

# **Studying user behavior through a participatory sensing framework in an urban context**

Dissertation submitted by Ngo Manh Khoi to apply for the joint doctorate degree from the Universitat Jaume I, Universidade Nova de Lisboa and Universität Münster.



**UNIVERSITAT  
JAUME·I**

## **European Joint Doctorate Marie Skłodowska-Curie in Geoinformatics**

Doctoral School from Universitat Jaume I



**Ngo Manh Khoi**

**Supervisors:**

Sven Casteleyn      Edzer Pebesma      Roberto Henriques  
(Universitat Jaume I)    (Universität Münster)    (Universidade Nova de Lisboa)

**Castelló de la Plana (Spain)  
November 2018**





This dissertation is funded by the European Commission within the Marie Skłodowska-Curie Actions (ITN-EJD). Grant Agreement num. 642332 - GEO-C - H2020-MSCA-ITN-2014.



*This thesis is dedicated to my family and my girlfriend.*



# Acknowledgments

I would like to thank Dr. Sven Casteleyn, Prof. Dr. Edzer Pebesma and Dr. Roberto Henriques for their guidance and support throughout this work. Their supervision and expertise within the field has been an inspiration during this time and has made my work both enjoyable and stimulating. This report is the result of a long process, with many tiring days, late nights working with the programming codes, sleepless nights, joyful moments of getting good results, sadness when something went wrong suddenly. This thesis is dedicated to my dear supervisors.

At the same time, I would like to thank the two reviewers for their valuable comments based on which I can significantly improve my thesis.

Throughout my Ph.D program, I am grateful for receiving the various kinds of support from Sergi Trilles Oliver and Auriol Degbelo. Without their helps, my research work would have not been successful.

Along with the professors, my colleagues are an important part of motivation. I would like to thank Diego, Mehdi, Fernando, Manuel and Mijail for giving me helps and inspirations.

My parents and my fiancée, Bich Diep, are the ones I often talked to when I felt most tired or frustrated. They always encouraged me to go forward, relieved me of worries and sadness. Without them, it was impossible for me to finish this thesis work.





# Resumen

En las ciudades modernas, existe una gran demanda de diferentes tipos de información urbana que pueden ser procesados y analizados con la finalidad de ayudar a la planificación urbana. Los métodos tradicionales de recopilación de información urbana, como son las redes de sensores inalámbricas, han mostrado diferentes debilidades, como su alto coste, la baja resolución espacial o el no adaptarse rápidamente a nuevos cambios. La llegada de nuevos dispositivos conectados a Internet con sensores integrados y capacidades multimedia, ha dado lugar a una alternativa para los métodos tradicionales de recopilación de datos: detección participativa. La que permite a un gran número de usuarios funcionar como si de sensores inteligentes el movimiento se tratase, que contribuyen con una gran variedad de información, desde observaciones subjetivas y opiniones hasta mediciones sensoriales y/o contenidos multimedia.

Aunque la detección participativa tiene varios potenciales para ser utilizados como un método para recopilar información a gran escala, también tiene varias debilidades propias que podrían entrometer su efectividad. En otras palabras, la detección participativa plantea varios desafíos de investigación, como la calidad de los datos recopilados, la distribución espacial desigual (y temporal) de los participantes, la privacidad y seguridad de los usuarios, la falta de una muestra adecuada de usuarios o su reutilización de los frameworks de detección participativa. Esta tesis tiene como objetivo abordar las siguientes tres preguntas de investigación: 1 / RQ1: ¿Cuáles son las características y funcionalidades necesarias de un framework de detección participativa? 2 / RQ2: ¿Cómo se comportan los participantes en un contexto de detección participativa? 3 / RQ3 - ¿Cómo pueden los organizadores motivar a los participantes a contribuir en la detección participativa?

Debido a la generalización de los dispositivos móviles y la aplicabilidad del

concepto de detección participativa, existe una lista amplia y creciente de aplicaciones de detección participativa. Esta tesis empieza con una revisión extensa de las aplicaciones de detección participativa. Las que, pueden clasificarse según la participación del usuario humano, el alcance de la recopilación de datos, el tipo de fuentes de los datos recopilados y la arquitectura del sistema de detección participativa. En términos de funcionalidad, las aplicaciones de detección participativa se pueden dividir en dos grupos: aplicaciones de función única y aplicaciones multipropósito (frameworks de detección participativa). Las aplicaciones de detección participativa de función única se han utilizado en escenarios como la medición del ruido, la monitorización del tráfico o de señales de radio, la monitorización ambiental o el reporte de incidentes en una ciudad. De forma desigual, cada uno de los frameworks de detección participativa revisados en la tesis puede funcionar en más de un escenario. Estos frameworks representan un cambio significativo en la realización del concepto de detección participativa, lo que resulta beneficioso por diferentes razones. En primer lugar, desde la perspectiva de los organizadores de detección participativa, el coste del desarrollo de software se reduce sustancialmente ya que hay menos plataformas de software para desarrollar y mantener. En segundo lugar, los usuarios finales tendrán que utilizar un número menor de plataformas, lo que les ayudará a familiarizarse con el concepto de detección participativa más rápidamente. En tercer lugar, estos frameworks pueden adaptarse fácilmente a las nuevas demandas de recopilación de datos modificando su estructura de datos interna y flujos de trabajo. Finalmente, estos frameworks permiten a los investigadores y otras partes interesadas estudiar los comportamientos y preferencias de los usuarios cuando los usuarios se enfrentan simultáneamente con una serie de tareas y solicitudes de detección. La revisión de los trabajos continúa con la comparación de aplicaciones y frameworks de detección participativa basados en varias características importantes: la flexibilidad de la herramienta, la presencia de incentivos, la selección de recolectores de datos, la entrega de resultados, la identidad de los recolectores de datos y la flexibilidad del proceso de detección. Estas características están directamente relacionadas con los desafíos de investigación mencionados. Posteriormente, la comparación ofrece como resultado una lista de 7 requisitos funcionales y 2 no funcionales para el marco de detección participativo. Los requisitos funcionales son: 1) Requisito F1 - Capacidad multiusuario; 2) Requisito F2 - Permitir interrupciones en la red; 3) Requisito F3 - Capacidad para identificar a los participantes;

4) Requisito F4 - Capacidad para especificar con flexibilidad el contenido de la campaña de detección; 5) Requisito F5 - Detección sensible al contexto; 6) Requisito F6 - Incentivo para los participantes; 7 / Requisito F7 - Seguridad y privacidad para los datos recopilados. Los requisitos no funcionales son: NF1 - Inclusividad y NF2 - Transparencia.

Una vez identificados los requisitos para la detección participativa, se propone un framework propio de detección participativa, llamado Citizense. Este framework incorpora todos los requisitos identificados con anterioridad en su diseño e implementación. Citizense está diseñado como un framework genérico de detección participativa orientado al usuario, con el objetivo de hacer que la detección participativa sea accesible para todos los usuarios, independientemente de su función (es decir, definición de campañas de recopilación de datos, recopilación de los datos necesarios) y experiencia. El principio de diseño central de Citizense es dividir un proceso complejo de participación en varias tareas de participación atómicas, donde cada tarea se ocupa de un solo tipo de dato o sensor. De esta manera, la campaña de detección participativa se puede diseñar a través de una interfaz gráfica de usuario donde los detalles técnicos complejos y las especificaciones se transforman en elementos visuales simples y comprensibles. Al mismo tiempo, el framework puede reutilizarse en diferentes escenarios y recopilar una amplia variedad de datos. Dos actores principales interactúan con Citizense: autores de campañas y recolectores de datos. Los primeros definen el proceso de detección participativa para recopilar ciertos datos y los últimos proporcionan a los primeros los datos necesarios.

El framework de Citizense adopta una arquitectura cliente-servidor tradicional, con una plataforma central que contiene la lógica de back-end (que incluye, gestión de incentivos y contextos, lógica de detección participativa general, manejo y procesamiento de resultados), almacena todos los datos (es decir, datos administrativos, campañas, resultados de participación) y sirve los resultados a las diversas aplicaciones de administración de campañas basadas en la Web ( Campaign Editor, Feedback Manager o Result Visualizer), y desde la parte del cliente, que consiste en la aplicación móvil, que se conecta al servidor central y permite a los ciudadanos participar en campañas. Hay 4 componentes principales en la arquitectura de Citizense: el objeto campaña, el servidor central, el administrador de la campaña y la aplicación móvil.

Los objetos campaña son centrales en el framework de Citizense: contienen

todos los detalles de una campaña de detección participativa. El editor de la campaña es capaz de crear un objeto campaña, luego lo almacena en el servidor central y lo envía a la aplicación móvil para su ejecución.

A este servidor central se le asignan tres tareas principales: (i) facilitar a los recolectores de datos las campañas relevantes para el contexto en el que se encuentren, (ii) procesar los datos enviados por los recolectores, (iii) enviar retroalimentación en tiempo real a los recolectores de datos. El administrador de la campaña tiene asignadas dos tareas principales: (i) crear los objetos de la campaña, y (ii) visualizar los resultados. Finalmente, la aplicación móvil se instala en los dispositivos de los recolectores de datos y se le asignan dos tareas principales: (i) recibir y visualizar objetos campaña, y (ii) ser el punto de entrada para los recolectores de datos.

Este framework fue probado por validar su escalabilidad, que es una característica importante de la detección participativa. Durante estas pruebas de estrés, el framework Citizense mostró estabilidad. Específicamente, el tiempo de respuesta se midió en tres escenarios diferentes: un gran número de campañas; campañas complicadas con una gran variedad de tipos de tareas y parámetros; y con un gran número de solicitudes simultáneas. Las pruebas de estrés afirmaron que Citizense se desempeñó bien en estas duras condiciones. Más concretamente, el tiempo de respuesta medio fue inferior a 500 milisegundos. Además, el despliegue en la nube permite de forma sencilla su ampliación, en caso de que se requiera un rendimiento más estricto.

Se implementó un framework operacional de Citizense en dos escenarios del mundo real, en los que las funcionalidades y características del framework se probaron con participantes reales y se obtuvieron resultados para el análisis y discusiones adicionales. En el primer despliegue, se usó Citizense para recopilar comentarios de los participantes del maratón anual de la ciudad de Castellón. Se creó una campaña mediante el creador de Citizense, la que estuvo disponible en español, catalán e inglés, y estaba formada de 28 preguntas. Esta campaña se caracterizó por tener sus preguntas vinculadas mediante transiciones condicionales e incondicionales. Hubo 180 participantes involucrados en este despliegue, entre ellos 140 participantes enviaron sus comentarios. Se recibieron 10 envíos realizadas en modo fuera de línea.

El segundo despliegue tuvo lugar en el campus de la Universitat Jaume I, en el que participaron 359 usuarios en 20 días. Entre estos 359 participantes, hubo

230 recolectores de datos (99 hombres y 131 mujeres). En este despliegue, hubo 44 campañas lanzadas secuencialmente, entre ellas 7 campañas se aplicaron con 3 mecanismos de incentivos diferentes (lotería, micropago fijo y micropago variable). Respecto a la autoría de las campañas, 10 campañas fueron creadas por el organizador y el resto (34 campañas) las crearon los usuarios, utilizando el administrador de campañas. Durante el período de 20 días, este despliegue recibió 4.167 solicitudes y 2.944 envíos de datos para las 44 campañas.

A través de las 7 campañas aplicadas con incentivos, se estudió la reacción de los usuarios. Se observó que los mecanismos de incentivo aumentan sustancialmente la participación de los recolectores de datos. Entre los mecanismos de incentivos implementados, el micropago fijo supero al micropago variable. El orden de aceptación general que se pudo fijar fue: micropago fijo > pago al estilo de lotería > micropago variable > sin incentivo. Además, se comprobó la reacción de los usuarios dependiendo del género y la edad. El mismo orden entre los mecanismos de incentivo mostrado de forma general, se puede establecer en los diferentes grupos de recolectores de datos masculinos, femeninos, jóvenes y ancianos. Finalmente, se observa que existe una correlación negativa entre la edad y la cantidad de envíos, a edad más avanzada menos envíos se presentan. Usando la hora y la ubicación de los envíos, se pueden obtener varias conclusiones sobre el comportamiento del usuario. En primer lugar, los recolectores de datos tienden a presentar datos durante la noche. Sin embargo, para ciertas campañas, como las adaptadas al contexto, los envíos se realizaron en el momento en que los usuarios accedieron a la campaña. En segundo lugar, distinguimos dos grupos de recolectores de datos: un grupo más activo de recolectores de datos, que usan regularmente la aplicación móvil, y un mayor número de recolectores de datos regulares, que usan la aplicación móvil con menos frecuencia. Tercer punto, si una campaña tenía un contexto geográfico, un porcentaje significativo de los envíos se realizaba dentro del área definida. En conclusión, la tesis ha abordado completamente las preguntas de investigación. La tesis ha contribuido a la creación de Citizense, un framework de detección participativa y al estudio de diferentes aspectos del comportamiento del usuario en el contexto de la detección participativa. También se presentan diferentes direcciones de futuras investigación.

**Palabras clave:** *detección participativa, comportamiento del usuario, incentivos, participación ciudadana, compromiso del usuario, ciudad inteligente.*



# Abstract

The widespread use of mobile devices has given birth to participatory sensing, a data collection approach leveraging the sheer number of device users, their mobility, intelligence and device's increasingly powerful computing and sensing capabilities. As a result, participatory sensing is able to collect various types of information at a high spatial and temporal resolution and it has many applications ranging from measuring cellular signal strength or road condition monitoring to observing the distribution of birds. However, in order to achieve better results from participatory sensing, some issues needed to be dealt with. On a high level, this thesis addressed two issues: (1) the design and development of a participatory sensing framework that allows users to flexibly create campaigns and at the same time collect different types of data and (2) the study of different aspects of the user behaviors in the context of participatory sensing.

In particular, the first contribution of the thesis is the design and development of *Citizensense*, a participatory sensing framework that facilitates flexible deployments of participatory sensing campaigns while at the same time providing intuitive interfaces for users to create sensing campaigns and collect a variety of data types. During the real-world deployments of *Citizensense*, it has shown its effectiveness in collecting different types of urban information and subsequently received appreciation from different stakeholders. The second contribution of the thesis is the in-depth study of user behavior under the presence of different monetary incentive mechanisms and the analysis of the spatial and temporal user behavior when participants are simultaneously exposed to a large number of participatory sensing campaigns. Concerning the monetary incentive, it is observed that participants prefer fixed micro-payment to other mechanisms (i.e., lottery, variable micro-payment); their participation was increased significantly when they were given this incentive. When taking part in the participatory sensing process, par-

ticipants exhibit certain spatial and temporal behaviors. They tend to primarily contribute in their free time during the working week, although the decision to respond and complete a particular participatory sensing campaign seems to be correlated to the campaign's geographical context and/or the recency of the participants' activities. Participants can be divided into two groups according to their behaviors: a smaller group of active participants who frequently perform participatory sensing activities and a larger group of regular participants who exhibit more intermittent behaviors

**Keywords:** *participatory sensing, user behavior, framework, monetary incentive, data collection.*



# Contents

<b>Resumen</b>	<b>v</b>
<b>Abstract</b>	<b>xi</b>
<b>Acronyms</b>	<b>xxiii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation . . . . .	1
1.2 Research challenges in participatory sensing . . . . .	4
1.3 Contributions . . . . .	7
1.4 The GEO-C joint doctorate program . . . . .	11
1.5 Thesis organization . . . . .	13
<b>2 Backgrounds and related works</b>	<b>15</b>
2.1 Participatory sensing . . . . .	15
2.1.1 Introduction . . . . .	15
2.1.2 Classification of participatory sensing . . . . .	17
2.1.3 Participatory sensing applications . . . . .	18
Noise monitoring . . . . .	18
Traffic monitoring . . . . .	19
Radio signal monitoring . . . . .	19
Environmental monitoring . . . . .	19
City incidents monitoring . . . . .	20
Other applications . . . . .	20
2.1.4 Participatory sensing frameworks . . . . .	21

2.1.5	Comparison of participatory sensing applications and frameworks . . . . .	25
2.2	The use of incentive in participatory sensing . . . . .	32
2.2.1	Classification of incentives . . . . .	32
2.2.2	Deployments of monetary incentive in a real context . . . . .	34
2.3	Users' spatial and temporal behaviors . . . . .	36
2.3.1	The use of spatial and temporal coverage in user recruitment for participatory sensing . . . . .	36
2.3.2	The users' spatial and temporal behavior in contributing data . . . . .	38
<b>3</b>	<b>The Citizense framework</b>	<b>41</b>
3.1	The design principle . . . . .	41
3.2	The Citizense architecture . . . . .	43
3.2.1	The campaign object . . . . .	43
3.2.2	The central platform - Storage layer . . . . .	45
3.2.3	The central platform - Application logic layer . . . . .	46
3.2.4	The central platform - Application layer . . . . .	47
3.2.5	The mobile application . . . . .	47
3.3	The implementation . . . . .	48
3.3.1	The campaign object . . . . .	48
3.3.2	The central platform - Storage layer . . . . .	48
3.3.3	The central platform - Application logic layer . . . . .	48
3.3.4	The central platform - Application layer . . . . .	52
3.3.4	Campaign creator . . . . .	52
3.3.4	Result visualization and feedback manager . . . . .	55
3.3.5	The mobile application . . . . .	58
3.3.6	Comparison of Citizense and other applications . . . . .	62
3.4	Technical testing of the framework . . . . .	65
3.4.1	Accuracy of measurement . . . . .	65
3.4.2	Scalability of the framework . . . . .	66
<b>4</b>	<b>Deploying Citizense in real-world scenarios</b>	<b>69</b>

4.1	The Castellón marathon . . . . .	69
4.1.1	The goals of the deployment . . . . .	69
4.1.2	The setup and progress of the deployment in the marathon . . . . .	69
4.1.3	The results . . . . .	70
4.1.4	Discussions . . . . .	71
4.2	The deployment in UJI campus . . . . .	72
4.2.1	The goals of the deployment . . . . .	72
4.2.2	The setup and progress of the deployment . . . . .	73
4.2.3	The results . . . . .	77
4.2.4	Discussions . . . . .	78
4.2.5	Limitations of the UJI deployment . . . . .	81
<b>5</b>	<b>The effectiveness of monetary incentives</b>	<b>83</b>
5.1	Raw data and descriptive analysis . . . . .	84
5.2	Overall statistical analysis . . . . .	90
5.3	Studying the influence of gender . . . . .	97
5.4	Studying the influence of data collectors' age . . . . .	100
5.5	Further analysis on incentive mechanisms . . . . .	104
5.6	Conclusions of the effectiveness of monetary incentive . . . . .	109
<b>6</b>	<b>Temporal and spatial user behavior</b>	<b>113</b>
6.1	Temporal user behavior: frequency of use . . . . .	114
6.2	Temporal user behavior: distribution of use . . . . .	117
6.3	Spatial user behavior . . . . .	119
6.4	Conclusions on the user behavior in participatory sensing . . . . .	126
<b>7</b>	<b>Conclusions</b>	<b>129</b>
7.1	Answers to the research questions . . . . .	130
7.2	Contributions of the thesis . . . . .	132
7.3	Results of the thesis . . . . .	133
7.4	Limitations . . . . .	134
7.5	Outlook . . . . .	135

<b>A</b>	<b>The campaign on usability and user feedback</b>	<b>139</b>
<b>B</b>	<b>The details of the campaigns in the UJI deployment</b>	<b>141</b>
	<b>Bibliography</b>	<b>145</b>

## List of tables

2.1	Analyzing the selected participatory sensing applications . . . . .	27
2.2	Analyzing the selected participatory sensing applications (continued)	28
2.3	Analyzing the selected participatory sensing applications (continued)	29
3.1	The campaign object. . . . .	49
3.2	The participant object. . . . .	50
3.3	The object result. . . . .	50
3.4	Definitions of the primitives used in the logical expression . . . . .	55
3.5	How can Citizense substitute other existing participatory sensing applications and frameworks . . . . .	64
4.1	The different parameters of the campaigns with incentives. . . . .	76
5.1	The results of the Mann-Whitney U test with and without BH correction for all pairs of incentive mechanisms, using the number of opens during the full period (all) and only weekdays. . . . .	94
5.2	Average openings per incentive mechanism, for all campaigns $C_1-C_4$ . . . . .	94
5.3	The results of the Mann-Whitney U test with and without BH correction for all pairs of incentive mechanisms, using the number of submissions during the full period (all) and the weekdays. . . . .	95
5.4	Average submissions per incentive mechanism, for all campaigns $C_1-C_4$ . . . . .	96
5.5	The results of the Mann-Whitney U tests with and without BH correction for all pairs of incentive mechanisms, using the number of submissions from female and male data collectors. . . . .	98

5.6	Average submissions per incentive mechanism by all female and male data collectors, for all campaigns $C_1$ – $C_4$ . . . . .	100
5.7	The results of the Mann-Whitney U tests, with and without BH correction for all pairs of incentive mechanisms, using the number of submissions from younger and older data collectors. . . . .	103
5.8	Average submissions per incentive mechanism by younger and older data collectors, for all campaigns $C_1$ – $C_4$ . . . . .	104
6.1	The on-target ratio of the geographically-related campaigns with different values of buffer distance $d$ . . . . .	124
A.1	The questions on the usability of the Citizense mobile application .	139
A.2	The questions on the the concept of participatory sensing . . . . .	140
B.1	Overview of all campaigns available during the deployment, sorted by the number of submissions. . . . .	142
B.2	Overview of all campaigns available during the deployment, sorted by the number of submissions (continued). . . . .	143

## List of figures

1.1	The structure of the GEO-C program. . . . .	12
1.2	The general structure of the thesis. . . . .	13
3.1	The Citizense application ecosystem. . . . .	42
3.2	The architecture of Citizense framework. . . . .	43
3.3	Activities performed by the data collectors using the central platform - application layer. . . . .	53
3.4	The graphical campaign creator. Creating a campaign, step 1: Specifying the various parameters of the campaign. . . . .	54
3.5	The graphical campaign creator. Creating a campaign, step 2: Defining the atomic sensing tasks and their transitions. . . . .	56
3.6	The visualization of the results. (left) The grouping by submission, each submission's status can be updated, at the bottom of the picture. (right) The grouping by task. . . . .	56
3.7	Visualizing the results on a map . . . . .	57
3.8	Activities performed by the data collectors using the mobile application. . . . .	59
3.9	The client mobile application. From left to right (1) An atomic sensing task. (2) The list of available campaigns. (3) The details of a participant's account. (4) The status/feedback of the submissions. . . . .	61
4.1	Example of the submissions during the deployment displayed by the Citizense result visualizer. (upper) The different text answers to a text-input task. (lower) The pictures and locations of the graffiti gathered by data collectors. . . . .	79

5.1	The number of list views of the campaigns. For each individual bar, dark ( <b>bottom part</b> ) and light color ( <b>top part</b> ) represent weekdays and weekends, respectively. . . . .	85
5.2	The number of opens of the campaigns. For each individual bar, dark ( <b>bottom part</b> ) and light color ( <b>top part</b> ) represent weekdays and weekends, respectively. . . . .	86
5.3	The number of submissions of the campaigns. For each individual bar, dark ( <b>bottom part</b> ) and light color ( <b>top part</b> ) represent weekdays and weekends, respectively. . . . .	87
5.4	The completion ratios of the campaigns based on all days. . . . .	88
5.5	The general completion ratios of the campaigns based on all days. . . . .	89
5.6	The attraction indexes of the campaigns based on all days. . . . .	89
5.7	The age count and average number of submissions per person per age, for campaigns $C_1-C_4$ . . . . .	102
5.8	The approximate order of the incentive mechanisms, for the overall data collectors and their subgroups based on number of submissions (significance level of 0.05). The significant differences between the mechanisms are marked by red arrows. . . . .	108
6.1	An illustrative example of requests sent by a typical data collector over time, each request is represented by a blue dot. A session is a cluster of requests, marked by the red ellipses. . . . .	115
6.2	The number of sessions (blue) and average number of request per session (red) for every data collector throughout the deployment considering a separation time $x = 600$ seconds, decreasingly ranked according to the number of sessions. . . . .	117
6.3	The distribution of all received requests and submissions throughout the deployment rendered in a 24-hour time frame. . . . .	118
6.4	The distribution of submissions from different campaigns throughout the deployment rendered in a 24-hour time frame. (upper) The campaigns having different patterns than the overall submission pattern. (lower) The campaigns having a similar pattern to the overall submission pattern. . . . .	119
6.5	The taxonomy of campaigns used in the Citizense deployment. . . . .	121



6.6 The data collectors' spatial pattern in different campaigns. (left) submissions from campaign 1 (geographically-neutral) addressing the topic of citizen participation, with the black rectangular frame covering the map extent on the right side. (right) submissions from campaign 14 (geographically-related) asking feedback on the city's only tram line, with the blue line representing the tram route and the purple area representing the buffer of 400 meters. . . . . 122

6.7 The submissions of the campaign "City tram line", represented by the red points. (left) The 400-meter catchments around the tram stations. (right) The 400-meter circular buffers around the tram stations. . . . . 126



# Acronyms

<b>AES</b>	Advanced Encryption Standard
<b>API</b>	Application programming interface
<b>APK</b>	Android Package
<b>BSSID</b>	Basic Service Set Identifier
<b>CPU</b>	Central processing unit
<b>CSM</b>	Cross-community mining and sensing
<b>CSS</b>	Cascading Style Sheets
<b>GPS</b>	Global Positioning System
<b>GSM</b>	Global System for Mobile communications
<b>HTML</b>	Hypertext Markup Language
<b>HTTP</b>	Hypertext Transfer Protocol
<b>ICT</b>	Information and communications technology
<b>iOS</b>	a mobile operating system created and developed by Apple
<b>JSON</b>	JavaScript Object Notation
<b>LTE</b>	Long-Term Evolution, a 4G mobile communications standard
<b>OCT</b>	Open City Toolkit
<b>OSM</b>	Open Street Map

<b>PR</b>	Public relations
<b>SSID</b>	Service Set Identifier
<b>SUS</b>	System Usability Scale
<b>UJI</b>	Universitat Jaume I
<b>WSN</b>	Wireless sensor network

# Chapter 1

## Introduction

The main goal of this introductory chapter is to familiarize the readers with the concept of participatory sensing. In this chapter, we define what is participatory sensing, why it is needed in the current urban context and the challenges facing participatory sensing. We continue with examples of existing participatory sensing applications and frameworks, and subsequently point out their strengths and weaknesses. As a result of this analysis, this thesis is dedicated to the introduction of Citizense participatory sensing framework and the study of the framework's effectiveness in collecting urban information as well as the participants' behavior in the participatory sensing context.

Chapter 1 is organized as follows: Section 1.1 presents the motivation for the development of participatory sensing. The research challenges in the field of participatory sensing are presented in Section 1.2. Section 1.3 details the contribution of this thesis to the participatory sensing body of knowledge. Section 1.4 describes the Geo-C Joint Doctorate program, aiming to enable Open cities. In this context, the Citizense framework acts as a tool for citizens to collect different types of urban data and communicate with the local authority. Finally, Section 1.5 details the outline of the thesis.

### 1.1 Motivation

The world population is growing at a considerably fast pace. Based on the number given by the United States Census Bureau, the world population is at 7,475

billions<sup>1</sup> (as of today 26 May 2018). According to well-established projections, the world population will reach at least 8.75 billions in 2050 (Lutz et al., 2017) and 11,184 billions in 2100 (United Nations, 2017). This growing population have caused problems such as depletion of natural resources and degradation of environment. Another problem with the growing population is its uneven distribution, with more people living in urban areas than anywhere else. It is reported that 54% of the world population are currently living in metropolitan areas, and this number is expected to reach 66% by 2050 (United Nations, 2014). As a result, the term "megacities" is coined to represent cities whose population exceed 10 millions inhabitants (Cross, 2001; United Nations, 2008). The list of megacities is also growing; some examples are Tokyo, Delhi and Shanghai (in Asia), Paris, London and Moscow (in Europe), Lagos (in Africa), New York (in North America) and São Paulo (in South America).

The process of urbanization can bring about positive changes as it allows easier access to education, health care, transportation, energy supply and water for a substantially large number of inhabitants. However, at the same time, this urbanization process has its inherent problems and issues. Examples are overcrowdedness, traffic congestion, environmental pollution, crime and safety issues, the impact of natural disasters, unemployment and lack of housing. The United Nations, in their report, suggested the use of Information and Communication Technology (ICT) to resolve the problems brought by urbanization (United Nations, 2014).

Specifically, this use of ICT in cities and urban areas are directly related to the concept of Smart Cities - a concept that aims to provide a sustainable, safe and liveable environment for all the inhabitants. Due to its importance, Smart Cities have gained the attention of many scholars, practitioners, the government and citizens; there have been many definitions of Smart cities (Albino et al., 2015). From a holistic point of view, the Smart cities concept typically includes 6 domains: smart economy, smart people, smart governance, smart mobility, smart environment and smart living (Rudolf et al., 2007). At the center of the Smart Cities concept is the demand to collect different types of information in an efficient and proper manner and later process, store and visualize the obtained information. Historically, information was obtained through the use of wireless sensor networks (WSN) (Yick et al., 2008). Recently, the advent of modern Internet-connected

---

<sup>1</sup>World Population clock - <http://www.census.gov/popclock/>

smartphones with powerful sensing and processing capabilities has given birth to a new data collecting approach: participatory sensing, which combines the mobility of people and the computing power of these devices. Specifically, participatory sensing is the concept in which a (large) group of people contribute (sensory) information to form a body of knowledge. To illustrate this participatory sensing concept, let us consider the following example of creating a noise map in the city. Noise pollution has been proven to cause many adverse effects on the human health, such as contribution to cardiovascular effects and an increased incidence of coronary artery disease (Münzel et al., 2018; Hoffmann et al., 2006). As a result, there is a demand for the control of noise pollution, especially in big cities and urban areas. A necessary condition for implementing noise control is the creation of a city's noise map. Traditionally, this noise map can be created by interpolating the measurements from the relatively few dedicated measuring stations across the city. Accurate noise measurements are recorded at these fixed locations, while the measurements in other locations has to be computed from these established values. In contrast, participatory noise measuring can quickly provide a high-resolution city-wide weather map compiled from thousands of in-situ measurements spread across the city produced by the participants' smartphones. In this noise measuring example, the participatory measuring approach shows its merits as noise data can be collected with high granularity in space and time (D'Hondt et al., 2013).

In parallel with the issue of measuring noise, there is a demand for reporting the presence of mosquitos in the city due to the surge of malaria caused by mosquitos. It is believed that computers and sensors are not effective in detecting the presence of mosquito, or it takes a considerable amount of time, money and efforts to build a system that can effectively detect mosquitos. In contrast, humans are quicker and more effective in detecting the mosquitos surrounding them. As a result, the task of detecting mosquitos can be completed much quicker and with higher accuracy by humans. Information on the presence of mosquitos produced by a large number of individuals might be aggregated into a comprehensive real-time map of mosquitos in the city <sup>2</sup>.

In these two simple examples, the advantages of participatory sensing are revealed. In the first example of noise measuring, participatory noise measuring

---

<sup>2</sup>Mosquito Alert - <http://webserver.mosquitoalert.com/static/tigapublic/spain.html>

leverages the vast amount of smartphones to produce a large number of measurements (or data in general). In the second example of reporting mosquitos, the involvement of humans are decisive when computers cannot handle the task efficiently. Furthermore, as the participants constantly move in the city, the spatial coverage of this participatory sensing is much more extensive than WSN. In terms of capital, participatory sensing is relatively affordable as it does not require any new investment in terms of hardware; it reuses the existing communication infrastructure (LTE, WiFi) which is becoming more pervasive over time. The big task (i.e., measuring noise and reporting mosquitos) in a city-wide scale is divided among the large number of participants, each participant is expected to carry out several simple measurements/observations. In the ideal case, if all the inhabitants of the city join this participatory sensing effort then the resulting high-quality noise map/mosquito map will cover the whole city with several reports in each location.

In general, participatory sensing is a promising approach to collect data, especially in urban context, as it takes advantage of the massive number of participants who own devices with advanced sensing and processing capabilities, their constant mobility and the pervasive communication network. As a result, the data produced by participatory sensing processes have high spatial and temporal resolution. Furthermore, participatory sensing can be applied in a variety of domains, as demonstrated in the surveys of (Khan et al., 2013; Guo et al., 2015; Jaimes et al., 2015; Lane et al., 2010).

## 1.2 Research challenges in participatory sensing

In order to fully exploit the strength of participatory sensing, there are several obstacles that must be overcome. Scholars have pointed out these obstacles in different works such as (Kanhere, 2011; Kapadia et al., 2009; Guo et al., 2015; Gao et al., 2015b; Guo et al., 2016; Montori et al., 2018b). The main research challenges in the field of participatory sensing are the followings:

- **Challenge 1 - Privacy of the participants (Christin, 2016; Montori et al., 2018b):** The prime actor behind participatory sensing are humans, who actively collect data for the common good of the society. During this process, they intentionally or unknowingly reveal the context of the data collection, which in some cases is sensitive. For example, the location and time of the data is added to the measurement to make it meaningful and complete. As



a result, some participants might lose their interest in joining participatory sensing activities as they fear that their data privacy is not protected. Certain privacy mechanisms need to be exercised to ensure the participants' privacy, such as using a unique pseudonym (Deng and Cox, 2009), periodic pseudonym (Christin et al., 2012), data perturbation (Zhu et al., 2013) and spatial cloaking (Agir et al., 2014).

- **Challenge 2 - Enhancing and guaranteeing the quality of the collected data (Budde et al., 2017):** The result of participatory sensing is built on the data contribution of the participants, therefore it can be affected by several factors, both subjective and objective. Subjective factors include the malicious intentions or unintentional errors from the participants while objective factors can be privacy regulations, device calibration and lack of information/participants. This challenge spans across several domains such as user recruitment (Yao et al., 2015), incentive mechanism (Peng et al., 2015; Gao et al., 2015a) and device calibration (Luo et al., 2018; Markert et al., 2016).
- **Challenge 3- The reusability of the participatory sensing platforms (Tangmunarunkit et al., 2015):** The list of participatory sensing applications has been growing. However, many of them are designed to function within limited use cases. As a result, there are a multitude of applications sharing the same functionalities, workflow or data to be collected. From the participants' point of view, a large number of applications with limited functionality, and their overlapping purposes can cause confusion and hinder their effective participation. In order to avoid the limitations of these applications, the idea of reusable participatory sensing applications has been developed and realized in several implementations (Brunette et al., 2013; Zaman and Meuter, 2015; Kim et al., 2013). The resulting participatory sensing applications are able to collect a variety of data types and they can be used across different scenarios.
- **Challenge 4 - Incentivizing the participants (Restuccia et al., 2016):** The results produced by the participatory sensing approach are meaningful only if there is a relevant number of participants producing an adequate number of contributions. In this process, participants might be negatively affected

by subjective obstacles (i.e., privacy concerns, distraction by other mobile apps) and objective obstacles (i.e., battery and mobile data consumption, memory occupation). Therefore, it is recommended to compensate the costs occurred during the participatory sensing process and motivate the participants through the use of incentives. The incentives can come in the form of monetary incentives (Reddy et al., 2010a) and non-monetary incentive, which includes reputation points (Ueyama et al., 2014; Silva et al., 2013b), gamification techniques (Arakawa and Matsuda, 2016), and the exchange of useful information (Yan et al., 2011).

- **Challenge 5 - Integrating participatory sensing with participatory social media (Guo et al., 2014):** Apart from the physical world, another valuable source of user-generated information is from the virtual societies (e.g., social media networks). For instance, for the same sensing object (e.g., an event in the city), it might be reflected in different ways through participatory sensing and social media networks. However, the reflections from the physical and virtual world usually complement each other, resulting in the need for the integration and fusion of the information from these worlds. In this direction, cross-community sensing and mining (CSM) is a new research area that addresses this demand. In this area, researchers try to link the virtual world and the physical world through tools such as location-based social networks (Lima and Musolesi, 2012) and event-based social networks<sup>3</sup>. Other works focus on the definition of cross-community modeling and measurement metrics, which can shed more light on the complex interaction between the physical and the virtual world. In parallel with these works, the question whether it is possible to transfer knowledge between one world (e.g., physical) to another (e.g., virtual) is crucial as it can enhance the performance of the latter (Tang et al., 2012).
- **Challenge 6 - Trust and reputation (Mousa et al., 2015):** As participatory sensing entirely relies on the voluntary data contribution of the users, it might be subjected to erroneous data, produced intentionally or unintentionally by the users. Therefore it is desirable to monitor the user behavior closely so that estimation on the users' honesty can be made. On one hand, malicious

---

<sup>3</sup><http://www.meetup.com>

users can launch different types of attack (e.g., corruption attack, on-off attack, GPS spoofing (Premchaiswadi et al., 2011), discrimination and Sybil attack (Wei et al., 2012)) to the participatory sensing system. On the other hand, several trust and reputation system have been built (Zhang et al., 2012; Irissappane et al., 2012)

- **Challenge 7 - Fairness (Montori et al., 2018b):** Fairness is crucial in participatory sensing as it helps to keep users being involved and prevent them from dropping out as they are not selected as data collector. This characteristic is in direct contrast with data trustworthiness, in which data quality takes precedence over broad participation of users. Fairness are often tackled in conjunction with incentive mechanism Lee and Hoh (2010a), in which a balance between optimal data quality and the well-being of users is the ultimate goal.
- **Challenge 8 - Interoperability (Montori et al., 2018b):** As the number of participatory sensing applications and the amount of data generate by these applications grow constantly, many of these applications function as silos. They have full control on the process of collecting and aggregating data and do not allow co-operation among each other. Therefore, it is desirable to increase the interoperability among applications by allowing them to exchange data, such as the works from (Soldatos et al., 2015) and (Montori et al., 2018a)

## 1.3 Contributions

Among the aforementioned research challenges, this thesis focuses on the issues of reusability in participatory sensing and incentivizing the participants. As the current trend in participatory sensing, reusable participatory sensing applications offer several benefits for different stakeholders. For end-users, they can have a smaller number of applications to deal with, resulting in better user experience. For the participatory sensing organizers, they can drastically reduce the cost of building and maintaining their software applications. For the researchers, they can study several aspects of the user behavior, interaction and preference as users are simultaneously exposed to a number of participatory sensing campaigns and topics.

Participatory sensing entirely relies on the voluntary involvement of users to collect a large amount of information in different spatial and temporal context. Furthermore, users have to bear tangible costs (e.g., energy consumption, data consumption) and intangible cost (e.g., the fear of losing privacy) caused by this data collection process. Therefore, it is desirable to provide users with certain incentives to compensate these costs.

Based on the focuses mentioned above, the following three research questions are formulated:

- RQ1 - What are the necessary characteristics and functionalities of a (multi-purpose) participatory sensing framework?

Due to the different benefits of reusable participatory sensing applications, it is desirable to create such a reusable participatory sensing application (participatory sensing framework). However, there are a large number of participatory sensing applications and several participatory sensing frameworks already existing and they greatly differ from each other in terms of functionality and characteristics. As a result, it is necessary to build such a framework that possesses all the necessary features. The answer to this research question will provide a list of characteristics and functionalities required for the implementation of the framework in question, which is one of the solution for research challenge 3.

- RQ2 - How do participants behave in a participatory sensing context?

The study of user behavior in participatory sensing will provide more understanding on how users perform their sensing tasks. Once these information are obtained, they can be used to improve the participatory sensing process as well as the applications. The answer to this research question can be partially used to address research challenge 2.

- RQ3 - How can organizers motivate the participants to contribute in participatory sensing through the use of monetary incentive?

This research question is directly related to research challenge 4. RQ3 seeks to understand what kinds of monetary incentive are effective in boosting the involvement of users in participatory sensing. Monetary incentive is selected as

it is generic and can potentially be incorporated in every participatory sensing applications.

This thesis has three main objectives. The first objectives is to identify and study the important characteristics of participatory sensing. Second, based on the identified characteristics, a multi-purpose participatory sensing framework is then designed and implemented, incorporating the identified characteristics. Finally, the resulting framework is then tested in real-world scenarios to verify its characteristics and to discover the different aspects of user behavior in the participatory sensing context. Due to the fact that the developed framework is used as a vehicle to study user behavior in participatory sensing, the research challenge 5 and 6 and the accuracy of results (part of research challenge 2) are not tackled in this thesis. However, certain aspects of other research challenges are thoroughly considered during the development of the framework. Research challenge 1 is partially addressed by the use of pseudonym for each data collectors in the framework. A rigid user recruitment based on location, time and user profile is proposed as one of the solutions to research challenge 2. Challenge 3 is fully addressed in this thesis by the creation of the Citizense multi-purpose participatory sensing framework. Subsequently, the research challenge 4 is one of the main focuses of this thesis as it seeks to study how the users react under the presence of (monetary) incentives. Research challenge 7 is partially considered through the theoretically fair nature of the monetary incentives implemented in the framework. Finally research challenge 8 is partially answered as the Citizense framework incorporates a data gateway which allows other applications to query data collected by this framework.

Concerning the research question RQ1, I have identified the set of requirements for a participatory sensing framework. This set stems from the extensive review of existing participatory sensing applications; it contains 7 functional requirements and 2 non-functional requirements. The functional requirements are the following:

- F1 - Multi-purpose: the participatory sensing framework should be able to collect different types of information (e.g., sensory input, multimedia content, input from human participants).
- F2 - Coping with network interruptions: the participatory sensing framework should be able to tolerate disruption in the communication network.

- F3 - Participant identity: The participatory sensing framework should be able to identify participants.
- F4 - Ability to flexibly specify content for the sensing campaign: the author of a sensing campaign should be able to manage the content of a sensing campaign.
- F5 - Context-aware information gathering: the participatory sensing framework should be able use the participants' context information to select eligible participants.
- F6 - Incentives for participants: Incentives should be used to motivate participants and compensate the costs resulted from the their sensing activities.
- F7 - Security and privacy for collected data: The collected data should be protected from unauthorized access.

