New data collection channel based on mobile phone APPs and applied to the demographic survey during the year 2020-2021

**Keywords:** Innovation, mobility, optimisation, quality

# Introduction

Over recent years the process of collecting data in statistical processes has undergone significant changes caused mainly by the speed of technological changes generated by the phenomenon of the internet and mobile phones. Over the last five years, the use of self-administered electronic online questionnaires became established as a new channel that has now been fully assimilated by most official Statistics Institutes. However, two years ago the boom in the use of mobile phones by a large part of today's society led us to launch a pilot project to evaluate the feasibility of using a new data collection channel based on building telephone software (apps) for mobile devices based on responsive design.

As a result of this project, an app was built for the “Time Budgets” survey and was used as an additional channel in the 2018 campaign. Analysis of the results obtained through the various channels offered to the public by Eustat revealed that both in terms of the quality of the answers and the number of questions answered, the data did not exhibit any bias affecting the production of statistics.

In 2019 it was decided to develop an app for the demographic survey that begins its data collection campaign in November this year. The purpose of the presentation, therefore, is to explain how the Basque Statistics Institute uses apps as a new data collection channel and to present the results obtained in both projects.

# Technical Architecture and design

All the Internet-based information collection channels must be fully integrated to enable the public to use any combination of these channels without information being lost. Consequently, the APP was developed in integration with the corresponding web questionnaire in such a way that the gathered information is stored directly at the same database storage location. This implies increased development complexity as regards synchronisation of the two channels, greater security measures and adaptation of the design according to the physical characteristics of the devices used in each channel.

## Channel synchronisation

Advances in telecommunications technology are facilitating a paradigm shift for traditional computer systems such that both information and processes can be hosted on external servers, which may be owned or rented, outside the organisations themselves. As a result, use of the software does not have to be linked to the private networks of each company, which are replaced by the internet, allowing access to be independent of the traditional physical workspace.

What is more, the mobile phone network has taken a spectacular technological leap forward over the last four years, improving service quality and allowing information of any type (images, videos...) and any size to be transmitted, within certain limits.

As such, we are addressing a scenario that allows statistics institutes to build innovative solutions to adapt to the changes that occur both in today's society and in the way we approach our work. Within this framework, our institute opted for a solution that integrates the web questionnaire channel with the mobile phone channel, with the following key elements:

* Single storage location for both channels
* Online updating for both channels
* Development of processes to synchronise the information entered in the two channels to avoid both loss of information and duplication of responses caused by members of the public using a combination of channels when they complete the official surveys

## Security measures

The institute has established optimal mechanisms to ensure the integrity and privacy of the information transmitted via this new communication channel. All HTTPS communication between the APP and the institute's servers is performed by the secure TLS protocol. Multiple mechanisms have been established for identification and subsequent authorisation to access the data:

* User credential and password system.
* Access control flags: periods, cancellations, status, etc.
* Registration of devices used via universal unique identifiers.
* Verification of communications transmitter by ID token encryption and decryption.

## Design

This new communication channel for mobile devices comprises:

* Web services: Responsible for sending and receiving information to/from the APP and updating the information on the institute's databases. All the information gathered from the APP is stored following the standards established by the institute for the rest of its channels, guaranteeing that this information can be accessed through the institute's other channels once it is saved.
* APP: Integrated app capable of authenticating users, managing the questionnaires shown to the app user, controlling questionnaire networks, performing validations and sending all the information gathered to the institute's servers. The App is capable of working without a permanent internet connection; the connection is only necessary for initial authentication and to send the information.

The App consists of the following main parts, designed to improve user experience:

* Authentication.
* Status summary where dates, completion percentages, etc. can be consulted.
* Navigable questionnaire map that allows you to view the various completed and unanswered questionnaire modules, with the option of returning to completed sections to make corrections.
* Questionnaire. Adapted to mobile devices to improve user experience, taking smaller screen size into consideration and making full use of the advantages of touch screens: questions are displayed individually with large texts and action and selection buttons; enabling actions, for example "next", after verifying that said action is possible following user interaction.
* Summaries of the questionnaire modules to allow all the questions and answers given for the corresponding module to be viewed on one screen, displaying interactive messages to validate the information and allowing access to correct errors.

# Results

In the full paper we will explain the results obtained in the data collection for the “Demographic” survey that begins in late November 2020 and ends in February 2021. The data that will be analysed and presented are the same as those analysed in the “Time Budgets” survey in 2019, i.e. for the various possible channels, the number of responses given, the time taken to complete the information and the quality of the data. As we do not have these data at the time of writing this abstract, we have included, by way of example, the analyses performed with the EPT survey - Time budgets for 2019 - and the results of the demographic survey will be described in the Full paper.

As the new channel was introduced as an experiment, it was not widely publicised and only a simple mention was included in the cover letter that the Institute sent to all the people who were selected to complete the survey, so that if the data obtained via this channel were not proportional to the other channels, there would be no bias or distortion of the overall results.

|  |  |
| --- | --- |
|  |  |

Figure 1. Graphical analysis of the data

Without suggesting that the analysed data is representative at this point of the survey, we can summarise the following observations:

The web channel has increased in comparison with the data collected in 2013, but the optimal percentage to be attained is 20-25%, which we are still far from achieving. It should also be noted that the data that must be completed are not traditional questions with different response options, but rather data that need to be entered with particular care.

The average number of activities provided via the four channels it is practically identical. It should be noted that with the self-completion methods, web questionnaire and app, they are only obliged to introduce a minimum of 10 activities and a recommended total of 15; however, the same data were obtained as with the methods supervised by interviewers. This means that self-completion is not generating bias as regards this variable.

In relation to the method that has the maximum and minimum number of activities entered, in both phase 1 and 2 the telephone and internet methods have similar ratios and stand out above the other two methods.

As regards the average age, it is clear that the respondents using self-completion methods are younger than those using traditional ones with the help of an interviewer. But, if we apply the "average age", the ages are similar, except for the visit method which is used for older citizens.

With regard to the minimum age, it is noteworthy that it is higher for the app, 14, compared to 10 for the other methods. By way of explanation we can suggest that children under 13 or 14 do not generally have mobile phones and they probably use the other methods with help from their parents.

Finally, the data on the duration variable are very significant, i.e. how long it takes to complete the whole survey: When it is the interviewer who fills it in by telephone or in person, the average length is 20 minutes, and yet when it is the citizen or respondent the length doubles in the case of the web questionnaire and almost triples in the case of apps. It can be assumed that this may be due to the fact that when the citizen fills it in directly and without any explanation from an interviewer, as it is the first time they have seen the web questionnaire and/or the app, they have to read the text of both the questions and the answers more slowly to fully understand what they are being asked and find the most appropriate option. In a telephone conversation with an interviewer, direct two-way communication is established helping the respondent to understand both the questions and the answers.