] F. Molinari and E. Ferro, "Framing web 2.0 in the process of public sector innovation: Going down the participation ladder," *European Journal of ePractice*, vol. 9, no. 1, pp. 20–34, 2009.

[6] E. Loukis, Y. Charalabidis, and A. Androutsopoulou, "Promoting open innovation in the public sector through social media monitoring," *Government Information Quarterly*, vol. 34, no. 1, pp. 99–109, 2017.

[7] I. Mergel, "A framework for interpreting social media interactions in the public sector," *Government Information Quarterly*, vol. 30, no. 4, pp. 327–334, 2013.

[8] D. C. Brabham, *Using crowdsourcing in government*. IBM Center for The Business of Government, 2013.

[9] J. Howe, "The rise of crowdsourcing," *Wired magazine*, vol. 14, no. 6, pp. 1–4, 2006.

[10] ——, *Crowdsourcing: How the power of the crowd is driving the future of business*. Random House, 2008.

[11] H. Chesbrough, W. Vanhaverbeke, and J. West, *New frontiers in open innovation*. Oup Oxford, 2014.

[12] I. Mergel, "Opening government: Designing open innovation processes to collaborate with external problem solvers," *Social Science Computer Review*, vol. 33, no. 5, pp. 599–612, 2015.

[13] A. Androutsopoulou, F. Mureddu, E. Loukis, and Y. Charalabidis, "Passive expert-sourcing for policy making in the european union," in *International Conference on Electronic Participation*. Springer, 2016, pp. 162–175.

[14] D. Geiger, M. Rosemann, E. Fielt, and M. Schader, "Crowdsourcing information systems-definition typology, and design," 2012.

[15] Y. Zhao and Q. Zhu, "Evaluation on crowdsourcing research: Current status and future direction," *Information Systems Frontiers*, vol. 16, no. 3, pp. 417–434, 2014.

[16] A. Androutsopoulou, N. Karacapilidis, E. Loukis, and Y. Charalabidis, "Towards an integrated and inclusive platform for open innovation in the public sector," in *International Conference on e-Democracy*. Springer, 2017, pp. 228–243.

[17] B. Y. Clark, N. Zingale, J. Logan, and J. L. Brudney, "A framework for using crowdsourcing in government," 2015.

[18] I. Mergel, S. I. Bretschneider, C. Louis, and J. Smith, "The challenges of challenge. gov: Adopting private sector business innovations in the federal government," in *System Sciences (HICSS), 2014 47th Hawaii International Conference on*. IEEE, 2014, pp. 2073–2082.

[19] T. Nam, "Suggesting frameworks of citizen-sourcing via government 2.0," *Government Information Quarterly*, vol. 29, no. 1, pp. 12–20, 2012.

[20] Y. Charalabidis, E. Loukis, and A. Androutsopoulou, "Fostering social innovation through multiple social media combinations," *Information Systems Management*, vol. 31, no. 3, pp. 225–239, 2014.

[21] V. Bekkers, A. Edwards, and D. de Kool, "Social media monitoring: Responsive governance in the shadow of surveillance?" *Government Information Quarterly*, vol. 30, no. 4, pp. 335–342, 2013.

[22] S. Fischer-Hübner, *IT-security and privacy: design and use of privacy-enhancing security mechanisms*. Springer-Verlag, 2001.

[23] J. Cannon, *Privacy: what developers and IT professionals should know*. Addison-Wesley Professional, 2004.

[24] A. Pfitzmann and M. Hansen, "A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management," 2010.

[25] ISO/IEC, "29100:2011(e) information technology - security techniques - privacy framework," Tech. Rep., 2011.

[26] C. Kalloniatis, E. Kavakli, and S. Gritzalis, "Addressing privacy requirements in system design: the pris method," *Requirements Engineering*, vol. 13, no. 3, pp. 241–255, 2008.

TABLE II
ADJACENCY MATRIX

	Active Crowdsourcing	Passive Crowdsourcing	Passive Expert-Sourcing
Authentication	✓	✓	✓
Authorisation	✓	✓	✓
Anonymity	–	✓	✓
Pseudonymity	✓	✓	X
Unlinkability	–	–	X
Undetectability	–	–	X
Unobservability	–	–	X

[27] B. Halder, “Crowdsourcing collection of data for crisis governance in the post-2015 world: potential offers and crucial challenges,” in *Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance*. ACM, 2014, pp. 1–10.

[28] H. To, G. Ghinita, and C. Shahabi, “A framework for protecting worker location privacy in spatial crowdsourcing,” *Proceedings of the VLDB Endowment*, vol. 7, no. 10, pp. 919–930, 2014.

[29] Y. Gong, Y. Guo, and Y. Fang, “A privacy-preserving task recommendation framework for mobile crowdsourcing,” in *Global Communications Conference (GLOBECOM), 2014 IEEE*. IEEE, 2014, pp. 588–593.

[30] M. Benisch, P. G. Kelley, N. Sadeh, and L. F. Cranor, “Capturing location-privacy preferences: quantifying accuracy and user-burden tradeoffs,” *Personal and Ubiquitous Computing*, vol. 15, no. 7, pp. 679–694, 2011.

[31] E. Toch, J. Cranshaw, P. H. Drielsma, J. Y. Tsai, P. G. Kelley, J. Springfield, L. Cranor, J. Hong, and N. Sadeh, “Empirical models of privacy in location sharing,” in *Proceedings of the 12th ACM international conference on Ubiquitous computing*. ACM, 2010, pp. 129–138.

[32] M. Gruteser and D. Grunwald, “Anonymous usage of location-based services through spatial and temporal cloaking,” in *Proceedings of the 1st international conference on Mobile systems, applications and services*. ACM, 2003, pp. 31–42.