The non-functional requirements are the following:

- NF1 - Inclusiveness: The framework should be easily understandable and usable, given the diverse backgrounds of participants in an urban environment, possibly those with limited computer knowledge.
- NF2 - Transparency: The participatory sensing process should be transparent to the participants (e.g., participants should be able to view the submitted results and to receive feedback on their submissions).

Based on this set of requirements, we have designed and developed *Citizensense*, a multi-purpose participatory sensing framework that aims to make participatory sensing accessible to all users, regardless of their role (i.e., defining data collection campaigns, gathering the required data) and computer skills. As a multi-purpose participatory sensing framework, *Citizensense* is designed as a union of other applications; it combines and extends the functionalities of existing participatory sensing applications and frameworks. Subsequently, *Citizensense* was deployed in two occasions in which the framework was used by hundreds of people. In these deployments, its designed functionalities were showcased and the framework

was appreciated by different stakeholders. Particularly in the second deployment, Citizense has been proved to be effective in collecting urban data.

By using the Citizense as a vehicle to study the participant's behaviors, we were able to answer research question RQ2 by studying the participants' behaviors and identifying their (spatial and temporal) patterns, which is under-investigated. It can be concluded that (i) the Citizense framework can effectively help participants to design data collecting processes and collect the required data, (ii) data collectors primarily contribute in their free time during the working week; much fewer submissions are done during the weekend, (iii) the decision to respond and complete a particular participatory sensing campaign seems to be correlated to the campaign's geographical context and/or the recency of the data collectors' activities, (iv) data collectors can be divided into two groups according to their behaviors: a smaller group of active data collectors who frequently perform participatory sensing activities and a larger group of regular data collectors who exhibit more intermittent behaviors.

Finally, through one of the deployments of Citizense, we studied the effectiveness of monetary incentives in motivating participants to contribute their data; this study addresses the third research question RQ3. It is concluded that (i) in general, monetary incentives work to improve participation rate; (ii) for the overall population, a general descending order in terms of effectiveness of the incentive mechanisms can be established: fixed micro-payment first, then lottery-style payout and finally variable micro-payment.

## **1.4 The GEO-C joint doctorate program**

Urbanization has been a key trend for centuries and is expected to continue throughout the 21st century as well. Cities have to continuously strive to provide a sustainable, safe and liveable environment for their ever-increasing populations. In recent years, the term 'smart cities' has been coined for initiatives that monitor and analyse different aspects of urban life, and manage service provision intelligently. A key gap in this area relates to how people can understand the processes driving smart cities and their services, and how they can gain a sense of control rather than being controlled by the services provided by a smart city. In this context, GEO-C, a Joint Doctorate program in GeoInformatics, was created to further research in this direction. Specifically, GEO-C aims to contribute methods and

tools to realize smart and open cities, in which all groups of society can participate on all levels and benefit in many ways.

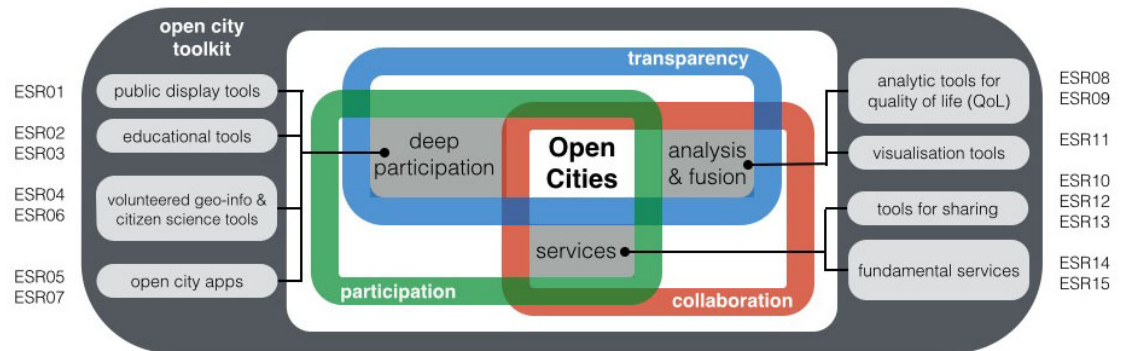


Figure 1.1: The structure of the GEO-C program.

The main goal of GEO-C is to make substantial scientific progress towards the notion of open cities, i.e. smart cities that are open to all citizens and facilitate participation on all levels. To do so, the GEO-C program has three main research strands: analysis and fusion, deep participation and services (see Figure 1.1); there will be 15 theses completed during the GEO-C program. The topic of this thesis (ESR 06) lies in the strand of deep participation as this thesis looks into the interactions among the citizens, the ICT platforms, the data available in the city and the local authorities.

Throughout the GEO-C program, different findings turned into reusable tools for cities are compiled in the Geo-C Open City Toolkit (OCT). Aiming to enable open cities, the OCT compiled and aggregated different tools useful for cities looking to increase citizens participation and transparency. My contribution to the OCT includes the Citizense participatory sensing framework and a guideline for effective usage of the framework.

My first contribution, the Citizense framework, consists of a front-end web interface, a front-end mobile application, and a back-end server. After thorough laboratory testings and several field deployments, this framework is added to the OCT in the category of "App and Services". As an open source software, the source code of the Citizense framework is accessible from the project's OCT repository (Geo-C Joint Doctorate program, 2018), and it is redistributed under the Apache v2 license, which effectively makes the code reusable. As a result,



interested users can download, customize and deploy Citizense in their own ICT infrastructure.

My second contribution to the OCT is the guideline for using the Citizense participatory sensing framework. This guideline provides detailed description about the Citizense framework, its functionalities and steps to create a participatory sensing campaign using this framework. This guideline is added to the OCT in the category of "Guidelines". The intended readers of the guideline are prospective users (to operate the framework) and software engineers (to modify or customize the framework).

## 1.5 Thesis organization

On a high level, the thesis is composed of three parts: the summary of the related works, the design and implementation of the Citizense framework and finally the discussions and analysis on the results of the Citizense deployments. In this structure, the research question RQ1 is fully answered in the first part. In the second part, the design and implementation of the Citizense framework will further clarify the answers to RQ1 as each characteristic and functionality will be explained using the Citizense framework. In the last part of the thesis, research question RQ2 will be addressed through the analysis of the user behaviors. Finally research question RQ3 will be answered by looking at the participants' responses to the monetary incentives administered in a Citizense deployment.

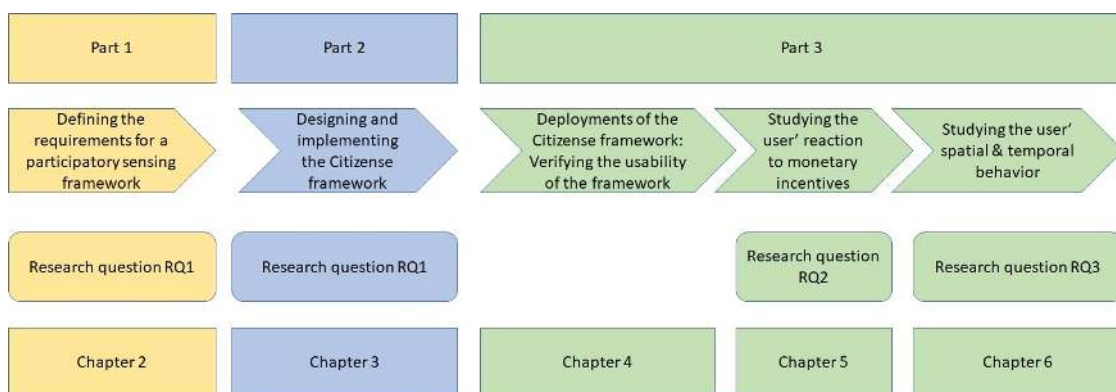


Figure 1.2: The general structure of the thesis.

In terms of chapters, the thesis is composed of 7 chapters. Chapter 1 introduces the readers to the content of the thesis. Chapter 2 is dedicated to the summary of

the related works for the thesis. This chapter addresses the topic of participatory sensing, the use of incentive in participatory sensing and the user behaviors in participatory sensing. In this chapter, a set of key requirements for participatory sensing frameworks is formed. Chapter 3 details the design and implementation of the Citizense framework, which incorporate the identified requirements. Each component of the framework is described in this chapter. Chapter 4 details the various real-world deployments of the Citizense framework and their results. In Chapter 5, the users' behaviors under the presence of monetary incentives are studied. Then in Chapter 6, the users' spatial and temporal behaviors are studied and the corresponding patterns are analyzed. Finally Chapter 7 concludes the thesis.

## Chapter 2

# Backgrounds and related works

### 2.1 Participatory sensing

In this chapter we review the works concerning the participatory sensing concept, the classification of participatory sensing, different participatory sensing applications and frameworks.

The rest of this chapter is organized as follows: In Chapter 2.1.1, the concept of participatory sensing is formally introduced. In addition, we also present the list of concepts and terms that are similar to participatory sensing. In Chapter 2.1.2, the classification of participatory sensing is provided. Chapter 2.1.3 describes examples of participatory sensing applications. Examples of participatory sensing frameworks are detailed in Chapter 2.1.4. Finally, in Chapter 2.1.5 we present a comparison of selected participatory sensing applications. At the end of this chapter, we formally define the set of key technical requirements that a participatory sensing application should possess.

#### 2.1.1 Introduction

The term "participatory sensing" was first coined in 2006 (Burke et al., 2006). In its definition, participatory sensing leverages modern mobile devices to create interactive and participatory sensor networks that allows people to collect, analyze and share information on certain topics. Specifically, this concept takes advantage of an increasing number of Internet-connected mobile devices incorporating embedded sensors with powerful sensing and computing capabilities; it also exploits

the mobility and intelligence of the device owners. From the definition, it is clear that the device owners (the people) are at the center of the data collecting process, which makes participatory sensing unique compared with other concepts such as WSN (Yick et al., 2008). An example of participatory sensing is the Waze application <sup>1</sup> in which users actively report real-time traffic and road information for the benefit of the whole community.

Apart from participatory sensing, there are several other concepts that represent the process of collecting information through the contributions from a large number of participants. Notably, participatory sensing shares some similarities with "citizen science", "crowdsensing", "volunteered geographic information" and "user-generated content". Each concept involves a variant of this information collecting process. For example, volunteered geographic information focuses on the collection of geographic information such as event location (Goodchild, 2007) while citizen science often includes the participation of non-expert users collecting environmental data across a large area over an extended period of time (Bonney et al., 2009). For the former, real-world implementations include OpenStreetMap <sup>2</sup> (OSM), a worldwide project to create free and editable map data, where users collectively contribute the geographical contents. For the latter, there are several implementations that allow users to document various features of the surrounding environments such as presence and behaviors of birds <sup>3</sup>, light pollution at night <sup>4</sup> and changes in the weather <sup>5</sup>; many of these implementations are indexed at scistarter <sup>6</sup> and CitizenScience.org <sup>7</sup>. Crowdsensing involves the collection of data sensed or generated from mobile devices which is later aggregated and analyzed in the cloud (Guo et al., 2015). While participatory sensing focuses on the active involvement of the users in the process of collecting and sharing data, *opportunistic sensing* represents a group of applications in which the users are not necessarily involved in that process (Campbell et al., 2006). However, the degree of user involvement in this data collecting and sharing process can vary among the applications or even among the users. As a result, mobile crowdsensing

---

<sup>1</sup>[www.waze.com](http://www.waze.com)

<sup>2</sup>OpenStreetMap - <https://www.openstreetmap.org>

<sup>3</sup>Main bird atlas - <https://www.maine.gov/ifw/fish-wildlife/maine-bird-atlas/index.html>

<sup>4</sup><https://www.globeatnight.org/>

<sup>5</sup>ISeeChange - <https://www.iseechange.org/>

<sup>6</sup><https://scistarter.com/>

<sup>7</sup><http://www.citizen-science.org/>

reflects this broad range of user participation in which participatory sensing and opportunistic sensing are at the two opposing sides (Ganti et al., 2011). Other related terminologies describing the crowdsensing process are community sensing (Krause et al., 2008), urban sensing (Campbell et al., 2006) and people-centric sensing (Campbell et al., 2008).

### 2.1.2 Classification of participatory sensing

Based on the main characteristics of participatory sensing, it can be classified into different groups. The main characteristics in question are the involvement of the human user, the scope of the data collection, the source of the collected data and the architecture of the participatory sensing system.

The involvement of the human user significantly affects the operation of the participatory sensing concept; this concept emphasizes on the active participation of the users in the data collecting process (Khan et al., 2013). In contrast, if this active participation of the user is not required, the data collecting process can be branded as opportunistic sensing (Capponi et al., 2017).

Concerning the scope of participatory sensing, it can be classified as personal sensing, social or group sensing and public sensing (Khan et al., 2013). Personal sensing helps the users to collect and monitor data for themselves. An example of this personal scope is from Lu et al. (2009), the authors developed a standalone mobile application that classifies the sound types exposed by the users. Group sensing allows the data to be collected and shared among members of a group. For example, Bao and Roy Choudhury (2010) developed MoVi, a mobile phone-based video highlighting system. In its largest scope, public participatory sensing aims to collect and share data, which usually is environmental data, with everyone for the common good. For example, NoiseTube is a participatory noise mapping tool that transforms several individual noise measurements into a noise map in certain parts of the city.

Concerning the data source, participatory sensing can take data from the embedded sensors in the mobile device (e.g., microphone, GPS receiver, accelerometer) or the mobile social networks (Amintoosi and Kanhere, 2014). It is known that participatory sensing favors the use of data from the device's embedded sensors (Zhang et al., 2016).

Participatory sensing applications differ from each other in terms of their internal

architectures. Due to the fact that the participatory sensing concept relies on the mobile devices to collect data and these devices have different sensing and computing capabilities, there are two approaches for the processing of data (Guo et al., 2013). In the first approach, data are processed in a centralized manner as the mobile devices send all the collected data to a central server for further processing. While in the second approach, the data are processed inside the mobile devices.

### **2.1.3 Participatory sensing applications**

As mentioned in Section 2.1.2, participatory sensing has varying scopes. Among them, the public participatory sensing attracted the most attention as its results and outcomes can potentially benefit a large population. As a consequence, public participatory sensing is also the focus of this thesis. Furthermore, due to its potential to collect different types of data at a high spatial and temporal resolution, the concept of participatory sensing has been realized in a large number of applications, including public sensing applications.

In this section, we are going to review a number of selected participatory sensing applications; they represent the diversity and complexity of participatory sensing. These applications differ from each other in terms of age, theme, domain and characteristics. For a more comprehensive survey of existing participatory sensing application, we refer the readers to (Khan et al., 2013; Lane et al., 2010; Guo et al., 2015).

#### **Noise monitoring**

In this category, we selected Earphone (Rana et al., 2010), NoiseTube (D'Hondt et al., 2013) and SoundOfTheCity (Ruge et al., 2013) as the representatives for a large number of participatory noise measuring applications. These three applications turn the mobile devices into noise measuring devices by collecting the noise measurements produced by the device's microphone and the GPS location of the corresponding measurements and later sending them to the back-end server for further processing and visualization. As a result, a collective noise pollution map can be generated based on the large number of individual submissions made by the participants.

### **Traffic monitoring**

Waze is a well-known mobile application for reporting real-time traffic and road information such as accidents, congestions and potential dangers on the road (Silva et al., 2013b). All contributions from the users will be labeled with pre-defined categories; each contribution typically includes a text description, a photo and the current GPS location of the user.

In the same way, CrowdOut (Aubry et al., 2014) is an application dedicated to the reporting of traffic-related issues in cities. This application allows users to describe an issue using text, pictures, date, time and GPS location in a fixed structure. While having similar purposes and type of collected data, CrowdOut and Waze differ from each other in terms of scope and origin. The former is a prototype created by university researchers and tested in a small scope while the latter is a completed product from the industry and used by millions of users worldwide.

### **Radio signal monitoring**

OpenSignal <sup>8</sup> is one of the most popular cellular signal strength monitoring applications incorporating the participatory sensing concept. It allows any smart-phone users to report various parameters of their received cellular signal (i.e., GSM, 3G, 4G LTE). Each report is attached with the exact GPS location of the signal measurement. Based on the numerous submissions from users, a detailed coverage map of the service providers can be generated for a city, a region or a country.

### **Environmental monitoring**

The participatory sensing concept can be used in environmental monitoring. In this example, a bird watching platform take advantage of people's observation on birds to collect various information that helps bird conservation (Wang et al., 2016). Specifically, users are asked to detail their observation with bird's sound, pictures, text description and the observation's location. These information are then processed in a central server and shared with the public. As a result, ecologists and biologists can benefit from the collected data while the public can raise their awareness on bird conservation.

---

<sup>8</sup><https://opensignal.com/>

BugMap is a research-oriented mobile application that enables users to report the location where a specific type of bug appears (Malek et al., 2018). Each report includes the date and time, geographic location, pictures of the bug, location of the bug (e.g., indoor, outdoor, green urban areas, bushes) and other parameters. Users will subsequently receive a feedback once their reports are analyzed by experts in this field.

MosquitoAlert <sup>9</sup> is a platform realizing the citizen science concept. Anyone with their mobile device can contribute data to the study of mosquito by reporting the presence of mosquito and especially the tiger mosquito and their breeding sites. Each report includes pictures, text description and location of the observation, among other details. Users also receive notifications and feedback after their report is validated by experts. To guarantee the quality of the collected information, users can also verify the submissions from their peers to accurately identify the tiger mosquitos among other types of mosquito.

### **City incidents monitoring**

In cities and urban area, there is a large demand for reporting various issues such as illegal car parking and potholes in the street. As a result, there exists an array of applications for this purpose; they allow citizens to quickly report issues related to their safety and comfort to the city authority. In this section, we analyze SeeClickFix (Mergel, 2012), FixMyStreet (Walravens, 2013), ImproveMy-City (Tsampoulatidis et al., 2013), PulsoDeLaCiudad (Sanchez et al., 2014) and Cityzen Mukherjee et al. (2014) as representatives of this group of applications.

Users can use the aforementioned applications to report any incident they feel important and relevant to the city authority, such as graffiti, fly tipping, broken paving slabs, or street lighting. Usually their reports fall in the pre-defined categories; each report often contains text description, pictures and the exact location of the incident. As a result, the city authority will have a global view on the different types of incidents in the city and their geographical distribution.

### **Other applications**

Biketastic is an application that enriches the cycling experience by allowing users to find new cycling routes and sense cycling-related information such as

---

<sup>9</sup><http://www.mosquitoalert.com/en/>



roughness and noisiness (Reddy et al., 2010c). The mobile application records location data (i.e., longitude and latitude) and speed for rendering the cycling route. The mobile device's accelerometer measures the variation in the z-axis to infer the roughness of the route. In the same way, the microphone measures the excessive noise level, which can be originated from large vehicles or heavy traffic.

BikeNet is a mobile application for mapping the cycling experience (Eisenman et al., 2010). This mobile application connects different sensors such as the GPS sensor, speedometer, inclinometer, magnetometer, pedal speed, lateral tilt, CO<sub>2</sub> sensor and the mobile device's camera and microphone to enable multifaceted sensing. As a result, cyclists can monitor their physical performance and at the same time record environmental data. The information on a route's pollution level, allergy level, noise level and roughness can be shared among the application users.

LiveCompare is an application for collecting grocery prices (Deng and Cox, 2009). By using the mobile device's camera for collecting and decoding the item's barcode and localization technique for pinpointing the stores' location, the application allows users to exchange price information on grocery items.

Wander is a smartphone app designed to sense sociability (Corcoran et al., 2017). This mobile application collects mobility trace from its users and other information presented in the in-app surveys. Subsequently the application combines the collected data with the users' demographical data to infer the users' movement through spaces and their interaction with other physical and social elements in the urban context.

#### **2.1.4 Participatory sensing frameworks**

Recently, the emergence of multi-purpose participatory sensing applications shows a significant change in the realization of the participatory sensing concept. While early applications were designed to address one or limited number of issues and data types, these multi-purpose participatory sensing applications are designed to work in various scenarios. In this section, we will review some representative examples from this group of multi-purpose participatory sensing applications.

EpiCollect <sup>10</sup> is a framework for generating forms that allow end users to collect

---

<sup>10</sup><http://www.epicollect.net/>

different types of data through their mobile devices which are later visualized and analyzed. In its most developed version (EpiCollect+), the framework enables the collection of different text input, multimedia contents (i.e., picture, video clip, sound recording, barcode) and GPS location. The data collection process can be performed without an active network connection. This active network connection is required only for loading forms on the mobile devices and data synchronization.

PRISM is a platform for remote sensing using mobile devices (Das et al., 2010). It allows programmers to define the tasks to be executed by participants in the form of executable binaries and later push these binaries into the participants' devices. Each task is created as a separate executable binary, which includes first level coarse-grained predicate, second level fine-grained predicate, the set up of the application and the processing of the collected data. This implementation offers the flexibility as code modules can be reused and multiple tasks can be performed using the same client in the mobile devices.

Medusa is a high-level language for specifying the different stages and connectors of a participatory sensing process (Ra et al., 2012). It employs three main principles: partitioned services, dumb smartphones and opt-in data transfers. The first principle states that the functionalities are spread in both the cloud server and the mobile devices while the second principle emphasizes on minimizing the amount of tasks to be performed by the mobile devices. Finally the third principle allows users to decide whether or not to transfer data from their device to the cloud server. Medusa uses XML language to describe the different stages such as "UploadVideo", "ExtractSummary" and "Recruit"; developers still have to manually compose the XML object. Comparing with other standalone participatory sensing applications, Medusa significantly reduces the code complexity perceived by the users; for the same tasks Medusa reduces it complexity by two orders of magnitude.

OpenDataKit is a platform for semiprofessional users to collect different types of data on various scenarios (Brunette et al., 2013). The data collected by OpenDataKit include text, pictures, measurements from embedded sensors in the Android mobile devices and external sensors. The requirements for data (the surveys) are scripted through XLSForm, a tool used to simplify the creation of forms. Once the scripted surveys are created, they will be pulled to the mobile devices by the users, who will complete the survey by collecting the data required by these surveys. This framework has been used in various use-cases such as

monitoring rainforests, observing elections and tracking Ebola outbreaks.

MOSDEN is a general data collection platform that allows application developers to define a plugin specifying the communication between the sensor and the MOSDEN server (Jayaraman et al., 2013). For example, the collection of three dimensional acceleration values can be configured through XML code. Furthermore, MOSDEN allows the integration of external sensors into the data collecting process. This platform can be used in different sensing purposes as the complexity of accessing, processing and storing sensory data is hidden from the developers, they can focus more on the application development. The MOSDEN platform also allows the storage of sensory data in the mobile devices while the device is in offline mode.

gPS is a generic participatory sensing system that allows system administrators to configure it into different instances, each serves in a particular scenario (Wu and Luo, 2014). In its generic form, gPS allows end users to submit various types of data such as subjective observations and opinions as well as sensory measurements. In one of its instance (i.e., imReporter), this participatory sensing system is customized into a mobile application that allows end users to report various information related to the presence of mosquitos (e.g., breeding sites, appearance of dengue spots) in Singapore. Each report contains the event category, the photo of the event, a short textual description, the severity of the event, the name of the reporter, the timestamp of the event and finally the event's location. In order to maintain the consistency and quality of the user-submitted reports, this imReporter application enables users to validate each other's report. Specifically, each user can approve or disapprove the report submitted by another user. Overtime, this endorsement relationship leads to a ranked list of user's credibility, in which users with higher ranks are more trusted by the system.

Szabó et al. (2013) developed a framework for collecting a variety of information using the participatory sensing concept. This framework is based on the Extensible Messaging and Presence Protocol (XMPP) and the publish-subscribe service model. The framework stands between producers, who actually sense the surrounding environment, and the consumers, who consume the collected data for different purposes (e.g., software developers, researchers). Apart from the consumers and producers, there exist service providers who directly subscribe to the producers, elaborate the data stream with more information and/or extract certain information from the producers' data stream. The framework has been

customized in different implementations, which were used in three separate use-cases: transit feed service, soccer information and smart campus. For example, the transit feed service allows end users to enrich the public transit data set with real-time information and observations, such as the crowdedness of the vehicles.

Mendez and Labrador (2014) proposed a theoretical framework for developing participatory sensing applications. This framework divides the design of a participatory sensing application into 5 modules: sample size determination, data collection, data verification, data visualization, and density maps generation modules. The first module, sample size determination, is tasked with deciding the adequate sample size so that the results from the participatory sensing process will be valid and comparable to other traditional measurement techniques. The data collection module deals with the users' concern on incentive, privacy and security. The third module, data verification module, verifies the data by removing abnormal measurement values caused by unintentional or intentional invalid measurements. The data visualization module is in charged of displaying the collected data in an understandable manner; interpolation techniques need to be revised to precisely interpolate 2D time series coming from participatory sensing. Finally, the fifth module, density maps generation modules, converts the cleaned data into desired density maps.

DisCoPar is a participatory sensing framework that allows users to visually compose their own sensing campaign by specifying the data types to be collected, the aggregation of the collected data and the feedback to be sent to the participants of their campaigns (Zaman and Meuter, 2015). Using flow-based programming, DisCoPar first offers several reusable citizen observatory components (synchronous processes) and then enables users to connect these processes through data streams. The DisCoPar framework can create different citizen observatories, which differ from each other in terms of use-cases component composition.

Ohmage is a open-source mobile to web participatory sensing platform that allows the collection of self-report data and sensory data (Tangmunarunkit et al., 2015). This platform contains four main components: the Ohmage backend, mobile data collection apps, Web-based data management and administration tools and Web-based data analysis and visualization tools. The platform allows users to define the project using XML scripts; users who are not familiar with programming languages can use a GUI-based project authoring tool to compose these XML scripts. The XML-based project will be sent to the Ohmage backend

and finally the mobile apps for execution.

One of the common denominators among these participatory sensing framework is the ability to function in different scenarios. For each of these frameworks, starting from a generic version, it can be further configured to fit in a particular use-case/scenario or used in various use-cases. As a result, these participatory sensing framework bring several benefits. First of all, the cost of software development will be significantly lowered as there are only one application or several variants of an application to be developed and maintained rather than several different software applications. For the end users, they will deal with a lower number of applications (with similar workflows and user interface) in their devices, which is essential considering their diverse background and ICT literacy. Secondly, these participatory sensing frameworks are able to accommodate a new demand for collecting data, should it arise, by modifying their internal data structure instead of creating another application. Thirdly, for the research community, these frameworks enable researchers and other stakeholders to study user behaviors and preferences when users are simultaneously faced with different sensing tasks and requests. The Citizense framework, whose architecture and implementation will be elaborated later in this thesis, proves to be more complete, combining all important features from other applications/frameworks and introducing some novelties. For example, ordinary users are able to compose sensing campaigns on a specific topic with several parameters (e.g., location restriction, time limit, incentives applied to participants, participant profile), which is not possible in the aforementioned participatory sensing frameworks.

### **2.1.5 Comparison of participatory sensing applications and frameworks**

In Chapter 1 we have pointed out the key research challenges in participatory sensing as *privacy of participants*, *enhancing and guaranteeing the quality of the collected data*, *the reusability of the participatory sensing framework* and *incentivizing the participants*. Each of the aforementioned applications in this chapter has different approaches in solving these challenges. Therefore, to achieve a better understanding of the state of the art in participatory sensing, a comparison among different selected participatory sensing applications is performed (see Table 2.1, 2.2 and 2.3). The applications used in this comparison were selected

from the growing list of participatory sensing examples, and at the same time they represent the diverse and complex nature of participatory sensing. We tried to keep the balance between contrasting factors such as the origin of the applications (industry or academia) and the age of the applications (applications developed several years ago or recently). This comparison was done along certain common features: flexibility of the tool, the presence of incentive, the selection of data collectors, result delivery, identity of data collectors and the flexibility of the sensing process. These features are directly related to the aforementioned research challenges. Specifically, the reusability of the participatory sensing framework can be addressed by increasing the flexibility of the software tool and the sensing process. One of the methods to enhance and guarantee the collected data is to select eligible participants, based on various parameters such as location, time and user profile. To ensure the completeness of the collected data and improve the users' participation, the submitted results should also be delivered in offline mode (i.e., when no Internet connection is available). For the convenience of the readers, useful metadata is listed: name and publication date of the application, measured phenomena, origin (industry or academia), functionality of the application (single-purpose or multi-purpose) and the popularity of the application (a research prototype or a complete product). Our *Citizensense* framework is placed in the last row of the table, with the planned features and functionalities.

From the comparison, it is clear that a majority of the selected applications do not include incentives as a tool to motivate participants. Only a few applications provide some incentive mechanisms in the form of gamification such as ranking and points. Furthermore, most of the applications do not have the ability to store the collected data in the mobile devices in case of a network disruption. Concerning the flexibility of the platform, many applications do not own this characteristic as they are designed to function in one or a limited number of use cases. The multi-purpose participatory sensing applications do offer this characteristic; however the degree of flexibility varies among them. Applications such as *PRISM* (Das et al., 2010) and *Medusa* (Ra et al., 2012) require users with good programming knowledge in order to create the sensing tasks. While this requirement is not present in *OpenDataKit*, this application still requires certain basic programming skills to create the tasks using *XLSForm*. The only application that does not require any programming skills is *Ohmage*; users can intuitively create the sensing tasks using the provided graphical user interface. Regarding the selection of participants,

Table 2.1: Analyzing the selected participatory sensing applications

Name	Time	Theme	Origin	Status	Tool flexibility	Incentive	User recruitment	Result delivery	Identity of participants	Campaign flexibility
Earphone (Rana et al., 2010)	2010	Noise	Academia	Prototype	No	Experience point	No	Immediate	Anonymous	No
NoiseTube (Maisonneuve et al., 2009)	2009	Noise	Academia	Complete	No	No	No	Immediate	Traceable	No
SoundOfTheCity (Ruge et al., 2013)	2013	Noise	Academia	Prototype	No	No	No	Immediate	Anonymous	No
Waze (Silva et al., 2013b)	2010	Traffic	Industry	Popular	No	Reputation point, rank	No	Immediate	Traceable	No
Matador (Carerras et al., 2013)	2013	General	Academia	Prototype	Partial	No	Location, time	Immediate	Unknown	Partial
OpenSignal (Open Signal, 2017)	2010	Signal strength	Industry	Popular	No	No	No	Immediate	Anonymous	No
COBWEB (Higgins et al., 2016)	2015	Environment	Academia	Complete	Partial	No	No	Flexible	Traceable	Partial
Bird Watching platform (Wang et al., 2016)	2016	Bird formation	in-Academia	Prototype	No	No	No	Immediate	Unknown	No
Noisemap (Schweizer et al., 2011)	2012	Noise	Academia	Prototype	No	Ranking	No	Immediate	Unknown	No

Table 2.2: Analyzing the selected participatory sensing applications (continued)

Name	Time	Theme	Origin	Status	Tool flexibility	Incentive	User recruitment	Result delivery	Identity of participants	Campaign flexibility
CrowdOut (Aubry et al., 2014)	2014	Road safety	Academia	Prototype	No	No	No	Immediate	Traceable	No
SeeClickFix (Mergel, 2012)	2008	City incidents	Industry	Popular	No	No	No	Immediate	Traceable	No
FixMyStreet (Walravens, 2013)	2012	City incidents	Industry	Popular	No	No	No	Immediate	Traceable	No
ImproveMyCity (Tsampoulidis et al., 2013)	2013	City incidents	Industry	Complete	No	No	No	Flexible	Unknown	No
Pulso La Ciudad (Sanchez et al., 2014)	2014	City incidents	Academia	Complete	No	No	No	Immediate	Anonymous	No
Cityzen (Mukherjee et al., 2014)	2014	City incidents	Academia	Prototype	No	Yes	No	Immediate	Traceable	No
PRISM (Das et al., 2010)	2010	Multiple	Academia	Prototype	Partial	No	Location	Immediate	Anonymous	Partial
Medusa (Ra et al., 2012)	2012	Multiple	Academia	Prototype	Partial	Yes	Location, time	Immediate	Traceable	Partial



Table 2.3: Analyzing the selected participatory sensing applications (continued)

Name	Time	Theme	Origin	Status	Tool flexibility	Incentive	User recruitment	Result delivery	Identity of participants	Campaign flexibility
Hive (Pavlov, 2013)	2013	Multiple	Academia	Prototype	Partial	No	Location	Immediate	Traceable	Partial
Sensr (Kim et al., 2013)	2013	Multiple	Academia	Prototype	Complete	No	No	Immediate	Anonymous	Partial
OpenDataKit (Brunette et al., 2013)	2013	Multiple	Academia	Complete	Partial	No	No	Immediate	Anonymous	Partial
EasyHarvest (Katsomallos and Lalis, 2014)	2014	Multiple	Academia	Prototype	Partial	No	No	Flexible	Anonymous	Partial
(Szabó et al., 2013)	2014	Multiple	Academia	Prototype	Partial	No	No	Immediate	Unknown	No
gPS (Wu and Luo, 2014)	2014	Multiple	Academia	Prototype	Partial	Trust	No	Immediate	Traceable	No
DisCoPar Zaman and Meuter (2015)	2015	Multiple	Academia	Prototype	Complete	No	Location, time	Immediate	Unknown	Dynamic
Ohmage (Tangmunarunkit et al., 2015)	2015	Multiple	Academia	Complete	Complete	No	No	Immediate	Traceable	Dynamic
EpiCollect <sup>a</sup>	n/a	Multiple	Academia	Prototype	Complete	No	No	Flexible	Anonymous	No
CitizenSense	2017	Multiple	Academia	Prototype	Complete	Intrinsic, extrinsic	Location, time, user profile	Flexible	Traceable	Dynamic

<sup>a</sup><http://www.epicollect.net/>

many applications do not have the ability to filter data collectors; any data collectors can make a submission regardless of their location, time and other parameters. Only a few application can select the data collectors based on their location and time. In general, it is observed that for each desirable feature in the comparison, it is implemented in some applications and left out in other applications. There is not a single application that simultaneously possessing all these desirable features.

As a result of the comparison, we propose the following technical requirements that should be simultaneously implemented by a participatory sensing framework. They stem from the comparison of the existing participatory sensing applications and frameworks discussed above.

- Requirement F1 - Multi-purpose ability: the participatory sensing framework should be able to collect different types of information (sensory and environmental measurements, multimedia content, various type of on-screen input from human participants) so that in one instance of the framework, different participatory sensing topics can be covered. As a result, the prospective participants have a single, all-in-one mobile application that can collect the data required for different participatory sensing campaigns, and campaign authors have a single channel to reach participants. This requirement corresponds to the criterion of tool flexibility.
- Requirement F2 - Coping with network interruptions: the participatory sensing framework should be able to tolerate disruption in the communication network. If the Internet connectivity is not available, the mobile application should store the collected data locally and send later when the communication network is available again. As a result, the mobile application can function as an offline data collection tool. This requirement matches with the comparison criterion of result delivery.
- Requirement F3 - Ability to differentiate participants: The participatory sensing framework should be able to identify participants. As a result, the system can differentiate the participants and keep track of their contribution, thus facilitate other functionalities such as user profiling, user ranking and deployment of incentive mechanisms. This requirement corresponds to the criteria of identity of the participants, user recruitment and incentive.

- Requirement F4 - Ability to flexibly specify content for the sensing campaign: the author of a sensing campaign (government authority, any interested organizations or individuals) should be able to manage the content of a sensing campaign (campaign-specific parameters, used sensors or other data to be gathered from participants). This requirement will make participatory sensing accessible to a broader audience and facilitate the 2-way communication between authorities (e.g., the government) and participants (e.g., citizens). Furthermore, it should be possible to flexibly specify the internal workflow of composed sensing tasks presented to the participants, allowing different paths through the workflow based for example on real-time input (e.g., previous answers) or context parameters. This requirement address the criterion of campaign flexibility.
- Requirement F5 - Context-aware sensing: the participatory sensing framework should be able to define and use context information for a particular sensing campaign, if needed. The context is any information (either implicit or explicit) that can be used to characterize the situation of the participants. The context can be spatial (a particular area(s) of interest), temporal (a specific period of time), user profile (age, gender, level of experience, etc.) or a mix of these. This requirement enables context-sensitive campaigns (only available on-site, and/or during a specific time period; presenting only selected individual sensing tasks), hereby preventing superfluous and irrelevant responses. This requirement corresponds with the criterion of user recruitment.
- Requirement F6 - Incentive for participants: to further motivate participants, incentives can be used to award them for their contribution. Incentive can take different forms, such as reputation points and/or monetary incentive, which represent intrinsic and extrinsic motivation. This requirement corresponds to the criterion of incentive in the comparison.
- Requirement F7 - Security and privacy for collected data: the collected data might contain sensitive information such as personal points of view, observations or opinions, which makes it a potential target for hackers. As a result, the collected data should be protected and securely stored. This requirement is always present in participatory sensing.

At the same time, the participatory sensing framework should comply with the following non-functional requirements:

- Requirement NF1 - Inclusiveness: The framework should be easily understandable and usable, given the diverse backgrounds of participants in an urban environment, possibly those with limited computer knowledge. Specifically, the framework should avoid asking users to work with programming code and/or script. The use of natural text and numeric values is preferred.
- Requirement NF2 - Transparency: The participatory sensing process should be transparent to the users. For example, users are able to view different aspects of the participatory sensing process such as the results of campaigns, the feedback on their submissions and the performance of other users.

## **2.2 The use of incentive in participatory sensing**

### **2.2.1 Classification of incentives**

The participatory sensing process is often hampered by natural factors such as battery life, more occupied memory in the participants' mobile devices and more mobile data consumption. Furthermore, the participants are adversely affected by their privacy concern as this participatory sensing process might reveal their identity, location and mobility patterns; they might be traced by the collected sensor readings (Christin, 2016). As a result, incentives are believed to motivate participants by directly compensating them for their various costs resulting from the participatory sensing process and at the same time maintain the trustworthiness of the results (Pouryazdan et al., 2017). In some special cases, apart from the reward for his/her own submission, a participant can be indirectly awarded based on the performance of other peers; the latter are recruited into the participatory sensing scheme by the recommendation of the former (Amintoosi et al., 2015). This policy is reported to simultaneously improve the welfare of several participants and increase the amount of participation.

Due to their potential, there is an abundance of works on incentives in participatory sensing. At the same time, there are some attempts to categorize these incentives based on their different dimensions (Gao et al., 2015b; Jaimes et al., 2015; Zhang et al., 2016; Restuccia et al., 2016). In Restuccia et al. (2016), the

incentives are classified into application-specific incentives or general-purpose incentives based on their characteristics. Application-specific incentives are specialized mechanisms adapted to a specific participatory sensing application; they are not designed to be generic and thus cannot be applied in other applications. An example for this category is LiveCompare (Deng and Cox, 2009), an application that utilizes the quid-pro-quo interactions among its users; they exchange photos of their common product of interest. In contrast, general purpose incentives, such as those studied in this article, can be applied to a wide range of participatory sensing applications. This category of incentives, which are much more common among participatory sensing applications, are further categorized into non-game theoretical and game theoretical incentives. This division is based on the use of mathematical models that represent the rational decision making process. For example, compete micro-payment as in (Reddy et al., 2010a) represents the non-game theoretical incentive category. Inside the game-theoretical category, there is a distinction between auction-based and non-auction-based incentives. For auction-based incentive, the examples are reverse auction-based RADP-VPC (Lee and Hoh, 2010a) and TSCM (Kantarci and Mouftah, 2014) while an example of non-auction-based incentive is the reward-based collaboration scheme (Duan et al., 2014).

In the work of Jaimes et al. (2015), the incentive mechanisms are categorized based on the type of incentives being used; the two main categories are monetary and non-monetary incentives. In the category of monetary incentives, there are two sub-categories of static monetary incentives and dynamic monetary incentives. Examples are macro-payment (Reddy et al., 2010a) and uniform payment (Musthag et al., 2011) for static monetary incentives and RADP-VPC (Lee and Hoh, 2010a) for dynamic monetary incentives. In the category of non-monetary incentives, the division is again based on the type of incentive being used; the three sub-categories are collective motives, social rewards and intrinsic motives and fun. Examples for these three sub-categories are (Deng and Cox, 2009; Kanjo, 2010; Han et al., 2011), respectively.

Zhang et al. (2016) proposed three incentive categories: entertainment, service and money. With respect to Jaimes et al.'s categorization (Jaimes et al., 2015), the first two categories can be branded as non-monetary incentives, while the latter are monetary incentives. Although the authors did not further categorize the incentives in their money category, these incentive can be approximately grouped into three

groups: posted price, auction-based and Stackelberg game-based, with examples of (Koutsopoulos, 2013; Reddy et al., 2010a; Duan et al., 2014), respectively.

Gao et al. (2015b) offered two different dimensions of incentives. They can be categorized by their employment purposes or by their incentive negotiation process. The employment purpose divides the incentives into user-centric and platform-centric groups; the former focuses on improving the recruitment process and the participants' motivation (e.g., (Lee and Hoh, 2010a)) while the latter focuses on improving the information gain of the platform and optimizing (minimizing) the payments to the participants (e.g., (Jaimes et al., 2012)). The incentive negotiation process divides the incentives into "Price-Decision-First" and "Data-Upload-First" groups; the first group allows participants to decide on their rewards before whether or not performing the actual sensing task (e.g., (Lee and Hoh, 2010b; Koutsopoulos, 2013)) while in the second group participants have to upload the sensing results before any decisions on rewards are made by the platform (e.g., (Faltings et al., 2014)).

### **2.2.2 Deployments of monetary incentive in a real context**

From our literature overview, most of the works we found on incentives in participatory sensing are on the theoretical side. For example, in a very recent survey of incentives (Restuccia et al., 2016) that addressed 43 works on the use of incentive, only 4 works discuss the results of real-world deployments of different incentive types. The rest of the works usually rely on computer simulation to propose an incentive model and/or to prove the effectiveness of such incentive model. In these work, the computer simulation depends on several parameters and probability values which are preset or varied within a preset range. For example, in Feng et al. (2014), the parameters of the computer simulation are the following: number of device — 400 to 1000, number of sensing task — 40, cost of each task — 0 to 50, maximum number of bids per user — 3. In some other works, a mathematical proof is used to validate an incentive mechanism (Koutsopoulos, 2013). However, we argue that these methods do not fully simulate the complexity of the participants' psychological behaviors and other environmental factors that may affect the participants and their interaction with the incentive mechanisms, such as the urgency of the task, the relevance of the task, personal preference or the effect of the "word of mouth". Therefore, in this article, we focus on the

deployments of incentive mechanisms in real life contexts. The main factors that determine the realities of the contexts in which an incentive mechanism is tested are, besides the incentive mechanisms themselves (i.e., type of incentive, the way of awarding, amount of reward), the content the sensing tasks, the diverse background of the participants and the large size of the participant pool.

Through our literature review, we found a couple of deployments of incentive mechanisms targeting human participants (Schweizer et al., 2012; Reddy et al., 2010a; Musthag et al., 2011; Ueyama et al., 2014; Stewart et al., 2009; Rula et al., 2014; Mao et al., 2013; Chon et al., 2013). A closer look into each deployment reveals its characteristics as well as its shortcomings. In Schweizer et al. (2012), although 49 participants are exposed to a realistic task of measuring the noise level, there is only one task to be completed. Furthermore, virtual credits within the prototype were used to motivate the participants instead of real monetary items. Similar as Schweizer et al. (2012), the research by Reddy et al. (2010a) exposed 55 participants to only one task of taking photos of the trash bin, and each of them have access to only one incentive mechanism among the five mechanisms: lump sum, low micro-payment, medium-micro-payment, high micro-payment and compete micro-payment. Although cash was used as the reward for participants, the relationship between the different content of the sensing task and the incentive mechanism was not studied in this work due to the availability of only one sensing task. In Musthag et al. (2011), 36 participants were divided into three groups, where each group has a different incentive mechanism assigned that allocates virtual credit instead (rather than real money). The same task of carrying wearable sensors and answering 20 questionnaires per day was performed by each group. In Chon et al. (2013) the authors divided the participants into two groups, each was faced with a different incentive mechanism. We argue that the participants in these researches did not have knowledge of other incentive mechanisms apart from the one to which they are exposed, therefore it might be not enough to draw conclusion on the effectiveness of a particular incentive mechanism, particularly in competition with others. Furthermore, compared with Reddy et al. (2010a), the number of participants are low and the diversity of incentive are less. In Ueyama et al. (2014), only 18 participants are involved in a research of combining monetary incentive and gamification-based techniques with the sensing task of gathering information at a specific point of interest. The largest number of participants was recorded in (Stewart et al., 2009) where thousands of participants were involved in a research

with extrinsic and intrinsic motivations for crowdsourcing in an enterprise context. Other examples, more specifically in participatory sensing, of works performed in a specific context are Rula et al. (2014), where 96 participants of a conference were given a specific task to perform in order to win rewards in the form of micro-payment and a lottery, and Mao et al. (2013), where participants were subjected to different incentive mechanisms for producing annotations of astronomical objects. In the last three examples, the participants are a non-heterogeneous group (i.e., employees of the company; conference goers; astronomers) asked to do a very specific sensing task in a particular context; they do not represent the diverse nature of people in a society performing a broad range of sensing tasks.

Our review process has shown that the recent deployments in real-life context of incentive mechanisms mentioned above have suffered from one or some of the following shortcomings: the sensing task does not address a realistic topic or is too specific, the context in which the experiment is too specific, a low number of tasks, a low number of participants and the participants' exposure to only one incentive mechanism. Furthermore, only two of the aforementioned related works (i.e., (Musthag et al., 2011; Rula et al., 2014)) substantiate their conclusions with a thorough statistical analysis of the results. Other works (i.e., (Schweizer et al., 2012; Reddy et al., 2010a; Ueyama et al., 2014; Stewart et al., 2009; Rula et al., 2014; Chon et al., 2013)) highlight the effectiveness of a particular incentive mechanism by presenting and discussing the results with numbers, percentage and charts, yet without underlying statistical analysis to strengthen conclusions.

## **2.3 Users' spatial and temporal behaviors in participatory sensing**

### **2.3.1 The use of spatial and temporal coverage in user recruitment for participatory sensing**

Naturally, the use of spatial and/or temporal coverage is crucial for user recruitment in participatory sensing. Based on the users' spatial and/or temporal coverage, participatory sensing organizers can select suitable participants for their data collecting process. In this direction, several works have addressed the issue of selecting participants based on multiple metrics (Reddy et al., 2008, 2009,



2010b); one metric among them is the participants' coverage (i.e., their availability in terms of space and time). In order to obtain this coverage, location points (i.e., longitude and latitude) and time points from the participants in a period of time are collected and studied.

The mobility of the participants is the primary source of information for obtaining their coverage. In Tuncay et al. (2013), the authors proposed a recruitment framework in which the suitability of a participant for a sensing campaign depends on his/her previous mobility profile. In the same way, Hamid et al. (2013) proposed a recruitment framework for urban sensing. This framework uses the vehicles' trajectories to compute the vehicles' coverage and subsequently select the suitable vehicles. It is observed that the aforementioned recruitment frameworks use the mobility pattern for deciding the suitability of the participants. However, we argue that there might be differences between the mobility pattern and the spatial submitting pattern of a participant. Apparently, the submitting pattern, in many cases, is a subset of the mobility pattern. For example, although a participant daily commuting route spans several kilometers and covers a number of points of interest/locations, he might only perform participatory sensing activities in certain locations (e.g., work place, transit stops, public parks). Therefore, it is recommended to study in further details the locations where participants actually contribute their data.

Other participatory sensing recruitment strategies incorporate non-spatio-temporal parameters such as user sociability and deadline-sensitivity. Sociability index is a user's willingness to contribute their information to the sensing process (Fiandrino et al., 2016). This index is built from the amount of data consumed through mobile social network applications or the time spent on these applications or the combination of them. The authors argued that users with higher sociability index are more likely to participate in a sensing task. In the simulations, the authors had shown that the recruitment strategy using this sociability index is more effective than a distance-based recruitment strategy. In Xiao et al. (2016), the authors proposed a recruitment method that finds a set of users that have minimal cost while having the expected completion time of the sensing tasks less than or equal to the given deadline.

While different user recruitment deal with the allocation of sensing tasks before the execution of of these tasks, campaigns orchestration aims to monitor the performance of users while they are performing their sensing activities and check

if their activities conform with the task's parameters defined earlier (D'Hondt et al., 2014). A sensing task is defined as a set of conditions on space, time, userID, data quantity and data density. During the execution of a task, the system checks every submission from users. If a submission does not satisfy one of the conditions, that submission is not accepted and the corresponding user will be notified. Otherwise the submission is accepted and the submission count is updated accordingly. Task organizers can consult the system and be informed about the current progress of the task, recommendations can be suggested to modify the task' conditions so that the initial goal of data quantity and data density can be reached.

### **2.3.2 The users' spatial and temporal behavior in contributing data**

Once participatory sensing data has been collected, the collected results are ready to be analyzed. Next to the collected content, of primarily importance for the organizers, the meta-data (e.g., when and where are users contributing) is also of particular interest, especially for the research community seeking to understand participants' spatial and temporal behavior. Such information might help organizers to better recruit the participants and generally better plan their sensing campaign, and hopefully obtain more and more qualitative results. Until now, the factors that drive the participants to contribute their observed data remain unclear (Massung et al., 2013). Stimulating factors such as incentives (Restuccia et al., 2016), gamification techniques (Arakawa and Matsuda, 2016) and their internal willingness (Jaimes et al., 2015) have been studied, but we found few works (i.e., (Shen et al., 2017; Grasso et al., 2017; Li et al., 2018; Sultan et al., 2017; Alt et al., 2010)) investigating the spatial and temporal behavior of participants. City Probe (Shen et al., 2017) is an application designed to help citizens in reporting certain city issues, whereby the exact location (latitude & longitude) of a report submission was stored. In its deployment, the participants were asked to report on the single issue of occupancy of the arcade in the city using this application. Finally, the map of all reports is compared with the map of the bus stations in the same studied area; the conclusion was that the location of the reports and the bus stations may have spatially negative correlation. We argue that this spatial behavior is applicable to only one specific issue; it is therefore not known if similar spatial behaviors hold when reporting other issues, or generally, collecting other

types of data.

The social media provides a good source of information to researchers, both the content of the information produced by users and the way they contribute data. Grasso et al. (2017) extracted the information on heat waves from social media data (tweets) and examined the relation between the social media activities and the spatio-temporal pattern of the heat waves. Due to the extremely low number of tweets with geo-location meta-data, the location of a tweet had to be inferred from the content of the tweet through entity recognition and natural language processing. As a result, the accuracy of this inference is not high, tweets can only be located at city level in the best case. Silva et al. (2013a) surveyed a large number of different approaches to study user and social behavior using data from social media. However, the data used to analyze the user behavior are not in close relation to participatory sensing. Qian et al. (2017) used geo-tagged Twitter data together with their time component to build the human mobility model in New York City and compare it with the ground-truth network built from official statistics. They concluded that while the results from geo-tagged Twitter data and from official statistics differ from each other quantitatively, these results are largely similar in terms of qualitative patterns.

Li et al. (2018) summarized a large-scale participatory noise measuring and analyzed the spatio-temporal distribution of the received measurements. Their analysis showed that the measurements' temporal distribution matches the human work-rest cycle (measurements were made primarily between 9.00am to 11.00pm) and the measurement were often conducted in residential areas. Unfortunately, the authors did not analyze these spatial and temporal patterns in more details, even though a large number of measurements was available.

In contrast, exact GPS locations are retrieved from the crowdsourced data to form the spatial patterns in bicycle routes (e.g., (Sultan et al., 2017)). However, these GPS traces are automatically created by the mobile device without any human intervention, and therefore the collected data does not represent the users' participation behaviors.

(Alt et al., 2010) described two small participatory sensing experiments in which 18 participants were asked to perform simple tasks such as taking a picture of the closest mailbox or checking the departure time of the next bus. Based on the analysis of the participants' spatial and temporal behaviors, the authors concluded that midday breaks are the participants' preferable time for performing the tasks

and participants tended to complete tasks near their homes. However, we argue that the small number of participants and the simplicity of the tasks significantly limit the validity of their conclusions.

In summary, we argue that although the aforementioned studies discussed the participants' spatial and temporal behaviors, there is room for significant improvement as each study has specific limitations. Furthermore, they all share the characteristic that participants were facing one or a few participatory sensing tasks of similar types and collect one or a few types of data, in a specific participatory sensing scenario. If the participants were given several sensing tasks of different nature, their spatial and temporal behaviors might be different. In this thesis, we will analyze the spatial and temporal behavior of the participants who were given multiple participatory sensing tasks, where each task has a unique content and context. This analysis is performed using the location and time of the submissions and the knowledge on the participants' profile.

## Chapter 3

# The Citizense framework

This chapter is dedicated to the design and implementation of the Citizense framework. In this chapter, we will present the details of each component of the framework, describe how the identified functional and non-functional requirements are realized into the framework and highlight its novelties compared with other existing participatory sensing applications and frameworks.

### 3.1 The design principle

Citizense is designed as a generic user-oriented participatory sensing framework, aiming to make participatory sensing accessible to all users, regardless of their role (i.e., defining data collection campaigns, gathering the required data) and expertise. The central design principle of Citizense is to divide a complex sensing process into several sequential atomic sensing tasks, where each task deals with a single type of data or sensor. In this way, the sensing campaign can be designed through a graphical-user-interface where complex technical details and specifications are transformed into simple and understandable visual elements. At the same time, the whole framework can be reused in different scenarios and collect a variety of data: a task requiring users to input text can be used to either describe the state of the entrance for people with disabilities, the extent of damage caused by a vandalism activity or the location of a mosquito breeding site; a task producing GPS location can be used to pin-point the location of a graffiti, to mark points along a path or to specify the location of illegal dumping in the city.

This principle distinguishes two different roles of the users: campaign authors

and data collectors. The former define the participatory sensing process to collect certain data and the latter provide the former with the required data; we will use these terms consistently throughout the thesis.

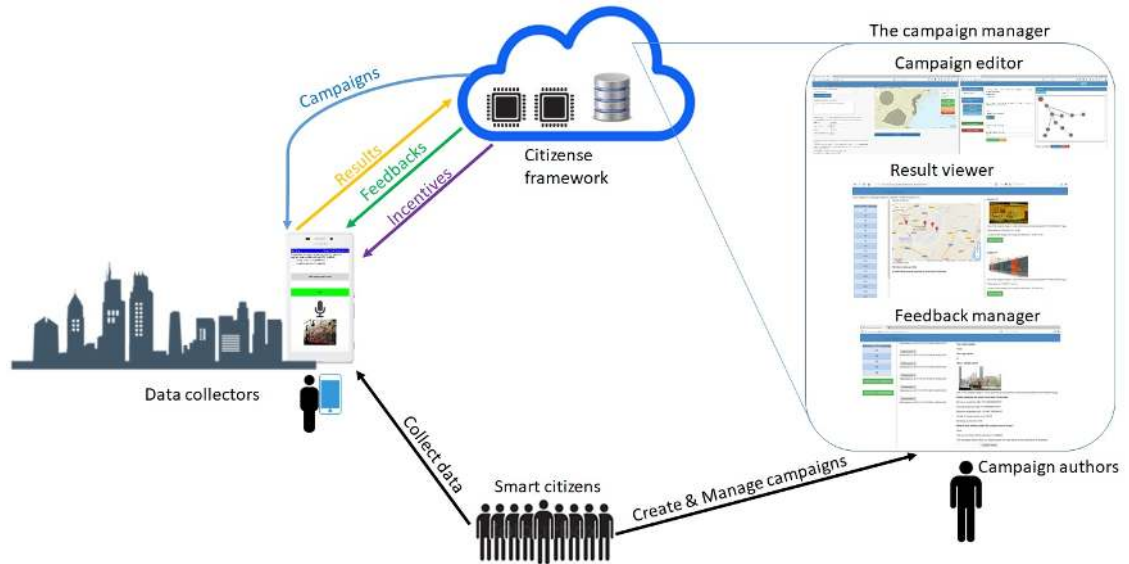


Figure 3.1: The Citizense application ecosystem.

Based on the identified requirements for participatory sensing framework (see Section 2.1.5), this section details the implementation of the Citizense framework, with a special focus on realizing the 7 functional requirements and 2 non-functional requirements mentioned above. The two non-functional requirements span the whole framework, therefore it can be recognized in different components and processes of the framework. For example, the requirement NF1 is translated into the user-centered design principles applied in the campaign manager and the mobile application. NF2 expects the users to receive direct feedback from campaign organizers as well as be able to view the aggregated results from participatory sensing campaigns; this requirement is realized in the result viewer (i.e., in the mobile application and the campaign manager) and the feedback process involving the web-based feedback manager and the push notification in the mobile application. Similarly, several functional requirements are also incorporated in more than one components of the framework (e.g., F1, F3, F5, F7). Other functional requirements are incorporated in a specific framework component. F2 is realized in the internal storage of the mobile application. F4 enables campaign authors to tailor-make their campaigns. F6 is implemented in the application logic

layer of the central server. Further text in this section will detail how the architecture and implementation of Citizense address the aforementioned requirements.

## 3.2 The Citizense architecture

From a technical point of view, the Citizense framework adopts a traditional client-server architecture, with a central platform that contains the back-end logic (i.e., context and incentive management, overall participatory sensing logic, results handling and processing), stores all data (i.e., administrative data, campaigns, participation results), and serves the various Web-based campaign management applications (i.e., Campaign Editor, Feedback Manager, Result Visualizer), and a client-side part, which consists of the Citizense mobile app, which connects to the central server-side platform and allows citizens to participate in campaigns.

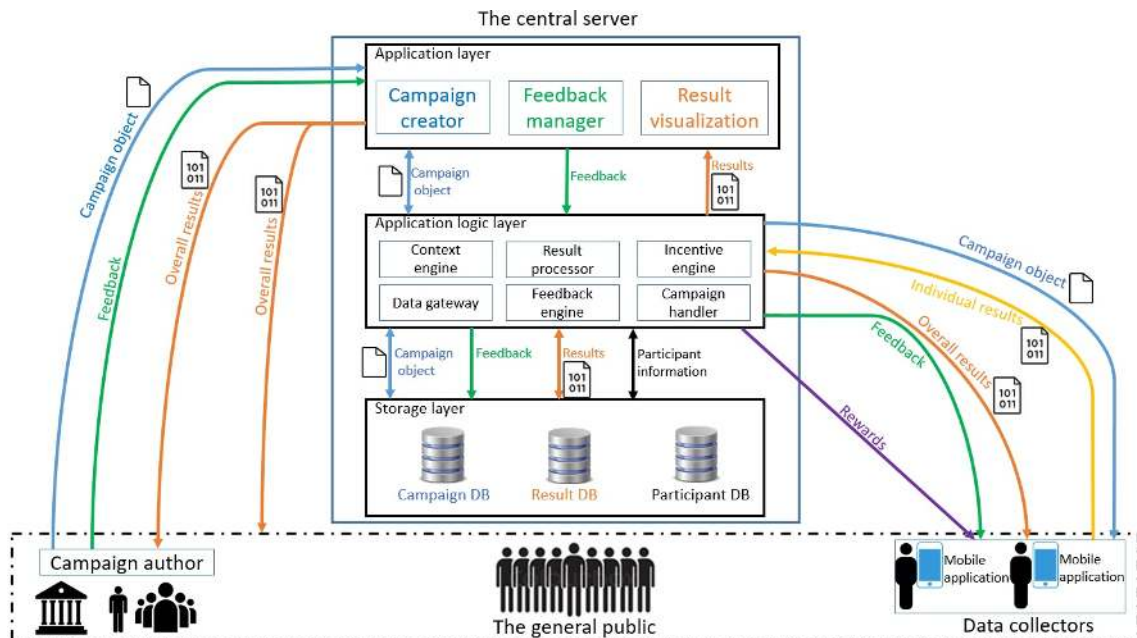


Figure 3.2: The architecture of Citizense framework.

### 3.2.1 The campaign object

The campaign object defines (the configuration of) a participatory sensing campaign, and are thus the blueprint for data communication of campaigns (i.e., campaign objects) and campaign results (i.e., feedback) between server and

clients, and internally between server-side components, and for the databases that respectively contain campaign configurations and participation results.

A campaign object dictates the tasks to be executed by data collectors. The Citizense framework is designed to collect users' subjective opinion and objective reflection of the surrounding environment. The former is realized in the group of textual feedback and the latter is realized in the groups of multimedia content and sensory measurements. The following list details all the types of task supported by the Citizense framework at the time of writing (requirement F1):

- Textual feedback
  1. Type 1 - Question with 1 text answer: This task presents data collectors with a question. They are expected to respond with a textual answer.
  2. Type 2 - Question with multiple text answers: This task presents data collectors with a main question and at least 2 sub-questions. They are expected to respond to each sub-question with a textual answer.
  3. Type 3 - Multiple-choice question with single answer: This task presents data collectors with a multiple-choice question and at least 2 mutually-exclusive options. They are expected to select only one option.
  4. Type 4 - Multiple-choice question with multiple answers: This task presents data collectors with a multiple-choice question and at least 2 options. They are expected to select at least one option.
  5. Type 7 - Question with one unbounded numeric value: This task presents data collectors with a question. They are expected to respond with an unbounded numeric value.
  6. Type 8 - Question with one bounded numeric value: This task presents data collectors with a question. They are expected to respond with a numeric value, bounded by the lower and upper limit defined by the campaign author.
  7. Type 10 - Date input task: This task presents data collectors with a question. They are expected to respond with a date value.
  8. Type 11 - Time input task: This task presents data collectors with a question. They are expected to respond with a time value.
- Multimedia content



1. Type 9 - Photo uploading task: This task asks data collectors to take a picture using the mobile device's camera.

In this group, the tasks of uploading sound recording and uploading video are anticipated.

- Sensory measurement

1. Type 6 - Noise measuring task: This task asks data collectors to perform a noise measurement. There will be two control buttons (i.e., start measuring and stop measuring) for the data collectors to use. As noise pollution is an emerging problem in urban areas, the inclusion of this task type in the framework is logical.
2. Type 12 - WiFi measurement task: This task asks data collectors to scan and measure different parameters (e.g., SSID, BSSID, received signal strength, channel width, frequency) of the surrounding WiFi networks, using one control button. This wireless network technology is ubiquitous in urban areas and large cities. Therefore, the measurement of WiFi network parameters is of paramount importance in the urban context.
3. Type 13 - Task with GPS location: This task asks data collectors to capture their current GPS location (i.e., latitude and longitude).

In this group, the tasks of measuring temperature, humidity, atmospheric pressure, magnetic field and luminosity are anticipated.

The task types in the group of textual feedback cover the basic forms of feedbacks generally produced by participants while the multimedia and sensory measurement tasks involve the sensors available in most smartphones nowadays.

### **3.2.2 The central platform - Storage layer**

The storage layer contains three different databases to store 1/ administrative data, 2/ campaigns (configurations), 3/ participation results. The campaign database stores the campaign objects created by the campaign authors through the graphical campaign manager. Each stored campaign is attached with meta-data such as the identity of the authors, the time of creation and time of last modification. The result database stores the results which passed the various checks

in the result processor, each result entry is attached with meta-data such as the submission's timestamp and submission mode (online or offline). The participant database stores the information given by the participant in registration step (i.e., age, gender, postcode) and the updated value of his experience point and earned rewards,

### **3.2.3 The central platform - Application logic layer**

The Application Logic Layer provides components that implement the main application logic and handles the control flow that defines the participatory sensing process. The main component is the Campaign Handler, which is responsible for 1/ creation and updating of campaigns, and communication with the storage layer for persistence, 2/ serving campaigns, and communication of campaigns, as campaign objects, towards the Citizense mobile app, 3/ internal communication with other Application Logical Layer components for specific logic (e.g., the Context engine, see further on). The Context Engine communicates with the Campaign Handler, and matches campaigns, as campaign objects, with the real-time context of data collectors. This includes matching spatial, temporal and personal information (requirement F5). For example, the Context Engine can select male data collectors currently near the city's central park (i.e., within a specific virtual boundary) who are between 30 and 65 years old who have previously answered at least 10 campaigns and specifically answered campaign "Smoking habit" as eligible data collector of a particular campaign. By doing so, the context engine Context Engine orchestrates campaigns according to the pre-defined space, time and user-profile parameters D'Hondt et al. (2014). The Result Processor is responsible to receive results from the Citizense mobile application, process the results (e.g., decryption - requirement F7), communicate with the storage layer for persistence and for retrieving aggregate results, and finally for communicating the aggregate results back to the Citizense mobile app. The Incentive Engine is responsible to handle incentives: based on the type of incentive, and in communication with the data storage layer, calculate the incentive for each completed campaign, and communicate it to the participant. Finally, the Feedback Engine handles feedback, originating from the campaign author. Depending on the type of feedback, the Feedback Engine internally handles the feedback (e.g., email to the participant), or communicates it to the Citizense client app for handling (e.g., push notification

on the participants mobile phone).

### **3.2.4 The central platform - Application layer**

The application layer provides a suite of applications that rely heavily on the functionality defined in the server-side logic layer, and allow users to realize the different functionalities offered by Citizense in an intuitive, easy-to-use way through a graphical user interface. Three applications are available, which together form the Campaign Manager ecosystem: the Campaign Creator, the Result Viewer, and the Feedback Manager. The Campaign Creator offers a graphical user interface on top of the application logic layer. It provides campaign creators with an intuitive way to create, edit/update and delete campaigns, and is responsible for assembling a campaign object (partial object in case of update/deletion) to communicate to the Campaign Handler. The Feedback Manager offers a graphical user interface on top of the Feedback Engine of the application logic layer, and allows the campaign author to select one of his campaigns, a single or all participants, and formulate the desired feedback. Finally, the Results Visualizer provides a graphical user interface (a viewer) on top of the results database (storage layer). Through these applications, campaign authors (e.g., citizens, researchers, local officials) can effortlessly and intuitively create and manage participatory sensing campaigns without the need for any advanced computer or programming skills.

### **3.2.5 The mobile application**

The Citizense mobile client app is installed on the citizens mobile device, communicates with the Citizense Central Platform, and forms the citizens' entry point for participation. The client application performs a negotiation process with the Central Platform (i.e., the Campaign Handler): it communicates context information, and receives a list of relevant campaigns, in the form of campaign objects. The mobile app are then capable of generating, based on the received list of campaign object, two views to the user: a list-based view, showing campaign only summary information for all campaigns (title, campaign author, promised rewards), and a detailed view for one campaign, which is a rendering of a campaign objects, task by task and taking into account task transitions. This rendering includes the relevant controls for the citizen to provide feedback for each atomic sensory task. Finally, the Citizense mobile app is capable of encrypting results, and communicates these

to the Result Processor of the Central Platform over the public communication network. The Citizense mobile app also includes a communication manager, capable of locally storing results in case of network disruptions, until network connection is resumed. Finally, in communication with the Central Platform (Incentive Engine), the mobile application is capable of rendering the citizen's gained rewards.

## **3.3 The implementation**

### **3.3.1 The campaign object**

Implemented in JSON format, a common and efficient data serialization format in mobile platforms (Sumaray and Makki, 2012) and for data transmission (Gao and Duan, 2011), a campaign object has the following key-value pairs, listed in Table 3.1:

### **3.3.2 The central platform - Storage layer**

The storage layer consists of three databases using MongoDB open-source cross-platform document-oriented database technology: participant database, campaign database and result database.

The participant database stores the profile of participants (both data collectors and campaign author), which are detailed in Table 3.2:

The campaign database stores the campaign; each campaign is stored as an object with the configuration detailed in Section 3.3.1.

The result database stores the submissions from data collectors, grouped by campaign. Each submission has the following configuration, detailed in Table 3.3:

### **3.3.3 The central platform - Application logic layer**

The application logic layer is implemented as 8 RESTful Java-based web services (hosted at <http://www.citizense.uji.es:8880/>), each service corresponds to a functionality of the central platform: receiving new registrations, logging in registered data collectors, serving campaigns, receiving submissions, querying data collectors' account, viewing public results, viewing feedbacks and serving data to external softwares.

Table 3.1: The campaign object.

	<b>Key</b>	<b>Meaning</b>	<b>Value</b>
1	Campaign_ID	The unique title of the campaign	Text
2	Campaign_Description	The textual description	Text
3	onetime	The campaign's execution mode (single/multiple)	Boolean
4	expiry	The campaign's time constraint	Boolean
5	startDate	The campaign's starting date	Date format
6	endDate	The campaign's ending date	Date format
7	geoBoolean	The campaign's location constraint	Boolean
8	PoI_GPS	An array of geographic points (Lat&Long)	JSON Array
9	PoI_WebMercator	An array of geographic points	JSON Array
10	initialTask	The ID of the initial task	Text
11	taskArray	The list of tasks	JSON Array
12	workflow	The list of transitions	JSON Array
13	authorID	The ID of the campaign author	Text
14	cytoscape	The visual structure of the campaign	JSON object
15	incentiveBoolean	The presence of extrinsic incentives	Boolean
16	incentiveList	The list of incentive mechanisms	JSON Array
17	showResult	The permission to show results publicly	Boolean
18	showAuthor	The permission to show the author's ID	Boolean
19	continuousSubmit	The permission to submit the results continuously	Boolean
20	followUp	The option to enable feedback	Boolean
21	userProfile	The option to enable filtering based on user profile	Boolean
22	userProfileValue	The list of filters on user profile	JSON Array

- /registration: This service receives registration requests from new data collectors. If the data collector's email has not been registered, the registration request will be accepted (requirement F3).
- /login: This service receives login request from data collectors who have

Table 3.2: The participant object.

	<b>Key</b>	<b>Meaning</b>	<b>Value</b>
1	firstName	Participant's first name	Text
2	lastname	Participant's last name	Text
3	userID	Participant's nickname in the Citizense platform	Text
4	yearOfBirth	Participant's year of birth	Integer
5	gender	Participant's gender	Boolean
6	timeOfRegister	The timestamp of first registration	Date and time value
7	deviceName	Name of the data collector's device	Text
8	APK level	Android version of the device	Integer
9	secretCode	The random secret code for incentive	Integer
10	expPoint	The experience point earned by the data collector	double
11	earnedMoney	The money earned from the incentives	double
12	earnedPrize	The prize(s) earned from the incentives	JSON Array

Table 3.3: The object result.

	<b>Key</b>	<b>Meaning</b>	<b>Value</b>
1	userID	The ID of the data collector	Text
2	submissionContent	The answers to the tasks of the campaign	JSON Array
3	submissionLat	Location of the submission (latitude)	double
4	submissionLong	Location of the submission (longitude)	double
5	submissionTime	The timestamp of the submission	Time value
6	submissionMode	The network condition when the submission was made	Boolean

registered earlier. If the data collector provides correct credentials (i.e., email and password), he will be logged in the Citizense framework from his mobile device (requirement F3).

- /getCampaign: This service receives download requests from data collectors. Each request includes the current location of the data collector and his

userID. For each campaign, the following parameters from the data collector will be checked against its corresponding parameters from the campaign: timestamp, location and user profile (requirement F5). These checks assist the campaign author to recruit the most suitable data collectors for the sensing campaign. In other words, they perform the campaign orchestration in the space and time domains by assigning campaigns to the exact data collectors. Specifically, if the request's timestamp falls within the campaign's temporal constraint, then the request passes the time check. If the request's location falls within the campaign's location restriction(s), then the request passes the location check, which is realized by the point-in-polygon operation. Using the data collector's userID, the data collector's profile will be retrieved from the participant database and checked against the campaign's profile filter(s). If the request passes all the checks, then the campaign will be served to the data collector.

- /submit: This service processes the submissions from data collectors. First, each encrypted submission is decoded using the AES 128-bit encryption algorithm and the pre-defined private key (requirement F7). Subsequently, the submission is checked against the corresponding campaign's location and time constraints. Specifically, if a campaign has location constraint(s) then its submissions' location will be checked, only submissions falling in the location constraint(s) are accepted. In the same way, the submission will be checked for time validity if its corresponding campaign has a time constraint, only submissions within the time constraint are accepted. Once the submission is accepted, the corresponding data collector has his profile updated with expPoint and rewards (if applicable).
- /viewAccount: This service processes the data collectors' request to view details of their account. Each request contains the data collector's userID, which is used to query all the details related to a data collector. This service returns the updated details of a data collector account to the mobile application.
- /viewResult: Once the request for public results is received, this service scans the campaign database and return all the public campaigns (campaigns whose results are available to the public). When the data collector selects a

campaign to view its results, another request is sent to this service and the data collector is served with all results from this campaign, grouped by tasks of the campaign.

- /viewFeedback: This service receives requests for feedback from data collectors, each request contains the userID, which is used to extract all feedbacks associated with this userID. The data collector will be served with all the feedbacks, sorted by the issuing date and time.
- /gateway: This service receives requests (including campaign IDs and task IDs) from external softwares, processes them and returns with either campaign configuration and/or campaign's anonymized results for further analysis and visualizations of the campaigns and their results.

### 3.3.4 The central platform - Application layer

This layer consists of three web-based graphical user interfaces (i.e., campaign creator, result visualization and feedback manager) for assisting the campaign authors to interact with the Citizense framework, for the purpose of managing the campaigns, viewing the results and issuing feedback/comments to the data collectors (see Figure 3.3).

#### Campaign creator

Campaign creator: This web-based graphical user-interface helps campaign authors to build their campaigns. This interface is built using HTML, CSS and jQuery (for rendering the web pages), ArcGIS API for JavaScript 3.20 (for specifying the location constraints) and Cytoscape complex-network visualization software platform (for visualizing/rendering the campaign structure). A campaign author can create a participatory sensing campaign through any platform (e.g., desktop, mobile device) in two steps by interacting with the web-based campaign editor (requirement NF1). In the first step (see Figure 3.4), the campaign author is expected to specify the general working parameters of the campaign, including visibility of the campaign results, location and time constraint, the textual description and profile of the eligible data collectors. The graphical user interface allows campaign authors to define every parameters of the campaign. In particular, the location constraints can be intuitively and precisely defined by drawing one or several areas



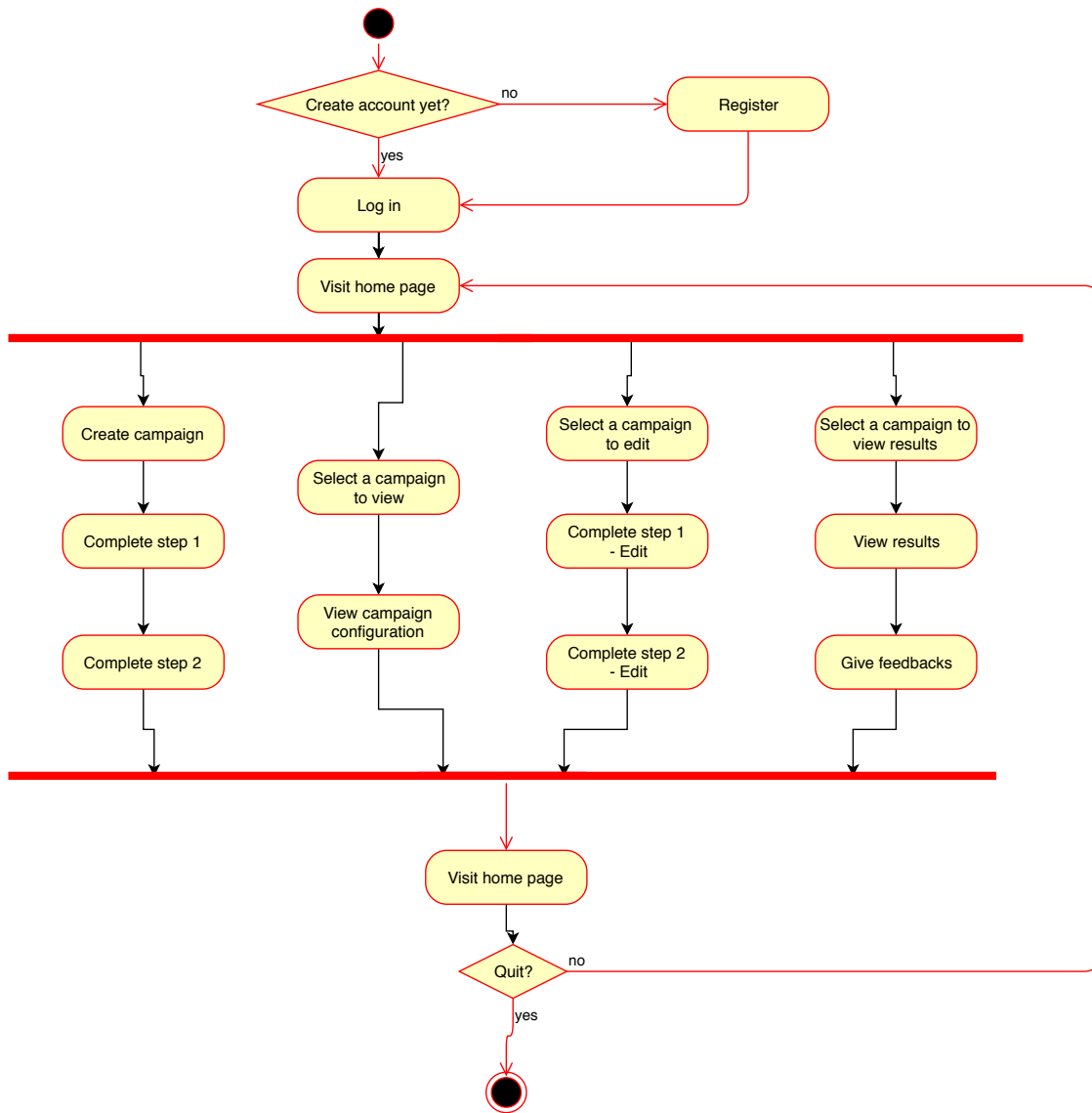


Figure 3.3: Activities performed by the data collectors using the central platform - application layer.

of arbitrary shape on the map to define where the campaign should be available (see Figure 3.4, right).

In the second step, to realize the multi-purpose design principle, campaign authors define the sequence of atomic sensing tasks to be completed by the data collectors (see Figure 3.5). Each task contains a textual description, an optional picture (which can guide data collectors to improve the quality of the

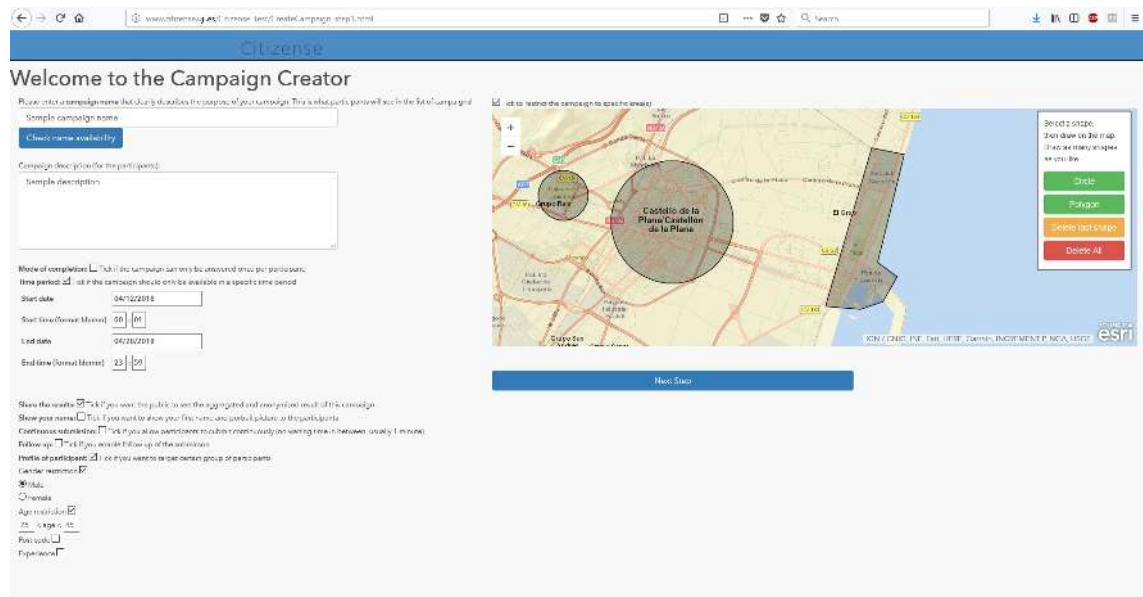


Figure 3.4: The graphical campaign creator. Creating a campaign, step 1: Specifying the various parameters of the campaign.

submission (Budde et al., 2017)) and other necessary parameters, depending on the task type (e.g., the different options in case of a multiple choice question). At the time of writing, the Citizenize framework supports 12 different task types (requirement F1 and F4), grouped into sensory input (noise measurement, GPS location and WiFi fingerprint), multimedia input (picture) and human input (text input, bounded and unbounded numeric input, multiple choice question, date input and time input). To make the campaign more flexible and better fit to the data collector's situation, the Citizenize campaign editor allows multiple branches within the sequence of the atomic sensing tasks through the use of unconditional and conditional transitions. For the unconditional transition, another task follows the current task unconditionally. For the conditional transition, the next task to be displayed depends on the result of the current task. In case of the default transition, the logical condition is "true" while in the other case, the logical condition is a JavaScript logical expression. Technically, the logical expression is composed of task ID, pre-defined primitives used in conjunction with JavaScript string and array methods and JavaScript comparison and logical operators. Table 3.4 gives the definition of the primitives used in the Citizenize framework:

For example, consider task Q26 of Type 4 - MultipleChoice with multiple an-

Table 3.4: Definitions of the primitives used in the logical expression

Primitive Name	Definition
value	The answer of the task if it produces only one answer
value[i]	The answer of a task identified by index i if the task produces several answers
values	The list contains all the answers of a task
count	The number of non-empty answers of a task

swers, it has five options (one of them is "Orange") and two tasks Q29, Q30 succeeding it, the conditional transition from Q26 to Q29 can be {"expression": "(Q26.values.contains('Orange') && Q26.count  $\geq$  2)", "next": "Q29"} and the transition from Q26 to Q30 can be {"expression": "! (Q26.values.contains('Orange') && Q26.count  $\geq$  2)", "next": "Q30"}, which mean that if the data collector selects at least two options in Q26 and one of them is "Orange" then next question is Q29 otherwise next question is Q30. The combination of conditional and unconditional transition can form a loop (see Figure 3.5, right), allowing data collectors to decide how many times they perform a particular task.

Different controls in this graphical user interface assist the campaign author with quick and easy management of the atomic sensing tasks (see Figure 3.5, right). For example, the graph-based visualization highlights the campaign's structure; buttons such as duplicating and deleting a specific task give campaign authors flexibility in defining atomic sensing tasks. In Figure 3.5, a loop between task Q10 and Q11 is formed by combining the two types of transition and red and blue arrows represent conditional and unconditional transitions, respectively. The Citizense campaign editor not only allows the campaign authors to create campaigns, but also to view, update and delete campaigns at any time (requirement F4).

### Result visualization and feedback manager

This web-based application helps campaign authors to view campaign results and issue feedbacks. It is built using HTML, CSS and jQuery (for rendering the web pages), Google Maps API (for rendering location information) and Firebase Cloud Messaging for sending cross-platform notifications and message.