[33] M. F. Mokbel, C.-Y. Chow, and W. G. Aref, “The new casper: Query processing for location services without compromising privacy,” in *Proceedings of the 32nd international conference on Very large data bases*. VLDB Endowment, 2006, pp. 763–774.

[34] G. Ghinita, P. Kalnis, A. Khoshgozaran, C. Shahabi, and K.-L. Tan, “Private queries in location based services: anonymizers are not necessary,” in *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*. ACM, 2008, pp. 121–132.

[35] G. Cormode, C. Procopiuc, D. Srivastava, E. Shen, and T. Yu, “Differentially private spatial decompositions,” in *Data engineering (ICDE), 2012 IEEE 28th international conference on*. IEEE, 2012, pp. 20–31.

[36] M. Bernstein, E. H. Chi, L. Chilton, B. Hartmann, A. Kittur, and R. C. Miller, “Crowdsourcing and human computation: systems, studies and platforms,” in *CHI’11 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2011, pp. 53–56.

[37] L. R. Varshney, “Privacy and reliability in crowdsourcing service delivery,” in *SRII Global Conference (SRII), 2012 Annual*. IEEE, 2012, pp. 55–60.

[38] H. Kajino, H. Arai, and H. Kashima, “Preserving worker privacy in crowdsourcing,” *Data Mining and Knowledge Discovery*, vol. 28, no. 5-6, pp. 1314–1335, 2014.

[39] D. Linders, “From e-government to we-government: Defining a typology for citizen coproduction in the age of social media,” *Government Information Quarterly*, vol. 29, no. 4, pp. 446–454, 2012.

[40] E. Ferro, E. N. Loukis, Y. Charalabidis, and M. Osella, “Evaluating advanced forms of social media use in government,” 2013.

[41] R. Medaglia and L. Zheng, “Mapping government social media research and moving it forward: A framework and a research agenda,” *Government Information Quarterly*, vol. 34, no. 3, pp. 496–510, 2017.

[42] E. Ferro, E. N. Loukis, Y. Charalabidis, and M. Osella, “Policy making 2.0: From theory to practice,” *Government Information Quarterly*, vol. 30, no. 4, pp. 359–368, 2013.

[43] A. V. Roman and H. T. Miller, “New questions for e-government: Efficiency but not (yet?) democracy,” in *Public Affairs and Administration: Concepts, Methodologies, Tools, and Applications*. IGI Global, 2015, pp. 2209–2227.

[44] Ø. Sæbø, J. Rose, and L. S. Flak, “The shape of eparticipation: Characterizing an emerging research area,” *Government information quarterly*, vol. 25, no. 3, pp. 400–428, 2008.

[45] M. Hosseini, A. Shahri, K. Phalp, J. Taylor, and R. Ali, “Crowdsourcing: A taxonomy and systematic mapping study,” *Computer Science Review*, vol. 17, pp. 43–69, 2015.

[46] G. Xintong, W. Hongzhi, Y. Song, and G. Hong, “Brief survey of crowdsourcing for data mining,” *Expert Systems with Applications*, vol. 41, no. 17, pp. 7987–7994, 2014.

[47] European parliament: Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general data protection regulation).

[48] C. Fu, Z. Shaobin, S. Guangjun, and G. Mengyuan, “Crowdsourcing leakage of personally identifiable information via sina microblog,” in *International Conference on Internet of Vehicles*. Springer, 2014, pp. 262–271.

[49] G. T. Duncan and R. W. Pearson, “Enhancing access to microdata while protecting confidentiality: Prospects for the future,” *Statistical Science*, pp. 219–232, 1991.

[50] A. Acquisti, S. Gritzalis, C. Lambrinoudakis, and S. di Vimercati, *Digital privacy: theory, technologies, and practices*. CRC Press, 2007.

[51] E. F. Stone, H. G. Gueutal, D. G. Gardner, and S. McClure, “A field experiment comparing information-privacy values, beliefs, and attitudes across several types of organizations.” *Journal of applied psychology*, vol. 68, no. 3, p. 459, 1983.

[52] N. Mohamed and I. H. Ahmad, “Information privacy concerns, antecedents and privacy measure use in social networking sites: Evidence from malaysia,” *Computers in Human Behavior*, vol. 28, no. 6, pp. 2366–2375, 2012.

[53] H. Krasnova, E. Kolesnikova, and O. Guenther, “” it won’t happen to me!”: self-disclosure in online social networks,” 2009.

[54] (2017, November). [Online]. Available: http://bit.ly/2Esh8us

[55] V. Diamantopoulou, C. Kalloniatis, S. Gritzalis, and H. Mouratidis, “Supporting privacy by design using privacy process patterns,” in *IFIP International Conference on ICT Systems Security and Privacy Protection*. Springer, 2017, pp. 491–505.

[56] V. Diamantopoulou, N. Argyropoulos, C. Kalloniatis, and S. Gritzalis, “Supporting the design of privacy-aware business processes via privacy process patterns,” in *Research Challenges in Information Science (RCIS), 2017 11th International Conference on*. IEEE, 2017, pp. 187–198.

[57] E. Gamma, *Design patterns: elements of reusable object-oriented software*. Pearson Education India, 1995.

[58] N. Yoshioka, H. Washizaki, and K. Maruyama, “A survey on security patterns,” *Progress in informatics*, vol. 5, no. 5, pp. 35–47, 2008.

[59] C. Alexander, *A pattern language: towns, buildings, construction*. Oxford university press, 1977.

[60] J. O. Borchers, “A pattern approach to interaction design,” in *Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques*. ACM, 2000, pp. 369–378.