During and after a campaign, the campaign author may at any time provide feedback to campaign participants using the Feedback Manager. This author

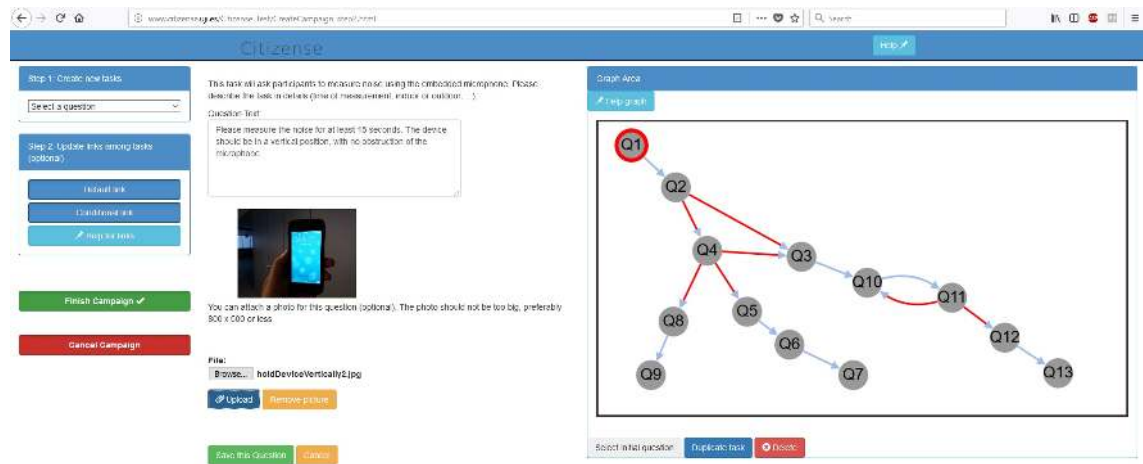


Figure 3.5: The graphical campaign creator. Creating a campaign, step 2: Defining the atomic sensing tasks and their transitions.

has the option to provide individual feedback to each single submission (e.g., useful for example in case of campaigns collecting complaints, suggestions for city improvements, etc.), or aggregate feedback to all participants at once (e.g., informing them about results of a voting campaign, the winner of a competition campaign, etc.). Once feedback is provided, participants are immediately informed by a push notification sent to his mobile device (requirement NF2).

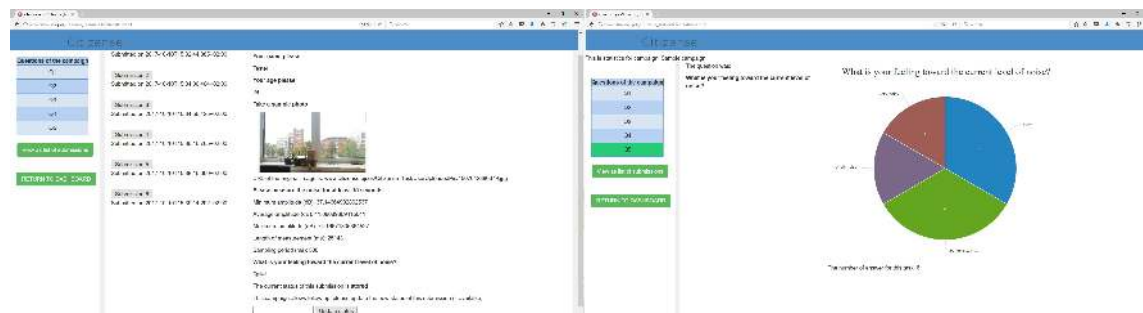


Figure 3.6: The visualization of the results. (left) The grouping by submission, each submission's status can be updated, at the bottom of the picture. (right) The grouping by task.

Once campaign is running, the general public can view live (aggregate) results of campaigns through the web-based Result Viewer, if such visualization of content were allowed by the campaign creator. Furthermore, anonymized detailed results are exclusively available to the campaign creator. These results can be organized

in two dimensions: by atomic sensing task, and by submission. The former (Figure 3.6 right) lists the aggregation and all the individual result per atomic sensing task in a campaign. It thus allows a task-centric overview of all answers, per task. The latter (Figure 3.6 left) lists all the individual submissions of a campaign, whereby the campaign author can select any submissions to see its details (results of each atomic sensing task in the order it was collected). This grouping strategy focuses on the participants, and allows to view all submitted data per participant.

For certain data types, the location where it is collected is an indispensable factor when they are analyzed. For example, the picture of an incident is meaningful when the location of that picture is available. Similarly, the sound measurement has to include location as part of its properties. The result visualization component allows results from tasks of Type 6 (i.e., noise measuring), Type 9 (i.e., uploading photo) and Type 12 (i.e., WiFi measurement) to be displayed on a map. Each response to these task is detailed with timestamp and location information (i.e., latitude and longitude), which is subjected to permission from the corresponding data collectors (see Figure 3.7).

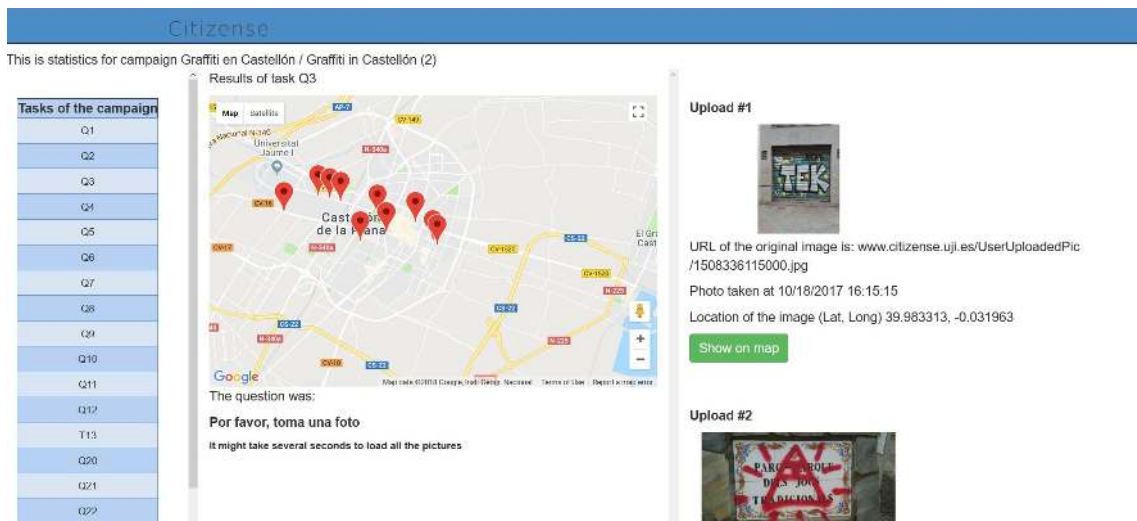


Figure 3.7: Visualizing the results on a map

### 3.3.5 The mobile application

Currently, the Citizense mobile application is implemented as an Android app, available in Google Play store <sup>1</sup>. This Android app needs the following Android permission to operate: INTERNET, ACCESS\_NETWORK\_STATE, ACCESS\_FINE\_LOCATION, ACCESS\_COARSE\_LOCATION, ACCESS\_WIFI\_STATE, CHANGE\_WIFI\_STATE, READ\_EXTERNAL\_STORAGE, CAMERA, WRITE\_EXTERNAL\_STORAGE and RECORD\_AUDIO. As the mobile application will be used by a large amount of users having diverse backgrounds and computer literacy, the main design approach for this component is user-centered design, with special focus on usability. First of all, the mobile application is multi-lingual: the language of the user interface is the same as the language of the device. Several established mobile application design principles were employed, such as decluttering, customizing the keyboard for a specific task type, meaningful error messages, speech-to-text to speed-up the input process and ease the barriers for people with disabilities.

Through this mobile application, the data collectors can perform 4 activities (see Figure 3.8):

- Complete a campaign: This is the main activity of the data collectors. In the background, an HTTP GET request is sent from the mobile application to the central platform's getCampaign service. If this request is sent successfully, meaning that there is an active Internet connection, the mobile application will check its internal storage for any stored results that was not submitted before due to the network interruptions (requirement F2). The central platform replies with the list of relevant campaigns for that data collector in the form of a JSON array, which will be displayed in the mobile application (see Figure 3.9 (1)). Data collectors can see the name of the campaigns, the incentives (if available) and the logo of the authority, if that campaign is created by the authority. The data collector then selects a campaign to open, and he will decide whether to continue with this campaign. If it is the case, he will complete the campaign and produce a submission. While executing the campaign, the mobile application provides a variety of simple control buttons to assist the data collectors. For example, for Type 6 - noise measuring task, the mobile application offers two control buttons (i.e., start measuring noise

---

<sup>1</sup><https://play.google.com/store/apps/details?id=geoc.esr06.citizenselisboa>

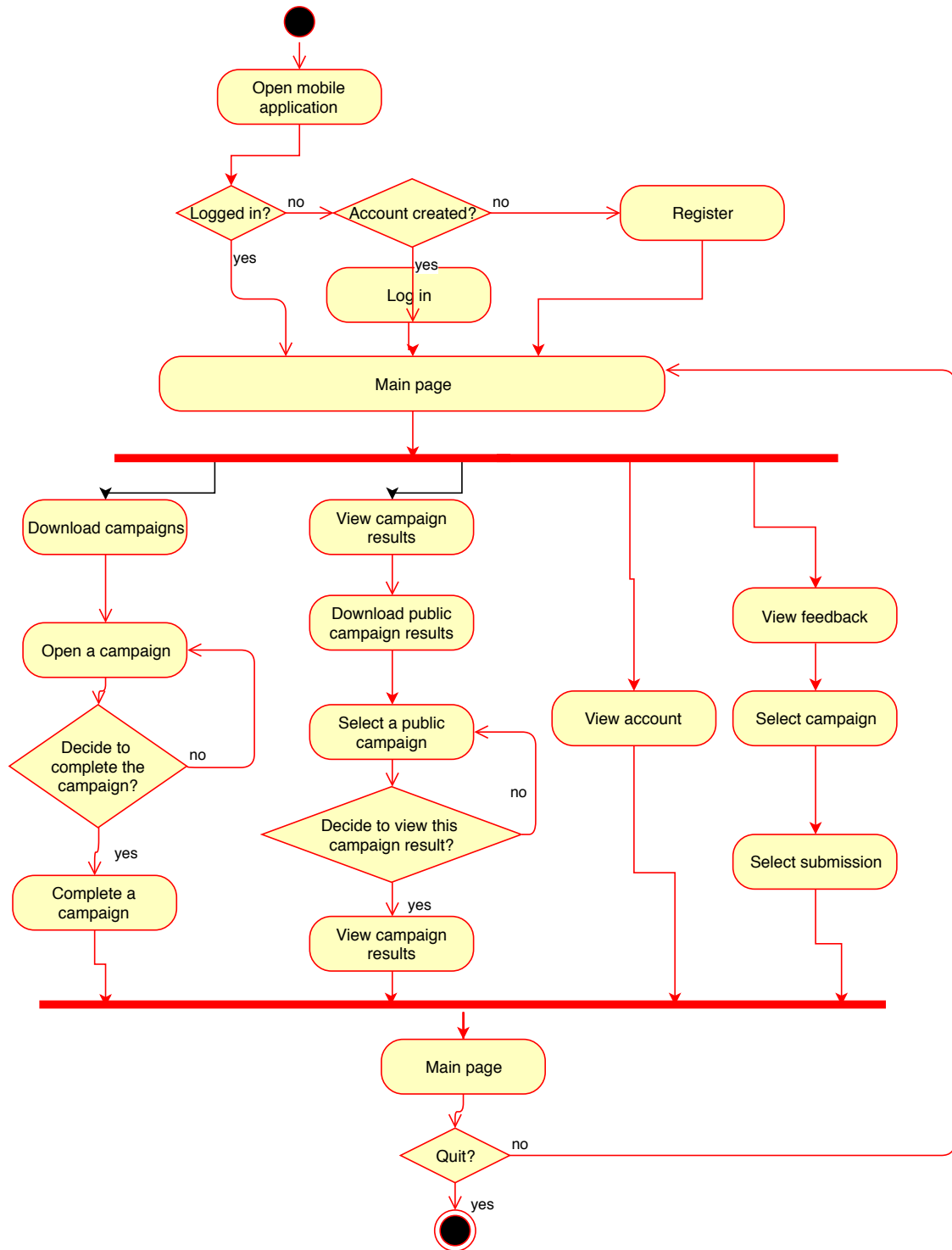


Figure 3.8: Activities performed by the data collectors using the mobile application.

and stop) to the data collectors, so that they can control their noise measuring process. In this type of task, the real-time raw reading from the device's microphone *raw\_reading* is transformed into the noise value *noise* (in dB) according to the following formula:  $noise = 20 * \log_{10} raw\_reading$ . The real-time raw reading is taken every 100 milliseconds. Together with the real-time noise value, other parameters (i.e., minimum noise value, maximum noise value, number of measurements) are recorded. Based on these parameters, the average noise value throughout the measurement is computed and included in the measurement results. The mobile application allows data collectors to measure noise for an extended period of time. For Type 8 - Question with one bounded numeric value, a slider will be displayed to help data collectors select the value within the pre-defined numeric range (see Figure 3.9 (2)). For Type 12 - WiFi sensor, the mobile application scans the available WiFi networks for the following parameters: SSID, BSSID, received signal strength, channel width, frequency. For tasks asking data collectors to input text (e.g., Type 1, Type 2), a speech-to-text button can be used to input text by voice rather than typing using the keyboard. This feature can speed up the input process, avoid typos and encourage people with disabilities to use the Citizense mobile application (requirement NF1). Finally, the AES 128-bit encryption algorithm is performed, using the pre-defined private key, to encrypt the submission before sending over the communication network (requirement F7). In case the network connection is not available, the mobile application stores the submission (in the form of a JSON object) in the memory of the mobile device (requirement F2). When the next HTTP GET request for campaign is sent successfully, this part of the memory is checked for the stored submission(s); they will be sent to the central platform.

- View campaign results: To view results of a campaign, an HTTP GET request is sent from the mobile application to the central platform. Immediately, the mobile application receives a list of public campaigns (i.e., campaigns that have their results available to the public), which will be displayed to the data collectors. A data collector then selects a campaign; in the background another HTTP GET request with the selected campaign's ID as a parameter is sent to the central platform. This request is processed and replied with the results of the selected campaign, extracted from the central platform's storage layer and grouped by campaign's tasks.



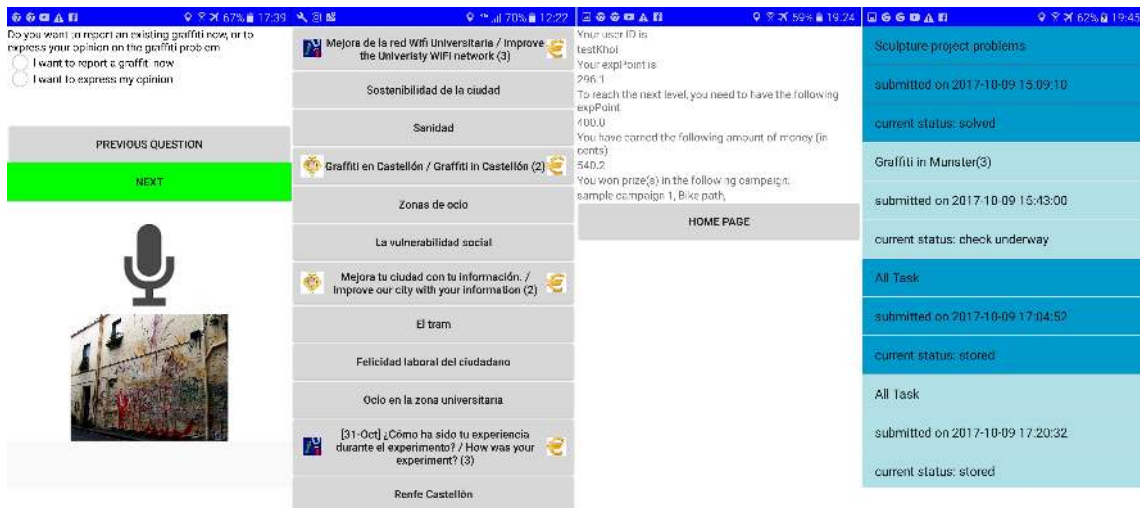


Figure 3.9: The client mobile application. From left to right (1) An atomic sensing task. (2) The list of available campaigns. (3) The details of a participant's account. (4) The status/feedback of the submissions.

- **View account:** Data collectors can view their account to see the details such as their earned incentives, level of experience points (expPoint) and number of submissions made (see Figure 3.9 (3)). Technically, an HTTP GET request, with the data collector's userID as a parameter, is sent from the mobile application to the /viewAccount service of the central platform's application logic layer. The participant database and the result database in the central platform's storage layer are queried to extract the necessary information on the data collector in question. Finally, the response from the central platform will be displayed on the mobile application, showing the requested information (i.e., userID, earned incentives, level of experience point and number of submissions made).
- **View submission feedback:** At any time, data collectors can view feedback on their submissions. When a data collector clicks on the button "View feedback", an HTTP GET request is sent from the mobile application to the /viewFeedback service of the central platform. Based on the userID of the data collector, this service will response with a list of submission feedback, associated with the campaign name and the submission timestamp, sorted by descending time (see Figure 3.9 (4)).

### 3.3.6 Comparison of Citizense and other participatory sensing applications

With this current implementation, Citizense are able to collect a variety of data types in different scenarios. Therefore, it can be considered a multi-purpose participatory sensing application. Specifically, the data types covered by Citizense span that of almost all applications described earlier. As a result, it can replace the single-purpose participatory sensing applications mentioned earlier in Chapter 2. For example, Citizense can be used to collect information on a traffic-related issue, which is the design goal of CrowdOut (Aubry et al., 2014). Data collectors can submit the location, time, text description and pictures of a traffic incident using tasks of Type 13, Type 11, Type 2 and Type 9, respectively.

Similarly, data collectors can use Citizense mobile application to measure the WiFi signal parameters at any location and later report the collected readings. Tasks of Type 12 and Type 13 will be involved in this scenario. With the future extension of Citizense that includes the measurement of cellular signal, Citizense will be able to completely replace OpenSignal in reporting cellular signal strength.

In the category of environmental monitoring, Citizense is also able to take over the job of MosquitoAlert and other similar applications. It allows users to report an appearance of tiger mosquitos by answering to a campaign that asking for the description of the mosquito, the picture of the mosquito in question, the location and time of the appearance. This campaign includes tasks of Type 2, Type 9, Type 13, Type 10 and Type 11, among other task types.

With a small extension of the Citizense framework that incorporates the task of uploading sound recordings, this framework can function in the field of bird watching as the platform developed by Wang et al. (2016) did.

Comparing with the aforementioned participatory sensing frameworks, Citizense offers at least similar functionality. Citizense retains the most important characteristics of these framework such as the possibility of designing a campaign, the graphical user interface, the variety of data types supported by each framework, security for the users and the ability to customize the framework. Specifically, Citizense enables users to create a sensing campaign through its graphical web-based Campaign Editor. Once a campaign is created, it can also be edited by the Campaign Editor. This graphical user interface allows user to interact with the

framework without any need for programming skills; users are required to provide only basic items such as text, numeric values, picture, on-screen selections and other simple interactions. At the time of writing, Citizense supports 13 different types of task, which cover all important data types such as text, (bounded and unbounded) numeric values, pictures, sound measurement, WiFi signal parameters and GPS location. The future extension of the framework which includes a more diverse set of task types is anticipated.

Table 3.5 summarizes the comparison between Citizense framework and the applications and framework mentioned in Table 2.1, 2.2 and 2.3.

The topic of user security is partially covered by Citizense framework. In the current implementation of Citizense, it offers each user a private user account. The delivery of the campaign results, which might contain sensitive and private information, is encrypted to ensure that unauthorized third-parties cannot have access to these information. When the results are visualized and presented to the general public, any information related to the data collectors is removed to preserve their privacy.

With regard to the existing frameworks, Citizense offer some novelties such as the conditional links between tasks, the speech-to-text feature and the communication from the campaign authors to the data collectors. Specifically, the existence of the conditional links between the tasks provides a more flexible workflow, tasks can fit better to the specific situation/condition of the data collectors. By combining conditional and unconditional links, a loop can be created among the tasks. This loop allows data collectors to dynamically adjust the size of their reply. For example, a task of Type 9 can be performed several times as the data collector wishes, to better describe a particular phenomenon/object. A fixed number of pictures given in advance by the campaign author cannot fit in the particular situation of all data collectors. Furthermore, campaign authors can create these conditional links by selecting the appropriate option in the graphical Campaign Editor.

The speech-to-text feature assists data collectors in the process of inputting long text and therefore improve their user experience. As the mobile device's screen is not adequately large in most cases, typing text manually can potentially lead to typos, which have several negative effects in the long term. Furthermore, this feature makes the framework more user-friendly towards people with disabilities as they can freely speak into the device's microphone without the need of typing. The communication from the campaign author to the data collectors serves several

Table 3.5: How can Citizense substitute other existing participatory sensing applications and frameworks

Name	Not supported	Partially supported	Fully supported	Note
Earphone (Rana et al., 2010)			X	
NoiseTube (Maisonneuve et al., 2009)			X	
SoundOfTheCity (Ruge et al., 2013)			X	
Waze (Silva et al., 2013b)		X		Computation of Estimated Time of Arrival (ETA)
Matador (Carreras et al., 2013)			X	
OpenSignal (Open Signal, 2017)		X		The task of capturing cellular signal is missing
COBWEB (Higgins et al., 2016)			X	
Bird watching platform (Wang et al., 2016)			X	
Noisemap (Schweizer et al., 2011)			X	
CrowdOut (Aubry et al., 2014)			X	
SeeClickFix (Mergel, 2012)			X	
FixMyStreet (Malravens, 2013)			X	
ImproveMyCity (Tsampoulaidis et al., 2013)			X	
PulsoDeLaCiudad (Sanchez et al., 2014)			X	
Cityzen (Mukherjee et al., 2014)			X	
PRISM (Das et al., 2010)			X	
Medusa (Ra et al., 2012)			X	
Hive (Pavlov, 2013)			X	
Sensr (Kim et al., 2013)			X	
OpenDataKit (Brunette et al., 2013)		X		External sensors are not supported
EasyHarvest (Katsomallos and Lalis, 2014)			X	
(Szabó et al., 2013)			X	
gPS (Wu and Luo, 2014)			X	
DisCopar Zaman and Meuter (2015)		X		Submission resolution is not supported
Ohmage (Tangmunarunkit et al., 2015)			X	
EpiCollect <sup>a</sup>			X	

<sup>a</sup><http://www.epicollect.net/>

purposes. First, it serves to engage the data collectors as data collectors can be contacted directly by campaign authors for the results of their previous sensing activities. Second, this communication channel can be used by the central server to send instant feedback to the data collectors in case they need some corrections (e.g., submitting too fast, submitting blank answers). In general, this communication channel helps to improve the data collectors' experience when doing participatory sensing and potentially improve the data quality.

## **3.4 Technical testing of the framework**

To test the framework, two types of tests were performed: technical tests, which verify technical features of the framework (particularly, accuracy and performance), and non-technical tests, which verify qualitative properties of the framework, particularly the feasibility of using the framework in a real-world setting. The latter is addressed in Chapter 4; the former we address in this subsection. We hereby note that *Citizensense* was developed as a prototype aimed to be used as a research vehicle to study user behavior (spatial, temporal, under monetary incentives); not as an end product capable of being released in the market place. As such, our focus was not on full fledged technical testing, and what we present here are only preliminary indications of accuracy and performance, rather than definite technical tests for market conditions. We elaborate on this in Chapter 7 (Outlook).

### **3.4.1 Accuracy of measurement**

The framework currently supports three sensory measurements: ambient noise measurement, WiFi signal strength measurement and capturing GPS location. Due to the fact that the measurement is done on the data collectors' device, the measurement's accuracy depends on the hardware of the device, and this accuracy has to be verified. We aim here to show that our approach at least has the potential to provide accurate results. Our method of verifying the measurement's accuracy is the following: we replicate the measurements done with professional equipments in the literature using our standard Android device (Samsung Galaxy A7, with Android version of 6.0.1).

In our noise measuring tests, we verified the accuracy of our algorithm that calculates the noise by testing in two different conditions. The first test was the

measurement of cabin noise level during the cruising phase of Airbus A321, a popular commercial aircraft. Our repeated measurements (10 measurements) read an average noise of 84 dB(A), which are in line with the range 80-85 dB(A) produced by professional equipments in the same aircraft type during the cruising phase (Ozcan and Nemlioglu, 2006). The second test was on the noise level in a quiet room. Our repeated measurements (10 measurements) resulted in an average in the range of 35-40 dB(A), which conforms to the established ambient noise reference values (sengpielaudio, 2018). These tests provide an initial indication that the noise measurements of the Citizense client application can produce data that are reliable enough for further discussion by different involved parties.

The accuracy of WiFi signal strength measurement was also verified. We performed a measurement replicating the settings of a measurement in Kaemarungsi (2006), using our standard Android device. The device was put one meter away in line of sight from the WiFi access point. Our repeated measurement (10 measurements) read an average of -32 dBm, which is in range of the results (-40 to -30 dBm) from the measurements with identical settings performed in Kaemarungsi (2006). We elaborate on accuracy of measurements in Chapter 7 (Outlook).

### **3.4.2 Scalability of the framework**

At the time of writing, the central server of Citizense is deployed on a standard infrastructure with Intel Xeon E5-2690 v2 3.0GHz CPU and 4GB of memory. We performed several stress tests to examine the scalability of the framework under different load conditions, in order to determine a sufficient performance for the case studies as described in Chapter 4 (Deployments). Among the monitored parameters, the processing time for each request is of importance, as serving a request is a time-critical operation and the processing time greatly affects the waiting time on the data collector's side (the latency). It is worth noting that this latency depends on the processing time in the server and the propagation time in the communication network; the latter is independent from the Citizense framework.

The processing time directly depends on the following factors: number of campaigns, the size of each campaign and the number of concurrent requests. For the first and the second factors, we filled the campaign database with 100 identical

campaigns, each contains 27 atomic sensing tasks which cover all the different task types supported by the Citizense framework and are connected by both conditional and unconditional transitions. After sending 1000 repeated single requests to the central server, on average the central server took 787.613 milliseconds to process each request, with the maximum of 1144 milliseconds, minimum of 310 milliseconds and standard deviation of 122.39. For the third factor, the Apache JMeter tool was used to launch a large number of concurrent requests to the central server simultaneously. We launched 1000 simultaneous requests to the central server containing 11 campaigns. The result was satisfying: the server was able to handle all the requests and 95% of the request was processed in less than 500 milliseconds. We performed another stress test in a more realistic condition: using JMeter, 100 simultaneous requests was sent to the central server containing 24 different campaigns, most of the campaigns have more than 15 tasks of different types. The result of this test was satisfying as well: the central server replied instantly to all the requests. Specifically, the processing time for all request is less than 500 milliseconds. It was therefore concluded that the Citizense's central server can independently withstand a load, should it arise, which sufficient for our planned deployments. Finally, it should be noted that the implementation was performed in such a way that it allows easy redeployment and upscaling in a cloud environment, as to prepare for real-world market conditions. We elaborate on scalability in Chapter 7 (Outlook).





## **Chapter 4**

# **Deploying Citizense in real-world scenarios**

This chapter details the two deployments of the Citizense framework in real-world conditions, in which the framework's functionalities and features are tested with human participants and results are obtained for further analysis and discussions in subsequent chapters.

### **4.1 The Castellón marathon**

#### **4.1.1 The goals of the deployment**

As approved by the city hall of Castellón and required by the organizers of the city marathon, Citizense was deployed in this event for the following two purposes. Firstly, this deployment aimed to collect information from the marathon participants (e.g., feedback on the marathon, demographical information). Secondly, we sought to test the usability of the Citizense mobile application as users were directly asked to evaluate different aspects of this component.

#### **4.1.2 The setup and progress of the deployment in the marathon**

An early version of the Citizense framework was deployed in the Castellón marathon (February 2017). The city marathon is an annual event; it attracts a

large number of participants and thousands of spectators. After receiving the permission from the city authority, the authors worked with the organizer of the marathon to customize the Citizense framework for this event.

The organizers of the marathon needed a multi-lingual mobile application that collects textual feedbacks from marathon athletes on various aspects of the marathon. Traditionally, they collected these information through paper-based survey. In their opinion, this paper-based process is inefficient, slow and it results in the waste of resource (e.g., materials and money). Their requirements matched with the capabilities of the Citizense framework; the framework is capable of collecting various types of data, including textual feedbacks.

A campaign, available in Spanish, Catalan and English, consisting of 28 questions was created through the Citizense campaign creator. This campaign have its questions linked by both conditional and unconditional transitions. A tailor-made version of the Citizense mobile application was made and hosted in Google Play store. Using the marathon as an opportunity to reach real users, we created another campaign to seek feedback on the mobile application from the users. This second campaign has two parts, the first part include the 10 standard questions from the System Usability Scale (SUS), in order to assess the usability of the Citizense mobile app (see Table A.1). The second parts includes questions for the users on the concept of participatory sensing. The details of this part can be found in Table A.2.

### **4.1.3 The results**

After a short promotion effort by the marathon organizers, the framework quickly received 180 registrations from the participants; these participants installed the customized Citizense mobile application on their mobile devices. Among them, 140 participants submitted their feedback on the marathon; the submission was of high quality, with no sensing task left blank. Moreover, there were 10 submissions made in offline mode (the data collectors didn't have Internet connection at the time of submission); these submissions were marked by the mobile application. The marathon's organizers praised the Citizense framework, as it allowed them to efficiently and quickly collect the runners' feedbacks, reduce costs compared to previous paper surveys, for which they needed staff, both during and after the event. The organizers furthermore indicated that insights gained through the

Results Visualization were to be used taken into account for the preparation of next event year's marathon.

Regarding the second campaign on the usability of the Citizense mobile application, we received submissions from 52 participants. On the adjective scale of Bangor et al. (2009), the result was average (i.e., "OK", on a scale of worst imaginable, poor, ok, good, excellent, best imaginable). Given the fact that many respondents just entered the "neutral" score for all questions, indicating they didn't answer the questions seriously (we cannot blame them, after just having run a marathon), and it was a first deployment of the Citizense system, we consider this result quite good. Furthermore, results were quite encouraging with respect to negative sentiment: only 11,6% answered negatively to the statement "I thought the system was easy to use" (Q3), 13,5% of the respondents thought that they needed help when using the Citizense mobile app (Q4), and I felt very confident using the system.), only 7,7% of the respondents said that they didn't feel confident when using the Citizense client application. The full list of questions and details of answers can be found in Table A.1.

#### **4.1.4 Discussions**

Although this deployment is relatively simple and small in terms of scope, it suggests two important observations: the approval of the local authority and the appreciation of the marathon organizer after the event.

Before the marathon event, we had a chance to present in details the idea of a multi-purpose participatory sensing framework and the development of the Citizense framework to the Castellón city hall. Due to the flexibility and the broad applicability of the Citizense framework, the city hall agreed that the framework will be deployed and tested in an official city event. The Castellón marathon was chosen as it is a mid-size event and its timing coincides with the completion of the first version of Citizense. Furthermore, the Citizense framework received the approval of the marathon organizers.

During and after the marathon, the results from the survey launched through the Citizense framework was immediately available to the marathon organizers. They were able to view the answers for all the questions in the survey through the web-based Citizense result visualization. Depending on the type of the question, the results were aggregated (e.g., multiple choice question) and/or displayed

individually (e.g., text or numeric question). For example, the question on the age of the participants has its results grouped and displayed in a pie chart while another question on the approximate amount of money spent during the event has its results displayed individually. The latter also has its results aggregated: the result visualization shows the maximum, minimum and average spending of all participants. The marathon organizers were satisfied with the various results collected by the Citizense framework. They confirmed that by using the Citizense framework, they were able to create and edit their survey in real-time. Compared with using the traditional paper-based, the Citizense framework allowed them to significantly reduce the cost of man power and material.

For the authors, this first deployment of the Citizense framework brought several benefits. First of all, the (early version of) framework was tested in real-world conditions (e.g., human participants, the diversity of the participants' devices and Android versions, the different stakeholders' requirements) and several functionalities of the framework were successfully confirmed. Secondly, the approval of the city hall and the appreciation of the event organizers suggested that the design and development of this framework can potentially serve a large group of different stakeholders. Finally, the participants' answers in the second campaign suggested us that incentive is one of the effective tools to incentivize their participation. Based on this observation, we conducted the deployment in the UJI campus, with one of the mail goals is to study the user behavior in the presence of monetary incentives.

## **4.2 The deployment in UJI campus**

### **4.2.1 The goals of the deployment**

The purpose of deploying the Citizense framework is three-fold. Firstly, the deployment serves to explore the framework's effectiveness in collecting both geographic and general information. We discuss this in the current chapter. Secondly, we seek to investigate the effectiveness of monetary incentives in general and identify the incentive mechanism(s) that is/are most successful in enhancing the participation of data collectors. Lastly, the purpose is to analyze the participants' spatial and temporal behavior based on the meta-data of data submission in this deployment.

### 4.2.2 The setup and progress of the deployment

The Citizense framework was presented to the students, professors, staffs and passers-by of the Universitat Jaume I's campus (Castellón, Spain). At the beginning of this 20-day deployment, a promotional activity was performed to recruit participants: emails concerning the deployment were sent to all members of the university (i.e., students, professors and administrative staffs), flyers were handed out and posters were visible in the most-visited places in the university campus (e.g., canteens, entrances, library, bus stops). Through this promotional activity, participants were encouraged to register in the Citizense framework and participate as campaign authors (to create campaigns for other participants to collect city-related information) and/or data collectors (to respond to the campaigns available in the framework). It is known that in general the university staff, professors and especially the students live across the province, not only in the city of Castellón.

On the first day of the deployment, 4 participatory sensing campaigns from the organizers were launched; these campaigns addressed relevant and common issues that can have impact on the lives of the local residents and the participants such as illegal graffiti and vandalism, cycling infrastructure in the city and improvement in the public furniture. While data collectors were responding to the campaigns from the organizers, campaign authors were encouraged to create campaigns for the data collectors, these campaigns are expected to gather meaningful information related to the city, which can be used by the city authority to improve different aspects of the city. During the course of the deployment, another 6 campaigns with similar theme were launched regularly by the organizers, totaling the number of campaigns from the organizers at 10 campaigns. At the same time, participants were encouraged to create their own campaigns.

Regarding the incentive, we use the following monetary incentive mechanisms in the Citizense deployment, which are supported by the Citizense framework (requirement F6) and they prioritize fairness among the data collectors: the payment depends on criteria that are theoretically fair to all participants.

- Mechanism M1 — Lottery: The participants who submitted at least one valid answer to the campaign have the chance to win the reward(s) defined in advance by the campaign author. The central platform randomly selects

the winning submission(s) and assigns the prize to the corresponding participant(s). This mechanism allows the campaign author to exactly determine the budget to be used in the campaign. The participants have equal chances to win the gift(s), however in the end only one or some lucky data collectors will actually receive a payout (i.e., the gift(s)). In this deployment, the data collectors are informed about the value of the monetary gift (20 Euro) before opening a campaign applied with this incentive mechanism.

- Mechanism M2 — Fixed micro-payment: A data collector is paid a fixed amount of money at the time he/she submits a valid answer of a campaign. This fixed amount of money is specified by the campaign author. In principle, the campaign author does not know the overall budget beforehand; yet this mechanism offers fair payments for the data collectors, i.e., they all get the same payment for each valid submission. To avoid an out of control overall cost, campaign authors may slightly restrict the fixed micro-payment scheme. In this deployment, before opening a campaign applied with this incentive mechanism, the data collectors are informed about the amount of the fixed micro-payment per valid submission which is paid upon submission (i.e., 25 Euro cents).
- Mechanism M3 — Variable micro-payment. The campaign author specifies the overall budget for the campaign; this budget is later divided by the number of valid submissions of the campaign, and each data collector gets an equal reward. While this mechanism allows the campaign author to know in advance the amount to spend, the data collectors do not know their payout beforehand, yet their reward is fair: they all get the same payment for each valid submission. In this deployment, before opening a campaign applied with this incentive mechanism, the data collectors are informed about the moment when their micro-payment is allocated (i.e., at the moment the time-limited campaign is finished). We used a budget of 20 Euro for each campaign applied with this incentive mechanism.

We selected 7 out of the 10 campaigns created by the organizers of the Citizense deployment, to which monetary incentives were applied. These campaigns are campaigns with number 1, 2, 3, 4, 6, 26 and 28 in Table B.1 and B.2. We further refer to these campaigns as  $C_1$ ,  $C_2$ ,  $C_3$ ,  $C_4$ ,  $C_5$ ,  $C_6$  and  $C_7$  (see Table 4.1).

These 7 campaigns were selected to have monetary incentive applied as they offered a good spectrum of complexity and length: some campaigns contain a couple of atomic tasks without any conditional transitions while others present data collectors with a considerably longer list of tasks connected by conditional transitions. All these 7 campaigns were available in both the local language (Spanish) and English. Campaigns  $C_1, C_2, C_3, C_4$  were launched on day 1 of the 20 day experiment (i.e., duration of 20 days);  $C_5, C_6$  on day 8 (i.e., duration of 12 days); and  $C_7$  on day 14 (i.e., duration of 6 days). We intentionally selected this launching schedule to observe the behavior of the data collectors over time. Campaigns  $C_6$  and  $C_7$  were one time campaigns (i.e., a data collector can complete a one-time campaign only once); other campaigns could be completed multiple times. Table 4.1 gives an overview of the details of campaigns  $C_1$ – $C_7$ . Each of these 7 campaigns was replicated into four functionally identical versions, where to each version one of the three aforementioned incentive mechanisms ( $M_1$ , lottery;  $M_2$ , fixed micro-payment and  $M_3$ , variable micro-payment), or the fourth mechanism  $M_4$ , a control mechanism which consists of no incentive (i.e., no reward), was applied. For each of these 7 campaigns, each data collector received only one version of an incentivized campaign (i.e.,  $M_1$ – $M_4$ ), randomly distributed and permanently assigned. As such, the data collectors, for each campaign, were divided into four randomly distributed groups of equal size, where each group was assigned with a particular incentive mechanism. As a result, each individual data collector obtains a fixed set of incentivized campaigns throughout the experiment, where each incentivized campaign is associated with a random incentive mechanism. Based on post-random assignment observation, the authors confirmed that all data collectors were exposed to different incentive mechanisms, for different campaigns, and were thus aware of all the incentive mechanisms used in the experiment. For example, participant  $P_1$  may receive the set of campaigns/incentive mechanisms ( $C_1 M_1, C_2 M_2, C_3 M_4, \dots$ ); participant  $P_2$  may receive ( $C_1 M_4, C_2 M_1, C_3 M_2, \dots$ ); participant  $P_3$  may receive ( $C_1 M_1, C_2 M_4, C_3 M_3, \dots$ ); etc. Participants  $P_1, P_2$  and  $P_3$  are all permanently assigned with their list of incentivized campaigns for the whole course of the experiment. We note that this experiment hereby differs from previous works in the fact that data collectors received different realistic sensing tasks with different incentive mechanisms at the same time.

Table 4.1: The different parameters of the campaigns with incentives.

Campaign	Theme	Duration	Remark
$C_1$	Graffiti in the city	20 days	
$C_2$	City incidents	20 days	
$C_3$	Different aspects of citizen participation	20 days	
$C_4$		WiFi signal strength	20 days
$C_5$	Street animals	12 days	
$C_6$	Use of bicycle	12 days	One-time campaign
$C_7$	Final survey	6 days	One-time campaign

During the course of the experiment, data collectors may view the list of currently available campaigns. In this list (see Figure 3.9 (2)), campaigns with incentive mechanisms are marked with an icon to indicate the presence of an incentive (a Euro sign for a fixed or variable micro-payment, a gift icon for a monetary gift in the form of a voucher) and the exact incentive mechanism used (indicated by the number of the incentive between bracket); exact details of the incentive (i.e., the amount of payout) are at this point not yet visible. Campaigns without incentive do not feature an incentive logo. In order to prevent bias, the list of campaigns is shuffled every time the data collectors view their list of campaigns (i.e., refresh the campaign list). When the data collector selects a campaign, he/she will be shown the description of the campaign (see Section 3.3.5), and full details of the incentive mechanism (if available). After viewing the details of each campaign, with or without incentive mechanism, data collectors can choose whether or not to continue with the campaign in question or go back the list of campaigns and select another one.

During the experiment, the data collectors can instantly view their earned rewards (cash and monetary gifts): for example, they can see the increased amount of their earned cash right after completing a campaign with incentive mechanism  $M_2$  or after the deadline of a time-limited campaign with incentive mechanism  $M_3$  (see Section 3.3.5). They are constantly informed that the rewards (i.e., cash and gifts) are deliverable immediately after the experiment. For every campaign in this experiment, we set up various metrics for further analysis, described in the next section: the number of times a campaign is viewed (in the campaign list), the number of times a campaign is opened and the number of times a campaign is submitted (by all data collectors).



### 4.2.3 The results

Throughout the deployment, 359 participants registered in the Citizense framework, among them 179 males and 180 females. These participants included administrative staff, teaching staff, researchers and mostly students; participants come from different academic domains such as engineering, social science and life science. The participants' age range from 18 to 63, with an average value of 23.769, median value of 22 and standard deviation of 6.637. Concerning the registration, 176 participants registered through the campaign manager and the rest (183 participants) registered through the mobile application. Due to the single credential system within the framework (requirement F3), 47 participants originally registered through the campaign manager also logged in to the mobile application. Therefore, in total there were 230 data collectors using the mobile application. Among them, there were 131 females (56.9%) and 99 male data collectors (43.1%). The data collectors' ages range from 18 to 63, with a mean value of 23.47 and standard deviation of 11.72. In other words, the majority of the data collectors are young adults (aged 18–25): students of the university; the rest includes older adults who are Ph.D students, lecturers, university staffs and passers-by. The educational background of the data collectors varies as they study at or are associated with different faculties of the university, namely technology and experimental sciences, law and economics, humanities and social sciences and health sciences. Most of the data collectors are locals; 7 data collectors are international students. All data collectors gave their informed consent for inclusion before they participated in the study, and the experiment complied with the ethics regulations of Universitat Jaume I (approval reference number 04/2018).

The participants created 34 campaigns, which addressed a variety of topics mostly related to the urban context. Together with the 10 campaigns launched by the organizers, the total number of campaigns was 44. Table B.1 and B.2 give an overview of all 44 campaigns available during the deployment, among which 7 campaigns from the organizers were applied with monetary incentive to study the data collectors' reactions to monetary incentive (i.e., campaign number 1, 2, 3, 4, 6, 26 and 28). Examples of user-created campaigns include public transportation and its efficiency, public furniture, key infrastructure in the city, the city's sustainability issues and cultural issues. The length of the campaigns created by participants ranges from 3 to 45 atomic sensing tasks, a typical campaign has 6 to 8 tasks.

After the manual data cleaning process which removed empty and duplicated submissions, the cleaned dataset contains 4167 requests sent by data collectors and 2944 submissions were received for all 44 available campaigns during the 20-day period of the experiments. Data collectors were encouraged to enable their GPS sensor when interacting with the Citizense mobile application. However, they had the right to disable this GPS sensor at any time, due to the enforced privacy regulations. Among the 2944 received submissions, 67.35% of them (1983 submissions) were attached with the GPS location while the rest (961 submissions - 32.65%) were made without the GPS location. Among the data collectors, 57.82% of the data collectors (133 data collectors) always had their GPS sensor enabled when collecting data while the rest (97 data collectors - 42.18% of the data collectors) had at least one of their submission without the location data.

#### **4.2.4 Discussions**

In order to discuss the effectiveness of Citizense to collect geographical and general data, we discern the campaign authors' ability to successfully use Citizense to create a variety of qualitative campaigns, and the data collectors' results with respect to the data collected.

In the 34 campaigns created by campaign authors, they incorporated all the campaign's features that were available during the deployment, such as the location and time restrictions, the visibility of the results and attaching multimedia content in atomic sensing tasks. Apart from the relevant topics covered by campaign authors, they used 11 out of the 12 supported task types throughout their campaigns; only the noise measurement type was not used. At the same time, we did not receive any requests for further explanation of the framework in the contact email throughout the Citizense deployment. This is an indication that the campaign editor is easy to use, the existing features are appreciated and no additional technical assistance is required, especially considering the diverse backgrounds of the campaign authors and their first time using this framework.

In terms of collected data, the deployment has gathered various data types: text (through typing, speech-to-text conversion and selecting pre-defined options), numeric values, pictures, date and time values, GPS locations and WiFi measurements. These data types correspond to the types of sensing task supported



Figure 4.1: Example of the submissions during the deployment displayed by the Citizense result visualizer. (upper) The different text answers to a text-input task. (lower) The pictures and locations of the graffiti gathered by data collectors.

by the Citizense framework; only audio was not recorded, as the corresponding task was not used by campaign authors. More specifically, among the 2944 submissions received there are 327 submissions containing embedded sensory measurements and 27 submissions containing pictures (see Figure 4.1, lower), 3714 text responses, 93 numeric responses and 30 date and time responses. On a higher level, these primitive data types were combined into geographic contents

and general contents. The geographic contents include time- and geo-tagged reports on the issue of illegal graffiti, street animals and potential dangers and malfunctions in the city's cycling infrastructure. In each of these reports, the data collectors provided a comprehensive set of information such as the GPS location of the object of interest, the time of appearance, the corresponding text description, the subjective opinion, the severity of the potential danger and a picture describing the object. General contents are the data collectors' subjective feedbacks, opinions and suggestions on various issues raised in the campaigns during the deployment, such as the comments on price of the city's tram line and its punctuality, complaints about the bike sharing system and suggestions for improving the infrastructure in the city's beach.

While the large amount of submissions and variety of data collected demonstrate the potential of Citizense to engage users, the qualitative involvement can be examined through the text input and pictures they produced. In this Citizense deployment, 100% of the submitted pictures are on topic and sharp; none of these pictures are regarded invalid or low quality. In contrast, in the work of Reddy et al. (2010a), the authors described an experiment in which college students were involving in the participatory sensing task of taking photos of the trash bins in their campus. The authors reported the ratio of invalid picture of  $6 \pm 4\%$ . 35 campaigns (79.54% of the campaigns launched in this deployment) had their text-input tasks completed at a ratio over 90%. In particular, the user-created campaign on social vulnerability achieved the completion rate of 100% for its text-input tasks. Furthermore, several data collectors provided very detailed text to elaborate their answers (see Figure 4.1, upper). The reasons for some campaigns to have low completion ratio for text-input tasks are the poor description of these tasks and/or these tasks are optional.

Next to the above quantitative indications, both data collectors and campaign authors were surveyed at the end of the deployment, in order to obtain their opinion on the usefulness of Citizense. This was done through a campaign in the Citizense framework and an email-based questionnaire, answered by 31 data collectors and 32 campaign authors, respectively. 30 data collectors (96.77% of those who responded to the campaign) agreed that Citizense is a useful platform to collect various types of city-related information while 31 campaign authors (96.87% of those who answered the survey) appreciated their abilities to create campaigns on the topics of their choice. Finally, in the post-deployment phase,

face-to-face interviews were conducted with 33 participants. All the interviewed participants appreciated the various functionalities, the simplicity and ease of use of the Citizense graphical campaign manager and the mobile application. All data collectors were satisfied with the way they contributed data through the Citizense client mobile application while all campaign authors were pleased with the information gathered and visualized by the Citizense framework. During these interviews, some participants acknowledged the behaviors identified by the analysis of the deployment's results (see Section 6.1).

In conclusion, the Citizense deployment showed that the Citizense campaign editor is effective in assisting the authors in designing their data collecting process. On the other hand, data collectors showed their willingness and involvement in providing a variety of high quality data. Both campaign authors and data collectors indicated to be satisfied and useful to collect city-related information.

#### **4.2.5 Limitations of the UJI deployment**

First of all, even though we do not anticipate deviations in the overall results, the lack of an iOS version for the Citizense mobile application has excluded part of the target population to collect data and input for our analysis. Should this iOS version be available, there would have been more data collectors and submissions for the study of data collectors' reaction to monetary incentives and the analysis of their behavior in participatory sensing context.

Secondly, the UJI deployment might have suffered from the homogeneity of the participants. They are mostly university students, and to a lesser extent staff; they all have at least the education level of undergraduate studies. If the scope of the case study were bigger, i.e., extended to the whole city, a more demographically diverse background of participants can be achieved.

Thirdly, regarding the analysis of the data collector's spatial behavior, it is based on the geo-tagged submissions (1983 submissions), which represent 67.35% of the total number of submissions as data collectors have the right to disable their mobile device's GPS sensor at any time. Therefore, our conclusion on the data collector's spatial behavior has to be interpreted with caution, although it is believed that those who submitted results without location data have their spatial behavior similar to the identified spatial behavior. This limitation, however, is shared with any other studies using mobile device produced data (e.g., tweets), as users always

have the option to disable their location sensor.

Finally, the number of campaigns with incentives in the UJI deployment was limited (7), and campaigns launched late in the experiment attracted few participants (3). A larger number of campaign with incentives would have given the data collectors more chances to interact with the campaigns and its incentive mechanisms, yielding the conclusions based on a larger dataset.

## Chapter 5

# The effectiveness of monetary incentives in motivating participants

There exists a wide range of incentives that can be applied to participatory sensing applications, and various classification schemes to classify them (see Section 2.2.1 for details). While all incentives have their merits, in this chapter we focus specifically on studying user behavior in participatory sensing under the presence of various monetary incentives. We elected to study this type of incentive for two main reasons: 1/ monetary incentive are generic, and thus they can be applied to any sensing task, independent of its type, purpose and content of the associated campaign, etc. Such incentives fit well with Citizense, which is designed as a generic, multi-purpose participatory sensing framework, and thus indeed supports a variety of sensing tasks and campaigns; 2/ although monetary incentives have been extensively covered in literature, most of the existing works address the theoretical side of this issue. On the other hand, the few works that do describe an experimental evaluation are performed in a specific context/conditions and/or for a specific task. Furthermore, they report few participants (i.e., typically 50 or less, with few exceptions). In this chapter, we are seeking empirical data on the behavior of users under different monetary incentives and over various sensing campaigns/tasks (i.e., not in a specific context nor under specific conditions).

This chapter first details the results concerning the campaign's metrics from the aforementioned UJI deployment in Chapter 5.1. Based on these metrics, the

effectiveness of the deployed incentive mechanisms is studied. Specifically, the proof for the effectiveness of the three deployed incentive mechanisms on the participation (i.e., opening a campaign, submitting results) of all data collectors can be found in Chapter 5.2. Subsequently, the set of data collectors is divided based on demographical parameters (e.g., age, gender) and the effectiveness of the three incentive mechanisms is studied in these subgroups. Chapter 5.3 studies the reaction of male and female data collectors on the three deployed incentive mechanisms while Chapter 5.4 studies the reaction of young and old data collectors on the same deployed incentive mechanisms. Chapter 5.5 provides an overview of the ranking of the incentive mechanisms for all data collectors and for the aforementioned subgroups. Lastly, Chapter 5.6 presents the final conclusions. The discussion in this chapter is based on our previous publication (Khoi et al., 2018).

## 5.1 Raw data and descriptive analysis

In this chapter, the number of times that campaigns with incentives were viewed (in the campaign list), opened, and submitted are detailed. To visualize this raw data, we use stacked bar charts with bottom parts (dark color) indicating weekday list view/open/submissions and top parts (light color) showing weekends.

Figure 5.1 shows the number of list views from the four versions of a campaign (i.e., with incentive mechanisms  $M_1$ – $M_4$ ), for all campaigns  $C_1$ – $C_7$ . Please note the differences in number of list views among the 4 versions of the same campaigns, for all campaigns. As mentioned before in Chapter 4.2.2, when a data collector views a (fixed) set of campaigns, the numbers of list views of all these campaigns will be incremented by 1. Therefore, the difference among the numbers of list views can be attributed to the more active participation of certain data collectors, who repeatedly view their assigned list of campaigns, thereby increasing the numbers of views of their assigned campaigns with associated incentive mechanism. The different durations among the campaigns result in the differences in number of overall views among  $C_1$ – $C_4$ ,  $C_5$ – $C_6$  and  $C_7$ ; each campaign in the first group received many more views than that of the second group and third group, respectively.

Figure 5.2 shows the number of opens from each version of a campaign, for all the campaigns  $C_1$ – $C_7$ . From this figure, it is clear that in most cases, campaigns with mechanism  $M_2$  were opened more than campaigns with other mechanisms.



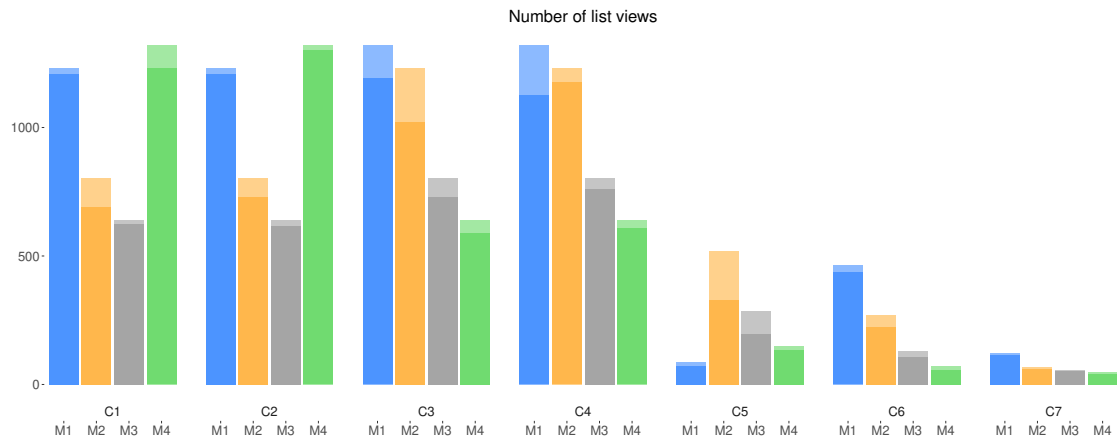


Figure 5.1: The number of list views of the campaigns. For each individual bar, dark (**bottom part**) and light color (**top part**) represent weekdays and weekends, respectively.

From Figures 5.1 and 5.2, although campaigns  $C_1$  and  $C_2$  with mechanism  $M_4$  were presented (viewed) more to the data collectors (more visible) than the same campaigns with other incentive mechanisms (i.e.,  $M_1$ ,  $M_2$  and  $M_3$ ), these campaigns ( $C_1$  and  $C_2$  with  $M_4$ ) have been opened only the same amount of times as  $C_1$  and  $C_2$  with mechanism  $M_3$ , and fewer than with mechanism  $M_1$  and  $M_2$ . In the same way, despite the fact that campaigns  $C_3$  and  $C_4$  with mechanism  $M_4$  received slightly less number of list views than  $C_3$  and  $C_4$  with  $M_3$ , the former ( $C_3$  and  $C_4$  with  $M_4$ ) received much less number of opens compared with the latter ( $C_3$  and  $C_4$  with  $M_3$ ). This observation is a first indication that campaigns lacking an incentive ( $M_4$ ) are less attractive to the data collectors. Finally, as for list views, we might observe an overall decrease in the total amount of opens among each campaign of the groups  $C_1$ – $C_4$ ,  $C_5$ – $C_6$  and campaign  $C_7$ , explainable by the subsequent shorter duration of campaigns.

Figure 5.3 shows the number of submissions from each version of a campaign with incentives, for all the considered campaigns. Similar to Figure 5.2, Figure 5.3 shows the higher performance of incentive mechanism  $M_2$  compared with other mechanisms, which is analyzed later in this chapter. We can also again observe the lower number of submissions from  $C_5$ ,  $C_6$  and especially  $C_7$  compared with that of  $C_1$ – $C_4$ , which can be explained by their later launching dates and shorter durations.

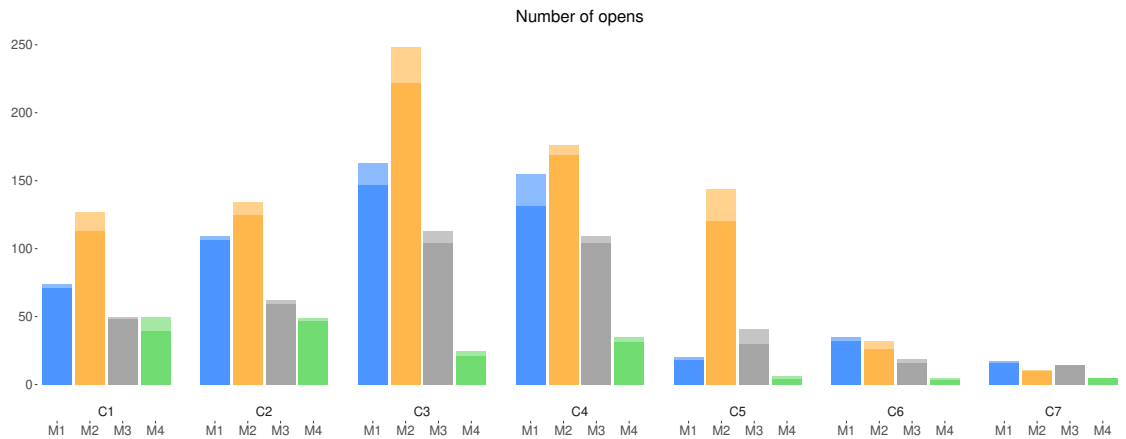


Figure 5.2: The number of opens of the campaigns. For each individual bar, dark (**bottom part**) and light color (**top part**) represent weekdays and weekends, respectively.

Considering Figures 5.1–5.3, we make two overall observations. First, we note that the amount of weekend interactions in all three charts is minimal, compared to weekday interactions; participants were thus notably less active during weekends, and weekend interactions thus do not appear to influence the overall (all days; weekday + weekend) trends. Second, there is an overall decrease in the number of interactions toward the end of the experiment; data collectors were not as active as in the beginning of the experiment. The decrease of participation in later campaigns can be partly explained by the diminishing duration of these campaigns ( $C_1$ – $C_4$  20 days,  $C_5$ – $C_6$  12 days and  $C_7$  6 days). However, extrapolating the duration may not fully explain the observed effect. This is also reflected when considering daily average participation. For example and based on daily average, the first 4 campaigns launched, two campaigns launched later and the last campaign receive 13.55, 9.20 and 5.16 submissions, respectively. Thus, there might be other factors that affect the data collectors' interest in participating toward the end of the experiment. It can be considered that wearing out of novelty and lack of continuous promotion may have contributed to this downward trend in the number of submissions.

Based on the three values (list view, open, submission), additional descriptive ratios are calculated. These ratios are calculated using the overall number of list view, open and submission of the corresponding campaigns (all days, i.e.,

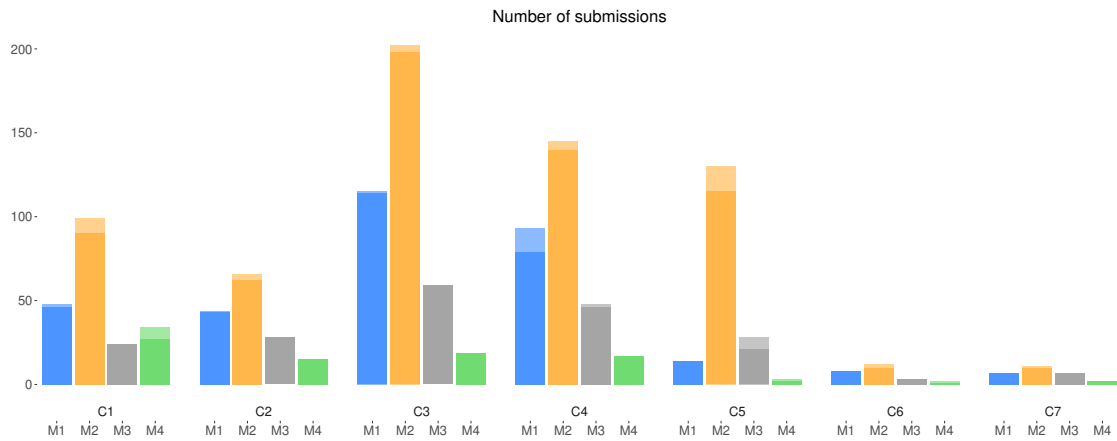


Figure 5.3: The number of submissions of the campaigns. For each individual bar, dark (**bottom part**) and light color (**top part**) represent weekdays and weekends, respectively.

weekdays + weekend) (later, we will show there is no significant difference in results when considering weekdays only versus all days). The completion ratio is defined as the ratio between the number of submissions and the number of opens of the same campaign. This ratio tells how an incentive mechanism potentially drives the data collectors to complete a particular campaign once they opened the campaign and learned the details of the incentive, and is depicted in Figure 5.4. In this figure, the completion ratio reaches 1 for mechanism  $M_2$  in campaign  $C_7$ . It can be explained by the fact that only the most motivated data collectors were still actively using the application at the end of the experiment, and they took their last chance to immediately earn the corresponding reward. In general, mechanism  $M_2$  achieved a higher completion ratio than the other three mechanisms. This can be explained by the fact that data collectors apparently prefer a fixed, certain reward, compared to an uncertain award. The situation for the other mechanisms is more mixed, with  $M_1$  seemingly having a slight edge over  $M_3$  and  $M_4$  possibly due to the fact that data collectors prefer, in lack of a fixed micro payout, to have a chance to win a valuable gift, than waiting for an unknown amount or no award. Interestingly, and in contrast to the general completion and attraction ratio (discussed further), mechanism  $M_4$  seems to perform well in some campaigns ( $C_1$ ,  $C_3$  and  $C_6$ ): once opened, there is a high possibility the data collector will also complete the campaign. This can be explained by the fact that data collectors,

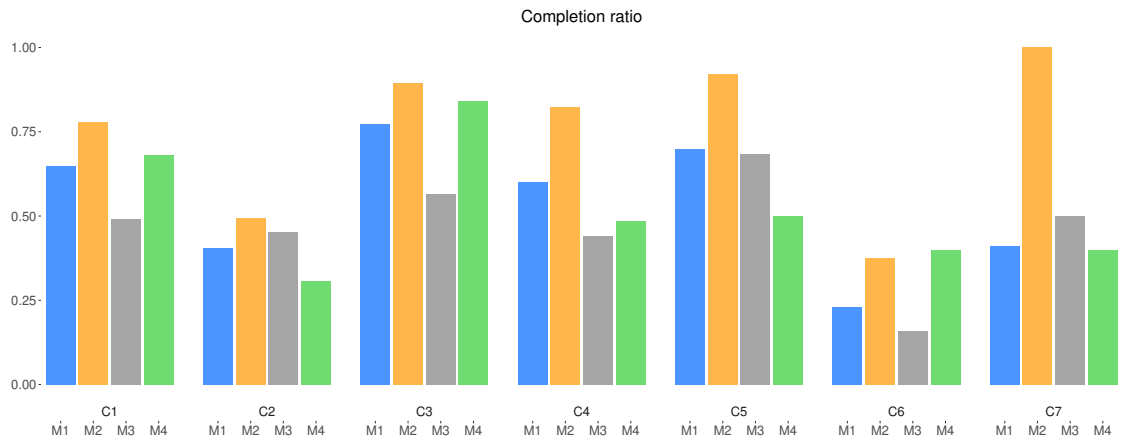


Figure 5.4: The completion ratios of the campaigns based on all days.

despite the lack of incentive, might be attracted by the content of these campaigns, and therefore open and complete them.

The general completion ratio is the ratio between the number of submissions and the number of list views of the same campaign. This ratio shows how the presence of an incentive mechanism drives the data collectors to complete a particular campaign. Figure 5.5 depicts the values of the general completion ratio among campaigns. In this figure,  $M_4$  generally is least effective, showing clearly how the presence of an incentive drives data collectors to contribute, and conversely, the lack of incentive deters them to contribute. On the other hand,  $M_2$  maintains its top position, showing that fixed micro-payment seem to drive data collectors more to contribute than other mechanisms. As before, the situation for  $M_1$  and  $M_3$  is less clear. In general, data collectors seem to be more interested in completing a campaign which surely provides them with rewards immediately, and do not appreciate getting nothing.

The attraction index is the ratio between the number of open and the number of list views of the same campaign. It represents the effectiveness of an incentive mechanism in attracting data collectors to open a particular campaign (yet without knowing its exact incentive details), and conversely, their low interest in campaigns without incentive. Figure 5.6 depicts the values of the attraction indexes among the versions of the campaigns with incentives. While mechanism  $M_2$  generally seems to outperform other mechanisms and  $M_4$  has the least successful performance, mechanism  $M_3$  generally slightly beats  $M_1$  in attracting data collectors to click

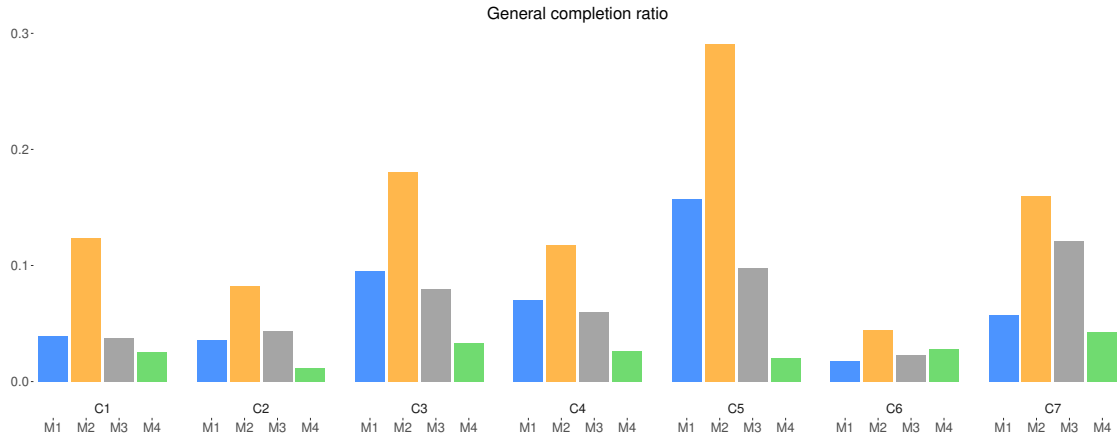


Figure 5.5: The general completion ratios of the campaigns based on all days.

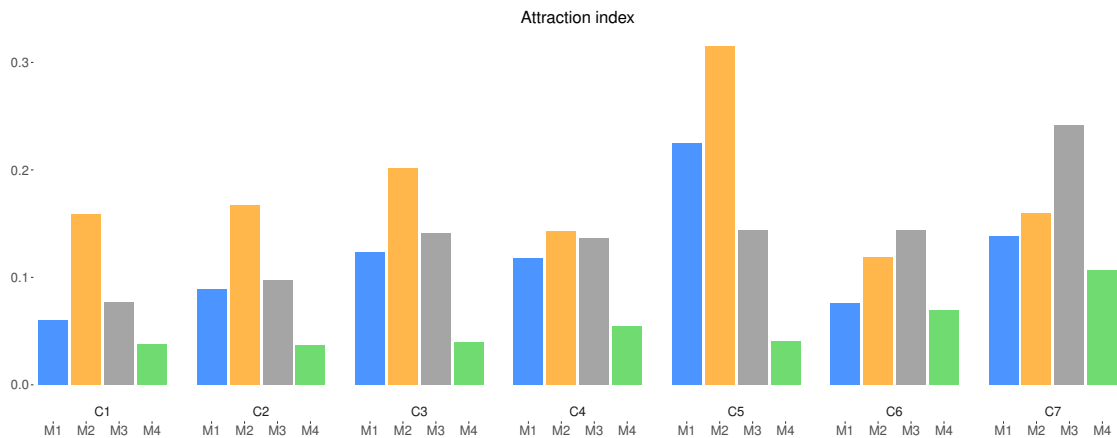


Figure 5.6: The attraction indexes of the campaigns based on all days.

on a campaign when they see it in the campaign list. It seems data collectors are more curious how much money is distributed over data collectors, perhaps hoping for a certain and potentially relatively high payout, compared to knowing what the (value of the) gift is. However, as the completion ratio already showed (Figure 5.4), once opened, there is no clear advantage of  $M_3$  over  $M_1$  (rather on the contrary). This might suggest that, once realizing the potential gains (and perhaps being disappointed) of the variable incentive mechanism  $M_3$ , users have a slight preference to complete a campaign with a chance for high payout.

Finally, in campaign  $C_7$ , we performed a final questionnaire among data collectors to qualitatively assess their opinion regarding incentive mechanisms. First, on

the question “Which type of incentives do you prefer?”, among the 31 respondents, 24 respondents selected “monetary rewards”, 6 respondents selected “no rewards needed” and 5 respondents selected “intangible rewards”. From the 24 respondents who selected “monetary rewards”, in a follow-up question, 15 indicated they preferred fixed micro-payment ( $M_2$ ), 7 respondents indicated that any monetary incentive mechanism is fine, 1 respondent selected lottery-style payout ( $M_1$ ) and 1 respondent selected variable micro-payment ( $M_3$ ). Even though the questionnaire was performed on a relatively small sample (31) of the total amount of participants, it indeed qualitatively confirms that a majority of data collectors prefer a monetary incentive (compared to no or intangible incentives), and a majority prefers a fixed micro-payment as concrete monetary incentive mechanism.

## 5.2 Overall statistical analysis

This chapter is dedicated to the overall statistical analysis of the raw data presented previously, in order to discover the general trends for the whole set of participants with respect to monetary incentive mechanisms. To this aim, we analyze the effectiveness of monetary incentive mechanisms by comparing different metrics of campaigns with incentives with those of the same campaign without incentive. Due to the setup of the experiment, whereby each campaign is assigned with each of the four incentives mechanisms, and subsequently these campaign-incentive couples are equally (randomly) distributed among participants, any possible effect of campaign topics is the same for every incentive. In this way, we can neglect the influence of a campaign topic on the decision to participate in a campaign among the data collectors. Furthermore, the broad range of campaign topics also help to minimize the aforementioned influence: some topics may be more appealing than others, but they’ll be so for each of the four incentive mechanisms. We particularly focus on the number of submissions, as this is the final and most important step in the data collecting process. It is worth noting that our analysis is based on the number of submissions of the 7 campaigns with incentives produced by 230 data collectors. This number of users is larger than that of most real-world deployments found in literature - see Section 2.2.2. Despite this, the UJI deployment still received a relatively small number of data collectors compared with the total number of potential data collectors (i.e., all university students and staffs). Therefore, we call for caution when interpreting the results

of our analysis, in particular for generalization towards the full population. Ideally, the experiment should be replicated to strengthen generalizability. In this analysis, we consider two variants, only weekdays and all days (weekends + weekdays), in order to verify if the changing routines and patterns of user behavior during weekends influence the overall findings in participatory sensing behavior toward incentives. Throughout this analysis, two-tailed hypothesis will be considered. We first verify if there is a preference for certain incentive mechanism(s) when opening a campaign:

- Hypothesis  $H1$ : Data collectors do not have a preference when **opening** a campaign  $C_i$  (with  $1 \leq i \leq 7$ ), with respect to the different incentives  $M_j$  (with  $1 \leq j \leq 4$ ).

We hereby consider two variants of  $H1$ , namely considering all data (denoted as Hypothesis  $H1_{all}$ ) and considering only weekdays (denoted as Hypothesis  $H1_{weekdays}$ ). We will consistently use this notation for further hypothesis in the article.

In order to test if the data in question are indeed independent, we perform a  $\chi^2$  test (using the CHISQ.TEST statistical function of MS Excel) by considering significance level 0.05 on the number of campaign opens, per campaign  $C_i$  and for each incentive mechanism  $M_j$ . This  $\chi^2$  test needs the observed values and expected values as its two inputs. The former is read from the corresponding campaign metric while the latter is the average of the number of opens from the 4 versions of the same campaign. The significance level of 0.05 is chosen as it is a typical value for significance level in statistical analysis. For the 20-day period, the result of the  $\chi^2$  test ( $p$ -value) for the campaigns with incentives ( $C_1$ – $C_7$ ) is less than  $1 \times 10^{-6}$  for  $C_1$ – $C_6$ , and 0.082 for  $C_7$ ; for weekdays only, the  $\chi^2$  test gives similar results:  $p$ -values are less than  $1 \times 10^{-5}$  for  $C_1$ – $C_6$  and 0.09 for  $C_7$ . As we perform the  $\chi^2$  on each campaign separately, our conclusion will not suffer from the multiple comparison problem. Consequently, with 95% confidence, we cannot accept Hypothesis  $H1_{all}$  nor  $H1_{weekdays}$  for all the campaigns, except for  $C_7$ : there are significant differences among the number of opens of the four versions of each campaign  $C_1$ – $C_6$ . In other words, for campaigns  $C_1$ – $C_6$ , data collectors seem to have a preference regarding the presence of an incentive when they open any of these campaigns, regardless of the campaign content. In contrast, there is no significant difference between the numbers of opens of the four versions

of campaign  $C_7$ . It can be explained that  $C_7$  was launched in the last week of the deployment when there were much less data collectors, as a result there is not enough data to show statistical significance. We now turned to test the following hypothesis:

- Hypothesis  $H2$ : Data collectors do not have a preference when **completing** a campaign  $C_i$  (with  $1 \leq i \leq 7$ ), with respect to the different incentives  $M_j$  (with  $1 \leq j \leq 4$ ).

Similarly and per campaign  $C_i$ , we perform a  $\chi^2$  test by considering significance level 0.05 on the number of submissions of the four variants of the campaign ( $M_1$ – $M_4$ ) for the full 20-day period (verifying Hypothesis  $H2_{all}$ ) and the weekdays (verifying Hypothesis  $H2_{weekdays}$ ). For the former, the results of the  $\chi^2$  test ( $p$ -value) is less than  $1 \times 10^{-6}$  for  $C_1$ – $C_5$ , 0.016 for  $C_6$  and 0.11 for  $C_7$ , and for the latter, the  $p$ -values are less than  $1 \times 10^{-6}$  for  $C_1$ – $C_5$ , 0.021 for  $C_6$  and 0.16 for  $C_7$ . Consequently, we cannot accept Hypothesis  $H2_{all}$  nor  $H2_{weekdays}$  for all the campaigns ( $p$ -value  $< 0.05$ ), except for  $C_7$ . With confidence level of 95%, data collectors seem to prioritize certain incentive mechanisms when they complete any of these campaigns with incentives, regardless of the campaign content. In contrast, there is again no significant difference between the number of submissions of the four versions of campaign with incentives  $C_7$ , mainly due to lower amount of data collectors participating at the end.

There is thus a statistically significant difference among different incentives, both in exploring (opening) and completing campaigns. Furthermore, the raw data (see Chapter 5.1) suggests that the presence of incentives seems to positively affect the behaviors of the data collectors (i.e.,  $M_4$  performs worst), and that among the incentive mechanisms,  $M_2$  performs best. Consequently, there is a further need to compare the incentive mechanisms themselves, in order to determine which incentive outperforms another. For this comparison, we selected the campaigns  $C_1$ ,  $C_2$ ,  $C_3$  and  $C_4$  as these were exposed to all the data collectors for the same amount of time (i.e., throughout the full length of the deployment). We first confirm a statistical difference in opening campaigns with different incentive mechanisms, over all considered campaigns:

- Hypothesis  $H3$ : Data collectors **open** the campaigns  $C_i$  (with  $1 \leq i \leq 4$ ) irrespective of the incentive mechanisms  $M_j$  (with  $1 \leq j \leq 4$ ).



Using R, a non-parametric Kruskal-Wallis test (used for comparing two or more independent samples of equal or different sample sizes) is performed on the number of opens of the campaigns  $C_1-C_4$ . This test needs the test variable and the grouping variable as the inputs. The former is the number of opens from all the 4 versions of the campaign  $C_1-C_4$  (16 numbers of open in total) and the latter is the list of the corresponding incentive index (e.g., 1 and 4). As we perform only one Kruskal-Wallis test on 4 numeric data vectors which produces only one conclusion, this test is excluded from the multiple comparison problem. This test gives the result  $p$ -value of 0.009 for all days and 0.006 for weekdays. Therefore, Hypothesis  $H_{3all}$  and  $H_{3weekdays}$  cannot be accepted ( $p$ -value  $< 0.05$ ) and the differences between the incentive mechanisms as perceived by the data collectors are confirmed. In other words, with 95% level of confidence, at least one mechanism statistically dominates another one. In a further step, we make a pair-wise comparisons of the incentive mechanisms based on the number of opens, in order to determine which incentive mechanism is statistically more attractive. For this purpose, the Mann-Whitney U test (used to determine whether two independent samples were selected from populations having the same distribution) was performed on each pair of incentive mechanisms, using the number of opens of campaigns  $C_1-C_4$ . This procedure of performing post-hoc analysis (e.g., Mann-Whitney U test) is known in literature for providing a good balance between Type 1 and Type 2 errors (Conover and Iman, 1979). In order to further prevent the higher probability of getting a Type 1 error when doing multiple comparisons simultaneously, several solutions are available, among which adjusting the  $p$ -value. Among the available  $p$ -value correction methods, the Benjamini-Hochberg correction (or "BH" in R) is selected due to its ability to control the false rate discovery. Table 5.1 details the results of these Mann-Whitney U tests with and without BH correction.

Table 5.1: The results of the Mann-Whitney U test with and without BH correction for all pairs of incentive mechanisms, using the number of opens during the full period (all) and only weekdays.

<b>Pairs of Mechanism</b>	$M_1-M_2$	$M_1-M_3$	$M_1-M_4$	$M_2-M_3$	$M_2-M_4$	$M_3-M_4$
corrected $p$ -value (all)	0.343	0.294	0.043	0.043	0.043	0.043
original $p$ -value (all)	0.343	0.245	0.029	0.029	0.029	0.029
corrected $p$ -value (weekdays)	0.343	0.132	0.044	0.044	0.044	0.044
original $p$ -value (weekdays)	0.343	0.110	0.029	0.029	0.029	0.029

The results ( $p$ -value) in Table 5.1 again confirm the differences among the performances of the incentive mechanisms in attracting the data collectors to open campaigns, both during the whole deployment and the weekdays. The trends of the results are very similar for both of these two aforementioned time periods (with and without BH correction), with respect to the significance level of 0.05. Specifically, with a level of confidence 95%, there are significant differences in performance between the pairs of incentive mechanisms  $M_1$  and  $M_4$ ,  $M_2$  and  $M_4$  and  $M_3$  and  $M_4$  and  $M_2$  and  $M_3$ . We can conclude that mechanisms  $M_1$ ,  $M_2$ ,  $M_3$  are all more effective than  $M_4$ ; in other words, incentive mechanism  $M_4$  has inferior performance and is thus least effective in driving the data collectors to open a campaign. Among mechanism  $M_1$ ,  $M_2$  and  $M_3$ , we can only conclude that  $M_2$  is more effective than  $M_3$ ; no statistical difference could be found between mechanism  $M_1$  on one hand, and  $M_2$  or  $M_3$  on the other hand. Nevertheless, the latter, in combination with the dominance of  $M_2$  over  $M_3$ , indicates that  $M_1$  must necessarily lie in between  $M_2$  and  $M_3$ . Indeed, this is confirmed by the overall average amount of openings for each incentive mechanism (see Table 5.2), and we can thus conclude a general descending order  $M_2 > M_1 > M_3 > M_4$  in terms of their effectiveness in stimulating campaign openings, for the whole duration of the deployment and for the weekdays.

Table 5.2: Average openings per incentive mechanism, for all campaigns  $C_1-C_4$ .

<b>Incentive Mechanism</b>	$M_1$	$M_2$	$M_3$	$M_4$
Average amount of openings (all)	125.25	171.25	83.5	39.75
Average amount of openings (weekdays)	113.75	157.25	78.75	34.5

In the same way, we compare the effectiveness of the incentive mechanisms in driving the data collectors to complete and submit sensing campaigns. We first confirm a statistical difference in completing campaigns with different incentive mechanisms, over all considered campaigns:

- Hypothesis  $H4$ : Data collectors **complete** the campaigns  $C_i$  (with  $1 \leq i \leq 4$ ) irrespective of the incentive mechanisms  $M_j$  (with  $1 \leq j \leq 4$ ).

A non-parametric Kruskal-Wallis test on the number of submissions of campaigns  $C_1-C_4$  gives the result  $p$ -value of 0.008 for the overall period and 0.009 for weekdays. Therefore, Hypothesis  $H4_{all}$  and Hypothesis  $H4_{weekdays}$  cannot be accepted ( $p$ -value  $< 0.05$ ); these non-acceptances once again confirm the differences between the incentive mechanisms as perceived by the data collectors, with the level of confidence of 95%. The result of this test also indicates that at least one mechanism statistically overwhelms another one.

We then make pair-wise comparisons of the incentive mechanisms based on the number of submissions of these aforementioned campaigns to determine which incentive mechanism is more effective in pushing data collectors to complete campaigns. As before, the results of the Mann-Whitney U test with and without BH correction on each pair of incentive mechanisms, using the number of submissions of campaigns  $C_1-C_4$ . Table 5.3 details the results of these statistical tests.

Table 5.3: The results of the Mann-Whitney U test with and without BH correction for all pairs of incentive mechanisms, using the number of submissions during the full period (all) and the weekdays.

Pairs of Mechanism	$M_1-M_2$	$M_1-M_3$	$M_1-M_4$	$M_2-M_3$	$M_2-M_4$	$M_3-M_4$
corrected $p$ -value (all)	0.240	0.343	0.044	0.044	0.044	0.044
original $p$ -value (all)	0.200	0.343	0.029	0.029	0.029	0.029
corrected $p$ -value (weekdays)	0.240	0.245	0.043	0.043	0.043	0.043
original $p$ -value (weekdays)	0.200	0.245	0.029	0.029	0.029	0.042

The results in Table 5.3 confirm the different performances of the incentives in pushing data collectors to complete campaigns. The trend of the results are identical for both the full time period and for weekdays, with respect to the significance level of 0.05. The use of BH correction ensures a low false discovery

rate and it does not change the overall results with respect to the significance level of 0.05. In particular, with 95% confidence, the pairs  $M_1$  and  $M_4$ ,  $M_2$  and  $M_4$  and  $M_3$  and  $M_4$ , and  $M_2$  and  $M_3$  have significant differences, as suggested by the corresponding  $p$ -values. Interestingly, these are the same pairs that have significantly different performance for opening campaigns, as indicated in Table 5.1. Based on the  $p$ -values in Table 5.3, we can conclude that  $M_4$  is least effective, and  $M_2$  is more effective than  $M_3$  in motivating data collectors to complete a campaign. The statistical test does not suggest any significant difference between  $M_1$  on one hand, and  $M_2$  and  $M_3$  on the other hand. Nevertheless, the latter, in combination with the dominance of  $M_2$  over  $M_3$ , indicates a general descending order  $M_2 > M_1 > M_3 > M_4$  during the whole length of the deployment and the weekdays in terms of effectiveness of incentive mechanisms to entice users to complete campaigns. This order is confirmed when looking at the average amount of submissions per incentive mechanism, shown in Table 5.4.

Table 5.4: Average submissions per incentive mechanism, for all campaigns  $C_1$ - $C_4$ .

Incentive Mechanism	$M_1$	$M_2$	$M_3$	$M_4$
Average amount of submissions (all)	77.75	133	41	19.25
Average amount of submissions (weekdays)	70.5	122.5	39.25	18.75

Furthermore, this order is in line with the answers given by data collectors when being explicitly asked about their preferences on incentive mechanisms at the end of the deployment (campaign  $C_7$ ). A considerable part of the data collectors (62.5% of the answered data collectors - 15 data collectors) said they prefer  $M_2$  while 4.16% opted for  $M_1$  and another 4.16% went for  $M_3$ . Although this reflection does not represent the whole set of data collectors, it does support the order of incentive mechanisms identified earlier.

As evidenced in this chapter, no different trends were found when including or excluding weekends from the data analysis. While Chapter 5.1 showed a much lower amount of activity during weekends, the different daily routines that participants exhibit during weekend (non-working) days does not influence their overall behavior toward incentive mechanisms in participatory sensing. Therefore, in the next sub-chapters, we only consider the full time period (20 days) of the

deployment for further analysis.

### 5.3 Studying the influence of gender

After studying the overall tendency toward certain incentive mechanisms in the whole deployment's population, and establishing a general order among incentive mechanisms, we now study the behavior of different demographic subgroups. In this sub-chapter, we seek to analyze how the data collectors' gender affects their behavior and the established general order among incentive mechanisms, using the number of submissions as the primary metric.

We first test for a statistically significant difference between the number of submissions of male and female data collectors:

- Hypothesis  $H5$ : Gender does not affect the number of submission when considering **completing** campaigns, for all campaigns and irrespective of incentive mechanisms.

As mentioned in Chapter 4.2.2, among the 230 data collectors there are 131 females and 99 males; they respectively submitted 1605 and 1180 times for all the 44 campaigns. Therefore, the average number of submission of a female data collector and a male data collector is 12.25 and 11.91, respectively. Concerning the campaign with incentives ( $C_1-C_7$ ), female and male data collectors made 703 and 690 submissions, resulting in the individual average of 5.36 and 6.96 submissions, respectively. In order to determine statistically significant difference, the Mann-Whitney U test was used on the pair of individual female submission vector and individual male submission vector, which results in  $p$ -value of 0.971, indicating very similar (close to 1) vector distributions. Consequently, the gender difference does not affect the general submission behavior of the data collectors.

Next, we verify Hypotheses  $H2$  and  $H4$ , considering the full time period of the deployment, separately for male and female data collectors, in order to confirm/-contradict the fact that not all incentive mechanism are considered equal (at least one mechanism dominates another), and the previously established general order among incentive mechanisms. For completeness and ease of comparison, we also repeat the results for all participants (male + female), as already provided in Chapter 5.2.

The  $\chi^2$  tests on the number of submissions from female data collectors give the results ( $p$ -value) of less than 0.05 for campaigns  $C_1$ – $C_6$  and 0.21 for campaign  $C_7$ . Similarly, the same tests for the number of submission from male data collectors give the  $p$ -values of less than 0.05 for campaigns  $C_1$ – $C_6$  and 0.22 for campaign  $C_7$ . As a result, with the significance level of 0.05, we cannot accept Hypothesis  $H2_{females}$  nor Hypothesis  $H2_{males}$  for all the mentioned campaigns, except campaign  $C_7$ , again due to its short lifetime and low number of submission. Both female and male data collectors are thus stimulated by at least one incentive mechanism.

Considering the 4 campaigns  $C_1$ – $C_4$ , to which all female and male data collectors where exposed throughout the full duration of the deployment, we first confirm a statistical difference in completing these campaigns with different incentive mechanisms by both female and male data collectors, for all considered campaigns. As in Chapter 5.2, the non-parametric Kruskal-Wallis tests on the combined number of submissions from female and male data collectors was used, resulting in  $p$ -values of 0.008 (female) and 0.003 (male), respectively. As a result, Hypothesis  $H4_{females}$  and Hypothesis  $H4_{males}$  cannot be accepted using the significance level of 0.05. With confidence level of 95%, these non-acceptances again confirm both the differences between the incentive mechanisms perceived by each gender-based group and the similar behaviors between these two genders.

For each gender, we then make pair-wise comparisons of the incentive mechanisms based on the number of submission of the aforementioned campaigns, using the Mann-Whitney U test with and without BH correction. Table 5.5 details the results of these tests.

Table 5.5: The results of the Mann-Whitney U tests with and without BH correction for all pairs of incentive mechanisms, using the number of submissions from female and male data collectors.

<b>Pairs of Mechanism</b>	$M_1$ – $M_2$	$M_1$ – $M_3$	$M_1$ – $M_4$	$M_2$ – $M_3$	$M_2$ – $M_4$	$M_3$ – $M_4$
corrected $p$ -value (female)	0.240	0.343	0.043	0.043	0.043	0.043
original $p$ -value (female)	0.200	0.343	0.029	0.029	0.020	0.029
corrected $p$ -value (male)	0.674	0.171	0.088	0.114	0.088	1.000
original $p$ -value (male)	0.561	0.114	0.029	0.057	0.029	1.000
corrected $p$ -value (all)	0.240	0.343	0.044	0.044	0.044	0.044
original $p$ -value (all)	0.200	0.343	0.029	0.029	0.029	0.029

For female data collectors, the results of these tests are similar to those of all data collectors with respect to the significance level of 0.05. The use of BH correction ensures a low false discovery rate and it does not change the overall result with respect to the significance level of 0.05. While the pairs  $M_1-M_2$  and  $M_1-M_3$  show no significant difference ( $p$ -value  $> 0.05$ ), the other pairs do have significant difference between them, at 95% level of confidence. Following a similar reasoning as for the whole deployment population, and referring to Table 5.6 showing the average number of submission per mechanism from female data collectors, we can conclude an order of  $M_2 > M_1 > M_3 > M_4$ , which is the same order as found for the whole deployment population.

For male data collectors, the result of the Mann-Whitney U test with BH correction shows a different picture. With the significance level of 0.05, we cannot reject the null hypothesis stating that for each pair of incentive mechanism  $M_i$  and  $M_j$ , male data collectors prefer one incentive over another. In other words, the male data collectors have no significant preference on a particular incentive mechanism, at 95% level of confidence. However, when looking closely at the  $p$ -values, the  $p$ -values of the pairs  $M_1-M_4$  and  $M_2-M_4$  are relatively small, which bears some similarities to those of the tests on all data collectors and female data collectors. If the BH correction is not performed, only the  $p$ -values from these two pairs are less than 0.05, which suggests that the difference in each of these pairs is significant and is similar to the result from all data collectors. The BH correction does not change the  $p$ -values from the other 4 pairs with respect to the significance level of 0.05. Such a situation, where corrected  $p$ -values yield different results than non-corrected  $p$ -values, call for careful interpretation of the results, and for replication of the experiments to clarify and confirm/disconfirm the results.

Although the differences among the incentive mechanisms perceived by male data collectors are not statistically significant, it is still desirable to establish a general order among these mechanisms. By looking at the average amount of submissions by male data collectors (5.6), which shows that the average amount of submission from  $M_2$  is the largest, followed by that of  $M_1$  and  $M_3$  and  $M_4$  in that order, a general order of preference by male data collectors can be established:  $M_2 > M_1 > M_3 > M_4$ . However, it is worth noting that this order is only indicative, as the Mann-Whitney U test with BH correction suggests that in the most pessimistic scenario, any two of these mechanisms are not statistically different from each other.

Table 5.6: Average submissions per incentive mechanism by all female and male data collectors, for all campaigns  $C_1-C_4$ .

Incentive Mechanism	$M_1$	$M_2$	$M_3$	$M_4$
Average amount of submissions (female)	43	62.75	28	11.25
Average amount of submissions (male)	34.75	70.25	13	8
Average amount of submissions (all)	77.75	133	41	19.25

The division of the whole population into two subgroups (male and female data collectors) results in a smaller sample size for the corresponding statistical analysis. As a result, although the conclusions presented in this Chapter 5.3 are valid for the population of the experiment, they might have limited value for generalization. Therefore, it is recommended to replicate this deployment and repeat this statistical analysis to strengthen the conclusions, and reach a better insight on the preference of data collectors for different genders.

## 5.4 Studying the influence of data collectors' age

A second demographic parameter, which is particularly relevant in the context of monetary incentives, is the age of participants. Particularly in a university campus, where a relevant amount of the population are students (which supposedly do not have an income), it might be worth studying if there is any difference between younger (students, no income) and older (possibly not students, with income) data collectors (note that, due to strict privacy regulations in Spain and ethical regulations in our university, we were not able to directly ask for income). Indeed, due to the relatively low monetary quantity of the used incentives, and due to participants' different daily activities and schedule, their perception and behavior toward monetary incentives may be different depending on their age. Therefore, we consider younger participants ( $\leq 25$ ), which generally are students, and older participants ( $> 25$ ) (consider that in Spain, students start university at 18, spend 4 years for a bachelor and 1 or 2 years for a master; possibly with an extra year). We verified the cut-off age by doing a similar analysis for cut-off age of 23 and 24, yielding the same results as 25). We hereby seek to analyze how the data collectors' age (younger versus older), representing student (no income) and non-students (with income), affects their behavior and the established general order



among incentive mechanisms, using the number of submissions as the primary metric.

First, we seek to determine if there is any correlation between age and amount of submissions:

- Hypothesis  $H6$ : There is no difference in the average number of submissions per person per age across the age spectrum, when **completing** a campaign, for campaigns  $C_1$ – $C_4$  and irrespective of incentive mechanism.

Figure 5.7 shows both the age of all data collectors and their average number of submissions per person per age from the four campaigns with incentives  $C_1$ – $C_4$ . Visual inspection immediately reveals that this average number of submissions per person per age is quite variable, and decreases especially for older ages (not considering outliers of age 34 and 39). To statistically determine any correlation, we performed a Kendall rank correlation analysis, which measures the relationship of the average number of submissions per person per age and the data collectors' age, in order to find the correlation between these two variables. This analysis considered the four campaigns with incentives  $C_1$ – $C_4$  which were exposed to all data collectors throughout the deployment. The Kendall correlation coefficient is  $-0.437$  with the corresponding  $p$ -value of  $0.002$ , meaning that there is a significant negative correlation between age and the average number of submissions per person per age. In other words, on average, older data collectors submit less results compared to their younger peers. The outliers in Figure 5.7 (age 34 and 39) all concern a single participant who submitted a large amount of answers for a single campaign, and thus do not represent a statistical interruption of the general trend.

In a further step, we aim to verify whether the previously established ordering for incentive mechanisms is valid for younger ( $\leq 25$ ) and older ( $> 25$ ) data collectors. As for gender, we therefore verify Hypotheses  $H2$  and  $H4$ , considering the full time period of the deployment, separately for younger and older data collectors. In the deployment, there were 195 younger data collectors and 35 older data collectors. Regarding all 44 campaigns available during the deployment, younger and older data collectors made 2446 and 339 submissions, resulting in the average number of submission of 12.54 and 9.68 per data collector, respectively. For the campaigns with incentives ( $C_1$ – $C_7$ ), younger and older data collectors submitted 1224 and 169 times, resulting in the individual average of 6.27 and 4.82, respectively.

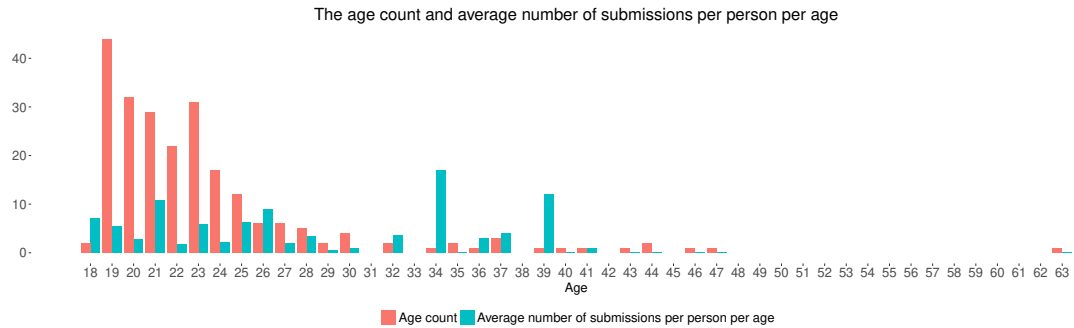


Figure 5.7: The age count and average number of submissions per person per age, for campaigns  $C_1$ – $C_4$ .

The  $\chi^2$  tests on the number of submissions from younger data collectors give the results ( $p$ -value) of less than 0.05 for all the 7 campaigns ( $C_1$ – $C_7$ ). Therefore, we cannot accept Hypothesis  $H2_{younger}$  for all the considered campaigns at 95% level of confidence; a preference for at least one incentive mechanism over another is present in all the considered campaigns. Note that when considering all data collectors (see Chapter 5.2—Hypothesis  $H2$ ), we could not find a significant difference for campaign  $C_7$ . On the other hand, the  $\chi^2$  tests on the number of submissions from older data collectors gives slightly different results: we found  $p$ -values of less than 0.05 for campaigns  $C_1$ – $C_5$ , 0.11 for campaign  $C_6$  and 0.57 for campaign  $C_7$ . Consequently, at 95% level of confidence, we cannot accept Hypothesis  $H2_{older}$  for all the considered campaigns, except  $C_6$  and  $C_7$ . In other words, older data collectors do have a preference regarding the presence of incentive for campaigns  $C_1$ – $C_5$ , for the other two campaigns there is no significant difference among the four incentive mechanisms. Again, we note a slight variation: for all data collectors (see Chapter 5.2—Hypothesis  $H2$ ), we did find a significant difference in incentive mechanisms for campaign  $C_6$ , while the insignificance for campaign  $C_7$  for older people seems to compensate its significance for younger people. Similar as for earlier analyses, we note that the importance of these variations should not be overstated, as the shorter duration and lower participation of campaigns  $C_5$ – $C_7$  may play a role.

In the same way as previous analyses, more attention is paid on campaigns  $C_1$ – $C_4$  due to their longer exposure (20 days) to the data collectors and higher participation. We first confirm a statistical difference in the number of submissions of campaigns with different incentive mechanisms, for all campaigns, by younger and

older data collectors. The Kruskal-Wallis test on the number of submissions from younger and older data collectors give the  $p$ -values of 0.006 and 0.09, respectively. As a result, Hypothesis  $H_{A_{younger}}$  cannot be accepted considering the significance level of 0.05. With 95% level of confidence, this non-acceptance again confirms the differences between the incentive mechanisms perceived by younger data collectors, which need to be analyzed further. For older data collectors, the result of the Kruskal-Wallis test suggests that there is no significant difference among the incentive mechanisms perceived by older data collectors. We further study the pair-wise comparisons of the four incentive mechanisms based on the number of submissions of campaigns  $C_1-C_4$ , using the Mann-Whitney U test. As before, we report the results with and without BH correction. Table 5.7 details the results of these tests, along with the results for all users, repeated from Chapter 5.2.

Table 5.7: The results of the Mann-Whitney U tests, with and without BH correction for all pairs of incentive mechanisms, using the number of submissions from younger and older data collectors.

<b>Pairs of Mechanism</b>	$M_1-M_2$	$M_1-M_3$	$M_1-M_4$	$M_2-M_3$	$M_2-M_4$	$M_3-M_4$
corrected $p$ -value (younger)	0.343	0.137	0.044	0.044	0.044	0.044
original $p$ -value (younger)	0.343	0.114	0.029	0.029	0.029	0.029
corrected $p$ -value (older)	0.460	0.660	0.170	0.290	0.290	0.660
original $p$ -value (older)	0.306	0.633	0.028	0.144	0.102	0.552
corrected $p$ -value (all)	0.240	0.343	0.044	0.044	0.044	0.044
original $p$ -value (all)	0.200	0.343	0.029	0.029	0.029	0.029

For younger data collectors, the use of BH correction ensures a low false discovery rate and it does not change the overall result with respect to the significance level of 0.05. Furthermore, the results of these tests are similar to those of all data collectors with respect to the significance level of 0.05. The tests show that all pairs of incentive mechanisms have a significant difference, except the pairs  $M_1-M_2$  and  $M_1-M_3$ . Table 5.8 conforms to these differences; the dominance of  $M_2$  over  $M_3$  suggests that  $M_1$  lies between these two mechanisms, and all these three mechanisms all dominate  $M_4$ . Therefore, the general order among the incentive mechanisms can be described as  $M_2 > M_1 > M_3 > M_4$ .

For older data collectors, the results of the Mann-Whitney U test with BH correction suggest that there is no significant difference among the incentive mechanisms,

which is in line with the conclusion from the Kruskal-Wallis test. In other words, in the most conservative interpretation, older data collectors do not show any preference on a particular incentive mechanism. Compared with the original  $p$ -value, the trend is similar, except for the pair  $M_1$ – $M_4$ , which suggests a significant difference in the most optimistic scenario (when BH correction is not used). Nevertheless, by referring to Table 5.8, we can establish the general order among the incentive mechanisms as  $M_2 > M_1 > M_3 > M_4$ , although the gaps between  $M_2$ ,  $M_1$ ,  $M_3$  and  $M_4$  are not significant. Again, we need to call for caution when interpreting this result, as few older data collectors (smaller sample size) were available; therefore, the results should be confirmed by repeated experiments.

Table 5.8: Average submissions per incentive mechanism by younger and older data collectors, for all campaigns  $C_1$ – $C_4$ .

Incentive Mechanism	$M_1$	$M_2$	$M_3$	$M_4$
Average amount of submissions (younger)	69	116.25	34	16.75
Average amount of submissions (older)	8.75	16.75	7	2.5
Average amount of submissions (all)	77.75	133	41	19.25

The division of the whole population into two subgroups of younger and older data collectors results in the smaller sample size for the corresponding statistical analysis. As a result, the conclusions presented earlier in this Chapter 5.4 might have limited value for generalization. Therefore, it is recommended to replicate this deployment in other scenarios and repeat this statistical analysis to reach a better insight on the behavior of data collectors in different age.

## 5.5 Further analysis and summary of results on incentive mechanisms

First of all, it is empirically clear that monetary incentives enhance the participation of the data collectors in the Citizense platform, both in terms of exploring (opening) campaigns and completing (submitting) campaigns. While the action of opening a campaign shows that a data collector is interested in that specific campaign, the submission matters more as that data collector finishes the sensing process and produces a result needed by the campaign author. To show the

importance of incentives, it was first statistically confirmed using two different tests, i.e., by not accepting Hypothesis  $H1$  and  $H2$  (for campaigns  $C_1-C_6$ ), and Hypothesis  $H3$  and  $H4$  (for campaigns  $C_1-C_4$ ), that both when considering the campaigns separately, and when considering the campaign jointly, there is a statistically significant difference between the different incentive mechanisms, both in opening and completing campaigns. We furthermore confirmed that including or excluding weekend, in which data collectors exhibit a different schedule with different daily activities, does not influence the result of the analysis. In absolute terms, and in many cases, the numbers of opens and submissions of the campaigns that are applied with mechanism  $M_1$ ,  $M_2$  and  $M_3$  (with incentive) are higher than those of the same campaign with mechanism  $M_4$  (without incentive), regardless of the campaign content. This suggests that participants actively seek out and prefer campaigns with incentives, and is in line with one of the conclusions from (Gao et al., 2015b). We note that regarding quality (e.g., the ratio of blank submissions, the length of textual answers), we proportionally did not observe a difference between campaigns with or without incentive. This is an indication of the fact that data collectors generally did not abuse the framework to get extra rewards by submitting blank answers or low quality information, which is in line with conclusions from (Gritz, 2004), where volunteers were doing similar tasks under the presence of material incentives.

Once statistically shown that there are significant differences in the performance of incentive mechanisms, we shifted our focus to establishing a ranking between different mechanisms: which monetary incentive mechanism performs better than another. Our analysis (the Mann-Whitney U tests with BH correction using the number of opens and submissions) confirm what the absolute numbers already suggested: in all cases, campaigns with incentive mechanisms perform significantly better than campaigns without, both when exploring (opening) and completing (submitting) them. Among the monetary incentive mechanisms ( $M_1$ ,  $M_2$  and  $M_3$ ), in absolute numbers (see Figures 5.2 and 5.3), a fixed micro-payment (mechanism  $M_2$ ) works more effectively than other incentive mechanisms, both to attract participants (opening campaigns) and to entice them to complete campaigns. Next in line is a lottery-based award, performing better than a variable micro-payment, yet we note that the completion ratio, and especially the general completion and attraction ratios show a mixed image; the latter even suggests a slight advantage for a variable micro-payment. This might indicate that even

if initially attracted and opening a micro-payment campaign, participants do not appreciate the potentially low payout and do not complete the campaign. The statistical analysis (the Mann-Whitney U tests with BH correction) however could only confirm the superiority of fixed micro-payment over variable micro-payment; no statistically significant difference could be shown between lottery-style incentive and both fixed and variable micro-payments. Nevertheless, by reasoning over the aforementioned statistically significant differences, and by comparing the overall average of openings/submissions per incentive mechanism, an overall descending ranking of  $M_2 > M_1 > M_3 > M_4$  could be established.

The different ratios (completion ratio, general completion ratio and attraction index) paint a similar picture of the performance of the incentive mechanisms. These ratios shows how the data collectors are driven to participate in a campaign at different stages in their data collection process: exploring the list of campaigns and selecting a campaign, learning the full details of the campaigns and its associated incentive, and completing the campaign to be eligible for the reward(s). Among the incentive mechanisms, mechanism  $M_2$  outperforms all other mechanisms in attracting data collectors to explore a campaign as well as complete it, regardless of the campaign content. Its top performance is clearly shown in Figures 5.4–5.6. On the other side of the ranking, mechanism  $M_4$  has the least successful performance; this outcome can be expected due to its no-reward mechanism. In between these two ends are the mechanism  $M_1$  and  $M_3$ , they have comparable performances and there is no clear winner among these two mechanisms. It seems that mechanism  $M_3$  (and obviously  $M_2$ ) is a bit more effective in attracting data collectors to open a campaign than  $M_1$ , which is evident by the larger attraction index (Figure 5.6). However, mechanism  $M_1$  slightly beats  $M_2$  in driving data collectors to complete a campaign although the differences are usually not large.

Once the general order of the incentive mechanisms was established for the complete set of data collectors, we shifted our focus to specific subgroups, based on demographic parameters, of the overall population: male versus female and younger versus older. Using the number of submissions, each subgroup's behaviors of completing campaigns are first compared and then examined separately. Concerning the gender of the data collectors, it does not affect the behavior of a data collector completing campaigns. For both female and male data collectors, a general order of  $M_2 > M_1 > M_3 > M_4$  can be established, although we observe difference between the different incentive mechanisms when ranking. In the

mechanism ordering for male data collectors, the differences among the incentive mechanisms are not statistically significant. For female data collectors, the three incentive mechanisms  $M_1$ ,  $M_2$  and  $M_3$  are closer together, with  $M_4$  performing significantly least effectively. Considering age, and using a cut-off age of 25, we could find the differences in the younger and older data collectors' behavior of completing a campaign. Firstly, younger data collectors significantly submit more results than older peers, and we found a negative correlation between average number of submissions and age. Secondly, although younger and older data collectors share the same general order of the mechanism ( $M_2 > M_1 > M_3 > M_4$ ), younger data collectors prefer most mechanism  $M_2$ , then  $M_1$ . Mechanism  $M_3$  ranks further below the first two mechanisms, and at the bottom of the order is  $M_4$  with a large difference to all other mechanisms. In contrast, the older data collectors are in favor of any kind of incentive mechanisms, hence the differences are not significant among  $M_1$ ,  $M_2$  and  $M_3$  and even  $M_4$ ; however  $M_4$  is clearly at the bottom of the order with the smallest average number of submissions. We hereby note that the results for older data collectors should be interpreted with caution, as the total amount of older data collectors was relatively small; repeated experiments should confirm these findings. Figure 5.8 provides an overview of the ranking of incentive mechanisms, and their relative differences, for the whole population, and different considered subgroups.

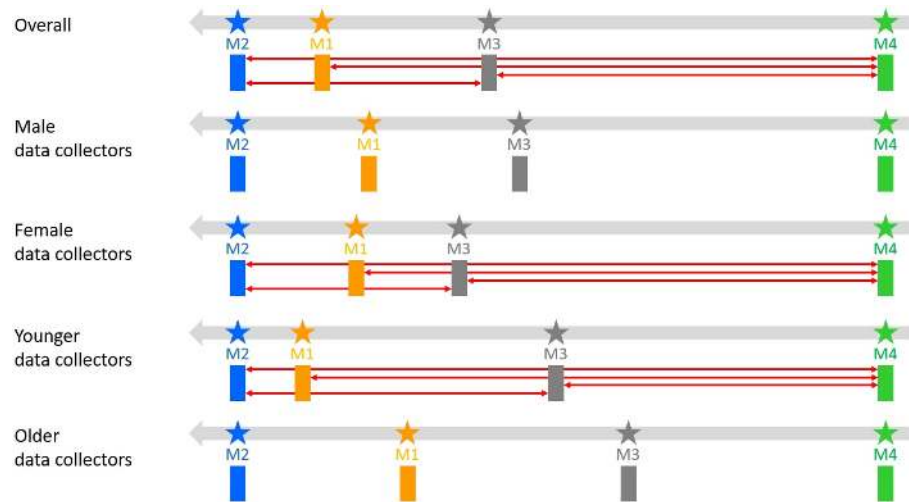


Figure 5.8: The approximate order of the incentive mechanisms, for the overall data collectors and their subgroups based on number of submissions (significance level of 0.05). The significant differences between the mechanisms are marked by red arrows.

As mentioned earlier, the statistical analysis performed in this thesis might suffer from the small sample size obtained from the UJI deployment. Due to this small sample, the conclusions might be vulnerable to noise (e.g., outlier) caused intentionally or unintentionally by (some of the) data collectors. As a result, and even though the presented conclusions are valid for the given population, they may have limited value for generalization. In order to remedy the negative effect of this small sample, it is recommended to replicate this deployment in a larger scope or in different scenario and/or on different set of data collectors. Subsequently, this statistical analysis will be performed again to further support or refute the conclusions reached in this thesis.

In general, data collectors seem to choose the incentive mechanism that gives them concrete, tangible and predictable rewards rather than the incentive mechanism that has some uncertainties (fixed versus variable micro-payment), the possibility of a high payout (lottery) does seem to be appealing as well. This empirical finding is consistent with a qualitative evaluation performed through campaign  $C_7$ , which had a question that directly asked data collectors which incentive mechanism they prefer after they have been exposed to all the different mechanisms. The majority of respondents (62.5%) chose fixed micro-payment (mechanism



$M_2$ ), 29.16% said that any reward is fine, 4.16% chose variable micro-payment (mechanism  $M_3$ ) and 4.16% chose lottery-style mechanism (mechanism  $M_1$ ). Nevertheless, we need to note that only 13.5% of the total amount of participants completed  $C_7$ , and thus the results should be interpreted with caution. More generally, the participation rate decreased toward the end of the deployment, as evidenced by the reduced participation to the campaigns launched toward the end of the deployment ( $C_5$ ,  $C_6$  and  $C_7$ ). The prolonged participation of data collectors is indeed a challenge for participatory sensing campaign organizers, which should be mitigated (e.g., by continuous promotional campaigns). It is hereby interesting to mention that we observed a slightly increased interest in campaigns (opening and completing) immediately after a mobile phone notification was sent regarding the availability of new campaigns (two times throughout the deployment).

## 5.6 Conclusions on the effectiveness of monetary incentive in participatory sensing

This deployment helped to study how the behavior of participants of a participatory sensing framework is influenced by the use of different monetary incentive mechanisms. In particular, through our Citizense participatory sensing framework, a variety of realistic participatory sensing tasks were deployed, using three monetary incentives: fixed micro-payment, variable micro-payment, and lottery style payout. In addition, a fourth control mechanism, no incentive/payout, was used. Fairness is a key characteristic in the incentive mechanisms: the mechanisms depend on criteria that are theoretically fair to all participants. These mechanisms were thoroughly tested in a real-world environment (a university campus), over a 20-day time period and with 230 data collectors.

By looking at the overall results, we see that the incentive mechanisms significantly increase the participation of the data collectors, while the framework maintains the quality of the submitted results: data collectors opened campaigns significantly more times and submitted more results when rewards through an incentive were available. Among the three implemented mechanisms, the statistical analysis shows that fixed micro-payment incentive (mechanism  $M_2$ ) outperforms variable micro-payment. A qualitative evaluation, even though performed on a small portion of participants, confirmed that the majority of data collectors prefers a

fixed micro-payment, over any other incentives. However, even though no statistically significant difference between the lottery-style payment latter and fixed/variable micro-payment could be found, the overall statistical analysis suggests an global decreasing performance ranking as follows: fixed micro-payment > lottery-style payout > variable micro-payment > no incentive. No difference in results was found when including or excluding weekends, showing that changed daily schedule and activities do not influence the overall results.

When considering different subgroups, according to demographical parameters (gender and age), the same performance ranking among incentive mechanisms could be found. However, different internal distances between individual incentive mechanisms have been observed. Most remarkably, male data collectors seem to only slightly prefer variable micro-payment over no incentive (no significant difference), and for older data collectors (>25), no significant difference could be found between any of the three incentive mechanisms, while they are all still significantly preferred to no-incentive mechanism. The latter however needs to be confirmed in repeated experiments, due to the smaller sample size of older data collectors. Generally, we also observed that there is a negative correlation between age and average amount of submissions: the older the participants, the less they submit on average.

Finally, based on the experience running our experiment, and the results obtained, we formulate three advices for participatory sensing campaign organizers: (1) continuous promotion is essential to attract participants and keep them interested; (2) monetary incentive mechanisms work better than not using an incentive, independent of type of incentive (3) we recommend predictable, tangible rewards (fixed micro-payment), rather than variable unpredictable rewards; a lottery-style reward may be an alternative.

This research performed a detailed, large-scale empirical study into the behavior of data collectors in the presence of monetary incentive mechanisms in participatory sensing. While this study investigated the overall trends, and studied particular subgroups based on gender and age, we call for further empirical research, first, to confirm and strengthen the results obtained to ensure generalizability, and second, focusing on other/additional relevant factors. In particular, in our study, we focused on one particular subjective reason to participate (i.e., monetary incentive), and minimized the influence of other (subjective and objective) factors. As such, this research leaves ample opportunity for future research, building upon

## 5.6. CONCLUSIONS OF THE EFFECTIVENESS OF MONETARY INCENTIVE<sup>111</sup>

and further refining the conclusions drawn from our large-scale empirical study into incentive mechanisms. In particular, the influence of demographic factors (e.g., age, level of income, cultural aspects, etc.), and the interplay between objective (e.g., costs to participate) and subjective reasons (e.g., subject of campaign, purpose, privacy, etc.) needs to be further studied.



## Chapter 6

# Temporal and spatial user behavior

This chapter seeks to study the spatial and temporal behavior of participatory sensing data collectors, when contributing data using a generic, multi-purpose participatory sensing framework. This study is expected to shed more light on how data collectors interact with the Citizense software tool in particular and how they behave in a participatory sensing context in general: when and where do participants contribute, how frequently are users contributing and are there differences among different contributors, are particular campaigns geographically or temporary dependent, etc. Such deeper insights open the door for various future improvements in participatory sensing, e.g., in enhancing data collectors' participation rate, for example through spatially or temporally targeted recruitment, promotional campaigns and/or incentive mechanisms; managing spatial and/or temporal spread of contributions.

Chapter 6.1 studies how frequently the data collectors interact with the mobile application throughout the deployment, using the timestamp of their request sent to the central server. Chapter 6.2 studies the daily submission pattern of data collectors during the whole deployment. Chapter 6.3 looks into the relation between the topic of a campaign and the location where data collectors make submissions for that campaign. Finally Chapter 6.4 presents the final conclusions. The discussion in this chapter is based on our previous publication (Khoi and Casteleyn, 2018).

## 6.1 Temporal user behavior: frequency of use

First, we are interested in studying the temporal interaction behavior of the user with the mobile application, or in other words, their frequency of use. In order to understand the performed analysis, consider the following typical usage pattern of the mobile application: at a certain moment in time, a data collector opens the mobile application, which sends a request for the list of current campaigns to the central server. Subsequently, the received list of campaigns is visualized on the user's mobile device. The user can now browse through the list and open a selected campaign to read its details. He then has two options: completing that campaign, after which he's again presented with the list of campaigns, or withdrawing from the selected campaign after reading its summary and returning to the list of campaigns. In either case, each time the user returns to the list of campaigns, it is refreshed by sending a request to the central server. This process might be repeated several times until the user finally exits the mobile application and comes back later. Therefore, based on the timestamp of the requests for the list of campaigns to the campaign server, and the time separation between them, we are able to determine usage sessions: periods of time in which the user is actively using the mobile application and potentially contributes data. Figure 6.1 exemplifies the behavior of a typical data collector (the timestamp of the requests) over time: certain requests are clustered in time, representing active, longer (usage) sessions, while others are more isolated, representing less active, shorter (usage) sessions.

We thus define a session as a cluster of requests coming from a single data collector, where the time difference between any two consecutive requests is at most  $x$  seconds. The delay  $x$  denotes a threshold high enough to include the time it takes to scroll through the list of downloaded campaigns, open a campaign to view its details and (possibly) complete it, plus any minor distraction the data collector may have (e.g., answering a text message, greeting a friend in the street). For our study, we select  $x$  to be 600 seconds (10 minutes) as this value allows data collectors to comfortably complete any available campaign, considering the average length of a campaign (i.e., at least 6 tasks) and the high-quality results produced by the data collectors (i.e., sharp pictures, detailed text). We then compute the number of sessions and the average number of request per session for every data collector. Due to the fact that the inequality in participation can

be found in various field (e.g., Wikipedia (Wilkinson and Huberman, 2007), VGI (Hochmair and Zielstra, 2015; Juhász and Hochmair, 2016), peer-to-peer file sharing (Hughes et al., 2005)), we expect to find a similar (uneven) distribution for the number of sessions among the data collectors in this deployment.



Figure 6.1: An illustrative example of requests sent by a typical data collector over time, each request is represented by a blue dot. A session is a cluster of requests, marked by the red ellipses.

Figure 6.2 shows the distribution of number of sessions and the average number of request per session for every data collector at the value  $x = 600$  seconds, decreasingly ordered according to the number of sessions. If we increase the value of  $x$ , the number of sessions decreases, as several sessions might form one bigger session due to the more tolerant waiting time between server requests. The most active data collector had 37 sessions ( $x = 600$  seconds) throughout the 20-day deployment, meaning that he/she had around 2 sessions per day on average. As we expected, Figure 6.2 clearly shows an uneven distribution of number of sessions among the data collectors. Specifically, this distribution follows a long-tail distribution (Anderson, 2006), with a smaller number of active data collectors having a large number of sessions and a much larger number of less active data collectors having a much smaller number of sessions (Hinz et al., 2011; Pan and Li, 2011). Specifically, while the most active data collector had 37 sessions, 227 of 230 data collectors (98.69%) had less than 30 sessions, 222 data collectors (96.52%) had less than 20 sessions, 201 data collectors (87.39%) had less than 10 sessions and 159 data collectors (69.13%) had at most 5 sessions. Such a distribution of performance is common for this type of participatory and voluntary works (Geiger and Halfaker, 2013). Among the number of average requests per session, there are some outliers. Furthermore, it can be observed that the average number of requests per session is slightly lower, and there is lower variation, among data collectors having more sessions (the smaller group of active data collectors, on the left part of Figure 6.2) than that of the data collectors having fewer sessions (the bigger group of regular data collectors, on the right part of Figure 6.2); most of the outliers are from these regular data collectors, which also brings the average number of requests per session slightly up. This

observation suggests two different behaviors of the data collectors: the regular data collectors interacted with the Citizense mobile application a few times, and each time they tried to discover (and complete) a larger number of campaigns (in some cases a very large number of campaigns) and later left the application for a longer time, while their active peers opened the mobile application more frequently, and each time selected (and completed) a smaller number of campaigns.

Additionally, this observation is confirmed by some of the interviewed data collectors after the deployment. 3 data collectors confirmed that they frequently opened the Citizense mobile application while 10 other acknowledged that they interacted with the Citizense mobile application just a few times, where each time they tried to explore several campaigns. Our further analysis has shown that in general, the total number of submissions made by an active data collector is larger than that of a regular data collector. To put this in perspective, note that a related study (Wang et al., 2015) gauged the number of visits to social network platforms among college students with different academic backgrounds. They concluded that the users made between 6 to 60 such visit per day, compared to 2 sessions per day for the most active data collector day for Citizense.

However, the 90-9-1 rule (Carron-Arthur et al., 2014) highlighting the participation inequality does not hold in this case. If data collectors are sorted by their number of submissions, those who are in the first percentile (1% - 3 data collectors) have made 343 submissions (11.65% of total number of submissions) while the next group of data collectors (9% - 21 data collectors) have made 1210 submissions (41.14% of total number of submissions) and the rest of data collectors (206 data collectors) have made 1391 submissions (47.24% of total number of submissions), which is in sharp contrast with the characteristic of the 90-9-1 rule stating that the group of 1% (creators), 9% (contributors) and 90% (lurkers) of the population contribute 90%, 9% and 1% of the content, respectively. A possible explanation for this contrast is that the incentives helped to increase the participation of the would-be lurkers, and to a lesser extent the contributors. Specifically, due to the extremely large number of the lurkers, a small increase in the number of submissions of each would-be lurkers results in a significant increase in their total number of submissions and this total outnumber that of the creators.



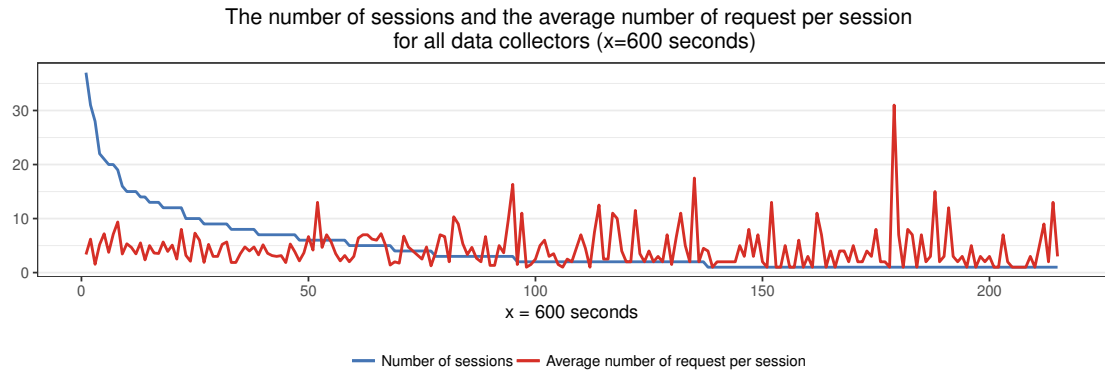


Figure 6.2: The number of sessions (blue) and average number of request per session (red) for every data collector throughout the deployment considering a separation time  $x = 600$  seconds, decreasingly ranked according to the number of sessions.

## 6.2 Temporal user behavior: distribution of use

Based on the timestamp of the requests and the submissions from the data collectors, we seek to answer the question: when are the participants interacting with the CitizenSense framework? We expect to find the participants' overall submission behaviors in terms of time. We further speculate there may be a relation between the content of a campaign and the temporal distribution of its corresponding submissions. From our analysis, we first note that the number of interactions (requests and submissions) in the weekend is insignificant compared with the total number of interactions (weekend requests and submissions constitute 5% and 4%, respectively). It is thus clear, perhaps somehow surprisingly, that data collectors are mainly active during the working days, and much less during weekends. From Figure 6.3, it is clear that the distribution of the data collectors' requests generally matches the human work-rest cycle (which is consistent with the findings of Li et al. (2018)), although several requests occurred late at night. A distinct peak around 19h00 can be identified, which corresponds to the hour work/classes generally finish, and the request intensity is generally higher in the evening (19h00 to 24h00) than during daytime (9h00 to 18h00). This observation suggests that data collectors interact more with the mobile application in the evening, during their free time, although a reasonable amount of requests is also made during the (working) day. As a submission would likely follow a request, the distribution

of the submissions follows the same pattern, with a distinct peak around 19h00 and more submissions delivered in the evening than during daytime. Overall, the submission intensity is relatively stable during mid-day, then slightly decreases in the afternoon and increase again in the evening. It reaches two peaks in the evening before gradually dropping to 0 when data collectors are sleeping.

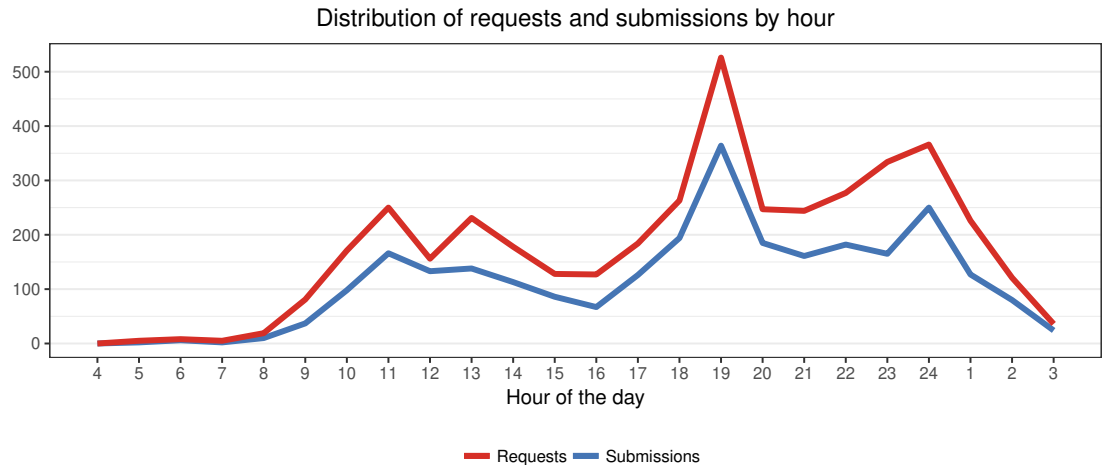


Figure 6.3: The distribution of all received requests and submissions throughout the deployment rendered in a 24-hour time frame.

A closer look into the temporal pattern of the submissions from different campaigns provides deeper insight in the data collectors' submission intensities. Indeed, they generally follow the overall pattern: submission intensities are higher in the evening than any other time of the day, with peaks occurring from 19h00 to 22h00. However, some campaigns have a different temporal pattern. In particular, one observed deviating pattern is that a primary peak occurs around mid-day, not in the evening (see Figure 6.4, upper), and a secondary peak during the evening. Interestingly, the campaigns whose submission patterns differ from the overall submission pattern discuss several aspects of the transportation such as the regional train service, the city's tram line, street crossings, the railway station and the transport efficiency. In contrast, this phenomenon is less visible in other campaigns whose topics are not related to the transportation theme (see Figure 6.4, lower) as they generally have a primary peak and more submissions in the evening. This difference might suggest that responding to campaigns related to the data collectors' current or recent activity and context (e.g., using public transport in

their commute to/from the university campus in the morning, noon (lunch), evening) might be triggered by the data collectors performing the activity, rather than in during fixed time periods (i.e., their free time in the evening). We further explore this hypothesis when considering the data collectors' spatial behavior in the next subsection.

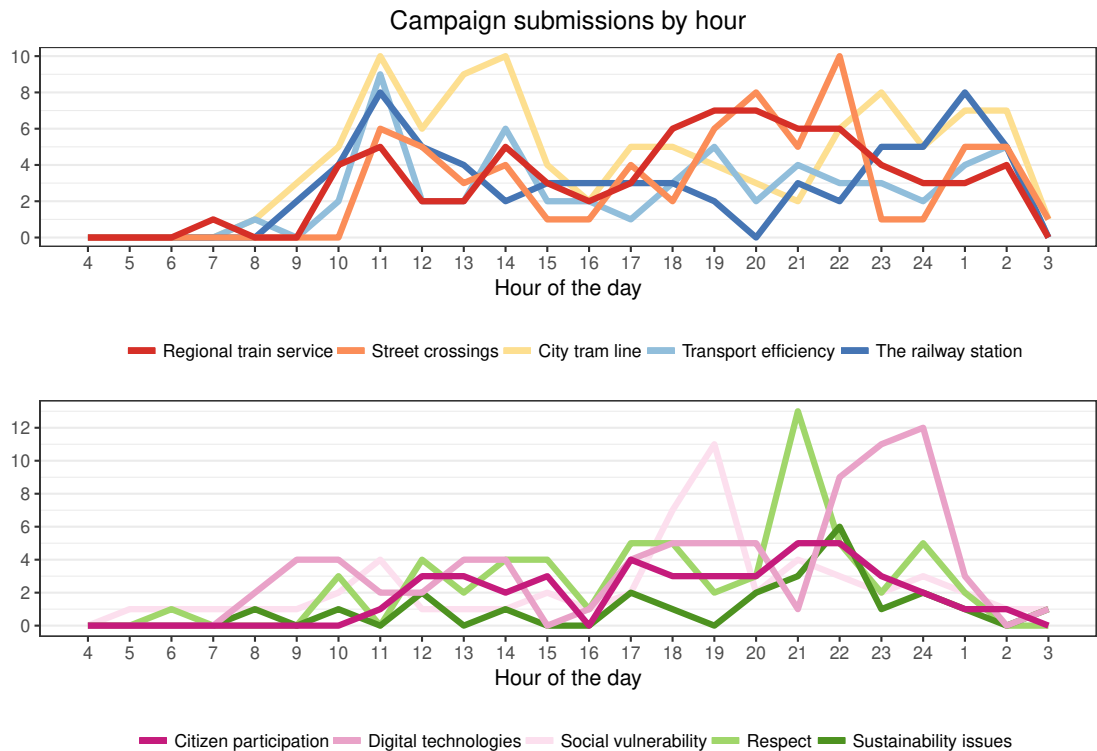


Figure 6.4: The distribution of submissions from different campaigns throughout the deployment rendered in a 24-hour time frame. (upper) The campaigns having different patterns than the overall submission pattern. (lower) The campaigns having a similar pattern to the overall submission pattern.

### 6.3 Spatial user behavior

As the data collectors live across the city (and the province) and commute to the university on a daily basis, their requests and submissions are spread across the city of Castellón, the nearby towns and the routes connecting them (i.e., railways and highways). The biggest cluster of submissions is in the city of

Castellón, as many campaigns addressed issues of the city on one hand, and data collectors spent much of their time in the city on the other hand. Therefore, we focus our attention to this cluster and study the relation between the locations of a campaign's submissions and the topic of that campaign.

With respect to spatial properties, the 44 campaigns available during the deployment can be divided into 2 groups: geographically-constrained campaigns and campaigns not geographically-constrained (see Figure 6.5). The former are campaigns having location constraint(s) defined by the campaign authors through the Citizense campaign manager (for an example, see Figure 3.4 and Section 3.3.4): a data collector can only download (i.e., each time he refreshes the list of available campaigns) and submit results for these campaigns if he complies with the defined location constraint(s) (i.e., he is inside the specified area(s)). In contrast, the latter are campaigns which do not have such location constraint(s): data collectors can download them and submit the results regardless of their current location. These campaigns which are not geographically constrained can be further divided into 2 sub-groups (see Figure 6.5): campaigns on a topic having a clear geographic boundary (e.g., a campaign on the city's only tram line, which has a fixed defined route), later referred to as geographically-related campaigns; and campaigns on a topic that does not have a geographic boundary (e.g., a campaign on the issue of citizen participation or the use of digital technologies), later referred to as geographically-neutral campaigns. Based on this classification, there are 3 geographically-constrained campaigns, 13 geographically-related campaigns and 28 geographically-neutral campaigns (see Table B.1).

As campaigns geographically constrained by the Citizense framework restrict responses to originate from specific area(s), they are not suitable to study the spatial distribution of responses. On the other hand, campaigns that are not geographically constrained leave participants free to submit responses anywhere in or outside the city; they are thus suitable to study the spatial distribution of responses in order to detect possible patterns. Due to the variety of the campaign and their content, we expect to see the different spatial distributions of the submissions from different campaigns. Specifically, we suspect a certain amount of submissions to fall within the boundaries of the geographically-related campaigns, reflecting the fact that the data collectors' current context (i.e., location) influences his participation behavior. We thus compare the spatial distributions of submissions from geographically-related campaigns to those of geographically-neutral campaigns.

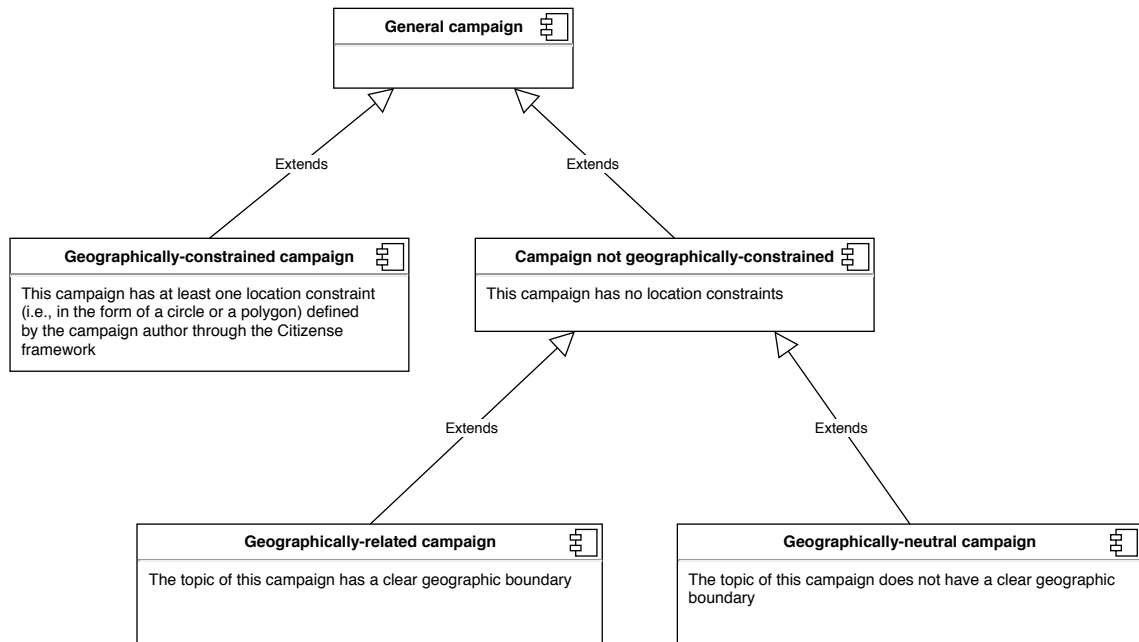


Figure 6.5: The taxonomy of campaigns used in the Citizense deployment.

To do so, we first formally define the geographic boundaries of the former campaigns. We then propose the on-target ratio as the number of submissions that fall within the boundary of a geographically-related campaign. Finally we compare the on-target ratio among the different geographically-related campaigns.

The boundary of a geographically-related campaign is the buffers surrounding the basic geographic shapes such as a point, polyline and polygon, depending on the topic of the campaign in question. For example, the campaign "City tram line" about Castellón's only tram line has its buffer in the form of a polygon surrounding a polyline, which represents the tram route. Similarly, the campaign "Regional train service" has its buffer in the form of polygon and circles surrounding the polyline and points which represent the railroad and the stations, respectively. The campaigns "University campus" and "Illegal graffiti" have the buffers encompassing the polygons representing the university campus and the city border, respectively. For our analysis, we select the initial buffer distance  $d = 100\text{m}$  and then we then gradually increase it to 200m and 400m (see Figure 6.6, right). There are several reasons behind this selection. First, we note that the embedded GPS receivers may have an error of 5 meters in ideal conditions and cell tower triangulation (used in low power mode in Android or when GPS is not available) often has

accuracy of a few hundreds of meters; these figures probably increase in less favorable conditions. Consequently, we need values for the buffer which are significantly bigger in order to compensate for these errors. Second, we seek to capture the submissions close to the geographic context of the campaign (i.e., the tram line, the station, the university campus) as these submissions are likely made due to the data collectors' preceding activities (e.g., commuting, taking classes in the university), hence the distances of 200m and 400m are reasonable (e.g., capturing users who perform the campaign while walking away or being driven away from the tram stop or train station). Third, taking into account the length of a typical campaign (i.e., 6 to 8 tasks), the required time to complete such campaign is around 1 minute; this time can be longer if the data collectors elaborate their answers and many data collectors did so (for an example, see Figure 4.1, upper). Combined with the different transportation modes of the data collectors (i.e., walking, personal car, public bus), the distances of 200m and 400m give data collectors sufficient time to complete these campaigns. Finally, the value of 400 meters is a typical distance for the catchment used in transportation engineering (Daniels and Mulley, 2013).

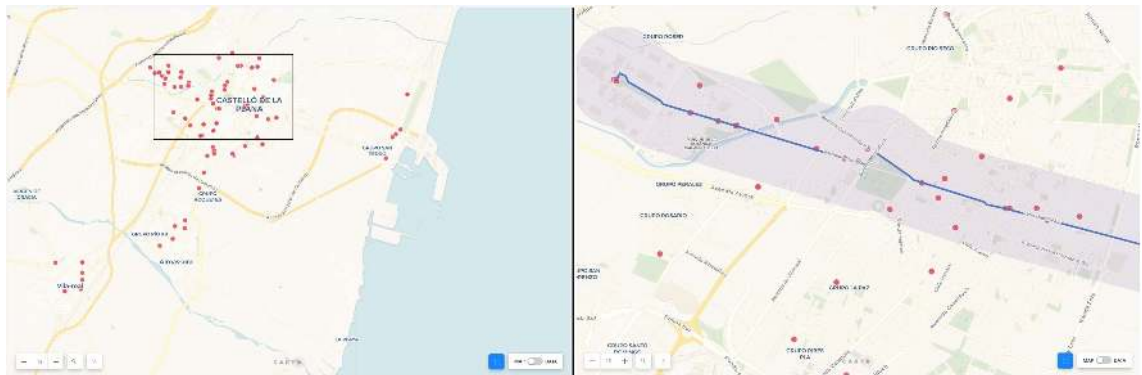


Figure 6.6: The data collectors' spatial pattern in different campaigns. (left) submissions from campaign 1 (geographically-neutral) addressing the topic of citizen participation, with the black rectangular frame covering the map extent on the right side. (right) submissions from campaign 14 (geographically-related) asking feedback on the city's only tram line, with the blue line representing the tram route and the purple area representing the buffer of 400 meters.

From visual inspection, we have identified a difference in the spatial distribution of the submissions between geographically-related campaigns and geographically-

neutral campaigns. Specifically, it is observed that the former have many of their submissions falling in the buffer related to the campaigns' topics; this observation will be quantified using the on-target ratio (introduced later in the text). For example, the campaign "City tram line" (campaign 14 in Table B.1) has several submissions in the vicinity of the tram route (this map extent includes 34 submissions, equivalent to 94.44% of all location-enabled submissions from this campaign, see Figure 6.6, right); coupled with the submissions' timestamps, it can be inferred that these submissions were made when the corresponding data collectors were using or recently used the tram service. This finding is consistent with the data collectors' temporal patterns, which showed peaks during typical commute times (see Section 6.2). In contrast, the campaign "Citizen participation" is geographically-neutral, as its topic does not have a clear geographic boundary, and shows submissions spread over the city (see Figure 6.6, left) and the region.

To quantify this result, we propose a campaign's on-target ratio as the ratio between the number of submissions falling in the campaign's buffer according to distance  $d$ , and the total number of submissions (with GPS location) of that campaign. This ratio shows how much the submissions are geographically concentrated in the campaign's buffer, or in other words, in the geographical neighborhood of the topic of the campaign. Table 6.1 details the on-target ratios of the 4 geographically-related campaigns (i.e., city tram line, regional train service, the intermodal station, university campus), with the different buffer distances  $d = 100\text{m}$ ,  $200\text{m}$  and  $400\text{m}$ . These campaigns were selected among the 13 geographically-related campaigns (see Table B.1) as their buffers cover a small area of the city or the region (in contrast to the other campaigns which cover the whole city of Castellón), and thus they allow the study of data collectors' spatial behavior with relatively high accuracy. For comparison purposes, we include the geographically-neutral campaign "Respect" (campaign 12 in Table B.1) for each geographically-related campaign, computing the on-target ratio with the same geographical buffer as the corresponding geographically-related campaigns (see Table 6.1). The "Respect" campaign was chosen because it has a similar number of submissions and duration (in terms of days) compared with the 4 geographically-related campaigns in Table 6.1.

Table 6.1 suggests that in general, a significant part of the data collectors tend to answer geographically-related campaigns when they are within the campaigns' buffer areas, even if these areas are small compared to the whole city or region.

Table 6.1: The on-target ratio of the geographically-related campaigns with different values of buffer distance  $d$ .

Campaign name	Campaign content	con-	Geographic buffer	Buffer			Catchment
				100m	200m	400m	400m
City tram line	The citizens' opinions on the tram service		Polyline - The tram route	52.77%	55.55%	61.11%	52.77%
Respect	How people respect each other		Neutral	7.69%	7.69%	7.69%	7.69%
Regional train service	Feedback on the current regional train service		Polyline - The railroad network and Points - The stations	45.90%	47.54%	54.09%	22.95%
Respect	How people respect each other		Neutral	4.61%	4.61%	4.61%	1.53%
The inter-modal station	Opinion on the current state of the station and ways to improve it		Point - The station	36.11%	41.66%	47.22%	44.44%
Respect	How people respect each other		Neutral	0%	0%	0%	0%
University campus	How to improve the campus's infrastructure		Polygon - The university campus	45.09%	50.98%	50.98%	n/a
Respect	How people respect each other		Neutral	6.55%	6.55%	6.55%	n/a

Remarkably, even when considering a buffer distance of only 100 meters, a high percentage of submissions already falls within the buffer area; subsequently doubling the buffer distance (i.e., from  $d = 100$  to 200, and from  $d = 200$  to 400) in comparison only slightly increases the on-target ratios (see Table 6.1). This denotes that a significant amount of people submit results in an area close by ( $d = 100\text{m}$ ) the topic of geographically-related campaigns, and the effect decreases the farther away. In comparison, the corresponding on-target ratio of the "Respect" campaign, which is geographically neutral, is very low, which shows that the



found spatial correlation for geographically-related campaigns is not a general effect. In other words, the submissions of a geographically-neutral campaign are spread across the city (or the province); they do not cluster around any specific geographic feature. Evidently, if the buffer distance is increased by large amounts (e.g., a few kilometers, the whole city), the on-target ratios for geographically-related campaigns increases significantly. However, we argue that this effect is no longer due to the topic of the geographically-related campaign, but simply because a larger area of the city/region is covered. Table 6.1 also suggests that for geographically-related campaigns, the on-target ratio is relatively high (in many cases over 50%), for all different topic addressed. Therefore, it can be inferred that the decision to open and complete a campaign can be influenced by the proximity to (the buffer of) the topic of the corresponding campaign.

Next to using buffers, the on-target ratio can also be computed using the concept of catchment, which is widely used in the field of transportation engineering (Andersen and Landex, 2008). For example, in the case of the "City tram line" campaign, the catchment-based on-target ratio is defined as the ratio between the number of submissions falling in the catchment of 400 meters (see Figure 6.7, left), which is a typical value for transit stop (Daniels and Mulley, 2013), and the total number of submissions (with GPS location) of that campaign. Consequently, this catchment-based on-target ratio is 52.77%, which is slightly lower than the on-target ratios using the corresponding tram route's buffer (see Table 6.1). For the sake of comparison, we also compute the on-target ratio of this campaign using the 400-meter circular buffer (Andersen and Landex, 2008) around the tram stations (see Figure 6.7, right) and get the value of 58.33%. Although slightly less outspoken than the corresponding values for the 400m buffer (see Table 6.1), both catchment-based and buffer-based on-target ratios support that a significant amount of data collectors have their decision to open and complete the "City tram line" campaign influenced by the proximity to the tram line.

For the campaigns "Regional train service" and "The intermodal station", we use the same value of 400 meters for the catchment as suggested in Daniels and Mulley (2013). The catchment-based on-target ratios for campaigns "Regional train service" and "The intermodal station" are 22.95% and 44.44 %. It is observed that for the campaign "Regional train service", its catchment-based on-target ratio drops considerably compared with the corresponding on-target ratio using the buffer of the railroad network as several submissions made while the train is

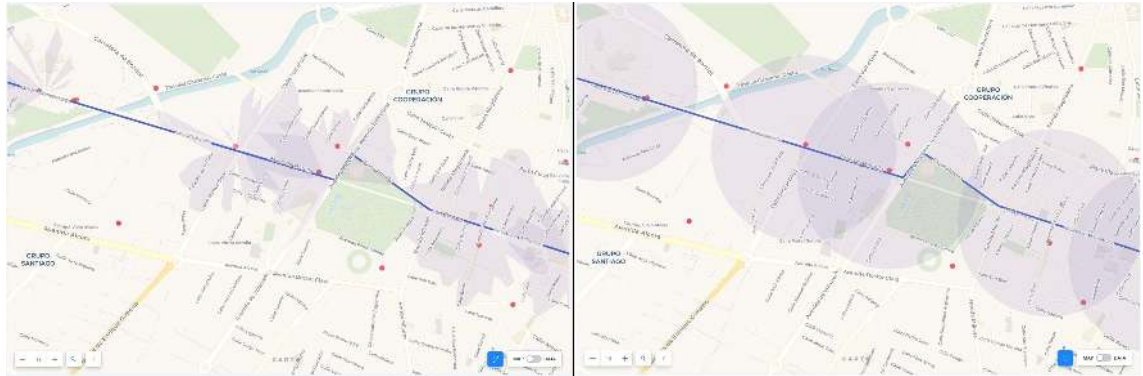


Figure 6.7: The submissions of the campaign "City tram line", represented by the red points. (left) The 400-meter catchments around the tram stations. (right) The 400-meter circular buffers around the tram stations.

moving (on the rail track and between the stations several kilometers away from each other) are excluded from the computation of the former. In conclusion, using the buffer-based and catchment-based on-target ratios and for all three considered transport-related campaigns, it can be concluded that a significant amount of data collectors are influenced in their decision to open and complete a campaign by the proximity to the geographical context of that campaign.

## 6.4 Conclusions on the user behavior in participatory sensing

Using Citizense, a real-world deployment with 359 participants, who have diverse demographical and educational backgrounds, was performed, in order to demonstrate Citizense meets its goals, and more importantly, to study the participants' spatial and temporal behavior when collecting data. For the former, the deployment showed that the framework's campaign editor effectively enabled campaign authors to intuitively design their desired participatory sensing campaigns that collect meaningful city-related information on one hand, and that the framework's mobile application effectively assisted data collectors in gathering various types of data through their mobile devices on the other hand. This result reaffirms the potentials, applicabilities and effectiveness of multi-purpose participatory sensing frameworks, with Citizense as an example, in collecting a large variety of data in different scenarios (Tangmunarunkit et al., 2015; Brunette et al., 2013).

#### 6.4. CONCLUSIONS ON THE USER BEHAVIOR IN PARTICIPATORY SENSING 127

Based on the results of campaigns receiving most interactions from data collectors, the analysis of the data collectors' spatial and temporal behavior revealed several patterns. First, and perhaps somewhat surprisingly, data collectors are much more active during weekdays compared to weekends. Second, data collectors as a whole tend to contribute data during their (evening) free time. However, for certain campaigns (e.g., campaigns related to public transportation), a different pattern emerges. Specifically, the submissions were made when data collectors found these campaigns relevant to their current context and activities: a considerable portion of the submissions were made during daytime when they are involved in the transportation activities. These observations provide new insights into the data collectors' temporal pattern, showing when they actively take part in their participatory sensing activities. Third, we discern two groups of data collectors: a smaller group of active data collectors, who regularly use the mobile application (the most active participants on average twice daily), and a larger number of regular data collectors, who use the mobile application less frequently to collect data. This uneven distribution of interactions in participatory sensing matches with the distribution of users' participation in other domains in which users voluntarily contribute their resources (Hughes et al., 2005), data and efforts (Hochmair and Zielstra, 2015; Juhász and Hochmair, 2016; Geiger and Halfaker, 2013). Furthermore, active data collectors perform less campaigns in each session, while regular data collectors perform more campaigns in a session. Overall, active data collectors collect more data. Fourth, when data collectors are not geographically confined when downloading campaigns and submitting their results, their submissions are spread across the city (or the province). However, if a campaign has a geographical context (i.e., a topic with clear geographic boundaries), it is observed that a significant part of that campaign's submissions (more than 50% in many cases) are within the buffer of that campaign's geographical context, even when the buffer is relatively small and data collectors are geographically free to submit their data anywhere. This suggests that a substantial part of the data collectors are bound to the geographical context of a campaign; data collectors complete these campaigns when the campaigns' contexts are geographically relevant to them. The observed behavior provides new understandings in how data collectors spatially behave in participatory sensing, which is under-investigated in the current state of the art. Interestingly, from the initial 100-meter buffer, the on-target ratios (submissions within the buffer w.r.t. all submissions) of such campaigns do not

increase significantly with increasing buffer size (unless the buffer is enlarged considerably), denoting that the geographical context effect is strongest close by the topic of the campaign and weakens the further away. These observations hold for several geographically-related campaigns, regardless of their topic and contents.

# Chapter 7

## Conclusions

As the urban population is growing steadily, cities are becoming larger in size and more complex in terms of physical and social structure. Naturally, there is a need to collect information from the surrounding environment and understand these information. This need is further fueled by the Smart city concept, which aims to provide a sustainable, safe and liveable environment for all the inhabitants. At the center of this Smart city concept is the demand for different types of urban information, which can be effectively collected via participatory sensing as this approach leverages the mobility and intelligence of the people and the computing power of their devices. Although participatory has achieved initial successes, which are reflected in several successful projects and the large number of participatory sensing applications and framework, there are still rooms for improvements.

This thesis is dedicated to the design and development of Citizense, a multi-purpose participatory sensing framework. The framework stemmed from an extensive review of literature in the field of participatory sensing, after which a list of functional and non-functional requirements is formulated. In other words, Citizense combines and extends features and functionalities of existing participatory sensing applications and frameworks. The aforementioned requirements are incorporated in the implementation of Citizense, in which each requirement is realized into a functionality or component of this framework.

Upon the completion of the Citizense implementation, the framework was deployed in two real-world deployments in which hundreds of users interact with different components of the framework. In these deployments, Citizense has showcased its designed features and functionalities. Specifically, it allowed the

creation of a variety of participatory sensing campaigns by ordinary users and the collection of different types of urban information. At the same time, Citizense received approval from the local government to be used in official events of the city and the appreciation from hundreds of users.

While it is known that monetary incentives generally help to motivate participants in participatory sensing as they compensate various physical and psychological costs associated in the data collection process, there is still a need to study the effectiveness of monetary incentives in closer details. Through the Citizense deployment in the UJI campus, it was possible to study the participants' reaction to different monetary incentive mechanisms. Finally, with a confidence level of 95%, this study presents a general order of preference among the experimented monetary incentive mechanisms. This order is also found in certain subgroups of data collectors grouped by demographic parameters (e.g., young data collectors, female data collectors). However, due to the relatively small size of the studied sample, the generalization of this conclusion should be taken with care. It is therefore strongly recommended to repeat the deployment of Citizense in order to both strengthen the conclusions and to further clarify the differences in certain subgroups (e.g., male data collectors, older data collectors) when the Benjamini–Hochberg procedure is used.

As a large number of users interacted with the Citizense framework, they exhibited certain spatial and temporal patterns. The Citizense deployment allows the study of where, when and how often data collectors interact with the Citizense mobile application. Due to the fact that more insights on user behavior in participatory sensing are obtained, it is possible to make the participatory sensing process more effective and/or improve the existing participatory sensing applications and frameworks.

## 7.1 Answers to the research questions

In the introductory chapter, the following three research questions were formulated in order to provide an overview about the work covered in the course of this thesis:

1. RQ1 - What are the necessary characteristics and functionalities of a participatory sensing framework?

2. RQ2 - How do participants behave in a participatory sensing context?
3. RQ3 - How can organizers motivate the participants to contribute in participatory sensing?

With respect to question RQ1, Chapter 2 first presented the diversity of the set of participatory sensing applications and frameworks. It then provided a comparison between a variety of participatory sensing applications and frameworks with the purpose of highlighting their strengths and weaknesses. At this stage, a list of 7 functional requirements and 2 non-functional requirements was formulated, which served as the foundation for the creation of the Citizense framework. Chapter 3 detailed the design and implementation of the Citizense framework and its components, in which all the aforementioned requirements were incorporated. Specifically, in Chapter 3, each functional and non-functional requirement is confirmed and highlighted by pointing out the component/functionality of the Citizense framework that realizes this requirement. In summary, the main contribution of Chapter 2 was the set of requirements for a participatory sensing framework, which is verified by the detailed description of the Citizense framework in Chapter 3.

In order to further verify the design and different functionalities of the Citizense framework, Chapter 4 details the two real-world deployments of Citizense. The first deployment was conducted in agreement with the city hall of Castellón, in which Citizense was used to collect the feedback from 140 marathon participants on various aspects of the marathon. This deployment is a milestone in the development of the Citizense framework as for the first time, several functionalities of the framework were verified in real-world conditions and more importantly, the framework received approval from the local authority to be launched during a public event and appreciation from the marathon's organizers for its efficiency in collecting data. In the second deployment, Citizense framework was deployed in the campus of UJI where it was used by 359 participants of different educational backgrounds. This deployment serves three purposes. Firstly, it examines the framework's effectiveness in collecting various types of urban information. Secondly, it allows the authors to study the reaction of data collectors to the monetary incentives. Thirdly, due to the large number of interactions (i.e., requests, submissions) from data collectors, the study of user behavior was made possible. The main contribution of Chapter 4 is the detailed description of the Citizense deployment in UJI, which creates the foundation for the answers to questions RQ2 and RQ3 in Chapter 5

and 6.

In order to answer research question RQ2, Chapter 6 discussed the data collectors' spatial and temporal behavior using the large number of interactions received from the UJI deployment. Specifically, the frequency of use and the distribution of use were built through the analysis of the timestamp of the request and the submissions, respectively. It is shown that in terms of frequency of use, there is a small group of active data collectors who regularly use the Citizense mobile application and a larger group of regular data collectors, who use the mobile application less frequently to collect data. In terms of distribution of use, data collectors tend to contribute data during their (evening) free time. However, certain campaigns (e.g., campaigns related to public transportation) attract the attention from data collectors when they found these campaigns relevant to their current context and activities. Finally, it is observed that data collectors tend to make their data contribution close to the campaign's geographical boundary, even though they are geographically free to submit those results at any location. Based on these observations and conclusions, several suggestions have been made to improve the participatory sensing process.

Regarding research question RQ3, Chapter 5 allowed the study of data collectors' reaction to the different monetary incentive mechanisms. Specifically, through the 7 campaigns applied with monetary incentives, a general order of preference among the different monetary incentive mechanisms can be suggested (at the confidence level of 95%): fixed micro-payment > lottery-style payout > variable micro-payment > no incentive. In general, this order does not change across different clusters of data collectors such as female and male data collectors or young and old data collectors.

## 7.2 Contributions of the thesis

The main contributions of this thesis were (1) the creation of Citizense, a multi-purpose participatory sensing framework which combines and extends features of existing participatory sensing applications and frameworks and (2) the study of different aspects of user behavior in the context of participatory sensing. The Citizense framework has shown to be an effective tool to collect a variety of data in the urban settings; at the same time it received appreciation from different stakeholders. Using the Citizense framework as a vehicle, a deployment of



Citizensense in real-world scenario allowed the study of different aspects of user behavior as data collectors were simultaneously exposed to a large number of participatory sensing campaigns. This deployment has shed light on how data collectors execute the campaigns spatially and temporally. These findings have extended the participatory sensing body of knowledge, especially on the topic of user behavior which is somehow under-investigated. This deployment also provided empirical insights on how different monetary incentive mechanisms motivate data collectors in their participatory sensing process.

## 7.3 Results of the thesis

Apart from the thesis, which compiles and summarizes the main findings throughout the Ph.D program, the following results have been achieved:

- **Journals with Impact Factor:**

1. **Khoi, Ngo M.**; Casteleyn, Sven; Moradi, M. M.; Pebesma, Edzer. 2018. "Do Monetary Incentives Influence Users' Behavior in Participatory Sensing?" *Sensors* 18, no. 5: 1426.
2. **Khoi, Ngo M.**; Casteleyn, Sven. 2018. "Analyzing Spatial and Temporal User Behavior in Participatory Sensing." *ISPRS Int. J. Geo-Inf.* 7, no. 9: 344.
3. **Khoi, Ngo Manh**, Sven Casteleyn and Mijail Naranjo-Zolotov. "Participatory sensing to empower citizens in smart cities". *Submitted to IEEE Computer Magazine, 2018.*

- **Conferences:**

1. **Khoi, Ngo Manh**, Luis E. Rodríguez-Pupo, and Sven Casteleyn. "Citizensense—A generic user-oriented participatory sensing framework." In *Selected Topics in Mobile and Wireless Networking (MoWNeT)*, 2017 International Conference on, pp. 1-8. IEEE, 2017.

- **Posters:**

1. The deployment of a generic user-oriented participatory sensing framework in a real-world context, AGILE 2018, Lund 12-15 June 2018, Sweden.

2. An overview of Citizense A generic user-oriented participatory sensing framework, AGILE 2017, Wageningen 9-12 May 2017, The Netherlands.

- **Stay:**

1. External semester at Institut für Geoinformatik, Universität Münster under the supervision of Prof. Dr. Edzer Pebesma, from 1 September 2017 to 28 February 2018.

- **Internship:**

1. Internship at the Castellón city hall, from 21 November 2016 to 20 December 2016, under the supervision of Juan Millan del Sol.

- **Software:**

1. Citizense web-base campaign creator: [www.citizense.uji.es](http://www.citizense.uji.es)
2. Citizense mobile application:  
[https://play.google.com/store/apps/details?id=geoc.esr06.citizense\\_munster](https://play.google.com/store/apps/details?id=geoc.esr06.citizense_munster)

- **Dataset:**

1. Collected data from the UJI deployment:

[http://www.citizense.uji.es:8880/gateway?type=campaign&campaign=\[CampaignID\]](http://www.citizense.uji.es:8880/gateway?type=campaign&campaign=[CampaignID])

2. Visualization of the collected results from the UJI deployment:

<http://www.citizense.uji.es/publicResultViewerTest.html>

## 7.4 Limitations

Participatory sensing, although having the capabilities of collecting a variety of information, entirely depends on the participation of a sufficiently large number of participants. If there are not enough participants, the result from the participatory sensing process might be incomplete and/or biased. Although the experiments attracted a reasonably large number of participants (compared to the amounts we

usually see in participatory sensing literature, which are around 50), it is recognised that the sample size should be increased in order to generalise conclusions. Additionally, to strengthen our statistical analysis, and avoiding false positives or negatives (i.e., increase the possibility of detecting ), i.e. type I (rejecting the null hypothesis) and type II (not rejecting the null hypothesis) errors. Finally, our analysis studied a single effect (monetary incentives), while other effects might play as well. More advanced statistical analysis could better take this into account.

Regarding the Citizense framework itself, it is noted that, although Citizense is meant to be deployed on a city-wide scale, scalability and performance have not been assessed to match this scale. For the context of this Phd, and the experiments performed within it, this was the goal, and thus not a priority nor necessity, but obviously more rigorous performance testing is needed before a potential real-life deployment. Possibly, code refactoring might be needed to ensure scalability on city-wide scale.

Finally, even though Citizense is meant as a multi-purpose participatory sensing platform, and therefore incorporates most features of existing participatory sensing frameworks, it does not include all. Some features, such as use of external sensors and campaign orchestration, are found in other frameworks/applications, yet not (yet) included in Citizense. This should not be considered a major limitation, as the Citizense implementation is prepared to be extended to include such support, and it is foreseen in future work.

## 7.5 Outlook

Naturally, the study presented in this thesis can be further extended. As a result, this final section provides an overview on potential next steps in terms of development and methodology.

First of all, new components can be developed to broaden the range of the Citizense framework. For example, currently, the Citizense client application for answering campaigns is only available as an Android application. A logical next step would be the creation of an iPhone app. Also, a web-based application might increase the reach of Citizense. Next, the functionality of Citizense can be extended in various ways. As already mentioned in limitations, features such as extending built-in sensor support (e.g., light intensity, magnetometer), support for external sensors, campaign orchestration and more advanced visualisation

of results (e.g., facilitate this functionality such as measurement resolution, grid size (in the Campaign Editor) and the submission rate projection (in the central server), which is completely feasible due to the open and extendable nature of the Citizense framework) could be included in the framework. Next to such features, which to some extent are already present in other participatory sensing applications/frameworks, additional features not present in other applications/frameworks can be useful as well. For example, concerning the temporal constraint, a recurring campaign (i.e., a campaign that are repeatedly available according to a certain condition) can be a useful addition. In summary, the Citizense framework is extensible in a variety of ways, each of which will extend its functionality and usefulness.

Secondly, the identified participants' spatial and temporal behavior suggest several improvements in the design and implementation of the Citizense framework. For example, based on the spatial and temporal of the data collectors, each group of data collectors can have a tailored communication strategy. This tailored communication strategy can be implemented as customized time- and location-dependent notification sent to each data collector. Based on the data collector's past activities, the notification might suggest that data collector to perform sensing activities in time periods or locations in which his/her submission brings more values or that data collector will likely respond. Using this knowledge, new recruitment strategies can be proposed. For example, a recruitment that selects data collectors who are active in the last  $X$  days or data collectors having more than  $X$  sessions daily. These filters can be combined with the quality of submissions evaluated automatically or by campaign organizers to further improve the data quality. Similarly, developers can include time- and location-dependent incentives, which adjust the amount of incentive based on the temporal (e.g., striving for more submissions during off-peak times) and spatial desirableness (e.g., striving for full spatial coverage) of the submissions, respectively.

Thirdly, although the thesis has drawn several conclusions on the user behavior in the participatory sensing context, it is strongly recommended that the deployment of Citizense framework is repeated. First of all, to confirm and strengthen the results obtained, and ensure generalizability. Second, also in different places and/or with different participants, to further deepen the results obtained, and to discover the relation between the user behavior in participatory sensing and socio-economic parameters such as the participants' culture, income, age and

profession.

Regarding the accuracy of the sensory measurements, the sensor calibration on the mobile application is required as each mobile device has a heterogeneous set of sensors (e.g., microphone, temperature sensor, luminosity sensor, barometer sensor, humidity sensor). Among the physical sensors currently supported by the Citizense framework, more focus should be paid on the calibration of the sound sensor, as it is more vulnerable to human errors and environmental factors. The calibration process should at least cover the most popular device models, as these devices span across a larger part of the population. As the framework is designed to function in (populous) urban environments, it is required to be able to cope with a sudden increase in demand and constantly high demand. Therefore, the framework is planned to be deployed on elastic cloud computing platforms so that it can take advantage of these platforms' scalability. It is also planned to revise the source code of the Citizense central server so that it is able to support parallel execution.

Finally, participatory sensing can be looked at from the angle of e-Participation due to the shared characteristics such as the participation of a large number of participants, the involvement of the local authority. In this direction, it is desired to study how participatory sensing can contribute to different levels of e-Participation, namely e-Consulting and e-Empowerment. In this regard, participatory sensing might extend the types of data returned by the participants and give them more flexibility in defining the topic of discussion.



## Appendix A

### The campaign on users' feedback in the Castellón marathon

Table A.1: The questions on the usability of the Citizense mobile application

Name	Question	One	Two	Three	Four	Five
Q1	I think that I would like to use this system frequently.	10	6	19	10	7
Q2	I found the system unnecessarily complex.	10	9	25	3	5
Q3	I thought the system was easy to use.	4	2	24	8	14
Q4	I think that I would need the support of a technical person to be able to use this system.	23	9	17	2	5
Q5	I found the various functions in this system were well integrated.	4	0	22	16	10
Q6	I thought there was too much inconsistency in this system.	2	4	30	10	6
Q7	I would imagine that most people would learn to use this system very quickly.	2	2	20	14	14
Q8	I found the system very cumbersome to use.	11	7	25	3	6
Q9	I felt very confident using the system.	2	2	24	13	11
Q10	I needed to learn a lot of things before I could get going with this system.	9	13	24	3	3

Table A.2: The questions on the the concept of participatory sensing

<b>Name</b>	<b>Question</b>	One	Two	Three	Four	Five
Q11	I think it would be useful to have this app available for other questionnaires related to the city and city events	4	2	23	12	10
Q12	I would like to be able to give my opinion on city and city events using this app	4	0	23	14	11
Q13	I feel more engaged in the city / city event(s) by being able to give my opinion using the app.	4	4	22	10	12
Q14	If it were possible, I would like to launch my own questionnaires using this app.	6	4	30	7	5
Q15	For me it is important to have incentives/rewards when giving feedback to the city and/or city events	2	6	23	8	13
Q16	I would like to receive concrete items as incentive/reward such as money, voucher and/or prize	3	7	22	8	12
Q17	I would like to receive intangible incentive/reward such as points, ranking and recognition among participants	2	13	27	5	12



## **Appendix B**

### **The details of the campaigns in the UJI deployment**

Table B.1: Overview of all campaigns available during the deployment, sorted by the number of submissions.

	<b>Campaign title</b>	<b>Submissions</b>	<b>Data collectors</b>	<b>Geographical scope</b>	<b>Days</b>	<b>Origin</b>
1	Citizen participation	370	195	Neutral	20	Organizer
2	Improve University's WiFi	332	147	Constrained	20	Organizer
3	Graffiti in Castellón	212	125	Geographically-related	20	Organizer
4	Street animal in the city	197	75	Geographically-related	20	Organizer
5	University campus	154	113	Geographically-related	20	Organizer
6	City incidents	153	90	Geographically-related	20	Organizer
7	Hygienic issues in the city	117	61	Geographically-related	20	Organizer
8	City's cycling infrastructure	105	55	Geographically-related	20	Organizer
9	Professors and teaching methods	104	75	Neutral	12	User
10	Regional train service	103	92	Geographically-related	18	User
11	Leisure destination	101	81	Neutral	14	User
12	Respect	83	75	Neutral	19	User
13	The intermodal station	76	72	Geographically-related	20	User
14	City tram line	73	55	Geographically-related	16	User
15	Transport efficiency	72	68	Neutral	18	User
16	Digital technologies	64	49	Neutral	20	User
17	Street crossings	61	58	Neutral	14	User

Table B.2: Overview of all campaigns available during the deployment, sorted by the number of submissions (continued).

<b>Campaign title</b>	<b>Submissions</b>	<b>Data collectors</b>	<b>Geographical scope</b>	<b>Days</b>	<b>Origin</b>
18 Nocturnal ambiance in Castellón	52	44	Geographically-related	18	User
19 Sustainability issues	47	42	Neutral	19	User
20 Social vulnerability	46	35	Neutral	19	User
21 Sales promotion	45	29	Neutral	20	Organizer
22 Food consumption	41	41	Neutral	15	User
23 Citizens' labor happiness	40	40	Neutral	20	User
24 Cultural promotion in the city	36	25	Geographically-related	19	User
25 Publications of the university	35	19	Neutral	15	User
26 Final survey	31	31	Neutral	6	Organizer
27 Health	29	22	Neutral	19	User
28 The use of sport apps	29	20	Neutral	12	User
29 Entertainment	16	16	Neutral	13	User
30 Public trash bins	15	14	Neutral	20	User
31 The use of social networks	12	10	Neutral	10	User
32 Faculty of medical science	10	8	Neutral	13	User
33 Road improvement	9	8	Neutral	13	User
34 Walk path for blind people	9	8	Geographically-related	13	User
35 Mathematics	9	6	Neutral	11	User
36 Public resource	9	2	Neutral	9	User
37 Football tickets	8	6	Neutral	14	User
38 Valencia football club	7	6	Neutral	10	User
39 Local bus service	7	5	Geographically-related	4	User
40 Free outdoor fitness	6	4	Constrained	6	User
41 Discotheque in Burriana	5	5	Constrained	10	User
42 Willingness to help foreigners	3	1	Neutral	10	User
43 Entertainment near university	2	2	Neutral	5	User
44 Entrance for disabled people	2	2	Neutral	3	User



## Bibliography

- AGIR, BERKER; PAPAIOANNOU, THANASIS G; NARENDULA, RAMMOHAN; ABERER, KARL and HUBAUX, JEAN-PIERRE (2014). User-side adaptive protection of location privacy in participatory sensing. *Geoinformatica*, **18(1)**, pp. 165–191.
- ALBINO, VITO; BERARDI, UMBERTO and DANGELICO, ROSA MARIA (2015). Smart cities: Definitions, dimensions, performance, and initiatives. *Journal of Urban Technology*, **22(1)**, pp. 3–21.
- ALT, FLORIAN; SHIRAZI, ALIREZA SAHAMI; SCHMIDT, ALBRECHT; KRAMER, URS and NAWAZ, ZAHID (2010). Location-based crowdsourcing: extending crowdsourcing to the real world. In: *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, pp. 13–22. ACM.
- AMINTOOSI, HALEH and KANHERE, SALIL S (2014). A reputation framework for social participatory sensing systems. *Mobile Networks and Applications*, **19(1)**, pp. 88–100.
- AMINTOOSI, HALEH; KANHERE, SALIL S and TORSHIZ, MASOOD NIAZI (2015). A socially-aware incentive scheme for social participatory sensing. In: *Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 2015 IEEE Tenth International Conference on*, pp. 1–6. IEEE.
- ANDERSEN, JONAS LOHMANN ELKJÆR and LANDEX, ALEX (2008). Catchment areas for public transport. *WIT Transactions on the Built Environment*, **101**, pp. 175–184.
- ANDERSON, CHRIS (2006). *The Long Tail: Why the Future of Business Is Selling More for Less*. Hyperion.

- ARAKAWA, YUTAKA and MATSUDA, YUKI (2016). Gamification mechanism for enhancing a participatory urban sensing: Survey and practical results. *Journal of Information Processing*, **24(1)**, pp. 31–38.
- AUBRY, ELIAN; SILVERSTON, THOMAS; LAHMADI, ABDELKADER and FESTOR, OLIVIER (2014). CrowdOut: a mobile crowdsourcing service for road safety in digital cities. In: *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2014 IEEE International Conference on*, pp. 86–91. IEEE.
- BANGOR, AARON; KORTUM, PHILIP and MILLER, JAMES (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies*, **4(3)**, pp. 114–123.
- BAO, XUAN and ROY CHOUDHURY, ROMIT (2010). Movi: mobile phone based video highlights via collaborative sensing. In: *Proceedings of the 8th international conference on Mobile systems, applications, and services*, pp. 357–370. ACM.
- BONNEY, RICK; COOPER, CAREN B; DICKINSON, JANIS; KELLING, STEVE; PHILLIPS, TINA; ROSENBERG, KENNETH V and SHIRK, JENNIFER (2009). Citizen science: a developing tool for expanding science knowledge and scientific literacy. *BioScience*, **59(11)**, pp. 977–984.
- BRUNETTE, WAYLON; SUNDT, MITCHELL; DELL, NICOLA; CHAUDHRI, ROHIT; BREIT, NATHAN and BORRIELLO, GAETANO (2013). Open data kit 2.0: expanding and refining information services for developing regions. In: *Proceedings of the 14th Workshop on Mobile Computing Systems and Applications*, p. 10. ACM.
- BUDDE, MATTHIAS; SCHANKIN, ANDREA; HOFFMANN, JULIEN; DANZ, MARCEL; RIEDEL, TILL and BEIGL, MICHAEL (2017). Participatory Sensing or Participatory Nonsense?: Mitigating the Effect of Human Error on Data Quality in Citizen Science. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, **1(3)**, p. 39.
- BURKE, JEFFREY A; ESTRIN, DEBORAH; HANSEN, MARK; PARKER, ANDREW; RAMANATHAN, NITHYA; REDDY, SASANK and SRIVASTAVA, MANI B (2006). Participatory sensing.

- CAMPBELL, ANDREW T; EISENMAN, SHANE B; LANE, NICHOLAS D; MILUZZO, EMILIANO and PETERSON, RONALD A (2006). People-centric urban sensing. In: *Proceedings of the 2nd annual international workshop on Wireless internet*, p. 18. ACM.
- CAMPBELL, ANDREW T; EISENMAN, SHANE B; LANE, NICHOLAS D; MILUZZO, EMILIANO; PETERSON, RONALD A; LU, HONG; ZHENG, XIAO; MUSOLESI, MIRCO; FODOR, KRISTÓF and AHN, GAHNG-SEOP (2008). The rise of people-centric sensing. *IEEE Internet Computing*, **12(4)**.
- CAPPONI, ANDREA; FIANDRINO, CLAUDIO; KLIAZOVICH, DZMITRY and BOUVRY, PASCAL (2017). Energy efficient data collection in opportunistic mobile crowd-sensing architectures for smart cities. In: *Computer Communications Workshops (INFOCOM WKSHPs), 2017 IEEE Conference on*, pp. 307–312. IEEE.
- CARRERAS, IACOPO; MIORANDI, DANIELE; TAMILIN, ANDREI; SSEBAGGALA, EMANUEL R and CONCI, NICOLA (2013). Matador: Mobile task detector for context-aware crowd-sensing campaigns. In: *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2013 IEEE International Conference on*, pp. 212–217. IEEE.
- CARRON-ARTHUR, BRADLEY; CUNNINGHAM, JOHN A. and GRIFFITHS, KATHLEEN M. (2014). Describing the distribution of engagement in an Internet support group by post frequency: A comparison of the 90-9-1 Principle and Zipf's Law. *Internet Interventions*, **1(4)**, pp. 165 – 168. ISSN 2214-7829. doi: <https://doi.org/10.1016/j.invent.2014.09.003>.  
<http://www.sciencedirect.com/science/article/pii/S221478291400027X>
- CHON, YOHAN; LANE, NICHOLAS D; KIM, YUNJONG; ZHAO, FENG and CHA, HOJUNG (2013). A large-scale study of mobile crowdsourcing with smartphones for urban sensing applications. In: *Proc. of ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp'13), Zurich, Switzerland*, p. 1.
- CHRISTIN, D.; ROSSKOPF, C.; HOLLICK, M.; MARTUCCI, L. A. and KANHERE, S. S. (2012). IncogniSense: An anonymity-preserving reputation framework for participatory sensing applications. In: *2012 IEEE International Conference on*

- Pervasive Computing and Communications*, pp. 135–143. doi: 10.1109/PerCom.2012.6199860.
- CHRISTIN, DELPHINE (2016). Privacy in mobile participatory sensing: Current trends and future challenges. *Journal of Systems and Software*, **116**, pp. 57–68.
- CONOVER, WJ and IMAN, RONALD L (1979). On multiple-comparisons procedures. *Los Alamos Sci. Lab. Tech. Rep. LA-7677-MS*, pp. 1–14.
- CORCORAN, JONATHAN; ZAHNOW, RENEE and ASSEMI, BEHRANG (2017). Wander: A Smartphone App for Sensing Sociability. *Applied Spatial Analysis and Policy*, pp. 1–20.
- CROSS, JOHN A (2001). Megacities and small towns: different perspectives on hazard vulnerability. *Global Environmental Change Part B: Environmental Hazards*, **3(2)**, pp. 63–80.
- DANIELS, RHONDA and MULLEY, CORINNE (2013). Explaining walking distance to public transport: The dominance of public transport supply. *Journal of Transport and Land Use*, **6(2)**, pp. 5–20.
- DAS, TATHAGATA; MOHAN, PRASHANTH; PADMANABHAN, VENKATA N; RAMJEE, RAMACHANDRAN and SHARMA, ASANKHAYA (2010). PRISM: platform for remote sensing using smartphones. In: *Proceedings of the 8th international conference on Mobile systems, applications, and services*, pp. 63–76. ACM.
- DENG, LINDA and COX, LANDON P (2009). Livecompare: grocery bargain hunting through participatory sensing. In: *Proceedings of the 10th workshop on Mobile Computing Systems and Applications*, p. 4. ACM.
- D'HONDT, ELLIE; ZAMAN, JESSE; PHILIPS, ELINE; BOIX, ELISA GONZALEZ and DE MEUTER, WOLFGANG (2014). Orchestration support for participatory sensing campaigns. In: *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 727–738. ACM.
- DUAN, L.; KUBO, T.; SUGIYAMA, K.; HUANG, J.; HASEGAWA, T. and WALRAND, J. (2014). Motivating Smartphone Collaboration in Data Acquisition and Distributed Computing. *IEEE Transactions on Mobile Computing*, **13(10)**, pp. 2320–2333. ISSN 1536-1233. doi: 10.1109/TMC.2014.2307327.



- D'HONDT, ELLIE; STEVENS, MATTHIAS and JACOBS, AN (2013). Participatory noise mapping works! An evaluation of participatory sensing as an alternative to standard techniques for environmental monitoring. *Pervasive and Mobile Computing*, **9(5)**, pp. 681–694.
- EISENMAN, SHANE B.; MILUZZO, EMILIANO; LANE, NICHOLAS D.; PETERSON, RONALD A.; AHN, GAHNG-SEOP and CAMPBELL, ANDREW T. (2010). BikeNet: A Mobile Sensing System for Cyclist Experience Mapping. *ACM Trans. Sen. Netw.*, **6(1)**, pp. 6:1–6:39. ISSN 1550-4859. doi: 10.1145/1653760.1653766. <http://doi.acm.org/10.1145/1653760.1653766>
- FALTINGS, B.; LI, J. J. and JURCA, R. (2014). Incentive Mechanisms for Community Sensing. *IEEE Transactions on Computers*, **63(1)**, pp. 115–128. ISSN 0018-9340. doi: 10.1109/TC.2013.150.
- FENG, ZHENNI; ZHU, YANMIN; ZHANG, QIAN; NI, LIONEL M and VASILAKOS, ATHANASIOS V (2014). TRAC: Truthful auction for location-aware collaborative sensing in mobile crowdsourcing. In: *INFOCOM, 2014 Proceedings IEEE*, pp. 1231–1239. IEEE.
- FIANDRINO, CLAUDIO; KANTARCI, BURAK; ANJOMSHOA, FAZEL; KLIAZOVICH, DZMITRY; BOUVRY, PASCAL and MATTHEWS, JEANNA (2016). Sociability-driven user recruitment in mobile crowdsensing internet of things platforms. In: *IEEE Global Communications Conference (GLOBECOM), Washington, DC, USA, 2016*, .
- GANTI, RAGHU K; YE, FAN and LEI, HUI (2011). Mobile crowdsensing: current state and future challenges. *IEEE Communications Magazine*, **49(11)**.
- GAO, HUI; LIU, CHI HAROLD; WANG, WENDONG; SHENG, ZHENG GUO; CHIN, ALVIN and TIAN, YE (2015a). Maximizing Data Credibility under Budget Constraint for Participatory Sensing. In: *High Performance Computing and Communications (HPCC), 2015 IEEE 7th International Symposium on Cyberspace Safety and Security (CSS), 2015 IEEE 12th International Conference on Embedded Software and Systems (ICCESS), 2015 IEEE 17th International Conference on*, pp. 134–139. IEEE.
- GAO, HUI; LIU, CHI HAROLD; WANG, WENDONG; ZHAO, JIANXIN R; SONG, ZHENG; SU, XIN; CROWCROFT, JON and LEUNG, KIN K (2015b). A Survey of Incentive

- Mechanisms for Participatory Sensing. *IEEE Communications Surveys and Tutorials*, **17(2)**, pp. 918–943.
- GAO, JING and DUAN, HUI-CHUAN (2011). Research on data transmission efficiency of JSON. *Computer Engineering and Design*, **32(7)**, pp. 2267–2270.
- GEIGER, R STUART and HALFAKER, AARON (2013). Using edit sessions to measure participation in Wikipedia. In: *Proceedings of the 2013 conference on Computer supported cooperative work*, pp. 861–870. ACM.
- GEO-C JOINT DOCTORATE PROGRAM (2018). Open City Toolkit. Internet.  
<http://giv-oct.uni-muenster.de:5000/>
- GOODCHILD, MICHAEL F (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal*, **69(4)**, pp. 211–221.
- GRASSO, VALENTINA; CRISCI, ALFONSO; MORABITO, MARCO; NESI, PAOLO and PANTALEO, GIANNI (2017). Public crowdsensing of heat waves by social media data. *Advances in Science and Research*, **14**, pp. 217–226.
- GRITZ, ANJA S (2004). The impact of material incentives on response quantity, response quality, sample composition, survey outcome and cost in online access panels. *International Journal of Market Research*, **46(3)**, pp. 327–345.
- GUO, B.; CHEN, C.; ZHANG, D.; YU, Z. and CHIN, A. (2016). Mobile crowd sensing and computing: when participatory sensing meets participatory social media. *IEEE Communications Magazine*, **54(2)**, pp. 131–137. ISSN 0163-6804. doi: 10.1109/MCOM.2016.7402272.
- GUO, B.; YU, Z.; ZHANG, D. and ZHOU, X. (2014). Cross-community sensing and mining. *IEEE Communications Magazine*, **52(8)**, pp. 144–152. ISSN 0163-6804. doi: 10.1109/MCOM.2014.6871682.
- GUO, BIN; WANG, ZHU; YU, ZHIWEN; WANG, YU; YEN, NEIL Y; HUANG, RUNHE and ZHOU, XINGSHE (2015). Mobile crowd sensing and computing: The review of an emerging human-powered sensing paradigm. *ACM Computing Surveys (CSUR)*, **48(1)**, p. 7.

- GUO, BIN; ZHANG, DAQING; YU, ZHIWEN; LIANG, YUNJI; WANG, ZHU and ZHOU, XINGSHE (2013). From the internet of things to embedded intelligence. *World Wide Web*, **16(4)**, pp. 399–420.
- HAMID, SHERIN ABDEL; TAKAHARA, GLEN and HASSANEIN, HOSSAM S (2013). On the recruitment of smart vehicles for urban sensing. In: *Global Communications Conference (GLOBECOM), 2013 IEEE*, pp. 36–41. IEEE.
- HAN, KYUNGSIK; GRAHAM, ERIC A; VASSALLO, DYLAN and ESTRIN, DEBORAH (2011). Enhancing motivation in a mobile participatory sensing project through gaming. In: *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on*, pp. 1443–1448. IEEE.
- HIGGINS, CHRISTOPHER I; WILLIAMS, JAMIE; LEIBOVICI, DIDIER G; SIMONIS, INGO; DAVIS, MASON J; MULDOON, CONOR; VAN GENUCHTEN, PAUL; O'HARE, GREGORY and WIEMANN, STEFAN (2016). Citizen OBServatory WEB (COB-WEB): A generic infrastructure platform to facilitate the collection of citizen science data for environmental monitoring. *Int. J. Spat. Data Infrastruct. Res.*, **11**, pp. 20–48.
- HINZ, OLIVER; ECKERT, JOCHEN and SKIERA, BERND (2011). Drivers of the long tail phenomenon: an empirical analysis. *Journal of management information systems*, **27(4)**, pp. 43–70.
- HOCHMAIR, HARTWIG H and ZIELSTRA, DENNIS (2015). Analysing user contribution patterns of drone pictures to the dronestagram photo sharing portal. *Journal of Spatial Science*, **60(1)**, pp. 79–98.
- HOFFMANN, BARBARA; MOEBUS, SUSANNE; STANG, ANDREAS; BECK, EVA-MARIA; DRAGANO, NICO; MÖHLENKAMP, STEPHAN; SCHMERMUND, AXEL; MEMMESHEIMER, MICHAEL; MANN, KLAUS; ERBEL, RAIMUND et al. (2006). Residence close to high traffic and prevalence of coronary heart disease. *European Heart Journal*, **27(22)**, pp. 2696–2702.
- HUGHES, DANNY; COULSON, GEOFF and WALKERDINE, JAMES (2005). Free riding on Gnutella revisited: the bell tolls? *IEEE distributed systems online*, **6(6)**.

- IRISSAPPANE, ATHIRAI ARAVAZHI; JIANG, SIWEI and ZHANG, JIE (2012). Towards a comprehensive testbed to evaluate the robustness of reputation systems against unfair rating attack. In: *UMAP Workshops*, volume 12.
- JAIMES, LUIS G; VERGARA-LAURENS, IDALIDES and LABRADOR, MIGUEL A (2012). A location-based incentive mechanism for participatory sensing systems with budget constraints. In: *Pervasive Computing and Communications (PerCom), 2012 IEEE International Conference on*, pp. 103–108. IEEE.
- JAIMES, LUIS G; VERGARA-LAURENS, IDALIDES J and RAIJ, ANDREW (2015). A survey of incentive techniques for mobile crowd sensing. *IEEE Internet of Things Journal*, **2(5)**, pp. 370–380.
- JAYARAMAN, PREM PRAKASH; PERERA, CHARITH; GEORGAKOPOULOS, DIMITRIOS and ZASLAVSKY, ARKADY (2013). Efficient opportunistic sensing using mobile collaborative platform mosden. *arXiv preprint arXiv:1310.4052*.
- JUHÁSZ, LEVENTE and HOCHMAIR, HARTWIG H (2016). User contribution patterns and completeness evaluation of Mapillary, a crowdsourced street level photo service. *Transactions in GIS*, **20(6)**, pp. 925–947.
- KAEMARUNGSU, KAMOL (2006). Distribution of WLAN received signal strength indication for indoor location determination. In: *Wireless Pervasive Computing, 2006 1st International Symposium on*, pp. 6–pp. IEEE.
- KANHERE, SALIL S (2011). Participatory sensing: Crowdsourcing data from mobile smartphones in urban spaces. In: *Mobile Data Management (MDM), 2011 12th IEEE International Conference on*, volume 2, pp. 3–6. IEEE.
- KANJO, EIMAN (2010). Noisespy: A real-time mobile phone platform for urban noise monitoring and mapping. *Mobile Networks and Applications*, **15(4)**, pp. 562–574.
- KANTARCI, BURAK and MOUFTAH, HUSSEIN T (2014). Trustworthy sensing for public safety in cloud-centric internet of things. *IEEE Internet of Things Journal*, **1(4)**, pp. 360–368.
- KAPADIA, APU; KOTZ, DAVID and TRIANDOPOULOS, NIKOS (2009). Opportunistic sensing: Security challenges for the new paradigm. In: *Communication Systems*

- and Networks and Workshops, 2009. COMSNETS 2009. First International*, pp. 1–10. IEEE.
- KATSOMALLOS, MANOS and LALIS, SPYROS (2014). EasyHarvest: Supporting the deployment and management of sensing applications on smartphones. In: *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2014 IEEE International Conference on*, pp. 80–85. IEEE.
- KHAN, WAZIR ZADA; XIANG, YANG; AALSALEM, MOHAMMED Y and ARSHAD, QURATULAIN (2013). Mobile phone sensing systems: A survey. *IEEE Communications Surveys & Tutorials*, **15(1)**, pp. 402–427.
- KHOI, NGO and CASTELEYN, SVEN (2018). Analyzing Spatial and Temporal User Behavior in Participatory Sensing. *ISPRS International Journal of Geo-Information*, **7(9)**, p. 344.
- KHOI, NGO MANH; CASTELEYN, SVEN; MORADI, M MEHDI and PEBESMA, EDZER (2018). Do Monetary Incentives Influence Users' Behavior in Participatory Sensing? *Sensors*, **18(5)**, p. 1426.
- KIM, SUNYOUNG; MANKOFF, JENNIFER and PAULOS, ERIC (2013). Sensr: evaluating a flexible framework for authoring mobile data-collection tools for citizen science. In: *Proceedings of the 2013 conference on Computer supported cooperative work*, pp. 1453–1462. ACM.
- KOUTSOPOULOS, IORDANIS (2013). Optimal incentive-driven design of participatory sensing systems. In: *Infocom, 2013 proceedings ieee*, pp. 1402–1410. IEEE.
- KRAUSE, ANDREAS; HORVITZ, ERIC; KANSAL, AMAN and ZHAO, FENG (2008). Toward community sensing. In: *Information Processing in Sensor Networks, 2008. IPSN'08. International Conference on*, pp. 481–492. IEEE.
- LANE, NICHOLAS D; MILUZZO, EMILIANO; LU, HONG; PEEBLES, DANIEL; CHOUDHURY, TANZEEM and CAMPBELL, ANDREW T (2010). A survey of mobile phone sensing. *IEEE Communications magazine*, **48(9)**.
- LEE, JUONG-SIK and HOH, BAIK (2010a). Dynamic pricing incentive for participatory sensing. *Pervasive and Mobile Computing*, **6(6)**, pp. 693–708.

- LEE, JUONG-SIK and HOH, BAIK (2010b). Sell your experiences: a market mechanism based incentive for participatory sensing. In: *Pervasive Computing and Communications (PerCom), 2010 IEEE International Conference on*, pp. 60–68. IEEE.
- LI, CHUNMING; LIU, YIN and HAKLAY, MUKI (2018). Participatory soundscape sensing. *Landscape and Urban Planning*, **173**, pp. 64–69.
- LIMA, ANTONIO and MUSOLESI, MIRCO (2012). Spatial dissemination metrics for location-based social networks. In: *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pp. 972–979. ACM.
- LU, HONG; PAN, WEI; LANE, NICHOLAS D; CHOUDHURY, TANZEEM and CAMPBELL, ANDREW T (2009). SoundSense: scalable sound sensing for people-centric applications on mobile phones. In: *Proceedings of the 7th international conference on Mobile systems, applications, and services*, pp. 165–178. ACM.
- LUO, C.; HONG, H.; CHAN, M. C.; LI, J.; ZHANG, X. and MING, Z. (2018). MPiLoc: Self-Calibrating Multi-Floor Indoor Localization Exploiting Participatory Sensing. *IEEE Transactions on Mobile Computing*, **17(1)**, pp. 141–154. ISSN 1536-1233. doi: 10.1109/TMC.2017.2698453.
- LUTZ, WOLFGANG; BUTZ, WILLIAM P and SAMIR, KC ED (2017). *World Population & Human Capital in the Twenty-First Century: An Overview*. Oxford University Press.
- MAISONNEUVE, NICOLAS; STEVENS, MATTHIAS; NIESSEN, MARIA E and STEELS, LUC (2009). NoiseTube: Measuring and mapping noise pollution with mobile phones. In: *Information technologies in environmental engineering*, pp. 215–228. Springer.
- MALEK, ROBERT; TATTONI, CLARA; CIOLLI, MARCO; CORRADINI, STEFANO; ANDREIS, DANIELE; IBRAHIM, AYA; MAZZONI, VALERIO; ERIKSSON, ANNA and ANFORA, GIANFRANCO (2018). Coupling Traditional Monitoring and Citizen Science to Disentangle the Invasion of *Halyomorpha halys*. *ISPRS International Journal of Geo-Information*, **7(5)**, p. 171.
- MAO, ANDREW; KAMAR, ECE; CHEN, YILING; HORVITZ, ERIC; SCHWAMB, MEGAN E; LINTOTT, CHRIS J and SMITH, ARFON M (2013). Volunteering

- versus work for pay: Incentives and tradeoffs in crowdsourcing. In: *First AAAI conference on human computation and crowdsourcing*, .
- MARKERT, JAN-FREDERIC; BUDDE, MATTHIAS; SCHINDLER, GREGOR; KLUG, MARKUS and BEIGL, MICHAEL (2016). Private Rendezvous-based Calibration of Low-Cost Sensors for Participatory Environmental Sensing. In: *Proceedings of the Second International Conference on IoT in Urban Space, Urb-IoT '16*, pp. 82–85. ACM, New York, NY, USA. ISBN 978-1-4503-4204-9. doi: 10.1145/2962735.2962754.  
<http://doi.acm.org/10.1145/2962735.2962754>
- MASSUNG, ELAINE; COYLE, DAVID; CATER, KIRSTEN F; JAY, MARC and PREIST, CHRIS (2013). Using crowdsourcing to support pro-environmental community activism. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 371–380. ACM.
- MENDEZ, DIEGO and LABRADOR, MIGUEL A (2014). A general framework for participatory sensing systems. *Journal of Networks*, **9(11)**, p. 2994.
- MERGEL, INES (2012). Distributed democracy: SeeClickFix.com for crowdsourced issue reporting.
- MONTORI, F.; BEDOGNI, L. and BONONI, L. (2018a). A Collaborative Internet of Things Architecture for Smart Cities and Environmental Monitoring. *IEEE Internet of Things Journal*, **5(2)**, pp. 592–605. ISSN 2327-4662. doi: 10.1109/JIOT.2017.2720855.
- MONTORI, FEDERICO; JAYARAMAN, PREM PRAKASH; YAVARI, ALI; HASSANI, ALIREZA and GEORGAKOPOULOS, DIMITRIOS (2018b). The Curse of Sensing: Survey of techniques and challenges to cope with sparse and dense data in mobile crowd sensing for Internet of Things. *Pervasive and Mobile Computing*, **49**, pp. 111 – 125. ISSN 1574-1192. doi: <https://doi.org/10.1016/j.pmcj.2018.06.009>.  
<http://www.sciencedirect.com/science/article/pii/S1574119217306417>
- MOUSA, HAYAM; MOKHTAR, SONIA BEN; HASAN, OMAR; YOUNES, OSAMA; HADHOUD, MOHIY and BRUNIE, LIONEL (2015). Trust management and reputation systems in mobile participatory sensing applications: A survey. *Computer Networks*, **90**, pp. 49–73.

- MUKHERJEE, TRIDIB; CHANDER, DEEPTHI; MONDAL, ANIRBAN; DASGUPTA, KOUS-TUV; KUMAR, AMIT and VENKAT, ASHWIN (2014). CityZen: A cost-effective city management system with incentive-driven resident engagement. In: *Mobile Data Management (MDM), 2014 IEEE 15th International Conference on*, volume 1, pp. 289–296. IEEE.
- MÜNDEL, THOMAS; SCHMIDT, FRANK P; STEVEN, SEBASTIAN; HERZOG, JOHANNES; DAIBER, ANDREAS and SØRENSEN, METTE (2018). Environmental noise and the cardiovascular system. *Journal of the American College of Cardiology*, **71(6)**, pp. 688–697.
- MUSTHAG, MOHAMED; RAIJ, ANDREW; GANESAN, DEEPAK; KUMAR, SANTOSH and SHIFFMAN, SAUL (2011). Exploring micro-incentive strategies for participant compensation in high-burden studies. In: *Proceedings of the 13th international conference on Ubiquitous computing*, pp. 435–444. ACM.
- OPEN SIGNAL (2017). Open Signal. Internet.  
<https://opensignal.com/>
- OZCAN, H KURTULUS and NEMLIOGLU, SEMIH (2006). In-cabin noise levels during commercial aircraft flights. *Canadian Acoustics*, **34(4)**, pp. 31–35.
- PAN, BING and LI, XIANG ROBERT (2011). The long tail of destination image and online marketing. *Annals of Tourism Research*, **38(1)**, pp. 132–152.
- PAVLOV, DIMITAR VALENTINOV (2013). Hive: An extensible and scalable framework for mobile crowdsourcing. *Diss. Imperial College London*.
- PENG, DAN; WU, FAN and CHEN, GUIHAI (2015). Pay as how well you do: A quality based incentive mechanism for crowdsensing. In: *Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pp. 177–186. ACM.
- POURYAZDAN, MARYAM; FIANDRINO, CLAUDIO; KANTARCI, BURAK; SOYATA, TOLGA; KLIAZOVICH, DZMITRY and BOUVRY, PASCAL (2017). Intelligent gaming for mobile crowd-sensing participants to acquire trustworthy big data in the internet of things. *IEEE Access*, **5**, pp. 22209–22223.



- PREMCHAIWADI, WICHIAN; ROMSAIYUD, WALISA and PREMCHAIWADI, NUCHARÉE (2011). Navigation without GPS: Fake location for mobile phone tracking. In: *ITS Telecommunications (ITST), 2011 11th International Conference on*, pp. 195–200. IEEE.
- QIAN, CHENG; KATS, PHILIPP; MALINCHIK, SERGEY; HOFFMAN, MARK; KETTLER, BRIAN; KONTOKOSTA, CONSTANTINE and SOBOLEVSKY, STANISLAV (2017). Geo-tagged social media data as a proxy for urban mobility. In: *International Conference on Applied Human Factors and Ergonomics*, pp. 29–40. Springer.
- RA, MOO-RYONG; LIU, BIN; LA PORTA, TOM F and GOVINDAN, RAMESH (2012). Medusa: A programming framework for crowd-sensing applications. In: *Proceedings of the 10th international conference on Mobile systems, applications, and services*, pp. 337–350. ACM.
- RANA, RAJIB KUMAR; CHOU, CHUN TUNG; KANHERE, SALIL S; BULUSU, NIRUPAMA and HU, WEN (2010). Ear-phone: an end-to-end participatory urban noise mapping system. In: *Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks*, pp. 105–116. ACM.
- REDDY, SASANK; ESTRIN, DEBORAH; HANSEN, MARK and SRIVASTAVA, MANI (2010a). Examining micro-payments for participatory sensing data collections. In: *Proceedings of the 12th ACM international conference on Ubiquitous computing*, pp. 33–36. ACM.
- REDDY, SASANK; ESTRIN, DEBORAH and SRIVASTAVA, MANI (2010b). Recruitment framework for participatory sensing data collections. In: *International Conference on Pervasive Computing*, pp. 138–155. Springer.
- REDDY, SASANK; SHILTON, KATIE; BURKE, JEFF; ESTRIN, DEBORAH; HANSEN, MARK and SRIVASTAVA, MANI (2008). Evaluating participation and performance in participatory sensing. *UrbanSense08*, 4.
- REDDY, SASANK; SHILTON, KATIE; BURKE, JEFF; ESTRIN, DEBORAH; HANSEN, MARK and SRIVASTAVA, MANI (2009). Using context annotated mobility profiles to recruit data collectors in participatory sensing. In: *International Symposium on Location-and Context-Awareness*, pp. 52–69. Springer.

- REDDY, SASANK; SHILTON, KATIE; DENISOV, GLEB; CENIZAL, CHRISTIAN; ESTRIN, DEBORAH and SRIVASTAVA, MANI (2010c). Biketastic: sensing and mapping for better biking. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1817–1820. ACM.
- RESTUCCIA, FRANCESCO; DAS, SAJAL K and PAYTON, JAMIE (2016). Incentive mechanisms for participatory sensing: Survey and research challenges. *ACM Transactions on Sensor Networks (TOSN)*, **12(2)**, p. 13.
- RUDOLF, GIFFINGER; CHRISTIAN, FERTNER; HANS, KRAMAR; ROBERT, KALASEK; NATAŠA, PICHLER-MILANOVIC and EVERT, MEIJERS (2007). Smart cities ranking of European medium-sized cities.
- RUGE, LUKAS; ALTAKROURI, BASHAR and SCHRADER, ANDREAS (2013). Soundofthecity-continuous noise monitoring for a healthy city. In: *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2013 IEEE International Conference on*, pp. 670–675. IEEE.
- RULA, JOHN P; NAVDA, VISHNU; BUSTAMANTE, FABIÁN E; BHAGWAN, RANJITA and GUHA, SAIKAT (2014). No one-size fits all: Towards a principled approach for incentives in mobile crowdsourcing. In: *Proceedings of the 15th Workshop on Mobile Computing Systems and Applications*, p. 3. ACM.
- SANCHEZ, LUIS; GUTIERREZ, VERONICA; GALACHE, JOSE A; SOTRES, PABLO; SANTANA, JUAN R and MUÑOZ, LUIS (2014). Engaging individuals in the smart city paradigm: Participatory sensing and augmented reality. *Interdisciplinary Studies Journal*, **3(4)**, p. 129.
- SCHWEIZER, IMMANUEL; BÄRTL, ROMAN; SCHULZ, AXEL; PROBST, FLORIAN and MÜHLÄUSER, MAX (2011). NoiseMap-real-time participatory noise maps. In: *Second International Workshop on Sensing Applications on Mobile Phones*, pp. 1–5. Citeseer.
- SCHWEIZER, IMMANUEL; MEURISCH, CHRISTIAN; GEDEON, JULIEN; BÄRTL, ROMAN and MÜHLHÄUSER, MAX (2012). Noisemap: multi-tier incentive mechanisms for participative urban sensing. In: *Proceedings of the Third International Workshop on Sensing Applications on Mobile Phones*, p. 9. ACM.

- SENGPIELAUDIO (2018). Decibel table loudness comparison chart. Internet.  
<http://www.siue.edu/~gengel/ece476WebStuff/SPL.pdf>
- SHEN, YANG TING; SHIU, YI SHIANG; LIU, WEI KUANG and LU, PEI WEN (2017). The participatory sensing platform driven by UGC for the evaluation of living quality in the city. In: *International Conference on Human Interface and the Management of Information*, pp. 516–527. Springer.
- SILVA, THIAGO H; DE MELO, PEDRO OLMO S VAZ; ALMEIDA, JUSSARA M and LOUREIRO, ANTONIO AF (2013a). Social media as a source of sensing to study city dynamics and urban social behavior: Approaches, models, and opportunities. In: *Ubiquitous Social Media Analysis*, pp. 63–87. Springer.
- SILVA, THIAGO H; DE MELO, PEDRO OS VAZ; VIANA, ALINE CARNEIRO; ALMEIDA, JUSSARA M; SALLES, JULIANA and LOUREIRO, ANTONIO AF (2013b). Traffic condition is more than colored lines on a map: characterization of waze alerts. In: *International Conference on Social Informatics*, pp. 309–318. Springer.
- SOLDATOS, JOHN; KEFALAKIS, NIKOS; HAUSWIRTH, MANFRED; SERRANO, MARTIN; CALBIMONTE, JEAN-PAUL; RIAHI, MEHDI; ABERER, KARL; JAYARAMAN, PREM PRAKASH; ZASLAVSKY, ARKADY; ŽARKO, IVANA PODNAR; SKORINKAPOV, LEA and HERZOG, REINHARD (2015). OpenIoT: Open Source Internet-of-Things in the Cloud. In: Ivana Podnar Žarko; Krešimir Pripužić and Martin Serrano (Eds.), *Interoperability and Open-Source Solutions for the Internet of Things*, pp. 13–25. Springer International Publishing, Cham. ISBN 978-3-319-16546-2.
- STEWART, OSAMUYIMEN; HUERTA, JUAN M and SADER, MELISSA (2009). Designing crowdsourcing community for the enterprise. In: *Proceedings of the ACM SIGKDD Workshop on Human Computation*, pp. 50–53. ACM.
- SULTAN, JODY; BEN-HAIM, GEV; HAUNERT, JAN-HENRIK and DALYOT, SAGI (2017). Extracting spatial patterns in bicycle routes from crowdsourced data. *Transactions in GIS*, **21(6)**, pp. 1321–1340.
- SUMARAY, AUDIE and MAKKI, S KAMI (2012). A comparison of data serialization formats for optimal efficiency on a mobile platform. In: *Proceedings of the 6th international conference on ubiquitous information management and communication*, p. 48. ACM.

- SZABÓ, RÓBERT; FARKAS, KÁROLY; ISPÁNY, MÁRTON; BENCZÚR, ANDRÁS A; BÁTFAI, NORBERT; JESZENSZKY, PÉTER; LAKI, SÁNDOR; VÁGNER, ANIKÓ; KOLLÁR, LAJOS; SIDLÓ, CS et al. (2013). Framework for smart city applications based on participatory sensing. In: *Cognitive Infocommunications (CogInfoCom), 2013 IEEE 4th International Conference on*, pp. 295–300. IEEE.
- TANG, JIE; LOU, TIANCHENG and KLEINBERG, JON (2012). Inferring Social Ties Across Heterogenous Networks. In: *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining, WSDM '12*, pp. 743–752. ACM, New York, NY, USA. ISBN 978-1-4503-0747-5. doi: 10.1145/2124295.2124382. <http://doi.acm.org/10.1145/2124295.2124382>
- TANGMUNARUNKIT, HONGSUDA; HSIEH, CHENG-KANG; LONGSTAFF, BRENT; NOLEN, S; JENKINS, JOHN; KETCHAM, CAMERON; SELSKY, JOSHUA; ALQUADDOOMI, FAISAL; GEORGE, DONY; KANG, JINHA et al. (2015). Ohmage: A general and extensible end-to-end participatory sensing platform. *ACM Transactions on Intelligent Systems and Technology (TIST)*, **6(3)**, p. 38.
- TSAMPOULATIDIS, IOANNIS; VERVERIDIS, DIMITRIOS; TSARCHOPOULOS, PANAGIOTIS; NIKOLOPOULOS, SPIROS; KOMPATSIARIS, IOANNIS and KOMNINOS, NICOS (2013). ImproveMyCity: an open source platform for direct citizen-government communication. In: *Proceedings of the 21st ACM international conference on Multimedia*, pp. 839–842. ACM.
- TUNCAY, GÜLİZ SERAY; BENINCASA, GIACOMO and HELMY, AHMED (2013). Autonomous and distributed recruitment and data collection framework for opportunistic sensing. *ACM SIGMOBILE Mobile Computing and Communications Review*, **16(4)**, pp. 50–53.
- UEYAMA, YOSHITAKA; TAMAI, MORIHIKO; ARAKAWA, YUTAKA and YASUMOTO, KEIICHI (2014). Gamification-based incentive mechanism for participatory sensing. In: *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2014 IEEE International Conference on*, pp. 98–103. IEEE.
- UNITED NATIONS (2008). *Urban Agglomerations 2007*. World Urbanization Prospects.
- UNITED NATIONS (2014). *World urbanization prospects: The 2014 revision, highlights*. World Urbanization Prospects.

- UNITED NATIONS (2017). *World Population Prospects: The 2017 Revision, Key Findings and Advance Tables*. World Urbanization Prospects.
- WALRAVENS, NILS (2013). Validating a business model framework for smart city services: The case of fixmystreet. In: *Advanced Information Networking and Applications Workshops (WAINA), 2013 27th International Conference on*, pp. 1355–1360. IEEE.
- WANG, JIANMING; LIU, XIAOFENG; LU, RUIHUA; WANG, CHENGFEI and WANG, CHENGCHENG (2016). Poster: Interactive Platform for Urban Bird Studies Using Participatory Sensing. In: *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services Companion*, pp. 85–85. ACM.
- WANG, YIRAN; NIIYA, MELISSA; MARK, GLORIA; REICH, STEPHANIE M and WARSCHAUER, MARK (2015). Coming of age (digitally): An ecological view of social media use among college students. In: *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pp. 571–582. ACM.
- WEI, WEI; XU, FENGYUAN; TAN, CHIU C and LI, QUN (2012). Sybildefender: Defend against sybil attacks in large social networks. In: *INFOCOM, 2012 Proceedings IEEE*, pp. 1951–1959. IEEE.
- WILKINSON, DENNIS M and HUBERMAN, BERNARDO A (2007). Assessing the value of cooperation in wikipedia. *arXiv preprint cs/0702140*.
- WU, FANG-JING and LUO, TIE (2014). A generic participatory sensing framework for multi-modal datasets. In: *Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 2014 IEEE Ninth International Conference on*, pp. 1–6. IEEE.
- XIAO, MINGJUN; WU, JIE; HUANG, HE; HUANG, LIUSHENG and HU, CHANG (2016). Deadline-sensitive user recruitment for probabilistically collaborative mobile crowdsensing. In: *Distributed Computing Systems (ICDCS), 2016 IEEE 36th International Conference on*, pp. 721–722. IEEE.
- YAN, TINGXIN; HOH, BAIK; GANESAN, DEEPAK; TRACTON, KENNETH; IWUCHUKWU, TOCH and LEE, JUONG-SIK (2011). CrowdPark: A crowdsourcing-based parking

- reservation system for mobile phones. *University of Massachusetts at Amherst Tech. Report*.
- YAO, HONG; ZHANG, CHANGKAI; LIU, CHAO; LIANG, QINGZHONG; YAN, XUESONG and HU, CHENGYU (2015). On participant selection for minimum cost participatory urban sensing with guaranteed quality of information. In: *International Conference on Collaborative Computing: Networking, Applications and Work-sharing*, pp. 183–194. Springer.
- YICK, JENNIFER; MUKHERJEE, BISWANATH and GHOSAL, DIPAK (2008). Wireless sensor network survey. *Computer networks*, **52(12)**, pp. 2292–2330.
- ZAMAN, J. and MEUTER, W. DE (2015). DisCoPar: Distributed components for participatory campaigning. In: *2015 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*, pp. 160–165. doi: 10.1109/PERCOMW.2015.7134012.
- ZHANG, LIZI; JIANG, SIWEI; ZHANG, JIE and NG, WEE KEONG (2012). Robustness of trust models and combinations for handling unfair ratings. In: *IFIP International Conference on Trust Management*, pp. 36–51. Springer.
- ZHANG, XINGLIN; YANG, ZHENG; SUN, WEI; LIU, YUNHAO; TANG, SHAOHUA; XING, KAI and MAO, XUFEI (2016). Incentives for mobile crowd sensing: A survey. *IEEE Communications Surveys & Tutorials*, **18(1)**, pp. 54–67.
- ZHU, J.; KIM, K.; MOHAPATRA, P. and CONGDON, P. (2013). An adaptive privacy-preserving scheme for location tracking of a mobile user. In: *2013 IEEE International Conference on Sensing, Communications and Networking (SECON)*, pp. 140–148. ISSN 2155-5486. doi: 10.1109/SAHCN.2013.6644972.